
ON RATES OF CONVERGENCE
IN THE PROBABILISTIC ANALYSIS
OF ALGORITHMS

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 12
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Jasmin Straub
aus Frankfurt am Main

Frankfurt am Main 2021
(D30)

vom Fachbereich 12 der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr.-Ing. Lars Hedrich

Gutachter: Prof. Dr. Ralph Neininger
Univ.-Prof. Dipl.-Ing. Dr. techn. Michael Drmota

Datum der Disputation: 02. August 2021

Contents

- 1 Introduction** **1**

- 2 Notation and Technical Preliminaries** **9**
 - 2.1 Notation and basic properties 9
 - 2.2 Probability metrics 14
 - 2.2.1 The Wasserstein metrics 15
 - 2.2.2 The Zolotarev metrics 16
 - 2.3 The contraction method 18

- 3 General Theorems for Convergence Rates** **21**
 - 3.1 Results 21
 - 3.1.1 A general transfer theorem for rates of convergence 22
 - 3.1.2 An improved transfer theorem for normal limit distributions 27
 - 3.2 Expansions of moments 33

- 4 Applications** **37**
 - 4.1 Some known results 37
 - 4.1.1 The number of key comparisons used by Quicksort 37
 - 4.1.2 Maxima in right triangles 40
 - 4.1.3 The size of m -ary search trees 46
 - 4.1.4 The number of leaves in d -dimensional quadtrees 49
 - 4.2 Periodic functions in mean and variance 51
 - 4.3 Local counters in binary search trees 56
 - 4.4 Convergence rates for Quickselect 63
 - 4.4.1 Quickselect for finding the smallest element 65
 - 4.4.2 Quickselect for finding a uniformly chosen element 69

4.4.3	Quickselect for finding the smallest element—a further approach . . .	72
5	A General Theorem with Relaxed Independence Conditions	79
5.1	A general theorem for convergence rates	80
5.2	A more convenient version of the general theorem	86
5.3	Applications	89
5.3.1	Refined Quicksort asymptotics	90
5.3.2	Importance sampling for estimating the number of Fibonacci matchings	92
	Concluding Remarks	97
	Bibliography	99
	List of Illustrations	107
	Deutsche Zusammenfassung	109

Acknowledgements

First of all, I would like to express my special thanks to my advisor Ralph Neininger whose courses awakened my interest in probability theory and introduced me into the field of the probabilistic analysis of algorithms. Thank you, Ralph, for your inspiring lectures, for your continuous support, guidance and feedback, for always taking the time to answer my questions and for giving me the opportunity to attend various conferences, workshops and summer schools.

Secondly, I would like to thank Michael Drmota from TU Wien for agreeing to read and review this thesis.

During the last years, various people from the university of Frankfurt helped me with mathematical, organizational and technical questions. Hence, special thanks go to the entire working group “Stochastik”. In particular, I thank Götz Kersting, Nicola Kistler and Anton Wakolbinger for their interesting and instructive lectures in the field of probability theory. Furthermore, I am very thankful to my PhD friends and colleagues for their support and company over the last nine years and for the enjoyable lunch and coffee breaks. Special thanks also go to Anna for numerous helpful discussions concerning mathematical and non-mathematical issues since the beginning of our studies.

Finally, I am grateful to my family and friends, especially to my parents, Kristina, and Jan, for their constant support and encouragement throughout my studies.

Chapter 1

Introduction

Recursive structures and self-similar objects arise in many areas of science: In the form of snow flakes, sunflowers, Romanesco broccoli and fern leaves, the phenomenon of self-similarity occurs numerously in nature. Simultaneously, it plays an important role in mathematics and computer science. In mathematics, for instance, popular objects exhibiting the phenomenon of self-similarity are the Koch snowflake or the Sierpiński triangle as well as various stochastic processes—just consider the Brownian motion and its scaling invariance or Galton–Watson trees, where the subtrees of the root are themselves independent Galton–Watson trees. In the field of computer science, a common technique is to divide a problem into several smaller problems of the same nature and to solve the original problem by combining the solutions of the smaller problems (*divide-and-conquer* method). An approach of this kind naturally involves recursive procedures and data structures. Examples range from popular sorting algorithms such as Quicksort and Mergesort to associated data structures such as binary search trees and related objects.

Over the last half century, the analysis of such recursive algorithms and data structures has received increasing attention. Generally speaking, the aim is to precisely analyze a certain quantity of interest (for the moment, just think of the number of key comparisons used by the sorting algorithm Quicksort, for instance). Apart from the analysis of the best-case and worst-case performance of algorithms, it is reasonable to investigate their performance on “typical inputs” of a given size. For this, a common approach is to think of the input as being randomly selected according to some specified probability distribution on the set of all possible inputs. By doing so, the quantity we are interested in becomes a random variable and will henceforth be denoted by Y_n (with n indicating the size of the input). As a concrete example, one can still think of Quicksort and consider Y_n to be the number of key comparisons used by this sorting algorithm, now with the additional assumption that the input consists of n randomly permuted distinct items. Instead of assuming a random input, it is also possible to randomize the algorithm itself, e.g., in the case of Quicksort, we can choose the pivot uniformly at random at each step. Either way, one can then use different analytic and stochastic methods to analyze the (random) quantity Y_n based on the chosen probabilistic model. In this context, the following four questions might be of

interest:

- (1) What is the expected value $\mathbb{E}[Y_n]$ of the quantity of interest? [average-case analysis]
- (2) Can we determine the variance $\text{Var}(Y_n)$? [dispersion analysis]
- (3) How can we quantify the probability that Y_n deviates substantially from its mean? [tail estimates, large deviations]
- (4) Can we normalize the sequence $(Y_n)_{n \geq 0}$ appropriately such that it converges (in distribution) to some limiting random variable? [distributional convergence analysis]

Of course, there is no universal approach to treat these issues, but there are several methods that can be useful to answer these questions. The approaches to address question (4) include, to name only a few, analytic methods using technical tools such as generating functions (see, e.g., Drmota [19], Flajolet and Sedgewick [29], Knuth [56] and Szpankowski [90]), martingale methods (see, e.g., Chauvin, Klein, Marckert and Rouault [8] and Régnier [77]) and the method of moments (see, e.g., Chern and Hwang [12]).

Another approach to determine the asymptotic behavior of a given sequence $(Y_n)_{n \geq 0}$ of recursive random variables is the so-called *contraction method* which was introduced by Rösler [80] in 1991 for the analysis of Quicksort and then further developed by Rösler [81] and Rachev and Rüschemdorf [76], see also Rösler and Rüschemdorf [83]. The name of the method refers to the fact that the limiting distribution is characterized as the fixed-point of a contracting map of measures. A general limit theorem for recursive algorithms and combinatorial structures based on the contraction method as well as numerous applications can be found in Neininger and Rüschemdorf [70].

A further step in the analysis of the asymptotic behavior of Y_n is to determine the *speed of convergence* towards its limiting distribution. This enables us to use the limit theorems for approximations with quantitative error bounds. Hence, another interesting issue going beyond question (4) is:

- (5) Can we bound the distance between the distribution of the normalized quantities and their limiting distribution in some appropriate metric? [rate of convergence]

Convergence rates can be bounded in various metrics and by various techniques. Prominent examples are the use of Fourier transforms or generating functions in combination with Berry–Esseen type inequalities formalizing the fact that two probability distributions are close if their transforms are (see, e.g., Feller [23, Chapter XVI], Petrov [74, Chapter V] and Flajolet and Sedgewick [29, Chapter IX]). This typically yields bounds measured by the Kolmogorov (also called uniform) metric. A related approach to obtain bounds in the Kolmogorov metric is based on the calculation of moments. In the context of the probabilistic analysis of algorithms, the systematic use of the method of moments to determine the speed of convergence was developed by Hwang [43] to study the convergence rate for the space requirement of random m -ary search trees. The strategy uses a refined

method of moments in combination with some inductive arguments, Fourier analysis and the Berry–Esseen smoothing inequality and can be seen as an extension of the method of moments used in Chern and Hwang [12] to show asymptotic normality. This refined method of moments is also applicable to other recursive random variables and was, for instance, used by Bai, Hwang and Tsai [4] to obtain a Berry–Esseen bound for the number of maxima in right triangles. Another powerful tool to prove asymptotic normality including convergence rates for a certain class of random combinatorial structures is Hwang’s *quasi power theorem*, see Hwang [41, 42] for the formulation of this theorem as well as many applications and Flajolet and Sedgewick [29, Section IX.5] for further applications. Although originally formulated for the one-dimensional case, Hwang’s quasi power theorem can be generalized to higher dimensions, still including a result on the speed of convergence (see Heuberger [35] for the two-dimensional case and Heuberger and Kropf [36] for higher dimensions).

Another common (but technically different) approach to obtain bounds on the distance between probability distributions is *Stein’s method*. Introduced by Charles Stein in the seminal paper [88], this method was originally formulated for normal limits, but can be modified to also obtain approximation results for non-normal distributions such as the Poisson distribution, see Chen [9]. For an overview of Stein’s method, we refer to Barbour, Holst and Janson [5], Chen, Goldstein and Shao [10] and Ross [84]. Since the bounds are expressed within metrics defined by using a set of test functions, typical metrics include the Kolmogorov metric, the Wasserstein metric or the total variation metric. Concrete examples of how to apply Stein’s method in the probabilistic analysis of random combinatorial structures can be found, e.g., in Janson, Łuczak and Ruciński [50] for random graphs, in Devroye [17] for random binary search trees and in Bai, Devroye, Hwang and Tsai [2], where a rate of convergence for the number of maxima in hypercubes is derived.

Although the contraction method was originally developed to show weak convergence, it is additionally possible to use ideas of this method to derive concrete bounds on the speed of convergence in some appropriate metric. An example of this procedure is given by Fill and Janson [24] who showed that the rate of convergence for the normalized number of key comparisons used by Quicksort is of order $O(n^{-1/2})$ in the Wasserstein ℓ_p metrics ($p \geq 1$) and of order $O(n^{-1/2+\epsilon})$ in the Kolmogorov metric. In the Zolotarev metric ζ_3 , the rate of convergence for the number of key comparisons was identified to be of order $\Theta(\log n/n)$ by Neininger and Rüschemdorf [69]. Further examples where ideas of the contraction method are used to obtain results on the rate of convergence in the Zolotarev metrics are given in Neininger and Rüschemdorf [71] and Fuchs, Hwang and Neininger [30].

Against this background, the subject of this thesis is to use ideas of the contraction method to develop general convergence theorems quantifying the convergence of appropriately normalized quantities towards their limiting distributions. The combinatorial structures considered are all of a recursive nature and cover various models of random trees, decomposable structures from stochastic geometry, sampling methods and complexities of recursive algorithms under either a random input or some randomization within the algorithm. As varied as the applications may be, they all have one thing in common: Their recursive

nature—which can also be seen as a self-similarity of the structures—provides a recurrence relation for the distributions of the quantities of interest. Such a recurrence is the starting point of our analysis. More specifically, from now on, we always consider a sequence $(Y_n)_{n \geq 0}$ of random vectors in \mathbb{R}^d ($d \in \mathbb{N}$) satisfying the distributional recursion

$$Y_n \stackrel{d}{=} \sum_{r=1}^K A_r(n) Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq n_0, \quad (1.1)$$

where $\stackrel{d}{=}$ denotes equality in distribution, $n_0 \in \mathbb{N}$, the coefficients $A_1(n), \dots, A_K(n)$ are random $(d \times d)$ -matrices and b_n is a d -dimensional random vector (the so-called toll term). The underlying idea of this recursion is that the original problem of size n can be split into K smaller subproblems of sizes $I_1^{(n)}, \dots, I_K^{(n)}$, the toll term b_n indicating the “cost” of this division and the merger. Concerning the number of subproblems and the subproblem sizes, we will always make the following assumptions:

- ◇ The number K of subproblems is a fixed integer $K \geq 1$. However, extensions to K being random and depending on n are possible.
- ◇ The vector $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$ of the subproblem sizes is a random vector in $\{0, \dots, n\}^K$.

Another integral part of our model is the assumption that the subproblems are of the same nature as the original problem, or formally:

$$(Y_n^{(r)})_{n \geq 0} \stackrel{d}{=} (Y_n)_{n \geq 0} \quad \text{for } r = 1, \dots, K. \quad (1.2)$$

Since this assumption guarantees the self-similarity between the initial structure and the parts into which the structure is decomposed, we will use the term *self-similarity condition* when referring to condition (1.2). Furthermore, we need some *conditional independence condition* ensuring that given the subproblem sizes, the subproblems behave independently. To be more precise, we assume that

$$(A_1(n), \dots, A_K(n), b_n, I^{(n)}), (Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0} \quad \text{are independent.} \quad (1.3)$$

Note, however, that dependencies between the coefficients $A_1(n), \dots, A_K(n), b_n$ and the subproblem sizes $I_1^{(n)}, \dots, I_K^{(n)}$ are allowed. Whenever we say that a sequence $(Y_n)_{n \geq 0}$ satisfies recursion (1.1), this includes all above-mentioned assumptions, from the recursion (1.1) itself to the self-similarity and conditional independence conditions (1.2) and (1.3).

As already mentioned, recurrences of this form come up in various fields, see Rösler and Rüschemdorf [83] and Neininger and Rüschemdorf [70] for many concrete examples. Just to name a few, possible applications range from complexity measures of recursive algorithms (e.g., the number of key comparisons used by Quicksort, Mergesort or Quickselect) to parameters of random trees (e.g., the size of tries and m -ary search trees, path lengths

in digital search trees, (PATRICIA) tries and m -ary search trees or the number of leaves in quadrees) to quantities of stochastic geometry (e.g., the number of maxima in right triangles). For all these examples, the contraction method can be used to derive distributional limit laws. However, it turns out that in some special cases of normal limits, the contraction method can be applied although the conditional independence condition (1.3) is not satisfied, see Neininger [68] and Müller and Neininger [66]. In order to also cover similar applications, we formulate a slightly weakened independence condition:

$$(A_1(n), \dots, A_K(n), I^{(n)}), (Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0} \quad \text{are independent.} \quad (1.4)$$

Note that, in contrast to the conditional independence condition (1.3), we allow dependencies between b_n and $(Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}$ here. Thus, condition (1.4) is slightly weaker than condition (1.3) and will be referred to as *partial conditional independence condition* in the following.

Up to the relaxed independence assumption (1.4), this is the framework of Neininger and Rüschendorf [70], where some general convergence results are shown for appropriate normalizations of the Y_n . The content of this thesis is to additionally study the rates of convergence in such general limit theorems. For this, we define the normalized sequence $(X_n)_{n \geq 0}$ by

$$X_n := C_n^{-1/2}(Y_n - M_n), \quad n \geq 0, \quad (1.5)$$

where M_n is a d -dimensional vector and C_n a positive definite $(d \times d)$ -matrix. Essentially, we choose M_n as the mean vector and C_n as the covariance matrix of Y_n if they exist or as the leading order terms in expansions of these moments as $n \rightarrow \infty$. The normalized quantities satisfy the following modified recursion:

$$X_n \stackrel{d}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \quad (1.6)$$

with

$$A_r^{(n)} := C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2} \left(b_n - M_n + \sum_{r=1}^K A_r(n) M_{I_r^{(n)}} \right) \quad (1.7)$$

and self-similarity and independence conditions as above.

In the context of the contraction method, the aim is to establish theorems that allow us to transfer the convergence of the coefficients to convergence in distribution of the quantities X_n to a limit X . In a nutshell, the procedure is as follows: First, one verifies that the coefficients given in (1.7) converge in an appropriate sense:

$$A_r^{(n)} \rightarrow A_r^*, \quad r = 1, \dots, K \quad \text{and} \quad b^{(n)} \rightarrow b^*, \quad n \rightarrow \infty.$$

If with $n \rightarrow \infty$, also the subproblem sizes $I_r^{(n)}$ become large, then we expect a potential limit X of X_n to satisfy the following distributional fixed-point equation which is obtained from (1.6) by letting formally $n \rightarrow \infty$:

$$X \stackrel{d}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*. \quad (1.8)$$

Here, $(A_1^*, \dots, A_K^*, b^*), X^{(1)}, \dots, X^{(K)}$ are independent and $X^{(r)}$ has the same distribution as X for $r = 1, \dots, K$. The idea then is to use the distributional fixed-point equation (1.8) to characterize the limiting distribution $\mathcal{L}(X)$ and to show convergence of X_n to this fixed-point, see Rösler and Rüschemdorf [83] and Neininger and Rüschemdorf [70] for details.

The aim of this thesis is to endow such general transfer theorems with bounds on the rates of convergence and to give specific bounds on rates of convergences for many concrete examples. Simply put, we are interested in statements of the following form: Given that the rate of convergence of the coefficients is of order $O(R(n))$, i.e.,

$$\sum_{r=1}^K \|A_r^{(n)} - A_r^*\|_s + \|b^{(n)} - b^*\|_s = O(R(n)),$$

where $\|\cdot\|_s$ denotes the usual L_s norm and $(R(n))_{n \geq 0}$ is a sequence tending to zero, can we find conditions that allow us to transfer this rate to a rate of convergence for the quantities X_n ?

As a distance measure between the distribution of X_n and the limiting distribution $\mathcal{L}(X)$, different probability metrics are possible. The approach we are following here is the use of probability metrics which are useful to show convergence within the framework of the contraction method. One class of metrics, the Zolotarev metrics, has proven to be particularly suitable and flexible in this context. Compared to, say, the Kolmogorov metric, the Zolotarev metrics have a sort of smoothing effect which becomes stronger as the index of the metric increases and which may smooth out certain phenomena resulting in a less sharp distance measure. However, the Zolotarev metrics can be considered as a universal tool to show weak convergence and enable us to provide bounds on the rate of convergence on a quite general level which are easy to apply and which cover a broad range of applications at once.

The organization of this thesis is as follows: In Chapter 2, the technical foundations for the general transfer theorems are laid by introducing some notation and by giving a brief overview of the different probability metrics we work with as well as the main ideas of the contraction method.

The following Chapter 3 deals with the derivation of general convergence theorems to transfer the convergence rates of the coefficients to a bound on the rate of the convergence $X_n \rightarrow X$ in the Zolotarev metrics. We start with a general transfer theorem (Theorem 3.1) and then present a refined version of this theorem (Theorem 3.4) which can be applied for

a large class of quantities with normal limits. Numerous examples of how to use those theorems in concrete applications are given in Chapter 4.

In the last part, Chapter 5, we derive a theorem similar to the refined transfer theorem in Chapter 3, with the difference that the conditional independence condition (1.3) is replaced by the (weaker) partial conditional independence condition (1.4) there and give further examples.

In the concluding remarks, we summarize the benefits and difficulties of the methodology developed in the present thesis and give some directions for further research.

Chapter 2

Notation and Technical Preliminaries

In this chapter, we introduce the notation and give a short overview of the technical basics and concepts of the contraction method. First, we present some basic notions and recall some essential facts on matrices and on random variables in Section 2.1. In the subsequent section, we introduce the spaces of distributions and the different probability metrics we work with (Section 2.2) and, lastly, describe the main ideas of the contraction method in Section 2.3.

2.1 Notation and basic properties

The following notation and conventions are used throughout this thesis:

- ◇ We denote by \mathbb{N} the set $\{1, 2, \dots\}$ of positive integers, by \mathbb{N}_0 the set $\{0, 1, 2, \dots\}$ of non-negative integers and by \mathbb{R} the set of real numbers.
- ◇ Unless otherwise specified, we use the convention $0 \cdot \infty = 0$ and $c/0 = \infty$ for any $c > 0$.

- ◇ As is common, we write

$$\mathbf{1}_{\{\text{expression}\}} = \begin{cases} 1, & \text{if expression is true,} \\ 0, & \text{if expression is false.} \end{cases}$$

- ◇ With $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$, we denote the minimum and the maximum of two real numbers x and y .
- ◇ Furthermore, the floor and ceiling functions are denoted by $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ and the absolute value function by $|\cdot|$.

We use the common asymptotic notation, i.e., for functions $f, g : \mathbb{N}_0 \rightarrow \mathbb{R}$, we write

- ◇ $f(n) = O(g(n))$ if there exist $C > 0$ and $n_0 \in \mathbb{N}$ such that $|f(n)| \leq C|g(n)|$ for all $n \geq n_0$,
- ◇ $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$,
- ◇ $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

If $g(n) \neq 0$ for n large enough, we further write

- ◇ $f(n) = o(g(n))$ if $f(n)/g(n) \rightarrow 0$ as $n \rightarrow \infty$,
- ◇ $f(n) \sim g(n)$ if $f(n)/g(n) \rightarrow 1$ as $n \rightarrow \infty$.

We always denote by “log” the natural logarithm to base e and set

$$\mathcal{H}_n = \sum_{k=1}^n \frac{1}{k} = \log(n) + \gamma + O\left(\frac{1}{n}\right),$$

where $\gamma = 0.5772156649\dots$ is the *Euler–Mascheroni constant*.

Since we only consider matrices with real-valued components in this thesis, the following definitions and properties refer to matrices in $\mathbb{R}^{d \times d'}$ ($d, d' \in \mathbb{N}$). We will use the following common notation:

- ◇ The identity matrix of size d is denoted by Id_d .
- ◇ For a matrix $A \in \mathbb{R}^{d \times d'}$, we denote by $A^T \in \mathbb{R}^{d' \times d}$ its transpose and say that A is *symmetric* if $A^T = A$.
- ◇ We call a square matrix A *regular* (also: *invertible*) if its determinant is non-zero and denote by A^{-1} its *inverse*.
- ◇ A square matrix is called *diagonal* if its off-diagonal elements are zero. Moreover, we denote by $\text{diag}(a_1, \dots, a_d)$ the diagonal matrix with diagonal elements $a_1, \dots, a_d \in \mathbb{R}$.

Concerning vector and matrix norms, we denote by $\|\cdot\|$ the Euclidean norm on \mathbb{R}^d and by $\|\cdot\|_{\text{op}}$ the corresponding operator norm (*spectral norm*) given by

$$\|A\|_{\text{op}} := \sup_{x \in \mathbb{R}^d, x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|$$

for any $(d \times d)$ -matrix A . It follows directly from this definition that for any square matrix $A \in \mathbb{R}^{d \times d}$ and $x \in \mathbb{R}^d$, we have $\|Ax\| \leq \|A\|_{\text{op}}\|x\|$. Furthermore, the spectral norm is submultiplicative, i.e., $\|AB\|_{\text{op}} \leq \|A\|_{\text{op}}\|B\|_{\text{op}}$ for any $(d \times d)$ -matrices A and B and

invariant under unitary (orthogonal) transformations. Apart from these properties, we will make use of the fact that for a symmetric matrix A , we have

$$\|A\|_{\text{op}} = \max\{|\lambda| : \lambda \text{ eigenvalue of } A\}.$$

Another matrix norm we will use is the *Frobenius norm* given by

$$\|A\|_{\text{F}} := \left(\sum_{i,j=1}^d A_{ij}^2 \right)^{1/2}$$

for every $A = (A_{ij})_{i,j=1,\dots,d} \in \mathbb{R}^{d \times d}$. Just as the spectral norm, the Frobenius norm is submultiplicative and unitarily invariant.

We say that a symmetric matrix A is *positive definite* (*positive semidefinite*, respectively) if $x^T A x > 0$ ($x^T A x \geq 0$, respectively) for any $x \neq 0$. Note that this definition includes the symmetry of A , i.e., when talking about a positive (semi)definite matrix, symmetry is always implied. We now summarize some important properties of positive (semi)definite matrices. One well-known fact is that positive (semi)definiteness can be characterized in terms of the eigenvalues: A symmetric matrix A is positive (semi)definite if and only if its eigenvalues are positive (non-negative). From this, it follows immediately that

- (1) a positive definite matrix is invertible and
- (2) we can declare the square root of a positive (semi)definite matrix. For this, let A be a positive (semi)definite matrix with eigenvalues $\lambda_1 \geq \dots \geq \lambda_d \geq 0$ and spectral decomposition $A = O \text{diag}(\lambda_1, \dots, \lambda_d) O^T$. Then, the matrix

$$A^{1/2} = O \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}) O^T$$

is positive (semi)definite and satisfies $A^{1/2} A^{1/2} = A$. Indeed, it can be shown that this is the unique positive (semi)definite matrix with this property which is why $A^{1/2}$ is called the (*principal*) *square root* of A .

Both the matrix inversion and the matrix square root are continuous functions on the space of positive definite matrices, i.e., sufficiently small perturbations in the input matrix result in small changes in the output matrix. In the course of this thesis, the following question arises at several points: Given a positive definite matrix A and a “small” matrix M_n with entries of order $O(R(n))$, what is the distance between $(A + M_n)^{-1/2}$ and $A^{-1/2}$ (measured in some arbitrary matrix norm, e.g., in the Frobenius norm)? This question is addressed by the following two lemmas.

Lemma 2.1 ([89, Corollary 4.19]). *Let A be a regular $(d \times d)$ -matrix and M_n a $(d \times d)$ -matrix with Frobenius norm $\|M_n\|_{\text{F}} = O(R(n))$ for some sequence $R(n) \rightarrow 0$ as $n \rightarrow \infty$. Then we have, as $n \rightarrow \infty$,*

$$\|(A + M_n)^{-1} - A^{-1}\|_{\text{F}} = O(R(n)).$$

Remark 2.2. Since any two matrix norms are equivalent, the Frobenius norm can be replaced by any other matrix norm. Furthermore, if finer approximations to $(A + M_n)^{-1}$ are needed, we can use a generalization of the geometric series called the Neumann series (see, e.g., Stewart [89, Theorem 4.20]), which implies for n large enough

$$(A + M_n)^{-1} = A^{-1}(\text{Id}_d + M_n A^{-1})^{-1} = A^{-1} \sum_{j=0}^{\infty} (-M_n A^{-1})^j.$$

From this representation, we can deduce more precise expansions than stated in Lemma 2.1, e.g., we have, as $n \rightarrow \infty$,

$$(A + M_n)^{-1} = A^{-1} - A^{-1} M_n A^{-1} + O(R^2(n)),$$

where the $O(R^2(n))$ -term can be read componentwise (or in any matrix norm).

We now formulate a similar result for the matrix square root.

Lemma 2.3. *Let A be a positive definite $(d \times d)$ -matrix and M_n a symmetric $(d \times d)$ -matrix with Frobenius norm $\|M_n\|_F = O(R(n))$ for some sequence $R(n) \rightarrow 0$ as $n \rightarrow \infty$. Then we have, as $n \rightarrow \infty$,*

$$\|(A + M_n)^{1/2} - A^{1/2}\|_F = O(R(n)).$$

Proof. As A is positive definite, we can choose n large enough such that also $A + M_n$ is positive definite and thus has a unique principal (positive definite) square root $(A + M_n)^{1/2}$. We now use a bound for the difference between the principal square roots of two positive definite matrices which can be found in Higham [37, Theorem 6.2] (a special case of a result of van Hemmen and Ando [91, Proposition 3.2]). This theorem states for any positive definite real matrices B and C that

$$\|B^{1/2} - C^{1/2}\| \leq (\lambda_{\min}(B)^{1/2} + \lambda_{\min}(C)^{1/2})^{-1} \|B - C\|$$

for any unitarily invariant norm, where we denote by λ_{\min} the smallest eigenvalue. Since the Frobenius norm is invariant under unitary operations, we have

$$\begin{aligned} \|(A + M_n)^{1/2} - A^{1/2}\|_F &\leq (\lambda_{\min}(A)^{1/2} + \lambda_{\min}(A + M_n)^{1/2})^{-1} \|M_n\|_F \\ &\leq \lambda_{\min}(A)^{-1/2} \|M_n\|_F = O(R(n)), \end{aligned}$$

which is the assertion. □

Remark 2.4. Combining Lemmas 2.1 and 2.3, we conclude: Let $(R(n))_{n \geq 0}$ be an arbitrary sequence converging to zero, A a positive definite matrix and M_n a symmetric matrix with entries of order $O(R(n))$. Then we have, as $n \rightarrow \infty$,

$$(A + M_n)^{-1/2} = A^{-1/2} + O(R(n)),$$

where the $O(R(n))$ -term can be read componentwise (or in any matrix norm).

In the remainder of this section, we fix some probability notation:

- ◇ For any random variable X , we denote by $\mathcal{L}(X)$ its distribution and write $X \stackrel{d}{=} Y$ if $\mathcal{L}(X) = \mathcal{L}(Y)$.
- ◇ Convergence in distribution is denoted by \xrightarrow{d} .
- ◇ For a random vector X in \mathbb{R}^d , we denote by $\mathbb{E}[X]$ its mean vector, by $\text{Cov}(X)$ its covariance matrix and for $d = 1$ by $\text{Var}(X)$ its variance.

Furthermore, the following notation is used: For a random vector X and some $0 < p < \infty$, we set

$$\|X\|_p := \mathbb{E}[\|X\|^p]^{(1/p)\wedge 1}$$

and for a random $(d \times d)$ -matrix A and $0 < p < \infty$, we define

$$\|A\|_p := \mathbb{E}[\|A\|_{\text{op}}^p]^{(1/p)\wedge 1}.$$

With this in mind, we call a random vector X (a random matrix A , respectively) L_p -integrable if $\|X\|_p < \infty$ ($\|A\|_p < \infty$, respectively). For $0 < p < \infty$, a sequence $(X_n)_{n \geq 0}$ of random vectors is said to converge in L_p to a random vector X if $\|X_n - X\|_p \rightarrow 0$ as $n \rightarrow \infty$. Similarly, L_p -convergence for random matrices is defined.

The following univariate distributions are needed in the course of this thesis:

- ◇ $\text{unif}\{m_1, \dots, m_n\}$: the *discrete uniform distribution* on the finite set $\{m_1, \dots, m_n\}$,
- ◇ $\text{unif}[a, b]$: the *continuous uniform distribution* on the interval $[a, b]$ ($a < b$),
- ◇ $\text{Bin}(n, p)$: the *binomial distribution* with n trials and success probability $p \in [0, 1]$,
- ◇ $\text{Mult}(n, p_1, \dots, p_k)$: the *multinomial distribution* with n trials and (non-negative) success probabilities p_1, \dots, p_k (with $p_1 + \dots + p_k = 1$),
- ◇ $\text{Hyp}(N, n, K)$: the *hypergeometric distribution* with n trials, population size N and K elements in the population that are classified as successes.

Furthermore, we will need the following well-known facts about the multivariate normal distribution:

- ◇ For some d -dimensional vector μ and a positive semidefinite $(d \times d)$ -matrix Σ , we denote by $\mathcal{N}(\mu, \Sigma)$ the *d -dimensional normal distribution* with mean vector μ and covariance matrix Σ , i.e., the probability distribution on \mathbb{R}^d with characteristic function

$$\varphi(t) = \exp\left(it^T \mu - \frac{1}{2}t^T \Sigma t\right). \quad (2.1)$$

The normal distribution $\mathcal{N}(0, \text{Id}_d)$ with mean vector $0 \in \mathbb{R}^d$ and covariance matrix Id_d is called the *d -dimensional standard normal distribution*.

- ◇ Let μ be an arbitrary d -dimensional vector and Σ a positive semidefinite $(d \times d)$ -matrix. If N has the d -dimensional standard normal distribution and if A is a $(d \times d)$ -matrix with $AA^T = \Sigma$ (this is, for instance, true if A is chosen as the square root of Σ), then the random vector $AN + \mu$ has the $\mathcal{N}(\mu, \Sigma)$ -distribution.
- ◇ We have the following *convolution property of the multivariate normal distribution*: Let X and Y be independent random vectors with distributions $\mathcal{N}(\mu_X, \Sigma_X)$ and $\mathcal{N}(\mu_Y, \Sigma_Y)$. Then, it follows directly from (2.1) that $X + Y$ has the normal distribution $\mathcal{N}(\mu_X + \mu_Y, \Sigma_X + \Sigma_Y)$.

2.2 Probability metrics

When studying rates of convergence for different random quantities towards their limits, an essential issue is the choice of the probability metric. One common probability metric on \mathbb{R} is the *Kolmogorov metric* (also: *uniform metric*) given by

$$\rho(\mu, \nu) := \sup_{x \in \mathbb{R}} |F_\mu(x) - F_\nu(x)|$$

for probability measures μ and ν on \mathbb{R} with distribution functions F_μ and F_ν . In the course of this thesis, two further probability metrics will be used, namely, the Wasserstein metrics and the Zolotarev metrics. The former metrics will be introduced in Section 2.2.1, the latter in Section 2.2.2. Since our theorems are formulated for random vectors in \mathbb{R}^d , the definitions in these sections are given for probability measures on \mathbb{R}^d (endowed with the usual Euclidean norm and the Borel σ -field), although the Wasserstein distance and the Zolotarev distance can be defined on more general spaces. For additional information on the metrics, we refer to Ambrosio, Gigli and Savaré [1], Dudley [22] and Villani [93] for the Wasserstein metrics and to Janson and Kaijser [49] for the Zolotarev metrics.

Before coming to the definition of the metrics, we introduce the spaces of probability measures we work with: Throughout this thesis, we denote by \mathcal{P}^d the space of all probability measures on \mathbb{R}^d and, for $0 < s < \infty$, by

$$\mathcal{P}_s^d := \{\mathcal{L}(X) \in \mathcal{P}^d : \|X\|_s < \infty\}$$

the subspace consisting of those probability measures with a finite moment of order s . Furthermore, we will need the following subspaces of \mathcal{P}_s^d :

$$\begin{aligned} \mathcal{P}_s^d(0) &:= \{\mathcal{L}(X) \in \mathcal{P}_s^d : \mathbb{E}[X] = 0\}, \quad s > 1, \\ \mathcal{P}_s^d(0, \text{Id}_d) &:= \{\mathcal{L}(X) \in \mathcal{P}_s^d : \mathbb{E}[X] = 0, \text{Cov}(X) = \text{Id}_d\}, \quad s > 2. \end{aligned}$$

For reasons of simplicity, we will also use the notation $\mathcal{P}_s^d(0, \text{Id}_d)$ for $s \leq 2$. For this, we use the convention

$$\mathcal{P}_s^d(0, \text{Id}_d) := \begin{cases} \mathcal{P}_s^d, & 0 < s \leq 1, \\ \mathcal{P}_s^d(0), & 1 < s \leq 2. \end{cases}$$

2.2.1 The Wasserstein metrics

In this section, we provide a brief overview of the Wasserstein distances on \mathbb{R}^d , having in mind that the presented results can be extended to more general metric spaces.

Definition 2.5. For all $0 < s < \infty$ and $\mu, \nu \in \mathcal{P}_s^d$, the *Wasserstein distance* ℓ_s between μ and ν is defined by

$$\ell_s(\mu, \nu) := \inf \{ \|W - Z\|_s : \mathcal{L}(W) = \mu, \mathcal{L}(Z) = \nu \}.$$

For L_s -integrable random variables X and Y in \mathbb{R}^d , we set

$$\ell_s(X, Y) := \ell_s(\mathcal{L}(X), \mathcal{L}(Y)).$$

Furthermore, we call a pair (X, Y) of L_s -integrable random variables an *optimal ℓ_s -coupling* of μ and ν if $\mathcal{L}(X) = \mu$, $\mathcal{L}(Y) = \nu$ and $\ell_s(\mu, \nu) = \|X - Y\|_s$.

We now summarize some well-known facts about the Wasserstein distance which will be useful later:

- ◇ It can be shown that the infimum in Definition 2.5 is a minimum, which means that for all $\mu, \nu \in \mathcal{P}_s^d$, there exists an optimal ℓ_s -coupling of μ and ν (see, e.g., Bickel and Freedman [6, Lemma 8.1]). If additionally $d = 1$ and $s \geq 1$, optimal ℓ_s -couplings can easily be constructed with the help of the associated quantile functions and a uniform random variable U on the unit interval: If F_μ^{-1} and F_ν^{-1} are the quantile functions of μ and ν , then the pair $(F_\mu^{-1}(U), F_\nu^{-1}(U))$ is an optimal ℓ_s -coupling of μ and ν (see Dall'Aglio [15] and Major [65, Theorem 8.1]).
- ◇ For any $0 < s < \infty$, the pair $(\mathcal{P}_s^d, \ell_s)$ is a complete metric space and convergence in ℓ_s is equivalent to weak convergence plus convergence of the s -th absolute moments (see, e.g., Bickel and Freedman [6] and Rachev [75]).
- ◇ For $\mu, (\mu_n)_{n \geq 1}$ in \mathcal{P}_s^d , there exist random variables W and W_n such that $\mathcal{L}(W) = \mu$, $\mathcal{L}(W_n) = \mu_n$ and (W_n, W) is an optimal ℓ_s -coupling of μ_n and μ for all $n \geq 1$, see, e.g., Ambrosio, Gigli and Savaré [1, Lemma 5.3.4]. Hence, we have ℓ_s convergence $X_n \rightarrow X$ if and only if there exist random variables W_n and W on a common probability space with $\mathcal{L}(W_n) = \mathcal{L}(X_n)$, $\mathcal{L}(W) = \mathcal{L}(X)$ and $W_n \rightarrow W$ in L_s .
- ◇ Furthermore, from the monotonicity of the L_s norms, we obtain a corresponding monotonicity statement for the Wasserstein metrics. To be more precise, for any $1 \leq p \leq s < \infty$ and an L_s -integrable random variable X , Jensen's inequality implies

$$\|X\|_p = \mathbb{E}[\|X\|^p]^{1/p} = \mathbb{E}[(\|X\|^s)^{p/s}]^{1/p} \leq \mathbb{E}[\|X\|^s]^{1/s} = \|X\|_s.$$

Thus, for $1 \leq p \leq s < \infty$ and $\mu, \nu \in \mathcal{P}_s^d$, we have

$$\ell_p(\mu, \nu) \leq \ell_s(\mu, \nu).$$

Further information on the ℓ_s metrics can be found in Ambrosio, Gigli and Savaré [1], Dudley [22] and Villani [93].

2.2.2 The Zolotarev metrics

The Zolotarev metrics for probability measures in a Banach space were introduced by Zolotarev in [96]. In this section, we define the Zolotarev metrics for probability measures on \mathbb{R}^d and summarize some of their properties.

First of all, it should be noted that the Zolotarev distance ζ_s can be defined for any real number $s > 0$, but only the values $0 < s \leq 3$ are relevant for our purposes. Hence, the following definitions and properties mostly refer to the general case $s > 0$, but some explanations focus on the values $s \leq 3$.

For some $s > 0$, we write $s = m + \alpha$, where $m := \lceil s \rceil - 1 \in \mathbb{N}_0$ and $\alpha := s - m \in (0, 1]$. Based on this decomposition of s , we define a class \mathcal{F}_s of test functions by

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

where $C^m(\mathbb{R}^d, \mathbb{R})$ denotes the set of all m times continuously differentiable functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$ and $D^m f$ the m -th derivative of such f . In particular, we have

- ◇ $D^m f = f$ and $\|\cdot\|_{\circ} = |\cdot|$ for $m = 0$,
- ◇ $D^m f = \nabla^T f$ (the transpose of the gradient of f , i.e., the Jacobian matrix of f) and $\|\cdot\|_{\circ} = \|\cdot\|$ for $m = 1$ and
- ◇ $D^m f = H_f$ (the Hessian matrix of f) and $\|\cdot\|_{\circ} = \|\cdot\|_{\text{op}}$ for $m = 2$.

Now, for a random vector $X = (X_1, \dots, X_d)$ in \mathbb{R}^d , some $k \in \mathbb{N}_0$ and non-negative integers $k_1 + \dots + k_d = k$, the expectation

$$\mathbb{E}[X_1^{k_1} \cdot \dots \cdot X_d^{k_d}]$$

is called a *mixed moment of order k* , i.e., the mixed moments of orders 1 and 2 are $\mathbb{E}[X_i]$, $i = 1, \dots, d$ and $\mathbb{E}[X_i X_j]$, $i, j = 1, \dots, d$.

Furthermore, for $s > 0$ and m chosen as above, we call a pair (X, Y) of L_s -integrable random vectors ζ_s -compatible if X and Y have the same mixed moments up to order m . In other words, for $2 < s \leq 3$, a pair (X, Y) of random vectors is ζ_s -compatible if both X and Y are L_s -integrable and if X and Y have identical mean vectors and identical covariance matrices. For $1 < s \leq 2$, the assumption of identical covariance matrices can be dropped and for $0 < s \leq 1$, we only need the L_s -integrability of X and Y to ensure ζ_s -compatibility. With this in mind, we can now define the Zolotarev distance. Unlike Zolotarev, we only define the Zolotarev distance ζ_s between ζ_s -compatible random vectors to a priori ensure that it is well-defined and finite.

