Concentration of Multivariate Random Recursive Sequences arising in the Analysis of Algorithms

Dissertation
zur Erlangung des Doktorgrades der
Naturwissenschaften

vorgelegt beim Fachbereich Informatik und Mathematik
der Johann Wolfgang Goethe-Universität
Frankfurt am Main

von
Tämur Ali Khan
aus Frankfurt am Main

Frankfurt am Main 2006
(D30)

vom Fachbereich Informatik und Mathematik
der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr.-Ing. Detlef Krömker
Gutachter:	Prof. Dr. Ralph Neininger Prof. Dr. Gábor Lugosi

Datum der Disputation: 19.03.2007

Contents

Introduction

Analysis of algorithms concerns with the evaluation of the efficiency of algorithms. Therefore the complexity of an algorithm is defined as a parameter which reflects the quantities most important for the efficiency of the algorithm. Mostly the running time is such a quantity, but also demand of ressources can be one. The complexity does not only depend on the algorithm but it also depends on the input, since quantities like running time and demand of ressources do so. Hence, if one wants to compare the complexities of two or more algorithms, solving the same problems, it does not suffice to compare their complexities for only one or few inputs. On the other hand it is often impossible to compare them for all inputs, because there are too many. E.g. sorting algorithms are theoretically able to sort lists of arbitrary length. Roughly speaking, the complexity of most algorithms increases with the length of the inputs. Thus, the complexity of an algorithm is analyzed, depending on the input length. Often asymptotic results for increasing input length are given.

One method to do this is the average case analysis of algorithms, a field of science founded by D.E. Knuth in 1963 and constantly developed, ever since. "The Art of Computer Programming" by Knuth (1997a, 1997b, 1998) is a three volume encyclopedical edition on that field. For average case analysis one defines on the set of all inputs of the same length a probability distribution and studies the expected complexity (average case complexity), determined by this distribution. Often this is the uniform distribution, but also other distributions might be of interest, possibly motivated by applications. Since the 1980's the law of the complexity under such a probabilistic model is also studied more detailed, than only its expectation. Furthermore the random output under such a probabilistic model is sometimes analyzed.

Another important method which is used a lot in Computer Science is the worst case analysis of algorithms. Here, the maximal complexity is studied, where the maximum is taken over all inputs of the same length. The maximal complexity is also called worst case complexity and every input yielding worst case complexity is called worst case input. The advantage of worst case analysis is that if the worst case complexity of an algorithm is identified to be small, then the complexity is small for every input.

Now, there are algorithms which have a small average case complexity and a large worst case complexity. E.g. for sorting a list of length n, quicksort has a small average case complexity of order $\Theta(n \ln(n))$ and a large worst case complexity of order $\Theta(n^2)$. An important principle of both, Computer Science and computer engineering, which is commonly used in such a situation is randomization, in order to avoid large complexities with high probability. The calculation progress of randomized algorithms is at some points randomized. In particular there are randomized algorithms where the random calculation progress yields that the complexity is random for every input, but which always return a correct result. Like randomized quicksort, where randomization is achieved by chosing the pivots at random. Such randomized algorithms are called Las Vegas algorithms. Furthermore there are random algorithms which only yield with high probability the correct result or a nearly correct result. These are called Monte Carlo algorithms. We will not discuss them any further and only mean Las Vegas algorithms by randomized algorithms as from now.

Randomized algorithms became more recognized about 30 years ago. For further information on that field one may confer Motwani and Raghavan (1995). So, randomization is another aspect in Computer Science where stochastics are used. For randomized algorithms the (maximal) expected complexity is studied. Furthermore other quantities of the complexity are analyzed, as variance, convergence in distribution after appropriate rescaling, rates of convergence and tail bounds. Beside the expected complexity, upper bounds on the right tail are of special interest for Computer Science, since small upper bounds guarantee that complexities much larger than the expected complexity only occur with small probability. If so, then it is reasonable to use a certain randomized algorithm with good average case complexity, even if its worst case complexity is bad.

A good example is randomized quicksort which has for every list of length n expected complexity of order $\Theta(n \ln(n))$ and Rösler (1991) showed for every n that large deviations only occur with very small probability.

Stochastic concentration inequalities are an important tool to study tail bounds for such algorithmic problems. As a survey one may confer McDiarmid (1998) or the lecture notes of Lugosi (2006). There are several approaches to concentration inequalities.

One is Chernoff's bounding technique. The idea is to estimate for a centered random variable X its moment generating function $\mathbb{E}\exp(sX)$ from above in order to get an upper bound on $\mathbb{P}(|X| > t)$ by Markov's inequality. For sums of bounded, independent random variables Chernoff's bounding technique yields Hoeffding's inequality immediately from Hoeffding's Lemma (see Lemma 2.3.4 and Hoeffding (1963)).

Azuma's inequality (see Azuma (1976)) is a tail bound result on martingales with bounded differences, which is also proved via Chernoff's bounding technique. Azuma's inequality can be used to estimate $\mathbb{P}(|X| > t)$ by defining a Doob martingale on X by an appropriate filtration and estimating its martingale differences. This strategy is called martingale method or method of bounded differences. If $X = f(X_1, \dots, X_n)$, where $X_1, \dots, X_n$ are independent and f is a measurable function with bounded differences, then $\mathbb{P}(|X| > t)$ can be estimated by the so called independent bounded differences inequality of McDiarmid (1989), which is built upon Azuma's inequality.

Further approaches to concentration inequalities are Talagrand's induction method introduced by Talagrand (1995), and entropy method developed by Ledoux (1995/97,1996).

In this thesis various sequences of multivariate random variables are studied with respect to tail bounds. Each sequence has a recursive structure. In chapters 1 and 2 these sequences arise from problems given by Computer Science. In chapter 3 supercritical multitype Galton-Watson processes are studied.

The upper tail bounds for these random structures and the method used to achieve them are thread of the contents of the chapters. In each chapter normalized versions of the multivariate random variables, denoted $\mathbf{Y}_n$, $n \geq 1$, are estimated according to Chernoff's bounding technique. Here, the multivariate moment generating function $\mathbb{E}\exp\langle \mathbf{s}, \mathbf{Y}_n \rangle$ is estimated inductively on n, by exploiting the recursive structure. In the context of algorithms this approach was first used by Rösler (1991) for a univariate recursive structure. It turns out that the most difficult task is to prove the inductive step for $\mathbf{s}$ close to $(0, \dots, 0)$. Essentially, this is done by a manipulation on $\mathbf{b}_n$, which is an additive term appearing in the recursive equation for $\mathbf{Y}_n$ (see (1.4), (2.3) and (3.18)).: Since $\mathbb{E}\,\mathbf{b}_n = (0, \ldots, 0)$ it is $\mathbb{E}\,\exp\langle \mathbf{s}, \mathbf{b}_n\rangle = 1 + \mathrm{O}(\|\mathbf{s}\|^2)$, as $\|\mathbf{s}\| \to 0$. We get an explicit constant by writing $\mathbb{E}\,\exp\langle \mathbf{s}, \mathbf{b}_n\rangle$ as Taylor series. This manipulation was similarly used in the proof of Bennett's inequality (see Bennett (1962)).

In chapter 1 we study minimax trees. We do worst case analysis for Snir's randomized algorithm for evaluating Boolean decision trees. We show that there is always an input for which the random complexity stochastically dominates the complexities of all other inputs of same length. For these random worst case complexities we give exact expectations, asymptotic of the variance, a limit law with uniquely characterized limit, and tail bounds. The results on expectation and variance and the limit law are based on the theory Galton-Watson processes (see Athreya and Ney (1972) and on contraction method (see Rösler (1991, 1992), Rachev and Rüschendorf (1995), Rösler and Rüschendorf (2001), and Neininger and Rüschendorf (2004)). Furthermore we derive a limit law for the value of a minimax tree under Pearl's model and show that the limit distribution has a continuous distribution function and that it fulfills some fixed point equation.

In chapter 2 we analyze tail bounds for the Wiener index of random binary search trees. Binary search trees are a fundamental data structure of Computer Science for preprocessing lists. In particular there is a well known equivalence between binary search trees and quicksort. Beside the above mentioned analysis via Chernoff's bounding technique we study upper tail bounds by the method of bounded differences. Furthermore we give a lower bound on the tails.

The worst case complexity in chapter 1 can be identified as the generation size of a supercritical 2-type Galton-Watson process, an approach by Karp and Zhang (1995). In chapter 3 we generalize the method used for analyzing the tails in chapter 1, in order to get a tail bound result on the generation size of supercritical multitype Galton-Watson processes with finite maximum family size. Furthermore we yield an upper bound on that tail bound result, which has the advantage that it is explicitly given in terms of the offspring distribution and we yield a tail bound result for supercritical multitype Galton-Watson processes with immigration.

Acknowledgements

I would like to thank my advisor Ralph Neininger for all his support. He proved to have a good sense in the choice of open problems, which he recommended me to focus my investigations on. He always helped me immediately when I asked him for help and he always encouraged me with a positive outlook, when I was sceptical with the success of my studies.

Thanks are also due to Luc Devroye for his commitment in the analysis presented in section 1.3. Without his support we probably were not able to find the limit law.

Finally, I would like to thank my father Shakir Ali Khan for proofreading the text carefully, even though he is not concerned with mathematical issues at all.

The results of chapter 1 are published in Ali Khan and Neininger (2004) and Ali Khan, Devroye and Neininger (2005).

Chapter 1

Probabilistic Analysis for Minimax Trees and Minimax Tree Evaluation

1.1 Survey of Game Trees

In this chapter we study game trees, which are trees being related to the analysis of game-searching methods for two-person perfect information games like Chess or Go. In this section we give a survey on various models of and results on game trees and point out how our results relate to already given ones.

In two-person perfect information games two players A and B start with an initial position and take alternate turns, choosing each time among $d \geq 2$ possible moves. A terminal position is reached after $2k$, $k \geq 0$, moves. It does not necessarily terminate the game it terminates the horizon of a player or machine searching for best possible moves. One would like to assign a value to each position that indicates the chances of each player winning the game when starting from that position. Although, assuming best possible moves of both players, it is deterministic how the game terminates, the horizon $2k$ of players or machines may be limited so that they cannot plan their moves up to the very end of the game. To overcome this problem one assigns values V to terminal positions, where large values of V indicate that the position favors player A, small values favor player B. Given the values of all $n = d^{2k}$ terminal nodes one can search for best possible moves for the initial position and calculate its value.

The possible moves and its terminal positions can be represented in a rooted, complete, d-ary tree with height $2k$, $k \geq 0$. The root represents the initial position and given a node represents a certain position, each of its d children represents one of the d possible moves from that position. The leaves are assigned with the same values $V_1, \ldots, V_n$, $n = d^{2k}$, as the terminal positions they are representing. All other nodes are labeled with $\vee$ on even levels and with $\wedge$ on odd levels, cf. Figure 1.1 for the case $d = 2$ and $k = 2$.

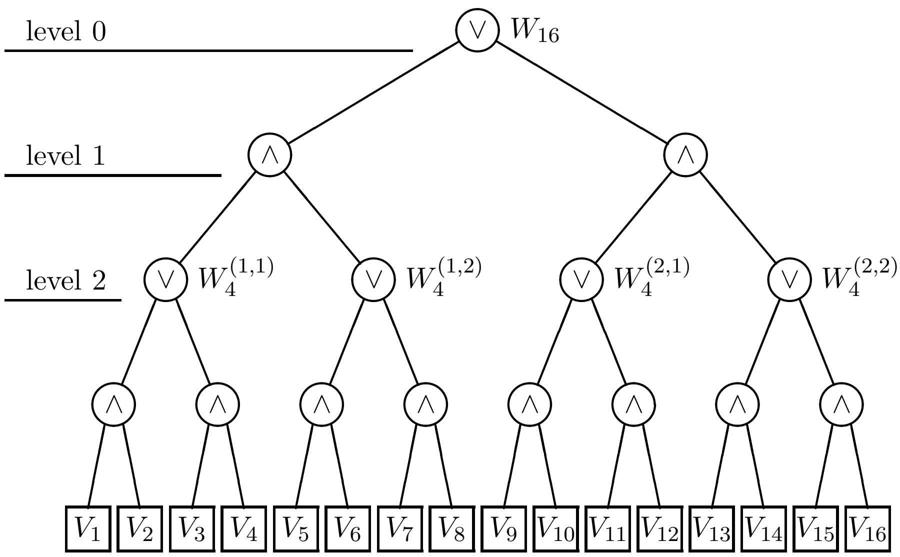

Figure 1.1: A minimax tree with branching degree 2 and height 4.

The value of a node is given as the value of the operator labeled at that node applied to the values of its children. This corresponds to player A always choosing the move with maximal value, player B always choosing a minimal value move. Thus from $V_1, \ldots, V_n$ one could first calculate the values of all nodes on level $2k - 1$ and successively determine the values on higher levels leading finally the root's value. These trees are called (d-ary) minimax trees. Sometimes in literature minimax trees are defined to have $\wedge$-labeled nodes on even levels and $\vee$-labeled nodes on odd levels. In this model a small value V indicates that the position favors player A and a large that it favors player B. Obviously both tree models are equivalent and can be transferred into each other, easily.

There are two important problems, concerning minimax trees. The first one is to calculate the root's value of a given minimax tree. This indicates for the start of the game the chances of each player to win. The root's value is also called value of the minimax tree. The second problem is to study the complexity of algorithms calculating the root's value. The complexity is defined as the number of leaves, an algorithm has to read, in order to calculate the root's value. Input of a minimax tree algorithm is the vector of leave values, $(V_1, \ldots, V_n)$, and output is the root's value. These two problems have been studied for various models of minimax trees. Next, we introduce some models and state given results concerning these two problems and finally relate our results to this context.

The first one are minimax trees where $V_1, \ldots, V_n$ only take values 0 and 1. These trees are also known as AND/OR trees or Boolean decision trees, since one may alternatively think of labels $\wedge$ and $\vee$ as boolean operators in this case. This is an important special case of minimax trees. The values of the leaves can be interpreted to indicate which player wins the game. Hence Boolean decision trees are minimax trees, where the terminal positions represent final positions of the game.

Snir (1985) implicitly proposed and analyzed the following randomized algorithm to evaluate a Boolean decision tree with branching degree $d = 2$: At each node one chooses randomly (with probability 1/2) one of its children and calculates its value recursively. If the result allows to identify the value of the node (that is a 0 for a $\wedge$-labeled node and a 1 for a $\vee$-labeled node) one is done, otherwise also the other child's value has to be calculated recursively in order to obtain the value of the node. Applying this to the root of the tree yields the value of the Boolean decision tree. For input $v \in \{0,1\}^n$ denote $C(v)$ the complexity of Snir's algorithm. Snir's Algorithm is a Las Vegas algorithm, i.e. that it always yields a correct output but complexity $C(v)$ is random. As pointed out in section 2.1 of Motwani and Raghavan (1995), Snir's analysis yields in particular

Theorem 1.1.1 (Snir (1985)) *We have*

$$\max_{v \in \{0,1\}^n} \mathbb{E}\, C(v) \leq n^{\log_3 4},$$

whereas for any deterministic version of Snir's algorithm there is an input, for which the algorithm has complexity n.

This documents that it is useful to randomize the algorithm, since linear worst case complexity is improved to sublinear worst case expected complexity. Snir's algorithm is naturally generalized to an algorithm for d-ary Boolean decision trees,

$d \geq 2$: For each node one chooses a random order of its children (with probability $1/d!$). The children are calculated recursively, one after another according to the chosen order, until the value of the node can be identified, the remaining children are discarded afterwards. This generalization is also called Snir's algorithm and $C(v)$ its complexity for a given input $v \in \{0,1\}^n$, $n = d^{2k}$.

Saks and Widgerson (1986) gave the exact order of the maximum expected complexity of Snir's algorithm and showed that it is optimal among all Las Vegas algorithms evaluating boolean decision trees:

Theorem 1.1.2 (Saks and Widgerson (1986)) *For fixed $d \geq 2$ denote* LAB *the set of all Las Vegas algorithms evaluating d-ary Boolean decision trees and* $\mathrm{com}(\mathcal{A}, v)$ *the complexity of an algorithm $\mathcal{A} \in$ LAB, given input $v \in \{0,1\}^n$, $n = d^{2k}$. Then*

$$\min_{\mathcal{A} \in \mathrm{LAB}} \max_{v \in \{0,1\}^n} \mathbb{E}\,\mathrm{com}(\mathcal{A}, v) = \max_{v \in \{0,1\}^n} \mathbb{E}\, C(v) = \Theta(n^{\alpha_d}),$$

for $\alpha_d = 1/2 \log_d((d^2 + 6d + 1 + (d-1)\sqrt{d^2 + 14d + 1})/8)$.

This result is essentially given in Theorem 5.4, Saks and Widgerson (1986).

Karp and Zhang (1995) showed for certain inputs, which may be denoted as regular inputs, that in particular every input $v' \in \{0,1\}^n$ with $\mathbb{E}\, C(v') = \max_v \mathbb{E}\, C(v)$ is a regular input and

Theorem 1.1.3 (Karp and Zhang (1995)) *For every $d \geq 2$, $k \geq 0$ and every regular input $v \in \{0,1\}^n$, $n = d^{2k}$, we have*

$$\mathbb{P}\left(\frac{C(v) - \mathbb{E}\, C(v)}{\mathbb{E}\, C(v)} > t \right) \leq \exp\left(-\ell_0 t^2\right),$$

for $t \geq 0$,

$$\mathbb{P}\left(C(v) < \frac{\mathbb{E}\, C(v)}{t} \right) \leq \ell_1 \exp\left(-\ell_2 t\right),$$

for $u_1 \leq t \leq u_2(\gamma/\sqrt{d})^{2k}$, where $\gamma > \sqrt{d}$ and

$$\ell_3\, \mathbb{E}\, C(v) \leq \sqrt{\mathrm{Var} C(v)} \leq \ell_4\, \mathbb{E}\, C(v),$$

for $\ell_4 > \ell_3$. $\ell_1, \ell_2, \ell_3, \ell_4, u_1, u_2$ explicitly known estimates depending on d are given.

Most important is that the first inequality yields that inputs with maximal expected complexity have a subgaussian right tail. In subsection 1.2.3 we will explain what are regular inputs.

The most frequently used algorithm for evaluating minimax trees with arbitrary nonnegative leaf values is $\alpha{-}\beta$ pruning (see Knuth and Moore (1975)). $\alpha{-}\beta$ pruning is a deterministic algorithm, which is just a deterministic version of Snir's algorithm, when applied on Boolean decision trees. Another algorithm for evaluating minimax trees with arbitrary nonnegative leaf values is called $\alpha - \beta$ pruning without deep cutoffs. $\alpha - \beta$ pruning without deep cutoffs is a simplification of $\alpha - \beta$ pruning. On the one hand it is easier to analyze than $\alpha - \beta$ pruning, on the other hand for every given input its complexity is at least as large as the complexity of $\alpha - \beta$ pruning. Thus $\alpha - \beta$ pruning without deep cutoffs has often been studied in order to get upper bound results for $\alpha - \beta$ pruning. $\alpha - \beta$ pruning without deep cutoffs works as follows for $d = 2$: Assume that the value W of a $\vee$-labeled node has to be calculated. Let W_ℓ and W_r be the values of its left and right child, respectively, and let $W_{r\ell}$ and W_{rr} be the values of the left and right child, respectively, of the node's right child, cf. Figure 1.2.

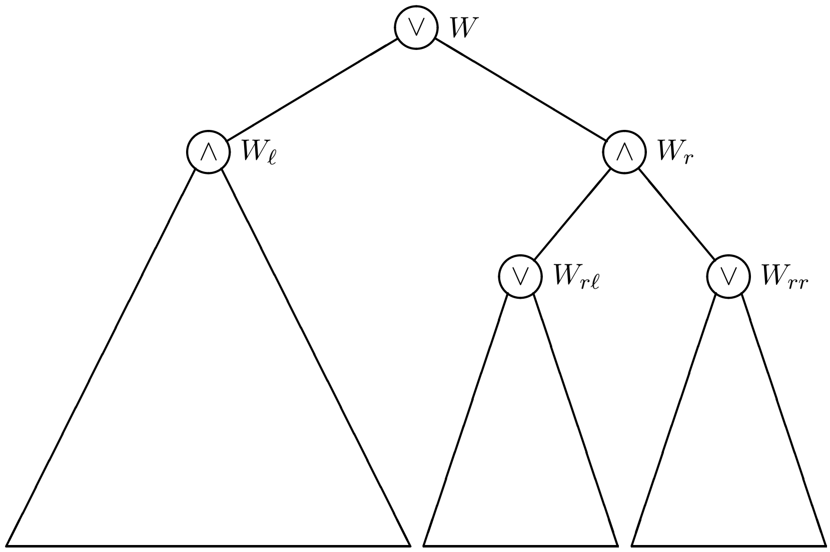

Figure 1.2: Subtree rooted at a $\vee$-node with value W.

In order to determine W, evaluate W_ℓ and $W_{r\ell}$ recursively. If $W_\ell \geq W_{r\ell}$, then $W = W_\ell$, since $W_\ell \geq W_{r\ell} \geq W_r$ and $W = W_\ell \vee W_r$, and one is done. Otherwise evaluate furthermore W_{rr} recursively and determine W by $W = W_\ell \vee (W_{r\ell} \wedge W_{rr})$. If W is the value of a $\wedge$-labeled node the procedure works equivalently. In that case one does not have to calculate W_{rr}, if $W_\ell \leq W_{r\ell}$. Applying this procedure to the root yields the value of the minimax tree. That some nodes do not have to be evaluated by the algorithm is visually phrased by saying, the minimax tree is "cutoff" at such a node. The phrase "without deep cutoffs" refers to the fact that a cutoff can only happen two levels below the node, currently evaluated. When evaluating a node by $\alpha - \beta$ pruning (with deep cutoffs), beside the above described cutoffs two levels below, furthermore cutoffs on deeper level might be done, by similar observations.

Another algorithm for minimax tree evaluation is SCOUT (see Pearl (1980)). SCOUT calculates the value of a $\vee$-valued ($\wedge$-valued) node by evaluating its left child recursively and next checking for each of its other $d-1$ children if it has a larger (smaller) value than the left child. This is done by assigning each leaf in the corresponding subtree a 1 if it has a larger value, a 0 else, and applying (deterministic) Snir's algorithm on the yielded subtree. Only the children which have larger (smaller) value than the left child are also evaluated recursively. In Pearl (1984) it is shown by an example that there is no dominating relation between the complexity of $\alpha - \beta$ pruning and SCOUT.

A traditional stochastic model for analyzing minimax tree algorithms is the i.i.d. model, in which the leaves' values $V_1, \ldots, V_n$ are independent and identically distributed random variables with a distribution $\mathcal{L}(V)$ having a distribution function $F_V(x) = \mathbb{P}(V \leq x)$ that is continuous. We denote by $C_1(d,k)$, $C_2(d,k)$ and $C_3(d,k)$ the complexity of $\alpha - \beta$ pruning, $\alpha - \beta$ pruning without deep cutoffs and SCOUT of a game tree with branching degree d and height $2k$ in the i.i.d. model. $C_1(d,k)$, $C_2(d,k)$ and $C_3(d,k)$ do not dependent upon $\mathcal{L}(V)$, since all three procedures depend only on the relative order of $V_1, \ldots, V_n$. On p. 314, Knuth and Moore (1975), it is given

Theorem 1.1.4 (Knuth and Moore (1975)) *For* $n = d^{2k}$,

$$p_{ij} = \binom{i-1+(j-1)/d}{i-1}^{-1},$$

r_d the largest Eigenvalue of matrix $[p_{ij}]_{1 \leq i,j \leq d}$, $a_d = \log_d(r_d)$ and c_d some positive constant, which can be specified, we have for fixed $d \geq 2$

$$\mathbb{E}\, C_2(d,k) \sim c_d n^{a_m}, \quad k \to \infty.$$

With this result they obtained furthermore

Theorem 1.1.5 (Knuth and Moore (1975)) *For $n = d^{2k}$, p_{ij} as above, r_d^* the largest Eigenvalue of matrix $[\sqrt{p_{ij}}]_{1\le i,j\le d}$, $a_d^* = \log_d(r_d^*)$ and c_d^* some positive constant, which can be specified, we have for fixed $d \ge 2$*

$$\mathbb{E}\, C_2(d,k) < c_d^* n^{a_m^*}.$$

Zhang (1984) analyzed complexities of the three mentioned minimax tree algorithms in the i.i.d. model and got results on variance and deviations and for SCOUT furthermore on expectation:

Theorem 1.1.6 (Zhang (1984)) *We have for $t > 0$ and all $k \ge 0$*

$$\mathbb{P}\left(\frac{C_2(d,k) - \mathbb{E}\, C_2(d,k)}{\mathbb{E}\, C_2(d,k)} \ge t\right) \le \exp(-\beta_d t^2),$$

where $\beta_d > 0$ is a constant depending on d and

$$\mathrm{Var} C_2(d,k) = \Theta\left(\mathbb{E}\, C_2(d,k)^2\right),$$

where the constant factors depend on d.

Theorem 1.1.7 (Zhang (1984)) *We have for $t > 0$ and all $k \ge 0$*

$$\mathbb{P}\left(\frac{C_1(d,k) - \mathbb{E}\, C_1(d,k)}{\mathbb{E}\, C_1(d,k)} \ge k^2 t\right) \le \exp(-\beta_d' t^2),$$

where $\beta_d' > 0$ is a constant depending on d and

$$\mathrm{Var} C_1(d,k) = O\left(k^2\, \mathbb{E}\, C_1(d,k)^2\right),$$

with the unspecified constant depending on d.

Theorem 1.1.8 (Zhang (1984)) *For $\varrho = (1-q)/q$, where q is the unique positive solution of $x = (1-x)^d$, we have*

$$\mathbb{E}\, C_3(d,k) = \Theta\left(\varrho^k\right),$$

for $t > 0$ and for all $k \ge 0$

$$\mathbb{P}\left(\frac{C_3(d,k) - \mathbb{E}\, C_3(d,k)}{\mathbb{E}\, C_3(d,k)} \ge t\right) \le \exp(-\beta_d'' t^2),$$

$\beta_d'' > 0$ is a constant depending on d and

$$\mathrm{Var} C_3(d,k) = O\left(\mathbb{E}\, C_3(d,k)^2\right),$$

with the unspecified constant depending on d.

Pearl (1980) analyzed the value of a minimax tree of height $2k$ in the i.i.d. modell, where the distribution of the leaves' values has a furthermore strictly increasing distribution function F_V on the range, where $0 < F_V < 1$. This special case of the i.i.d. model may be called Pearl's model. He showed:

Theorem 1.1.9 (Pearl (1980)) *Denote W_n, $n = d^{2k}$, the value of minimax tree with branching degree d in Pearl's model. Then*

$$W_n \to q_V, \quad k \to \infty,$$

in probability, with $q_V = F_V^{-1}(q)$ and q being the unique positive solution of $x = (1-x)^d$.

In games like chess, different moves which can be made from the same position are usually positively correlated. If a position favors a player, it is more likely that the following position favors the same player. But in the i.i.d. model values of siblings are independent. There are other models of random minimax trees, where random leaves' values are constructed in a way that siblings' values are positively correlated. See Knuth and Moore (1975) and Newborn (1977) for two such models. Another model with positively correlated sibling values is the incremental model: Every edge of the minimax tree is assigned with a random value. The edge values are independent and identically distributed as edge X. The value of a leaf is the sum of the values of all edges along the path from this leaf to the root. Nao (1982) developed this model for distribution $\mathcal{L}(X)$ determined by $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = -1) = p \in (0, 1)$. Denote $\widetilde{W}_n = \widetilde{W}_n(X)$, $n = d^k$, the value of a d-ary minimax tree height $k \geq 0$ in the incremental model and note that height k can also be odd, now. Devroye and Kamoun (1996) gave limiting results of $\widetilde{W}_n(X)$, for bounded X, bounded and nonegative X and Bernoulli distributed X:

Theorem 1.1.10 (Devroye and Kamoun (1996)) *In the incremental model with bounded edge variable X, we have for $n = d^k$*

$$\lim_{k\to\infty} \frac{\mathbb{E}\,\widetilde{W}_n}{k} = c < \infty,$$

where c is a positive constant, depending on $\mathcal{L}(X)$.

Theorem 1.1.11 (Devroye and Kamoun (1996)) *For the incremental model let edge variable X be Bernoulli(p) distributed. Then there is a $0 < p_d < 1$, such* *that for* $0 \le p \le p_d$ *we have*

$$\lim_{k\to\infty} \mathbb{P}\left(\widetilde{W}_{d^{2k}} = 0\right) > 0,$$

$$\lim_{k\to\infty} \mathbb{P}\left(\widetilde{W}_{d^{2k+1}} = 0\right) > 0$$

and for $p > p_d$

$$\lim_{k\to\infty} \mathbb{P}\left(\widetilde{W}_{d^{2k}} = 0\right) = \lim_{k\to\infty} \mathbb{P}\left(\widetilde{W}_{d^{2k+1}} = 0\right) = 0.$$

Furthermore

$$p_d \le 1 - d^{-1/(d+1)} \to 0, \quad d \to \infty.$$

They even obtained a law of large numbers:

Theorem 1.1.12 (Devroye and Kamoun (1996)) *For the incremental model let edge variable* X *be bounded and nonnegative with* $\mathbb{P}(X > 0) > p_d$*, where* p_d *is defined in Theorem 1.1.11. Then we have for* c *given in Theorem 1.1.10,* $n = d^k$*,*

$$\lim_{k\to\infty} \frac{\mathbb{E}\,\widetilde{W}_n}{k} = c$$

and

$$\lim_{k\to\infty} \frac{\widetilde{W}_n}{\mathbb{E}\,\widetilde{W}_n} = \lim_{k\to\infty} \frac{\widetilde{W}_n}{kc} = 1$$

almost surely as $k \to \infty$.

In the second section of this chapter we analyze Snir's algorithm. We show that for every height $2k$ their is a worst case input not only having maximal expected complexity but even more having maximal complexity in stochastic order among all inputs of Boolean decision trees of height $2k$. For this worst case complexity we derive exact expectation, asymptotic growth of the variance including the evaluation of the leading constant and a limit law with uniquely described limiting distribution, as $k \to \infty$. Our main finding is an improvement of Karp and Zhang's tail bound $\exp(-\text{const } t^2)$ for $t > 0$, which is stated in Theorem 1.1.3, to $\exp(-\text{const } t^\kappa)$, with $2 \le \kappa < 1/(1-\alpha_d) \ge 1/(1-\alpha_2) \doteq 4.06$ and α_d given in Theorem 1.1.2 (see Theorems 1.2.6 and 1.2.7). For ease of notation the analysis is done for Boolean decision trees with branching degree $d = 2$. It can be transferred to Boolean decision trees with arbitrary branching degrees, easily, and the results for that case are stated in subsection 1.2.6.

In the third section we derive a limit law for the value W_n of a minimax tree with branching degree d and height $2k$ in Pearl's model after appropriate rescaling. We show that the limiting distribution has a continuous distribution function and it fulfills some fixed point equation.

1.2 Probabilistic Analysis for Randomized Boolean Decision Tree Evaluation

We study Snir's algorithm on Boolean decision trees where all nodes on even level are labeled $\wedge$ and all nodes on odd level are labeled $\vee$, since in literature it is more common to define decision trees in this way (not so minimax trees in general). Furthermore, for ease of description the analysis in this section is done for binary decision trees. It can be transferred to d-ary decision trees, $d \geq 2$, easily, and the results therefore are stated in subsection 1.2.6.

In subsection 1.2.1 we will explain how to obtain input $v^\star \in \{0,1\}^n$ that $C(v^\star)$ is maximal in stochastic order, $C(v) \preceq C(v^\star)$ for all $v \in \{0,1\}^n$. Here, $X \preceq Y$ for random variables X, Y denotes that the corresponding distribution functions F_X, F_Y satisfy $F_X(x) \geq F_Y(x)$ for all $x \in \mathbb{R}$, or, equivalently, that there are realizations X', Y' of the distributions $\mathcal{L}(X), \mathcal{L}(Y)$ of X, Y on a joint probability space such that we pointwise have $X' \leq Y'$.

From this perspective it is reasonable to consider $C(v^\star)$ as the worst case complexity of the randomized algorithm and to analyze its asymptotic probabilistic behavior. Since $v^\star$ is a regular input in the sense of Karp and Zhang, also their 2-type Galton-Watson process approach applies.

The tail bound $\exp(-\text{const } t^\kappa)$, with $2 \leq \kappa < 1/(1-\alpha_d) > 4.06$ is based on a direct, inductive estimate of the moment generating function. In particular therefore we need the recursive description of subsection 1.2.4. Our approach is also applicable to any regular input as well as to other related problems.

1.2.1 Worst case input

In this subsection we explain how a worst case input $v^\star$ is constructed. We first have a look at the case $k = 1$ and $v \in \{0,1\}^4$ such that the decision tree is evaluated to 1 at the root. Clearly both children of the root have to lead to an evaluation of 1. Now each pair of external nodes attached to the children needs to have at least one value 1. Note that the algorithm reads in both pairs of external nodes until it finds the first one. Hence there will in total be read two 1's no matter how $v \in \{0,1\}^4$ is drawn among the choices that lead to an evaluation of 1 for the decision tree. Clearly, to maximize the number of 0's being read we choose in each pair of external nodes one 0 and one 1. Then both 0's are being read independently with probability 1/2. Hence, $v_1 = (0,1,0,1)$ stochastically maximizes $C(v)$ for all $v \in \{0,1\}^4$ such that the decision tree evaluates 1, see Figure 1.3.

Analogously look at the case $k = 1$ and $v \in \{0,1\}^4$ such that the decision tree is evaluated to 0. Clearly, one child of the root has to have the value 0, whose external nodes attached need to have both values 0. If we choose also value 0 for the other child of the root, we are lead to $v = (0,0,0,0)$, and the algorithm reads exactly 2 external nodes with values both 0. Therefore, to stochastically maximize $C(v)$ we choose the second child of the root with value 1 and again its external nodes attached with values 0 and 1. Then, $v_0 = (0,0,0,1)$ stochastically maximizes $C(v)$ for all $v \in \{0,1\}^4$ for which the decision tree evaluates to 0, see Figure 1.3.

Since we have $C(v_0) \preceq C(v_1)$, it follows that $v^\star = (0,1,0,1)$ is a choice with $C(v) \preceq C(v^\star)$ for all $v \in \{0,1\}^4$. For general $k \geq 2$ a corresponding $v^\star = v^\star(k)$ can recursively be constructed from $v^\star(k-1)$ as follows: Each component 0 in $v^\star(k-1)$ is replaced by the block $0,0,0,1$, whereas each 1 is replaced by the block $0,1,0,1$. For example, for $k = 3$, this yields

$$\begin{aligned} v^\star \;=\; & (0,0,0,1,0,0,0,1,0,0,0,1,0,1,0,1,\\ & 0,0,0,1,0,1,0,1,0,0,0,1,0,1,0,1,\\ & 0,0,0,1,0,0,0,1,0,0,0,1,0,1,0,1,\\ & 0,0,0,1,0,1,0,1,0,0,0,1,0,1,0,1). \end{aligned}$$

In Proposition 1.2.1 we show that this construction yields a $v^\star$ with $C(v) \preceq C(v^\star)$ for all $v \in \{0,1\}^n$ and $k \geq 1$.

If we would only want to stochastically maximize the cost over all $v \in R_0(n) \subset \{0,1\}^n$ that evaluate to a 0 at the root, the same recursive construction of replacing digits by corresponding blocks, starting with $v_0 = (0,0,0,1)$, yields a $v_\star \in R_0(n)$ such that $C(v) \preceq C(v_\star)$ for all $v \in R_0(n)$.

$v_\star(k)$ and $v^\star(k) \in \{0,1\}^n$, $k \geq 1$, are the regular inputs yielded by the patterns $(a_0, b_0, c_0, d_0) = (0,0,0,1)$ and $(a_1, b_1, c_1, d_1) = (0,1,0,1)$. Every input $v(k) \in \{0,1\}$, $k \geq 1$, which is constructed recursively by two patterns (a_0, b_0, c_0, d_0), $(a_1, b_1, c_1, d_1) \in \{0,1\}^4$ in the way described above, is a regular input, studied by Karp and Zhang (1995).

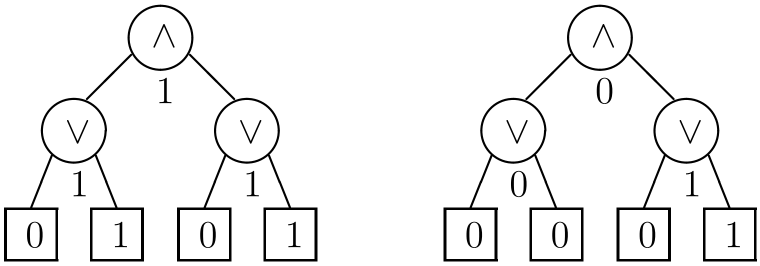

Figure 1.3: *Shown are decision trees for $k = 1$ evaluating at the root to 1 and 0, respectively, together with a choice for the external nodes that stochastically maximizes the number of external nodes read by the algorithm.*

1.2.2 Results

We assume that we have $n = 2^{2k}$ with $k \geq 1$ and denote by $v^\star \in \{0,1\}^n$ an input as constructed in section 2.

Proposition 1.2.1 *For $v^\star \in \{0,1\}^n$ as defined in section 2 we have $C(v) \preceq C(v^\star)$ for all $v \in \{0,1\}^n$.*

The stochastic worst case behavior $C(v^\star)$ of Snir's algorithm has the following asymptotic properties: The subsequent theorems describe the behavior of mean, variance, limit distribution, and large deviations of $C(v^\star)$. For the mean we have:

Theorem 1.2.2 *The expectation of $C(v^\star)$ is given by $\mathbb{E}\, C(v^\star) = c_1 n^\alpha - c_2 n^\beta$, with*

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

We denote for sequences $(a_k), (b_k)$ by $a_k \sim b_k$ asymptotic equivalence, i.e., $a_k/b_k \to 1$ as $k \to \infty$. Then we have for the variance of $C(v^\star)$:

Theorem 1.2.3 *The variance of $C(v^\star)$ satisfies asymptotically $\mathrm{Var}\, C(v^\star) \sim r\, n^{2\alpha}$ as $k \to \infty$, where $r \doteq 0.0938$. The constant r can also be given in closed form.*

For random variables X, Y we denote by $X \stackrel{d}{=} Y$ equality in distribution, i.e., $\mathcal{L}(X) = \mathcal{L}(Y)$. Then we have the following limit law for $C(v^\star)$:

Theorem 1.2.4 *For $C(v^\star)$ we have after normalization convergence in distribution,*

$$\frac{C(v^\star)}{n^\alpha} \longrightarrow C, \quad k \to \infty,$$

where the distribution of C *is given as* $\mathcal{L}(C) = \mathcal{L}(G^{[1]})$ *and* $\mathcal{L}(\mathbf{G}) = \mathcal{L}(G^{[0]}, G^{[1]})$ *is characterized by* $\mathbb{E}\,\|\mathbf{G}\|^2 < \infty$, $\mathbb{E}\,\mathbf{G} = (c_0, c_1)$, *with* $c_0 = 1/2 + 5/(2\sqrt{33})$, *and*

$$\mathbf{G} \stackrel{d}{=} \frac{1}{4^\alpha}\left\{\mathbf{G}^{(1)} + \mathbf{G}^{(2)} + \begin{bmatrix} B_1 B_2 & 0 \\ 1 - B_2 & 0 \end{bmatrix}\mathbf{G}^{(3)} + \begin{bmatrix} 0 & B_1 \\ B_1 & 0 \end{bmatrix}\mathbf{G}^{(4)}\right\},$$

with $\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(4)}$, B_1, B_2 *independent with* $\mathcal{L}(\mathbf{G}^{(r)}) = \mathcal{L}(\mathbf{G})$, *for* $r = 1, \ldots, 4$, *and* $\mathcal{L}(B_1) = \mathcal{L}(B_2) = B(1/2)$. *Here,* $B(1/2)$ *denotes the Bernoulli*$(1/2)$ *distribution.*

For the estimate of the tails we rely on Chernoff's bounding technique. We need to follow a bivariate setting for the vector $(C(v^\star), C(v_\star))$ as introduced in subsection 1.2.4. The following bound on the moment generating function is obtained:

Proposition 1.2.5 *It exists a sequence* $(\mathbf{Y}_k)_{k\geq 0} = (Y_k^{[0]}, Y_k^{[1]})_{k\geq 0}$ *of bivariate random variables with marginal distributions* $\mathcal{L}((C(v^\star) - \mathbb{E}\,C(v^\star))/n^\alpha)$, $\mathcal{L}((C(v_\star) - \mathbb{E}\,C(v_\star))/n^\alpha)$ *such that for all* $2 \geq q > 1/\alpha \doteq 1.33$ *there is a* $K > 0$ *with*

$$\mathbb{E}\,\exp\langle \mathbf{s}, \mathbf{Y}_k\rangle \leq \exp(K\|\mathbf{s}\|^q) \tag{1.1}$$

for all $\mathbf{s} \in \mathbb{R}^2$ *and* $k \geq 0$. *An explicit value for* $K = K_q$ *is given in (1.5).*

The bound on the moment generating function in the previous proposition implies upper tail bounds via Chernoff bounds:

Theorem 1.2.6 *For all* $2 \leq \kappa < 1/(1-\alpha) \doteq 4.06$ *there exists an* $L > 0$ *such that for any* $t > 0$ *and* $n = 2^{2k}$

$$\mathbb{P}\left(\frac{C(v^\star) - \mathbb{E}\,C(v^\star)}{n^\alpha} > t\right) \leq \exp(-Lt^\kappa). \tag{1.2}$$

An explicit value for L *is given in (1.6). The same bound applies to the left tail.*

Karp and Zhang (1995) used Azuma's inequality to get the first inequality in Theorem 1.1.3 Since $\mathbb{E}\,C(v^\star) = c_1 n^\alpha + o(n^\alpha)$, this inequality can be restated to

$$\mathbb{P}\left(\frac{C(v^\star) - \mathbb{E}\,C(v^\star)}{n^\alpha} > t\right) \leq \exp(-L' t^2)$$

for an explicitly known L'. For $\kappa = 2$ the prefactor $L = L_2$ in Theorem 1.2.6 can also be evaluated and satisfies $L_2 > 2L'$. It is yielded by Jones (2004) that one cannot improve the upper bound $1/(1-\alpha)$ upon exponent κ. This is pointed out in section 3.5.