Definition 2.6. The *Zolotarev distance* ζ_s between two ζ_s -compatible random vectors X and Y in \mathbb{R}^d (between $\mathcal{L}(X)$ and $\mathcal{L}(Y)$, respectively) is defined by

$$\zeta_s(X, Y) := \zeta_s(\mathcal{L}(X), \mathcal{L}(Y)) := \sup_{f \in \mathcal{F}_s} |\mathbb{E}[f(X) - f(Y)]|. \quad (2.2)$$

First of all, note that the term on the right hand side of (2.2) is finite and does not depend on the joint distribution of (X, Y) , but only on the marginal distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ if X and Y are ζ_s -compatible. This can be shown by a Taylor expansion of f (see Lemma 2 in Zolotarev [96]). Hence, the Zolotarev distance is a distance between probability distributions. However, just as in the definition of the Wasserstein metrics, it is convenient to also consider it as a distance of random variables, keeping in mind that only the marginal distributions of the random variables matter.

We now summarize some basic and well-known facts concerning the Zolotarev distance ζ_s :

- ◊ For $0 < s \leq 3$, the pair $(\mathcal{P}_s^d(0, \text{Id}_d), \zeta_s)$ is a complete metric space and, furthermore, ζ_s -convergence implies weak convergence (see, e.g., Drmota, Janson and Neininger [20, Theorem 5.1] for a more general version of this statement).
- ◊ The Zolotarev metric ζ_s is $(s, +)$ -ideal (see Zolotarev [96]), i.e., for a ζ_s -compatible pair (X, Y) , an independent and L_s -integrable random variable Z and some real number $c \neq 0$, we have

$$\zeta_s(X + Z, Y + Z) \leq \zeta_s(X, Y), \quad \zeta_s(cX, cY) = |c|^s \zeta_s(X, Y).$$

- ◊ From the $(s, +)$ -ideality, it follows directly (e.g., by induction on n) that for independent X_1, \dots, X_n and independent Y_1, \dots, Y_n such that (X_i, Y_i) is ζ_s -compatible for $i = 1, \dots, n$, we have

$$\zeta_s\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) \leq \sum_{i=1}^n \zeta_s(X_i, Y_i). \quad (2.3)$$

- ◊ Finally, we will need the following bound (see Zolotarev [97]): Let (X, Y) be a pair of ζ_s -compatible d -dimensional random vectors and A a $(d \times d)$ -matrix. Then we have

$$\zeta_s(AX, AY) \leq \|A\|_{\text{op}}^s \zeta_s(X, Y). \quad (2.4)$$

The following theorem by Drmota, Janson and Neininger [20, Lemma 5.7] provides an upper bound of the Zolotarev distance ζ_s by the Wasserstein metric ℓ_s .

Theorem 2.7 ([20, Lemma 5.7]). *Let (X, Y) be a pair of ζ_s -compatible random vectors. For $s > 1$, we have*

$$\zeta_s(X, Y) \leq \left((\mathbb{E}[\|X\|^s])^{1-1/s} + (\mathbb{E}[\|Y\|^s])^{1-1/s} \right) \ell_s(X, Y).$$

For $0 < s \leq 1$, we have

$$\zeta_s(X, Y) \leq \ell_s(X, Y). \quad (2.5)$$

Note that for $0 < s \leq 1$, it follows from the Kantorovich–Rubinstein theorem that the ‘ \leq ’ in (2.5) can be replaced by ‘ $=$ ’, see Kantorovich and Rubinstein [51] (for compact metric spaces) and Dudley [21, 22] (for the generalization to separable metric spaces). Further properties and results concerning the Zolotarev distances can be found in Janson and Kaijser [49], Rachev [75] and Zolotarev [96, 97].

2.3 The contraction method

In this section, we give a brief overview of the contraction method as a tool to show weak convergence of random recursive structures. As already described in the introduction, the starting point when working with the contraction method is a sequence $(Y_n)_{n \geq 0}$ (here: in \mathbb{R}^d) satisfying the distributional recursion (1.1). With the scaling

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

with some d -dimensional vector M_n and some positive definite $(d \times d)$ -matrix C_n , the recursive decomposition (1.1) of Y_n leads to a similar distributional recursion for the scaled quantities:

$$X_n \stackrel{d}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \quad (2.6)$$

with modified coefficients

$$A_r^{(n)} := C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2} \left(b_n - M_n + \sum_{r=1}^K A_r(n) M_{I_r^{(n)}} \right) \quad (2.7)$$

and independence relations and self-similarity conditions as in (1.1). Building on this recursive decomposition of X_n , the further strategy consists of the following key steps:

(1) **Derivation of a distributional fixed-point equation for the limit:**

First, one verifies that the coefficients given in (2.7) converge in an appropriate sense to some limiting coefficients:

$$A_r^{(n)} \rightarrow A_r^*, \quad b^{(n)} \rightarrow b^*.$$

If with $n \rightarrow \infty$, also the subproblem sizes $I_r^{(n)}$ become large, then we expect a potential limit X of X_n to satisfy the following distributional fixed-point equation which is obtained from (2.6) by replacing all appearing quantities by their limits ($n \rightarrow \infty$):

$$X \stackrel{d}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*, \quad (2.8)$$

where $X^{(1)}, \dots, X^{(K)}, (A_1^*, \dots, A_K^*, b^*)$ are independent and $X^{(1)}, \dots, X^{(K)}$ are identically distributed as X .

(2) **Existence and uniqueness of the solution of the fixed-point equation (2.8):**

To this end, the solutions of (2.8) can be considered as the fixed-points of the following map from the space \mathcal{P}^d of probability distributions on \mathbb{R}^d to itself:

$$T : \mathcal{P}^d \rightarrow \mathcal{P}^d, \quad \mu \mapsto \mathcal{L} \left(\sum_{r=1}^K A_r^* Z^{(r)} + b^* \right), \quad (2.9)$$

where $Z^{(1)}, \dots, Z^{(K)}, (A_1^*, \dots, A_K^*, b^*)$ are independent and $Z^{(1)}, \dots, Z^{(K)}$ are identically distributed with distribution μ . Subsequently, we choose an appropriate subspace of \mathcal{P}^d and a complete metric on this subspace such that the map T (restricted to the subspace) has contraction properties. Then, Banach's fixed-point theorem implies the existence and uniqueness of the solution of the distributional equation (2.8) in this subspace.

(3) **Convergence of the scaled quantities X_n to this unique solution:**

After having shown the existence of a unique solution X of equation (2.8) in a suitable subspace, we show convergence of the scaled quantities X_n to this unique solution X in the chosen metric. From this convergence, depending on the chosen metric, further statements (such as weak convergence and convergence of moments) can be derived.

In the light of this procedure, there are several possible choices for the probability metric mentioned in steps (2) and (3), see, e.g., Rachev and Rüschendorf [76]. Early work on the contraction method (see, e.g., Rösler [80, 82] and Neininger [67]) makes use of the Wasserstein ℓ_s distances, mainly the ℓ_2 metric. However, the ℓ_s setting has several limitations. One problem is that it typically cannot be used for an important class of problems that lead to normal limit laws. For instance, assume that the limiting equation (2.8) is of the form

$$X \stackrel{d}{=} \sum_{r=1}^K A_r^* X^{(r)}, \quad (2.10)$$

with $d = 1$ and $(A_1^*)^2 + \dots + (A_K^*)^2 = 1$ almost surely (this holds true for many applications with normal limits). Then, by the convolution property of the normal distribution, every centered normal distribution solves equation (2.10) and the ℓ_s setting does not apply. Instead, recurrences leading to fixed-point equations of the form (2.10) can be studied by use of the ζ_s metrics with $2 < s \leq 3$ (see, e.g., Neininger and Rüschendorf [70] and Rösler and Rüschendorf [83]).

As described above, when working with the contraction method, we have to find conditions under which T has a unique fixed-point in a suitable subspace of \mathcal{P}^d (see step (2) above). The following theorem provides such conditions.

Theorem 2.8 ([70, Corollary 3.4]). *Let T be given as in (2.9) with L_s -integrable coefficients A_1^*, \dots, A_K^*, b^* satisfying $\sum_{r=1}^K \mathbb{E}[\|A_r^*\|_{\text{op}}^s] < 1$ for some $0 < s \leq 3$. Further assume that $\mathbb{E}[b^*] = 0$ if $s > 1$ and additionally $\mathbb{E}[b^*(b^*)^T + \sum_{r=1}^K A_r^*(A_r^*)^T] = \text{Id}_d$ for $s > 2$. Then, T has a unique fixed-point in $\mathcal{P}_s^d(0, \text{Id}_d)$.*

The proof of this theorem can be accomplished by Banach's fixed-point theorem in the following way: First, note that the restriction of the map T to $\mathcal{P}_s^d(0, \text{Id}_d)$ maps into $\mathcal{P}_s^d(0, \text{Id}_d)$ by the assumptions that $\mathbb{E}[b^*] = 0$ if $s > 1$ and additionally $\mathbb{E}[b^*(b^*)^T + \sum_{r=1}^K A_r^*(A_r^*)^T] = \text{Id}_d$

for $s > 2$. Now, for arbitrary $\mu, \nu \in \mathcal{P}_s^d(0, \text{Id}_d)$, we find

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

Together with $\sum_{r=1}^K \mathbb{E}[\|A_r^*\|_{\text{op}}^s] < 1$, this implies that the restriction of T to $\mathcal{P}_s^d(0, \text{Id}_d)$ is a ζ_s -contraction. Since $(\mathcal{P}_s^d(0, \text{Id}_d), \zeta_s)$ is a complete metric space, Banach's fixed-point theorem implies the assertion.

Once having established the existence of a unique solution X of equation (2.8) in a suitable subspace, it remains to study the distance between the sequence X_n and X in the chosen metric (step (3) above). In this context, it is possible to not only show convergence but also to obtain (upper) bounds on the rate of convergence, see Fill and Janson [24], Neininger and Rüschendorf [69, 71], Fuchs, Hwang and Neininger [30] as well as Neininger and Sulzbach [73, Corollary 21] for some examples where ideas of the contraction method are used to obtain rates of convergence in the Wasserstein metrics or in the Zolotarev metrics.

The aim of this thesis is to make use of related ideas in a general framework in order to provide general convergence theorems with upper bounds on the rate of convergence in the Zolotarev metrics ζ_s , $0 < s \leq 3$. This is part of the subsequent chapter.

Chapter 3

General Theorems for Convergence Rates

In this chapter, we develop a general framework to bound convergence rates for sequences of recursive random variables in the Zolotarev distances based on ideas of the contraction method. The chapter is organized as follows: The main theorems are worked out in Section 3.1, whereby we start with a general version in Section 3.1.1 and then give a refined version for normal limits in Section 3.1.2. After having presented these general transfer theorems, we provide an additional lemma which will be useful in many applications where only asymptotic expansions of the moments are known (Section 3.2). The presented results were announced in the extended abstract [72].

3.1 Results

We return to the situation outlined in the introduction and in the previous chapter, where our starting point was a sequence $(Y_n)_{n \geq 0}$ of random vectors in \mathbb{R}^d satisfying the distributional recursion

$$Y_n \stackrel{d}{=} \sum_{r=1}^K A_r(n) Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq n_0, \quad (3.1)$$

with $n_0 \in \mathbb{N}$, $K \in \mathbb{N}$, random $(d \times d)$ -matrices $A_1(n), \dots, A_K(n)$, a d -dimensional random vector b_n , random subproblem sizes $I_r^{(n)} \in \{0, \dots, n\}$ for $r = 1, \dots, K$, the self-similarity condition (1.2) and the conditional independence condition (1.3). Since we will work in the space \mathcal{P}_s^d , we further assume that all appearing random variables are L_s -integrable. As already indicated, we normalize the quantities Y_n in the following way:

$$X_n := C_n^{-1/2}(Y_n - M_n), \quad n \geq 0, \quad (3.2)$$

where M_n is a d -dimensional vector and C_n a positive definite $(d \times d)$ -matrix. As recalled in Chapter 2, when working with the Zolotarev distances, it might be necessary to fix

the mean and the covariance matrix of the quantities we work with to guarantee their ζ_s -compatibility. In concrete terms, this means that:

- ◇ For $2 < s \leq 3$, we have to control the mean and the covariances of X_n . For this, we assume that there exists an $n_1 \in \mathbb{N}_0$ such that $\text{Cov}(Y_n)$ is positive definite for $n \geq n_1$ and choose $M_n = \mathbb{E}[Y_n]$ for $n \geq 0$, $C_n = \text{Cov}(Y_n)$ for $n \geq n_1$ and $C_n = \text{Id}_d$ for $n < n_1$. With this choice, we have $\mathbb{E}[X_n] = 0$ and $\text{Cov}(X_n) = \text{Id}_d$ for any $n \geq n_1$.
- ◇ For $1 < s \leq 2$, we just have to control the first moments. Hence, we choose $M_n = \mathbb{E}[Y_n]$ for $n \geq 0$, such that $\mathbb{E}[X_n] = 0$ for any $n \geq 0$. Concerning the matrix C_n , we just assume that it is positive definite and set $n_1 = 0$.
- ◇ For $0 < s \leq 1$, we just assume that $M_n \in \mathbb{R}^d$, C_n is positive definite and set $n_1 = 0$.

The normalized quantities satisfy the modified recursion

$$X_n \stackrel{d}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \quad (3.3)$$

with

$$A_r^{(n)} := C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2} \left(b_n - M_n + \sum_{r=1}^K A_r(n) M_{I_r^{(n)}} \right), \quad (3.4)$$

$(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}, I^{(n)})$, $(X_n^{(1)})_{n \geq 0}, \dots, (X_n^{(K)})_{n \geq 0}$ independent and $X_j^{(r)} \stackrel{d}{=} X_j$ for all $j \geq 0$ and $r = 1, \dots, K$.

Starting with this recursive decomposition of X_n , the idea now is to find general conditions that allow us to transfer the convergence rates of the coefficients $A_r^{(n)}$ and $b^{(n)}$ to bounds on the convergence rates for the sequence $(X_n)_{n \geq 0}$. We start with a general transfer theorem in Section 3.1.1 and then present an improved version of this theorem for normal limits in Section 3.1.2.

3.1.1 A general transfer theorem for rates of convergence

The following theorem provides a general framework to bound convergence rates for the sequence $(X_n)_{n \geq 0}$. For the proof, we need some technical conditions which guarantee that the sizes $I_r^{(n)}$ of the subproblems grow with n . More precisely, we will assume that there exists some monotonically decreasing sequence $R(n) > 0$ with $R(n) \rightarrow 0$ such that

$$\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_s = O(R(n)), \quad n \rightarrow \infty, \quad (3.5)$$

for all $\ell \in \mathbb{N}$ and $r = 1, \dots, K$ and that

$$\|\mathbf{1}_{\{I_r^{(n)} = n\}} A_r^{(n)}\|_s \rightarrow 0, \quad n \rightarrow \infty, \quad (3.6)$$

for all $r = 1, \dots, K$.

Our first result is a direct extension of the main Theorem 4.1 in Neininger and Rüschen-
dorf [70].

Theorem 3.1. *Let $(X_n)_{n \geq 0}$ be given as in (3.2) with the choices for M_n and C_n specified there and with $(Y_n)_{n \geq 0}$ satisfying the distributional recurrence (3.1), where all appearing quantities are assumed to be L_s -integrable for some $0 < s \leq 3$. We further assume that there exist L_s -integrable random variables A_1^*, \dots, A_K^*, b^* and some monotonically decreasing sequence $R(n) > 0$ with $R(n) \rightarrow 0$ such that, as $n \rightarrow \infty$,*

$$\|b^{(n)} - b^*\|_s + \sum_{r=1}^K \|A_r^{(n)} - A_r^*\|_s = O(R(n)). \quad (3.7)$$

If conditions (3.5) and (3.6) are satisfied and if

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^K \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^s \right] < 1, \quad (3.8)$$

then we have, as $n \rightarrow \infty$,

$$\zeta_s(X_n, X) = O(R(n)),$$

where $\mathcal{L}(X)$ is given as the unique solution in $\mathcal{P}_s^d(0, \text{Id}_d)$ of the equation

$$X \stackrel{\text{d}}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*, \quad (3.9)$$

with $(A_1^*, \dots, A_K^*, b^*), X^{(1)}, \dots, X^{(K)}$ independent and $X^{(r)} \stackrel{\text{d}}{=} X$ for $r = 1, \dots, K$.

Remark 3.2. Note that recurrence (3.1) is only a distributional recursion, i.e., the appearing coefficients are not necessarily defined on a common probability space for different n . Hence, strictly speaking, we should talk about convergence in ℓ_s of $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ to $(A_1^*, \dots, A_K^*, b^*)$ instead of convergence in L_s . However, provided that we have ℓ_s -convergence of the coefficients, we can find versions of these coefficients on a common probability space which converge in L_s , see Section 2.2.1. Thus, when talking about L_s -convergence of the coefficients and when estimating the orders in (3.7), we always assume that the quantities are appropriately embedded into a common probability space.

Proof of Theorem 3.1. As a first step, we show that there is indeed a unique solution of equation (3.9) in $\mathcal{P}_s^d(0, \text{Id}_d)$: Using condition (3.7), the assumption that R is monotonically decreasing and condition (3.8), we have

$$\mathbb{E} \left[\sum_{r=1}^K \|A_r^*\|_{\text{op}}^s \right] = \lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{r=1}^K \|A_r^{(n)}\|_{\text{op}}^s \right] \leq \limsup_{n \rightarrow \infty} \mathbb{E} \left[\sum_{r=1}^K \frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^s \right] < 1.$$

Furthermore, condition (3.7) implies $\mathbb{E}[b^*] = \lim_{n \rightarrow \infty} \mathbb{E}[b^{(n)}] = 0$ if $s > 1$ and additionally

$$\mathbb{E}[b^*(b^*)^\top] + \mathbb{E}\left[\sum_{r=1}^K A_r^*(A_r^*)^\top\right] = \text{Id}_d$$

if $s > 2$. Thus, Theorem 2.8 states that equation (3.9) (or rather the associated map T defined in (2.9)) has a unique fixed-point $\mathcal{L}(X)$ in $\mathcal{P}_s^d(0, \text{Id}_d)$. To establish a rate of convergence to this fixed-point, we introduce the accompanying sequence

$$Z_n^* := \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} X^{(r)} + b^{(n)},$$

where $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)}, b^{(n)})$, $X^{(1)}, \dots, X^{(K)}$ are independent and $X^{(r)}$ is identically distributed as X for $r = 1, \dots, K$. Here, for $2 < s \leq 3$, the sequence $(T_n)_{n \geq 0}$ is chosen such that Z_n^* has the same covariance structure as X_n . To be more precise, for $2 < s \leq 3$, we choose $T_n := (\text{Cov}(X_n))^{1/2}$ (i.e., $T_n = \text{Id}_d$ for $n \geq n_1$ and $T_n = (\text{Cov}(Y_n))^{1/2}$ for $n < n_1$). For $s \leq 2$, we do not need to control the covariance of Z_n^* and set $T_n := \text{Id}_d$ for $n \geq 0$.

Then, Z_n^* is L_s -integrable, we have $\mathbb{E}[Z_n^*] = 0$ for $s > 1$ and in the case $s > 2$ additionally $\text{Cov}(Z_n^*) = \text{Cov}(X_n) = \text{Id}_d$ for $n \geq n_1$. Hence, X_n , Z_n^* and X are ζ_s -compatible for $n \geq n_1$. Applying the triangle inequality, we have for $n \geq n_1$

$$\zeta_s(X_n, X) \leq \zeta_s(X_n, Z_n^*) + \zeta_s(Z_n^*, X). \quad (3.10)$$

The idea now is to treat the two terms of the right-hand side of (3.10) separately, starting with the second summand $\zeta_s(Z_n^*, X)$. For this purpose, we switch to the Wasserstein metric ℓ_s : From

$$\|Z_n^*\|_s \leq \sum_{r=1}^K \|A_r^{(n)} T_{I_r^{(n)}}\|_s \|X\|_s + \|b^{(n)}\|_s$$

and condition (3.7), it follows directly that $\sup_{n \geq 0} \|Z_n^*\|_s < \infty$. Thus, Theorem 2.7 implies that $\zeta_s(Z_n^*, X) \leq C_s \ell_s(Z_n^*, X)$ for some constant $C_s > 0$ and it suffices to bound the Wasserstein distance $\ell_s(Z_n^*, X)$. For this distance, we have by definition of the Wasserstein metric ℓ_s

$$\begin{aligned} \ell_s(Z_n^*, X) &\leq \left\| \left(\sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} X^{(r)} + b^{(n)} \right) - \left(\sum_{r=1}^K A_r^* X^{(r)} + b^* \right) \right\|_s \\ &\leq \sum_{r=1}^K \|A_r^{(n)} T_{I_r^{(n)}} - A_r^*\|_s \|X^{(r)}\|_s + \|b^{(n)} - b^*\|_s \\ &\leq \sum_{r=1}^K \left(\|A_r^{(n)} T_{I_r^{(n)}} - A_r^{(n)}\|_s + \|A_r^{(n)} - A_r^*\|_s \right) \|X\|_s + \|b^{(n)} - b^*\|_s \\ &= \sum_{r=1}^K \left(\|\mathbf{1}_{\{I_r^{(n)} < n_1\}} A_r^{(n)} (T_{I_r^{(n)}} - \text{Id}_d)\|_s + \|A_r^{(n)} - A_r^*\|_s \right) \|X\|_s + \|b^{(n)} - b^*\|_s. \end{aligned}$$

From this, using conditions (3.5) and (3.7), we obtain $\ell_s(Z_n^*, X) = O(R(n))$.

To bound the first summand $\zeta_s(X_n, Z_n^*)$ in (3.10), we use some standard estimates for the Zolotarev distances presented in Section 2.2.2. Denoting by Υ_n the joint distribution of $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}, I^{(n)})$, $\alpha = (\alpha_1, \dots, \alpha_K)$, $j = (j_1, \dots, j_K)$ and $\Delta(n) := \zeta_s(X_n, X)$, we obtain by conditioning on the coefficients, by using the s -ideality of ζ_s and the bounds (2.3) and (2.4) that, for $n \geq n_1$,

$$\begin{aligned}
\zeta_s(X_n, Z_n^*) &= \zeta_s \left(\sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} X^{(r)} + b^{(n)} \right) \\
&= \sup_{f \in \mathcal{F}_s} \left| \int \mathbb{E} \left[f \left(\sum_{r=1}^K \alpha_r X_{j_r}^{(r)} + \beta \right) \right] - \mathbb{E} \left[f \left(\sum_{r=1}^K \alpha_r T_{j_r} X^{(r)} + \beta \right) \right] d\Upsilon_n(\alpha, \beta, j) \right| \\
&\leq \int \zeta_s \left(\sum_{r=1}^K \alpha_r X_{j_r}^{(r)} + \beta, \sum_{r=1}^K \alpha_r T_{j_r} X^{(r)} + \beta \right) d\Upsilon_n(\alpha, \beta, j) \\
&\leq \int \sum_{r=1}^K \|\alpha_r\|_{\text{op}}^s \zeta_s \left(X_{j_r}^{(r)}, T_{j_r} X^{(r)} \right) d\Upsilon_n(\alpha, \beta, j) \\
&\leq \left(\mathbb{E} \sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)}=n\}} \|A_r^{(n)}\|_{\text{op}}^s \right) \Delta(n) + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\text{op}}^s \Delta(I_r^{(n)}) \right] \\
&\quad + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} < n_1\}} \|A_r^{(n)}\|_{\text{op}}^s \sup_{k < n_1} \zeta_s(X_k, T_k X) \right]. \tag{3.11}
\end{aligned}$$

Note that the last summand is of order $O(R(n))$ by condition (3.5). Hence, putting everything together and introducing the notation

$$p_n := \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)}=n\}} \|A_r^{(n)}\|_{\text{op}}^s \right],$$

we obtain from (3.10) and (3.11) that

$$\Delta(n) \leq p_n \Delta(n) + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\text{op}}^s \Delta(I_r^{(n)}) \right] + O(R(n)). \tag{3.12}$$

From (3.8), there exists a $\delta > 0$ such that

$$\mathbb{E} \left[\sum_{r=1}^K \frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^s \right] \leq 1 - \delta$$

for all n sufficiently large and from (3.6) we have $p_n < \delta/2$ for n large. We now choose some $C > 0$ and $n_2 \geq n_1$ sufficiently large such that for $n \geq n_2$ all these inequalities are

satisfied and the $O(R(n))$ term in (3.12) is bounded by $CR(n)$. By setting

$$L := \frac{2C}{\delta} \vee \max \left\{ \frac{\Delta(n)}{R(n)} : n \leq n_2 \right\},$$

we now obtain $\Delta(n) \leq LR(n)$ by induction: For $n \leq n_2$, by definition of L , the assertion is true. For $n > n_2$, solving for $\Delta(n)$ in (3.12), we find

$$\begin{aligned} \Delta(n) &\leq \frac{1}{1-p_n} \left(\mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\text{op}}^s \Delta(I_r^{(n)}) \right] + CR(n) \right) \\ &\leq \frac{1}{1-\delta/2} \left(\mathbb{E} \left[\sum_{r=1}^K \|A_r^{(n)}\|_{\text{op}}^s LR(I_r^{(n)}) \right] + CR(n) \right) \\ &= \frac{1}{1-\delta/2} \left(L \mathbb{E} \left[\sum_{r=1}^K \|A_r^{(n)}\|_{\text{op}}^s \frac{R(I_r^{(n)})}{R(n)} \right] R(n) + CR(n) \right) \\ &\leq \frac{1}{1-\delta/2} (L(1-\delta) + C) R(n) \\ &\leq LR(n), \end{aligned}$$

which completes the proof. \square

Remark 3.3. In applications, the convergence rate of the coefficients (conditions (3.5) and (3.7)) is often faster than the convergence rate of the quantities X_n , see Section 4.4 for an example. In these cases, it is often possible to perform the induction step in the proof of Theorem 3.1 although condition (3.8) does not hold. To be more precise, we may assume that $\mathbb{E}[\sum_{r=1}^K \|A_r^*\|_{\text{op}}^s] < 1$ (this is to ensure the existence of a unique fixed-point) and that

$$\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_s + \|b^{(n)} - b^*\|_s + \|A_r^{(n)} - A_r^*\|_s = O(\tilde{R}(n))$$

for every $\ell \in \mathbb{N}$, $r = 1, \dots, K$ and as $n \rightarrow \infty$. Then, instead of condition (3.8), it is sufficient to find some $C > 0$ such that

$$\mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^s \right] \leq 1 - p_n - \frac{\tilde{R}(n)}{CR(n)}$$

for all large n with $p_n := \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} = n\}} \|A_r^{(n)}\|_{\text{op}}^s \right]$ as above. In other words, it is enough to show

$$\mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^s \right] \leq 1 - p_n - \varepsilon_n$$

for all large n with some positive $\varepsilon_n = \Omega(\tilde{R}(n)/R(n))$. We will make use of this observation in several applications, for instance, when analyzing the number of key comparisons of the Quickselect algorithm in Section 4.4.

3.1.2 An improved transfer theorem for normal limit distributions

We now consider the special case where the sequence $(X_n)_{n \geq 0}$ is L_3 -integrable and satisfies recursion (3.3) with $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L_3} (A_1^*, \dots, A_K^*, b^*)$ for some L_3 -integrable coefficients A_1^*, \dots, A_K^*, b^* with

$$b^* = 0, \quad \sum_{r=1}^K A_r^* (A_r^*)^T = \text{Id}_d \quad (3.13)$$

almost surely. Theorem 2.8 implies that, if $\mathbb{E}[\sum_{r=1}^K \|A_r^*\|_{\text{op}}^3] < 1$, equation (3.9) has a unique solution in the space $\mathcal{P}_3^d(0, \text{Id}_d)$. Furthermore, e.g., by using characteristic functions, it is easily checked that this unique solution is the standard normal distribution $\mathcal{N}(0, \text{Id}_d)$.

In this special case of normal limits, it is possible to derive a refined version of Theorem 3.1. Instead of the technical condition (3.5), we now need the weaker condition

$$\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_3^3 = O(R(n)), \quad n \rightarrow \infty, \quad (3.14)$$

for all $\ell \in \mathbb{N}$ and $r = 1, \dots, K$. Moreover, condition (3.7) concerning the convergence rates of the coefficients can be weakened, which is formulated in the following theorem.

Theorem 3.4. *Let $(X_n)_{n \geq 0}$ be given as in (3.2) with M_n and C_n denoting the mean vector and the covariance matrix of Y_n for $n \geq n_1$ and with $(Y_n)_{n \geq 0}$ satisfying the distributional recurrence (3.1), where all appearing quantities are assumed to be L_3 -integrable. We further assume that the coefficients $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ converge in L_3 to some limiting coefficients $(A_1^*, \dots, A_K^*, b^*)$ satisfying (3.13) and that for some $R(n) > 0$ monotonically decreasing with $R(n) \rightarrow 0$ we have, as $n \rightarrow \infty$,*

$$\left\| \sum_{r=1}^K A_r^{(n)} (A_r^{(n)})^T - \text{Id}_d \right\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3 = O(R(n)). \quad (3.15)$$

If the technical conditions (3.6) and (3.14) are satisfied for $s = 3$ and if

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^K \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^3 \right] < 1, \quad (3.16)$$

then we have, as $n \rightarrow \infty$,

$$\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d)) = O(R(n)).$$

Theorem 3.4, when applicable, often improves over Theorem 3.1 by a factor 3 in the exponent, see Remark 4.7 for an example. This is caused by the additional exponents in (3.15) in comparison to (3.7).

Remark 3.5. Similarly to Remark 3.3, in some applications, it is possible to perform the induction step in the proof of Theorem 3.4 although condition (3.16) is not satisfied, see Section 4.1.2 and Section 4.3 for examples. To this end, we assume that

$$\left\| \sum_{r=1}^K A_r^{(n)} (A_r^{(n)})^\top - \text{Id}_d \right\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3 + \sum_{r=1}^K \|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_3^3 = O(\tilde{R}(n))$$

for every $\ell \in \mathbb{N}$ and as $n \rightarrow \infty$. Then, instead of condition (3.16), it is sufficient to check that $\mathbb{E}[\sum_{r=1}^K \|A_r^*\|_{\text{op}}^3] < 1$ and to show

$$\mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^3 \right] \leq 1 - p_n - \varepsilon_n$$

for all large n with $p_n := \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} = n\}} \|A_r^{(n)}\|_{\text{op}}^s \right]$ and some positive $\varepsilon_n = \Omega(\tilde{R}(n)/R(n))$. We will make use of this modified condition in several applications, for instance, when analyzing the number of maxima in right triangles (Section 4.1.2).

Proof of Theorem 3.4. Similarly as in the proof of Theorem 3.1, we start with an accompanying sequence $(Z_n^*)_{n \geq 0}$ defined by

$$Z_n^* := \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} N^{(r)} + b^{(n)}, \quad n \geq 0,$$

where $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)}, b^{(n)}, N^{(1)}, \dots, N^{(K)})$ are independent, $\mathcal{L}(N^{(r)}) = \mathcal{N}(0, \text{Id}_d)$ for $r = 1, \dots, K$ and $T_n = (\text{Cov}(X_n))^{1/2}$ for $n \geq 0$. Thus, Z_n^* is L_3 -integrable, $\mathbb{E}[Z_n^*] = 0$ and $\text{Cov}(Z_n^*) = \text{Id}_d$ for all $n \geq n_1$, implying that X_n , Z_n^* and $\mathcal{N}(0, \text{Id}_d)$ are ζ_3 -compatible for $n \geq n_1$. By the triangle inequality, we have

$$\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d)) \leq \zeta_3(X_n, Z_n^*) + \zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)).$$

In the remaining part of this proof, we will show that the second summand $\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d))$ is of order $O(R(n))$. With this bound, after having handled the first summand $\zeta_3(X_n, Z_n^*)$ just as in the proof of Theorem 3.1, the same inductive argument as used there finally implies the assertion.

Hence, it remains to show the bound $\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)) = O(R(n))$: Using the convolution property of the multivariate normal distribution, we obtain the representation

$$Z_n^* = \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} N^{(r)} + b^{(n)} \stackrel{d}{=} G_n N + b^{(n)}, \quad (3.17)$$

where G_n is a $(d \times d)$ -matrix satisfying

$$G_n G_n^\top = \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} T_{I_r^{(n)}}^\top (A_r^{(n)})^\top,$$

$\mathcal{L}(N) = \mathcal{N}(0, \text{Id}_d)$ and N is independent of $(G_n, b^{(n)})$. Furthermore, we have $\mathbb{E}[Z_n^*] = 0$, $\text{Cov}(Z_n^*) = \text{Id}_d$ for $n \geq n_1$, $\|b^{(n)}\|_3^3 = O(R(n))$ and

$$\begin{aligned} \|G_n G_n^\text{T} - \text{Id}_d\|_{3/2}^{3/2} &= \left\| \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} T_{I_r^{(n)}}^\text{T} (A_r^{(n)})^\text{T} - \text{Id}_d \right\|_{3/2}^{3/2} \\ &= O \left(\left\| \sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} < n_1\}} A_r^{(n)} (T_{I_r^{(n)}} T_{I_r^{(n)}}^\text{T} - \text{Id}_d) (A_r^{(n)})^\text{T} \right\|_{3/2}^{3/2} \right. \\ &\quad \left. + \left\| \sum_{r=1}^K A_r^{(n)} (A_r^{(n)})^\text{T} - \text{Id}_d \right\|_{3/2}^{3/2} \right) \\ &= O \left(\sum_{r=1}^K \|\mathbf{1}_{\{I_r^{(n)} < n_1\}} A_r^{(n)}\|_3^3 + \left\| \sum_{r=1}^K A_r^{(n)} (A_r^{(n)})^\text{T} - \text{Id}_d \right\|_{3/2}^{3/2} \right) \\ &= O(R(n)) \end{aligned}$$

by conditions (3.14) and (3.15). As a consequence, the subsequent Lemma 3.6 implies $\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)) = O(R(n))$, which finishes the proof. \square

The following lemma served as the main component of the preceding proof.

Lemma 3.6. *Let $(Z_n^*)_{n \geq 0}$ be a sequence of L_3 -integrable d -dimensional random vectors satisfying*

$$Z_n^* \stackrel{d}{=} G_n N + b^{(n)},$$

with some random $(d \times d)$ -matrix G_n and some d -dimensional random vector $b^{(n)}$ such that $\mathbb{E}[Z_n^] = 0$, $\text{Cov}(Z_n^*) = \text{Id}_d$, $\mathcal{L}(N) = \mathcal{N}(0, \text{Id}_d)$ and N being independent of $(G_n, b^{(n)})$. Furthermore, we assume that, as $n \rightarrow \infty$,*

$$\|G_n G_n^\text{T} - \text{Id}_d\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3 = O(R(n))$$

for appropriate $R(n)$. Then we have, as $n \rightarrow \infty$,

$$\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)) = O(R(n)).$$

The proof of Lemma 3.6 builds upon ideas of Neininger and Rüschemdorf [71].

Proof of Lemma 3.6. As the matrix $G_n G_n^\text{T}$ is positive semidefinite, we can decompose it in the following way (to simplify notation, we do not explicitly indicate the dependence on n in this decomposition): Let $\lambda_1 \geq \dots \geq \lambda_m \geq 1 > \lambda_{m+1} \geq \dots \geq \lambda_d \geq 0$ be the (random) eigenvalues of $G_n G_n^\text{T}$. Then, with a suitable (random) orthogonal matrix O , we have

$$\begin{aligned} G_n G_n^\text{T} &= O \text{diag}(\lambda_1, \dots, \lambda_d) O^\text{T} \\ &= O \text{diag}(1, \dots, 1, \lambda_{m+1}, \dots, \lambda_d) O^\text{T} + O \text{diag}(\lambda_1 - 1, \dots, \lambda_m - 1, 0, \dots, 0) O^\text{T} \\ &= B_n B_n^\text{T} + C_n C_n^\text{T}, \end{aligned}$$

where we define the random $(d \times d)$ -matrices $B_n := O \operatorname{diag}(1, \dots, 1, \sqrt{\lambda_{m+1}}, \dots, \sqrt{\lambda_d}) O^T$ and $C_n := O \operatorname{diag}(\sqrt{\lambda_1 - 1}, \dots, \sqrt{\lambda_m - 1}, 0, \dots, 0) O^T$. Hence, we can decompose Z_n^* in the following way:

$$Z_n^* \stackrel{d}{=} G_n N + b^{(n)} \stackrel{d}{=} B_n N + C_n N' + b^{(n)} =: \hat{Z}_n^*,$$

where $(B_n, C_n, b^{(n)})$, N and N' are independent with $\mathcal{L}(N) = \mathcal{L}(N') = \mathcal{N}(0, \operatorname{Id}_d)$. Analogously, we decompose the multivariate standard normal distribution: With the same random variables N and N' as in the definition of \hat{Z}_n^* , we write

$$N \stackrel{d}{=} B_n N + D_n N' =: \hat{N},$$

where the matrix $D_n := O \operatorname{diag}(0, \dots, 0, \sqrt{1 - \lambda_{m+1}}, \dots, \sqrt{1 - \lambda_d}) O^T$ is chosen such that $B_n B_n^T + D_n D_n^T = \operatorname{Id}_d$. The idea behind this decomposition is to construct \hat{Z}_n^* and \hat{N} on a common probability space such that they both contain the summand $B_n N$ (main part) and only differ in the (small) terms $C_n N'$, $D_n N'$ and $b^{(n)}$.