1.2.3 Karp and Zhang's 2-type branching process

For the analysis of $C(v^\star)$ note that whenever the algorithm has to evaluate the value of a node at a certain depth that yields a 1, according to the discussion of subsection 1.2.1, the algorithm has to evaluate the values of two nodes of depths two levels below that each yield a 1, and $B_3 + B_4$ nodes of depths two levels below that each yield a 0, cf. Figure 1.3. Here, B_3, B_4 are independent Bernoulli $B(1/2)$ distributed random variables. Analogously, when the algorithm has to evaluate the value of a node at a certain depth that yields 0, two levels below it has to evaluate B_1 nodes yielding a 1 and $2+B_1B_2$ nodes yielding a 0, where B_1, B_2 are independent $B(1/2)$ distributed random variables. Here, the event $\{B_1 = 1\}$ corresponds to the algorithm first checking the right child of the node to be evaluated and $\{B_2 = 1\}$ to first checking the left child of that child, cf. Figure 1.3. Since at each node the child being evaluated first is independently drawn from all other choices, this gives rise to the following 2-type Galton-Watson branching process.

We have individuals of type 0 and 1 where the population of the k-th generation corresponds to the number of nodes at depth $2k$ that are read by the algorithm. We consider processes starting either with an individual of type 1 or type 0 and assume that the algorithm is applied to the worst case inputs $v^\star$ and $v_\star$, respectively. Then we have the following offspring distributions: An individual of type 1 has an offspring of 2 individuals of type 1 and $B_3 + B_4$ individuals of type 0. An individual of type 0 has an offspring of B_1 individuals of type 1 and $2 + B_1B_2$ individuals of type 0. We denote the number of individuals of type 0 and 1 in generation k by $(V_n^{[i]}, W_n^{[i]})$, when starting with an individual of type $i = 0, 1$, where $n = 2^{2k}$. Note that for $v^\star, v_\star \in \{0,1\}^n$ we have the representations

$$C(v^\star) \stackrel{d}{=} V_n^{[1]} + W_n^{[1]}, \quad C(v_\star) \stackrel{d}{=} V_n^{[0]} + W_n^{[0]}.$$

This is the approach of Karp and Zhang (1995) for regular inputs like $v^\star, v_\star$. Hence, part of the analysis of $C(v^\star)$ can be reduced to the application of the theory of multitype branching processes; see for general reference Harris (1963) and Athreya and Ney (1972), and for a survey on the application of branching processes to tree structures and tree algorithms see Devroye (1998). Obviously for every regular input v complexity $C(v)$ can be represented ba some 2-type Galton-Watson process.

However, we will also use a recursive description of the problem. This will be given in the next subsection and enables to use as well results from the probabilistic analysis of recursive algorithms by the contraction method.

1.2.4 The recursive point of view

It is convenient to work as well with a recursive description of the distributions $\mathcal{L}(C(v_\star))$ and $\mathcal{L}(C(v^\star))$. For this, we define the distributions of a bivariate random sequence $(\mathbf{G}_n) = (G_n^{[0]}, G_n^{[1]})$ for all $n = 2^{2k}$, $k \geq 0$ by $\mathbf{G}_1 = (1,1)$ and, for $k \geq 1$,

$$\mathbf{G}_n \stackrel{d}{=} \mathbf{G}_{n/4}^{(1)} + \mathbf{G}_{n/4}^{(2)} + \begin{bmatrix} B_1 B_2 & 0 \\ 1 - B_2 & 0 \end{bmatrix} \mathbf{G}_{n/4}^{(3)} + \begin{bmatrix} 0 & B_1 \\ B_1 & 0 \end{bmatrix} \mathbf{G}_{n/4}^{(4)},$$

where $\mathbf{G}_{n/4}^{(1)}, \ldots, \mathbf{G}_{n/4}^{(4)}$, B_1, B_2 are independent, B_1, B_2 are Bernoulli $B(1/2)$ distributed and $\mathcal{L}(\mathbf{G}_{n/4}^{(1)}) = \cdots = \mathcal{L}(\mathbf{G}_{n/4}^{(4)}) = \mathcal{L}(\mathbf{G}_{n/4})$. It can directly be checked by induction that the marginals of $\mathbf{G}_n$ satisfy $\mathcal{L}(G_n^{[0]}) = \mathcal{L}(C(v_\star))$ and $\mathcal{L}(G_n^{[1]}) = \mathcal{L}(C(v^\star))$. Note that $G_n^{[0]}$ and $G_n^{[1]}$ become dependent, firstly, since we have coupled the offspring distributions using for the second component again B_1 and $1 - B_2$ instead of B_3 and B_4, cf. subsection 1.2.3, and, secondly, since the first component of $\mathbf{G}_{n/4}^{(3)}$ contributes to both components of $\mathbf{G}_n$. Sequences satisfying recursive equations as $(\mathbf{G}_n)$ are being dealt with in a probabilistic framework, the so called contraction method; see Rösler (1991, 1992), Rachev and Rüschendorf (1995), Rösler and Rüschendorf (2001), and Neininger and Rüschendorf (2004).

1.2.5 Proofs

In this section we sketch the proofs of the results stated in subsection 1.2.2

Proof of Proposition 1.2.1: We denote by $R_0(n), R_1(n) \subset \{0,1\}^n$ the sets of vectors at the external nodes at depth $2k$ that yield an evaluation at the root of the decision tree of value 0 and 1, respectively. From the discussion in subsection 1.2.1 we have

$$C(v) \preceq C(v_\star), \quad v \in R_0(n), \quad \text{and} \quad C(v) \preceq C(v^\star), \quad v \in R_1(n).$$

Hence, it remains to show that $C(v_\star) \preceq C(v^\star)$. This is shown by induction on $k \geq 1$. For $k = 1$ this can directly be checked. For the step $k - 1 \to k$ assume that we have $C(v_\star(k-1)) \preceq C(v^\star(k-1))$. It suffices to find realizations of the quantities $(V_n^{[1]}, W_n^{[1]})$ and $(V_n^{[0]}, W_n^{[0]})$ on a joint probability space with $V_n^{[0]} + W_n^{[0]} \leq V_n^{[1]} + W_n^{[1]}$ almost surely, $n = 2^{2k}$.

For this we use B, B', $(V_{n/4}^{[i],(j)}, W_{n/4}^{[i],(j)})$ for $i = 1, 2$, $j = 1, \ldots, 4$ being independent for each $i = 0, 1$ and with B, B' Bernoulli $B(1/2)$ distributed, $\mathcal{L}(V_{n/4}^{[i],(j)}) =$

$\mathcal{L}(V_{n/4}^{[i]})$, $\mathcal{L}(W_{n/4}^{[i],(j)}) = \mathcal{L}(W_{n/4}^{[i]})$ for $i = 1, 2$ and $j = 1, \dots, 4$. By the induction hypothesis we may assume that we have versions of these random variates with $V_{n/4}^{[0],(j)} + W_{n/4}^{[0],(j)} \leq V_{n/4}^{[1],(j)} + W_{n/4}^{[1],(j)}$ for $j = 1, \dots, 4$. With this coupling we define $(V_n^{[1]}, W_n^{[1]})$ and $(V_n^{[0]}, W_n^{[0]})$ according to the values of B, B': On $\{B = 1, B' = 0\}$ we set

$$\begin{aligned}
\begin{pmatrix} V_n^{[0]} \\ W_n^{[0]} \end{pmatrix} &:= \begin{pmatrix} V_{n/4}^{[0],(2)} \\ W_{n/4}^{[0],(2)} \end{pmatrix} + \begin{pmatrix} V_{n/4}^{[0],(3)} \\ W_{n/4}^{[0],(3)} \end{pmatrix} + BB' \begin{pmatrix} V_{n/4}^{[0]),(4)} \\ W_{n/4}^{[0],(4)} \end{pmatrix} + B \begin{pmatrix} V_{n/4}^{[1],(1)} \\ W_{n/4}^{[1],(1)} \end{pmatrix}, \\
\begin{pmatrix} V_n^{[1]} \\ W_n^{[1]} \end{pmatrix} &:= B \begin{pmatrix} V_{n/4}^{[0],(3)} \\ W_{n/4}^{[0],(3)} \end{pmatrix} + \begin{pmatrix} V_{n/4}^{[1],(1)} \\ W_{n/4}^{[1],(1)} \end{pmatrix} + B' \begin{pmatrix} V_{n/4}^{[0],(4)} \\ W_{n/4}^{[0],(4)} \end{pmatrix} + \begin{pmatrix} V_{n/4}^{[1],(2)} \\ W_{n/4}^{[1],(2)} \end{pmatrix}
\end{aligned}$$

and obtain $V_n^{[0])} + W_n^{[0]} \leq V_n^{[1]} + W_n^{[1]}$. On the remaining sets $\{B = 0, B' = 0\}$, $\{B = 0, B' = 1\}$, and $\{B = 1, B' = 1\}$ similar couplings of $(V_n^{[0]}, W_n^{[0]})$, $(V_n^{[1]}, W_n^{[1]})$ can be defined with $V_n^{[0]} + W_n^{[0]} \leq V_n^{[1]} + W_n^{[1]}$. ■

Proof of Theorem 1.2.2: Assume that a generation has (w_0, w_1) individuals of type 0 and 1. Then, by the definition on the offspring distribution in section 4, the expected number of individuals in the subsequent generation is given by

$$M \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}, \quad M := \begin{bmatrix} 9/4 & 1 \\ 1/2 & 2 \end{bmatrix}.$$

Since $C(v^\star) = C(v^\star(k))$ is the sum of the individuals at generation k for the process started with an individual of type 1 we obtain

$$\mathbb{E}\, C(v^\star) = (1, 1) M^k \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The matrix M has the eigenvalues $\lambda_1 = (17 + \sqrt{33})/8$ and $\lambda_2 = (17 - \sqrt{33})/8$ and its k-th power can be evaluated to

$$M^k = \frac{1}{2\sqrt{33}} \begin{bmatrix} (\sqrt{33}+1)\lambda_1^k + (\sqrt{33}-1)\lambda_2^k & 8(\lambda_1^k - \lambda_2^k) \\ 4(\lambda_1^k - \lambda_2^k) & (\sqrt{33}-1)\lambda_1^k + (\sqrt{33}+1)\lambda_2^k \end{bmatrix}.$$

From this, $\mathbb{E}\, C(v^\star)$ and various constants needed subsequently can be read off. Note, that $\lambda_1^k = n^\alpha$ with α given in Theorem 1.2.2 and $n = 2^{2k}$. ■

Before proving Theorem 1.2.3 it is convenient to first prove Theorem 1.2.4.

Proof of Theorem 1.2.4: The 2-type branching process defined in section 4 is supercritical, nonsingular, and positive regular. Hence, a theorem of Harris (1963) implies that

$$\frac{1}{n^\alpha}\begin{pmatrix} V_n^{[1]} \\ W_n^{[1]} \end{pmatrix} \longrightarrow Y \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}$$

almost surely, as $k \to \infty$, where Y is a nonnegative random variable and (ν_1, ν_2) a deterministic vector that could also be further specified. Thus we obtain

$$\frac{C(v^\star)}{n^\alpha} \longrightarrow C$$

in distribution, as $k \to \infty$, with $\mathcal{L}(C) = \mathcal{L}((\nu_1 + \nu_2)Y)$.

On the other hand the recursive formulation of subsection 1.2.4 leads after the normalization $(X_n^{[0]}, X_n^{[1]}) = \mathbf{X}_n := \mathbf{G}_n / n^\alpha$ to

$$\mathbf{X}_n \stackrel{d}{=} \sum_{r=1}^{4} A^{(r)} \mathbf{X}_{n/4}^{(r)},$$

for $k \geq 1$, where $A^{(r)} = A^{(2)} = (1/4^\alpha) I_2$, with the 2×2 identity matrix I_2, and

$$A^{(3)} = \frac{1}{4^\alpha}\begin{bmatrix} B_1 B_2 & 0 \\ 1 - B_2 & 0 \end{bmatrix}, \quad A^{(4)} = \frac{1}{4^\alpha}\begin{bmatrix} 0 & B_1 \\ B_1 & 0 \end{bmatrix}, \tag{1.3}$$

where $\mathbf{X}_{n/4}^{(1)}, \ldots, \mathbf{X}_{n/4}^{(4)}$, B_1, B_2 are independent with $\mathcal{L}(\mathbf{X}_{n/4}^{(r)}) = \mathcal{L}(\mathbf{X}_{n/4})$ for $r = 1, \ldots, 4$ and $\mathcal{L}(B_1) = \mathcal{L}(B_2) = B(1/2)$. It follows from the contraction method that $\mathbf{X}_n$ converges weakly and with all mixed second moments to some G, that can be characterized as in Theorem 1.2.4. For details, how to apply the contraction method, see Theorem 4.1 in Neininger (2001). Thus, we have $C(v^\star)/n^\alpha \to G^{[1]}$ in distribution. ■

Proof of Theorem 1.2.3: As shown in the proof of Theorem 1.2.4 we have the convergence $\mathbf{X}_n = \mathbf{G}_n / n^\alpha \to \mathbf{G}$ for all mixed second moments. This, in particular, implies $\operatorname{Var} X_n^{[1]} \to \operatorname{Var} G^{[1]}$. The variances of $G^{[1]}$ can be obtained from the distributional identity for $\mathbf{G}$ stated in Theorem 1.2.4. Then we obtain $\operatorname{Var} C(v^\star) = \operatorname{Var}(n^\alpha X_n^{[1]}) \sim r n^{2\alpha}$ with $r = \operatorname{Var} G^{[1]}$. ■

Proof of Proposition 1.2.5: For $(Y_n^{[0]}, Y_n^{[1]}) = \mathbf{Y}_n = (1/n^\alpha)(\mathbf{G}_n - \mathbb{E}\,\mathbf{G}_n)$ we have marginals $\mathcal{L}(Y_n^{[1]}) = \mathcal{L}((C(v^\star) - \mathbb{E}\,C(v^\star))/n^\alpha)$ and $\mathcal{L}(Y_n^{[0]}) = \mathcal{L}((C(v_\star) -$

$\mathbb{E}\, C(v_\star))/n^\alpha)$. The distributional recurrence for $\mathbf{G}_n$ from subsection 1.2.4 implies the relation

$$\mathbf{Y}_n \stackrel{d}{=} \sum_{r=1}^{4} A^{(r)} \mathbf{Y}^{(r)}_{n/4} + \mathbf{b}_n, \quad k \geq 1, \tag{1.4}$$

with $\mathbf{Y}^{(1)}_{n/4}, \ldots, \mathbf{Y}^{(4)}_{n/4}, B_1, B_2$ independent, $\mathcal{L}(\mathbf{Y}^{(r)}_{n/4}) = \mathcal{L}(\mathbf{Y}_{n/4})$, for $r = 1, \ldots, 4$, $\mathcal{L}(B_1) = \mathcal{L}(B_2) = B(1/2)$ and

$$\mathbf{b}_n = \frac{4^\alpha}{n^\alpha}\, \mathbb{E}\, \mathbf{G}_{n/4} \sum_{r=1}^{4} A^{(r)} - \frac{4^\alpha}{n^\alpha}\, \mathbb{E}\, \mathbf{G}_n.$$

The matrices $A^{(r)}$ are given in (1.3).

We prove the assertion by induction on k. For $k = 0$ we have $\mathbf{Y}_1 = 0$, thus the assertion is true. Assume the assertion is true for some $n/4 = 2^{2(k-1)}$. Then, conditioning on $(A^{(1)}, \ldots, A^{(4)}, \mathbf{b}_n)$, denoting the distribution of this vector by σ_n, and using the induction hypothesis, we obtain

$$\begin{aligned}
\mathbb{E} \exp\langle \mathbf{s}, \mathbf{Y}_n \rangle &= \int \exp\langle \mathbf{s}, \beta_n \rangle \prod_{r=1}^{4} \mathbb{E} \exp\langle \mathbf{s}, a^{(r)} \mathbf{Y}_{n/4} \rangle d\sigma_n(a^{(1)}, \ldots, a^{(4)}, \beta_n) \\
&\leq \int \exp\langle \mathbf{s}, \beta_n \rangle \prod_{r=1}^{4} \exp(K \| a^{(r)T} s \|^q) d\sigma_n(a^{(1)}, \ldots, a^{(4)}, \beta_n) \\
&\leq \int \exp\left(\langle \mathbf{s}, \beta_n \rangle + K \|\mathbf{s}\|^q \sum_{r=1}^{4} \|a^{(r)}\|^q_{\mathrm{op}} \right) d\sigma_n(a^{(1)}, \ldots, a^{(4)}, \beta_n) \\
&= \mathbb{E} \exp(\langle \mathbf{s}, \mathbf{b}_n \rangle + K \|\mathbf{s}\|^q U) \exp(K \|\mathbf{s}\|^q),
\end{aligned}$$

with $U := \sum_{r=1}^{4} \left(\|A^{(r)}\|^q_{\mathrm{op}} \right) - 1 = 4^{-\alpha q}(2 + B_1 B_2 + (1 - B_2) + B_1) - 1$ and $\|A\|_{\mathrm{op}} = \sup_{\|x\|=1} \|Ax\|$ for matrices A. Hence, the proof is completed by showing

$$\sup_{k \geq 1} \mathbb{E} \exp(\langle \mathbf{s}, \mathbf{b}_n \rangle + K \|\mathbf{s}\|^q U) \leq 1,$$

for some appropriate $K > 0$. We denote $\xi := -\operatorname{esssup} U = 1 - 4^{1-\alpha q}$, thus $q > 1/\alpha$ implies $\xi > 0$.

Small $\|\mathbf{s}\|$: First we consider small $\|\mathbf{s}\|$ with $\|\mathbf{s}\| \leq c / \sup_{k \geq 1} \|\mathbf{b}_n\|_{2,\infty}$ for some $c > 0$, where $\|\mathbf{b}_n\|_{2,\infty} := \| \, \|\mathbf{b}_n\| \, \|_\infty$, the inner norm being the Euclidean norm. Note that throughout we have $n = n(k) = 2^{2k}$. For these small $\|\mathbf{s}\|$ we have

$$\mathbb{E} \exp((\langle \mathbf{s}, \mathbf{b}_n \rangle + K \|\mathbf{s}\|^q U) \leq \exp(-K \|\mathbf{s}\|^q \xi) \mathbb{E} \exp\langle \mathbf{s}, \mathbf{b}_n \rangle$$

and, with $\mathbb{E}\,\langle \mathbf{s}, \mathbf{b}_n\rangle = 0$,

$$\begin{aligned}
\mathbb{E}\,\exp\langle \mathbf{s}, \mathbf{b}_n\rangle &= \mathbb{E}\left[1 + \langle \mathbf{s}, \mathbf{b}_n\rangle + \sum_{k=2}^{\infty} \frac{\langle \mathbf{s}, \mathbf{b}_n\rangle^k}{k!}\right] \\
&= 1 + \mathbb{E}\,\langle \mathbf{s}, \mathbf{b}_n\rangle^2 \sum_{k=2}^{\infty} \frac{\langle \mathbf{s}, \mathbf{b}_n\rangle^{k-2}}{k!} \\
&\leq 1 + \|\mathbf{s}\|^2 \mathbb{E}\,\|\mathbf{b}_n\|^2 \sum_{k=2}^{\infty} \frac{c^{k-2}}{k!} \\
&= 1 + \|\mathbf{s}\|^2 \mathbb{E}\,\|\mathbf{b}_n\|^2 \frac{e^c - 1 - c}{c^2}.
\end{aligned}$$

Using $\exp(-K\|\mathbf{s}\|^q\xi) \leq 1/(1 + K\|\mathbf{s}\|^q\xi)$ and with $\Psi(c) = (e^c - 1 - c)/c^2$ we obtain

$$\mathbb{E}\,\exp(\langle \mathbf{s}, \mathbf{b}_n\rangle + K\|\mathbf{s}\|^q U) \leq \frac{1 + \|\mathbf{s}\|^2 \mathbb{E}\|\mathbf{b}_n\|^2 \Psi(c)}{1 + K\|\mathbf{s}\|^q \xi}.$$

Hence, we have to choose K with

$$K \geq \frac{\|\mathbf{s}\|^{2-q}\Psi(c)}{\xi} \sup_{k\geq 1} \mathbb{E}\,\|\mathbf{b}_n\|^2.$$

With $\|\mathbf{s}\| \leq c/\sup_{k\geq 1}\|\mathbf{b}_n\|_{2,\infty}$ and $q \geq 2$ a possible choice is

$$K = \frac{\sup_{k\geq 1}\mathbb{E}\,\|\mathbf{b}_n\|^2}{\sup_{k\geq 1}\|\mathbf{b}_n\|_{2,\infty}^{2-q}}\ \frac{\Psi_q(c)}{\xi},$$

with $\Psi_q(c) = (e^c - 1 - c)/c^q$.

Large $\|\mathbf{s}\|$: For general $\mathbf{s} \in \mathbb{R}^2$ we have

$$\langle \mathbf{s}, \mathbf{b}_n\rangle + K\|\mathbf{s}\|^q U \leq \|\mathbf{s}\|\|\mathbf{b}_n\| - \|\mathbf{s}\|^q K\xi \leq \|\mathbf{s}\|\|\mathbf{b}_n\|_{2,\infty} - \|\mathbf{s}\|^q K\xi,$$

and this is less than zero if

$$\|\mathbf{s}\|^{q-1} \geq \frac{\sup_{k\geq 1}\|\mathbf{b}_n\|_{2,\infty}}{K\xi} = \frac{\sup_{k\geq 1}\|\mathbf{b}_n\|_{2,\infty}^{3-q}}{\sup_{k\geq 1}\mathbb{E}\,\|\mathbf{b}_n\|^2\Psi_q(c)}.$$

If $\|\mathbf{s}\|$ satisfies the latter inequality we call it large. Thus, for large $\|\mathbf{s}\|$ we have $\sup_{k\geq 1}\mathbb{E}\,\exp(\langle \mathbf{s}, \mathbf{b}_n\rangle + K\|\mathbf{s}\|^q U) \leq 1$.

In order to overlap the regions for small and large $\|\mathbf{s}\|$ we need

$$\Psi_1(c) \geq \frac{\sup_{k\geq 1}\|\mathbf{b}_n\|_{2,\infty}^2}{\sup_{k\geq 1}\mathbb{E}\,\|\mathbf{b}_n\|^2}.$$

The right hand side of the latter display can be evaluated explicitly for our problem and equals 104/77. Thus, this inequality is true for, e.g., $c = 1.53$. Hence, with the explicit value

$$K := K_q = \frac{\sup_{k\geq 1} \mathbb{E}\,\|\mathbf{b}_n\|^2}{\sup_{k\geq 1} \|\mathbf{b}_n\|_{2,\infty}^{2-q}} \frac{e^{1.53} - 2.53}{1.53^q(1 - 4^{1-q\alpha})} \tag{1.5}$$

the proof is completed. ■

Proof of Theorem 1.2.6: By Chernoff's bounding technique we have, for $u > 0$ and with Proposition 1.2.5,

$$\begin{aligned} \mathbb{P}\left(\frac{C(v^\star) - \mathbb{E}\,C(v^\star)}{n^\alpha} > t\right) &= \mathbb{P}(\exp(uY_n^{[1]}) > \exp(ut))) \\ &\leq \mathbb{E}\,\exp(uY_n^{[1]} - ut) \\ &= \mathbb{E}\,\exp(\langle(0,u), \mathbf{Y}_n\rangle - ut) \\ &\leq \exp(K_q u^q - ut), \end{aligned}$$

for all q, K_q as in Proposition 3.5 and (1.5). Minimizing over $u > 0$ we obtain the bound

$$\mathbb{P}\left(\frac{C(v^\star) - \mathbb{E}\,C(v^\star)}{n^\alpha} > t\right) \leq \exp(-Lt^\kappa),$$

for $1 < \kappa < 1/(1-\alpha)$, with

$$L = L_\kappa = K_{\kappa/(\kappa-1)}^{1-\kappa} \frac{(\kappa-1)^{\kappa-1}}{\kappa^\kappa} \tag{1.6}$$

and $K_{\kappa/(\kappa-1)}$ given in (1.5). This completes the tail bound. ■

1.2.6 d-ary Boolean decision trees

The analysis can be carried over to Snir's algorithm for d-ary Boolean decision trees. A worst case input $v^\star \in \{0,1\}^n$ with $n = d^{2k}$ can be constructed similarly. Then we have similar results for $C(v^\star)$:

Theorem 1.2.7 *For the worst case complexity $C(v^\star)$ of evaluating an d-ary*

Boolean decision tree we have the following asymptotics:

$$\begin{aligned}
\mathbb{E}\, C(v^\star) &= c_1^{(d)} n^{\alpha_d} + c_2^{(d)} n^{\beta_d},\\
\mathrm{Var}\; C(v^\star) &\sim r_d n^{2\alpha_d},\\
\frac{C(v^\star)}{n^{\alpha_d}} &\to C_d,\\
\mathbb{P}\left(\frac{C(v^\star) - \mathbb{E}\, C(v^\star)}{n^{\alpha_d}} > t\right) &\leq \exp(-L^{(d)} t^\kappa), \quad t > 0,
\end{aligned}$$

with constants $c_1^{(d)}$, α_d, β_d, r_d, $L^{(d)} > 0$, $c_2^{(d)} \in \mathbb{R}$, *and* $2 \leq \kappa < \kappa_d = 1/(1-\alpha_d)$, α_d *given in Theorem 1.1.2.*

Numerical values for α_d, r_d and κ_d are listed in Table 1. The distribution of C_d is given as $\mathcal{L}(C_d) = \mathcal{L}(G^{[1]})$ and $\mathcal{L}(\mathbf{G}) = \mathcal{L}(G^{[0]}, G^{[1]})$ is characterized by $\mathbb{E}\, \|\mathbf{G}\|^2 < \infty$, $\mathbb{E}\, \mathbf{G} = (c_0^{(d)}, c_1^{(d)})$ and

$$\begin{aligned}
\mathbf{G} \stackrel{d}{=} \frac{1}{d^{2\alpha_d}} \Bigg\{ & \sum_{r=1}^{d} \mathbf{G}^{(r)} + \sum_{r=1}^{d-1} \begin{bmatrix} 0 & \mathbf{1}_r(U_0) \\ \mathbf{1}_r(U_0) & 0 \end{bmatrix} \bar{\mathbf{G}}^{(r)} \\
& + \sum_{r,\ell=1}^{d-1} \begin{bmatrix} \mathbf{1}_r(U_0)\mathbf{1}_\ell(U_r) & 0 \\ 1 - \mathbf{1}_\ell(U_r) & 0 \end{bmatrix} \mathbf{G}^{(r,\ell)} \Bigg\},
\end{aligned}$$

with $\mathcal{L}(\mathbf{G}^{(r)}) = \mathcal{L}(\bar{\mathbf{G}}^{(r)}) = \mathcal{L}(\mathbf{G}^{(r,\ell)}) = \mathcal{L}(G)$ and $\mathbf{G}^{(r)}$, $\bar{\mathbf{G}}^{(r)}$, $\mathbf{G}^{(r,\ell)}$, U_r independent with $\mathcal{L}(U_r) = \mathrm{unif}\{0, \ldots, d-1\}$ for all r, ℓ. Here, we denote $\mathbf{1}_i(Y) := \mathbf{1}_{\{i \leq Y\}}$ for integer i and a random variable Y, and we have

$$c_0^{(d)} = \frac{1}{2} + \frac{d+3}{2\sqrt{16d + (d-1)^2}}, \quad c_1^{(d)} = \frac{1}{2} + \frac{3d+1}{2\sqrt{16d + (d-1)^2}}.$$

---

d	2	3	4	5	6	7	8
α_d	0.754	0.759	0.765	0.769	0.774	0.778	0.781
r_d	0.0938	0.0847	0.0782	0.0731	0.0689	0.0652	0.0619
κ_d	4.060	4.154	4.247	4.336	4.419	4.497	4.571
d	9	10	11	12	13	14	15
α_d	0.785	0.788	0.790	0.793	0.795	0.798	0.800
r_d	0.0590	0.0564	0.0541	0.0519	0.0499	0.0481	0.0464
κ_d	4.641	4.707	4.769	4.829	4.886	4.940	4.993
d	16	17	20	30	40	50	100
α_d	0.802	0.804	0.809	0.821	0.830	0.837	0.856
r_d	0.0448	0.0433	0.0394	0.0304	0.0247	0.0209	0.0117
κ_d	5.043	5.091	5.226	5.596	5.885	6.123	6.928

Table 1: *Numerical values of the quantities α_d, r_d and κ_d appearing in Theorem 1.2.7 for various values of d.*

1.3 A Limit Law for the Root Value of Minimax Trees

In this section we study minimax trees with real valued leaves, where all nodes on even levels are labeled $\vee$ and all nodes on odd levels are labeled $\wedge$. We are not concerned with the complexity of algorithms to determine the root's value of such a tree, but with the root's value itself. We derive a limit law for W_n, the root's value under Pearl's model.

Recall that in Pearl's model the leaves' values $V_1, \dots, V_n$ are independent and identically distributed random variables with a distribution $\mathcal{L}(V)$ having a distribution function $F_V(x) = \mathbb{P}(V \leq x)$ that is continuous and strictly increasing on the range, where $0 < F_V < 1$.

We denote the distribution function of W_n by F_n. Note that this is defined for all $n = d^{2k}$ with $k \in \mathbb{N}_0$ and that we have $F_1 = F_V$. Moreover, for $k \geq 1$, we have $F_n = f \circ F_{n/d^2}$ with

$$f(x) = \Big(1 - (1-x)^d\Big)^d, \quad x \in [0,1]. \tag{1.7}$$

This is implied by the recursive structure of the tree: The values of the d^2 nodes on level 2 are independent and identically distributed with distribution $\mathcal{L}(W_{n/d^2})$. We denote these values by $W_{n/d^2}^{(i,j)}$ with $i, j = 1, \dots, d$, see Figure 1.1 for the case $d = 2$. Hence, by independence we have

$$\begin{aligned} F_n(x) &= \mathbb{P}\left(\bigvee_{i=1}^{d} \bigwedge_{j=1}^{d} W_{n/d^2}^{(i,j)} \le x \right) = \left(1 - \left(1 - \mathbb{P}\left(W_{n/d^2}^{(i,j)} \le x\right)\right)^d\right)^d \\ &= f(F_{n/d^2}(x)). \end{aligned}$$

Function f has the fixed points 0 and 1 and q defined in Theorem 1.1.9 as the unique positive solution of $x = (1-x)^d$ as the only fixed point in the open unit interval $(0,1)$. Recall that Pearl (1980) showed $W_n \to q_V$ in probability, as $k \to \infty$ for $q_V = F_V^{-1}(q)$, see Theorem 1.1.9. We denote the slope of f in q by $\xi = f'(q)$. Then the following limit law holds.

Theorem 1.3.1 *With F_V, q and ξ as above and $d \ge 2$ we have the following convergence in distribution for the value W_n of the minimax tree in Pearl's model. With $\alpha = \log(\xi)/\log(d^2) \in (0,1)$,*

$$n^\alpha (F_V(W_n) - q) \xrightarrow{\mathcal{L}} W, \quad k \to \infty. \tag{1.8}$$

The random variable W does not depend upon $\mathcal{L}(V)$, has a continuous distribution function F_W with $0 < F_W < 1$, $F_W(0) = q$ and

$$F_W(x) = f\left(F_W(x/\xi)\right), \quad x \in \mathbb{R}, \tag{1.9}$$

where f is the function defined in (1.7).

An approximation of the limit distribution function F_W is plotted in Figure 1.4 for the cases $d = 2, \dots, 10$.

Further analysis of F_W is done in the Diploma thesis of Meiners (2006). He showed in Theorem 3.2.1, Meiners (2006) that $F_W \in \mathcal{C}^\infty$ and that its power series in 0 converges on $\mathbb{C}$.

Note that the transformation $F_V(W_n)$ of W_n in (1.8) allows to rewrite $F_V(W_n)$ as follows: The random variable $F_V(W_n)$ is distributed as the root's value W'_n of a minimax tree with same branching degree and height where the independent, identically distributed leaves now have distribution $\mathcal{L}(V') = \mathcal{L}(F_V(V)) = \mathrm{unif}[0,1]$, the uniform distribution on $[0,1]$. Hence without loss of generality one may assume that $\mathcal{L}(V) = \mathrm{unif}[0,1]$.

In subsection 1.3.1 we collect some properties of f in section 1.3.1, since later on the recurrence relation $F_n = f \circ F_{n/d^2}$ is exploited. Subsection 1.3.2 contains the

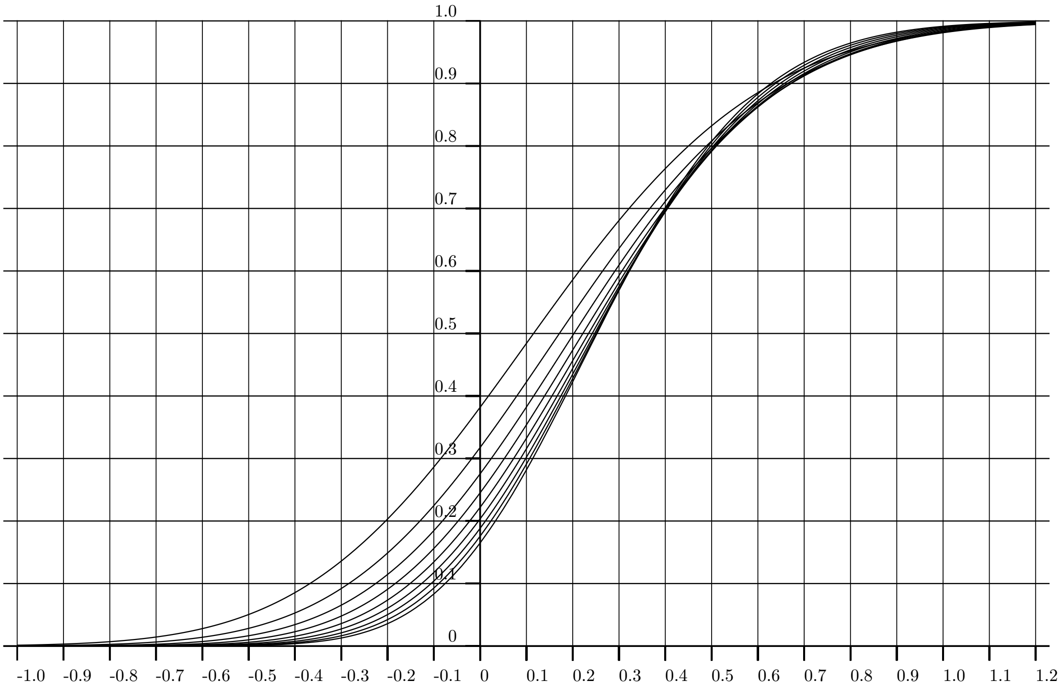

Figure 1.4: Approximations of the limit distribution function F_W for $d = 2, \ldots, 10$. They can be distinguished by $F_W(0) = q_d$ being decreasing in d. As approximations the functions g_6 defined in (1.13) are plotted.

proof of Theorem 1.3.1 and subsection 1.3.3 states a further result on the limit W, given by the Diploma thesis of Meiners (2006).

1.3.1 Technical preliminaries

We collect some properties of the function f defined in (1.7).

Lemma 1.3.2 *There is a unique $q \in (0,1)$ with $f(q) = q$. We have $\xi = f'(q) = d^2q^2/(1-q)^2 \in (1, d^2)$. Furthermore, for $z := 1 - 1/(d+1)^{1/d}$, we have*

$$f''(x) \begin{cases} > 0 & \text{for } 0 < x < z, \\ = 0 & \text{for } x = z, \\ < 0 & \text{for } z < x < 1. \end{cases} \tag{1.10}$$

We have $q < z$, thus $f''(q) > 0$.

Proof: For $0 < x < 1$ we have

$$f'(x) = d^2(1-x)^{d-1}(1-(1-x)^d)^{d-1}, \tag{1.11}$$

$$f''(x) = d^2(d-1)(1-x)^{d-2}(1-(1-x)^d)^{d-2}((d+1)(1-x)^d - 1). \tag{1.12}$$

So, the (in-)equalities in (1.10) follow with $z = z_d = 1 - 1/(d+1)^{1/d}$. For existence and uniqueness of the fixed point q of f in (0,1) we first show:

Claim: $f(z_d) - z_d > 0$ for all $d \geq 2$.

The claim follows for $d = 2, 3$ by explicit calculation. Furthermore we have $f(z_d) = (1 - 1/(d+1))^d \downarrow 1/e$ as $d \to \infty$, hence $f(z_d) \geq 1/e$ for all $d \geq 4$. It is easily seen that z_d is decreasing in d, thus $z_d \leq z_4$ for all $d \geq 4$. Consequently, for all $d \geq 4$

$$f(z_d) - z_d \geq \frac{1}{e} - z_4 = \frac{1}{e} + 1 - \frac{1}{5^{1/4}} > 0,$$

which implies the claim.

Since $f(0) = f'(0) = 0$, there exists $0 < \varepsilon < z_d$ with $f(x) - x < 0$ for all $0 < x \leq \varepsilon$. Together with the previous claim, continuity and the intermediate value theorem we obtain a fixed point of f in (ε, z_d). We denote by $q = q_d$ the smallest fixed point of f in $(0, z_d)$, which exists by continuity and satisfies $q > \varepsilon > 0$. Then we have $f(x) < x$ for all $x \in (0, q)$. For $x \in (q, z)$ we have $f(x) > x$ by convexity of f on $[0, z]$: Otherwise there was an $x \in (q, z)$ with $f(x) \leq x$. For arbitrary $y \in (0, q)$, and $\lambda \in (0,1)$ with $q = \lambda y + (1-\lambda)x$ this implied $f(q) \leq \lambda f(y) + (1-\lambda)f(x) < \lambda y + (1-\lambda)x = q$, a contradiction. Similarly, concavity of f on $[z, 1]$ implies $f(x) > x$

for all $x \in (z, 1)$: For all such x there is a $\lambda \in (0, 1)$ with $x = \lambda z + (1 - \lambda)1$ thus $f(x) \geq \lambda f(z) + (1 - \lambda) f(1) > \lambda z + (1 - \lambda)1 = x$. Altogether, q is the unique fixed point of f in $(0, 1)$.

It remains to prove that $\xi = \xi_d = f'(q) = d^2 q^2/(1 - q)^2 \in (1, d^2)$. For this note that the function $u_d : [0, 1] \to [0, 1]$, $x \mapsto (1 - x)^d$, has a unique fixed point in $(0, 1)$. Since $f = u_d \circ u_d$ this fixed point must be $q = q_d$, hence we obtain the relation $q = (1 - q)^d$. Using this relation in (1.11) implies $\xi = f'(q) = d^2 q^2/(1 - q)^2$. Moreover, since $u_{d'} \leq u_d$ for all $2 \leq d \leq d'$ the sequence $(q_d)_{d \geq 2}$ is decreasing. Thus $q_d \leq q_2 = (3 - \sqrt{5})/2 < 1/2$ for all $d \geq 2$, hence $\xi_d < d^2$. Finally, $q = (1 - q)^d$, $f''(q) > 0$ and the representation (1.12) imply $q > 1/(d + 1)$, hence $q/(1 - q) > 1/d$ and $\xi > d^2/d^2 = 1$. ■

In the following, it is convenient to extend function f defined in (1.7) to the real line by setting $f(x) = 0$ for $x < 0$ and $f(x) = 1$ for $x > 1$. We denote the iterations of f by $f_k = f \circ f_{k-1}$ for $k \geq 1$ and $f_0(x) = x$ for all $x \in \mathbb{R}$. In particular, we have $f_1 = f$. Using $F_n = f \circ F_{n/d^2}$ we obtain for $n = d^{2k}$ that $F_n = f_k \circ F_1 = f_k \circ F_V$.