By definition of the Zolotarev metric ζ_3 , we have

$$\zeta_3(Z_n^*, \mathcal{N}(0, \operatorname{Id}_d)) = \zeta_3(\hat{Z}_n^*, \hat{N}) = \sup_{f \in \mathcal{F}_3} \left| \mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})] \right|,$$

where the set \mathcal{F}_3 contains all functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ whose second derivative is Lipschitz continuous with Lipschitz constant 1. In view of this, our goal is to bound $|\mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]|$ uniformly in f . To this end, for arbitrary $f \in \mathcal{F}_3$, we use Taylor expansion around N and obtain for $x \in \mathbb{R}^d$ that

$$f(x) = f(N) + (x - N)^T \nabla f(N) + \frac{1}{2} (x - N)^T H_f(N) (x - N) + R(x, N),$$

where ∇f and H_f denote the gradient and the Hessian matrix of f and the remainder term satisfies $|R(x, N)| \leq \frac{1}{2} \|x - N\|^3$. Thus, we have

$$\begin{aligned} f(\hat{Z}_n^*) - f(\hat{N}) &= (\hat{Z}_n^* - \hat{N})^T \nabla f(N) + \frac{1}{2} (\hat{Z}_n^* - N)^T H_f(N) (\hat{Z}_n^* - N) \\ &\quad - \frac{1}{2} (\hat{N} - N)^T H_f(N) (\hat{N} - N) + R(\hat{Z}_n^*, N) - R(\hat{N}, N). \end{aligned} \quad (3.18)$$

We now study the expectations of these summands: For the first summand, we have

$$\begin{aligned} \mathbb{E}[(\hat{Z}_n^* - \hat{N})^T \nabla f(N)] &= \mathbb{E}[(C_n - D_n)N' + b^{(n)}]^T \nabla f(N) \\ &= \mathbb{E}[(C_n - D_n)N' + b^{(n)}]^T \mathbb{E}[\nabla f(N)] = 0, \end{aligned}$$

since N is independent of the other quantities, N' is independent of (C_n, D_n) and $\mathbb{E}[N'] = \mathbb{E}[b^{(n)}] = 0$. For the second summand in (3.18), we define $F_n := B_n - \operatorname{Id}_d$ and obtain

$$\begin{aligned} &\mathbb{E}[(\hat{Z}_n^* - N)^T H_f(N) (\hat{Z}_n^* - N)] \\ &= \mathbb{E}[(F_n N + C_n N' + b^{(n)})^T H_f(N) (F_n N + C_n N' + b^{(n)})] \\ &= \mathbb{E}[(F_n N)^T H_f(N) (F_n N)] + \mathbb{E}[(F_n N)^T H_f(N) (C_n N')] + \mathbb{E}[(F_n N)^T H_f(N) b^{(n)}] \\ &\quad + \mathbb{E}[(C_n N')^T H_f(N) (F_n N)] + \mathbb{E}[(C_n N')^T H_f(N) (C_n N')] + \mathbb{E}[(C_n N')^T H_f(N) b^{(n)}] \\ &\quad + \mathbb{E}[(b^{(n)})^T H_f(N) (F_n N)] + \mathbb{E}[(b^{(n)})^T H_f(N) (C_n N')] + \mathbb{E}[(b^{(n)})^T H_f(N) b^{(n)}]. \end{aligned}$$

Since N , N' and $(F_n, C_n, b^{(n)})$ are independent with $\mathbb{E}[N'] = 0$, we have

$$\mathbb{E}[(F_n N)^\top H_f(N)(C_n N')] = 0.$$

The same argument applies to the terms $\mathbb{E}[(C_n N')^\top H_f(N)(F_n N)]$, $\mathbb{E}[(C_n N')^\top H_f(N)b^{(n)}]$ and $\mathbb{E}[(b^{(n)})^\top H_f(N)(C_n N')]$. Analogously, we obtain for the third summand in (3.18)

$$\begin{aligned} & \mathbb{E}[(\hat{N} - N)^\top H_f(N)(\hat{N} - N)] \\ &= \mathbb{E}[(F_n N + D_n N')^\top H_f(N)(F_n N + D_n N')] \\ &= \mathbb{E}[(F_n N)^\top H_f(N)(F_n N)] + \mathbb{E}[(D_n N')^\top H_f(N)(D_n N')]. \end{aligned}$$

This implies, together with $\mathbb{E}[(F_n N)^\top H_f(N)b^{(n)}] = \mathbb{E}[(b^{(n)})^\top H_f(N)(F_n N)]$,

$$\begin{aligned} & \mathbb{E}[(\hat{Z}_n^* - N)^\top H_f(N)(\hat{Z}_n^* - N) - (\hat{N} - N)^\top H_f(N)(\hat{N} - N)] \\ &= \mathbb{E}[(C_n N')^\top H_f(N)(C_n N')] - \mathbb{E}[(D_n N')^\top H_f(N)(D_n N')] + \mathbb{E}[(b^{(n)})^\top H_f(N)b^{(n)}] \\ & \quad + 2 \mathbb{E}[(F_n N)^\top H_f(N)b^{(n)}]. \end{aligned}$$

Note that we have $C_n C_n^\top - D_n D_n^\top = G_n G_n^\top - \text{Id}_d$ by definition of C_n and D_n . Furthermore, the assumption that Z_n^* has covariance matrix Id_d implies that $\mathbb{E}[G_n G_n^\top + b^{(n)}(b^{(n)})^\top] = \text{Id}_d$. Thus, with the independence of N , N' and $(C_n, D_n, b^{(n)})$ and $\mathbb{E}[N'_i N'_j] = \mathbf{1}_{\{i=j\}}$ for $i, j = 1, \dots, d$, we have

$$\begin{aligned} & \mathbb{E}[(C_n N')^\top H_f(N)(C_n N')] - \mathbb{E}[(D_n N')^\top H_f(N)(D_n N')] + \mathbb{E}[(b^{(n)})^\top H_f(N)b^{(n)}] \\ &= \sum_{i,j=1}^d \mathbb{E}[H_f(N)_{ij}] \mathbb{E}[(C_n N')_i (C_n N')_j - (D_n N')_i (D_n N')_j + b_i^{(n)} b_j^{(n)}] \\ &= \sum_{i,j=1}^d \mathbb{E}[H_f(N)_{ij}] \mathbb{E}[(C_n C_n^\top - D_n D_n^\top)_{ij} + (b^{(n)}(b^{(n)})^\top)_{ij}] \\ &= \sum_{i,j=1}^d \mathbb{E}[H_f(N)_{ij}] \mathbb{E}[(G_n G_n^\top + b^{(n)}(b^{(n)})^\top - \text{Id}_d)_{ij}] \\ &= 0. \end{aligned}$$

Bringing together the previous results, we find that

$$\begin{aligned} \left| \mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})] \right| &= \left| \mathbb{E}[(F_n N)^\top H_f(N)b^{(n)}] + \mathbb{E}[R(\hat{Z}_n^*, N)] - \mathbb{E}[R(\hat{N}, N)] \right| \\ &\leq \mathbb{E} \left[|(F_n N)^\top H_f(N)b^{(n)}| \right] + \mathbb{E}[|R(\hat{Z}_n^*, N)|] + \mathbb{E}[|R(\hat{N}, N)|]. \end{aligned}$$

We now bound these three terms. For this, without loss of generality, we may assume that $H_f(0) = 0$: If this is not the case, we consider the function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ defined by $g(x) := f(x) - \frac{1}{2}x^\top H_f(0)x$ for $x \in \mathbb{R}^d$. Then, by definition of g , we have $H_g(0) = 0$ and

$\mathbb{E}[g(\hat{Z}_n^*) - g(\hat{N})] = \mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]$, since \hat{Z}_n^* and \hat{N} both have mean vector 0 and covariance matrix Id_d . The assumption $H_f(0) = 0$ implies, together with the Lipschitz property of the second derivative of f , $\|H_f(N)\|_{\text{op}} \leq \|N\|$. Hence, using the Cauchy–Schwarz inequality, the independence of $(F_n, b^{(n)})$ and N and Hölder’s inequality, we have

$$\begin{aligned} \mathbb{E}\left[\left|(F_n N)^T H_f(N) b^{(n)}\right|\right] &\leq \mathbb{E}\left[\|F_n\|_{\text{op}} \|N\| \|H_f(N)\|_{\text{op}} \|b^{(n)}\|\right] \\ &\leq \mathbb{E}\left[\|N\|^2\right] \mathbb{E}\left[\|F_n\|_{\text{op}} \|b^{(n)}\|\right] \\ &\leq d \|F_n\|_{3/2} \|b^{(n)}\|_3 \\ &\leq d \|G_n G_n^T - \text{Id}_d\|_{3/2} \|b^{(n)}\|_3, \end{aligned}$$

where the last step follows by $\|G_n G_n^T - \text{Id}_d\|_{\text{op}} = \max\{|\lambda_1 - 1|, |\lambda_d - 1|\}$, $\|F_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_d < 1\}} |\sqrt{\lambda_d} - 1|$ and the identity $|\sqrt{a} - 1| \leq |a - 1|$ for any real number $a \geq 0$. The first remainder term is bounded by

$$\begin{aligned} \mathbb{E}[|R(\hat{Z}_n^*, N)|] &\leq \frac{1}{2} \mathbb{E}[\|\hat{Z}_n^* - N\|^3] \\ &= \frac{1}{2} \mathbb{E}[\|F_n N + C_n N' + b^{(n)}\|^3] \\ &= O\left(\mathbb{E}[\|F_n\|_{\text{op}}^3] + \mathbb{E}[\|C_n\|_{\text{op}}^3] + \mathbb{E}[\|b^{(n)}\|^3]\right) \\ &= O\left(\|G_n G_n^T - \text{Id}_d\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3\right), \end{aligned}$$

since $\|C_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_1 > 1\}} \sqrt{|\lambda_1 - 1|} \leq \|G_n G_n^T - \text{Id}_d\|_{\text{op}}^{1/2}$ and $\|F_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_d < 1\}} |\sqrt{\lambda_d} - 1| \leq \|G_n G_n^T - \text{Id}_d\|_{\text{op}}^{1/2}$ (note that we have $|\sqrt{a} - 1| \leq \sqrt{|a - 1|}$ for any $a \geq 0$). With the same arguments, we obtain for the second remainder term

$$\begin{aligned} \mathbb{E}[|R(\hat{N}, N)|] &\leq \frac{1}{2} \mathbb{E}[\|F_n N + D_n N'\|^3] = O(\|F_n\|_3^3 + \|D_n\|_3^3) \\ &= O(\|G_n G_n^T - \text{Id}_d\|_{3/2}^{3/2}), \end{aligned}$$

as $\|D_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_d < 1\}} \sqrt{|\lambda_d - 1|} \leq \|G_n G_n^T - \text{Id}_d\|_{\text{op}}^{1/2}$. This implies

$$\begin{aligned} \left|\mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]\right| &\leq \mathbb{E}\left[\left|(F_n N)^T H_f(N) b^{(n)}\right|\right] + \mathbb{E}[|R(\hat{Z}_n^*, N)|] + \mathbb{E}[|R(\hat{N}, N)|] \\ &= O\left(\|G_n G_n^T - \text{Id}_d\|_{3/2} \|b^{(n)}\|_3 + \|G_n G_n^T - \text{Id}_d\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3\right) \\ &= O(R(n)). \end{aligned}$$

Note that the constants in the O-notation do not depend on the function f , i.e., we have

$$\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)) = \sup_{f \in \mathcal{F}_3} \left|\mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]\right| = O(R(n)),$$

which is the assertion. \square

3.2 Expansions of moments

In applications to problems arising in theoretical computer science or in other fields, where the recurrence (3.1) is explicitly given, one usually has no direct means to identify the orders of the terms $\|b^{(n)} - b^*\|_s$ and $\|A_r^{(n)} - A_r^*\|_s$. This is due to the fact that the mean vector M_n and the covariance matrix C_n , for the cases $1 < s \leq 2$ and $2 < s \leq 3$, respectively, which are used for the normalization (3.2) are typically not exactly known or too involved to be amenable to explicit calculations. As a substitute, one usually has asymptotic expansions of these sequences as $n \rightarrow \infty$.

With this in mind, the goal of the present section is to provide a theorem which allows us to also handle examples where only asymptotic expansions of the moments are known. Thereby, we focus on the univariate case with $A_r(n) = 1$ for all $r = 1, \dots, K$ and provide tools to apply the general Theorems 3.1 and 3.4 on the basis of expansions of the mean and variance. We assume that

$$\mathbb{E}[Y_n] = \mu(n) = f(n) + O(e(n)), \quad \text{Var}(Y_n) = \sigma^2(n) = g(n) + O(h(n)), \quad (3.19)$$

with $e(n) = o(f(n))$, $h(n) = o(g(n))$ and a non-negative function g . The scaling introduced in (3.2) with the special choices $A_r(n) = 1$ for all $r = 1, \dots, K$ leads to the scaled recurrence for X_n given in (3.3) with coefficients

$$A_r^{(n)} = \frac{\sigma(I_r^{(n)})}{\sigma(n)}, \quad b^{(n)} = \frac{1}{\sigma(n)} \left(b_n - \mu(n) + \sum_{r=1}^K \mu(I_r^{(n)}) \right) \quad (3.20)$$

for $n \geq n_1$. Additionally, we consider the corresponding quantities

$$\bar{A}_r^{(n)} = \frac{g^{1/2}(I_r^{(n)})}{g^{1/2}(n)}, \quad \bar{b}^{(n)} = \frac{1}{g^{1/2}(n)} \left(b_n - f(n) + \sum_{r=1}^K f(I_r^{(n)}) \right). \quad (3.21)$$

To connect Theorems 3.1 and 3.4 to recurrences with known expansions, we use the following notion.

Definition 3.7. A sequence $(a(n))_{n \geq 0}$ of non-negative real numbers is called *essentially non-decreasing* if there exists a $c > 0$ such that $a(m) \leq c a(n)$ for all $0 \leq m < n$.

To apply Theorems 3.1 and 3.4, we have to identify the orders of $\|A_r^{(n)} - A_r^*\|_s$, $\|b^{(n)} - b^*\|_s$, $\|(A_1^{(n)})^2 + \dots + (A_K^{(n)})^2 - 1\|_{3/2}$ and handle the terms appearing in (3.8) and (3.16). The following lemma can be useful for this.

Lemma 3.8. *In the situation of Theorem 3.1 (Theorem 3.4, respectively), with $A_r^{(n)}$, $b^{(n)}$ given in (3.20), $\bar{A}_r^{(n)}$, $\bar{b}^{(n)}$ given in (3.21), and the expansions for $\mu(n)$, $\sigma^2(n)$ given in (3.19), the following holds for $s \geq 1$ and as $n \rightarrow \infty$.*

If the sequence $h/g^{1/2}$ is essentially non-decreasing and $\mathbb{P}(I_r^{(n)} < \ell)^{1/s} = O(h(n)/g^{1/2}(n))$ for any $\ell \in \mathbb{N}$, then we have for $r = 1, \dots, K$

$$\|A_r^{(n)} - A_r^*\|_s \leq \|\bar{A}_r^{(n)} - A_r^*\|_s + O\left(\frac{h(n)}{g(n)}\right). \quad (3.22)$$

If the sequence h is essentially non-decreasing and $\mathbb{P}(I_r^{(n)} < \ell)^{2/3} = O(h(n))$ for any $\ell \in \mathbb{N}$ and $r = 1, \dots, K$, then we have

$$\left\| \sum_{r=1}^K (A_r^{(n)})^2 - 1 \right\|_{3/2} \leq \left\| \sum_{r=1}^K (\bar{A}_r^{(n)})^2 - 1 \right\|_{3/2} + O\left(\frac{h(n)}{g(n)}\right). \quad (3.23)$$

If the sequence e is essentially non-decreasing and $\mathbb{P}(I_r^{(n)} < \ell)^{1/s} = O(e(n))$ for any $\ell \in \mathbb{N}$ and $r = 1, \dots, K$, then we have

$$\|b^{(n)} - b^*\|_s \leq \|\bar{b}^{(n)} - b^*\|_s + O\left(\frac{h(n)}{g(n)} + \frac{e(n)}{g^{1/2}(n)}\right). \quad (3.24)$$

If the sequence g/h is essentially non-decreasing, $(r(n))_{n \geq 0}$ is a non-negative sequence satisfying $\mathbb{P}(I_r^{(n)} < \ell) = O(r(n))$ for any $\ell \in \mathbb{N}$, $r = 1, \dots, K$ and

$$T(n) := \sum_{r=1}^K \mathbb{E} \left[\frac{g^{s/2-1}(I_r^{(n)})h(I_r^{(n)})R(I_r^{(n)})}{g^{s/2}(n)R(n)} \right] + \frac{r(n)}{g^{s/2}(n)R(n)},$$

then we have

$$\sum_{r=1}^K \mathbb{E} \left[\frac{\sigma^s(I_r^{(n)})R(I_r^{(n)})}{\sigma^s(n)R(n)} \right] \leq \sum_{r=1}^K \mathbb{E} \left[\frac{g^{s/2}(I_r^{(n)})R(I_r^{(n)})}{g^{s/2}(n)R(n)} \right] + O(T(n)). \quad (3.25)$$

Proof. We only show (3.22), since the other bounds can be shown similarly. First, with the triangle inequality, we obtain

$$\|A_r^{(n)} - A_r^*\|_s \leq \|\bar{A}_r^{(n)} - A_r^*\|_s + \|A_r^{(n)} - \bar{A}_r^{(n)}\|_s,$$

thus it suffices to show that

$$\|A_r^{(n)} - \bar{A}_r^{(n)}\|_s = O\left(\frac{h(n)}{g(n)}\right).$$

To this end, note that $\sigma^2(n) = g(n) + O(h(n))$ implies $\sigma(n) = g^{1/2}(n) + O(h(n)/g^{1/2}(n))$, i.e., we can find a sequence $(d(n))_{n \geq 0}$ with $d(n) = O(h(n)/g^{1/2}(n))$ such that $\sigma(n) = g^{1/2}(n) + d(n)$ for any $n \geq 0$. With this in mind, we write

$$A_r^{(n)} = \frac{\sigma(I_r^{(n)})}{\sigma(n)} = \frac{g^{1/2}(I_r^{(n)})}{g^{1/2}(n)} \cdot \frac{g^{1/2}(n)}{\sigma(n)} + \frac{d(I_r^{(n)})}{\sigma(n)} = \bar{A}_r^{(n)} \left(1 - \frac{d(n)}{\sigma(n)}\right) + \frac{d(I_r^{(n)})}{\sigma(n)}$$

and obtain

$$\|A_r^{(n)} - \bar{A}_r^{(n)}\|_s \leq \frac{|d(n)|}{\sigma(n)} \|\bar{A}_r^{(n)}\|_s + \frac{1}{\sigma(n)} \|d(I_r^{(n)})\|_s.$$

Since $\bar{A}_r^{(n)} \rightarrow A_r^*$ in L_s , we have $\|\bar{A}_r^{(n)}\|_s = O(1)$, hence the first summand in the latter display is of order $O(g(n)/h(n))$. To bound the second summand, we use that $d(n) = O(h(n)/g^{1/2}(n))$, i.e., we can find some $c > 0$ and some $n_2 \in \mathbb{N}$ such that $|d(n)| \leq ch(n)/g^{1/2}(n)$ for $n \geq n_2$. This implies

$$\begin{aligned} \|d(I_r^{(n)})\|_s &\leq \|\mathbf{1}_{\{I_r^{(n)} < n_2\}} d(I_r^{(n)})\|_s + \|\mathbf{1}_{\{I_r^{(n)} \geq n_2\}} d(I_r^{(n)})\|_s \\ &\leq O(\|\mathbf{1}_{\{I_r^{(n)} < n_2\}}\|_s) + O\left(\left\|\frac{h(I_r^{(n)})}{g^{1/2}(I_r^{(n)})}\right\|_s\right) = O\left(\frac{h(n)}{g^{1/2}(n)}\right), \end{aligned}$$

where the last step follows from the assumptions that $h/g^{1/2}$ is essentially non-decreasing and that $\mathbb{P}(I_r^{(n)} < \ell)^{1/s} = O(h(n)/g^{1/2}(n))$ for any $\ell \in \mathbb{N}$. \square

Note that in applications, the terms on the right hand side in the estimates (3.22)–(3.25) can easily be bounded when expansions as in (3.19) with explicit functions e, f, g, h are available. Numerous examples of this procedure are given in the following chapter.

Chapter 4

Applications

This chapter provides some examples of typical applications of Theorems 3.1 and 3.4. We start by deriving some known results in Section 4.1 to illustrate in detail how to apply our framework of the previous sections. Subsequently, we use the presented theorems to analyze certain quantities in binomial splitting processes in Section 4.2, local counters in binary search trees in Section 4.3 and different cost measures of the Quickselect algorithm in Section 4.4.

4.1 Some known results

In this section, we consider some recursive quantities for which the rate of convergence has already been identified (either in the Zolotarev metric or in the Kolmogorov metric). In all examples, we can apply the theorems from Chapter 3 to rederive those results without much effort.

4.1.1 The number of key comparisons used by Quicksort

As a first example, we analyze the number Y_n of key comparisons used by the divide-and-conquer sorting algorithm *Quicksort* (see Hoare [39]) on a list of n randomly permuted (distinct) numbers. In short, Quicksort works as follows: After having selected an arbitrary element p of the list—the so-called pivot—the main idea of Quicksort is to classify the remaining elements into elements smaller than p and elements larger than p . While doing so, the list is successively reordered in a way such that, in the end, all small elements come before p and all large elements after p . After this partitioning step, the pivot p is at its final position and the strategy is recursively repeated on both parts. In practice, there are lots of concrete implementations of Quicksort, but for our purposes, we just assume that the partitioning strategy preserves the randomness and the independence between the subfiles and uses exactly $n - 1$ key comparisons in the first partitioning step.

Under these assumptions, the number Y_n of key comparisons needed by the Quicksort algorithm to sort n randomly permuted (distinct) numbers satisfies $Y_0 = 0$ and the distributional recursion

$$Y_n \stackrel{d}{=} Y_{I_n}^{(1)} + Y_{n-1-I_n}^{(2)} + n - 1, \quad n \geq 1, \quad (4.1)$$

where $(Y_n^{(1)})_{n \geq 0}, (Y_n^{(2)})_{n \geq 0}, I_n$ are independent, I_n is uniformly distributed on $\{0, \dots, n-1\}$ and $Y_j^{(r)}$ has the same distribution as Y_j for $j \geq 0$ and $r = 1, 2$. Hence, equation (4.1) is covered by our general recurrence (3.1) with $n_0 = 1, K = 2, A_r(n) = 1$ for $r = 1, 2, I_1^{(n)} = I_n, I_2^{(n)} = n - 1 - I_n$, and $b_n = n - 1$. For the expectation and variance of Y_n , exact expressions are known which imply the asymptotic expansions (see Hoare [39] and Knuth [55])

$$\mathbb{E}[Y_n] = 2n \log n + (2\gamma - 4)n + O(\log n), \quad (4.2)$$

$$\text{Var}(Y_n) = \sigma^2 n^2 - 2n \log n + O(n), \quad (4.3)$$

where $\gamma = 0.57721\dots$ denotes the Euler–Mascheroni constant and $\sigma^2 := 7 - 2\pi^2/3 > 0$. Since $\text{Var}(Y_n)$ is positive for $n \geq 3$, we introduce the normalized quantities $X_0 := X_1 := X_2 := 0$ and

$$X_n := \frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \quad n \geq 3. \quad (4.4)$$

Limit theorems $X_n \rightarrow X$ for these standardized quantities have been derived by different methods by Régnier [77] and Rösler [80]. Rösler [80] also found that the scaled limit $C := \sigma X$ satisfies the distributional fixed-point equation

$$C \stackrel{d}{=} UC^{(1)} + (1 - U)C^{(2)} + \varphi(U) \quad (4.5)$$

with $U, C^{(1)}, C^{(2)}$ independent, U uniform on the unit interval, $C^{(1)}$ and $C^{(2)}$ having the same distribution as C and $\varphi(u) := 2u \log u + 2(1 - u) \log(1 - u) + 1$ for $u \in [0, 1]$. Lower and upper bounds for the rate of convergence in $X_n \rightarrow X$ have been studied for various metrics in Fill and Janson [24] and Neininger and Rüschemdorf [69].

We can now apply our framework to partly rederive these results on the convergence rates. Note that the quantities Y_n are bounded, thus L_s -integrable for any $s > 0$. Furthermore, with the recursive decomposition (4.1) of Y_n and the normalization (4.4), we are exactly in the situation of Theorem 3.1. To apply this theorem, we need to find an $0 < s \leq 3$ and a sequence $R(n) \rightarrow 0$ with (3.5)–(3.8). To bound the L_s norms $\|A_r^{(n)} - A_r^*\|_s$ and $\|b^{(n)} - b^*\|_s$ appearing in (3.7), we use Lemma 3.8 and, in view of (4.2) and (4.3), choose

$$\begin{aligned} f(n) &= 2n \log n + (2\gamma - 4)n, & e(n) &= \log n, \\ g(n) &= \sigma^2 n^2, & h(n) &= n \log n. \end{aligned}$$

With these functions, we obtain for the quantities defined in (3.21) that

$$\begin{aligned}\bar{A}_1^{(n)} &= \frac{I_n}{n}, & \bar{A}_2^{(n)} &= \frac{n-1-I_n}{n}, \\ \bar{b}^{(n)} &= \frac{1}{\sigma} \left(2 \frac{I_n}{n} \log \frac{I_n}{n} + 2 \frac{n-1-I_n}{n} \log \frac{n-1-I_n}{n} + \frac{n-1}{n} + O\left(\frac{\log n}{n}\right) \right).\end{aligned}$$

With the embedding $I_n = \lfloor nU \rfloor$ with U uniformly distributed on the unit interval $[0, 1]$, we have

$$A_1^* = U, \quad A_2^* = 1 - U, \quad b^* = \frac{1}{\sigma} (2U \log U + 2(1-U) \log(1-U) + 1) = \frac{1}{\sigma} \varphi(U)$$

with φ given in (4.5). We now apply Lemma 3.8 which allows us to bound the L_s norms $\|A_r^{(n)} - A_r^*\|_s$ and $\|b^{(n)} - b^*\|_s$ by the corresponding L_s norms $\|\bar{A}_r^{(n)} - A_r^*\|_s$, $\|\bar{b}^{(n)} - b^*\|_s$ and some additional error terms. Inserting the values of $\bar{A}_r^{(n)}$, A_r^* , $\bar{b}^{(n)}$ and b^* , we find for $r = 1, 2$ and any $s \geq 1$ that

$$\|\bar{A}_r^{(n)} - A_r^*\|_s = O\left(\frac{1}{n}\right)$$

and, together with Proposition 3.2 of Rösler [80],

$$\|\bar{b}^{(n)} - b^*\|_s = O\left(\frac{\log n}{n}\right).$$

Moreover, with $R(n) := \log n/n$, we have

$$\frac{h(n)}{g(n)} = O(R(n)) \quad \text{and} \quad \frac{e(n)}{g^{1/2}(n)} = O(R(n)),$$

thus Lemma 3.8 implies that condition (3.7) is satisfied for our choice of the sequence R . Also the technical conditions (3.5) and (3.6) are clearly satisfied, since I_n is uniform on $\{0, \dots, n-1\}$. To verify condition (3.8), we use Lemma 3.8 again: For the term $T(n)$ given there, we find $T(n) = O(\log(n)/n) \rightarrow 0$ and for the first term on the right hand side of (3.25), we obtain

$$\sum_{r=1}^2 \mathbb{E} \left[\frac{g^{s/2}(I_r^{(n)}) R(I_r^{(n)})}{g^{s/2}(n) R(n)} \right] = \sum_{r=1}^2 \mathbb{E} \left[\left(\frac{I_r^{(n)}}{n} \right)^{s-1} \frac{\log I_r^{(n)}}{\log n} \right].$$

Note that the latter expression has a limit superior of less than 1 if and only if $s > 2$. Hence, Theorem 3.1 is applicable for $2 < s \leq 3$ and yields that, as $n \rightarrow \infty$,

$$\zeta_s(X_n, X) = O\left(\frac{\log n}{n}\right). \quad (4.6)$$

The bound (4.6) had previously been shown for $s = 3$ by Neininger and Rüschemdorf [69], where also the optimality of the order was shown, i.e., that $\zeta_3(X_n, X) = \Theta(\log(n)/n)$.

In Section 4.4, we additionally discuss bounds on rates of convergence for various cost measures of the related Quickselect algorithm under different models for the rank to be selected.

4.1.2 Maxima in right triangles

As a second example, we consider the number Y_n of maxima of n independent, uniform samples in a right triangle in \mathbb{R}^2 with corners $(0, 0)$, $(1, 0)$ and $(0, 1)$, see Bai, Hwang, Liang and Tsai [3]. For a set S of points in \mathbb{R}^2 , we say that a point $(x_1, x_2) \in S$ is *maximal in S* if there is no other point $(y_1, y_2) \in S$ with $y_1 \geq x_1$ and $y_2 \geq x_2$, and refer to the set of maximal points in S as *maxima of S* , see Figure 4.1 for an example.

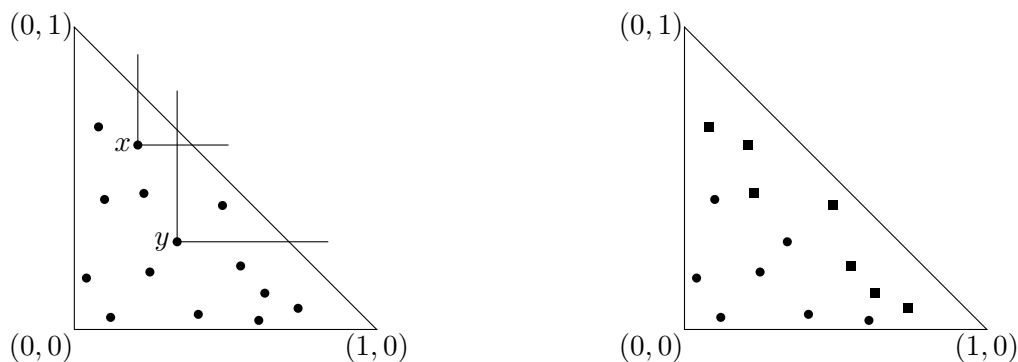


FIGURE 4.1: On the left: x is a maximal point, y is not (since its upper right quadrant is non-empty). On the right: the maxima of the sample set, marked by squares.

To apply our framework, the first step is to find a recursive decomposition of the form (3.1) for the number Y_n of maxima. To this end, we denote by (U_n, V_n) the point that maximizes the sum of the components in the sample of the n points and follow the strategy of Bai et al. [3, Proposition 1] who divided the original triangle into smaller parts as shown in Figure 4.2.

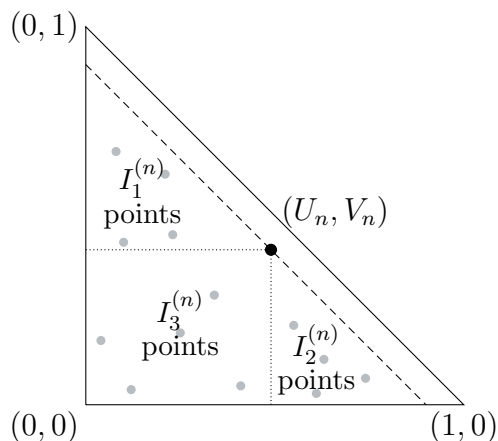


FIGURE 4.2: Division of the original triangle into smaller regions at the point (U_n, V_n) maximizing the sum of the components.

Denoting by $I_1^{(n)}$, $I_2^{(n)}$ and $I_3^{(n)}$ the numbers of points in the respective regions (see Figure 4.2), the number of maxima satisfies $Y_0 = 0$ and the recursion

$$Y_n \stackrel{d}{=} Y_{I_1^{(n)}}^{(1)} + Y_{I_2^{(n)}}^{(2)} + 1, \quad n \geq 1,$$

with independence and self-similarity conditions as in (3.1). The sizes $I_1^{(n)}$ and $I_2^{(n)}$ of the two subproblems can be described as the first two components of a mixed multinomial distribution: Given $(U_n, V_n) = (u, v)$, the vector $(I_1^{(n)}, I_2^{(n)}, I_3^{(n)})$ of the subproblem sizes is multinomially

$$\text{Mult}\left(n-1, \frac{u^2}{(u+v)^2}, \frac{v^2}{(u+v)^2}, \frac{2uv}{(u+v)^2}\right)$$

distributed. Note that the fraction $\frac{U_n}{U_n+V_n}$ is uniformly distributed on the unit interval. Hence, the vector of the subproblem sizes has a mixed multinomial distribution

$$\mathcal{L}((I_1^{(n)}, I_2^{(n)}, I_3^{(n)})) = \text{Mult}(n-1, U^2, (1-U)^2, 2U(1-U))$$

with U uniformly on $[0, 1]$.

It is known that the mean and variance of Y_n satisfy (see Bai et al. [3, Theorem 3])

$$\begin{aligned} \mathbb{E}[Y_n] &= \sqrt{\pi n} + O(1), \\ \text{Var}(Y_n) &= \sigma^2 \sqrt{n} - \frac{\pi}{4} + O(n^{-1/2}), \end{aligned}$$

where $\sigma^2 = (2 \log 2 - 1)\sqrt{\pi} > 0$. Furthermore, they have shown by the method of moments that the distribution of the normalized quantities $\frac{Y_n - \sqrt{\pi n}}{\sigma n^{1/4}}$ is asymptotically normal, see also Neininger and Rüschemdorf [70] for a different argument via contraction method. Bai, Hwang and Tsai [4, Theorem 1] improved this result by deriving the convergence rate $O(n^{-1/4})$ for the Kolmogorov metric in the central limit theorem for the number of maxima. We rederive this result for the Zolotarev metric ζ_3 by applying our refined Theorem 3.4 and obtain the following result.

Theorem 4.1. *The number Y_n of maxima of n independent, uniform samples in a right triangle with corners $(0, 0)$, $(1, 0)$ and $(0, 1)$ satisfies, as $n \rightarrow \infty$,*

$$\zeta_3\left(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \mathcal{N}(0, 1)\right) = O(n^{-1/4}).$$

Proof. As $\text{Var}(Y_n)$ is positive for $n \geq 2$, we introduce the normalized quantities $X_0 := X_1 := 0$ and

$$X_n := \frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \quad n \geq 2,$$

which satisfy the modified recursion (3.3)

$$X_n \stackrel{d}{=} \sum_{r=1}^2 A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq 1,$$

with $A_1^{(n)}$, $A_2^{(n)}$ and $b^{(n)}$ defined in (3.4). For these coefficients, we find $A_1^{(n)} \rightarrow U^{1/2} =: A_1^*$, $A_2^{(n)} \rightarrow (1 - U)^{1/2} =: A_2^*$ and $b^{(n)} \rightarrow 0 =: b^*$ in ℓ_3 (see Neininger and Rüschemdorf [70]), thus we are in the situation of Theorem 3.4. To apply this theorem, we have to estimate the orders of $\|(A_1^{(n)})^2 + (A_2^{(n)})^2 - 1\|_{3/2}$ and $\|b^{(n)}\|_3$. For this, we use Lemma 3.8 with

$$\begin{aligned} f(n) &= \sqrt{\pi n}, & e(n) &= 1, \\ g(n) &= \sigma^2 \sqrt{n}, & h(n) &= 1. \end{aligned}$$

With these functions, Lemma 3.8 yields

$$\left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2} \leq \left\| \sum_{r=1}^2 (\bar{A}_r^{(n)})^2 - 1 \right\|_{3/2} + O\left(\frac{1}{n^{1/2}}\right)$$

and

$$\|b^{(n)}\|_3 \leq \|\bar{b}^{(n)}\|_3 + O\left(\frac{1}{n^{1/2}} + \frac{1}{n^{1/4}}\right),$$

where the quantities $\bar{A}_1^{(n)}$, $\bar{A}_2^{(n)}$ and $\bar{b}^{(n)}$ are defined in (3.21):

$$\bar{A}_r^{(n)} = \left(\frac{I_r^{(n)}}{n}\right)^{1/4}, \quad \bar{b}^{(n)} = \frac{1}{\sigma n^{1/4}} \left(1 - \sqrt{\pi n} + \sum_{r=1}^2 \sqrt{\pi I_r^{(n)}}\right).$$

Hence, we have

$$\left\| \sum_{r=1}^2 (\bar{A}_r^{(n)})^2 - 1 \right\|_{3/2} = \frac{1}{n^{1/2}} \left\| (I_1^{(n)})^{1/2} + (I_2^{(n)})^{1/2} - n^{1/2} \right\|_{3/2} = O\left(\frac{1}{n^{1/2}}\right)$$

and

$$\|\bar{b}^{(n)}\|_3 = \frac{1}{\sigma n^{1/4}} \left\| 1 + \sqrt{\pi} \left((I_1^{(n)})^{1/2} + (I_2^{(n)})^{1/2} - n^{1/2} \right) \right\|_3 = O\left(\frac{1}{n^{1/4}}\right),$$

where the last step in each case follows from Lemma 4.2 below. Obviously, condition (3.6) is satisfied and we have

$$\left\| \mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)} \right\|_3^3 = O\left(\text{Var}(Y_n)^{-3/2} \mathbb{P}(I_r^{(n)} < \ell)\right) = O(n^{-3/4})$$

for any $\ell \in \mathbb{N}$ and $r = 1, 2$. To sum up, our findings so far are

$$\left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2}^3 + \|b^{(n)}\|_3^3 + \sum_{r=1}^2 \left\| \mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)} \right\|_3^3 = O\left(\frac{1}{n^{3/4}}\right)$$

for any $\ell \in \mathbb{N}$ and as $n \rightarrow \infty$. Consequently, it remains to check condition (3.16)—which is not satisfied for $R(n) = n^{-1/4}$. However, we can use the weakened condition stated in

Remark 3.5 with $R(n) = n^{-1/4}$ and $\tilde{R}(n) = n^{-3/4}$. As described there, it is sufficient to show that

$$\mathbb{E} \left[\sum_{r=1}^2 \frac{R(I_r^{(n)})}{R(n)} (A_r^{(n)})^3 \right] \leq 1 - \varepsilon(n) \quad (4.7)$$

for all large n with some positive $\varepsilon(n) = \Omega(n^{-1/2})$. However, applying Lemma 3.8 with $g(n) = \sigma^2 \sqrt{n}$ and $h(n) = 1$ only results in showing

$$\mathbb{E} \left[\sum_{r=1}^2 \frac{R(I_r^{(n)})}{R(n)} (A_r^{(n)})^3 \right] \leq n^{-1/2} \mathbb{E} \left[\sum_{r=1}^2 (I_r^{(n)})^{1/2} \right] + O(n^{-1/2}) \leq 1 + O(n^{-1/2}).$$

Considering that, we choose $g(n) = \sigma^2 \sqrt{n} - \frac{\pi}{4}$ and $h(n) = n^{-1/2}$. With this choice, we obtain $T(n) = O(1/n)$ for the quantity $T(n)$ defined in Lemma 3.8 (note that we have $\mathbb{P}(I_r^{(n)} < \ell) = O(n^{-1/2})$ for $\ell \in \mathbb{N}$ and $r = 1, 2$). Consequently, the estimate (3.25), together with Lemma 4.5 below, yields

$$\begin{aligned} \mathbb{E} \left[\sum_{r=1}^2 \frac{R(I_r^{(n)})}{R(n)} (A_r^{(n)})^3 \right] &\leq \mathbb{E} \left[\sum_{r=1}^2 \mathbf{1}_{\{I_r^{(n)} \geq 2\}} \frac{(\sigma^2 (I_r^{(n)})^{1/2} - \pi/4)^{3/2}}{(\sigma^2 n^{1/2} - \pi/4)^{3/2}} \left(\frac{I_r^{(n)}}{n} \right)^{-1/4} \right] + O\left(\frac{1}{n}\right) \\ &\leq 1 - \frac{3\pi}{8\sigma^2} n^{-1/2} + o(n^{-1/2}), \end{aligned}$$

and the assertion follows. \square

The preceding proof shows that a straightforward application of our theorems only provides the rate $O(n^{-1/4+\varepsilon})$ for any $\varepsilon > 0$. However, by making use of the weaker condition stated in Remark 3.5, we were able to avoid the ε in the exponent of the rate.

The rest of this section states some basic facts about mixed binomial/trinomial distributions which were used in the preceding proof. The following statement was applied to determine the convergence rate of the coefficients.

Lemma 4.2. *Let $(I_1^{(n)}, I_2^{(n)}, I_3^{(n)})$ be a random vector with mixed trinomial distribution $\text{Mult}(n-1, U^2, (1-U)^2, 2U(1-U))$, where U is uniform on $[0, 1]$. Then, for any $s > 0$ and as $n \rightarrow \infty$, we have*

$$\left\| (I_1^{(n)})^{1/2} + (I_2^{(n)})^{1/2} - n^{1/2} \right\|_s = O(1).$$

Proof. Proposition 4 in Bai, Hwang and Tsai [4] states that for any non-negative integer r ,

$$\mathbb{E} \left[\left| (I_1^{(n)})^{1/2} + (I_2^{(n)})^{1/2} - n^{1/2} \right|^{2r} \right] \leq \frac{(2r)! 4^r e^{2r}}{(\log(2r+1))^{2r}}.$$

The corresponding result for arbitrary exponents $s > 0$ follows from this by Jensen's inequality. \square

The following two lemmas serve as a basis for the subsequent Lemma 4.5 which was used in the final step of the proof of Theorem 4.1.