For the quantities $n^\alpha(F_V(W_n) - q)$ of Theorem 1.3.1 we obtain with the relation $n^\alpha = \xi^k$

$$\begin{aligned}
\mathbb{P}(n^\alpha(F_V(W_n) - q) \leq x) &= \mathbb{P}\left(W_n \leq F_V^{-1}\left(q + \frac{x}{\xi^k}\right)\right) \\
&= F_n \circ F_V^{-1}\left(q + \frac{x}{\xi^k}\right) \\
&= f_k\left(q + \frac{x}{\xi^k}\right).
\end{aligned}$$

Thus, the functions $g_k : \mathbb{R} \to \mathbb{R}$ defined by

$$g_k(x) = f_k\left(q + \frac{x}{\xi^k}\right), \quad x \in \mathbb{R}, \tag{1.13}$$

are the distribution functions of $n^\alpha(F_V(W_n) - q)$ for $n = d^{2k}$, $k \geq 0$.

Subsequently we will need bounds for g_k valid locally around $x = 0$ and uniformly in $k \geq 0$.

Lemma 1.3.3 *Denote $h_1(x) := q + x$ and $h_2(x) := q + x + cx^2$ for $x \in \mathbb{R}$ with $c := 1 + f''(q)/(2\xi(\xi - 1)) > 1$. Then it exists an $\varepsilon > 0$ such that for all $k \geq 0$ and $|x| < \varepsilon$*

$$h_1(x) \leq g_k(x) \leq h_2(x).$$

Proof: We prove the assertion by induction on k. For $k = 0$ we have, for all $x \in \mathbb{R}$,

$$h_1(x) = q + x = g_0(x) \leq h_2(x).$$

Assume that the assertion is true for some $k-1 \geq 0$ and $\varepsilon > 0$. Since f is increasing and $|x|/\xi < \varepsilon$ for all $|x| < \varepsilon$ we obtain

$$g_k(x) = f_k\left(q + \frac{x}{\xi^k}\right) = f\left(f_{k-1}\left(q + \frac{x/\xi}{\xi^{k-1}}\right)\right) = f\left(g_{k-1}\left(\frac{x}{\xi}\right)\right) \geq f\left(h_1\left(\frac{x}{\xi}\right)\right),$$

and analogously

$$g_k(x) \leq f\left(h_2\left(\frac{x}{\xi}\right)\right).$$

Thus, the induction proof is completed by showing that for some $\varepsilon > 0$ we have

$$f\left(h_1\left(\frac{x}{\xi}\right)\right) \geq h_1(x), \quad f\left(h_2\left(\frac{x}{\xi}\right)\right) \leq h_2(x), \tag{1.14}$$

for all $|x| < \varepsilon$.

Taylor expansion of $x \mapsto f(h_i(x/\xi))$ around $x = 0$ yields for each $i = 1, 2$

$$f(h_i(x/\xi)) = q + x + \frac{1}{2}\left(\frac{h_i''(0)}{\xi} + \frac{f''(q)}{\xi^2}\right) x^2 + O(x^3),$$

for all x in a bounded neighborhood of 0. We have

$$\frac{1}{2}\left(\frac{h_1''(0)}{\xi} + \frac{f''(q)}{\xi^2}\right) = \frac{1}{2}\frac{f''(q)}{\xi^2} > 0$$

by Lemma 1.3.2. From $h_2''(0) = 2c$ and the definition of c it follows

$$\frac{1}{2}\left(\frac{h_2''(0)}{\xi} + \frac{f''(q)}{\xi^2}\right) = \frac{f''(q)}{2\,\xi\,(\xi-1)} + \frac{1}{\xi} < \frac{f''(q)}{2\,\xi\,(\xi-1)} + 1 = c.$$

Thus, there exists an $\varepsilon > 0$ with (1.14) for all $|x| < \varepsilon$. ■

1.3.2 Proof of Theorem 1.3.1

Convergence in distribution: We show that $n^\alpha(F_V(W_n) - q)$ converges in distribution by showing that its distribution functions g_k, $n = d^{2k}$, convergence pointwise to a distribution function g.

Fix $x \in \mathbb{R}$. Since $q < z$ and $f'(q) = \xi > 1$ there is $k_0(x)$ such that $0 < q + x/\xi^k < z$, for all $k \geq k_0(x)$. By Lemma 1.3.2 the function f is convex on $[0, z]$ and satisfies $f(q) = q$. Hence, for all $k \geq k_0(x)$

$$f\left(q + \frac{x}{\xi^k}\right) \geq f(q) + f'(q)\frac{x}{\xi^k} = q + \frac{x}{\xi^{k-1}}$$

and, since f_{k-1} is monotone increasing,

$$g_k(x) = f_k\left(q + \frac{x}{\xi^k}\right) = f_{k-1}\left(f\left(q + \frac{x}{\xi^k}\right)\right) \geq f_{k-1}\left(q + \frac{x}{\xi^{k-1}}\right) = g_{k-1}(x). \tag{1.15}$$

Thus, the sequence $(g_k(x))_{k \geq k_0(x)}$ is monotone increasing and upper bounded, hence convergent. We denote its limit by

$$g(x) := \lim_{k\to\infty} g_k(x), \quad x \in \mathbb{R}.$$

Since g_k is nondecreasing for all $k \geq 1$ its limit g is a nondecreasing function. Since $g_k(0) = f_k(q) = q$ for every $k \geq 0$, we have $g(0) = q$. Continuity of f and $g_k(x) = f(g_{k-1}(x/\xi))$ yields, with $k \to \infty$, the functional equation $g(x) = f(g(x/\xi))$.

Monotonicity of g and $0 \leq g \leq 1$ imply that $\lim_{x\to\infty} g(x)$ and $\lim_{x\to-\infty} g(x)$ exist. Continuity of f and $\xi > 0$ yield with the functional equation for g that

$$\lim_{x\to-\infty} g(x) = f\Big(\lim_{x\to-\infty} g(x)\Big), \quad \lim_{x\to\infty} g(x) = f\Big(\lim_{x\to\infty} g(x)\Big).$$

Hence, both limits are fixed points of f. Lemma 1.3.3 and convergence of g_k yield, with ε as in Lemma 1.3.3,

$$h_1(x) < g(x) < h_2(x), \quad -\varepsilon < x < \varepsilon. \tag{1.16}$$

In a left neighborhood of 0 we have $h_2 < q$. Thus, for some $x < 0$ we have $g(x) < q$, and for appropriate $x > 0$ we have $g(x) > h_1(x) > q$. Since f has only the fixed points 0, q and 1 we obtain $\lim_{x\to-\infty} g(x) = 0$ and $\lim_{x\to\infty} g(x) = 1$.

Hence, $\bar{g}(x) = \lim_{y\downarrow x} g(y)$ for $x \in \mathbb{R}$ is a distribution function with $g_k(x) \to \bar{g}(x)$ for all continuity points x of $\bar{g}$. This implies that $n^\alpha(F_V(W_n) - q) \to W$ in distribution with a random variable W with distribution function $F_W = \bar{g}$.

Note that up to now we only know $g(x) = \bar{g}(x)$ for continuity points x of $\bar{g}$. (We will see below that g is continuous, hence $g(x) = \bar{g}(x) = F_W(x)$ for all $x \in \mathbb{R}$.) ■

Continuity of g: We show that g is continuous in all $x \in \mathbb{R}$ by distinguishing the three cases $x < 0$, $x = 0$ and $x > 0$. Note that for all $x \in \mathbb{R}$ it is sufficient to show that there exists a $\delta > 0$ with

$$\sup\left\{g_k'(y) \Big| |x - y| < \delta, k \geq 0\right\} =: C < \infty. \tag{1.17}$$

From this we obtain $|g_k(x) - g_k(y)| \leq C|x - y|$ for all $k \in \mathbb{N}$ and $|x - y| < \delta$, hence $|g(x) - g(y)| \leq C|x - y|$, in particular g is continuous in x.

Case $x < 0$: The chain rule and induction imply

$$g_k'(x) = \frac{1}{\xi^k} \prod_{i=0}^{k-1} f'\left(f_i\left(q + \frac{x}{\xi^k}\right)\right). \tag{1.18}$$

For $x \le 0$ we have $f_i(q + x/\xi^k) \le q$. Since f' is monotone increasing on $(-\infty, q]$ we obtain $g_k'(x) \le (f'(q)/\xi)^k = 1$ for all $x \le 0$ and $k \ge 0$. Hence, for all $x < 0$ we have (1.17) with $C = 1$.

Case $x = 0$: By Lemma 1.3.3 and $\bar{g}(x) = g(x)$ for all $x < 0$ we obtain

$$\begin{aligned}\mathbb{P}(W < 0) &= \lim_{\ell\to\infty} \mathbb{P}\left(W \le -\frac{1}{\ell}\right) = \lim_{\ell\to\infty} \bar{g}\left(-\frac{1}{\ell}\right) = \lim_{\ell\to\infty} g\left(-\frac{1}{\ell}\right) \\ &\ge \lim_{\ell\to\infty} h_1\left(-\frac{1}{\ell}\right) = q. \end{aligned}\tag{1.19}$$

Since g is a monotone function it has at most countably many discontinuity points. Hence there exists a sequence $(x_\ell)_{\ell\ge1}$ of continuity points of g with $x_\ell \downarrow 0$. Then, with Lemma 1.3.3 we obtain

$$\begin{aligned}\mathbb{P}(W > 0) &= 1 - \lim_{\ell\to\infty} \mathbb{P}\left(W \le x_\ell\right) = 1 - \lim_{\ell\to\infty} \bar{g}\left(x_\ell\right) = 1 - \lim_{\ell\to\infty} g\left(x_\ell\right) \\ &\ge 1 - \lim_{\ell\to\infty} h_2\left(x_\ell\right) = 1 - q. \end{aligned}\tag{1.20}$$

Inequalities (1.19) and (1.20) together imply $\mathbb{P}(W = 0) = 0$, hence g is continuous in $x = 0$. Since we have $g(0) = q$ this implies $F_W(0) = g(0) = q$.

Case $x > 0$: We first show the following assertion:

Claim: There exists a $0 < \varepsilon \le z - q$ such that g_k' is a monotone increasing function on $[0, \varepsilon]$ for all $k \ge 0$.
The claim is shown as follows: Since g is continuous in 0 and $g(0) = q < z$ there exists a $0 < \varepsilon < z - q$ with $g(y) \le z$ for all $0 \le y \le \varepsilon$. By monotonicity of the f_i, we have for all $k \ge 0$, $0 \le i \le k$ and $0 < y' < y \le \varepsilon$

$$f_i(q + y'/\xi^k) \le f_i(q + y/\xi^k) \le f_i(q + y/\xi^i) = g_i(y) \le g(y) \le z. \tag{1.21}$$

For the second last inequality in the latter display note that $(g_i)_{i\ge0}$ is increasing on $(-\infty, z - q)$, cf. (1.15). Since f' is monotone increasing on $(-\infty, z]$ this yields

$$f'(f_i(q + y'/\xi^k)) \le f'(f_i(q + y/\xi^k)), \tag{1.22}$$

thus by (1.18) we obtain $g_k'(y') \le g_k'(y)$ which implies the claim.

Now, assume g is discontinuous in some $x' > 0$. Let ε be as in the previous claim. Note that all the points x'/ξ^k, $k \geq 0$, are discontinuities of g by the functional equation $g(x) = f(g(x/\xi))$ and continuity of f. Hence there exists a discontinuity $0 < x < \varepsilon/2$ of g. By (1.17), we have for all $0 < \delta < (\varepsilon/2 - x) \wedge x$,

$$\sup\left\{g_k'(y) \,\middle|\, y : |y - x| < \delta,\, k \geq 0\right\} = \infty. \tag{1.23}$$

Fix such a δ. By (1.23) and the claim we have $g_m'(x + \delta) \geq 4/\varepsilon$ for a sufficiently large m. Now, the claim implies $g_m'(y) \geq 4/\varepsilon$ for all $y \in [\varepsilon/2, \varepsilon]$. Then,

$$g_m(\varepsilon) - g_m(\varepsilon/2) = \int_{\varepsilon/2}^{\varepsilon} g_m'(y)\, dy \geq \int_{\varepsilon/2}^{\varepsilon} \frac{4}{\varepsilon}\, dy = 2. \tag{1.24}$$

This is a contradiction, since g_m is a distribution function. ■

$\mathbf{0 < F_W < 1}$**:** Assume that $F_W(x) = g(x) \in \{0, 1\}$ for some $x \in \mathbb{R}$. Then $g(x/\xi^k) = g(x)$ for all $k \geq 0$. Hence by continuity of g, we obtain $g(0) \in \{0, 1\}$. Since $g(0) = q \in (0, 1)$ this is a contradiction. ■

1.3.3 Further result on the limit W

In the Diploma thesis of Meiners (2006) further analysis of F_W is done. He shows for f_d, q and ξ as above:

Theorem 1.3.4 (Meiners (2006)) *Distribution function F_W can be extended to a function which is holomorph on $\mathbb{C}$. Hence it has a power series around the center 0,*

$$F_W(x) = \sum_{n \geq 0} a_n x^n,$$

and F_W converges on $\mathbb{C}$. The coefficients a_n are given by

$$a_0 = q, \quad a_1 = 1$$

and

$$a_n = \frac{1}{\xi^n - \xi} \sum_{k=2}^{d^2} \tilde{c}_k \sum_{j_1 + \cdots + j_k = n, j_1, \ldots, j_k \geq 1} a_{j_1} \cdots a_{j_k}, \quad n \geq 2,$$

where $\tilde{c}_0, \ldots, \tilde{c}_k$ are the coefficients of polynom f_d, centered at q. In particular $F_W \in \mathcal{C}^\infty$.

Chapter 2

Tail Bounds for the Wiener index of Random Binary Search Trees

2.1 Introduction and Main Results

The Wiener index of a connected graph is the sum of all distances between all unordered pairs of vertices of the graph. The distance between two vertices is defined as the minimum number of edges connecting them. This index was introduced by chemist Wiener (1947), in order to study relations between organic compounds and the index of their molecular graphs. In particular for trees it is much studied by mathematicians and chemists (cf. Dobrynin, Entringer and Gutman (2001) for survey) but comparably little work has been done for random trees.

Entriger, Meir, Moon and Székely (1994) studied the Wiener index of simply generated families of trees. Given such a family, a simply generated random tree of order n is uniformly distributed on all trees of that family having n nodes. Entriger, Meir, Moon and Székely (1994) proved that the expected Wiener index of the simply generated random tree of order n is asymptotically $Kn^{5/2}$, where K is a constant depending on the simply generated family. Several important tree families are simply generated. For some of them, like ordinary rooted trees, rooted labeled trees and rooted binary trees, they gave even more exact formulæ for the expected Wiener index.

Neininger (2002) analyzed the Wiener index of random binary search trees and random recursive trees. A random binary search tree of order n is generated by a random permutation of numbers $1, \dots, n$, whereas a random recursive tree of order n is uniformly distributed on all recursive trees with n nodes (see Knuth (1998)). The internal path length of a rooted tree is defined as the sum of the distances between the root and all nodes. Neininger (2002) showed that the recursive structure of binary search trees leads to a bivariate distributional recurrence of Wiener index and internal path length for the random binary search tree: Denote (W_n, P_n) Wiener index and internal path length of the random binary search tree of order n, $w_n = \mathbb{E}\, W_n$, $p_n = \mathbb{E}\, P_n$ and I_n and $J_n = n - 1 - I_n$ the cardinalities of the left and right subtree of the root. I_n and J_n are uniformly distributed on $\{0, \dots, n-1\}$. He showed

$$\begin{pmatrix} W_n \\ P_n \end{pmatrix} \stackrel{d}{=} \begin{bmatrix} 1 & n - I_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W_{I_n} \\ P_{I_n} \end{pmatrix} + \begin{bmatrix} 1 & n - J_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W'_{J_n} \\ P'_{J_n} \end{pmatrix} + \begin{pmatrix} 2 I_n J_n + n - 1 \\ n - 1 \end{pmatrix}, \quad (2.1)$$

where (W_i, P_i), (W'_j, P'_j), $0 \le i, j \le n-1$, I_n are independent and $\mathcal{L}((W'_j, P'_j)) = \mathcal{L}((W_j, P_j))$. This distributional recursion enabled Neininger (2002) to study the Wiener index via contraction method. For W_n he obtained exact expectation, asymptotic of the variance and L_2-convergence of

$$\mathbf{Y}_n = \left(\frac{W_n - w_n}{n^2}, \frac{P_n - p_n}{n} \right),$$

where the bivariate limit distributions are characterized uniquely, such that all its mixed moments can be calculated. Note that it was already stated in Hwang and Neininger (2002) that

$$w_n = 2n^2 H_n - 6n^2 + 8n H_n - 10n + 6 H_n, \quad (2.2)$$

where H_n denotes the n-th harmonic number $H_n = \sum_{i=1}^{n} 1/i$. Furthermore he showed that decomposing the random recursive tree of order n (see Mahmoud and Smythe (1994)) into the subtree rooted at the node labeled 2 and the rest of the tree leads to a distributional recursion, similar to (2.1) and he obtained analog results for random recursive trees.

Janson (2003) proved a limit law for the Wiener index of Galton-Watson trees, conditioned on total population size n, as $n \to \infty$, where offspring distribution $\mathcal{L}(X)$ satisfies $\mathbb{E}\, X = 1$ and $\mathrm{Var} X < \infty$. He showed convergence in distribution and with all moments, characterized the limit via a normalized Brownian excursion and obtained a formula for all moments. Aldous (1991) showed that, beside some extreme cases usually not considered, simply generated random trees are distributed as conditioned Galton-Watson trees. Thus, the limit law of Janson (2003) can also be interpreted as a result on simply generated random trees.

Wagner (2006) studied the Wiener index of rooted and unrooted degree-restricted trees. Given a set $\mathcal{D} \subseteq \mathbb{N}$, $1 \in \mathcal{D}$, the family of rooted (unrooted) degree-restricted trees, consists of all rooted (unrooted) trees, for which the degree (number of connected nodes) of every node is in $\mathcal{D}$. This model might be of particular interest for chemists, since molecular graphs are degree-restricted. Given a family of rooted or a family of unrooted degree-restricted trees, a random degree restricted tree of order n is uniformly distributed on all trees of that family, having n nodes. If set $\{d-1|d \in \mathcal{D}, d \neq 1\}$ has greatest common divisor 1, then the expected Wiener index of a the random degree-restricted tree is asymptotically $Kn^{5/2}$, where K is a constant depending on $\mathcal{D}$. Constant K is the same for the family of rooted and the family of unrooted degree-restricted trees, determined by $\mathcal{D}$. Wagner (2006) used the method of Entriger, Meir, Moon and Székely (1994) in order to obtain these results.

In this chapter we are analyzing deviations from the expectation of W_n, the Wiener index of random binary search trees. As an upper bound we obtain the following result:

Theorem 2.1.1 *Let $L_0 \doteq 5.0177$ be the largest root of $e^L = 6L^2$ and $c = (L_0 - 1)/(24L_0^2) \doteq 0.0066$. Then we have for every $t > 0$ and every $n \geq 0$*

$$\mathbb{P}\left(\frac{W_n - w_n}{n^2} > t\right) \leq \begin{cases} \exp(-1/36t^2), & \text{for } 0 \leq t \leq 8.82 \\ \exp(-1/96t^2), & \text{for } 8.82 < t \leq 48L_0 \doteq 240.848 \\ \exp(-ct^2), & \text{for } 48L_0 < t \leq 24L_0^2 \doteq 604.256 \\ \exp(-t(\ln(t) - \ln(4e)), & \text{for } 24L_0^2 < t. \end{cases}$$

The same bound applies for the left tail.

We use the notation $\ln^{(k)}(n)$, where $\ln^{(1)}(n) = \ln(n)$ and $\ln^{(k+1)}(n) = \ln(\ln^{(k)}(n))$. Replacing t by tw_n/n^2 and availing $w_n/n^2 = 2\ln n + O(1)$, Theorem 2.1.1 yields in particular this corollary:

Corollary 2.1.2 *For every $t > 0$ we have for every $n \geq 0$*

$$\mathbb{P}\left(|W_n - w_n| > tw_n\right) \leq n^{-2t\left(\ln^{(2)}(n)+\ln(t)-\ln(2e)+o(1)\right)}.$$

Furthermore we have a lower bound on the tails of W_n:

Theorem 2.1.3 *For fixed* $0 < t \leq 1$ *we have*

$$\mathbb{P}\left(|W_n - w_n| > tw_n\right) \geq \mathbb{P}\left(W_n - w_n > tw_n\right) \geq n^{-8t\left(\ln^{(2)}(n)+O\left(\ln^{(3)}(n)\right)\right)},$$

as $n \to \infty$.

We are going to analyze upper tail bounds by two different methods.

In section 2.2 we introduce our analysis via Chernoff's bounding technique. For this method it is crucial to estimate the moment generating function $\mathbb{E}\exp\langle \mathbf{s}, \mathbf{Y}_n\rangle$, as done by Proposition 2.2.1, in order to get an upper tail bound for W_n via Markov's inequality. Upper tail bounds via Chernoff's bounding technique for P_n, the internal path length of the random binary search tree, are given essentially by Rösler (1992) and explicitly by Fill and Janson (2002). They obtained their estimate of the moment generating function inductively, by using the univariate distributional recurrence

$$P_n \stackrel{d}{=} P_{I_n} + P'_{J_n} + n - 1,$$

for P_i, P'_j, I_n and J_n as in (2.1). Proposition 2.2.1 is also proved by induction on n, now using the bivariate recurrence (2.1) and also different arguments for the inductive step. This proposition is leading to Theorem 2.1.1 and thus corollary 2.1.2.

In section 2.3 we introduce our analysis via the method of bounded differences. The idea is to define by an appropriate filtration a Doob Martingale on W_n and to estimate the martingale differences. Recursion (2.1) is used again for this estimate. Tails of P_n have been analyzed with the method of bounded differences by Hayward and McDiarmid (1996). We are transferring this method on the analysis of W_n and obtain Theorem 2.3.1, which is a slightly weaker estimate than Corollary 2.1.2. The upper tail bounds for P_n, given by Fill and Janson (2002) and by Hayward and McDiarmid (1996) are in the same relationship with each other.

In section 2.4 we prove lower bound result Theorem 2.1.3. We will show that there is a class of binary search trees, having untypically large Wiener indices and that the random binary search tree is in that class with probability at least as large as the right hand side of the inequality in the theorem. This proof is geared to McDiarmid and Hayward's (1996) analysis on the lower tail bounds for P_n.

2.2 Analysis via Chernoff's bounding technique

As already pointed out, it is crucial to prove the following proposition, to obtain Theorem 2.1.1:

Proposition 2.2.1 *Let L_0 be as in Theorem 2.1.1 and $\mathbf{s} \in \mathbb{R}^2$. Then for every $n \geq 1$*

$$\mathbb{E}\exp\langle \mathbf{s}, \mathbf{Y}_n\rangle \leq \begin{cases} \exp\left(9\|\mathbf{s}\|^2\right), & \text{for } 0 \leq \|\mathbf{s}\| \leq 1/2 \\ \exp(24\|\mathbf{s}\|^2), & \text{for } 1/2 < \|\mathbf{s}\| \leq L_0 \\ \exp(4e^{\|\mathbf{s}\|}), & \text{for } L_0 < \|\mathbf{s}\| \end{cases}$$

For $1 \leq i \leq n-1$ and $j = j(i) = n-i-1$ we denote

$$\begin{aligned} a_n^{(1)}(i) &= \begin{bmatrix} (i/n)^2 & i(n-i)/n^2 \\ 0 & i/n \end{bmatrix}, \\ a_n^{(2)}(i) &= a_n^{(1)}(j), \\ C_n^{(1)}(i) &= \frac{1}{n^2}\left(w_i + (n-i)p_i + w_j + (n-j)p_j - w_n + 2ij + n - 1\right), \\ C_n^{(2)}(i) &= \frac{1}{n}\left(p_i + p_j - p_n + n - 1\right) \end{aligned}$$

and $\mathbf{C}_n(i) = (C_n^{(1)}(i), C_n^{(2)}(i))$. With this notation (2.1) is equivalent to distributional recurrence

$$\mathbf{Y}_n \stackrel{d}{=} A_n^{(1)}\mathbf{Y}_{I_n} + A_n^{(2)}\mathbf{Y}'_{J_n} + \mathbf{b}_n, \tag{2.3}$$

for

$$\left(A_n^{(1)}, A_n^{(2)}, \mathbf{b}_n\right) = \left(a_n^{(1)}(I_n), a_n^{(2)}(I_n), \mathbf{C}_n(I_n)\right),$$

where $\mathbf{Y}_i$, $\mathbf{Y}'_j$, $0 \leq i,j \leq n-1$, I_n are independent and $\mathcal{L}(\mathbf{Y}'_j) = \mathcal{L}(\mathbf{Y}_j)$. This will be used in the proof of Proposition 2.2.1 and therefore the following two estimates are needed:

Lemma 2.2.2 *Let U be uniformly distributed on $[0,1]$ and couple I_n, $n \geq 1$, by choosing $I_n = \lfloor Un \rfloor$ a.s. Then we have*

$$\left\|A_n^{(1)T}A_n^{(1)}\right\|_{\text{op}} + \left\|A_n^{(2)T}A_n^{(2)}\right\|_{\text{op}} - 1 < -U(1-U) \quad \text{a.s.},$$

for every $n \geq 1$.

Lemma 2.2.3 *We have*

$$\sup_{n\geq 0}\max_{1\leq i\leq n-1}\|\mathbf{C}_n(i)\| = 1.$$

Proof of Lemma 2.2.2: For $x \in [0,1]$ we set

$$M(x) = \begin{bmatrix} x^2 & x(1-x) \\ 0 & x \end{bmatrix},$$

and get

$$\begin{aligned}
\left\| M(x)^T M(x) \right\|_{\mathrm{op}} &= x^2 \left(1 - x + x^2 + \sqrt{(1+x^2)(1-x)^2} \right) \\
&\leq x^2 \left(1 - x + x^2 + \sqrt{(1+x)^2(1-x)^2} \right) \\
&= x^2(2-x).
\end{aligned}$$

Furthermore we define $\xi = \xi(U,n) \in [0, 1/n)$ by

$$\xi = U - \frac{\lfloor Un \rfloor}{n}.$$

Hence it is $I_n/n = U - \xi$, $J_n/n = 1 - 1/n - U + \xi$ and

$$A_n^{(1)} = M(U-\xi), \quad A_n^{(2)} = M(1 - 1/n - U + \xi).$$

Thus

$$\begin{aligned}
&\left\| A_n^{(1)T} A_n^{(1)} \right\|_{\mathrm{op}} + \left\| A_n^{(2)T} A_n^{(2)} \right\|_{\mathrm{op}} - 1 \\
&\leq (U-\xi)^2 (2 - U + \xi) + \left(1 - \frac{1}{n} - U + \xi \right)^2 \left(1 + \frac{1}{n} + U - \xi \right) - 1 \\
&= -U(1-U) + U^2 \frac{3}{n} + U \frac{3 - 2n - 6\xi n - 2\xi n^2}{n^2} \\
&\quad + \frac{\xi n^3 - 3\xi n + 3\xi^2 n^2 + \xi^2 n^3 + 2\xi n^2 + 1 - n - n^2}{n^3},
\end{aligned}$$

and the proof is completed by showing

$$\begin{aligned}
&U^2 \frac{3}{n} + U \frac{3 - 2n - 6\xi n - 2\xi n^2}{n^2} \\
&+ \frac{\xi n^3 - 3\xi n + 3\xi^2 n^2 + \xi^2 n^3 + 2\xi n^2 + 1 - n - n^2}{n^3} < 0 \quad \text{a.s.}
\end{aligned} \tag{2.4}$$

For proving this, we define for every deterministic $\xi \in [0, 1/n)$, $n \in \mathbb{N}$, a function $r_{\xi,n} : [\xi, 1 - 1/n + \xi] \to \mathbb{R}$ by

$$\begin{aligned}
r_{\xi,n}(u) &= u^2 \frac{3}{n} + u \frac{3 - 2n - 6\xi n - 2\xi n^2}{n^2} \\
&\quad + \frac{\xi n^3 - 3\xi n + 3\xi^2 n^2 + \xi^2 n^3 + 2\xi n^2 + 1 - n - n^2}{n^3}
\end{aligned}$$

Convexity of $r_{\xi,n}$ is given by $r''_{\xi,n}(u) = 6/n$, so we have for every $u \in [\xi, 1 - 1/n + \xi]$

$$r_{\xi,n}(u) \leq r_{\xi,n}(\xi) \vee r_{\xi,n}(1 - 1/n + \xi) < 0, \quad \forall\, n \geq 2.$$

Since for $\xi = \xi(U, n)$ we have $U \in [\xi, 1 - 1/n + \xi]$ a.s. This yields in particular that (2.4) is a.s. true. Furthermore the assertion is trivial for $n = 1$, which completes the proof. ■

Proof of Lemma 2.2.3: Since $\sup_{n\geq 0} \max_{1\leq i\leq n-1} \|\mathbf{C}_n(i)\| \geq \sup_{n\geq 0} C_n^{(2)}(0) = 1$ it suffices to prove

$$\sup_{n\geq 0} \max_{1\leq i\leq n-1} \|\mathbf{C}_n(i)\| \leq 1. \tag{2.5}$$

For fixed $n \geq 1$ and every $0 \leq i \leq n-1$ we define $f(i) = C_n^{(1)}(i) + C_n^{(2)}(i)$, $g(i) = C_n^{(2)}(i) - C_n^{(1)}(i)$ and will prove that $-1 \leq f(i), g(i) \leq 1$. This yields $|C_n^{(1)}(i)| + |C_n^{(2)}(i)| \leq 1$ and thus (2.5).

$-1 \leq f(i) \leq 1$: At first we show that f has increasing increments, thus it is convex. Using formula (2.2) and

$$p_n = 2(n+1)H_n - 4n,$$

one gets by straightforward calculation

$$\begin{aligned} f(i) = 1 &+ \frac{6i(n-i-1) + (4n+6)((i+1)H_i + (n-i)H_{n-i-1} - (n+1)H_n)}{n^2} \\ &+ \frac{12n+7}{n^2} \end{aligned}$$

Hence with

$$(i+2)H_{i+1} - (i+1)H_i = H_{i+1} + 1$$

and

$$(i+1)(n-i-2) - i(n-i-1) = n - 2i - 2$$

we get

$$f(i+1) - f(i) = \frac{6(n-2i-2) + (4n+6)(H_{i+1} - H_{n-i-1})}{n^2}.$$

Thus

$$\begin{aligned}
&f(i+1) - f(i) \leq f(i+2) - f(i+1) \\
&\Leftrightarrow 6(n-2i-2) + (4n+6)(H_{i+1} - H_{n-i-1}) \\
&\qquad \leq 6(n-2(i+1)-2) + (4n+6)(H_{i+2} - H_{n-i-2}) \\
&\Leftrightarrow 12 \leq (4n+6)(H_{i+2} - H_{i+1} + H_{n-i-1} - H_{n-i-2}) \\
&\Leftrightarrow 12 \leq (4n+6)\left(\frac{1}{i+2} + \frac{1}{n-i-1}\right).
\end{aligned}$$

The last inequality is true, because minimizing the right hand side over i yields

$$(4n+6)\left(\frac{1}{i+2} + \frac{1}{n-i-1}\right) \geq 16 + \frac{8}{n+1},$$

thus f is convex. Furthermore f is symmetric at $(n-1)/2$, since $C_n^{(1)}$ and $C_n^{(2)}$ obviously are, hence, for $0 \leq i \leq n-1$,

$$f(\lfloor (n-1)/2 \rfloor) \leq f(i) \leq f(0).$$

So $f(\lfloor (n-1)/2 \rfloor) \geq -1$ and $f(0) \leq 1$ provide $-1 \leq f(i) \leq 1$.

$-1 \leq g(i) \leq 1$: Again, via straightforward calculation we get

$$g(i) = 1 - 6\frac{i(n-i-1) + (i+1)H_i + (n-i)H_{n-i-1} - (n+1)H_n}{n^2} - \frac{6n+7}{n^2}.$$

Analogue calculations and same arguments yield

$$-1 \leq g(\lfloor (n-1)/2 \rfloor) \leq g(i) \leq g(0) \leq 1,$$

which completes the proof. ■

Proof of Propostion 2.2.1: The assertion follows from the next result by choosing $L = \|\mathbf{s}\|$: For every $L > 0$, denote

$$K_L = \begin{cases} 9 & \text{for } L \leq 0.49 \\ 24 & \text{for } 0.49 < L \leq L_0 \\ 4e^L/L^2 & \text{for } L_0 < L. \end{cases}$$

Then

$$\mathbb{E} \exp\langle \mathbf{s}, \mathbf{Y}_n \rangle \leq \exp\left(K_L \|\mathbf{s}\|^2\right), \tag{2.6}$$

for every $\|\mathbf{s}\| \leq L$, $n \geq 0$. This will be proved by induction on n. For $n = 0$ we have $\mathbf{Y}_0 = (0,0)$ and the assertion is true. Assume the assertion is true for some $L > 0$,

$\|\mathbf{s}\| \leq L$ and every $0 \leq i \leq n-1$. Then, conditoning on $I_n = \lfloor Un \rfloor = i$ and using distributional recurrence (2.3) we obtain for $j = n-i-1$ and $\|\mathbf{s}\| \leq L$

$$\begin{aligned}
\mathbb{E} \exp \langle \mathbf{s}, \mathbf{Y}_n \rangle &= \frac{1}{n} \sum_{i=0}^{n-1} \exp \langle \mathbf{s}, \mathbf{C}_n(i) \rangle \, \mathbb{E} \exp \left\langle \mathbf{s}, a_n^{(1)}(i) \, \mathbf{Y}_i \right\rangle \mathbb{E} \exp \left\langle \mathbf{s}, a_n^{(2)}(i) \, \mathbf{Y}_j \right\rangle \\
&\leq \frac{1}{n} \sum_{i=0}^{n-1} \exp \langle \mathbf{s}, \mathbf{C}_n(i) \rangle \exp \left(K_L \left\| a_n^{(1)}(i)^T \, \mathbf{s} \right\|^2 + K_L \left\| a_n^{(2)}(i)^T \, \mathbf{s} \right\|^2 \right) \\
&\leq \frac{1}{n} \sum_{i=0}^{n-1} \exp \left(\langle \mathbf{s}, \mathbf{C}_n(i) \rangle + K_L \|\mathbf{s}\|^2 \sum_{r=1}^{2} \left\| a_n^{(r)}(i)^T a_n^{(r)}(i) \right\|_{\mathrm{op}} \right) \\
&= \mathbb{E} \exp \left(\langle \mathbf{s}, \mathbf{b}_n \rangle + K_L \|\mathbf{s}\|^2 \sum_{r=1}^{2} \left\| A_n^{(r)T} A_n^{(r)} \right\|_{\mathrm{op}} \right) \\
&\leq \mathbb{E} \exp \big(\langle \mathbf{s}, \mathbf{b}_n \rangle + K_L \|\mathbf{s}\|^2 (1 - U(1-U)) \big) \\
&= \mathbb{E} \exp \big(\langle \mathbf{s}, \mathbf{b}_n \rangle - K_L \|\mathbf{s}\|^2 U(1-U) \big) \exp \big(K_L \|\mathbf{s}\|^2 \big) .
\end{aligned}$$

We applied induction hypothesis in the second line, using

$$\|a_n^{(r)}(i)^T \mathbf{s}\| \leq \|a_n^{(r)}(i)^T a_n^{(r)}(i)\|_{\mathrm{op}}^{1/2} \, \|\mathbf{s}\| \leq \|\mathbf{s}\| \leq L,$$

since $\|a_n^{(r)}(i)^T a_n^{(r)}(i)\|_{\mathrm{op}} \leq 1$ for $r = 1, 2,\ 0 \leq i \leq n-1$, and Lemma 2.2.2 in the fifth line. Hence the proof is completed by showing

$$\sup_{n \geq 0} \mathbb{E} \exp \big(\langle \mathbf{s}, \mathbf{b}_n \rangle - K_L \|\mathbf{s}\|^2 U(1-U) \big) \leq 1.$$

Next we are studying the two cases $L \leq 0.49$ and $L > 0.49$.

$L \leq 0.49$: Cauchy-Schwarz inequality yields

$$\begin{aligned}
&\mathbb{E} \exp \big(\langle \mathbf{s}, \mathbf{b}_n \rangle - K_L \|\mathbf{s}\|^2 U(1-U) \big) \\
&\leq \mathbb{E} \exp \left(2 \langle \mathbf{s}, \mathbf{b}_n \rangle \right)^{1/2} \, \mathbb{E} \exp \big(-2 K_L \|\mathbf{s}\|^2 U(1-U) \big)^{1/2} ,
\end{aligned}$$

thus it suffices to prove

$$\mathbb{E} \exp \left(2 \langle \mathbf{s}, \mathbf{b}_n \rangle \right) \, \mathbb{E} \exp \big(-2 K_L \|\mathbf{s}\|^2 U(1-U) \big) \leq 1.$$

With $\|\mathbf{b}_n\| \leq 1$ a.s. by Lemma 2.2.3 and $\mathbb{E}\langle \mathbf{s}, \mathbf{b}_n \rangle = 0$ we obtain

$$\begin{aligned}
\mathbb{E} \exp\left(2\langle \mathbf{s}, \mathbf{b}_n\rangle\right) &= \mathbb{E}\left(1 + 2\langle \mathbf{s}, \mathbf{b}_n\rangle + \sum_{k=2}^{\infty} \frac{(2\langle \mathbf{s}, \mathbf{b}_n\rangle)^k}{k!}\right) \\
&= 1 + \mathbb{E}\langle \mathbf{s}, \mathbf{b}_n\rangle^2 \sum_{k=2}^{\infty} \frac{2^k \langle \mathbf{s}, \mathbf{b}_n\rangle^{k-2}}{k!} \\
&\leq 1 + \|\mathbf{s}\|^2 \sum_{k=2}^{\infty} \frac{2^k (1/2)^{k-2}}{k!} \\
&= 1 + \|\mathbf{s}\|^2 4(e-2) \qquad (2.7)
\end{aligned}$$

and with $K_L = 9$

$$\mathbb{E} \exp\left(-2U(1-U)K_L\|\mathbf{s}\|^2\right) \leq 1 - 3\|\mathbf{s}\|^2 + \frac{27}{5}\|\mathbf{s}\|^4, \qquad (2.8)$$

using

$$\exp(-x) \leq 1 - x + \frac{x^2}{2},$$

for $x \geq 0$. Furthermore

$$\begin{aligned}
&\left(1 + \|\mathbf{s}\|^2 4(e-2)\right)\left(1 - 3\|\mathbf{s}\|^2 + \frac{27}{5}\|\mathbf{s}\|^4\right) \leq 1 \\
\Leftrightarrow\ &\|\mathbf{s}\|^2\left(\frac{108(e-2)}{5}\|\mathbf{s}\|^4 + \left(\frac{147}{5} - 12e\right)\|\mathbf{s}\|^2 + 4e - 11\right) \leq 0 \\
\Leftrightarrow\ &\|\mathbf{s}\| \leq \left(\frac{60e - 147 + 3\left(2600e - 560e^2 - 2879\right)^{1/2}}{216e - 432}\right)^{1/2} \doteq 0.491.
\end{aligned}$$

Thus (2.7) and (2.8) yield that (2.6) is true for $K_L = 9$, $\|\mathbf{s}\| \leq L \leq 0.49$.