Lemma 4.3. *For a random variable $I_1^{(n)}$ with mixed binomial distribution $\text{Bin}(n-1, U^2)$, where U is uniform on $[0, 1]$, we have that $\mathbb{E}[(I_1^{(n)})^{1/2}] \leq \frac{1}{2}n^{1/2}$ for all $n \geq 1$.*

Proof. Denoting by B the Euler beta function, we obtain by conditioning on U

$$\begin{aligned} \mathbb{E}[(I_1^{(n)})^{1/2}] &= \sum_{k=1}^{n-1} k^{1/2} \binom{n-1}{k} \int_0^1 u^{2k} (1-u^2)^{n-1-k} du \\ &= \frac{1}{2} \sum_{k=1}^{n-1} k^{1/2} \binom{n-1}{k} B\left(k + \frac{1}{2}, n-k\right) \\ &= \frac{4^n}{2n \binom{2n}{n}} \sum_{k=1}^{n-1} k^{1/2} 4^{-k} \binom{2k}{k}. \end{aligned}$$

The assertion now follows inductively. For this, note that the statement is obviously true for $n = 1$ and that, for any $n \geq 1$, we have

$$\begin{aligned} \mathbb{E}[(I_1^{(n+1)})^{1/2}] &= \frac{4^{n+1}}{2(n+1) \binom{2(n+1)}{n+1}} \left(\sum_{k=1}^{n-1} k^{1/2} 4^{-k} \binom{2k}{k} + n^{1/2} 4^{-n} \binom{2n}{n} \right) \\ &\leq \frac{4^{n+1}}{2(n+1) \binom{2(n+1)}{n+1}} \left(n^{3/2} 4^{-n} \binom{2n}{n} + n^{1/2} 4^{-n} \binom{2n}{n} \right) \\ &= \frac{n+1}{2n+1} n^{1/2}, \end{aligned}$$

which is smaller than $\frac{1}{2}(n+1)^{1/2}$ and finishes the proof. \square

Lemma 4.4. *For a random variable $I_1^{(n)}$ with mixed binomial distribution $\text{Bin}(n-1, U^2)$, where U is uniform on $[0, 1]$, we have that $\mathbb{E}[\mathbf{1}_{\{I_1^{(n)} \geq 1\}} (I_1^{(n)})^{-1/2}] \rightarrow 0$ as $n \rightarrow \infty$.*

Proof. Similarly as in the proof of the previous lemma, we obtain the representation

$$\mathbb{E}[\mathbf{1}_{\{I_1^{(n)} \geq 1\}} (I_1^{(n)})^{-1/2}] = \frac{4^n}{2n \binom{2n}{n}} \sum_{k=1}^{n-1} k^{-1/2} 4^{-k} \binom{2k}{k}.$$

We now use the bounds $\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq 2\sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ to obtain $\binom{2n}{n} \geq 4^{n-1} (\pi n)^{-1/2}$ and $\binom{2k}{k} \leq 2 \cdot 4^k (\pi k)^{-1/2}$, and therefore

$$\mathbb{E}[\mathbf{1}_{\{I_1^{(n)} \geq 1\}} (I_1^{(n)})^{-1/2}] \leq \frac{4}{\sqrt{n}} \sum_{k=1}^{n-1} \frac{1}{k} = O\left(\frac{\log n}{n^{1/2}}\right),$$

which is approaching 0 as $n \rightarrow \infty$. \square

Lemma 4.5. *Let $I_1^{(n)}$ denote some random variable with mixed binomial distribution $\text{Bin}(n-1, U^2)$, where U is uniform on $[0, 1]$. As $n \rightarrow \infty$, we have*

$$\mathbb{E} \left[\mathbf{1}_{\{I_1^{(n)} \geq 2\}} \frac{(\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{3/2}}{(\sigma^2 n^{1/2} - \pi/4)^{3/2}} \left(\frac{I_1^{(n)}}{n} \right)^{-1/4} \right] \leq \frac{1}{2} - \frac{3\pi}{16\sigma^2} n^{-1/2} + o(n^{-1/2}).$$

Proof. We start with a lower bound for the denominator: For $n \geq 2$, e.g., by the mean value theorem, we have

$$(\sigma^2 n^{1/2} - \pi/4)^{3/2} n^{-1/4} \geq \left((\sigma^2 n^{1/2})^{3/2} - \frac{3\pi}{8} (\sigma^2 n^{1/2})^{1/2} \right) n^{-1/4} = \sigma^3 n^{1/2} - \frac{3\pi\sigma}{8}.$$

Similarly, we continue by upper bounding the nominator: On the event $\{I_1^{(n)} \geq 2\}$, the term $\sigma^2(I_1^{(n)})^{1/2} - \pi/4$ is positive and we find that

$$\begin{aligned} (\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{3/2} &\leq (\sigma^2(I_1^{(n)})^{1/2})^{3/2} - \frac{3\pi}{8} (\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{1/2} \\ &= \sigma^3(I_1^{(n)})^{3/4} - \frac{3\pi}{8} (\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{1/2} \\ &\leq \sigma^3(I_1^{(n)})^{3/4} - \frac{3\pi\sigma}{8} (I_1^{(n)})^{1/4} + \frac{3\pi^2}{64} (\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{-1/2}. \end{aligned}$$

Thus, we obtain with Lemma 4.3 and Lemma 4.4

$$\begin{aligned} \mathbb{E} \left[\mathbf{1}_{\{I_1^{(n)} \geq 2\}} (\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{3/2} (I_1^{(n)})^{-1/4} \right] \\ \leq \mathbb{E} \left[\sigma^3(I_1^{(n)})^{1/2} - \frac{3\pi\sigma}{8} + \mathbf{1}_{\{I_1^{(n)} \geq 2\}} \frac{3\pi^2}{64} (\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{-1/2} (I_1^{(n)})^{-1/4} \right] \\ \leq \frac{1}{2} \sigma^3 n^{1/2} - \frac{3\pi\sigma}{8} + o(1). \end{aligned}$$

Putting these estimates together, we find

$$\begin{aligned} \mathbb{E} \left[\mathbf{1}_{\{I_1^{(n)} \geq 2\}} \frac{(\sigma^2(I_1^{(n)})^{1/2} - \pi/4)^{3/2}}{(\sigma^2 n^{1/2} - \pi/4)^{3/2}} \left(\frac{I_1^{(n)}}{n} \right)^{-1/4} \right] &\leq \frac{\frac{1}{2} \sigma^3 n^{1/2} - \frac{3\pi\sigma}{8} + o(1)}{\sigma^3 n^{1/2} - \frac{3\pi\sigma}{8}} \\ &= \frac{1}{2} - \frac{3\pi}{16\sigma^2} n^{-1/2} + o(n^{-1/2}), \end{aligned}$$

which is the assertion. \square

4.1.3 The size of m -ary search trees

In this section, we consider m -ary search trees, a class of data structures generalizing the concept of binary search trees. In short, an m -ary search tree is constructed similarly to a binary search tree, with the difference that each node in the m -ary search tree has the capacity to contain $m - 1$ keys. More specifically, for $m \geq 2$ and a given input sequence of n distinct keys, an m -ary search tree is constructed as follows (see, e.g., Mahmoud [59]): If $n < m$, then the tree consists of a single node containing the keys in increasing order. Otherwise, the first $m - 1$ keys are inserted into the root (in increasing order) and are used to split the remaining keys up into m different groups:

- ◇ Those keys with values smaller than the smallest key in the root go to the first subtree of the root.
- ◇ Those keys with values between the $(k - 1)$ -th smallest and k -th smallest key in the root go to the k -th subtree of the root ($k = 2, \dots, m - 1$).
- ◇ Those keys with values larger than the largest key in the root go to the m -th subtree of the root.

With this procedure, we obtain m subtrees of the root which are, by definition, themselves m -ary search trees. To illustrate the construction of m -ary search trees, Figure 4.3 shows an example for $m = 4$ and $n = 11$. We call the total number of nodes of an m -ary search tree the *size* (also: *space requirement*) of this tree. The size of the tree in Figure 4.3 is 6.

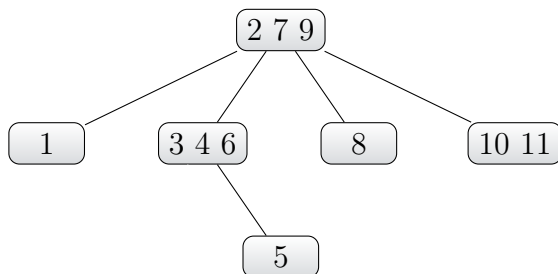


FIGURE 4.3: An m -ary search tree for $m = 4$, constructed from the input sequence $(9, 2, 7, 4, 3, 10, 8, 6, 11, 1, 5)$ of length $n = 11$.

We now denote by Y_n the size of an m -ary search tree constructed from an input sequence consisting of n randomly permuted distinct numbers. Since $Y_n = n$ is deterministic in the binary search tree case ($m = 2$), we restrict our analysis to $m \geq 3$. In this case, $Y_0 = 0$, $Y_1 = \dots = Y_{m-1} = 1$ and Y_n satisfies the recursion (see, e.g., Neininger and Rüschemdorf [70])

$$Y_n \stackrel{d}{=} \sum_{r=1}^m Y_{I_r^{(n)}}^{(r)} + 1, \quad n \geq m,$$

with independence and self-similarity conditions as in (3.1). Thus, the sequence $(Y_n)_{n \geq 0}$ is an instance of recurrence (3.1) with $K = m$, $A_1(n) = \cdots = A_m(n) = 1$, $n_0 = m$ and $b_n = 1$. For a representation of $I^{(n)} = (I_1^{(n)}, \dots, I_m^{(n)})$, we define for independent, identically unif $[0, 1]$ distributed random variables U_1, \dots, U_{m-1} their spacings in $[0, 1]$ by

$$S_1 = U_{(1)}, S_2 = U_{(2)} - U_{(1)}, \dots, S_m = 1 - U_{(m-1)},$$

where $U_{(1)}, \dots, U_{(m-1)}$ denote the order statistics of U_1, \dots, U_{m-1} . Then, $I^{(n)}$ has the mixed multinomial distribution:

$$\mathcal{L}(I^{(n)}) = \text{Mult}(n - m + 1, S_1, \dots, S_m).$$

By this we mean that given $(S_1, \dots, S_m) = (s_1, \dots, s_m)$, the vector $I^{(n)}$ is multinomially $\text{Mult}(n - m + 1, s_1, \dots, s_m)$ distributed. Expectations and variances for Y_n have been studied by, among others, Knuth [56, § 6.2.4], Mahmoud and Pittel [63], Mahmoud [59] and Chern and Hwang [12]. We have

$$\mathbb{E}[Y_n] = \mu n + O(1 + n^{\alpha-1}), \quad m \geq 3, \quad (4.8)$$

$$\text{Var}(Y_n) = \sigma^2 n + O(1 + n^{2\alpha-2}), \quad 3 \leq m \leq 26. \quad (4.9)$$

Here, the constants $\mu, \sigma^2 > 0$ and α depend on m with $\alpha < 1$ for $m \leq 13$, $1 < \alpha < 4/3$ for $14 \leq m \leq 19$, and $4/3 < \alpha < 3/2$ for $20 \leq m \leq 26$, see Table 4.1.

m	3...13	14...19	20...26	27...
α	-3...0.95	1.04...1.31	1.34...1.49	1.51...

TABLE 4.1: Approximate values of $\alpha = \alpha_m$ depending on m (see, e.g., Mahmoud [59, Table 3.1]).

It is known that Y_n normalized by mean and standard deviation converges in distribution to the standard normal distribution for $3 \leq m \leq 26$ (this was shown by an inductive approximation approach in Mahmoud and Pittel [63] and Lew and Mahmoud [58], by the method of moments in Chern and Hwang [12] and by the contraction method in Neininger and Rüschemdorf [70]), whereas the standardized sequence has no weak limit for $m > 26$ due to dominant periodicities, see Chern and Hwang [12]. Moreover, the rate of convergence in the central limit law for $3 \leq m \leq 26$ for the Kolmogorov metric was identified by Hwang [43]: By a refinement to the method of moments, he derived the rate $O(n^{-1/2})$ for $3 \leq m \leq 19$ and $O(n^{-3(3/2-\alpha)})$ for $20 \leq m \leq 26$. The following result shows that the application of our Theorem 3.4 yields the same rates of convergence for the Zolotarev metric ζ_3 (up to an ε for $3 \leq m \leq 19$).

Theorem 4.6. *The size Y_n of an m -ary search tree constructed from n uniformly permuted distinct numbers satisfies, for any $\varepsilon > 0$ and as $n \rightarrow \infty$,*

$$\zeta_3\left(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \mathcal{N}(0, 1)\right) = \begin{cases} O(n^{-1/2+\varepsilon}), & 3 \leq m \leq 19, \\ O(n^{-3(3/2-\alpha)}), & 20 \leq m \leq 26. \end{cases}$$

Proof. To apply Theorem 3.4, we have to estimate the orders of $\|\sum_{r=1}^m (A_r^{(n)})^2 - 1\|_{3/2}$ and $\|b^{(n)}\|_3$ with $A_r^{(n)}$ and $b^{(n)}$ defined in (3.4). For this, we apply Lemma 3.8. From (4.8) and (4.9), we obtain that we can choose

$$\begin{aligned} f(n) &= \mu n, & e(n) &= 1 \vee n^{\alpha-1}, \\ g(n) &= \sigma^2 n, & h(n) &= 1 \vee n^{2(\alpha-1)}, \end{aligned}$$

for the quantities appearing in Lemma 3.8. Hence, with $\bar{A}_r^{(n)}$ given in (3.21), we obtain

$$\left\| \sum_{r=1}^m (\bar{A}_r^{(n)})^2 - 1 \right\|_{3/2} = \left\| \sum_{r=1}^m \frac{I_r^{(n)}}{n} - 1 \right\|_{3/2} = \frac{m-1}{n} = O\left(\frac{1}{n}\right)$$

and $O(h(n)/g(n)) = O(n^{-(1 \wedge (3-2\alpha))})$. Consequently, Lemma 3.8 implies

$$\left\| \sum_{r=1}^m (A_r^{(n)})^2 - 1 \right\|_{3/2}^{3/2} = O(n^{-((3/2) \wedge (3(3/2-\alpha)))}).$$

Similarly, we obtain for $\bar{b}^{(n)}$ given in (3.21)

$$\|\bar{b}^{(n)}\|_3 = \frac{1}{\sigma\sqrt{n}} \left\| 1 - \mu n + \sum_{r=1}^m \mu I_r^{(n)} \right\|_3 = \frac{1}{\sigma\sqrt{n}} \|1 - \mu(m-1)\|_3 = O(n^{-1/2})$$

and $O(e(n)/g^{1/2}(n)) = O(n^{-((1/2) \wedge (3/2-\alpha))})$. From this, it follows by Lemma 3.8 that

$$\|b^{(n)}\|_3^3 = O(n^{-((3/2) \wedge (3(3/2-\alpha)))}).$$

Hence, the rate of convergence in condition (3.15) is of order $O(n^{-((3/2) \wedge (3(3/2-\alpha)))})$ and it remains to check condition (3.16): For $3 \leq m \leq 19$ and $R(n) = n^{-1/2+\varepsilon}$, Lemma 3.8 (in combination with the convergence $I_r^{(n)}/n \rightarrow S_r$ and dominated convergence) implies, as $n \rightarrow \infty$,

$$\sum_{r=1}^m \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} |A_r^{(n)}|^3 \right] \leq \sum_{r=1}^m \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \left(\frac{I_r^{(n)}}{n} \right)^{3/2} \right] + o(1) \rightarrow \sum_{r=1}^m \mathbb{E} [(S_r)^{1+\varepsilon}] < 1.$$

Analogously, for $20 \leq m \leq 26$ and $R(n) = n^{-3(3/2-\alpha)}$, condition (3.16) is satisfied and the assertion follows. \square

Remark 4.7. Using Theorem 3.1 instead of Theorem 3.4 in the latter proof is also possible but leads to a bound $O(n^{-(3/2-\alpha)})$ for $20 \leq m \leq 26$, missing the factor 3 appearing in Theorem 4.6.

4.1.4 The number of leaves in d -dimensional quadtrees

As a last example in this section, we now discuss rates of convergence for the number of leaves in d -dimensional random point quadtrees, where a similar behavior as in Theorem 4.6 appears.

The quadtree structure is due to Finkel and Bentley [25] and can be considered as a generalization of the binary search tree to multidimensional data. Given a set x_1, \dots, x_n of points in the unit hypercube $[0, 1]^d$, the corresponding quadtree is constructed as follows (see, e.g., Mahmoud [59] or Chern, Fuchs and Hwang [11]): The first point x_1 is placed at the root and splits the unit hypercube into 2^d smaller regions (quadrants), each corresponding to one of the 2^d subtrees of the root. Next, the remaining points x_2, \dots, x_n are assigned to the respective quadrants (subtrees) and the 2^d subtrees are then built recursively by the same procedure. An example of this construction for 2-dimensional data is given in Figure 4.4. If the points are uniformly and independently chosen from $[0, 1]^d$, the associated quadtree is called a d -dimensional random point quadtree.

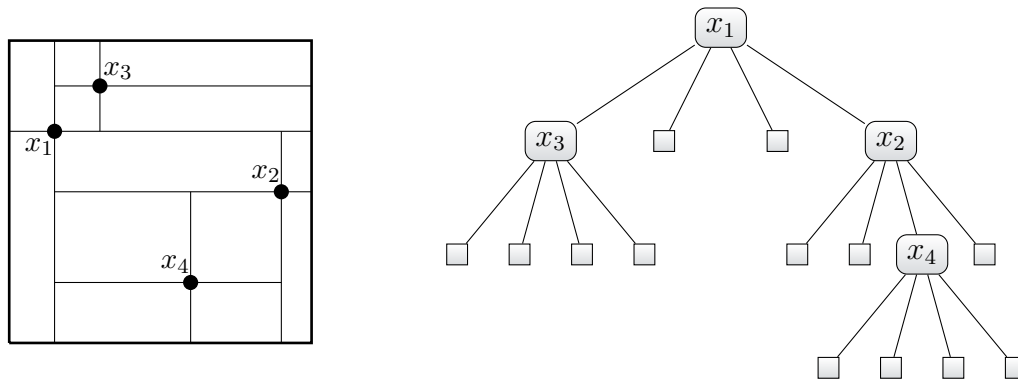


FIGURE 4.4: A 2-dimensional quadtree of size $n = 4$, constructed from the points x_1, x_2, x_3, x_4 . Each box (external node) in the tree corresponds to one of the regions on the left.

We now denote by Y_n the number of leaves (i.e., the number of nodes without children) in a d -dimensional random point quadtree with n nodes. The tree in Figure 4.4, for example, has two leaves (x_3 and x_4). The quantity Y_n satisfies $Y_0 = 0$, $Y_1 = 1$ and

$$Y_n \stackrel{d}{=} \sum_{r=1}^{2^d} Y_{I_r^{(n)}}^{(r)}, \quad n \geq 2,$$

with the usual independence and self-similarity assumptions, i.e., Y_n satisfies recurrence (3.1) with $n_0 = 2$, $K = 2^d$, $A_1(n) = \dots = A_K(n) = 1$ and $b_n = 0$. The vector $I^{(n)}$ has a mixed multinomial distribution,

$$\mathcal{L}(I^{(n)}) = \text{Mult}(n-1, \langle U \rangle_1, \dots, \langle U \rangle_{2^d}),$$

where U is uniformly distributed over the hypercube $[0, 1]^d$ and $(\langle U \rangle_1, \dots, \langle U \rangle_{2^d})$ is the vector of the volumes of the 2^d quadrants which are generated by cutting $[0, 1]^d$ with the d hyperplanes through U perpendicular to the axes.

Expectations, variances and limit laws for Y_n have been studied by Flajolet, Labelle, Laforest and Salvy [27] and Chern, Fuchs and Hwang [11] for general $d \geq 1$. The case $d = 1$ is the binary search tree case, see Devroye [16] and Flajolet, Gourdon and Martínez [26]. Here, we focus on the regime $1 \leq d \leq 8$, where a central limit theorem holds. For $d > 8$, more pronounced periodicities prevent asymptotic normality; see Chern, Fuchs and Hwang [11] and Fuchs, Müller and Sulzbach [32] for asymptotic distributional results for $d > 8$. We have, see [16, 27, 11],

$$\mathbb{E}[Y_n] = \mu n + \begin{cases} O(1), & d = 1, \\ O(n^\alpha + n^\delta), & d \geq 2, \end{cases} \quad (4.10)$$

with a positive constant μ depending on d , $\alpha = 2 \cos(2\pi/d) - 1$ and arbitrary $\delta > 0$. The values of α are displayed in Table 4.2.

d	2	3	4	5	6	7	8
α	-3	-2	-1	-0.38	0	0.24	0.41

TABLE 4.2: Approximate values of $\alpha = \alpha_d = 2 \cos(2\pi/d) - 1$ depending on d .

The variance satisfies, see Chern, Fuchs and Hwang [11],

$$\text{Var}(Y_n) = \sigma^2 n + \begin{cases} O(n^{2/3}), & 1 \leq d \leq 7, \\ O(n^{2(\sqrt{2}-1)}), & d = 8, \end{cases} \quad (4.11)$$

with $\sigma^2 > 0$ depending on d . A rate of convergence in the central limit law for $1 \leq d \leq 8$ for the Kolmogorov metric has been identified by Chern, Fuchs and Hwang [11] showing the order $O(n^{-1/2})$ when $1 \leq d \leq 7$ and the order $O(n^{-3(3/2-\sqrt{2})}) \approx O(n^{-0.257})$ when $d = 8$. By applying Theorem 3.4, we rederive this result for the Zolotarev metric ζ_3 with an additional ε in the exponent for $2 \leq d \leq 7$.

Theorem 4.8. *The number Y_n of leaves in a d -dimensional random point quadtree with n points inserted satisfies, as $n \rightarrow \infty$,*

$$\zeta_3\left(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \mathcal{N}(0, 1)\right) = \begin{cases} O(n^{-1/2}), & d = 1, \\ O(n^{-1/2+\varepsilon}), & 2 \leq d \leq 7, \\ O(n^{-3(3/2-\sqrt{2})}), & d = 8, \end{cases}$$

Proof. For $2 \leq d \leq 8$, the proof follows along the same lines as the proof of Theorem 4.6, and is omitted here. For the binary search tree case $d = 1$, we can use the exact mean and variance of Y_n given in Devroye [16] in combination with Remark 3.5 to avoid the ε in the exponent of the rate. \square

4.2 Periodic functions in mean and variance

We now discuss some examples where the asymptotic expansions of the mean and the variance include periodic functions instead of fixed constants. This is the case for several quantities in binomial splitting processes such as tries, PATRICIA tries and digital search trees (see, e.g., Drmota [19], Knuth [56] and Mahmoud [59] for an overview of those structures). Throughout this section, let $(Y_n)_{n \geq 0}$ be an L_3 -integrable sequence satisfying the recursion

$$Y_n \stackrel{d}{=} Y_{I_1^{(n)}}^{(1)} + Y_{I_2^{(n)}}^{(2)} + b_n, \quad n \geq n_0, \quad (4.12)$$

with $(I^{(n)}, b_n)$, $(Y_n^{(1)})_{n \geq 0}$ and $(Y_n^{(2)})_{n \geq 0}$ independent and $(Y_n^{(r)})_{n \geq 0} \stackrel{d}{=} (Y_n)_{n \geq 0}$ for $r = 1, 2$. Apart from that, we assume that $I_1^{(n)}$ has essentially the binomial distribution $\text{Bin}(n, 1/2)$ and that $I_1^{(n)} + I_2^{(n)}$ is essentially n . In concrete terms, this means that $I_1^{(n)}$ is binomially $\text{Bin}(n - k, 1/2)$ distributed and $I_2^{(n)} = n - k - I_1^{(n)}$ for some fixed $k \in \mathbb{N}_0$. Extensions to the asymmetric case with success probability $p \neq 1/2$ are possible. Mostly, these binomial recurrences are asymptotically normally distributed, see Hwang, Fuchs and Zacharovas [44], Jacquet and Szpankowski [47, 48], Neininger and Rüschemdorf [70] and Schachinger [85] for some examples.

Our first theorem covers the case of linear mean and variance, i.e., we assume that, as $n \rightarrow \infty$,

$$\mathbb{E}[Y_n] = nP_1(\log_2 n) + O(1), \quad (4.13)$$

$$\text{Var}(Y_n) = nP_2(\log_2 n) + O(1), \quad (4.14)$$

for some differentiable and 1-periodic functions P_1, P_2 with Lipschitz continuous derivatives and $P_2 > 0$. Possible applications would start with the analysis of the number of internal nodes of a trie for n keys in the symmetric Bernoulli model (see, e.g., Fuchs, Hwang and Zacharovas [31], Jacquet and Régnier [46], Neininger and Rüschemdorf [70] and the references therein for the mean, variance and limit laws) or the number of leaves in a random digital search tree of n keys (see, e.g., Flajolet and Sedgewick [28], Hwang, Fuchs and Zacharovas [44] and Kirschenhofer and Prodinger [52] for the mean and variance of the number of leaves). For a sequence $(Y_n)_{n \geq 0}$ satisfying the above assumptions, the following theorem provides a bound on the rate of convergence in the Zolotarev ζ_3 metric.

Theorem 4.9. *Let $(Y_n)_{n \geq 0}$ be L_3 -integrable and satisfy recursion (4.12) with $\|b_n\|_3 = O(1)$, (4.13) and (4.14). Then, for any $\varepsilon > 0$ and as $n \rightarrow \infty$, we have*

$$\zeta_3\left(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \mathcal{N}(0, 1)\right) = O(n^{-1/2+\varepsilon}).$$

Proof. Since P_2 is positive, continuous and 1-periodic, there exists some $n_1 \in \mathbb{N}_0$ such that $\text{Var}(Y_n) > 0$ for $n \geq n_1$. As in (3.2), we define the normalized quantities $X_n := Y_n - \mathbb{E}[Y_n]$

for $n < n_1$ and

$$X_n := \frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \quad n \geq n_1.$$

With $A_r^{(n)}$ and $b^{(n)}$ given in (3.4), we have the L_3 -convergences

$$A_1^{(n)} \rightarrow \frac{1}{\sqrt{2}}, \quad A_2^{(n)} \rightarrow \frac{1}{\sqrt{2}}, \quad b^{(n)} \rightarrow 0,$$

thus we are in the situation of Theorem 3.4. To apply this theorem, we have to estimate the orders of $\left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2}$ and $\|b^{(n)}\|_3$. For this purpose, we use Lemma 3.8 with

$$\begin{aligned} f(n) &= nP_1(\log_2 n), & e(n) &= 1, \\ g(n) &= nP_2(\log_2 n), & h(n) &= 1, \end{aligned}$$

where we use the convention $0 \cdot P_i(\log_2(0)) := 0$ for $i = 1, 2$. With these functions, we obtain

$$\left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2} \leq \left\| \sum_{r=1}^2 (\bar{A}_r^{(n)})^2 - 1 \right\|_{3/2} + O\left(\frac{1}{n}\right)$$

and

$$\|b^{(n)}\|_3 \leq \|\bar{b}^{(n)}\|_3 + O\left(\frac{1}{n} + \frac{1}{n^{1/2}}\right),$$

where the quantities $\bar{A}_r^{(n)}$, $r = 1, 2$, and $\bar{b}^{(n)}$ are defined in (3.21):

$$\begin{aligned} \bar{A}_r^{(n)} &= \sqrt{\frac{I_r^{(n)} P_2(\log_2 I_r^{(n)})}{n P_2(\log_2 n)}}, \\ \bar{b}^{(n)} &= \frac{1}{\sqrt{n P_2(\log_2 n)}} \left(b_n - n P_1(\log_2 n) + \sum_{r=1}^2 I_r^{(n)} P_1(\log_2 I_r^{(n)}) \right). \end{aligned}$$

Hence, we have

$$\left\| \sum_{r=1}^2 (\bar{A}_r^{(n)})^2 - 1 \right\|_{3/2} = \frac{1}{n P_2(\log_2 n)} \left\| \sum_{r=1}^2 I_r^{(n)} P_2(\log_2 I_r^{(n)}) - n P_2(\log_2 n) \right\|_{3/2} = O\left(\frac{1}{n}\right)$$

and

$$\|\bar{b}^{(n)}\|_3 = \frac{1}{\sqrt{n P_2(\log_2 n)}} \left\| b_n + \sum_{r=1}^2 I_r^{(n)} P_1(\log_2 I_r^{(n)}) - n P_1(\log_2 n) \right\|_3 = O\left(\frac{1}{n^{1/2}}\right),$$

where we used the assumption $\|b_n\|_3 = O(1)$ and Lemma 4.11 below. It is easily seen that the technical condition (3.6) is satisfied and that

$$\left\| \mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)} \right\|_3^3 = O\left(\text{Var}(Y_n)^{-3/2} \mathbb{P}(I_r^{(n)} < \ell)\right) = O(n^{-3/2})$$

for any $\ell \in \mathbb{N}$ and $r = 1, 2$. Hence, it remains to check condition (3.16): With Lemma 3.8 and $R(n) = n^{-1/2+\varepsilon}$, we have as $n \rightarrow \infty$

$$\begin{aligned} \sum_{r=1}^2 \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} |A_r^{(n)}|^3 \right] &\leq \sum_{r=1}^2 \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \left(\frac{I_r^{(n)} P_2(\log_2 I_r^{(n)})}{n P_2(\log_2 n)} \right)^{3/2} \right] + O\left(\frac{1}{n}\right) \\ &\rightarrow 2 \left(\frac{1}{2}\right)^{1+\varepsilon} < 1, \end{aligned}$$

and the assertion follows. \square

Remark 4.10. From the proof of Theorem 4.9, it follows that

$$\left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3 + \sum_{r=1}^2 \|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_3^3 = O\left(\frac{1}{n^{3/2}}\right)$$

for any $\ell \in \mathbb{N}$ and as $n \rightarrow \infty$. Thus, with Remark 3.5 (see also Section 4.1.2), it might be possible to remove the ε in the exponent of the rate in Theorem 4.9 in some concrete applications.

The following lemma was used to bound the convergence rates of the coefficients in the proof of Theorem 4.9. Note that a related bound can be found in Neininger and Rüschemdorf [70, Proof of Corollary 5.7], there however without a rate of convergence.

Lemma 4.11. *Let P be a differentiable, 1-periodic function with Lipschitz continuous derivative P' and B_n binomially $\text{Bin}(n, 1/2)$ distributed. Then we have, for any $s \geq 1$ and as $n \rightarrow \infty$,*

$$\|B_n P(\log_2 B_n) + (n - B_n) P(\log_2(n - B_n)) - n P(\log_2 n)\|_s = O(1).$$

Proof. Note that for $B_n \in \{0, n\}$, the integrand has value 0, which is why the following estimates are carried out on the event $\{0 < B_n < n\}$. We set $R_n = B_n - \frac{n}{2}$. By the mean value theorem, there exists some (random) ξ_n between $\frac{n}{2}$ and $B_n = \frac{n}{2} + R_n$ such that

$$\log_2 B_n = \log_2 \left(\frac{n}{2} + R_n \right) = \log_2 \left(\frac{n}{2} \right) + \frac{R_n}{\log 2 \xi_n}.$$

Using the 1-periodicity of P and once again the mean value theorem, there exists some (random) χ_n between $\log_2 n$ and $\log_2 n + \frac{R_n}{\log 2 \xi_n}$ such that

$$\begin{aligned} P(\log_2 B_n) &= P\left(\log_2 \left(\frac{n}{2}\right) + \frac{R_n}{\log 2 \xi_n}\right) = P\left(\log_2 n + \frac{R_n}{\log 2 \xi_n}\right) \\ &= P(\log_2 n) + \frac{R_n}{\log 2 \xi_n} P'(\chi_n). \end{aligned}$$

Analogously, there exists some $\tilde{\xi}_n$ between $\frac{n}{2}$ and $n - B_n = \frac{n}{2} - R_n$ and some $\tilde{\chi}_n$ between $\log_2 n$ and $\log_2 n - \frac{R_n}{\log 2 \tilde{\xi}_n}$ such that

$$P(\log_2(n - B_n)) = P\left(\log_2\left(\frac{n}{2} - R_n\right)\right) = P(\log_2 n) - \frac{R_n}{\log 2 \tilde{\xi}_n} P'(\tilde{\chi}_n).$$

Hence, we obtain (with the convention $\frac{1}{\xi_n} = \frac{1}{\tilde{\xi}_n} = 0$ on the event $\{B_n \in \{0, n\}\}$)

$$\begin{aligned} & \left\| B_n P(\log_2 B_n) + (n - B_n) P(\log_2(n - B_n)) - n P(\log_2 n) \right\|_s \\ &= \left\| \left(\frac{n}{2} + R_n\right) \frac{R_n}{\log 2 \xi_n} P'(\chi_n) - \left(\frac{n}{2} - R_n\right) \frac{R_n}{\log 2 \tilde{\xi}_n} P'(\tilde{\chi}_n) \right\|_s \\ &\leq \left\| \frac{n R_n}{2 \log 2} \left(\frac{1}{\xi_n} P'(\chi_n) - \frac{1}{\tilde{\xi}_n} P'(\tilde{\chi}_n) \right) \right\|_s + \left\| \frac{R_n^2}{\log 2} \left(\frac{1}{\xi_n} P'(\chi_n) + \frac{1}{\tilde{\xi}_n} P'(\tilde{\chi}_n) \right) \right\|_s. \end{aligned} \quad (4.15)$$

For the first summand in (4.15), we obtain by the Cauchy–Schwarz inequality

$$\left\| \frac{n R_n}{2 \log 2} \left(\frac{1}{\xi_n} P'(\chi_n) - \frac{1}{\tilde{\xi}_n} P'(\tilde{\chi}_n) \right) \right\|_s \leq \frac{n}{2 \log 2} \|R_n\|_{2s} \left\| \frac{1}{\xi_n} P'(\chi_n) - \frac{1}{\tilde{\xi}_n} P'(\tilde{\chi}_n) \right\|_{2s},$$

which is of order $O(1)$ since $\|R_n\|_{2s} = \frac{1}{2} n^{1/2} \left\| \frac{B_n - \frac{n}{2}}{\sqrt{n/2}} \right\|_{2s} = O(n^{1/2})$ and

$$\left\| \frac{1}{\xi_n} P'(\chi_n) - \frac{1}{\tilde{\xi}_n} P'(\tilde{\chi}_n) \right\|_{2s} \leq \left\| \frac{1}{\xi_n} (P'(\chi_n) - P'(\tilde{\chi}_n)) \right\|_{2s} + \left\| P'(\tilde{\chi}_n) \left(\frac{1}{\xi_n} - \frac{1}{\tilde{\xi}_n} \right) \right\|_{2s}$$

is of order $O(n^{-3/2})$. This can be seen, e.g., by applying once again the Cauchy–Schwarz inequality and by using the following facts valid for any $r \geq 1$:

- ◊ Since ξ_n is between $\frac{n}{2}$ and B_n , we obtain

$$\left\| \frac{1}{\xi_n} \right\|_r \leq \left\| \mathbf{1}_{\{B_n > 0\}} \frac{1}{\min\{n/2, B_n\}} \right\|_r \leq \mathbb{P}\left(B_n \leq \frac{n}{3}\right) + \left(\frac{3}{n}\right)^r.$$

By some standard Chernoff bound, we find that $\mathbb{P}(B_n \leq n/3)$ is exponentially small, which implies that $\left\| \frac{1}{\xi_n} \right\|_r = O(1/n)$. Analogously, we also find $\left\| \frac{1}{\tilde{\xi}_n} \right\|_r = O(1/n)$.

- ◊ Furthermore, by the Lipschitz property of P' (with Lipschitz constant L , say), we obtain

$$\|P'(\chi_n) - P'(\tilde{\chi}_n)\|_r \leq L \|\chi_n - \tilde{\chi}_n\|_r \leq L \left\| \frac{R_n}{\log 2} \left(\frac{1}{\xi_n} + \frac{1}{\tilde{\xi}_n} \right) \right\|_r = O(n^{-1/2}).$$

- ◊ To bound the term $\left\| P'(\tilde{\chi}_n) \left(\frac{1}{\xi_n} - \frac{1}{\tilde{\xi}_n} \right) \right\|_{2s}$, we make use of the boundedness of P' and the observation that, on $\{0 < B_n < n\}$, we have (recall that ξ_n is between $\frac{n}{2}$ and B_n and $\tilde{\xi}_n$ between $\frac{n}{2}$ and $n - B_n$)

$$\left| \frac{1}{\xi_n} - \frac{1}{\tilde{\xi}_n} \right| \leq \left| \frac{1}{B_n} - \frac{1}{n - B_n} \right| = \frac{2|B_n|}{B_n(n - B_n)}.$$

Since the quadratic function $x \mapsto x(n-x)$ is monotonically increasing for $x \leq n/2$ and monotonically decreasing for $x \geq n/2$, we obtain, by splitting into the events $\{\frac{n}{3} \leq B_n \leq \frac{2n}{3}\}$ and $\{0 < B_n < \frac{n}{3}\} \cup \{\frac{2n}{3} < B_n < n\}$,

$$\left\| \mathbf{1}_{\{0 < B_n < n\}} \frac{1}{B_n(n-B_n)} \right\|_r^r \leq \left(\frac{9}{2n^2} \right)^r + \mathbb{P}\left(B_n < \frac{n}{3}\right) + \mathbb{P}\left(B_n > \frac{2n}{3}\right) = O(n^{-2r}),$$

which implies $\left\| \frac{1}{\xi_n} - \frac{1}{\tilde{\xi}_n} \right\|_{2s} = O(n^{-3/2})$.

In a similar manner, we bound the second summand in (4.15):

$$\left\| \frac{R_n^2}{\log 2} \left(\frac{1}{\xi_n} P'(\chi_n) + \frac{1}{\tilde{\xi}_n} P'(\tilde{\chi}_n) \right) \right\|_s \leq \frac{1}{\log 2} \|R_n^2\|_{2s} \left\| \frac{1}{\xi_n} P'(\chi_n) + \frac{1}{\tilde{\xi}_n} P'(\tilde{\chi}_n) \right\|_{2s},$$

and the latter term is bounded since $\|R_n^2\|_{2s} = O(n)$ and $\left\| \frac{1}{\xi_n} P'(\chi_n) \right\|_{2s} = O(1/n)$. This completes the proof. \square

We now consider the case where our quantities Y_n satisfy recursion (4.12) with b_n being essentially n . We assume that, as $n \rightarrow \infty$, we have

$$\mathbb{E}[Y_n] = n \log_2(n) + nP_1(\log_2 n) + O(1), \quad (4.16)$$

$$\text{Var}(Y_n) = nP_2(\log_2 n) + O(1), \quad (4.17)$$

for some differentiable and 1-periodic functions P_1, P_2 with Lipschitz continuous derivatives and $P_2 > 0$. With minor adjustments in the proof, the error terms $O(1)$ appearing in (4.16) and (4.17) can be replaced by $O(\log^\alpha n)$ with arbitrary $\alpha > 0$. This covers, for example, the external path length of random tries and related digital tree structures constructed from n random binary strings under appropriate independence assumptions (see, e.g., Fuchs, Hwang and Zacharovas [31, 44], Kirschenhofer, Prodinger and Szpankowski [53] and the references therein).

Theorem 4.12. *Let $(Y_n)_{n \geq 0}$ be L_3 -integrable and satisfy (4.12) with $\|b_n - n\|_3 = O(1)$, (4.16) and (4.17). Then, for any $\varepsilon > 0$ and as $n \rightarrow \infty$, we have*

$$\zeta_3 \left(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \mathcal{N}(0, 1) \right) = O(n^{-1/2+\varepsilon}).$$

Proof. The proof works along the same lines as the proof of Theorem 4.9, except that we use Lemma 3.8 with

$$\begin{aligned} f(n) &= n \log_2(n) + nP_1(\log_2 n), & e(n) &= 1, \\ g(n) &= nP_2(\log_2 n), & h(n) &= 1, \end{aligned}$$

with the conventions $0 \cdot \log_2(0) := 0$ and $0 \cdot P_i(\log_2(0)) := 0$ for $i = 1, 2$. Here again, we obtain

$$\left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2} \leq \left\| \sum_{r=1}^2 (\bar{A}_r^{(n)})^2 - 1 \right\|_{3/2} + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right),$$

where $\bar{A}_r^{(n)}$ is defined as in the proof of Theorem 4.9 for $r = 1, 2$. For the toll term $b^{(n)}$, we find

$$\|b^{(n)}\|_3 \leq \|\bar{b}^{(n)}\|_3 + O(n^{-1/2}),$$

where $\bar{b}^{(n)}$ is now given by

$$\frac{1}{\sqrt{nP_2(\log_2 n)}} \left(b_n - n \log_2(n) - nP_1(\log_2 n) + \sum_{r=1}^2 I_r^{(n)} \log_2(I_r^{(n)}) + I_r^{(n)} P_1(\log_2 I_r^{(n)}) \right).$$

Considering the following Lemma 4.13 and the assumption $\|b_n - n\|_3 = O(1)$, we have

$$\|b_n - n \log_2(n) + I_1^{(n)} \log_2(I_1^{(n)}) + I_2^{(n)} \log_2(I_2^{(n)})\|_3 = O(1),$$

which implies, together with Lemma 4.11, $\|\bar{b}^{(n)}\|_3 = O(n^{-1/2})$. The assertion follows just as in Theorem 4.9. \square

Lemma 4.13. *Let B_n be binomially $\text{Bin}(n, 1/2)$ distributed. Then we have, for any $s \geq 1$ and as $n \rightarrow \infty$,*

$$\|n - n \log_2(n) + B_n \log_2(B_n) + (n - B_n) \log_2(n - B_n)\|_s = O(1).$$

Proof. We use the same notation as in the proof of Lemma 4.11 to obtain

$$\begin{aligned} & \|n - n \log_2(n) + B_n \log_2(B_n) + (n - B_n) \log_2(n - B_n)\|_s \\ &= \left\| n - n \log_2(n) + \left(\frac{n}{2} + R_n\right) \left(\log_2\left(\frac{n}{2}\right) + \frac{R_n}{\log 2 \xi_n}\right) + \left(\frac{n}{2} - R_n\right) \left(\log_2\left(\frac{n}{2}\right) - \frac{R_n}{\log 2 \tilde{\xi}_n}\right) \right\|_s \\ &= \left\| \left(\frac{n}{2} + R_n\right) \frac{R_n}{\log 2 \xi_n} - \left(\frac{n}{2} - R_n\right) \frac{R_n}{\log 2 \tilde{\xi}_n} \right\|_s, \end{aligned}$$

which is of order $O(1)$ by the same arguments as in the proof of Lemma 4.11. \square

4.3 Local counters in binary search trees

In this section, we analyze different quantities in a binary search tree built from a random permutation of $\{1, \dots, n\}$. As in Devroye [16], we consider a slightly modified construction of the binary search tree: We assign a random time stamp U_i to each key $i \in \{1, \dots, n\}$, where the time stamps U_1, \dots, U_n are assumed to be independent and uniformly distributed on $[0, 1]$ (and distinct). We think of the data in terms of pairs $(1, U_1), \dots, (n, U_n)$ and insert key i before key j if and only if the corresponding time stamps satisfy $U_i < U_j$ ($i, j = 1, \dots, n$). We now introduce and analyze a class of parameters defined on such

a random binary search tree, the so-called local counters. *Local counters of order k* are quantities Y_n which can be written in the form, see Devroye [16],

$$Y_n = \sum_{i=1}^n f(U_{i-k}, \dots, U_i, \dots, U_{i+k}) \quad (4.18)$$

for some fixed constant $k \geq 0$ and a function $f : [0, 1]^{2k+1} \rightarrow \{0, 1\}$ only depending on the relative order of the input values (meaning that f is invariant under changes in the input values that preserve their relative order). For simplicity of notation, we use the convention $U_i = 0$ for $i \leq 0$ or $i \geq n + 1$. The random variable Y_n can be interpreted as the number of nodes in the binary search tree with some given property only depending on the insertion order of “environing” keys. More precisely, we say that the node storing key i has the given property if and only if $f(U_{i-k}, \dots, U_i, \dots, U_{i+k}) = 1$. To decide whether this is true, it suffices to know the relative order of the time stamps of the keys $i - k, \dots, i, \dots, i + k$. One example of a local counter of order 1 would be the number of leaves since the node storing key i is a leaf if and only if the keys $i - 1$ and $i + 1$ are inserted before i (i.e., $U_{i-1} < U_i$ and $U_{i+1} < U_i$). Further examples of local counters include the number of nodes with one child, the number of nodes with two children, the number of 2-protected nodes (nodes which are neither leaves nor parents of leaves), the number of nodes with exactly k (left) descendants and the number of fringe trees that are copies of a fixed binary tree with k nodes, see Devroye [16], Holmgren and Janson [40] and Mahmoud and Ward [64], where also moments and asymptotic normality are derived.

The definition (4.18) naturally allows a recursive decomposition of local counters: As explained above, Y_n can be understood as the total number of nodes in the binary search tree with some given property. This number can be decomposed recursively by adding the numbers of nodes with the given property in the two subtrees of the root to the indicator that the root itself has the given property. To apply our theorems, we need the additional assumption that the event that the root has the property does not influence the occurrence of the property in the subtrees of the root, in the sense that we have

$$Y_n \stackrel{d}{=} Y_{I_n}^{(1)} + Y_{n-1-I_n}^{(2)} + b_n, \quad n \geq n_0,$$

with $(Y_n^{(1)})_{n \geq 0}$, $(Y_n^{(2)})_{n \geq 0}$ and (I_n, b_n) independent, $Y_n^{(r)} \stackrel{d}{=} Y_n$ for $n \geq 0$ and $r = 1, 2$, I_n uniformly distributed on $\{0, \dots, n - 1\}$ and $b_n \in \{0, 1\}$. As we want to study various local counters multivariately, we analyze the expectation, variance and covariances of different local counters before stating the main theorem.

Lemma 4.14. *Let Y_n and Y'_n be local counters of orders k and k' , respectively. Then, there exists some $n_2 \in \mathbb{N}$ such that we have*

$$\begin{aligned} \mathbb{E}[Y_n] &= \mu n + c_\mu, \\ \text{Var}(Y_n) &= \sigma^2 n + c_\sigma, \\ \text{Cov}(Y_n, Y'_n) &= \varsigma n + c_\varsigma, \end{aligned}$$

for $n \geq n_2$ with some explicitly computable constants $\mu \in [0, 1]$, $\sigma^2 \geq 0$ and $\varsigma, c_\mu, c_\sigma, c_\varsigma \in \mathbb{R}$.

Proof. By definition (4.18), Y_n can be written as $Y_n = \sum_{i=1}^n f(U_{i-k}, \dots, U_i, \dots, U_{i+k})$ with some function f as specified above. In the following, for $i = 1, \dots, n$, we use the notation $Z_i := f(U_{i-k}, \dots, U_i, \dots, U_{i+k})$. Then, for $i = k+1, \dots, n-k$, the random variables Z_i are identically distributed, which implies for $n \geq 2k$

$$\begin{aligned} \mathbb{E}[Y_n] &= \sum_{i=k+1}^{n-k} \mathbb{E}[Z_i] + \sum_{i=1}^k \mathbb{E}[Z_i] + \sum_{i=n-k+1}^n \mathbb{E}[Z_i] \\ &= \mathbb{E}[Z_{k+1}] \cdot n - 2k \cdot \mathbb{E}[Z_{k+1}] + \sum_{i=1}^k \mathbb{E}[Z_i] + \sum_{i=1}^k \mathbb{E}[Z_{n-i+1}]. \end{aligned}$$

Note that for $i = 1, \dots, k$, the term $\mathbb{E}[Z_{n-i+1}]$ is a fixed constant independent of n since $Z_{n-i+1} = f(U_{n-i+1-k}, \dots, U_n, 0, \dots, 0)$ has the same distribution as $f(U_1, \dots, U_{i+k}, 0, \dots, 0)$. Thus, the statement for the mean is satisfied with

$$\mu = \mathbb{E}[Z_{k+1}], \quad c_\mu = -2k \cdot \mathbb{E}[Z_{k+1}] + \sum_{i=1}^k \mathbb{E}[Z_i] + \sum_{i=1}^k \mathbb{E}[Z_{n-i+1}].$$

For the variance of Y_n , we additionally use that Z_i and Z_j are independent whenever $|i-j| > 2k$ and that the pair (Z_i, Z_{i+j}) has the same joint distribution as the pair (Z_{k+1}, Z_{k+1+j}) for $i = k+1, \dots, n-3k$ and $j = 1, \dots, 2k$. Thus, we have for $n \geq 4k$

$$\begin{aligned} \text{Var}(Y_n) &= \sum_{i=1}^n \text{Var}(Z_i) + 2 \sum_{i < j} \text{Cov}(Z_i, Z_j) \\ &= (n-2k) \text{Var}(Z_{k+1}) + \sum_{i=1}^k \text{Var}(Z_i) + \sum_{i=n-k+1}^n \text{Var}(Z_i) + 2 \sum_{i=k+1}^{n-3k} \sum_{j=1}^{2k} \text{Cov}(Z_i, Z_{i+j}) \\ &\quad + 2 \sum_{i=1}^k \sum_{j=i+1}^{i+2k} \text{Cov}(Z_i, Z_j) + 2 \sum_{i=n-3k+1}^n \sum_{j=i+1}^{(i+2k) \wedge n} \text{Cov}(Z_i, Z_j) \\ &= \left(\text{Var}(Z_{k+1}) + 2 \sum_{j=1}^{2k} \text{Cov}(Z_{k+1}, Z_{k+1+j}) \right) n + c_\sigma, \end{aligned}$$

where the symbol c_σ is defined as

$$\begin{aligned} c_\sigma &= -2k \cdot \text{Var}(Z_{k+1}) - 8k \sum_{j=1}^{2k} \text{Cov}(Z_{k+1}, Z_{k+1+j}) + \sum_{i=1}^k \text{Var}(Z_i) + \sum_{i=1}^k \text{Var}(Z_{n-i+1}) \\ &\quad + 2 \sum_{i=1}^k \sum_{j=i+1}^{i+2k} \text{Cov}(Z_i, Z_j) + 2 \sum_{i=n-3k+1}^n \sum_{j=i+1}^{(i+2k) \wedge n} \text{Cov}(Z_i, Z_j). \end{aligned}$$

Arguments similar to those for the mean show that the terms combined in c_σ are indeed fixed constants independent of n for $n \geq 4k$. In a similar manner, we derive the

covariance $\text{Cov}(Y_n, Y'_n)$. Since Y'_n is a local counter of order k' , we find some function f' as specified in definition (4.18) such that we can write $Y'_n = \sum_{i=1}^n Z'_i$ with $Z'_i := f'(U_{i-k'}, \dots, U_i, \dots, U_{i+k'})$. As Z_i and Z'_j are independent for $|i-j| > k+k'$ and as the pair (Z_i, Z'_j) has the same distribution as $(Z_{k+2k'+1}, Z'_{k+2k'+1+j-i})$ for $i = k+2k'+1, \dots, n-k-2k'$ and $j = i-k-k', \dots, i+k+k'$, we obtain for $n \geq 2k+4k'$

$$\begin{aligned} \text{Cov}(Y_n, Y'_n) &= \sum_{i,j=1}^n \text{Cov}(Z_i, Z'_j) \\ &= \sum_{i=1}^{k+2k'} \sum_{j=(i-k-k') \vee 1}^{i+k+k'} \text{Cov}(Z_i, Z'_j) + \sum_{i=k+2k'+1}^{n-k-2k'} \sum_{j=i-k-k'}^{i+k+k'} \text{Cov}(Z_i, Z'_j) \\ &\quad + \sum_{i=n-k-2k'+1}^n \sum_{j=i-k-k'}^{(i+k+k') \wedge n} \text{Cov}(Z_i, Z'_j) \\ &= \left(\sum_{j=k'+1}^{2k+3k'+1} \text{Cov}(Z_{k+2k'+1}, Z'_j) \right) n + c_\zeta, \end{aligned}$$

with c_ζ defined by

$$\begin{aligned} c_\zeta &= -(2k+4k') \sum_{j=k'+1}^{2k+3k'+1} \text{Cov}(Z_{k+2k'+1}, Z'_j) + \sum_{i=1}^{k+2k'} \sum_{j=(i-k-k') \vee 1}^{i+k+k'} \text{Cov}(Z_i, Z'_j) \\ &\quad + \sum_{i=n-k-2k'+1}^n \sum_{j=i-k-k'}^{(i+k+k') \wedge n} \text{Cov}(Z_i, Z'_j). \end{aligned}$$

Here again, all appearing sums are independent of n for $n \geq 2k+4k'$, which completes the proof. \square

Now that we have determined the covariance structure, we can analyze any number of local counters multivariately. For this, we assume $Y_n = (Y_{1,n}, \dots, Y_{d,n})$ to be a d -dimensional vector of local counters of orders k_1, \dots, k_d satisfying the recursion

$$Y_n \stackrel{d}{=} Y_{I_n}^{(1)} + Y_{n-1-I_n}^{(2)} + b_n, \quad n \geq n_0, \quad (4.19)$$

with $(Y_n^{(1)})_{n \geq 0}$, $(Y_n^{(2)})_{n \geq 0}$ and (I_n, b_n) independent, $Y_n^{(r)} \stackrel{d}{=} Y_n$ for $n \geq 0$ and $r = 1, 2$, I_n uniformly distributed on $\{0, \dots, n-1\}$ and some random vector b_n in $\{0, 1\}^d$. By Lemma 4.14, there exists a $(d \times d)$ -matrix A such that the covariance matrix of Y_n has the form

$$\text{Cov}(Y_n) = An + O(1), \quad (4.20)$$

where the $O(1)$ -term can be read componentwise. The following theorem provides a bound on the rate of convergence in the Zolotarev distance ζ_3 for the multivariate quantities Y_n .

Theorem 4.15. *Let $Y_n = (Y_{1,n}, \dots, Y_{d,n})$ be a vector of local counters of orders k_1, \dots, k_d satisfying the distributional recursion (4.19). Furthermore, assume that the matrix A defined by (4.20) is positive definite. Then we have, for any $\varepsilon > 0$ and as $n \rightarrow \infty$,*

$$\zeta_3(\text{Cov}(Y_n)^{-1/2}(Y_n - \mathbb{E}[Y_n]), \mathcal{N}(0, \text{Id}_d)) = O(n^{-1/2+\varepsilon}).$$

Proof. Since $\text{Cov}(Y_n) = An + O(1)$ with some positive definite matrix A , we have that $\text{Cov}(Y_n)$ is positive definite for large n , say for $n \geq n_1$. Thus, we set $C_n = \text{Cov}(Y_n)$ for $n \geq n_1$, $C_n = \text{Id}_d$ for $n < n_1$, $M_n = \mathbb{E}[Y_n]$ for $n \geq 0$ and introduce the normalized quantities

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

The normalized sequence satisfies the modified recursion

$$X_n \stackrel{d}{=} A_1^{(n)} X_{I_1^{(n)}}^{(1)} + A_2^{(n)} X_{I_2^{(n)}}^{(2)} + b^{(n)}, \quad n \geq n_0,$$

with $I_1^{(n)} = I_n$, $I_2^{(n)} = n - 1 - I_n$,

$$A_r^{(n)} := C_n^{-1/2} C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2} \left(b_n - M_n + M_{I_1^{(n)}} + M_{I_2^{(n)}} \right)$$

and independence and self-similarity relations as before. By Lemma 2.3, we obtain

$$C_n^{1/2} = (An + O(1))^{1/2} = A^{1/2}n^{1/2} + O(n^{-1/2})$$

and, by Lemma 2.1,

$$C_n^{-1/2} = n^{-1/2} (A^{1/2} + O(n^{-1}))^{-1} = A^{-1/2}n^{-1/2} + O(n^{-3/2}).$$

Modeling all quantities on a common probability space such that $I_1^{(n)}/n$ converges almost surely to a uniformly distributed random variable U in $[0, 1]$, we have the L_3 -convergences $A_1^{(n)} \rightarrow \sqrt{U} \text{Id}_d$, $A_2^{(n)} \rightarrow \sqrt{1-U} \text{Id}_d$ and $b^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. Thus, we are in the situation of Section 3.1.2 and obtain the limiting equation

$$X \stackrel{d}{=} \sqrt{U} X^{(1)} + \sqrt{1-U} X^{(2)},$$

with U uniformly distributed on $[0, 1]$ and $X^{(1)}$, $X^{(2)}$ and U independent. To apply Theorem 3.4, we need to determine the orders of $\|A_1^{(n)}(A_1^{(n)})^\top + A_2^{(n)}(A_2^{(n)})^\top - \text{Id}_d\|_{3/2}^{3/2}$ and $\|b^{(n)}\|_3^3$. For $r = 1, 2$, we find

$$\begin{aligned} A_r^{(n)}(A_r^{(n)})^\top &= C_n^{-1/2} C_{I_r^{(n)}} C_n^{-1/2} \\ &= (A^{-1/2}n^{-1/2} + O(n^{-3/2})) (A I_r^{(n)} + O(1)) (A^{-1/2}n^{-1/2} + O(n^{-3/2})) \\ &= \frac{I_r^{(n)}}{n} \text{Id}_d + O(n^{-1}) \end{aligned}$$

and hence

$$\|A_1^{(n)}(A_1^{(n)})^\top + A_2^{(n)}(A_2^{(n)})^\top - \text{Id}_d\|_{3/2}^{3/2} = O(n^{-3/2}).$$

Furthermore, denoting by μ the d -dimensional vector such that $M_n = \mu n + O(1)$, we obtain

$$b^{(n)} = C_n^{-1/2} \left(b_n - \mu n + \mu I_1^{(n)} + \mu I_2^{(n)} + O(1) \right) = C_n^{-1/2} (b_n + O(1)).$$

Thus, we have

$$\|b^{(n)}\|_3^3 \leq \|C_n^{-1/2}\|_{\text{op}}^3 \cdot \|b_n + O(1)\|_3^3 = O(n^{-3/2}).$$

As the technical conditions (3.6) and (3.14) are satisfied, it remains to check condition (3.16): With $R(n) = n^{-1/2+\varepsilon}$ and as $n \rightarrow \infty$, we have

$$\mathbb{E} \left[\frac{R(I_1^{(n)})}{R(n)} \|A_1^{(n)}\|_{\text{op}}^3 \right] = \mathbb{E} \left[\left(\frac{I_1^{(n)}}{n} \right)^{-1/2+\varepsilon} \|A_1^{(n)}\|_{\text{op}}^3 \right] \rightarrow \mathbb{E}[U^{1+\varepsilon}] < 0.5,$$

which finishes the proof. \square

In numerous applications, the ε in the convergence rate of Theorem 4.15 can be removed. Note that we only needed $\text{Cov}(Y_n) = An + O(1)$ in the proof of Theorem 4.15, but we could use the representation $\text{Cov}(Y_n) = An + B$ for some matrices A, B in combination with Remark 3.5 to avoid the ε in certain applications. This is particularly true, see Theorem 4.16 below, if the covariance matrix of Y_n is of the form $A(n+1)$ for large n and some positive definite matrix A —which is true for a large class of examples (see, e.g., Holmgren and Janson [40] and the references therein). The following theorem considers this special case. In particular, it covers applications such as the number Y_n of 2-protected nodes in a random binary search tree of size n , for which Mahmoud and Ward [64] have shown that $\mathbb{E}[Y_n] = 11/30 n - 19/30$ for $n \geq 4$, $\text{Var}(Y_n) = 29/225(n+1)$ for $n \geq 8$ and that the standardized Y_n converge in distribution to a normal limit.

Theorem 4.16. *Let $Y_n = (Y_{1,n}, \dots, Y_{d,n})$ be a vector of local counters of orders k_1, \dots, k_d satisfying the distributional recursion (4.19). Furthermore, assume that there exist d -dimensional vectors μ and ν , a positive definite $(d \times d)$ -matrix A and some $n_1 \in \mathbb{N}_0$ such that the mean and the covariance matrix of Y_n are of the form $\mathbb{E}[Y_n] = \mu n + \nu$ and $\text{Cov}(Y_n) = A(n+1)$ for $n \geq n_1$. If additionally $\|b_n - \mu + \nu\|_3^3 = O(1/n)$, then we have, as $n \rightarrow \infty$,*

$$\zeta_3(\text{Cov}(Y_n)^{-1/2}(Y_n - \mathbb{E}[Y_n]), \mathcal{N}(0, \text{Id}_d)) = O(n^{-1/2}).$$

Proof. Just as in the previous proof, we use the normalization

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

with $C_n = \text{Cov}(Y_n)$ for $n \geq n_1$, $C_n = \text{Id}_d$ for $n < n_1$, $M_n = \mathbb{E}[Y_n]$ for $n \geq 0$ and check the conditions of Theorem 3.4. Since $A_1^{(n)}(A_1^{(n)})^\top + A_2^{(n)}(A_2^{(n)})^\top = \text{Id}_d$ on the event $\{I_1^{(n)}, I_2^{(n)} \geq n_1\}$, we obtain, as $n \rightarrow \infty$,

$$\begin{aligned} \left\| \sum_{r=1}^2 A_r^{(n)}(A_r^{(n)})^\top - \text{Id}_d \right\|_{3/2}^{3/2} &= O\left(\left\| \mathbf{1}_{\{I_1^{(n)} < n_1\}} \left(\frac{1}{n+1} A^{-1} + \frac{I_2^{(n)} + 1}{n+1} \text{Id}_d - \text{Id}_d \right) \right\|_{3/2}^{3/2} \right) \\ &= O\left(\mathbb{E} \left[\mathbf{1}_{\{I_1^{(n)} < n_1\}} \left\| \frac{1}{n+1} A^{-1} - \frac{I_1^{(n)} + 1}{n+1} \text{Id}_d \right\|_{\text{op}}^{3/2} \right] \right) \\ &= O(n^{-5/2}). \end{aligned}$$

Similarly, using the representation $M_n = \mu n + \nu$ for $n \geq n_1$ and the assumption $\|b_n - \mu + \nu\|_3^3 = O(n^{-1})$, we obtain

$$\|b^{(n)}\|_3^3 \leq \|C_n^{-1/2}\|_{\text{op}}^3 \cdot \|b_n - M_n + M_{I_1^{(n)}} + M_{I_2^{(n)}}\|_3^3 = O(n^{-5/2}).$$

We now use Theorem 3.4 with $R(n) = n^{-1/2}$. Note that condition (3.16) is not satisfied for $R(n) = n^{-1/2}$, but we can use the weakened condition stated in Remark 3.5. \square

To conclude this section, we now give a multivariate example. For this, we consider a random binary search tree with n nodes constructed as described at the beginning of this section and denote by L_{0n} the number of nodes with no left descendant and by L_{1n} the number of nodes with exactly one left descendant, $n \geq 0$. Following Devroye [16], we have the local counter representations

$$L_{0n} = \sum_{i=1}^n \mathbf{1}_{\{U_{i-1} < U_i\}}, \quad L_{1n} = \sum_{i=1}^n \mathbf{1}_{\{U_{i-2} < U_i < U_{i-1}\}},$$

since the node storing key i has no left descendant if and only if the key $i-1$ is inserted before i (i.e., $U_{i-1} < U_i$) and it has exactly one left descendant (namely the node with key $i-1$) if the key i is inserted before $i-1$, but after $i-2$ (i.e., $U_{i-2} < U_i < U_{i-1}$). Note that we use the convention $U_i = 0$ for $i \leq 0$ again to simplify notation. In Devroye [16], it is shown that, for $n \geq 2$,

$$\mathbb{E}[L_{0n}] = \frac{1}{2}(n+1), \quad \mathbb{E}[L_{1n}] = \frac{1}{6}(n+1),$$

and that the standardized quantities have a limiting normal distribution. In order to study the quantities multivariately, we additionally derive their covariance structure with Lemma 4.14 and find that we have $\text{Cov}((L_{0n}, L_{1n})) = (n+1)\Gamma$ for $n \geq 4$ with

$$\Gamma = \frac{1}{360} \begin{pmatrix} 30 & -15 \\ -15 & 28 \end{pmatrix}.$$

Furthermore, defining $Y_n := (L_{0n}, L_{1n})$ for $n \geq 0$, we have $Y_0 = (0, 0)$ and we directly obtain the following distributional recurrence:

$$Y_n \stackrel{d}{=} Y_{I_n}^{(1)} + Y_{n-1-I_n}^{(2)} + b_n, \quad n \geq 1,$$

where $(Y_j^{(1)})_{j \geq 0}$, $(Y_j^{(2)})_{j \geq 0}$ and I_n are independent, $Y_j^{(r)}$ has the same distribution as Y_j for $j \geq 0$ and $r = 1, 2$, $\mathcal{L}(I_n) = \text{unif}\{0, \dots, n-1\}$ and $b_n = (\mathbf{1}_{\{I_n=0\}}, \mathbf{1}_{\{I_n=1\}})$. Since the matrix Γ is positive definite, we can apply Theorem 4.16 to obtain the following result.

Theorem 4.17. *Denoting by $Y_n := (L_{0n}, L_{1n})$ the vector of the numbers of nodes with no left descendant and with exactly one left descendant, respectively, in a random binary search tree with n nodes, we have, as $n \rightarrow \infty$,*

$$\zeta_3(\text{Cov}(Y_n)^{-1/2}(Y_n - \mathbb{E}[Y_n]), \mathcal{N}(0, \text{Id}_2)) = O(n^{-1/2}).$$

4.4 Convergence rates for Quickselect

In this section, we study the number of key comparisons and the number of key exchanges used by Hoare's *Quickselect* algorithm [38] for finding the m -th smallest element in a uniformly distributed random permutation of $\{1, \dots, n\}$. Briefly summarized, Quickselect works as follows: First, we choose an arbitrary (here, the rightmost) element p of the list, the so-called pivot, which is used to partition the remaining elements into elements smaller than p (left sublist) and elements larger than p (right sublist). After this partitioning step, the procedure continues on the sublist which contains the m -th smallest element or stops if the pivot itself has the desired rank. To be concrete, the following algorithm (Algorithm 1) shows pseudocode for this method.

Algorithm 1 Quickselect (finds the m -th smallest element among the elements $A[\text{left}], \dots, A[\text{right}]$)

```

1: procedure QUICKSELECT( $A$ , left, right,  $m$ )
2:   if left = right then
3:     return  $A[\text{right}]$ 
4:   else
5:      $p \leftarrow A[\text{right}]$ 
6:      $r \leftarrow$  PARTITION( $A$ , left, right - 1,  $p$ )
       // the procedure PARTITION returns the rank  $r$  of the pivot  $p$  among the elements
       //  $A[\text{left}], \dots, A[\text{right}]$  and moves the keys such that the keys  $A[\text{left}], \dots, A[\text{left} + r - 2]$ 
       // are smaller than  $p$  and  $A[\text{left} + r - 1], \dots, A[\text{right} - 1]$  are larger than  $p$ 
7:     if  $m = r$  then return  $p$ 
8:     else if  $m < r$  then return QUICKSELECT( $A$ , left, left +  $r - 2$ ,  $m$ )
9:     else if  $m > r$  then return QUICKSELECT( $A$ , left +  $r - 1$ , right - 1,  $m - r$ )

```

Note that there exist various versions of Hoare’s Quickselect algorithm which differ in the concrete partitioning strategy used. In this section, we assume that our partitioning procedure preserves the randomness and the independence between the subfiles and uses exactly $n - 1$ key comparisons and a mixed hypergeometrically distributed number of exchanges. This is true, e.g., for the procedure presented in Wild [94, Algorithm 9]. Based on Hoare’s “crossing pointers” technique, this partitioning procedure makes use of two pointers scanning the list from left and right until misplaced elements are found. For the sake of completeness, Algorithm 2 shows the concrete structure of the partitioning procedure. However, our asymptotic results (Theorems 4.18 and 4.19 below) also hold for slightly different partitioning procedures (e.g., for the versions presented in Sedgewick [86, Program 1], Cormen, Leiserson and Rivest [13, Section 8.1] or Wild and Nebel [95, Algorithm 1]).

Algorithm 2 Partitioning Procedure (compares the keys $A[\text{lower}], \dots, A[\text{upper}]$ with the pivot p , moves the keys that are smaller than p to the lower positions and those that are larger than p to the higher positions and returns the rank of the pivot p among the elements $A[\text{lower}], \dots, A[\text{upper}], p$).

```

1: procedure PARTITION( $A$ , lower, upper,  $p$ )
2:    $k \leftarrow$  lower;  $g \leftarrow$  upper
3:   while  $k \leq g$  do
4:     while true do
5:       if  $A[k] \leq p$  then
6:          $k \leftarrow k + 1$ 
7:       if  $k > g$  then break outer while
8:       else break inner while
9:     while true do
10:      if  $k == g$  then  $g \leftarrow g - 1$ ; break outer while
11:      if  $A[g] > p$  then  $g \leftarrow g - 1$ 
12:      else break inner while
13:      swap  $A[k]$  and  $A[g]$ 
14:       $k \leftarrow k + 1$ ;  $g \leftarrow g - 1$ 
15:   return  $k - \text{lower} + 1$ 

```

The limiting distribution of the number $Y_{n,m}$ of key comparisons used by Quickselect for finding the m -th smallest element in a randomly permuted input of length n (with $m \sim cn$ for some $0 \leq c \leq 1$) was studied independently by Grübel and Rösler [34] and Kodaj and Móri [57], the latter showing (among other things) that the rate of convergence for the normalized number $Y_{n, \lfloor nc \rfloor} / n$ of comparisons measured in the Wasserstein ℓ_1 distance is $\Theta(\log n / n)$. Hwang and Tsai [45] have shown that for $m = o(n)$, the limiting distribution of the normalized number $(Y_{n,m} - n) / n$ of comparisons is the Dickman distribution D satisfying the distributional recursion ($\mathcal{L}(U) = \text{unif}[0, 1]$, U and D independent)

$$D \stackrel{d}{=} UD + U.$$

This result was further quantified by Goldstein [33] who derived concrete non-asymptotic upper and lower bounds of order $\log n/n$ for the Wasserstein ℓ_1 distance between the normalized number $(Y_{n,m}-n)/n$ of key comparisons and its limiting (Dickman) distribution. From the definition of the Zolotarev ζ_1 distance (which coincides with the Wasserstein ℓ_1 metric, see Section 2.2.2), it follows immediately that the rate $\log n/n$ cannot be improved (see Theorem 1.3 in Goldstein [33]). This is because the function $x \mapsto x$ is in \mathcal{F}_1 , i.e., we have

$$\zeta_1\left(\frac{Y_{n,m}-n}{n}, D\right) \geq \left| \frac{\mathbb{E}[Y_{n,m}]}{n} - 1 - \mathbb{E}[D] \right| = \Omega\left(\frac{\log n}{n}\right)$$

for any fixed $m \neq 2$ (the last step follows by using $\mathbb{E}[D] = 1$ and the exact expression for the expected number $\mathbb{E}[Y_{n,m}]$ of comparisons given in Knuth [54]). This raises the question whether the rate $\log n/n$ can be improved if we consider the centered quantities $(Y_{n,m} - \mathbb{E}[Y_{n,m}])/n$ instead.

For the number $W_{n,m}$ of key exchanges made by Quickselect for finding the m -th smallest element in a randomly permuted input of length n (with Hoare's partitioning method), Hwang and Tsai [45] showed that for $m = o(n)$, the normalized number $W_{n,m}/n$ of exchanges converges in distribution to a limit W satisfying the distributional recursion

$$W \stackrel{d}{=} UW + U(1 - U)$$

with U uniformly on $[0, 1]$ and independent of W . Mahmoud [61] analyzed the limiting distribution of the number of exchanges made by Quickselect to find an element with a randomly selected rank (using Lomuto's partition algorithm) and showed that different partition algorithms lead to different results. Furthermore, for the number of key exchanges when selecting a uniform rank with Hoare's "crossing pointers" technique, Dadoun and Neininger [14] have provided a rate of convergence in the Wasserstein ℓ_p metrics and in the Kolmogorov metric.

In what follows, we will use our general Theorem 3.1 in order to derive convergence rates in the Zolotarev ζ_s distances for both the number of key comparisons and the number of key exchanges. We first consider the case $m = 1$ where we always recurse on the left sublist and then the case where an independent and uniformly distributed rank is selected.

4.4.1 Quickselect for finding the smallest element

Throughout this section, let $Y_n^{[c]}$ and $Y_n^{[e]}$ denote the numbers of key comparisons and key exchanges used by Quickselect (with the partitioning strategy given in Algorithm 2) for finding the smallest element in a list of n uniformly permuted distinct data. Denoting by I_n the size of the left sublist and by $b_c(n)$ and $b_e(n)$ the numbers of key comparisons and key exchanges during the first partitioning step, we have $Y_0^{[c]} = 0$ and $Y_n^{[c]}$ satisfies the following distributional recursion for $\iota = c, e$:

$$Y_n^{[\iota]} \stackrel{d}{=} Y_{I_n}^{[\iota]} + b_\iota(n), \quad n \geq 1,$$

where $(I_n, b_\iota(n))$ and $(Y_n^{[\iota]})_{n \geq 0}$ are independent and I_n is uniformly distributed on the set $\{0, \dots, n-1\}$. Using the partitioning procedure described in Algorithm 2, we have $n-1$ key comparisons during the first step, i.e., $b_c(n) = n-1$ for $n \geq 1$. To count the number of exchanges, we notice that we need one exchange for each misplaced pair of keys. More precisely, given $I_n = j$, the number of exchanges coincides with the number of large elements among the first j elements of the permutation. Thus, for the number of exchanges, we obtain $b_e(n) = Z_n$ for $n \geq 1$, where, conditional on $I_n = j$, the random variable Z_n is hypergeometrically $\text{Hyp}(n-1, j, n-1-j)$ distributed. In Knuth [54], it is shown that

$$\mu_c(n) := \mathbb{E}[Y_n^{[c]}] = 2n - 2\mathcal{H}_n, \quad n \geq 1.$$

Furthermore, using Lemma 5.3 in Hwang and Tsai [45] together with the fact that we have $\mathbb{E}[b_e(n)] = (n-2)/6$ for $n \geq 2$, we obtain

$$\mu_e(n) := \mathbb{E}[Y_n^{[e]}] = \frac{1}{3}n - \frac{1}{2}\mathcal{H}_n + \frac{1}{12}, \quad n \geq 2.$$

For $\iota = c, e$, we define the normalized quantities $X_0^{[\iota]} := 0$ and

$$X_n^{[\iota]} := \frac{1}{n}(Y_n^{[\iota]} - \mu_\iota(n)), \quad n \geq 1.$$

Then we have

$$X_n^{[\iota]} \stackrel{d}{=} A_1^{(n)} X_{I_n}^{[\iota]} + b_\iota^{(n)}, \quad n \geq 1,$$

where $(A_1^{(n)}, b_\iota^{(n)}, I_n)$ and $(X_n^{[\iota]})_{n \geq 0}$ are independent,

$$A_1^{(n)} = \frac{I_n}{n} \quad \text{and} \quad b_\iota^{(n)} = \frac{1}{n}(b_\iota(n) - \mu_\iota(n) + \mu_\iota(I_n)).$$

Writing $I_n = \lfloor nU \rfloor$ for some uniformly distributed U on $[0, 1]$, we have the joint L_s -convergence

$$\left(\frac{I_n}{n}, \frac{Z_n}{n} \right) \rightarrow (U, U(1-U))$$

as $n \rightarrow \infty$, $1 \leq s < \infty$ (see, e.g., Dadoun and Neininger [14]). Thus, this leads to the limiting equation

$$X^{[\iota]} \stackrel{d}{=} A_1^* X^{[\iota]} + b_\iota^* \tag{4.21}$$

with $A_1^* = U$ uniformly distributed on $[0, 1]$, $b_c^* := 2U - 1$, $b_e^* := U(1-U) + \frac{1}{3}U - \frac{1}{3}$ and $X^{[\iota]}$ independent of U . Note that a criterion of Vervaat [92] states that there exists a unique solution of equation (4.21) among all probability distributions on the real line. Applying Theorem 3.1, we now derive the following convergence rates.

Theorem 4.18. *The number $Y_n^{[c]}$ of key comparisons and the number $Y_n^{[e]}$ of key exchanges used by Quickselect for finding the smallest element in a random permutation of n distinct elements satisfy, as $n \rightarrow \infty$,*

$$\zeta_s\left(\frac{1}{n}(Y_n^{[c]} - \mathbb{E}[Y_n^{[c]}]), X^{[c]}\right) = \begin{cases} O(\log n/n), & s = 1, \\ O(1/n), & s \in (1, 2), \\ \Theta(1/n), & s = 2, \end{cases}$$

and, for $s \in [1, 2]$,

$$\zeta_s\left(\frac{1}{n}(Y_n^{[e]} - \mathbb{E}[Y_n^{[e]}]), X^{[e]}\right) = O\left(\frac{1}{n^{1/2}}\right),$$

where the distribution of $X^{[c]}$ ($X^{[e]}$, respectively) is the unique solution of equation (4.21) among all probability distributions on the real line.

Proof. We start with the number $Y_n^{[c]}$ of key comparisons by proving the easier lower bound for $s = 2$: By definition of the Zolotarev metric, we have for square integrable V and W with identical first moments

$$\zeta_2(V, W) \geq \frac{1}{2} |\mathbb{E}[V^2] - \mathbb{E}[W^2]|,$$

since $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \frac{1}{2}x^2$ is in \mathcal{F}_2 . Thus, we only need information on the second moments of $X_n^{[c]}$ and $X^{[c]}$ to obtain a lower bound for the quantity $\zeta_2(X_n^{[c]}, X^{[c]})$. Using the limiting equation (4.21), we find

$$\mathbb{E}[(X^{[c]})^2] = \mathbb{E}[U^2]\mathbb{E}[(X^{[c]})^2] + \mathbb{E}[(2U - 1)^2] = \frac{1}{3}\mathbb{E}[(X^{[c]})^2] + \frac{1}{3}.$$

Solving for $\mathbb{E}[(X^{[c]})^2]$, we obtain $\mathbb{E}[(X^{[c]})^2] = \frac{1}{2}$. Furthermore, by Theorem 1 in Mahmoud, Modarres and Smythe [62], we have as $n \rightarrow \infty$,

$$\mathbb{E}[(X_n^{[c]})^2] = \frac{1}{n^2} \text{Var}(Y_n^{[c]}) = \frac{1}{2} - \frac{9}{2n} + O\left(\frac{\log n}{n^2}\right).$$

Altogether, we obtain

$$\zeta_2(X_n^{[c]}, X^{[c]}) \geq \frac{1}{2} \left| \mathbb{E}[(X_n^{[c]})^2] - \mathbb{E}[(X^{[c]})^2] \right| = \Theta\left(\frac{1}{n}\right).$$

We now prove the upper bound for $s \in [1, 2]$ by applying Theorem 3.1. With the embedding $I_n = \lfloor nU \rfloor$, we have

$$\|A_1^{(n)} - A_1^*\|_s = \left\| \frac{\lfloor nU \rfloor}{n} - U \right\|_s \leq \frac{1}{n}.$$

Furthermore, applying Lemma 3.8 with $f(n) = 2n - 2\log(n)$, $e(n) = 1$, $g(n) = n^2$ and $h(n) = 0$, we obtain

$$\|b_c^{(n)} - b_c^*\|_s \leq \|\bar{b}_c^{(n)} - b_c^*\|_s + O\left(\frac{1}{n}\right),$$

where for the first summand, we have

$$\begin{aligned} \|\bar{b}_c^{(n)} - b_c^*\|_s &= \left\| \frac{1}{n} (n - 1 - (2n - 2 \log(n)) + (2I_n - 2 \log(I_n \vee 1))) - (2U - 1) \right\|_s \\ &\leq \left\| \frac{1}{n} \right\|_s + 2 \left\| \frac{I_n}{n} - U \right\|_s + \frac{2}{n} \left\| \log \left(\frac{n}{I_n \vee 1} \right) \right\|_s = O\left(\frac{1}{n}\right). \end{aligned}$$

For the last estimate, observe that $n/(I_n \vee 1)$ converges almost surely to U^{-1} as $n \rightarrow \infty$ and that we can use the bound $[nU] \vee 1 \geq \frac{1}{2}nU$ to see that $\log\left(\frac{n}{I_n \vee 1}\right)$ is bounded by $\log(2U^{-1})$ for any $n \geq 1$. Thus, by dominated convergence, we obtain that $\mathbb{E}\left[\log^s\left(\frac{n}{I_n \vee 1}\right)\right]$ is bounded in n .

Moreover, the technical conditions (3.5) and (3.6) are satisfied for $R(n) = 1/n$ and by dominated convergence, we have for $s > 1$ and as $n \rightarrow \infty$

$$\mathbb{E}\left[\frac{R(I_n)}{R(n)} |A_1^{(n)}|^s\right] = \mathbb{E}\left[\left(\frac{I_n}{n}\right)^{s-1}\right] \rightarrow \mathbb{E}[U^{s-1}] = \frac{1}{s} < 1.$$

However, for $s = 1$, condition (3.8) is not satisfied—neither for $R(n) = 1/n$ nor for $R(n) = \log n/n$ —but we can use the weaker condition in Remark 3.3 with $R(n) = \log n/n$ and $\tilde{R}(n) = 1/n$: From the inequality $\sum_{i=2}^{n-1} \log i \leq n \log n - n$, it follows that

$$\mathbb{E}\left[\mathbf{1}_{\{I_n \geq 1\}} \frac{R(I_n)}{R(n)} A_1^{(n)}\right] = \mathbb{E}\left[\frac{\log(I_n \vee 1)}{\log n}\right] \leq 1 - \frac{1}{\log n},$$

which completes the proof for the number $Y_n^{[c]}$ of key comparisons.