$L > 0.49$: Again, with $\|\mathbf{b}_n\| \leq 1$ we get

$$\mathbb{E} \exp\left(\langle \mathbf{s}, \mathbf{b}_n\rangle - K_L\|\mathbf{s}\|^2 U(1-U)\right) \leq \exp(\|\mathbf{s}\|)\mathbb{E} \exp\left(-K_L\|\mathbf{s}\|^2 U(1-U)\right).$$

It is proved in Section 4 of Fill and Janson (2001) that the right hand side of the latter inequality is smaller than 1 if $0.42 \leq \|\mathbf{s}\| \leq 2$ and $K_L = 24$, respectively if $2 \leq \|\mathbf{s}\| \leq L$ and $K_L = 4e^L/L^2$. Thus for $K_L = 24L^2 \vee 4e^L/L^2$ we have $\mathbb{E} \exp\langle \mathbf{s}, \mathbf{Y}_n\rangle \leq \exp(K_L\|\mathbf{s}\|^2)$, for every $\|\mathbf{s}\| \leq L$, $n \geq 0$. Since $24L^2 \geq 4e^L/L^2$, for $L \leq L_0$ and $24L^2 \leq 4e^L/L^2$, for $L > L_0$, this completes the proof. ∎

Proof of Theorem 2.1.1: With Chernoff's bounding technique we have for $u > 0$

$$\begin{aligned}\mathbb{P}\left(\frac{W_n - w_n}{n^2} > t\right) &= \mathbb{P}\left(\exp\left(u\frac{W_n - w_n}{n^2}\right) > \exp(ut)\right)\\ &\leq \mathbb{E}\exp\left(u\frac{W_n - w_n}{n^2} - ut\right)\\ &= \mathbb{E}\exp\left(\langle (u,0), \mathbf{Y}_n\rangle - ut\right)\\ &\leq \exp\left(K_u u^2 - ut\right),\end{aligned}$$

for all $n \geq 0$ and K_u as in the proof of Proposition 2.2.1. Minimizing over $u > 0$ we obtain the bounds

$$\mathbb{P}\left(\frac{W_n - w_n}{n^2} > t\right) \leq \begin{cases} \exp(-1/36t^2), & \text{for } 0 \leq t \leq 8.82 \\ \exp(-1/96t^2), & \text{for } 8.82 < t \leq 48L_0 \\ \exp(-t(\ln(t) - \ln(4e)), & \text{for } 24L_0^2 < t. \end{cases}$$

and choosing $u = t/(24L_0)$ for $2 < u \leq L_0$ we obtain the bound

$$\mathbb{P}\left(\frac{W_n - w_n}{n^2} > t\right) \leq \exp\left(-\frac{L_0 - 1}{24L_0^2}\, t^2\right),$$

for $48L_0 < t \leq 24L_0^2$. This completes the proof. ■

Proof of Corollary 2.1.2: Choosing $t_n = tw_n/n^2 = 2t\ln(n) + O(1)$ we get from Theorem 2.1.1

$$\begin{aligned}\mathbb{P}\left(|W_n - w_n| > tw_n\right) &\leq \exp(-(2t\ln(n) + O(1))(\ln(2t\ln(n) + O(1)) - \ln(4e))\\ &= \exp\left(-2t\ln(n)\left(\ln^{(2)}(n) + \ln(t) - \ln(2e) + o(1)\right) + O\left(\ln^{(2)}(n)\right)\right)\\ &= \exp\left(-2t\ln(n)\left(\ln^{(2)}(n) + \ln(t) - \ln(2e) + o(1)\right)\right),\end{aligned}$$

where we used $\ln(x + O(1)) = \ln(x) + o(1)$, as $x \to \infty$, in the second line and $O(1)O(\ln^{(2)}(n)) = -2t\ln(n)o(1)$ in the third line. This completes the proof. ■

2.3 Analysis via method of bounded differences

Applying the method of bounded differences we obtain the following result:

Theorem 2.3.1 *Let $t = t_n$ satisfy $0 < t \leq 1$. Then as $n \to \infty$*

$$\mathbb{P}\left(|W_n - w_n| > tw_n\right) \leq n^{-2t\left(\ln^{(2)}(n) + \ln(t) + O(\ln^{(3)}(n))\right)}.$$

For ease of description we embed the random binary search tree underlying (W_n, P_n) in the complete infinite binary tree. The nodes of the complete infinite binary tree may be numbered $1, 2, 3, \ldots$, level by level and left to right. So, for instance the left most node in level k is node number 2^k, for every $k \geq 0$. If node m belongs to the binary search tree, let S_m be the size of the subtree of the binary search tree, rooted at node m. If node m does not belong to the random binary search tree, set $S_m = 0$. S_m is called size of node m. Denote $\mathbf{H}_k$ the vector of sizes of all nodes up to level k,

$$\mathbf{H}_k = (S_1, \ldots, S_{2^{k+1}-1}).$$

$\mathbf{H}_k$ determines up to level k, which nodes belong to the random binary search tree and furthermore the sizes of the nodes in level k. $\mathbf{H}_k$ is called k-history. Given a deterministic k-history $\mathbf{h}$ with $\mathbb{P}(\mathbf{H}_k = \mathbf{h}) > 0$ we define for random variables X and Y, defined on the same finite probability space and events E of this space

$$\begin{aligned} \mathbb{P}_{\mathbf{h}}(X) &= \mathbb{P}(X|\mathbf{H}_k = \mathbf{h}), \\ \mathbb{E}_{\mathbf{h}}(X) &= \mathbb{E}(X|\mathbf{H}_k = \mathbf{h}) \\ \mathbb{E}_{\mathbf{h}}(X|E) &= \mathbb{E}(X|\{\mathbf{H}_k = \mathbf{h}\} \cap E) \end{aligned}$$

and conditional expectation $\mathbb{E}_{\mathbf{h}}(X|Y)$ by

$$\mathbb{P}(\mathbb{E}_{\mathbf{h}}(X|Y) = \mathbb{E}_{\mathbf{h}}(X|Y = y)) = \mathbb{P}(Y = y),$$

for every y in the codomain of Y.

In order to prove Theorem 2.3.1 we are going to estimate for fix n the differences of martingale

$$M_k = \mathbb{E}_{\mathbf{h}}(W_n|H_{k_1+k}), \quad 0 \leq k \leq k_2 - k_1, \tag{2.9}$$

where $k_1 < k_2$ are positive integers and $\mathbf{h}$ is a deterministic k_1-history. Therefore the following estimates are done:

Equation (2.1) yields that given $I_n = i$ (thus $J_n = n - i - 1 = j$) Wiener index W_n is

$$W_i + (n-i)P_i + W_j + (n-j)P_j + 2ij + n - 1$$

and internal path length P_n is

$$P_i + P_j + n - 1.$$

For 1-history $\mathbf{h} = (n, i, n-i-1)$ it is $\{I_n = i\} = \{\mathbf{h} = \mathbf{H}_1\}$. Hence the previous two expressions and Lemma 2.2.3 yield

$$\begin{aligned}
&|\,\mathbb{E}_{\mathbf{h}}\,(W_n) - w_n| \\
&= |w_i + (n-i)p_i + w_j + (n-j)p_j + 2ij + n - 1 - w_n| \\
&= \left|C_n^{(1)}(i)\right| n^2 \\
&\leq n^2 \qquad (2.10)
\end{aligned}$$

and

$$\begin{aligned}
|\,\mathbb{E}_{\mathbf{h}}\,(P_n) - p_n| &= |p_i + p_j + n - 1 - p_n| \\
&= \left|C_n^{(2)}(i)\right| n \\
&\leq n. \qquad (2.11)
\end{aligned}$$

These two inequalities lead to the following crucial estimate:

Lemma 2.3.2 *For a random binary search tree of size n, $k < n$, let $\mathbf{h}$ be an arbitrary deterministic k-history. Then*

$$|\,\mathbb{E}_{\mathbf{h}}\,(W_n) - w_n| \leq kn^2.$$

Proof: For $k = 1$ the result is given by the inequality (2.10). For $k \geq 2$ and a fix $(k+1)$-history $\mathbf{h}'$ let $\mathbf{h}$ be the corresponding k-history and $s(1), \ldots, s(2^k)$ be the sizes of nodes at level k. Then we get for suitable 1-histories $\mathbf{h}(1), \ldots, \mathbf{h}(2^k)$ of random binary search trees of orders $s(1), \ldots, s(2^k)$

$$\begin{aligned}
&|\,\mathbb{E}_{\mathbf{h}'}\,(W_n) - \,\mathbb{E}_{\mathbf{h}}\,(W_n)| \\
&= \left| \sum_{m=1}^{2^k} \left\{ \,\mathbb{E}_{\mathbf{h}(m)}\left(W_{s(m)}\right) - w_{s(m)} + (n - s(m))\left[\,\mathbb{E}_{\mathbf{h}(m)}\left(P_{s(m)}\right) - p_{s(m)}\right]\right\} \right| \\
&\leq \sum_{m=1}^{2^k} \left\{ s(m)^2 + (n - s(m))\, s(m) \right\} \\
&= n \sum_{m=1}^{2^k} s(m) \\
&\leq n^2.
\end{aligned}$$

In the third line we used (2.10) and (2.11). The prove is completed by induction on k with triangle-inequality. ■

This enables us to prove the upcoming lemma. Therefore denote L_{nk} the maximal size of nodes at level k, that is

$$L_{nk} = \max\left\{S_{2^k+m} | \, 0 \le m \le 2^k - 1\right\}.$$

Lemma 2.3.3 *Let $k_1 < k_2$ be positive integers, $\alpha > 0$ and $\mathbf{h}$ be a k_1-history, for which $L_{nk_1} \le \alpha n$. Then for any $t > 0$*

$$\mathbb{P}_{\mathbf{h}}(| \, \mathbb{E}_{\mathbf{h}}(W_n|\mathbf{H}_{k_2}) - \, \mathbb{E}_{\mathbf{h}}(W_n)| \ge t) \le 2\exp\left(\frac{-t^2}{2\alpha(k_2-k_1)n^4}\right).$$

Note that $M_0 = \mathbb{E}_{\mathbf{h}}(W_n)$ and $M_{k_2-k_1} = \mathbb{E}_{\mathbf{h}}(W_n|H_{k_2-k_1})$ (see (2.9)) and that Lemma 2.3.3 is essentially the estimate for martingale $(M_0, \ldots, M_{k_2-k_1})$ mentioned in section 2.1. For proving this lemma we use Hoeffding's inequality and a version of Azuma's inequality:

Lemma 2.3.4 (Hoeffding (1963)) *Let X be a random variable with $\mathbb{E}\, X = 0$, $a \le X \le b$. Then for $u > 0$,*

$$\mathbb{E}\,\exp(uX) \le \exp(u^2(b-a)^2/8).$$

Lemma 2.3.5 (Hayward and McDiarmid (1996)) *Let $\mathcal{F}_0$ be the trivial σ-algebra, $\mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_n$ a filtration, X an integrable random variable and $(X_0, \ldots, X_n)$ the corresonding Doob martingale, i.e. $X_k = \mathbb{E}(X|\mathcal{F}_k)$. Suppose that for each $1 \le k \le n$ there is a constant c_k such that*

$$\mathbb{E}(\exp(u(X_k - X_{k-1})|\mathcal{F}_{k-1}) \le \exp(c_k^2 u^2),$$

for every u. Then we have for every $t > 0$

$$\mathbb{P}(|X_n - X_0| \ge t) \le 2\exp\left(\frac{-t^2}{4\sum_{k=1}^n c_k^2}\right).$$

Proof of Lemma 2.3.3: If we can show

$$\mathbb{E}_{\mathbf{h}}(\exp(u(M_k - M_{k-1})|\mathbf{H}_{k_1+k-1}) \le \exp\left(\frac{\alpha}{2}n^4u^2\right), \tag{2.12}$$

for every $1 \le k \le k_2 - k_1$, then the result follows from Lemma 2.3.5. For fix $k \ge 0$ set $m = k_1 + k$ and let $\mathbf{h}'$ a possible m-history extending of $\mathbf{h}$. Define the random variable T by

$$T = \mathbb{E}_{\mathbf{h}'}(W_n|\mathbf{H}_{m+1}) - \mathbb{E}_{\mathbf{h}'}(W_n).$$

Then inequality (2.12) is equivalent to showing for any possible extension $\mathbf{h}'$ and any u that

$$\mathbb{E}_{\mathbf{h}'}(\exp(uT)) \le \exp\left(\frac{\alpha}{2} n^4 u^2\right).$$

Given $\mathbf{H}_m = \mathbf{h}'$ the nodes at level m have deterministic sizes, say $s(1), \dots, s(2^m)$, and the subtrees rooted at these nodes are independent. Thus

$$T \stackrel{d}{=} \sum_{i=1}^{2^m} \left\{ \mathbb{E}\left(W_{s(i)} \middle| \mathbf{H}(i)\right) - w_{s(i)} + (n - s(i)) \left[\mathbb{E}\left(Q_{s(i)} \middle| \mathbf{H}(i)\right) - q_{s(i)} \right] \right\},$$

where $\mathbf{H}(1), \dots, \mathbf{H}(2^m)$ are random 1-histories, induced by $\mathcal{L}(\mathbf{H}_{m+1}|\mathbf{h}')$ and the 2^m summands on the right hand side are independent. Furthermore

$$\mathbb{E}\left\{ \mathbb{E}\left(W_{s(i)} \middle| \mathbf{H}(i)\right) - w_{s(i)} + (n - s(i)) \left[\mathbb{E}\left(Q_{s(i)} \middle| \mathbf{H}(i)\right) - q_{s(i)} \right] \right\} = 0$$

and

$$\left| \mathbb{E}\left(W_{s(i)} \middle| \mathbf{H}(i)\right) - w_{s(i)} + (n - s(i)) \left[\mathbb{E}\left(Q_{s(i)} \middle| \mathbf{H}(i)\right) - q_{s(i)} \right] \right| \le n s_i,$$

for $1 \le i \le 2^m$, as is implicitly given by the calculation in the proof of Lemma 2.3.2. Each $s(i) \le \alpha n$, by assumption $L_{nk_1} \le \alpha n$, and thus $\sum_i s(i)^2 \le \alpha n \sum_i s(i) \le \alpha\, n^2$. Together with Hoeffding's inequality this yields

$$\begin{aligned}
&\mathbb{E}_{\mathbf{h}'} \exp(uT) \\
&= \prod_{i=1}^{2^j} \mathbb{E} \exp\left(\mathbb{E}\left(W_{s(i)} \middle| \mathbf{H}(i)\right) - w_{s(i)} + (n - s(i)) \left[\mathbb{E}\left(Q_{s(i)} \middle| \mathbf{H}(i)\right) - q_{s(i)} \right] \right) \\
&\le \prod_{i=1}^{2^j} \exp\left(u^2 \frac{n^2 s(i)^2}{2} \right) \\
&\le \exp\left(u^2 \frac{\alpha\, n^4}{2} \right).
\end{aligned}$$

This completes the proof. ■

The lemma stated next is essentially given by Devroye (1986):

Lemma 2.3.6 (Hayward and McDiarmid (1996)) *For any* $0 < \alpha < 1$ *and any integer* $k \geq \ln(1/\alpha)$ *it is*

$$\mathbb{P}(L_{nk} \geq \alpha n) \leq \alpha \left(\frac{2e\ln(1/\alpha)}{k} \right)$$

In particular the probability that we have a k_1-history for which Lemma 2.3.3 is not applicable is estimated by this lemma, if $k_1 \geq \ln(1/\alpha)$. Hence, together with Lemma 2.3.2 we are able to prove the next one:

Lemma 2.3.7 *Let* n, k_1, *and* u *be positive integers. Then for any* $0 < \alpha \leq 1$ *and integer* $k_2 > k_1$ *such that* $\ln(1/\alpha) \leq k_1$, $k_2 \geq \ln(n/2)$ *we have*

$$\begin{aligned}
&\mathbb{P}\left(|W_n - w_n| \geq k_1 n^2 + u\right) \\
&\leq \frac{2}{n}\left(\frac{2e\ln(n/2)}{k_2}\right)^{k_2} + \alpha\left(\frac{2e\ln(1/\alpha)}{k_1}\right)^{k_1} + 2\exp\left(\frac{-u^2}{2(k_2-k_1)\alpha\, n^4}\right)
\end{aligned}$$

Proof: Denote $R_n = \mathbb{E}\left(W_n|\mathbf{H}_{k_2}\right)$ and $\mathcal{H}$ the set of k_1-histories $\mathbf{h}$ with $L_{nk_1} \leq \alpha n$. Then

$$\begin{aligned}
&\mathbb{P}(|W_n - w_n| \geq k_1 n^2 + u) \\
&\leq \mathbb{P}(|R_n - w_n| \geq k_1 n^2 + u \text{ and } \mathbf{H}_{k_1} \in \mathcal{H}) \\
&\quad + \mathbb{P}(R_n \neq W_n) + \mathbb{P}(\mathbf{H}_{k_1} \notin \mathcal{H}) \\
&= \sum_{\mathbf{h}\in\mathcal{H}} \mathbb{P}_{\mathbf{h}}(|R_n - w_n| \geq k_1 n^2 + u)\, \mathbb{P}(\mathbf{H}_{k_1} = \mathbf{h}) \\
&\quad + \mathbb{P}(R_n \neq W_n) + \mathbb{P}(\mathbf{H}_{k_1} \notin \mathcal{H}) \\
&\leq \sum_{\mathbf{h}\in\mathcal{H}} \mathbb{P}_{\mathbf{h}}(|R_n - \mathbb{E}_{\mathbf{h}}(W_n)| \geq u)\, \mathbb{P}(\mathbf{H}_{k_1} = \mathbf{h}) \\
&\quad + \mathbb{P}(L_{nk_2} \geq 2) + \mathbb{P}(L_{nk_1} > \alpha n).
\end{aligned}$$

For the last inequality, we used

$$|R_n - w_n| \leq |R_n - \mathbb{E}_{\mathbf{h}}(W_n)| + |\mathbb{E}_{\mathbf{h}}(W_n) - w_n| \leq |R_n - \mathbb{E}_{\mathbf{h}}(W_n)| + k_1 n^2$$

by Lemma 2.3.2 and $\{W_n = R_n\} \supseteq \{L_{nk_2} \leq 1\}$. The result now follows from Lemmas 2.3.3 and 2.3.6. ∎

Choosing the parameters in Lemma 2.3.7 appropriately finally leads to Theorem 2.3.1:

Proof of Theorem 2.3.1: Without loss of generality we can assume that $t_n \geq 5\ln^{(2)}(n)/\ln(n)$, since otherwise the estimate in Theorem 2.3.1 might be 1. We choose

$$\begin{aligned}
u &= \left\lceil \frac{tn^2 \ln(n)}{\ln^{(2)}(n)} \right\rceil, \\
k_1 &= \left\lfloor 2t\ln(n) - 2\frac{u}{n^2} \right\rfloor = 2t\ln(n)\left(1 + O\left(1/\ln^{(2)}(n)\right)\right), \\
k_2 &= \left\lceil \ln(n)\ln^{(2)}(n) \right\rceil, \\
\alpha &= \frac{t^2}{\ln^{(2)}(n)^5}.
\end{aligned}$$

Observe that

$$\begin{aligned}
k_1 n^2 + u &\leq 2tn^2\ln(n) - u \\
&\leq 2tn^2\ln(n) - 7tn^2 \\
&\leq tw_n,
\end{aligned}$$

for sufficiently large n. It is proved by Hayward and McDiarmid (1996), for $k_1 = 2t\ln(n)\left(1 + O\left(1/\ln^{(2)}(n)\right)\right)$ and α and k_2 as above that $k_1 \geq \ln(1/\alpha)$, for sufficiently large n,

$$\frac{2}{n}\left(\frac{2e\ln(n/2)}{k_2}\right)^{k_2} \leq \exp\left(-k_2\ln^{(3)}(n)\right) \tag{2.13}$$

and

$$\alpha\left(\frac{2e\ln(1/\alpha)}{k_1}\right)^{k_1} \leq \exp\left(-2t\ln(n)\left(\ln^{(2)}(n) + \ln(t) + O\left(\ln^{(3)}(n)\right)\right)\right). \tag{2.14}$$

$k_1 \geq \ln(1/\alpha)$ yields that Lemma 2.3.7 is applicable, whereas (2.13) proves that the first summand on the right hand side of the inequality in this lemma is smaller than required and (2.14) proves that the second summand is exactly as required. With

$$\begin{aligned}
2\exp\left(\frac{-u^2}{2(k_2-k_1)\alpha\, n^4}\right) &\leq 2\exp\left(-\frac{1+o(1)}{2}\frac{t^2n^4\ln(n)^2/\ln^{(2)}(n)^2}{\ln(n)\ln^{(2)}(n)\,t^2/\ln^{(2)}(n)^5 n^4}\right) \\
&= 2\exp\left(-\frac{1+o(1)}{2}\ln(n)\ln^{(2)}(n)^2\right) \qquad (2.15)
\end{aligned}$$

we have that the third summand is also smaller than required, which completes the proof. ■

2.4 Lower Bound

In this section we prove Theorem 2.1.3, which is a lower bound on $\mathbb{P}(W_n > (1+t)w_n)$. Roughly speaking, the Wiener index of a binary search tree of order n is rather large, if it has two nodes which have a large distance and both nodes have large sizes. Based on this observation we define for every fix $t > 0$ a class of binary search trees of order n. Every tree in that class has two nodes, with sufficiently large distance and large sizes, such that conditioned on the event that the random binary search tree is in that class, event $\{W_n > (1+t)w_n\}$ has probability tending to 1, as $n \to \infty$. Moreover the probability that the random binary search tree is in that class is at least as large as the right hand side of the inequality stated in Theorem 2.1.3. We have to define this class carefully in order to assure these two conflicting properties.