Similarly, for the number $Y_n^{[e]}$ of key exchanges, we apply Lemma 3.8 with $f(n) = \frac{1}{3}n$, $e(n) = \log(n)$, $g(n) = n^2$ and $h(n) = 0$ and obtain

$$\|b_e^{(n)} - b_e^*\|_s \leq \|\bar{b}_e^{(n)} - b_e^*\|_s + O\left(\frac{\log(n)}{n}\right),$$

with

$$\begin{aligned} \|\bar{b}_e^{(n)} - b_e^*\|_s &= \left\| \frac{1}{n} \left(Z_n - \frac{1}{3}n + \frac{1}{3}I_n \right) - \left(U(1 - U) + \frac{1}{3}U - \frac{1}{3} \right) \right\|_s \\ &\leq \left\| \frac{Z_n}{n} - U(1 - U) \right\|_s + \frac{1}{3} \left\| \frac{I_n}{n} - U \right\|_s = O\left(\frac{1}{n^{1/2}}\right), \end{aligned}$$

since we have $\|Z_n/n - U(1 - U)\|_s = O(n^{-1/2})$ for $s \geq 1$ (see Dadoun and Neininger [14, Lemma 3.2] with small changes in the proof, since their partitioning procedure slightly differs from ours). Furthermore, with $R(n) = n^{-1/2}$ and as $n \rightarrow \infty$,

$$\mathbb{E}\left[\frac{R(I_n)}{R(n)} |A_1^{(n)}|^s\right] = \mathbb{E}\left[\left(\frac{I_n}{n}\right)^{s-1/2}\right] \rightarrow \mathbb{E}[U^{s-1/2}] = \frac{1}{s+1/2} < 1.$$

Thus, for the number $Y_n^{[e]}$ of key exchanges, the conditions of Theorem 3.1 are satisfied for $R(n) = n^{-1/2}$ and $s \in [1, 2]$. \square

4.4.2 Quickselect for finding a uniformly chosen element

In this section, we consider the case where the rank R_n to be selected is uniformly distributed on $\{1, \dots, n\}$ and independent of the permutation of the data. Again, we denote by $Y_n^{[c]}$ ($Y_n^{[e]}$, respectively) the number of key comparisons (key exchanges, respectively), by I_n the size of the left sublist and by $b_c(n)$ and $b_e(n)$ the numbers of key comparisons and key exchanges during the first partitioning step. Furthermore, we denote by J_n the size of the sublist containing the element we are looking for (with the convention $J_n = 0$ if the chosen pivot itself has the desired rank R_n). Then, J_n is a size-biased version of I_n with distribution

$$\mathbb{P}(J_n = 0) = \frac{1}{n}, \quad \mathbb{P}(J_n = j) = \frac{2j}{n^2}, \quad j = 1, \dots, n-1. \quad (4.22)$$

For $\iota = c, e$, we have $Y_0^{[\iota]} = 0$ and the recursive relation

$$Y_n^{[\iota]} \stackrel{d}{=} Y_{J_n}^{[\iota]} + b_\iota(n), \quad n \geq 1,$$

where $(J_n, b_\iota(n))$ and $(Y_n^{[\iota]})_{n \geq 0}$ are independent, J_n has the distribution given in (4.22), $b_c(n) = n-1$ and $b_e(n)$ has a mixed hypergeometric $\text{Hyp}(n-1, I_n, n-1-I_n)$ distribution. However, the number $b_e(n)$ of key exchanges during the first partitioning step coincides not only with the number of large elements among the first I_n elements of the permutation but also with the number of misplaced elements in the sublist the algorithm recurses on (at least if the element to be selected is not the pivot itself). Thus, we can and will make use of the fact that conditional on $\{J_n = j\}$, $j = 1, \dots, n-1$, we have that $b_e(n)$ is hypergeometrically $\text{Hyp}(n-1, j, n-1-j)$ distributed. Mahmoud, Modarres and Smythe [62, Theorem 2] have shown that, as $n \rightarrow \infty$,

$$\mu_c(n) := \mathbb{E}[Y_n^{[c]}] = 3n - 8 \log n + O(1).$$

Furthermore, similarly as in Mahmoud [60] (where he analyzed the number of data moves which is essentially twice the number of key exchanges), we obtain

$$\mu_e(n) := \mathbb{E}[Y_n^{[e]}] = \frac{1}{2}n - \frac{5}{3}\mathcal{H}_n + \frac{49}{18} - \frac{5\mathcal{H}_n}{3n} + \frac{1}{18n} = \frac{1}{2}n + O(\log n).$$

Just as in the previous section, for $\iota = c, e$, we define the centered quantities $X_0^{[\iota]} := 0$ and

$$X_n^{[\iota]} := \frac{1}{n}(Y_n^{[\iota]} - \mu_\iota(n)), \quad n \geq 1.$$

Then we have

$$X_n^{[\iota]} \stackrel{d}{=} A_1^{(n)} X_{J_n}^{[\iota]} + b_\iota^{(n)}, \quad n \geq 1,$$

where $(A_1^{(n)}, b_\iota^{(n)}, J_n)$ and $(X_n^{[\iota]})_{n \geq 0}$ are independent,

$$A_1^{(n)} = \frac{J_n}{n} \quad \text{and} \quad b_\iota^{(n)} = \frac{1}{n}(b_\iota(n) - \mu_\iota(n) + \mu_\iota(J_n)).$$

Since J_n/n converges to \sqrt{U} (a size-biased version of a uniform random variable U on $[0, 1]$) and $b_e(n)/n$ to $\sqrt{U}(1 - \sqrt{U})$, this suggests the limiting equation

$$X^{[l]} \stackrel{d}{=} A_1^* X^{[l]} + b_l^*, \quad (4.23)$$

where $A_1^* = \sqrt{U}$ with U uniformly distributed on $[0, 1]$, $b_c^* := 3\sqrt{U} - 2$, $b_e^* := \frac{1}{2}\sqrt{U} + \sqrt{U}(1 - \sqrt{U}) - \frac{1}{2}$ and $X^{[l]}$ is independent of U . Applying Theorem 3.1, we now derive the corresponding convergence rates for the Zolotarev metrics.

Theorem 4.19. *The number $Y_n^{[c]}$ of key comparisons and the number $Y_n^{[e]}$ of key exchanges used by Quickselect for finding a uniformly chosen element in a random permutation of n distinct elements satisfy, as $n \rightarrow \infty$,*

$$\zeta_s \left(\frac{1}{n} (Y_n^{[c]} - \mathbb{E}[Y_n^{[c]}]), X^{[c]} \right) = \begin{cases} O(\log n/n), & s = 1, \\ O(1/n), & s \in (1, 2), \\ O(\sqrt{\log n}/n), & s = 2, \end{cases}$$

and for $s \in [1, 2]$

$$\zeta_s \left(\frac{1}{n} (Y_n^{[e]} - \mathbb{E}[Y_n^{[e]}]), X^{[e]} \right) = O\left(\frac{1}{n^{1/2}}\right),$$

where the distribution of $X^{[c]}$ ($X^{[e]}$, respectively) is the unique solution of equation (4.23) in the space of all probability distributions on the real line.

Proof. We start by determining the order of $\|A_1^{(n)} - A_1^*\|_s$. For this, we use the embedding

$$J_n = \mathbf{1}_{\{U \geq 1/n\}} \lceil \sqrt{n^2 U - n + 1/4} - 1/2 \rceil \quad \text{and} \quad A_1^* = \sqrt{U}$$

with a uniform random variable U on $[0, 1]$. Note that the quantile function of A_1^* is given by $u \mapsto \sqrt{u}$ and the quantile function of J_n by $u \mapsto \mathbf{1}_{\{u \geq 1/n\}} \lceil \sqrt{n^2 u - n + 1/4} - 1/2 \rceil$ for $0 < u < 1$, which justifies this embedding (see the construction of optimal ℓ_s -couplings in Section 2.2.1). Thus, we find

$$\begin{aligned} \|A_1^{(n)} - A_1^*\|_s^s &= \int_0^1 \left| \mathbf{1}_{\{u \geq 1/n\}} \frac{1}{n} \lceil \sqrt{n^2 u - n + 1/4} - 1/2 \rceil - \sqrt{u} \right|^s du \\ &= \int_{1/n}^1 \left(\sqrt{u} - \frac{1}{n} \lceil \sqrt{n^2 u - n + 1/4} - 1/2 \rceil \right)^s du + O(n^{-s/2-1}) \\ &= \sum_{k=1}^{n-1} \int_{v_{k-1}}^{v_k} \left(\sqrt{u} - \frac{1}{n} \lceil \sqrt{n^2 u - n + 1/4} - 1/2 \rceil \right)^s du + O(n^{-s/2-1}), \end{aligned}$$

where $v_k := (k^2 + k + n)/n^2$ for $k = 0, \dots, n-1$. Since $\lceil \sqrt{n^2 u - n + 1/4} - 1/2 \rceil = k$ for $v_{k-1} < u \leq v_k$, we have ($k = 1, \dots, n-1$)

$$\begin{aligned} \int_{v_{k-1}}^{v_k} \left(\sqrt{u} - \frac{1}{n} \lceil \sqrt{n^2 u - n + 1/4} - 1/2 \rceil \right)^s du &= \int_{v_{k-1}}^{v_k} \left(\sqrt{u} - \frac{k}{n} \right)^s du \\ &\leq (v_k - v_{k-1}) \left(\sqrt{v_k} - \frac{k}{n} \right)^s = \frac{2k}{n^{2+s}} (\sqrt{k^2 + k + n} - k)^s \leq \frac{2k}{n^{2+s}} \left(\frac{n}{k} \right)^s = 2 \frac{k^{1-s}}{n^2}. \end{aligned}$$

This implies, for $1 \leq s < 2$,

$$\|A_1^{(n)} - A_1^*\|_s^s \leq \frac{2}{n^2} \sum_{k=1}^{n-1} k^{1-s} + O(n^{-s/2-1}) = O(n^{-s}), \quad (4.24)$$

i.e., $\|A_1^{(n)} - A_1^*\|_s = O(n^{-1})$ for $1 \leq s < 2$. For $s = 2$, similarly to (4.24), we find

$$\|A_1^{(n)} - A_1^*\|_2^2 \leq \frac{2}{n^2} \sum_{k=1}^{n-1} \frac{1}{k} + O(n^{-2}) = O\left(\frac{\log n}{n^2}\right).$$

Note that the order $\log n/n^2$ is already the true rate for $s = 2$ with the above embedding since we have

$$\|A_1^{(n)} - A_1^*\|_2^2 \geq \sum_{k=1}^{n-1} (v_k - v_{k-1}) \left(\sqrt{v_{k-1}} - \frac{k}{n}\right)^2 + O(n^{-2}) = \Theta\left(\frac{\log n}{n^2}\right).$$

Furthermore, applying Lemma 3.8 with $f(n) = 3n - 8 \log(n)$, $e(n) = 1$, $g(n) = n^2$ and $h(n) = 0$, we obtain

$$\|b_c^{(n)} - b_c^*\|_s \leq \|\bar{b}_c^{(n)} - b_c^*\|_s + O(n^{-1}),$$

where for the first summand, we have (similarly as in the proof of Theorem 4.18)

$$\begin{aligned} \|\bar{b}_c^{(n)} - b_c^*\|_s &= \left\| \frac{1}{n} (n - 1 - (3n - 8 \log(n)) + (3J_n - 8 \log(J_n \vee 1))) - (3\sqrt{U} - 2) \right\|_s \\ &\leq \left\| \frac{1}{n} \right\|_s + 3 \left\| \frac{J_n}{n} - \sqrt{U} \right\|_s + \frac{8}{n} \left\| \log\left(\frac{n}{J_n \vee 1}\right) \right\|_s \\ &= \begin{cases} O(1/n), & s \in [1, 2), \\ O(\sqrt{\log n}/n), & s = 2. \end{cases} \end{aligned}$$

For $s > 1$, condition (3.8) is satisfied for $R(n) = n^{-1}$ since as $n \rightarrow \infty$

$$\mathbb{E} \left[\frac{R(J_n)}{R(n)} |A_1^{(n)}|^s \right] = \mathbb{E} \left[\left(\frac{J_n}{n}\right)^{s-1} \right] \rightarrow \mathbb{E}[(\sqrt{U})^{s-1}] = \frac{2}{s+1} < 1.$$

For $s = 1$, we proceed in the same way as in the proof of Theorem 4.18 and use the inequality $\sum_{k=2}^{n-1} k \log k \leq \frac{1}{2} n^2 \log n - \frac{1}{4} n^2$ to show that the condition stated in Remark 3.3 is satisfied with $R(n) = \log n/n$ and $\tilde{R}(n) = 1/n$:

$$\mathbb{E} \left[\mathbf{1}_{\{J_n \geq 1\}} \frac{R(J_n)}{R(n)} A_1^{(n)} \right] = \mathbb{E} \left[\frac{\log(J_n \vee 1)}{\log n} \right] \leq 1 - \frac{1}{2 \log n}.$$

Similarly as in the proof of Theorem 4.18, the result for the number of key exchanges follows immediately from the previous observations and Theorem 3.1 if we show that

$$\left\| \frac{b_e(n)}{n} - \sqrt{U}(1 - \sqrt{U}) \right\|_s = O(n^{-1/2})$$

for $s \geq 1$. For this, recall that on the event $\{J_n = j\}$, $j = 1, \dots, n-1$, $b_e(n)$ is hypergeometrically $\text{Hyp}(n-1, j, n-1-j)$ distributed. We can now proceed in very much the same way as it was done in Dadoun and Neininger [14, Lemmas 3.1 and 3.2]. For this, we denote by $Z_n^{(j)}$, $j = 1, \dots, n-1$, a random variable with hypergeometric $\text{Hyp}(n-1, j, n-1-j)$ distribution and use Theorem 3.1 in Serfling [87] to obtain the following moment inequalities for the hypergeometric distribution: For $j = 1, \dots, n-1$ and $s > 0$, we have

$$\mathbb{E} \left[\left| \frac{Z_n^{(j)}}{n} - \frac{j(n-1-j)}{n(n-1)} \right|^s \right] \leq \frac{\Gamma(s/2+1)}{2^{s/2+1}} n^{-s/2}, \quad (4.25)$$

where Γ denotes Euler's gamma function. Hence, we find for $s \in [1, 2]$

$$\begin{aligned} & \left\| \frac{b_e(n)}{n} - \sqrt{U}(1 - \sqrt{U}) \right\|_s \\ & \leq \left\| \frac{b_e(n)}{n} - \frac{J_n(n-1-J_n)}{n(n-1)} \right\|_s + \left\| \frac{J_n(n-1-J_n)}{n(n-1)} - \sqrt{U}(1 - \sqrt{U}) \right\|_s = O(n^{-1/2}) \end{aligned}$$

by the uniform bound (4.25) and the estimates

$$\begin{aligned} \left\| \frac{J_n(n-1-J_n)}{n(n-1)} - \sqrt{U}(1 - \sqrt{U}) \right\|_s & \leq \left\| \frac{J_n}{n} - \sqrt{U} \right\|_s + \left\| \frac{J_n^2}{n(n-1)} - U \right\|_s \\ & \leq \left\| \frac{J_n}{n} - \sqrt{U} \right\|_s + \left\| \frac{J_n^2}{n^2} - U \right\|_s + O\left(\frac{1}{n}\right) \end{aligned}$$

and (with the same notation and arguments as before)

$$\begin{aligned} \left\| \frac{J_n^2}{n^2} - U \right\|_s & = \sum_{k=1}^{n-1} \int_{v_{k-1}}^{v_k} \left(u - \frac{k^2}{n^2}\right)^s du + O(n^{-(s+1)}) \\ & \leq \sum_{k=1}^{n-1} (v_k - v_{k-1}) \left(v_k - \frac{k^2}{n^2}\right)^s + O(n^{-(s+1)}) \\ & \leq \sum_{k=1}^{n-1} \frac{2k}{n^2} \left(\frac{2}{n}\right)^s + O(n^{-(s+1)}) = O(n^{-s}). \end{aligned}$$

This completes the proof. \square

Remark 4.20. The same argument as in the proof of Theorem 4.18 together with Theorem 2 in Mahmoud, Modarres and Smythe [62] shows that we have $\zeta_2(X_n^{[c]}, X^{[c]}) = O(1/n)$.

4.4.3 Quickselect for finding the smallest element—a further approach

In Section 4.4.1, we showed that the rate of convergence for the normalized number of key comparisons used by Quickselect for finding the smallest element in a list of n randomly

permuted distinct elements is of order $O(1/n)$ in the Zolotarev metric ζ_s ($1 < s \leq 2$) and of order $O(\log n/n)$ in the Zolotarev metric ζ_1 (coinciding with the Wasserstein metric ℓ_1). So far, however, it is not clear whether the latter rate is optimal or whether it might be possible to obtain the rate $O(1/n)$ also for $s = 1$. Throughout this section, let Y_n denote the number of key comparisons used by Quickselect for finding the smallest element in a list of n uniformly permuted distinct data.

The following approach uses an inequality by Rio [79, Theorem 3.1] (see also Rio [78, Theorem 1] for a similar version with different constants) which allows us to transfer the ζ_s rates ($1 < s \leq 2$) from Theorem 4.18 to ℓ_1 rates. Rio's result states the following: For any $s > 1$, there exists a constant $c_s > 0$ such that

$$\ell_s(X, Y) \leq (c_s \zeta_s(X, Y))^{1/s} \quad (4.26)$$

for all L_s -integrable random variables X and Y . For $1 < s \leq 2$, we can choose $c_s = 2s$. Our aim now is to use this inequality to transfer the ζ_s rates to ℓ_s rates and then, by the monotonicity of the Wasserstein ℓ_s metrics, to ℓ_1 rates. For this, we need to make the bounds in Theorem 4.18 explicit.

Theorem 4.21. *The number Y_n of key comparisons used by Quickselect for finding the smallest element in a random permutation of n distinct elements satisfies, for $1 < s \leq 2$ and $n \geq 1$,*

$$\zeta_s\left(\frac{1}{n}(Y_n - \mathbb{E}[Y_n]), X\right) \leq \frac{Cs}{(s-1)n},$$

where, e.g., the constant C can be chosen as $C = 5$ and the distribution of X is the unique solution of the equation $X \stackrel{d}{=} UX + 2U - 1$ with X and U independent and U uniformly distributed on $[0, 1]$.

Proof. The proof follows along the same lines as the proof of Theorem 3.1, except that we use concrete non-asymptotic bounds here. However, note that our goal is to show the existence of such a constant C rather than choosing it as small as possible. Just as in the previous sections, we define the normalized quantities $X_0 := 0$ and

$$X_n := \frac{1}{n}(Y_n - \mu(n)), \quad n \geq 1,$$

with $\mu(n) = \mathbb{E}[Y_n] = 2n - 2\mathcal{H}_n$. Then we have

$$X_n \stackrel{d}{=} \frac{I_n}{n} X_{I_n} + b^{(n)},$$

with independent I_n and $(X_n)_{n \geq 0}$ and $b^{(n)} := \frac{1}{n}(n - 1 - \mu(n) + \mu(I_n))$. For $n \geq 1$, we introduce the accompanying sequence

$$Z_n^* := \frac{I_n}{n} X + b^{(n)},$$

where I_n and X are independent. Note that Z_n^* is L_s -integrable and centered. Hence, the quantities X_n , Z_n^* and X are ζ_s -compatible for $n \geq 1$. Applying the triangle inequality, we have

$$\zeta_s(X_n, X) \leq \zeta_s(X_n, Z_n^*) + \zeta_s(Z_n^*, X). \quad (4.27)$$

Defining $\Delta_s(n) := \zeta_s(X_n, X)$ and $b_n(k) := \frac{1}{n}(n-1-\mu(n)+\mu(k))$, we obtain by conditioning on I_n that, for $n \geq 1$,

$$\begin{aligned} \zeta_s(X_n, Z_n^*) &= \zeta_s\left(\frac{I_n}{n}X_{I_n} + b^{(n)}, \frac{I_n}{n}X + b^{(n)}\right) \\ &= \sup_{f \in \mathcal{F}_s} \left| \frac{1}{n} \sum_{k=0}^{n-1} \left(\mathbb{E}\left[f\left(\frac{k}{n}X_k + b_n(k)\right)\right] - \mathbb{E}\left[f\left(\frac{k}{n}X + b_n(k)\right)\right] \right) \right| \\ &\leq \frac{1}{n} \sum_{k=0}^{n-1} \zeta_s\left(\frac{k}{n}X_k + b_n(k), \frac{k}{n}X + b_n(k)\right) \\ &\leq \frac{1}{n} \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^s \Delta_s(k). \end{aligned} \quad (4.28)$$

To bound the second summand in (4.27), we use Theorem 2.7 to show that $\zeta_s(Z_n^*, X) \leq 2\ell_s(Z_n^*, X)$ for $n \geq 1$ (note that $\mathbb{E}[|X|^s] \leq 1$ by definition of X and $\mathbb{E}[|Z_n^*|^s] \leq 1$ by Lemma 4.24 at the end of this section). With the embedding $I_n = \lfloor nU \rfloor$, the notation $b^* = 2U - 1$ and Lemma 4.23 below, we have

$$\begin{aligned} \ell_s(Z_n^*, X) &\leq \left\| \left(\frac{I_n}{n}X + b^{(n)} \right) - (UX + b^*) \right\|_s \\ &\leq \left\| \frac{I_n}{n} - U \right\|_s \|X\|_s + \|b^{(n)} - b^*\|_s \\ &\leq \frac{1}{\sqrt{3n}} \frac{1}{\sqrt{2}} + \sqrt{\frac{13}{3}} \cdot \frac{1}{n} \\ &= \frac{1 + \sqrt{26}}{\sqrt{6n}} \leq \frac{5}{2n}. \end{aligned}$$

We now choose

$$K_s := \frac{5s}{s-1}$$

and obtain $\Delta_s(n) \leq \frac{K_s}{n}$ by induction: For $n = 1$, we have $\Delta_s(1) = \zeta_s(0, X) \leq \mathbb{E}[|X|^s] \leq K_s$.

For $n > 1$, using the induction hypothesis, we obtain from (4.27) and (4.28) that

$$\begin{aligned}
\Delta_s(n) &\leq \zeta_s(X_n, Z_n^*) + \zeta_s(Z_n^*, X) \\
&\leq \frac{1}{n} \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^s \Delta_s(k) + \frac{5}{n} \\
&\leq \frac{K_s}{n^{1+s}} \sum_{k=1}^{n-1} k^{s-1} + \frac{5}{n} \\
&\leq \frac{K_s}{n^{1+s}} \cdot \frac{1}{s} \cdot n^s + \frac{5}{n} \\
&= \frac{K_s}{n},
\end{aligned}$$

which finishes the proof. \square

Remark 4.22. Note that the bounds of the just presented proof (except for the induction step at the end) also apply to the case $s = 1$, where we obtain the bound

$$\zeta_1\left(\frac{1}{n}(Y_n - \mathbb{E}[Y_n]), X\right) \leq \frac{5\mathcal{H}_n}{n}, \quad n \geq 1.$$

We now want to transfer the ζ_s rates $\zeta_s(X_n, X) \leq \frac{5s}{(s-1)n}$ from Theorem 4.21 to ℓ_1 rates. By the monotonicity of the ℓ_s metrics and Rio's inequality (4.26), we obtain for $1 < s \leq 2$ and $n \geq 1$

$$\ell_1(X_n, X) \leq \ell_s(X_n, X) \leq (2s \cdot \zeta_s(X_n, X))^{1/s},$$

which—in combination with Theorem 4.21—gives the bound

$$\ell_1(X_n, X) \leq \left(\frac{10s^2}{(s-1)n}\right)^{1/s} \leq 10 \left(\frac{1}{(s-1)n}\right)^{1/s} =: 10f_n(s).$$

We now minimize f_n on $(1, 2]$ and find that f_n is minimal for $s^* = 1 + (W(n/e))^{-1}$ with minimal value $f_n(s^*) = e^{-W(n/e)} = \frac{e}{n}W(n/e)$ for $n \geq 8$, where W denotes the Lambert W function (the inverse relation of the function $x \mapsto xe^x$). Note that we have $W(x) \geq \log x - \log \log x$ for $x \geq e$, which gives for $n \geq 8$

$$\ell_1(X_n, X) \leq 10f_n(s^*) = 10e^{-W(n/e)} \leq 10e \frac{\log n}{n}.$$

However, with this approach, we cannot improve the rate $\log n/n$ since we have, on the other hand,

$$f_n(s^*) = \frac{e}{n}W\left(\frac{n}{e}\right) \geq \frac{e}{n} \left(\log(n/e) - \log \log(n/e)\right) \sim e \frac{\log n}{n}.$$

Thus, in conclusion, it can be said that we derived concrete non-asymptotic upper bounds for the rate of convergence in the Wasserstein ℓ_1 metric in this section, but it remains unclear whether the rate $\log n/n$ is optimal.

To conclude this section, we present the two technical lemmas which were used in the proof of Theorem 4.21. The first one gives concrete non-asymptotic upper bounds for the convergence rates of the coefficients.

Lemma 4.23. *Let $I_n = \lfloor nU \rfloor$ for some uniform random variable U on $[0, 1]$, $b^{(n)} = \frac{1}{n}(n-1-2n+2\mathcal{H}_n+2I_n-2\mathcal{H}_{I_n})$ and $b^* = 2U-1$. Then we have, for $s \in [1, 2]$,*

$$\left\| \frac{I_n}{n} - U \right\|_s \leq \frac{1}{\sqrt{3n}}$$

and

$$\|b^{(n)} - b^*\|_2^2 = \frac{13}{3n^2} - \frac{4}{n^3}\mathcal{H}_n.$$

Proof. To show the first assertion, we calculate

$$\left\| \frac{I_n}{n} - U \right\|_s^s = \sum_{k=0}^{n-1} \int_{k/n}^{(k+1)/n} (u - k/n)^s du = \frac{1}{s+1} n^{-s}$$

and use the fact that $(s+1)^{-1/s} \leq 1/\sqrt{3}$ for $s \in [1, 2]$. For the second part, we obtain (with the convention $\mathcal{H}_0 = 0$)

$$\begin{aligned} \|b^{(n)} - b^*\|_2^2 &= \frac{1}{n^2} \mathbb{E} \left[(-1 + 2(I_n - nU) + 2(\mathcal{H}_n - \mathcal{H}_{I_n}))^2 \right] \\ &= \frac{1}{n^2} \sum_{k=0}^{n-1} \int_{k/n}^{(k+1)/n} (-1 + 2(k - nu) + 2(\mathcal{H}_n - \mathcal{H}_k))^2 du \\ &= \frac{13}{3n^2} - \frac{8}{n^3} \sum_{k=0}^{n-1} (\mathcal{H}_n - \mathcal{H}_k) + \frac{4}{n^3} \sum_{k=0}^{n-1} (\mathcal{H}_n - \mathcal{H}_k)^2. \end{aligned}$$

Using the identities $\sum_{k=0}^{n-1} (\mathcal{H}_n - \mathcal{H}_k) = n$ and $\sum_{k=0}^{n-1} (\mathcal{H}_n - \mathcal{H}_k)^2 = 2n - \mathcal{H}_n$ (which can be proved by induction, e.g.), we find

$$\|b^{(n)} - b^*\|_2^2 = \frac{13}{3n^2} - \frac{4}{n^3}\mathcal{H}_n,$$

which is the second assertion. □

In the proof of Theorem 4.21, we further used that $\mathbb{E}[|Z_n^*|^s] \leq 1$ which is addressed by the following lemma.

Lemma 4.24. *Let X be the unique solution of the equation $X \stackrel{d}{=} UX + 2U - 1$ with U uniform on $[0, 1]$ and independent of X , $I_n = \lfloor nU \rfloor$, $b^{(n)} = \frac{1}{n}(n-1-2n+2\mathcal{H}_n+2I_n-2\mathcal{H}_{I_n})$ and $Z_n^* = \frac{I_n}{n}X + b^{(n)}$. Then we have $\mathbb{E}[|Z_n^*|^s] \leq 1$ for any $n \geq 1$ and $s \in [1, 2]$.*

Proof. By the monotonicity of the L_s norms, it is enough to show $\|Z_n^*\|_2 \leq 1$ for any $n \geq 1$. Since X is centered and independent of $(I_n, b^{(n)})$, we obtain

$$\|Z_n^*\|_2^2 = \mathbb{E} \left[\left(\frac{I_n}{n} X + b^{(n)} \right)^2 \right] = \frac{1}{n^2} \mathbb{E}[I_n^2] \mathbb{E}[X^2] + \mathbb{E}[(b^{(n)})^2].$$

Using the elementary identities $\sum_{k=0}^{n-1} (\mathcal{H}_n - \mathcal{H}_k) = n$, $\sum_{k=0}^{n-1} k(\mathcal{H}_n - \mathcal{H}_k) = n(n-1)/4$ and $\sum_{k=0}^{n-1} (\mathcal{H}_n - \mathcal{H}_k)^2 = 2n - \mathcal{H}_n$, the second summand can be explicitly computed:

$$\mathbb{E}[(b^{(n)})^2] = \frac{1}{n^3} \sum_{k=0}^{n-1} (2k - (n+1) + 2\mathcal{H}_n - 2\mathcal{H}_k)^2 = \frac{1}{3} - \frac{2}{n} + \frac{17}{3n^2} - \frac{4\mathcal{H}_n}{n^3} \leq \frac{1}{3}.$$

This, together with $\mathbb{E}[X^2] = 0.5$ and $\mathbb{E}[I_n^2] = (n-1)(2n-1)/6 \leq n^2/3$, implies

$$\|Z_n^*\|_2^2 \leq \frac{1}{6} + \frac{1}{3} = \frac{1}{2},$$

which finishes the proof. □

Chapter 5

A General Theorem with Relaxed Independence Conditions

In the previous sections, we always analyzed quantities $(Y_n)_{n \geq 0}$ satisfying the recurrence relation

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

Apart from various other conditions, we always assumed the conditional independence condition (1.3), i.e., that

$$(A_1(n), \dots, A_K(n), b_n, I^{(n)}), (Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0} \quad \text{are independent.}$$

However, recent work shows that in some special cases with normal limits, it is possible to use ideas of the contraction method even though the random variables $(Y_n^{(r)})_{n \geq 0}$ and the toll term b_n show dependencies (see Neininger [68] for refined Quicksort asymptotics or Müller and Neininger [66] for the composition of cyclic urns). Therefore, the aim of this chapter is to derive a general convergence theorem similar to Theorem 3.4, with the difference that we replace the conditional independence condition (1.3) by the (weaker) partial conditional independence condition (1.4), i.e., by the assumption that

$$(A_1(n), \dots, A_K(n), I^{(n)}), (Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0} \quad \text{are independent.}$$

Note that this weakened independence condition allows additional dependencies between the toll function b_n and the quantities $(Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}$. We start with a general (multivariate) convergence theorem in Section 5.1 and then state a more convenient form for univariate applications in Section 5.2. Several examples of how these theorems can be used are given in Section 5.3.

5.1 A general theorem for convergence rates

In this section, we consider a sequence $(Y_n)_{n \geq 0}$ of d -dimensional random vectors satisfying the distributional recursion

$$Y_n \stackrel{d}{=} \sum_{r=1}^K A_r(n) Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq n_0, \quad (5.1)$$

where $n_0 \in \mathbb{N}$, the coefficients $A_1(n), \dots, A_K(n)$ are random $(d \times d)$ -matrices, b_n is a d -dimensional random vector, $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$ is a random vector in $\{0, \dots, n\}^K$ and all appearing quantities have finite third absolute moments. Furthermore, we assume that the self-similarity condition (1.2) and the partial conditional independence condition (1.4) are satisfied. Apart from the fact that the conditional independence condition (1.3) is replaced by the weaker partial conditional independence condition (1.4) now, this is exactly the framework of the previous sections.

As before, we assume that there exists some $n_1 \in \mathbb{N}_0$ such that the covariance matrix of Y_n is positive definite for $n \geq n_1$ and define the normalized sequence $(X_n)_{n \geq 0}$ by

$$X_n := C_n^{-1/2}(Y_n - M_n), \quad n \geq 0, \quad (5.2)$$

where M_n is chosen as the mean vector of Y_n and C_n as the covariance matrix of Y_n for $n \geq n_1$ (and $C_n = \text{Id}_d$ for $n < n_1$). The normalized quantities satisfy the following modified recursion:

$$X_n \stackrel{d}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \quad (5.3)$$

with

$$A_r^{(n)} := C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2} \left(b_n - M_n + \sum_{r=1}^K A_r(n) M_{I_r^{(n)}} \right) \quad (5.4)$$

and self-similarity and independence relations as in (5.1).

To obtain a convergence result $X_n \rightarrow \mathcal{N}(0, \text{Id}_d)$, we assume that the coefficients appearing in (5.4) converge appropriately, i.e., that there exist L_3 -integrable random variables A_1^*, \dots, A_K^* such that, as $n \rightarrow \infty$,

$$(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L_3} (A_1^*, \dots, A_K^*, 0),$$

with $A_1^*(A_1^*)^\top + \dots + A_K^*(A_K^*)^\top = \text{Id}_d$ almost surely. Then, from (5.3), we expect a limit X of X_n to satisfy the distributional fixed-point equation

$$X \stackrel{d}{=} \sum_{r=1}^K A_r^* X^{(r)}, \quad (5.5)$$

where $(A_1^*, \dots, A_K^*), X^{(1)}, \dots, X^{(K)}$ are independent and $X^{(r)} \stackrel{d}{=} X$ for $r = 1, \dots, K$. As explained earlier, under the additional assumption $\sum_{r=1}^K \mathbb{E}[\|A_r^*\|_{\text{op}}^3] < 1$, the standard normal distribution is the unique solution of equation (5.5) in the space $\mathcal{P}_3^d(0, \text{Id}_d)$ of L_3 -integrable probability distributions with mean vector 0 and covariance matrix Id_d .

The subject of this section is to derive a general convergence result corresponding to Theorem 3.4 for normal limits, except that the independence condition (1.3) required there is replaced by the weaker condition (1.4) allowing dependence on b_n now. In return, to bound the Zolotarev distance $\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d))$, we need slightly amended conditions on the convergence rates of the coefficients here, namely that

$$\left\| \sum_{r=1}^K \mathbb{E}[A_r^{(n)}(A_r^{(n)})^T] - \text{Id}_d \right\|_{\text{op}} + \left\| \sum_{r=1}^K A_r^{(n)}(A_r^{(n)})^T - \text{Id}_d \right\|_{3/2}^{3/2} + \|b^{(n)}\|_3 = O(R(n)) \quad (5.6)$$

for some monotonically decreasing sequence $R(n) > 0$ with $R(n) \rightarrow 0$. Furthermore, we now need the technical conditions

$$\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_2^2 + \|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_3^3 = O(R(n)), \quad n \rightarrow \infty, \quad (5.7)$$

for all $\ell \in \mathbb{N}$ and $r = 1, \dots, K$ and

$$\|\mathbf{1}_{\{I_r^{(n)} = n\}} A_r^{(n)}\|_3 \rightarrow 0, \quad n \rightarrow \infty, \quad (5.8)$$

for all $r = 1, \dots, K$. Under these assumptions, we can deduce the following result which, in the main, corresponds to Theorem 3.4, the difference being only the somewhat stronger conditions (5.6) and (5.7) as compensation for the relaxed independence assumption.

Theorem 5.1. *Let $(X_n)_{n \geq 0}$ be given as in (5.2) with $(Y_n)_{n \geq 0}$ satisfying the distributional recurrence (5.1). Furthermore, assume that the coefficients $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ defined in (5.4) converge to $(A_1^*, \dots, A_K^*, 0)$ in the L_3 norm as $n \rightarrow \infty$ with $\sum A_r^*(A_r^*)^T = \text{Id}_d$ almost surely. If conditions (5.6), (5.7) and (5.8) are satisfied and if*

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^K \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^3 \right] < 1, \quad (5.9)$$

then we have, as $n \rightarrow \infty$,

$$\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d)) = O(R(n)).$$

Remark 5.2. Before giving the proof of this theorem, we make two short comments:

- (1) Looking at condition (5.6) in Theorem 5.1, we find the assumption $\|b^{(n)}\|_3 = O(R(n))$, i.e., compared to the corresponding condition (3.15) for the toll function in Theorem 3.4 ($\|b^{(n)}\|_3^3 = O(R(n))$), there is no additional factor 3 in the exponent here—which is due to the fact that we allow additional dependencies. It is, however, not clear whether the proof of Theorem 5.1 might be modified to “regain” the factor 3.

- (2) In some applications, we are only interested in showing (weak) convergence rather than estimating the rate of convergence. In this case, the formulation of Theorem 5.1 is more complex than necessary. More specifically, if no rates are needed, we can replace condition (5.9) by the assumption that

$$\sum_{r=1}^K \mathbb{E}[\|A_r^*\|_{\text{op}}^3] < 1.$$

Furthermore, condition (5.6) can be dropped and instead of condition (5.7), it is enough to assume that

$$\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_3 \rightarrow 0$$

as $n \rightarrow \infty$ for any $\ell \in \mathbb{N}$ and $r = 1, \dots, K$. With these modified conditions, similar arguments as in the proof of Theorem 5.1 can be used to show that $\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d))$ converges to zero as $n \rightarrow \infty$ (and hence, X_n converges in distribution to the d -dimensional standard normal distribution).

In the rest of this section, we give the proof of Theorem 5.1 which follows along the same lines as the proof of Theorems 3.1 and 3.4. To handle the additional dependencies, we adopt and generalize the main ideas used in the proof of Theorem 1.1 in Neininger [68] for refined Quicksort asymptotics. Recall the bound

$$\zeta_3(V_1 + W_1, V_2 + W_2) \leq \zeta_3(V_1, V_2) + \zeta_3(W_1, W_2)$$

for independent pairs (V_1, V_2) and (W_1, W_2) both being ζ_3 -compatible. For the proof of Theorem 5.1, we need a slightly stronger version of this bound which also applies if the appearing quantities are dependent. The following statement can be found in Müller and Neininger [66, Lemma 3.4] (see also Neininger [68, Lemma 2.1] for a similar statement for the univariate case).

Lemma 5.3. *Let V_1, V_2, W_1 and W_2 be random vectors in \mathbb{R}^d such that the pairs (V_1, V_2) and $(V_1 + W_1, V_2 + W_2)$ are ζ_3 -compatible. Then we have*

$$\zeta_3(V_1 + W_1, V_2 + W_2) \leq \zeta_3(V_1, V_2) + \sum_{i=1}^2 \left(\|V_i\|_3^2 \|W_i\|_3 + \frac{\|V_i\|_3 \|W_i\|_3^2}{2} + \frac{\|W_i\|_3^3}{2} \right).$$

With the help of this statement, we can now give the proof of Theorem 5.1.

Proof of Theorem 5.1. To establish a rate of convergence, we set

$$Z_n := \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} N^{(r)},$$

where $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)})$, $N^{(1)}, \dots, N^{(K)}$ are independent, $N^{(r)}$ has the d -dimensional standard normal distribution for $r = 1, \dots, K$ and the deterministic sequence $(T_n)_{n \geq 0}$

of positive semidefinite matrices is defined by $T_n = (\text{Cov}(X_n))^{1/2}$ for $n \geq 0$ (i.e., $T_n = \text{Id}_d$ for $n \geq n_1$). Note that X_n and Z_n are not necessarily ζ_3 -compatible. In fact, we have

$$\begin{aligned} \text{Cov}(Z_n) &= \sum_{r=1}^K \mathbb{E}[A_r^{(n)} T_{I_r^{(n)}} T_{I_r^{(n)}}^T (A_r^{(n)})^T] \\ &= \sum_{r=1}^K \mathbb{E}[A_r^{(n)} (A_r^{(n)})^T] + \sum_{r=1}^K \mathbb{E}[\mathbf{1}_{\{I_r^{(n)} < n_1\}} A_r^{(n)} (T_{I_r^{(n)}} T_{I_r^{(n)}}^T - \text{Id}_d) (A_r^{(n)})^T] \\ &= \text{Id}_d + O(R(n)) \end{aligned} \quad (5.10)$$

by (5.6) and (5.7), where the $O(R(n))$ -term can be read componentwise (or in operator norm). Hence, Z_n does not necessarily have covariance matrix Id_d , but we can choose a deterministic sequence $(M_n)_{n \geq 0}$ of matrices converging to Id_d (componentwise or in operator norm) such that $\text{Cov}(M_n Z_n) = \text{Id}_d$ for all large enough n . To be more precise, since $\text{Cov}(Z_n)$ is positive definite for large n (say for $n \geq n_1$, where n_1 may need to be enlarged), we can choose $M_n = (\text{Cov}(Z_n))^{-1/2}$ for $n \geq n_1$. With this choice, we have $M_n = (\text{Id}_d + O(R(n)))^{-1/2} = \text{Id}_d + O(R(n))$ by (5.10) and Remark 2.4. As a result, we have $\|M_n - \text{Id}_d\|_{\text{op}} = O(R(n))$. In the following, we will use the notation $K_n := M_n - \text{Id}_d$, such that the quantity

$$Z_n^* := (\text{Id}_d + K_n) Z_n$$

has covariance matrix Id_d for $n \geq n_1$. Hence, the quantities X_n , Z_n^* and $\mathcal{N}(0, \text{Id}_d)$ are ζ_3 -compatible for $n \geq n_1$. Applying the triangle inequality, we have, for $n \geq n_1$,

$$\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d)) \leq \zeta_3(X_n, Z_n^*) + \zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)). \quad (5.11)$$

We start by estimating the first summand in (5.11). For this, we introduce the notation $Q_n = A_1^{(n)} X_{I_1^{(n)}}^{(1)} + \dots + A_K^{(n)} X_{I_K^{(n)}}^{(K)}$ and use Lemma 5.3 to obtain

$$\begin{aligned} \zeta_3(X_n, Z_n^*) &= \zeta_3(Q_n + b^{(n)}, Z_n + K_n Z_n) \\ &\leq \zeta_3(Q_n, Z_n) + \|Q_n\|_3^2 \|b^{(n)}\|_3 + \frac{1}{2} \|Q_n\|_3 \|b^{(n)}\|_3^2 + \frac{1}{2} \|b^{(n)}\|_3^3 \\ &\quad + \left(\|K_n\|_{\text{op}} + \frac{1}{2} \|K_n\|_{\text{op}}^2 + \frac{1}{2} \|K_n\|_{\text{op}}^3 \right) \|Z_n\|_3^3. \end{aligned} \quad (5.12)$$

Using the triangle inequality together with $\|X_n\|_3 = O(1)$ (see Lemma 5.4 below) and the fact that $A_r^{(n)}$ converges in L_3 for $r = 1, \dots, K$ (from which follows that $\|A_r^{(n)}\|_3$ is bounded in n for any $r = 1, \dots, K$), we find

$$\|Q_n\|_3 \leq \sum_{r=1}^K \|A_r^{(n)} X_{I_r^{(n)}}^{(r)}\|_3 = O(1).$$

In a similar manner, we find $\|Z_n\|_3 = O(1)$, which implies, together with inequality (5.12) and $\|b^{(n)}\|_3 + \|K_n\|_{\text{op}} = O(R(n))$,

$$\zeta_3(X_n, Z_n^*) \leq \zeta_3(Q_n, Z_n) + O(R(n)).$$

Similarly as in the proof of Theorem 3.1, we condition on $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)})$ and use the notation $\Delta(n) := \zeta_3(X_n, \mathcal{N}(0, \text{Id}_d))$ to obtain, for $n \geq n_1$,

$$\begin{aligned} \zeta_3(Q_n, Z_n) &\leq \left(\mathbb{E} \sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)}=n\}} \|A_r^{(n)}\|_{\text{op}}^3 \right) \Delta(n) + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\text{op}}^3 \Delta(I_r^{(n)}) \right] \\ &\quad + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} < n_1\}} \|A_r^{(n)}\|_{\text{op}}^3 \sup_{k < n_1} \zeta_3(X_k, T_k N^{(r)}) \right]. \end{aligned} \quad (5.13)$$

Note that the last summand is of order $O(R(n))$ by condition (5.7). To bound the second summand $\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d))$ in (5.11), we use Lemma 3.6: By the convolution property of the normal distribution, we have $Z_n^* \stackrel{d}{=} G_n N$ with

$$G_n := \left(\sum_{r=1}^K (\text{Id}_d + K_n) A_r^{(n)} T_{I_r^{(n)}} T_{I_r^{(n)}}^\top (A_r^{(n)})^\top (\text{Id}_d + K_n)^\top \right)^{1/2},$$

N having the d -dimensional standard normal distribution and being independent of G_n . By definition of K_n , we have $\text{Cov}(Z_n^*) = \text{Id}_d$ for $n \geq n_1$. In addition, conditions (5.6) and (5.7) imply that $\|G_n G_n^\top - \text{Id}_d\|_{3/2}^{3/2} = O(R(n))$. Thus, by Lemma 3.6, we obtain $\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)) = O(R(n))$. Hence, putting everything together and introducing the notation $p_n := \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)}=n\}} \|A_r^{(n)}\|_{\text{op}}^3 \right]$, we have

$$\Delta(n) \leq p_n \Delta(n) + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\text{op}}^3 \Delta(I_r^{(n)}) \right] + O(R(n)). \quad (5.14)$$

Using the assumptions $p_n \rightarrow 0$ (condition (5.8)) and (5.9), we can now proceed in exactly the same way as in the proof of Theorem 3.1 and complete the proof inductively. \square

To bound the right hand side of (5.12) in the proof of Theorem 5.1, we used the assumption that $\|X_n\|_3$ is bounded in n . The following lemma shows that this assumption is indeed satisfied under the conditions of Theorem 5.1 (note that condition (5.9) implies that $\sum \mathbb{E}[\|A_r^*\|_{\text{op}}^3] < 1$).

Lemma 5.4. *Let $(X_n)_{n \geq 0}$ be given as in (5.2) with $(Y_n)_{n \geq 0}$ satisfying the distributional recurrence (5.1). Furthermore, assume that the coefficients $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ defined in (5.4) converge to $(A_1^*, \dots, A_K^*, 0)$ in the L_3 norm as $n \rightarrow \infty$ with $\sum A_r^* (A_r^*)^\top = \text{Id}_d$ almost surely and $\sum \mathbb{E}[\|A_r^*\|_{\text{op}}^3] < 1$ and that the technical condition (5.8) is satisfied. Then we have, as $n \rightarrow \infty$, $\|X_n\|_3 = O(1)$.*

Proof. The proof follows along the same lines as the proof of Lemma 2.3 in Neininger [68], except that we consider a more general setting with multivariate quantities and an arbitrary

number K of subproblems here. Clearly, by the triangle inequality and the fact that X_n satisfies recursion (5.3) for $n \geq n_0$, we have for $n \geq n_0$

$$\|X_n\| \leq \sum_{r=1}^K \|A_r^{(n)}\|_{\text{op}} \|X_{I_r^{(n)}}^{(r)}\| + \|b^{(n)}\| = \sum_{r=1}^K \Lambda_{r,n} + \|b^{(n)}\|,$$

where we use the notation $\Lambda_{r,n} = \|A_r^{(n)}\|_{\text{op}} \|X_{I_r^{(n)}}^{(r)}\|$ for $n \geq n_0$ and $r = 1, \dots, K$. Thus, we obtain for $n \geq n_0$

$$\|X_n\|^3 \leq \left(\sum_{r=1}^K \Lambda_{r,n} \right)^3 + 3 \left(\sum_{r=1}^K \Lambda_{r,n} \right)^2 \|b^{(n)}\| + 3 \left(\sum_{r=1}^K \Lambda_{r,n} \right) \|b^{(n)}\|^2 + \|b^{(n)}\|^3. \quad (5.15)$$

We now analyze the expectations of these summands separately. For the last summand, we have $\mathbb{E}[\|b^{(n)}\|^3] = o(1)$ since $b^{(n)}$ converges to zero in L_3 . The second and third summand in (5.15) can be estimated by Hölder's inequality. Just as in [68], we define the quantities

$$\beta_n := 1 \vee \max_{0 \leq j \leq n} \mathbb{E}[\|X_j\|^3]$$

for $n \geq 0$. Recall that the coefficients $A_r^{(n)}$ are L_3 -convergent for $r = 1, \dots, K$, which is why their L_3 norms are uniformly bounded by some constant D . Hence, conditioning on $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)})$ gives the bound

$$\|\Lambda_{r,n}\|_3 \leq \left(\mathbb{E}[\|A_r^{(n)}\|_{\text{op}}^3 \beta_n] \right)^{1/3} \leq D\beta_n^{1/3}.$$

In combination with Hölder's inequality, the L_3 -convergence of $b^{(n)}$ to zero and the fact that $\beta_n \geq 1$, we obtain

$$\mathbb{E}[\|\Lambda_{r,n}\| b^{(n)}\|^2] \leq \|\Lambda_{r,n}\|_3 \|b^{(n)}\|_3^2 = o(1)\beta_n = o(1)(\beta_{n-1} + \mathbb{E}[\|X_n\|^3])$$

for $r = 1, \dots, K$. Similarly, we have for $r, s = 1, \dots, K$

$$\mathbb{E}[\|\Lambda_{r,n}\Lambda_{s,n}\| b^{(n)}\|] \leq \|\Lambda_{r,n}\|_3 \|\Lambda_{s,n}\|_3 \|b^{(n)}\|_3 = o(1)(\beta_{n-1} + \mathbb{E}[\|X_n\|^3]).$$

For the first summand in (5.15), we write

$$\mathbb{E} \left[\left(\sum_{r=1}^K \Lambda_{r,n} \right)^3 \right] = \sum_{r=1}^K \mathbb{E}[\Lambda_{r,n}^3] + 3 \sum_{r=1}^K \sum_{s \neq r} \mathbb{E}[\Lambda_{r,n}^2 \Lambda_{s,n}] + \sum_{r=1}^K \sum_{s \neq r} \sum_{t \neq r,s} \mathbb{E}[\Lambda_{r,n} \Lambda_{s,n} \Lambda_{t,n}].$$

For $r = 1, \dots, K$, again by conditioning on $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)})$ and assumption (5.8), we obtain

$$\mathbb{E}[\Lambda_{r,n}^3] = \mathbb{E} \left[\|A_r^{(n)}\|_{\text{op}}^3 \|X_{I_r^{(n)}}^{(r)}\|^3 \right] \leq \mathbb{E}[\|A_r^{(n)}\|_{\text{op}}^3] \beta_{n-1} + o(1)\mathbb{E}[\|X_n\|^3].$$

Similarly, for $r, s = 1, \dots, K$, $r \neq s$, by conditioning on $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)})$ and using the fact that $(X_n^{(r)})_{n \geq 0}$ and $(X_n^{(s)})_{n \geq 0}$ are independent as well as Hölder's inequality and the boundedness of $\|A_r^{(n)}\|_3$ by D , we have

$$\begin{aligned} \mathbb{E}[\Lambda_{r,n}^2 \Lambda_{s,n}] &= \mathbb{E}\left[\|A_r^{(n)}\|_{\text{op}}^2 \|X_{I_r^{(n)}}^{(r)}\|^2 \|A_s^{(n)}\|_{\text{op}} \|X_{I_s^{(n)}}^{(s)}\|\right] \\ &\leq D^3 \max_{0 \leq j \leq n} \mathbb{E}[\|X_j\|^2] \max_{0 \leq j \leq n} \mathbb{E}[\|X_j\|] = O(1), \end{aligned}$$

since $\mathbb{E}[\|X_j\|^2] = d$ for j large enough (recall that X_j has mean vector 0 and covariance matrix Id_d for $j \geq n_1$) and, by Jensen's inequality, $\mathbb{E}[\|X_j\|] \leq \sqrt{d}$ for $j \geq n_1$. Finally, the same arguments imply

$$\mathbb{E}[\Lambda_{r,n} \Lambda_{s,n} \Lambda_{t,n}] \leq D^3 \left(\max_{0 \leq j \leq n} \mathbb{E}[\|X_j\|] \right)^3 = O(1)$$

for any pairwise different indices $r, s, t = 1, \dots, K$. Putting all estimates together, we have

$$\mathbb{E}[\|X_n\|^3] \leq \left(\sum_{r=1}^K \mathbb{E}[\|A_r^{(n)}\|_{\text{op}}^3] + o(1) \right) \beta_{n-1} + O(1) + o(1) \mathbb{E}[\|X_n\|^3].$$

We now solve this inequality for $\mathbb{E}[\|X_n\|^3]$ and use the fact that the vector $(A_1^{(n)}, \dots, A_K^{(n)})$ converges in L_3 to (A_1^*, \dots, A_K^*) to find

$$\mathbb{E}[\|X_n\|^3] \leq \frac{1}{1 + o(1)} \left(\sum_{r=1}^K \mathbb{E}[\|A_r^*\|_{\text{op}}^3] + o(1) \right) \beta_{n-1} + O(1). \quad (5.16)$$

Note that we assumed that the sum $\mathbb{E}[\|A_1^*\|_{\text{op}}^3 + \dots + \|A_K^*\|_{\text{op}}^3]$ is smaller than 1. Thus, based on (5.16), we can find some constants $C > 0$, $\delta \in (0, 1)$ and $n_2 \in \mathbb{N}$ such that for $n \geq n_2$, we have

$$\mathbb{E}[\|X_n\|^3] \leq (1 - \delta) \beta_{n-1} + C. \quad (5.17)$$

To show that $\mathbb{E}[\|X_n\|^3] = O(1)$, we now check that $\mathbb{E}[\|X_n\|^3] \leq \beta_{n_2} \vee (C/\delta)$ by induction on n : For $n < n_2$, the assertion is clearly true. For $n \geq n_2$, using (5.17) and the induction hypothesis, we have

$$\mathbb{E}[\|X_n\|^3] \leq (1 - \delta) \beta_{n-1} + C \leq (1 - \delta) \left(\beta_{n_2} \vee \frac{C}{\delta} \right) + C \leq \beta_{n_2} \vee \frac{C}{\delta},$$

which finishes the proof. \square

5.2 A more convenient version of the general theorem

As mentioned in Remark 5.2 (2), Theorem 5.1 is more complex than necessary if no bounds on the convergence rate are needed or, moreover, if the quantities are one-dimensional.

Therefore, this section contains an additional theorem for the univariate case which is a more convenient version of Theorem 5.1 if no rates of convergence are needed. In fact, the following theorem is a generalization of the central limit theorem given in Corollary 5.2 in Neininger and Rüschemdorf [70], with the difference that it applies even if we allow dependencies between the toll term b_n and $(Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}$.

From now on, we focus on the univariate case $d = 1$ and consider the special situation where Y_n satisfies the distributional recursion

$$Y_n \stackrel{d}{=} \sum_{r=1}^K Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq n_0, \quad (5.18)$$

where $I^{(n)}, (Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}$ are independent, $Y_j^{(r)}$ has the same distribution as Y_j for $j \geq 0$ and $r = 1, \dots, K$, the subproblem sizes $I_r^{(n)}$ are in $\{0, 1, \dots, n\}$ and satisfy $\mathbb{P}(I_r^{(n)} = n) \rightarrow 0$ as $n \rightarrow \infty$ and all appearing quantities are L_3 -integrable. With the same arguments as in the proof of Theorem 5.1, we derive the following theorem.

Theorem 5.5. *Let $(Y_n)_{n \geq 0}$ be a sequence of random variables in \mathbb{R} that satisfies recursion (5.18). Suppose that, for some positive functions f and g and as $n \rightarrow \infty$,*

$$\mathbb{E}[Y_n] = f(n) + o(g^{1/2}(n)), \quad \text{Var}(Y_n) = g(n) + o(g(n)). \quad (5.19)$$

Further assume that for all $r = 1, \dots, K$ and as $n \rightarrow \infty$, we have the L_3 -convergences

$$\frac{g^{1/2}(I_r^{(n)})}{g^{1/2}(n)} \rightarrow A_r^*, \quad \frac{1}{g^{1/2}(n)} \left(b_n - f(n) + \sum_{r=1}^K f(I_r^{(n)}) \right) \rightarrow 0 \quad (5.20)$$

with $(A_1^)^2 + \dots + (A_K^*)^2 = 1$ almost surely and $\mathbb{P}(\exists r : A_r^* = 1) < 1$. If the technical condition*

$$\frac{\mathbb{P}(I_r^{(n)} < \ell)}{g^{3/2}(n)} \rightarrow 0 \quad (5.21)$$

is satisfied for any $\ell \in \mathbb{N}$, $r = 1, \dots, K$ and as $n \rightarrow \infty$, then we have, as $n \rightarrow \infty$,

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

Remark 5.6. This theorem generalizes the central limit theorem in Neininger and Rüschemdorf [70, Corollary 5.2], where the toll function b_n is assumed to be independent of the quantities $(Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}$. Surprisingly, we do not need any extra conditions as compensation for the weakened independence assumption.

Proof. Since we have $\text{Var}(Y_n) = g(n) + o(g(n))$ for some positive function g , we can find some constant $n_1 \in \mathbb{N}_0$ such that $\text{Var}(Y_n)$ is positive for $n \geq n_1$. As before, we define the standardized quantities by

$$X_n := \frac{Y_n - \mu(n)}{\sigma(n)}, \quad n \geq 0,$$

where $\mu(n) := \mathbb{E}[Y_n]$, $\sigma^2(n) := \text{Var}(Y_n)$ for $n \geq n_1$ and $\sigma(n) = 1$ for $n < n_1$. The statement of the theorem follows directly from the asymptotic expansions in (5.19) and Slutsky's theorem if we show that the normalized quantities X_n converge in distribution to the standard normal distribution. Thus, it is sufficient to show that the Zolotarev distance $\Delta(n) := \zeta_3(X_n, \mathcal{N}(0, 1))$ converges to zero as $n \rightarrow \infty$. Since we will proceed in the same way as in the proof of Theorem 5.1, we just point out the main arguments. Here, the sequence $(X_n)_{n \geq 0}$ satisfies the modified recursion

$$X_n \stackrel{d}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \quad (5.22)$$

with $I^{(n)}, (X_n^{(1)})_{n \geq 0}, \dots, (X_n^{(K)})_{n \geq 0}$ independent, $X_j^{(r)}$ identically distributed as X_j for $j \geq 0$ and $r = 1, \dots, K$ and

$$A_r^{(n)} = \frac{\sigma(I_r^{(n)})}{\sigma(n)}, \quad b^{(n)} = \frac{1}{\sigma(n)} \left(b_n - \mu(n) + \sum_{r=1}^K \mu(I_r^{(n)}) \right).$$

By conditions (5.19), (5.20) and (5.21), we have $A_r^{(n)} \rightarrow A_r^*$ and $b^{(n)} \rightarrow 0$ in L_3 for $r = 1, \dots, K$. Similarly as in the proof of Theorem 5.1, we now define

$$Z_n := \sum_{r=1}^K A_r^{(n)} \tau_{I_r^{(n)}} N^{(r)},$$

where $I^{(n)}, N^{(1)}, \dots, N^{(K)}$ are independent, the deterministic non-negative sequence $(\tau_n)_{n \geq 0}$ is defined by $\tau_n^2 = \text{Var}(X_n)$ for $n \geq 0$ and $N^{(r)}$ has the standard normal distribution for $r = 1, \dots, K$. Then, Z_n is centered and has variance

$$\text{Var}(Z_n) = \sum_{r=1}^K \mathbb{E}[(A_r^{(n)})^2 \tau_{I_r^{(n)}}^2] = \sum_{r=1}^K \left(\mathbb{E}[(A_r^{(n)})^2] + \mathbb{E}[\mathbf{1}_{\{I_r^{(n)} < n_1\}} (A_r^{(n)})^2 (\tau_{I_r^{(n)}}^2 - 1)] \right).$$

We now observe that for any $r = 1, \dots, K$, the latter summand in the above sum converges to zero by condition (5.21), since this condition and Jensen's inequality imply that $\mathbb{E}[\mathbf{1}_{\{I_r^{(n)} < n_1\}} (A_r^{(n)})^2] \rightarrow 0$. Together with the fact that $A_r^{(n)}$ converges in the L_3 norm (and thus also in L_2) to A_r^* with $(A_1^*)^2 + \dots + (A_K^*)^2 = 1$ almost surely, we obtain that $\text{Var}(Z_n)$ converges to 1. Hence, we can choose a deterministic sequence $(\kappa_n)_{n \geq 0}$ with $\kappa_n \rightarrow 0$ such that

$$Z_n^* := (1 + \kappa_n) Z_n$$

has variance 1 for $n \geq n_1$ (where n_1 may need to be enlarged). As a consequence, X_n, Z_n^* and $\mathcal{N}(0, 1)$ are ζ_3 -compatible for $n \geq n_1$ and we can apply the triangle inequality to obtain

$$\Delta(n) = \zeta_3(X_n, \mathcal{N}(0, 1)) \leq \zeta_3(X_n, Z_n^*) + \zeta_3(Z_n^*, \mathcal{N}(0, 1)), \quad n \geq n_1.$$

We use the same notation $Q_n := A_1^{(n)} X_{I_1^{(n)}}^{(1)} + \dots + A_K^{(n)} X_{I_K^{(n)}}^{(K)}$ as before and Lemma 5.3 to find

$$\begin{aligned} \zeta_3(X_n, Z_n^*) &\leq \zeta_3(Q_n, Z_n) + \|Q_n\|_3^2 \|b^{(n)}\|_3 + \frac{1}{2} \|Q_n\|_3 \|b^{(n)}\|_3^2 + \frac{1}{2} \|b^{(n)}\|_3^3 \\ &\quad + \left(|\kappa_n| + \frac{1}{2} |\kappa_n|^2 + \frac{1}{2} |\kappa_n|^3 \right) \|Z_n\|_3^3 \\ &= \zeta_3(Q_n, Z_n) + o(1), \end{aligned}$$

since $b^{(n)}$ converges to zero in the L_3 norm, κ_n converges to zero and $\|Z_n\|_3$ and $\|Q_n\|_3$ are bounded in n , the latter by Lemma 5.4, again (note that we have $\sum \mathbb{E}[(A_r^*)^3] < 1$ by the assumptions $\sum (A_r^*)^2 = 1$ almost surely and $\mathbb{P}(\exists r : A_r^* = 1) < 1$ and that the technical condition (5.8) is satisfied since we assumed $\mathbb{P}(I_r^{(n)} = n) \rightarrow 0$ for $r = 1, \dots, K$, thus Lemma 5.4 is applicable). Conditioning on $I^{(n)}$ implies that, for $n \geq n_1$,

$$\begin{aligned} \zeta_3(Q_n, Z_n) &\leq \left(\sum_{r=1}^K \mathbb{P}(I_r^{(n)} = n) \right) \Delta(n) + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} (A_r^{(n)})^3 \Delta(I_r^{(n)}) \right] \\ &\quad + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} < n_1\}} (A_r^{(n)})^3 \sup_{k < n_1} \zeta_3(X_k, \tau_k N^{(r)}) \right] \\ &= o(1) \Delta(n) + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} (A_r^{(n)})^3 \Delta(I_r^{(n)}) \right] + o(1), \end{aligned}$$

where we used the assumption $\mathbb{P}(I_r^{(n)} = n) \rightarrow 0$ for $r = 1, \dots, K$ and the technical condition (5.21) in the last step. Furthermore, we have $\zeta_3(Z_n^*, \mathcal{N}(0, 1)) \rightarrow 0$. This can be shown similarly as in the proof of Theorem 3.1 by showing that $\ell_3(Z_n^*, \mathcal{N}(0, 1)) \rightarrow 0$ and using Theorem 2.7 together with the fact that $\|Z_n^*\|_3$ is bounded in n . Collecting all estimates, we obtain that

$$\Delta(n) \leq o(1) \Delta(n) + \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} (A_r^{(n)})^3 \Delta(I_r^{(n)}) \right] + o(1). \quad (5.23)$$

From this, the statement follows by a standard argument (see, e.g., Neininger and Rüschendorf [70, pp. 390–391]). \square

5.3 Applications

As already mentioned, possible examples of distributional recurrences with dependent toll function start with the refined convergence results for Quicksort in Neininger [68] and for cyclic urns in Müller and Neininger [66], see also Fuchs, Müller and Sulzbach [32]. Since [68] only contains a convergence result for the Zolotarev metric without a rate of

convergence, we take up this example in Section 5.3.1 and use Theorem 5.1 to estimate the rate of convergence. Furthermore, in Section 5.3.2, we present another application of Theorems 5.1 and 5.5 concerning recent work of Diaconis and Kolesnik [18] on Fibonacci permutations.

5.3.1 Refined Quicksort asymptotics

In this section, we come back to the Quicksort algorithm already analyzed in Section 4.1.1. We set $K_0 = 0$ and denote by K_n , $n \geq 1$, the number of key comparisons needed by Quicksort to sort the list (U_1, \dots, U_n) , where $(U_i)_{i \geq 1}$ is a sequence of independent and uniformly on the unit interval distributed random variables. With the normalization

$$C_n := \frac{K_n - \mathbb{E}[K_n]}{n+1}, \quad n \geq 0,$$

Régnier [77] showed that the sequence $(C_n)_{n \geq 0}$ is a martingale converging almost surely (and in L_p) to some non-degenerate limit C which satisfies the distributional fixed-point equation (see Rösler [80])

$$C \stackrel{d}{=} UC^{(1)} + (1-U)C^{(2)} + \varphi(U)$$

with U , $C^{(1)}$, $C^{(2)}$ independent, U uniform on the unit interval, $C^{(1)}$ and $C^{(2)}$ having the same distribution as C and $\varphi(u) := 2u \log u + 2(1-u) \log(1-u) + 1$ for $u \in [0, 1]$.

The aim of this section is to further quantify the almost sure convergence $C_n \rightarrow C$ by analyzing the error term $C_n - C$. It is known that this error term satisfies (see Bindjeme and Fill [7, Theorem 1.4])

$$\|C_n - C\|_2^2 = 2 \frac{\log n}{n} + O\left(\frac{1}{n}\right) \quad (5.24)$$

and (see Neininger [68, Theorem 1.1])

$$\sqrt{\frac{n}{2 \log n}} (C_n - C) \xrightarrow{d} \mathcal{N}(0, 1)$$

as $n \rightarrow \infty$. We now show that the application of Theorem 5.1 provides a rate of convergence in the Zolotarev metric ζ_3 for the latter convergence without much effort. For this, we need some of the results deduced in [7] and [68]. First of all, we use the notation $Y_n := C_n - C$, $n \geq 0$, for the error terms. Note that we chose this notation, although differing from the notation used in [7] and [68], to guarantee that the notation is in accordance with the formulation of our theorems. We then observe that the error term Y_n can be decomposed recursively. Equation (12) in [68] states a sample-pointwise recurrence for the error term Y_n (see also equation (2.6) in [7]), from which we obtain that Y_n satisfies the distributional recursion

$$Y_n \stackrel{d}{=} \frac{I_n + 1}{n + 1} Y_{I_n}^{(1)} + \frac{n - I_n}{n + 1} Y_{n-1-I_n}^{(2)} + b_n, \quad n \geq 1, \quad (5.25)$$

with $I_n, (Y_n^{(1)})_{n \geq 0}, (Y_n^{(2)})_{n \geq 0}$ independent, I_n uniformly distributed on $\{0, \dots, n-1\}$, $Y_j^{(r)}$ distributed as Y_j for $j \geq 0$ and $r = 1, 2$ and some toll function b_n which is not independent of $(Y_n^{(1)})_{n \geq 0}$ and $(Y_n^{(2)})_{n \geq 0}$. Since the concrete representation of b_n is not needed in the following, we omit the details here and refer to [7] and [68]. Certainly, recurrence (5.25) is an instance of recursion (5.1) with $n_0 = 1$, $K = 2$, $I_1^{(n)} = I_n$, $I_2^{(n)} = n - 1 - I_n$ and $A_r(n) = (I_r^{(n)} + 1)/(n + 1)$ for $r = 1, 2$.

We then define the standardized error terms by

$$X_n := \frac{Y_n}{\sigma(n)}, \quad n \geq 0,$$

where $\sigma^2(n) := \text{Var}(Y_n) > 0$ for $n \geq 0$. Note that $\sigma^2(n) = 2 \log n/n + O(1/n)$ by (5.24) and the fact that both components C_n and C of Y_n are centered. The normalized quantities satisfy recursion (5.3) with the same parameters as above and

$$A_1^{(n)} = \frac{(I_n + 1) \sigma(I_n)}{(n + 1) \sigma(n)}, \quad A_2^{(n)} = \frac{(n - I_n) \sigma(n - 1 - I_n)}{(n + 1) \sigma(n)}, \quad b^{(n)} = \frac{b_n}{\sigma(n)}.$$

For these coefficients, we obtain the L_3 -convergences (see [68])

$$A_1^{(n)} \rightarrow \sqrt{U_1} =: A_1^*, \quad A_2^{(n)} \rightarrow \sqrt{1 - U_1} =: A_2^*, \quad b^{(n)} \rightarrow 0.$$

Thus, we are in the situation of Section 5.1 and now check the conditions of Theorem 5.1 with $R(n) = \log^{-1/2}(n)$. First of all, Lemma 2.2 in [68] states that, as $n \rightarrow \infty$,

$$\|b^{(n)}\|_3 = O\left(\frac{1}{\sqrt{\log n}}\right) = O(R(n)).$$

To estimate the order of $\|(A_1^{(n)})^2 + (A_2^{(n)})^2 - 1\|_{3/2}$, we use the arguments of Lemma 3.8 with $g(n) = 2n \log n$ and $h(n) = n$ to obtain

$$\begin{aligned} \left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2} &\leq \frac{1}{n \log n} \left\| I_n \log\left(\frac{I_n}{n}\right) + (n - 1 - I_n) \log\left(\frac{n - 1 - I_n}{n}\right) \right\|_{3/2} + O\left(\frac{1}{\log n}\right) \\ &= O\left(\frac{1}{\log n}\right). \end{aligned}$$

From this, we also obtain

$$\left| \sum_{r=1}^2 \mathbb{E}[(A_r^{(n)})^2] - 1 \right| \leq \left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_1 \leq \left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2} = O\left(\frac{1}{\log n}\right).$$

Thus, condition (5.6) is satisfied. Since $A_1^{(n)}$ and $A_2^{(n)}$ are uniformly bounded and I_n is uniform on $\{0, \dots, n-1\}$, the technical conditions (5.7) and (5.8) are clearly satisfied.

Furthermore, we can use the arguments of Lemma 3.8, the fact that I_n/n converges almost surely to U_1 and the dominated convergence theorem to show that

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^2 \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} (A_r^{(n)})^3 \right] = 2 \mathbb{E}[U_1^{3/2}] = \frac{4}{5} < 1,$$

such that all assumptions of Theorem 5.1 are satisfied and we obtain the following result.

Theorem 5.7. *For the number K_n of key comparisons used by Quicksort to sort the list (U_1, \dots, U_n) with $(U_i)_{i \geq 1}$ independent and uniformly distributed on the unit interval and the almost sure limit C of $C_n = (K_n - \mathbb{E}[K_n])/(n+1)$, we have, as $n \rightarrow \infty$,*

$$\zeta_3 \left(\frac{C_n - C}{\sqrt{\text{Var}(C_n - C)}}, \mathcal{N}(0, 1) \right) = O \left(\frac{1}{\sqrt{\log n}} \right).$$

5.3.2 Importance sampling for estimating the number of Fibonacci matchings

The basis of this application is the paper [18] by Diaconis and Kolesnik from which we adopt the notation and some of their results. The set \mathcal{F}_n of *Fibonacci matchings of size n* is defined by

$$\mathcal{F}_n = \{ \pi \in \mathcal{S}_n : |\pi(i) - i| \leq 1 \text{ for } 1 \leq i \leq n \},$$

where \mathcal{S}_n denotes the set of permutations of $\{1, \dots, n\}$ (note that Diaconis and Kolesnik use the notation $\mathcal{F}_{n,1}$ instead of \mathcal{F}_n). The set \mathcal{F}_4 of Fibonacci matchings of size $n = 4$ is displayed in Figure 5.1.

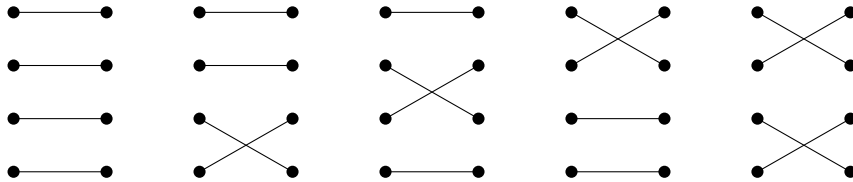


FIGURE 5.1: The 5 Fibonacci matchings of size $n = 4$.

One can observe that the cardinality of the set \mathcal{F}_n is easily computed by considering whether $\pi(1) = 1$ or $\pi(1) = 2$ and coincides with the $(n+1)$ -th Fibonacci number (which explains the name). Although the number of Fibonacci matchings is known, Diaconis and Kolesnik [18] present different importance sampling algorithms for estimating the size of \mathcal{F}_n . These algorithms in each step match the current index with an index chosen uniformly at random among the remaining allowable indices. To be more precise, Diaconis and Kolesnik present three such algorithms differing in the order the indices are matched (see Diaconis and Kolesnik [18] for details):

- ◇ The *random algorithm* A_r matches the indices in uniformly random order,
- ◇ the *fixed-order algorithm* A_f matches them in fixed order from top to bottom and

- ◇ the *greedy algorithm* A_g matches them in a certain greedy order. More precisely, A_g always matches the smallest unmatched index among those indices with the maximal number of remaining choices. This means that algorithm A_g always starts by matching index 2 uniformly at random with one of 1, 2, 3. If $\pi(2) \in \{1, 2\}$ (i.e., either $\pi(2) = 1$ and consequently $\pi(1) = 2$ or vice versa), then the next index to be matched is index 4 (uniformly among 3, 4, 5), since this is the smallest index with 3 remaining choices. Otherwise, i.e., if $\pi(2) = 3$, then the assignments $\pi(1) = 1$ and $\pi(3) = 2$ are forced and the next index to be matched is index 5 (uniformly among 4, 5, 6).

Having introduced the different algorithms, we now summarize some of the results given in [18]. For any Fibonacci matching $\pi \in \mathcal{F}_n$, we denote by $P_r(\pi)$, $P_f(\pi)$ and $P_g(\pi)$ the probability of π under the algorithm A_r , A_f and A_g , respectively. For Π_n chosen uniformly at random from the set \mathcal{F}_n of Fibonacci matchings, we have

$$\mathbb{E}[-\log(P_\iota(\Pi_n))] = \mu_\iota n + O(1), \quad \text{Var}(-\log(P_\iota(\Pi_n))) = \sigma_\iota^2 n + O(1), \quad (5.26)$$

for $\iota = r, f, g$, where $\mu_\iota \in (0.49, 0.51)$ and $\sigma_\iota^2 > 0$ can be computed exactly (we refer to [18] for the concrete values). Furthermore, for $\iota = r, f, g$ and as $n \rightarrow \infty$, we have the central limit theorem [18, Theorem 1.1]

$$\frac{-\log(P_\iota(\Pi_n)) - \mu_\iota n}{\sigma_\iota \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (5.27)$$

While the proof of this statement can be accomplished with Corollary 5.2 in Neininger and Rüschemdorf [70] for the former two algorithms (see [18, Theorems 3.2 and 3.4]), there is no obvious way of applying this corollary in the greedy case due to arising dependencies. Instead, Diaconis and Kolesnik use arguments from renewal theory to show (5.27) for the greedy algorithm A_g . However, using the theorems of this chapter, we can also handle the additional dependencies arising in the greedy case.

This is why we focus on algorithm A_g from now on and define the random variable

$$Y_n = -\log(P_g(\Pi_n)),$$

where, as before, Π_n is uniformly distributed on \mathcal{F}_n and $P_g(\pi)$ denotes the probability of π under the greedy algorithm A_g for any $\pi \in \mathcal{F}_n$. Recall that algorithm A_g always starts by matching index 2 uniformly at random with one of 1, 2, 3. Consequently, for a fixed Fibonacci permutation π , the probability that index 2 is matched correctly with $\pi(2)$ by A_g equals $1/3$. Depending on the value of $\pi(2)$, the resulting number of indices that are neither matched nor forced is $n - 2$ or $n - 3$ afterwards. Thus, we obtain that

$$Y_n \stackrel{d}{=} Y_{I_1^{(n)}} + \log 3,$$

where $I_1^{(n)}$ takes the values $n - 2$ and $n - 3$ with probabilities $2|\mathcal{F}_{n-2}|/|\mathcal{F}_n|$ and $|\mathcal{F}_{n-3}|/|\mathcal{F}_n|$, respectively, and is independent of $(Y_j)_{j \geq 0}$. However, using this recursion, our theorems

do not apply, since there is only one subproblem of almost the same size as the original problem (i.e., $A_1^* = 1$).

Instead, to obtain a recursion to which our framework applies, we now divide the permutation at the middle, more precisely at index $k_n = \lfloor n/2 \rfloor$, instead of dividing it at the top. To guarantee that the two resulting subproblems are still of the same nature and independent, we make use of the same trick as Diaconis and Kolesnik for the fixed-order case and consider a slightly modified algorithm A'_g : This modified algorithm A'_g proceeds in exactly the same way as algorithm A_g , with the difference that if there are exactly two unmatched indices left at the end (indices $n-1$ and n), then A'_g matches index n uniformly at random with one of $n-1, n, \varepsilon$, where ε can be interpreted as an error message. Thus, algorithm A'_g either returns a Fibonacci matching or an error message. However, denoting by $P'_g(\pi)$ the probability of π under the modified algorithm A'_g , the probabilities $P'_g(\pi)$ and $P_g(\pi)$ differ, if at all, by a factor 1.5 for any $\pi \in \mathcal{F}_n$. Hence, we obtain assertion (5.27) for the greedy case if we prove the corresponding central limit theorem for

$$Y_n = -\log(P'_g(\Pi_n)),$$

where Π_n is again uniform on \mathcal{F}_n . The advantage of this modified random variable is that we can now decompose Y_n in a way which is compatible with our theorems. For this, recall that $k_n = \lfloor n/2 \rfloor$ and consider whether $\pi(k_n) = k_n - 1$, $\pi(k_n) = k_n$ or $\pi(k_n) = k_n + 1$. In the first case, the resulting subproblem sizes are $k_n - 2$ and $n - k_n$, whereas they are $k_n - 1$ and $n - k_n$ in the second case and $k_n - 1$ and $n - k_n - 1$ in the third case. Hence, we obtain the recursive decomposition

$$Y_n \stackrel{d}{=} Y_{I_1^{(n)}}^{(1)} + Y_{I_2^{(n)}}^{(2)} + b_n,$$

where the vector $I^{(n)} = (I_1^{(n)}, I_2^{(n)})$ contains the subproblem sizes and has distribution

$$\mathbb{P}(I_1^{(n)} = i_1, I_2^{(n)} = i_2) = \frac{1}{|\mathcal{F}_n|} \begin{cases} |\mathcal{F}_{k_n-2}| \cdot |\mathcal{F}_{n-k_n}|, & i_1 = k_n - 2, i_2 = n - k_n, \\ |\mathcal{F}_{k_n-1}| \cdot |\mathcal{F}_{n-k_n}|, & i_1 = k_n - 1, i_2 = n - k_n, \\ |\mathcal{F}_{k_n-1}| \cdot |\mathcal{F}_{n-k_n-1}|, & i_1 = k_n - 1, i_2 = n - k_n - 1, \end{cases}$$

$I^{(n)}$, $(Y_j^{(1)})_{j \geq 0}$ and $(Y_j^{(2)})_{j \geq 0}$ are independent, $Y_j^{(r)}$ has the same distribution as Y_j for $j \geq 0$ and $r = 1, 2$ and b_n is a random variable taking the values 0 and $\log 3$. However, note that b_n is not independent of $(Y_j^{(2)})_{j \geq 0}$, which is the reason why Corollary 5.2 of [70] does not apply, but we can use Theorem 5.5 instead: The sequence $(Y_n)_{n \geq 0}$ satisfies recursion (5.18) as well as condition (5.19) with $f(n) = \mu_g n$ and $g(n) = \sigma_g^2 n$. Furthermore, conditions (5.20) and (5.21) are obviously satisfied with $A_1^* = A_2^* = 1/\sqrt{2}$. Thus, Theorem 5.5 implies

$$\frac{Y_n - \mu_g n}{\sigma_g \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

giving another proof of (5.27) for the greedy case. Note that we can also apply Theorems 3.4 and 5.1, respectively, if we are further interested in the convergence rate for the limit theorems (5.27) in the Zolotarev ζ_3 metric. This is done in the following theorem for

the random algorithm A_r , but corresponding results can also be derived for the other algorithms.

Theorem 5.8. *Let $P(\pi)$ be the probability of π under the random algorithm A_r for any Fibonacci permutation π . Further set $Y_n = -\log(P(\Pi_n))$, where Π_n is uniformly distributed on the set \mathcal{F}_n of Fibonacci matchings of length n . Then we have, for any $\varepsilon > 0$ and as $n \rightarrow \infty$,*

$$\zeta_3\left(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\text{Var}(Y_n)}}, \mathcal{N}(0, 1)\right) = O(n^{-1/2+\varepsilon}).$$

Proof. The quantity Y_n satisfies $Y_0 = Y_1 = 0$ and the distributional recursion (see [18, Proof of Theorem 3.2])

$$Y_n \stackrel{d}{=} Y_{I_1^{(n)}}^{(1)} + Y_{I_2^{(n)}}^{(2)} + b_n, \quad n \geq 2,$$

where the vector $I^{(n)} = (I_1^{(n)}, I_2^{(n)})$ contains the subproblem sizes, $I^{(n)}$, b_n , $(Y_j^{(1)})_{j \geq 0}$ and $(Y_j^{(2)})_{j \geq 0}$ are independent, $Y_j^{(r)}$ has the same distribution as Y_j for $j \geq 0$ and $r = 1, 2$ and b_n is a random variable taking the values $\log 2$ and $\log 3$. We now observe that $\text{Var}(Y_n)$ is positive for $n \geq 3$ and define the standardized quantities X_n as in (3.2). These normalized quantities satisfy the modified recursion (3.3) with coefficients $A_1^{(n)}$, $A_2^{(n)}$ and $b^{(n)}$ given in (3.4). From this and from the expansions of the mean and variance of Y_n in (5.26), it is easily seen that we are in the situation of Theorem 3.4 and that

$$\left\| \sum_{r=1}^2 (A_r^{(n)})^2 - 1 \right\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3 = O(n^{-3/2}).$$

Furthermore, for $R(n) = n^{-1/2+\varepsilon}$, the technical conditions are clearly satisfied and we have (with $\mathcal{L}(U) = \text{unif}[0, 1]$)

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^2 \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} (A_r^{(n)})^3 \right] = 2 \mathbb{E}[U^{1+\varepsilon}] < 1,$$

thus the assertion follows from Theorem 3.4. □

Concluding Remarks

In Chapter 3 of this thesis, we derived a general convergence theorem (Theorem 3.1) to bound the rates of convergence for recursive quantities in the Zolotarev distances. The main ingredients for this theorem are the recursive decomposition (3.1), the identification of a rate of convergence for the coefficients (3.7) and condition (3.8) which allows it to transfer the rate of the coefficients to a rate for the normalized quantities X_n . In the special case of normal limits, we were able to give a refined version of this result (Theorem 3.4), often improving over the general theorem by an additional factor 3 in the exponent of the rate.

In Chapter 4, we gave various examples of how to apply the general theorems. The applications presented in this chapter have shown that our theorems cover a wide range of recursive quantities and can be used to provide first, sometimes rough, bounds on the rate of convergence in the Zolotarev metrics in many concrete examples. For the number of maxima in right triangles, for instance, a straightforward application of Theorem 3.4 leads to the rate $O(n^{-1/4+\varepsilon})$ (with arbitrary $\varepsilon > 0$), see Section 4.1.2, where we also show that it is possible to obtain the rate $O(n^{-1/4})$ with a little more effort. The reason for this is that condition (3.8) (condition (3.16), respectively) is sometimes stronger than necessary, since it does not take into account that the convergence rates of the coefficients might be faster than those of the quantities X_n . An alternative (weaker) condition is formulated in Remark 3.3 (Remark 3.5, respectively) and is applied in several situations to improve the rate of convergence, in some instances avoiding the additional ε in the exponent of the rate.

In Chapter 5, we derived a convergence result similar to the refined version for the normal case (Theorem 3.4), with the difference that we assumed a relaxed conditional independence condition there allowing additional dependencies between the toll term and the subproblems (see Theorem 5.1). As compensation for the relaxed independence condition, we need slightly stronger assumptions on the convergence rates of the coefficients, missing the additional factor 3 in the exponent of $\|b^{(n)}\|_3$ compared to Theorem 3.4. It remains an open question whether this factor 3 might be “regained” by appropriately adjusting the proof of Theorem 5.1.

Finally, recall that the application of our general framework requires that the normalized quantities X_n given in (3.2) have mean vector $\mathbb{E}[X_n] = 0$ for $s > 1$ and additionally covariance matrix $\text{Cov}(X_n) = \text{Id}_d$ for $s > 2$ and n large enough. Hence, in these cases, the mean vector (and the covariance matrix, respectively) appear in the definition of X_n —and

thus also in the definition of the coefficients (3.4). As a result, to determine the order of the convergence rate of these coefficients, it is necessary that either the exact expressions of the mean and covariance matrix or at least asymptotic expansions with second order error terms are amenable (see Section 3.2). In the applications presented in this thesis and in numerous further examples, we have sufficiently precise control of these moments.

Bibliography

- [1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [2] Z.-D. Bai, L. Devroye, H.-K. Hwang, and T.-H. Tsai. Maxima in hypercubes. *Random Structures Algorithms*, 27(3):290–309, 2005.
- [3] Z.-D. Bai, H.-K. Hwang, W.-Q. Liang, and T.-H. Tsai. Limit theorems for the number of maxima in random samples from planar regions. *Electron. J. Probab.*, 6:no. 3, 41 pp. (electronic), 2001.
- [4] Z.-D. Bai, H.-K. Hwang, and T.-H. Tsai. Berry-Esseen bounds for the number of maxima in planar regions. *Electron. J. Probab.*, 8:no. 9, 26, 2003.
- [5] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. Oxford Science Publications.
- [6] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 1981.
- [7] P. Bindjeme and J. A. Fill. Exact L^2 -distance from the limit for QuickSort key comparisons (extended abstract). In *23rd Intern. Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA'12)*, Discrete Math. Theor. Comput. Sci. Proc., AQ, pages 339–348. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2012.
- [8] B. Chauvin, T. Klein, J.-F. Marckert, and A. Rouault. Martingales and profile of binary search trees. *Electron. J. Probab.*, 10:no. 12, 420–435, 2005.
- [9] L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probability*, 3(3):534–545, 1975.
- [10] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

- [11] H.-H. Chern, M. Fuchs, and H.-K. Hwang. Phase changes in random point quadtrees. *ACM Trans. Algorithms*, 3(2):Art. 12, 51, 2007.
- [12] H.-H. Chern and H.-K. Hwang. Phase changes in random m -ary search trees and generalized quicksort. *Random Structures Algorithms*, 19(3-4):316–358, 2001. Analysis of algorithms (Krynica Morska, 2000).
- [13] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to algorithms*. The MIT Electrical Engineering and Computer Science Series. MIT Press, Cambridge, MA; McGraw-Hill Book Co., New York, 1990.
- [14] B. Dadoun and R. Neininger. A statistical view on exchanges in Quickselect. In *ANALCO14—Meeting on Analytic Algorithmics and Combinatorics*, pages 40–51. SIAM, Philadelphia, PA, 2014.
- [15] G. Dall’Aglia. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (3)*, 10:35–74, 1956.
- [16] L. Devroye. Limit laws for local counters in random binary search trees. *Random Structures Algorithms*, 2(3):303–315, 1991.
- [17] L. Devroye. Applications of Stein’s method in the analysis of random binary search trees. In *Stein’s method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 247–297. Singapore Univ. Press, Singapore, 2005.
- [18] P. Diaconis and B. Kolesnik. Randomized sequential importance sampling for estimating the number of perfect matchings in bipartite graphs. *Adv. in Appl. Math.*, to appear, preprint available as arXiv:1907.02333.
- [19] M. Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.
- [20] M. Drmota, S. Janson, and R. Neininger. A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.*, 18(1):288–333, 2008.
- [21] R. M. Dudley. *Probabilities and metrics*. Matematisk Institut, Aarhus Universitet, Aarhus, 1976. Convergence of laws on metric spaces, with a view to statistical testing, Lecture Notes Series, No. 45.
- [22] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [23] W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

-
- [24] J. A. Fill and S. Janson. Quicksort asymptotics. *J. Algorithms*, 44(1):4–28, 2002. Analysis of algorithms.
- [25] R. Finkel and J. Bentley. Quad trees: A data structure for retrieval on composite keys. *Acta Inf.*, 4:1–9, 03 1974.
- [26] P. Flajolet, X. Gourdon, and C. Martínez. Patterns in random binary search trees. *Random Structures Algorithms*, 11(3):223–244, 1997.
- [27] P. Flajolet, G. Labelle, L. Laforest, and B. Salvy. Hypergeometrics and the cost structure of quadtrees. *Random Structures Algorithms*, 7(2):117–144, 1995.
- [28] P. Flajolet and R. Sedgewick. Digital search trees revisited. *SIAM J. Comput.*, 15(3):748–767, 1986.
- [29] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.
- [30] M. Fuchs, H.-K. Hwang, and R. Neininger. Profiles of random trees: limit theorems for random recursive trees and binary search trees. *Algorithmica*, 46(3-4):367–407, 2006.
- [31] M. Fuchs, H.-K. Hwang, and V. Zacharovas. An analytic approach to the asymptotic variance of trie statistics and related structures. *Theoret. Comput. Sci.*, 527:1–36, 2014.
- [32] M. Fuchs, N. S. Müller, and H. Sulzbach. Refined asymptotics for the number of leaves of random point quadtrees. In *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 110 of *LIPICs. Leibniz Int. Proc. Inform.*, pages Art. No. 23, 16. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018.
- [33] L. Goldstein. Non-asymptotic distributional bounds for the Dickman approximation of the running time of the Quickselect algorithm. *Electron. J. Probab.*, 23:Paper No. 100, 13, 2018.
- [34] R. Grübel and U. Rösler. Asymptotic distribution theory for Hoare’s selection algorithm. *Adv. in Appl. Probab.*, 28(1):252–269, 1996.
- [35] C. Heuberger. Hwang’s quasi-power-theorem in dimension two. *Quaest. Math.*, 30(4):507–512, 2007.
- [36] C. Heuberger and S. Kropf. Higher dimensional quasi-power theorem and Berry-Esseen inequality. *Monatsh. Math.*, 187(2):293–314, 2018.
- [37] N. J. Higham. *Functions of matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008. Theory and computation.

- [38] C. A. R. Hoare. Algorithm 65: Find. *Commun. ACM*, 4(7):321–322, July 1961.
- [39] C. A. R. Hoare. Quicksort. *Comput. J.*, 5:10–15, 1962.
- [40] C. Holmgren and S. Janson. Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electron. J. Probab.*, 20:no. 4, 51, 2015.
- [41] H.-K. Hwang. *Théoremes limites pour les structures combinatoires et les fonctions arithmétiques*. PhD thesis, Palaiseau, Ecole polytechnique, 1994.
- [42] H.-K. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European J. Combin.*, 19(3):329–343, 1998.
- [43] H.-K. Hwang. Second phase changes in random m -ary search trees and generalized quicksort: convergence rates. *Ann. Probab.*, 31(2):609–629, 2003.
- [44] H.-K. Hwang, M. Fuchs, and V. Zacharovas. Asymptotic variance of random symmetric digital search trees. *Discrete Math. Theor. Comput. Sci.*, 12(2):103–165, 2010.
- [45] H.-K. Hwang and T.-H. Tsai. Quickselect and the Dickman function. *Combin. Probab. Comput.*, 11(4):353–371, 2002.
- [46] P. Jacquet and M. Régnier. Normal limiting distribution of the size of tries. In *Performance '87 (Brussels, 1987)*, pages 209–223. North-Holland, Amsterdam, 1988.
- [47] P. Jacquet and W. Szpankowski. Asymptotic behavior of the Lempel–Ziv parsing scheme and digital search trees. *Theoret. Comput. Sci.*, 144(1-2):161–197, 1995. Special volume on mathematical analysis of algorithms.
- [48] P. Jacquet and W. Szpankowski. Analytical de-Poissonization and its applications. *Theoret. Comput. Sci.*, 201(1-2):1–62, 1998.
- [49] S. Janson and S. Kaijser. Higher moments of Banach space valued random variables. *Mem. Amer. Math. Soc.*, 238(1127):vii+110, 2015.
- [50] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
- [51] L. V. Kantorovič and G. v. Rubinštejn. On a space of completely additive functions. *Vestnik Leningrad. Univ.*, 13(7):52–59, 1958.
- [52] P. Kirschenhofer and H. Prodinger. Eine Anwendung der Theorie der Modulfunktionen in der Informatik. *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II*, 197(4-7):339–366, 1988.
- [53] P. Kirschenhofer, H. Prodinger, and W. Szpankowski. On the variance of the external path length in a symmetric digital trie. *Discrete Appl. Math.*, 25(1-2):129–143, 1989. *Combinatorics and complexity (Chicago, IL, 1987)*.

- [54] D. E. Knuth. Mathematical analysis of algorithms. In *Information processing 71 (Proc. IFIP Congress, Ljubljana, 1971), Vol. 1: Foundations and systems*, pages 19–27, 1972.
- [55] D. E. Knuth. *The art of computer programming. Vol. 1*. Addison-Wesley, Reading, MA, 1973. Fundamental algorithms, Second edition.
- [56] D. E. Knuth. *The art of computer programming. Vol. 3*. Addison-Wesley, Reading, MA, 1998. Sorting and searching, Second edition.
- [57] B. Kodaj and T. F. Móri. On the number of comparisons in Hoare’s algorithm “FIND”. *Studia Sci. Math. Hungar.*, 33(1-3):185–207, 1997.
- [58] W. Lew and H. M. Mahmoud. The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.*, 23(5):1050–1074, 1994.
- [59] H. M. Mahmoud. *Evolution of random search trees*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.
- [60] H. M. Mahmoud. Average-case analysis of moves in Quick Select. In *ANALCO09—Workshop on Analytic Algorithmics and Combinatorics*, pages 35–40. SIAM, Philadelphia, PA, 2009.
- [61] H. M. Mahmoud. Distributional analysis of swaps in Quick Select. *Theoret. Comput. Sci.*, 411(16-18):1763–1769, 2010.
- [62] H. M. Mahmoud, R. Modarres, and R. T. Smythe. Analysis of QUICKSELECT: an algorithm for order statistics. *RAIRO Inform. Théor. Appl.*, 29(4):255–276, 1995.
- [63] H. M. Mahmoud and B. Pittel. Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms*, 10(1):52–75, 1989.
- [64] H. M. Mahmoud and M. D. Ward. Asymptotic distribution of two-protected nodes in random binary search trees. *Appl. Math. Lett.*, 25(12):2218–2222, 2012.
- [65] P. Major. On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.*, 8(4):487–517, 1978.
- [66] N. Müller and R. Neininger. Refined asymptotics for the composition of cyclic urns. *Electron. J. Probab.*, 23:Paper no. 117, 20, 2018.
- [67] R. Neininger. On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms*, 19(3-4):498–524, 2001. Analysis of algorithms (Krynica Morska, 2000).
- [68] R. Neininger. Refined Quicksort asymptotics. *Random Structures Algorithms*, 46(2):346–361, 2015.

- [69] R. Neininger and L. Rüschendorf. Rates of convergence for Quicksort. *J. Algorithms*, 44(1):51–62, 2002. Analysis of algorithms.
- [70] R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14(1):378–418, 2004.
- [71] R. Neininger and L. Rüschendorf. On the contraction method with degenerate limit equation. *Ann. Probab.*, 32(3B):2838–2856, 2004.
- [72] R. Neininger and J. Straub. Convergence Rates in the Probabilistic Analysis of Algorithms. In M. Drmota and C. Heuberger, editors, *31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020)*, volume 159 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:13, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- [73] R. Neininger and H. Sulzbach. On a functional contraction method. *Ann. Probab.*, 43(4):1777–1822, 2015.
- [74] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York-Heidelberg, 1975. Translated from the Russian by A. A. Brown, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Band 82.
- [75] S. T. Rachev. *Probability metrics and the stability of stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1991.
- [76] S. T. Rachev and L. Rüschendorf. Probability metrics and recursive algorithms. *Adv. in Appl. Probab.*, 27(3):770–799, 1995.
- [77] M. Régnier. A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.*, 23(3):335–343, 1989.
- [78] E. Rio. Distances minimales et distances idéales. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(9):1127–1130, 1998.
- [79] E. Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):802–817, 2009.
- [80] U. Rösler. A limit theorem for “Quicksort”. *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.
- [81] U. Rösler. A fixed point theorem for distributions. *Stochastic Process. Appl.*, 42(2):195–214, 1992.
- [82] U. Rösler. On the analysis of stochastic divide and conquer algorithms. *Algorithmica*, 29(1-2):238–261, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).

- [83] U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).
- [84] N. Ross. Fundamentals of Stein’s method. *Probab. Surv.*, 8:210–293, 2011.
- [85] W. Schachinger. Asymptotic normality of recursive algorithms via martingale difference arrays. *Discrete Math. Theor. Comput. Sci.*, 4(2):363–397, 2001.
- [86] R. Sedgewick. Implementing quicksort programs. *Commun. ACM*, 21(10):847–857, Oct. 1978.
- [87] R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *Ann. Statist.*, 2:39–48, 1974.
- [88] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, Vol. II: Probability theory, pages 583–602, 1972.
- [89] G. W. Stewart. *Matrix algorithms. Vol. I*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1998. Basic decompositions.
- [90] W. Szpankowski. *Average case analysis of algorithms on sequences*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2001. With a foreword by Philippe Flajolet.
- [91] J. L. van Hemmen and T. Ando. An inequality for trace ideals. *Comm. Math. Phys.*, 76(2):143–148, 1980.
- [92] W. Vervaat. On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.*, 11(4):750–783, 1979.
- [93] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [94] S. Wild. *Dual-Pivot Quicksort and Beyond: Analysis of Multiway Partitioning and Its Practical Potential*. PhD thesis, Technische Universität Kaiserslautern, 2016.
- [95] S. Wild and M. E. Nebel. Average case analysis of Java 7’s dual pivot Quicksort. In *Algorithms—ESA 2012*, volume 7501 of *Lecture Notes in Comput. Sci.*, pages 825–836. Springer, Heidelberg, 2012.
- [96] V. M. Zolotarev. Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Teor. Veroyatnost. i Primenen.*, 21(4):741–758, 1976.

- [97] V. M. Zolotarev. Ideal metrics in the problem of approximating the distributions of sums of independent random variables. *Teor. Verojatnost. i Primenen.*, 22(3):449–465, 1977.

List of Figures

4.1	The set of maximal points in a right triangle	40
4.2	Recursive decomposition of the number of maxima	40
4.3	An m -ary search tree for $m = 4$	46
4.4	A 2-dimensional quadtrees of size $n = 4$	49
5.1	The 5 Fibonacci matchings of size $n = 4$	92

List of Tables

4.1	The values $\alpha = \alpha_m$ depending on m	47
4.2	The values $\alpha = \alpha_d$ depending on d	50

Deutsche Zusammenfassung

Die Kontraktionsmethode ist eine Technik, mit der Verteilungskonvergenz passend skaliert rekursiver Zufallsgrößen gezeigt werden kann. Seit ihrer Einführung vor 30 Jahren durch Uwe Rösler [80] zur Analyse der Anzahl an Schlüsselvergleichen des Sortieralgorithmus *Quicksort* wurde die Kontraktionsmethode sukzessive weiterentwickelt, sodass mittlerweile eine große Klasse von rekursiven Zufallsgrößen mit dieser Methode untersucht werden kann. Üblicherweise wird die Kontraktionsmethode angewandt, um Verteilungskonvergenz zu zeigen. Es ist jedoch auch möglich, Ideen der Kontraktionsmethode zu verwenden, um darüber hinaus *Konvergenzraten* in einer passenden Metrik zu bestimmen. Als konkretes Beispiel ist in diesem Zusammenhang die Anzahl der Schlüsselvergleiche von *Quicksort* zu nennen, für welche Fill und Janson [24] die Rate $O(n^{-1/2})$ in den Wasserstein- ℓ_p -Metriken ($p \geq 1$) und Neininger und Rüschemann [69] die Rate $\Theta(\log n/n)$ in der Zolotarev-Metrik ζ_3 zeigten.

Thema dieser Dissertation ist es, das Grundprinzip der Kontraktionsmethode zu verwenden, um universelle Konvergenztheoreme zum Bestimmen von Konvergenzraten passend normalisierter rekursiver Größen herzuleiten. Ausgangspunkt unserer Analyse ist die Annahme, dass die zu untersuchenden Größen Y_n eine gewisse Selbstähnlichkeit aufweisen. Genauer gesagt nehmen wir an, dass $(Y_n)_{n \geq 0}$ eine Folge zufälliger Vektoren in \mathbb{R}^d ist, welche folgendermaßen rekursiv zerlegt werden kann:

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

wobei $\stackrel{d}{=}$ Verteilungsgleichheit bezeichne, $n_0 \in \mathbb{N}$ eine beliebige natürliche Zahl sei, die Koeffizienten $A_1(n), \dots, A_K(n)$ zufällige $(d \times d)$ -Matrizen seien und b_n ein d -dimensionaler Zufallsvektor sei. Die dieser Verteilungsrekursion zugrundeliegende Vorstellung ist, dass das ursprüngliche Problem der Größe n in K (kleinere) Teilprobleme mit Größen $I_1^{(n)}, \dots, I_K^{(n)}$ zerlegt werden kann, wobei der Toll-Term b_n die „Kosten“ des Aufteilens und des anschließenden Zusammensetzens bezeichnet. Weitere grundlegende Annahmen unseres Modells sind, dass die Anzahl K der Teilprobleme eine feste natürliche Zahl und der Vektor $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$ der Teilproblemgrößen ein zufälliger Vektor in $\{0, \dots, n\}^K$ ist. Abgesehen davon stellt die schon angesprochene Selbstähnlichkeit der Größen Y_n eine zentrale

Voraussetzung unseres Modells dar. Formal bedeutet dies, dass

$$(Y_n^{(r)})_{n \geq 0} \stackrel{d}{=} (Y_n)_{n \geq 0} \quad \text{für } r = 1, \dots, K \quad (\text{A.2})$$

gilt. Schließlich fordern wir noch die folgende Unabhängigkeitsbedingung:

$$(A_1(n), \dots, A_K(n), b_n, I^{(n)}, (Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}) \quad \text{sind unabhängig.} \quad (\text{A.3})$$

Abhängigkeiten zwischen den Koeffizienten $A_1(n), \dots, A_K(n), b_n$ und den Teilproblemgrößen $I_1^{(n)}, \dots, I_K^{(n)}$ sind hingegen zulässig – und treten auch in vielen Anwendungen auf. Unter gewissen Voraussetzungen ist es außerdem möglich, die Unabhängigkeitsbedingung (A.3) leicht abzuschwächen. Wir betrachten deshalb neben (A.3) noch eine ähnliche Bedingung, und zwar:

$$(A_1(n), \dots, A_K(n), I^{(n)}, (Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}) \quad \text{sind unabhängig.} \quad (\text{A.4})$$

Der Unterschied zwischen (A.3) und (A.4) liegt darin, dass bei letzterer Abhängigkeiten zwischen dem Toll-Term b_n und den Größen $(Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}$ erlaubt sind.

Rekursionen der Form (A.1) treten in verschiedenen Gebieten auf, siehe zum Beispiel Rösler und Rüschemdorf [83] sowie Neining und Rüschemdorf [70] für zahlreiche konkrete Anwendungen. Beispiele reichen von Komplexitätsmaßen rekursiver Algorithmen (etwa die Anzahl an Schlüsselvergleichen von Quicksort, Mergesort oder Quickselect) über verschiedene Parameter zufälliger Bäume (etwa die Größe von Tries oder m -nären Suchbäumen) bis hin zu geometrischen Problemen (zum Beispiel die Anzahl an Maxima in rechtwinkligen Dreiecken). Für alle genannten Beispiele kann die Kontraktionsmethode verwendet werden, um Verteilungskonvergenz der passend skalierten Größen zu zeigen. In der vorliegenden Arbeit soll nun zusätzlich noch die Konvergenzgeschwindigkeit in einer passenden Metrik betrachtet werden.

Dafür definieren wir die normalisierten Größen $(X_n)_{n \geq 0}$ durch

$$X_n := C_n^{-1/2}(Y_n - M_n), \quad n \geq 0, \quad (\text{A.5})$$

wobei M_n ein d -dimensionaler Vektor und C_n eine positiv definite $(d \times d)$ -Matrix ist. Im Wesentlichen wählen wir M_n als den Erwartungswertvektor und C_n als die Kovarianzmatrix von Y_n , falls diese Momente existieren. Die rekursive Darstellung (A.1) der Größen Y_n lässt sich dann auf die normalisierten Größen X_n übertragen und wir erhalten

$$X_n \stackrel{d}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \quad (\text{A.6})$$

mit modifizierten Koeffizienten

$$A_r^{(n)} := C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2} \left(b_n - M_n + \sum_{r=1}^K A_r(n) M_{I_r^{(n)}} \right) \quad (\text{A.7})$$

und Selbstähnlichkeits- sowie Unabhängigkeitsvoraussetzungen wie oben.

Ausgehend von dieser rekursiven Darstellung der Größen X_n besteht die Kontraktionsmethode grundsätzlich aus folgenden Schritten:

(1) **Aufstellen einer stochastischen Fixpunktgleichung für den Grenzwert:**

Zunächst einmal prüft man, dass die in (A.7) gegebenen Koeffizienten in einem passenden Sinne konvergieren:

$$A_r^{(n)} \rightarrow A_r^*, \quad r = 1, \dots, K \quad \text{und} \quad b^{(n)} \rightarrow b^*, \quad n \rightarrow \infty.$$

Wenn nun mit $n \rightarrow \infty$ auch die Teilproblemgrößen $I_r^{(n)}$ groß werden, dann würden wir erwarten, dass ein potenzieller Grenzwert X von X_n folgende Fixpunktgleichung erfüllt (welche sich aus (A.6) ergibt, wenn man alle auftretenden Terme durch ihre Grenzwerte ersetzt):

$$X \stackrel{d}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*. \quad (\text{A.8})$$

Hierbei sind $(A_1^*, \dots, A_K^*, b^*), X^{(1)}, \dots, X^{(K)}$ unabhängig und $X^{(r)}$ hat die gleiche Verteilung wie X für $r = 1, \dots, K$.

(2) **Existenz und Eindeutigkeit der Lösung von (A.8):**

Hierfür betrachten wir die Lösungen von (A.8) als Fixpunkte der folgenden Abbildung, wobei \mathcal{P}^d den Raum aller Wahrscheinlichkeitsmaße auf \mathbb{R}^d bezeichne:

$$T : \mathcal{P}^d \rightarrow \mathcal{P}^d, \quad \mu \mapsto \mathcal{L}\left(\sum_{r=1}^K A_r^* Z^{(r)} + b^*\right), \quad (\text{A.9})$$

wobei $Z^{(1)}, \dots, Z^{(K)}, (A_1^*, \dots, A_K^*, b^*)$ unabhängig seien und $Z^{(1)}, \dots, Z^{(K)}$ Verteilung μ haben. Wir wählen dann einen passenden Teilraum von \mathcal{P}^d sowie eine vollständige Metrik auf diesem Teilraum, sodass die Abbildung T eine Kontraktion darstellt und wir mit dem Banach'schen Fixpunktsatz die Existenz und Eindeutigkeit der Lösung von (A.8) in dem gewählten Teilraum erhalten.

(3) **Konvergenz der normalisierten Größen gegen diesen Fixpunkt:**

Nachdem die Existenz einer eindeutigen Lösung X von Gleichung (A.8) in einem passenden Teilraum sichergestellt wurde, wird die Konvergenz der normalisierten Größen X_n gegen diese eindeutige Lösung X in der in (2) gewählten Metrik gezeigt. Daraus können dann – je nach Metrik – weitere Aussagen abgeleitet werden (zum Beispiel schwache Konvergenz und Konvergenz der Momente).

Ziel dieser Arbeit ist es, in Schritt (3) nicht nur Konvergenz in der gewählten Metrik zu zeigen, sondern auch eine Konvergenzrate zu bestimmen. Grob gesagt sind wir interessiert

an Aussagen folgender Art: Angenommen, die Konvergenzrate der Koeffizienten (gemessen mithilfe der üblichen L_s -Norm $\|\cdot\|_s$) ist von der Ordnung $O(R(n))$, d.h.

$$\sum_{r=1}^K \|A_r^{(n)} - A_r^*\|_s + \|b^{(n)} - b^*\|_s = O(R(n))$$

mit einer Nullfolge $(R(n))_{n \geq 0}$, können wir Bedingungen angeben, sodass wir die Rate $O(R(n))$ auch für die Größen X_n erhalten? Als Maß des Abstands zwischen der Verteilung von X_n und der Grenzverteilung $\mathcal{L}(X)$ sind prinzipiell verschiedene Metriken denkbar. Wir verwenden hier die sogenannten Zolotarev-Metriken ζ_s , welche sich im Zusammenhang mit der Kontraktionsmethode als sehr nützlich erwiesen haben. Insbesondere erlaubt die Verwendung der Zolotarev-Metriken es, sehr allgemeine Konvergenzresultate aufzustellen, welche viele Beispiele gleichzeitig abdecken.

Bevor wir zu den Konvergenzresultaten kommen, führen wir nun verschiedene Räume von Wahrscheinlichkeitsmaßen sowie die Zolotarev-Metriken kurz ein. Wie schon erwähnt bezeichnen wir mit \mathcal{P}^d den Raum aller Wahrscheinlichkeitsmaße auf \mathbb{R}^d und für $0 < s < \infty$ mit

$$\mathcal{P}_s^d := \{\mathcal{L}(X) \in \mathcal{P}^d : \|X\|_s < \infty\}$$

den Teilraum der Wahrscheinlichkeitsmaße mit endlichem s -ten Moment. Außerdem definieren wir

$$\mathcal{P}_s^d(0, \text{Id}_d) := \begin{cases} \mathcal{P}_s^d, & 0 < s \leq 1, \\ \{\mathcal{L}(X) \in \mathcal{P}_s^d : \mathbb{E}[X] = 0\}, & 1 < s \leq 2, \\ \{\mathcal{L}(X) \in \mathcal{P}_s^d : \mathbb{E}[X] = 0, \text{Cov}(X) = \text{Id}_d\}, & 2 < s \leq 3. \end{cases}$$

Die Zolotarev-Metrik ζ_s auf $\mathcal{P}_s^d(0, \text{Id}_d)$ wird dann folgendermaßen definiert ($0 < s \leq 3$): Zunächst schreiben wir $s = m + \alpha$ mit $m := \lceil s \rceil - 1 \in \{0, 1, 2\}$ und $\alpha := s - m \in (0, 1]$. Damit definieren wir eine Klasse \mathcal{F}_s von Test-Funktionen durch

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

wobei $C^m(\mathbb{R}^d, \mathbb{R})$ die Menge aller m -fach stetig differenzierbaren Funktionen $f : \mathbb{R}^d \rightarrow \mathbb{R}$ und $D^m f$ die m -te Ableitung einer solchen Funktion f bezeichne. Der *Zolotarev-Abstand* ζ_s zwischen zwei Zufallsvariablen X und Y mit Wahrscheinlichkeitsverteilungen $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{P}_s^d(0, \text{Id}_d)$ ist dann gegeben durch

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

Darauf aufbauend können wir nun die Hauptresultate formulieren. Unser erstes Resultat bezieht sich auf ein allgemeines Konvergenztheorem aus Neininger und Rüschendorf [70, Theorem 4.1], welches zahlreiche Anwendungen hat. Das folgende Resultat entspricht im Wesentlichen diesem Theorem, liefert aber zusätzlich noch eine Konvergenzrate in der Zolotarev-Metrik.

Theorem A.1. Sei $(X_n)_{n \geq 0}$ definiert wie in (A.5), wobei $(Y_n)_{n \geq 0}$ die Rekursion (A.1) erfülle, alle dort auftretenden Größen L_s -integrierbar seien (für ein $0 < s \leq 3$) und M_n und C_n so gewählt seien, dass $\mathcal{L}(X_n) \in \mathcal{P}_s^d(0, \text{Id}_d)$ für n groß genug gilt. Außerdem nehmen wir an, dass L_s -integrierbare Zufallsvariable A_1^*, \dots, A_K^*, b^* und eine monoton fallende Folge $R(n) > 0$ mit $R(n) \rightarrow 0$ existieren mit

$$\|b^{(n)} - b^*\|_s + \sum_{r=1}^K \|A_r^{(n)} - A_r^*\|_s + \sum_{r=1}^K \|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_s = O(R(n)) \quad (\text{A.10})$$

für alle $\ell \in \mathbb{N}$ und $n \rightarrow \infty$. Ist außerdem $\|\mathbf{1}_{\{I_r^{(n)} = n\}} A_r^{(n)}\|_s \rightarrow 0$ für $r = 1, \dots, K$ sowie

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^K \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^s \right] < 1, \quad (\text{A.11})$$

dann gilt für $n \rightarrow \infty$

$$\zeta_s(X_n, X) = O(R(n)),$$

wobei $\mathcal{L}(X)$ als die eindeutige Lösung der Gleichung (A.8) in $\mathcal{P}_s^d(0, \text{Id}_d)$ gegeben ist.

Neben dem Bestimmen einer Konvergenzrate für die Konvergenzen $A_r^{(n)} \rightarrow A_r^*$ sowie $b^{(n)} \rightarrow b^*$ (jeweils in der L_s -Norm) müssen wir also noch die folgenden zwei technischen Bedingungen sicherstellen:

$$\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_s = O(R(n)) \quad \text{und} \quad \|\mathbf{1}_{\{I_r^{(n)} = n\}} A_r^{(n)}\|_s \rightarrow 0$$

für alle $\ell \in \mathbb{N}$ und $r = 1, \dots, K$. Diese Bedingungen stellen sicher, dass die Teilproblemgrößen $I_r^{(n)}$ mit n wachsen und sind in Anwendungen meist erfüllt. Die Annahme (A.11) sichert uns einerseits die Existenz einer eindeutigen Lösung der Gleichung (A.8) in $\mathcal{P}_s^d(0, \text{Id}_d)$ und dient andererseits dazu, die Rate $O(R(n))$ der Koeffizienten auf die Konvergenz $X_n \rightarrow X$ zu übertragen.

In der vorliegenden Arbeit wird das soeben vorgestellte Resultat etwa zur Bestimmung von Konvergenzraten für verschiedene Komplexitätsmaße (Schlüsselvergleiche, Schlüsselaustausche) von Quicksort und Quickselect verwendet. Darüber hinaus deckt Theorem A.1 viele Anwendungen mit normalverteilten Grenzwerten ab – ein konkretes Beispiel in diesem Kontext ist etwa die Analyse der Größe von zufälligen m -nären Suchbäumen. Hier stellt sich jedoch heraus, dass die von Theorem A.1 gelieferte Konvergenzrate nicht optimal ist. Vor diesem Hintergrund geben wir neben dem allgemeinen Theorem A.1 noch eine weitere, verbesserte Version dieses Theorems an, welche eine große Klasse von Größen mit normalverteiltem Grenzwert X abdeckt und im Vergleich zum allgemeinen Theorem teilweise bessere Raten liefert.

Dafür nehmen wir wieder an, dass Y_n die Rekursion (A.1) erfüllt (wobei nun alle auftretenden Größen L_3 -integrierbar seien) und dass ein $n_1 \in \mathbb{N}_0$ existiert mit der Eigenschaft,

dass die Kovarianzmatrix von Y_n für $n \geq n_1$ positiv definit ist. Darauf aufbauend definieren wir die normalisierten Größen wie in (A.5), wobei M_n den Erwartungswertvektor und C_n für $n \geq n_1$ die Kovarianzmatrix von Y_n bezeichne. Unser zweites Theorem bezieht sich nun auf den Spezialfall, dass die in (A.7) gegebenen Koeffizienten $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ in L_3 gegen $(A_1^*, \dots, A_K^*, b^*)$ konvergieren, wobei

$$b^* = 0, \quad \sum_{r=1}^K A_r^* (A_r^*)^T = \text{Id}_d \quad (\text{A.12})$$

fast sicher gilt. Gilt zusätzlich $\mathbb{E}[\sum_{r=1}^K \|A_r^*\|_{\text{op}}^3] < 1$, dann hat Gleichung (A.8) eine eindeutige Lösung im Raum $\mathcal{P}_3^d(0, \text{Id}_d)$ – und zwar die d -dimensionale Standardnormalverteilung $\mathcal{N}(0, \text{Id}_d)$. Für diesen Spezialfall lässt sich folgende verfeinerte Version von Theorem A.1 angeben.

Theorem A.2. *Sei $(X_n)_{n \geq 0}$ definiert wie in (A.5), wobei $(Y_n)_{n \geq 0}$ die rekursive Darstellung (A.1) erfülle, alle dort auftretenden Größen L_3 -integrierbar seien und M_n und C_n den Erwartungswertvektor bzw. die Kovarianzmatrix von Y_n bezeichnen ($n \geq n_1$). Wir nehmen außerdem an, dass die Koeffizienten $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ in L_3 gegen $(A_1^*, \dots, A_K^*, b^*)$ mit (A.12) konvergieren und dass für ein monoton fallendes $R(n) > 0$ mit $R(n) \rightarrow 0$*

$$\left\| \sum_{r=1}^K A_r^{(n)} (A_r^{(n)})^T - \text{Id}_d \right\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3 + \sum_{r=1}^K \|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_3^3 = O(R(n)) \quad (\text{A.13})$$

für alle $\ell \in \mathbb{N}$ und $n \rightarrow \infty$ gilt. Ist außerdem $\|\mathbf{1}_{\{I_r^{(n)} = n\}} A_r^{(n)}\|_3 \rightarrow 0$ für $r = 1, \dots, K$ sowie

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^K \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^3 \right] < 1,$$

dann gilt für $n \rightarrow \infty$

$$\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d)) = O(R(n)).$$

Vergleichen wir die Bedingungen (A.10) und (A.13) der beiden Theoreme A.1 und A.2, so fällt auf, dass in (A.13) noch ein zusätzlicher Faktor 3 im Exponenten auftritt. Dies führt dazu, dass in den Normalverteilungsfällen mit der Zusatzbedingung (A.12) zwar prinzipiell beide Theoreme anwendbar sind, Theorem A.2 aber teilweise eine um den Faktor 3 im Exponenten bessere Rate liefert.

Der Hauptbestandteil des Beweises von Theorem A.2 ist folgendes Lemma.

Lemma A.3. *Sei $(Z_n^*)_{n \geq 0}$ eine Folge mit $\mathcal{L}(Z_n^*) \in \mathcal{P}_3^d(0, \text{Id}_d)$ und*

$$Z_n^* \stackrel{d}{=} G_n N + b^{(n)},$$

wobei G_n eine zufällige $(d \times d)$ -Matrix, $b^{(n)}$ ein zufälliger d -dimensionaler Vektor und N multivariat standardnormalverteilt sowie unabhängig von $(G_n, b^{(n)})$ sei. Gilt

$$\|G_n G_n^T - \text{Id}_d\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3 = O(R(n))$$

für $n \rightarrow \infty$, dann folgt

$$\zeta_3(Z_n^*, \mathcal{N}(0, \text{Id}_d)) = O(R(n)).$$

Schließlich befasst sich die Arbeit noch mit dem Fall, dass die Unabhängigkeitsbedingung (A.3) ersetzt wird durch die schwächere Forderung (A.4). Mit dieser abgeschwächten Unabhängigkeitsannahme erhalten wir folgendes Resultat.

Theorem A.4. Sei $(X_n)_{n \geq 0}$ definiert wie in (A.5), wobei M_n und C_n für $n \geq n_1$ den Erwartungswertvektor bzw. die Kovarianzmatrix von Y_n bezeichnen und $(Y_n)_{n \geq 0}$ die rekursive Darstellung (A.1) mit der abgeschwächten Unabhängigkeitsbedingung (A.4) erfüllt, wobei alle auftretenden Größen L_3 -integrierbar seien. Wir nehmen außerdem an, dass die Koeffizienten $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ in L_3 gegen $(A_1^*, \dots, A_K^*, b^*)$ mit (A.12) konvergieren, und dass für ein monoton fallendes $R(n) > 0$ mit $R(n) \rightarrow 0$

$$\left\| \sum_{r=1}^K \mathbb{E}[A_r^{(n)} (A_r^{(n)})^T] - \text{Id}_d \right\|_{\text{op}} + \left\| \sum_{r=1}^K A_r^{(n)} (A_r^{(n)})^T - \text{Id}_d \right\|_{3/2}^{3/2} + \|b^{(n)}\|_3 = O(R(n)) \quad (\text{A.14})$$

sowie

$$\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_2^2 + \|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\|_3^3 = O(R(n)), \quad n \rightarrow \infty, \quad (\text{A.15})$$

für alle $\ell \in \mathbb{N}$ und $r = 1, \dots, K$ gilt. Ist außerdem $\|\mathbf{1}_{\{I_r^{(n)} = n\}} A_r^{(n)}\|_3 \rightarrow 0$ für $r = 1, \dots, K$ sowie

$$\limsup_{n \rightarrow \infty} \sum_{r=1}^K \mathbb{E} \left[\frac{R(I_r^{(n)})}{R(n)} \|A_r^{(n)}\|_{\text{op}}^3 \right] < 1,$$

dann gilt für $n \rightarrow \infty$

$$\zeta_3(X_n, \mathcal{N}(0, \text{Id}_d)) = O(R(n)).$$

Dieses Theorem entspricht im Wesentlichen Theorem A.2, mit dem Unterschied, dass die zusätzlich erlaubten Abhängigkeiten zwischen Toll-Term und $(Y_n^{(1)})_{n \geq 0}, \dots, (Y_n^{(K)})_{n \geq 0}$ zu leicht verschärften Bedingungen (A.14) und (A.15) im Vergleich zu (A.13) führen.