Proof of Theorem 2.1.3: Since we just study the event that the random binary search tree is in the above mentioned class, we will define this class only implicitly by defining event A below. Therefore we denote for fixed $t > 0$

$$\begin{aligned}
\lambda &= \frac{\ln^{(3)}(n)}{\ln^{(2)}(n)}, \\
\kappa &= 8 + 24\lambda, \\
k &= \lfloor \kappa t \ln(n) \rfloor, \\
s &= \left\lfloor \frac{\lambda n}{t \ln(n)} \right\rfloor .
\end{aligned}$$

Recall that S_i is the size of the subtree rooted at node i, respectively 0 if no such node exists, and that node $2^m + 1$ is the second leftmost node in level m. Let A be the event that $S_2 = \lfloor (n+1)/2 \rfloor$ and that $S_{2^m+1} \leq s - 1$, for $2 \leq m \leq k$, see figure 2.1.

Thus under event A we have $S_3 = \lceil (n-3)/2 \rceil$ and $S_{2^k} \geq n/2 - (k-1)s$. Having two large subtrees this far away from each other will yield that W_n is sufficiently

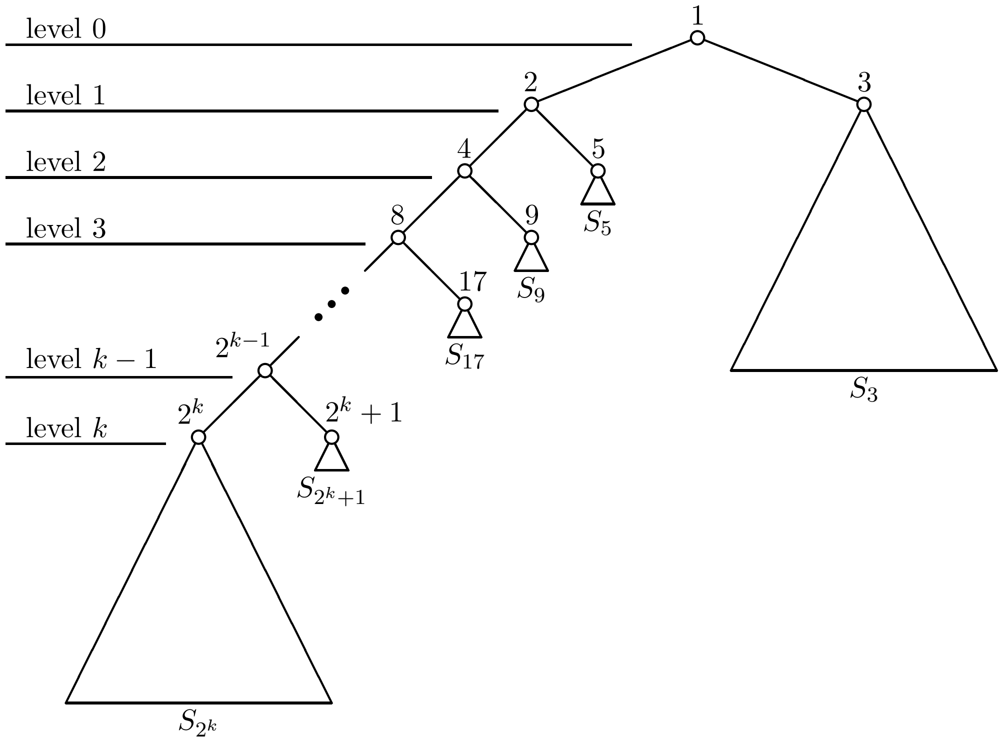

Figure 2.1: Under event A we have subtree sizes $S_3 = \lceil (n-3)/2 \rceil$ and $S_{2^m+1} \le s-1$, for $2 \le m \le k$, thus $S_{2^k} \ge n/2 - (k-1)s$.

large. But first note that

$$\begin{aligned}
\mathbb{P}(A) &\ge \frac{1}{n}\left(\frac{s}{(n+1)/2}\right)^{k-1} \\
&\ge \frac{1}{n}\left(\frac{s}{n}\right)^{k-1} \\
&= \exp(-(k-1)(\ln(n/s)) - \ln(n)) \\
&\ge \exp\left(-8t\ln(n)\left(\ln^{(2)}(n) + O\left(\ln^{(3)}(n)\right)\right)\right). \qquad (2.16)
\end{aligned}$$

As from now we will assume w.l.o.g. that n is even, since all further calculations are almost the same if n is odd.

The distance between two nodes in a tree can be visualized as the minimal number of edges one has to pass in order to get from one node to the other. From that point of view the Wiener index of a tree can be calculated by counting how often each edge is passed when summing up all node distances. In our modell the edge above node i is passed $S_i(n-S_i)$ times. Thus

$$W_n = \sum_{i\in\mathbb{N}} S_i(n-S_i),$$

where exactly $n-1$ summands on the right hand side are nonzero. We set

$$W_n' = \sum_{m=1}^{k} S_{2^m}(n-S_{2^m}).$$

and $W_n'' = W_n - W_n'$ and will estimate W_n' and W_n'' seperately under event A. Descriptively W_n' is the number of passings of the edges above the nodes 2^m, $1 \le m \le k$. For $(s_2,\dots,s_k) \in M = \{1,\dots,s\}^{k-1}$ let $A(s_2,\dots,s_k)$ be the event that $S_3 = \lceil (n+1)/2 \rceil$ and that $S_{2^m+1} = s_m - 1$, for $2 \le m \le k$. Thus

$$A = \bigcup_{(s_2,\dots,s_k)\in M} A(s_2,\dots,s_k).$$

Denote $\sigma_1 = 0$ and $\sigma_m = \sigma_{m-1} + s_m$ for $2 \le m \le k$. Then $(m-1) \le \sigma_m \le (m-1)s$ and under event $A(s_2,\dots,s_k)$ we have

$$\begin{aligned}
W_n' &= \sum_{m=1}^{k} \left(\frac{n}{2}+\sigma_m\right)\left(\frac{n}{2}-\sigma_m\right) \\
&= \sum_{m=1}^{k} \left(\frac{n^2}{4}-\sigma_m^2\right) \\
&\ge \frac{kn^2}{4} - s^2 \sum_{m=1}^{k} (m-1)^2 \\
&\ge \frac{kn^2}{4}\left(1-\frac{4}{3}\frac{k^2s^2}{n^2}\right) \\
&\ge \left((1+3\lambda)2\ln(n)t - \frac{1}{4}\right) n^2 \left(1-\frac{4}{3}\kappa^2\lambda^2\right) \\
&= t\,2n^2\ln(n)\left(1+3\lambda-\frac{1}{t8\ln(n)}\right)\left(1-\frac{4}{3}\kappa^2\lambda^2\right) \\
&\ge t\,2n^2\ln(n)\,(1+\lambda)\,, \qquad (2.17)
\end{aligned}$$

for sufficiently large n. In the last step we used

$$\begin{aligned}
\left(1+3\lambda-\frac{1}{t8\ln(n)}\right)\left(1-\frac{4}{3}\kappa^2\lambda^2\right) &\ge (1+2\lambda)\left(1-\frac{4}{3}\kappa^2\lambda^2\right) \\
&\ge 1+\lambda,
\end{aligned}$$

for sufficiently large n.

In order to estimate W_n'' under event $A(s_2, \dots, s_k)$ via Chebychev's inequality, we will use

$$\begin{aligned}\mathbb{E}\left(W_n''|A(s_2,\dots,s_k)\right) \geq w_{n/2-1} &+ \left(\frac{n}{2}+1\right)p_{n/2-1}\\ &+ w_{n/2-\sigma_k} + \left(\frac{n}{2}+\sigma_k\right)p_{n/2-\sigma_k}\\ &+ \sum_{m=2}^{k}\left(w_{s_m-1} + (n-s_m+1)p_{s_m-1}\right). \end{aligned} \tag{2.18}$$

This inequality is valid, since the right hand side is the number of passings of all edges belonging to subtrees rooted at either node 3 (first row) or node 2^k (second row) or node 2^m+1, $2 \leq m \leq k$, (third row). With $H_x \geq \ln(x)$ we get for $x \leq n$

$$\begin{aligned} w_x + (n-x)p_x &\geq 2x^2\ln(x) - 6x^2 + o(x^2) + (n-x)\left(2x\ln(x) - 4x\right)\\ &\geq n(2x\ln(x) - 6x + o(x)). \end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}\left(W_n''|A(s_2,\dots,s_k)\right) &\geq 2n\left(\frac{n}{2}-1\right)\ln\left(\frac{n}{2}-1\right) + 2n\left(\frac{n}{2}-\sigma_k\right)\ln\left(\frac{n}{2}-\sigma_k\right)\\ &\quad+\sum_{m=2}^{k}2n(s_m-1)\ln(s_m-1) - 6n^2 + o(n^2)\\ &\geq 2n(n-\sigma_k-1)\ln\left(\frac{n}{2}-\sigma_k\right)\\ &\quad+2n(k-1)(\hat{s}-1)\ln(\hat{s}-1) - 6n^2 + o(n^2),\end{aligned}$$

where $\hat{s} = 1/(k-1)\sum_{m=2}^{k} s_m$. And with $\sigma_k = (k-1)\hat{s} \leq (k-1)s$

$$\begin{aligned}(n-\sigma_k-1)\ln\left(\frac{n}{2}-\sigma_k\right) &\geq (n-(k-1)\hat{s})\left(\ln(n) + \ln\left(1-\frac{2(k-1)s}{n}\right) - \ln(2)\right)\\ &= n\ln(n) - \log(2)n - (k-1)\hat{s}\ln(n) + o(n).\end{aligned}$$

Together this yields

$$\begin{aligned}
\mathbb{E}\left(W_n''|A(s_2,\ldots,s_k)\right) &\geq 2n^2\ln(n) - 2n(k-1)(\hat{s}-1)\ln\left(\frac{n}{\hat{s}-1}\right) \\
&\quad - (6+2\ln(2))n^2 - 2n(k-1)\ln(n) + o(n^2) \\
&\geq 2n^2\ln(n) - 2n(k-1)(s-1)\ln\left(\frac{n}{s-1}\right) \\
&\quad - (6+2\ln(2))n^2 + o(n^2) \\
&= 2n^2\ln(n) - 2\kappa\lambda n^2\ln\left(\frac{t\ln(n)}{\lambda}\right) - (6+2\ln(2))n^2 + o(n^2) \\
&\geq 2n^2\ln(n) - (16+o(1))n^2\ln^{(3)}(n),
\end{aligned}$$

for sufficiently large n. In the second line we used that $x \mapsto x\ln x$ is increasing for $x \leq 1/e$ and that $\hat{s}-1 < s < 1/e$ for large n. Similarly to (2.18) we have

$$\begin{aligned}
\mathrm{Var}(W_n''|A(s_2,\ldots,s_k)) &= \mathrm{Var}\left(W_{n/2-1} + \left(\frac{n}{2}+1\right)P_{n/2-1}\right) \\
&\quad + \mathrm{Var}\left(W_{n/2-\sigma_k} + \left(\frac{n}{2}+\sigma_k\right)P_{n/2-\sigma_k}\right) \\
&\quad + \sum_{m=2}^{k}\mathrm{Var}\left(W_{s_m-1} + (n-s_m+1)P_{s_m-1}\right)
\end{aligned}$$

and for $x \leq n$

$$\begin{aligned}
\mathrm{Var}\left(W_x + (n-x)P_x\right) &= \mathrm{Var}(W_x) + (n-x)^2\mathrm{Var}(P_x) + 2(n-x)\mathrm{Cov}(W_x,P_x) \\
&= O(x^4) + n^2O(x^2) + 2nO(x^3),
\end{aligned}$$

since $\mathrm{Var}(W_n) = O(n^4)$, $\mathrm{Cov}(W_n,P_n) = O(n^3)$ (see Hwang and Neininger (2002)) and $\mathrm{Var}(P_n) = O(n^2)$. Thus

$$\mathrm{Var}(W_n''|A(s_2,\ldots,s_k)) = O\left(n^4\right)$$

and hence by Chebychev's inequality

$$\mathbb{P}\left(W_n'' \geq 2n^2\ln(n) - 17n^2\ln^{(3)}(n)|A(s_2\ldots,s_k)\right) \to 1 \quad \text{as } n \to \infty. \tag{2.19}$$

This convergence is uniform over all $(s_2,\ldots,s_k) \in M$. For sufficiently large n,

$$t\,2n^2\ln(n)\,(1+\lambda) + 2n^2\ln(n) - 17n^2\ln^{(3)}(n) > (1+t)w_n. \tag{2.20}$$

Using estimates (2.16), (2.17), (2.19) and (2.20) we get

$$
\begin{aligned}
&\mathbb{P}(W_n > (1+t)w_n) \\
&\geq \mathbb{P}(W_n > (1+t)w_n | A)\mathbb{P}(A) \\
&= \sum_{(s_2,\ldots,s_k)\in M} \mathbb{P}(W_n > (1+t)w_n | A(s_2,\ldots,s_k))\mathbb{P}(A(s_2,\ldots,s_k)) \\
&\geq \sum_{(s_2,\ldots,s_k)\in M} \mathbb{P}(W_n'' > 2n^2\ln(n) - 17n^2\ln^{(3)}(n) | A(s_2\ldots,s_k))\mathbb{P}(A(s_2,\ldots,s_k)) \\
&= (1+o(1))\mathbb{P}(A) \\
&= \exp\left(-8t\ln(n)\left(\ln^{(2)}(n) + O\left(\ln^{(3)}(n)\right)\right)\right).
\end{aligned}
$$

This completes the proof. ■

Chapter 3

Tail Bounds for the Generation Size of Supercritical Multitype Galton-Watson Processes

3.1 Introduction

A singletype Galton-Watson process is a Markov chain $(Z_n)_{n\geq 0}$ on nonnegative integers, with $Z_0 = 1$ and

$$Z_n = X^{(n,1)} + \cdots + X^{(n,Z_{n-1})},$$

where $Z_{n-1}, X^{(n,1)}, X^{(n,2)}, \ldots$ are independent and $X^{(n,1)}, X^{(n,2)}, \ldots$ furthermore identically distributed according to some probability distribution μ on $\mathbb{N}_0$. Z_n can be thought as the number of individuals of a population at time n. So, during a time step all individuals propagate independent of each other and of the past according to distribution μ and die immediately after propagating. Hence μ is called offspring distribution of the process.

A d-type Galton-Watson process, $d \geq 1$, is a Markov chain $(\mathbf{Z}_n^{[i]})_n \geq 0$ on $\mathbb{N}_0^d$, $\mathbf{Z}_n^{[i]} = (Z_n^{[i]}(1), \ldots, Z_n^{[i]}(d))$, for $1 \leq i \leq d$. Here $\mathbf{Z}_0^{[i]} = \mathbf{e}_i$, where $\mathbf{e}_i$ denotes the i-th unit vector and

$$\mathbf{Z}_n^{[i]} = \sum_{j=1}^{Z_{n-1}^{[i]}(1)} \mathbf{X}^{[1],(n,j)} + \cdots + \sum_{j=1}^{Z_{n-1}^{[i]}(d)} \mathbf{X}^{[d],(n,j)},$$

where $\mathbf{Z}_{n-1}^{[i]}$, $\mathbf{X}^{[k],(n,j)}$, $j \geq 1$, $1 \leq k \leq d$, are independent and for every k vector $\mathbf{X}^{[k],(n,j)}$ is distributed according to some probability distribution $\mu^{[k]}$ on $\mathbb{N}_0^d$. $\mathbf{Z}_n^{[i]} = (Z_n^{[i]}(1), \ldots, Z_n^{[i]}(d))$ can be thought as the vector of numbers of individuals of type $1, \ldots, d$ at time n, when starting with an individual of type i. So, at time 0 there is only a single type i individual and during a time step, all individuals propagate independent of each other and of the past, according to some distribution $\mu^{[k]}$, depending on their type k. Although $\mu = (\mu^{[1]}, \ldots, \mu^{[d]})$ is not a probability distribution, but a vector of probability distributions, it is called offspring distribution, since it describes the propagation mechanism, like the offspring distribution does for the singletype Galton-Watson process. Offspring distribution μ is called bounded, if there is an $\ell \in \mathbb{N}$, such that $\mu^{[i]}(\{1, \ldots, \ell\}) = 1$, for $1 \leq i \leq d$. If a Galton-Watson process has a bounded offspring distribution, it has a finite maximum family size at time n, $n \geq 0$. A d-type Galton-Watson process is called singular, if each particle has a.s. exactly one offspring, otherwise it is called nonsingular. Denote $(X^{[i]}(1), \ldots, X^{[i]}(d)) = \mathbf{X}^{[i]} = \mathbf{Z}_1^{[i]}$ and mean matrix $M = [\mathbb{E}\, X^{[i]}(j)]_{1 \leq i,j \leq d}$. If it exists some $n \geq 1$ such that matrix M^n has only positive entries, then M is called strictly positive and the d-type Galton-Watson process is called positive regular. Frobenius theorem yields that if M is strictly positive, it has a largest eigenvalue $\varrho > 0$ and associated right and left eigenvectors $\mathbf{u} = (u_1, \ldots, u_d)$ and $\mathbf{v} = (v_1, \ldots, v_d)$, respectively exist, which are positive in each component and may be normalized, so that $\sum_i u_i = 1$ and $\sum_i u_i v_i = 1$. If $\varrho > 1$, the process is called supercritical. A nonsingular, positive regular d-type Galton-Watson process does not a.s. extinct, if and only if it is supercritical. See Harris (1963) for these results. Furthermore

Theorem 3.1.1 (Keesten, H. and Stigum, B. (1966)) *Let $(\mathbf{Z}_n^{[i]})$ be a nonsingular, positive regular, supercritical d-type Galton-Watson process. Then*

$$\lim_{n \to \infty} \frac{\mathbf{Z}_n^{[i]}}{\varrho^n} = \mathbf{v} W^{[i]} \quad a.s.,$$

where $W^{[i]}$ is a nonnegative random variable, such that

$$\mathbb{P}(W^{[i]} > 0) > 0 \Longleftrightarrow \mathbb{E}\, X^{[k]}(j) \ln X^{[k]}(j) < \infty \quad \forall 1 \leq k, j \leq d.$$

If $\mathbb{P}(W^{[i]} > 0) > 0$, then it is furthermore $\mathbb{E}\, W^{[i]} = u_i$.

Note that for singletype Galton-Watson processes this theorem yields: If $\varrho = \mathbb{E}\, Z_1 >$

1, then

$$\lim_{n\to\infty} \frac{Z_n}{\varrho^n} = W \quad \text{a.s.},$$

where W is a nonnegative random variable, such that

$$\mathbb{P}(W>0)>0 \Longleftrightarrow \mathbb{E}\, Z_1 \ln Z_1 < \infty.$$

If $\mathbb{P}(W>0)>0$, then it is furthermore $\mathbb{E}\, W = 1$.

Tails of both Z_n, $n \geq 0$, and of limit W have been analyzed for singletype Galton-Watson processes with $\varrho > 1$ and finite maximum family sizes.

Theorem 3.1.2 (Biggins, J.D. and Bingham, N.H.) *Let $(Z_n)_n$ be a single-type Galton-Watson process with finite maximum family size $m = \operatorname{ess\,sup} Z_1$ and $m > \varrho > 1$. Then there is a real analytic, multiplicatively periodic function $F : (0,\infty) \to (0,\infty)$ with period m/ϱ, such that for*

$$\tilde{\kappa} = 1 + \frac{1}{\ln(m)/\ln(\varrho) - 1} > 1$$

we have

$$\mathbb{P}\left(W - \mathbb{E}\, W > t\right) = \exp\left(-t^{\tilde{\kappa}}(F(t+1) + o(1))\right), \qquad t \to \infty.$$

F is bounded, since it is real analytic and multiplicatively periodic. Hence this theorem yields in particular that there are positive constants $\tilde{\alpha}_0$ and $\tilde{\alpha}_1$ with

$$\exp(-\tilde{\alpha}_1 t^{\tilde{\kappa}}) \leq \mathbb{P}\left(W - \mathbb{E}\, W > t\right) \leq \exp(-\tilde{\alpha}_0 t^{\tilde{\kappa}}), \qquad \forall\, t > 0.$$

Karp and Zhang gave a comparable result for Z_n:

Theorem 3.1.3 (Karp, R. and Zhang, Y. (1995)) *Let Z_n, ϱ, m, $\tilde{\kappa}$ be as in Theorem 3.1.2. Then*

$$\mathbb{P}\left(\frac{Z_n - \mathbb{E}\, Z_n}{\varrho^n} > t\right) \leq \exp\left(-\alpha_0(t+1)^{\tilde{\kappa}} + c_0\right), \qquad t \geq 1,$$

where $\alpha_0 > 0$ and c_0 are positve constants depending on ϱ and m and

$$\mathbb{P}\left(\frac{Z_n - \mathbb{E}\, Z_n}{\varrho^n} > t\right) \geq \exp\left(-\alpha_1(t+1)^{\tilde{\kappa}} + c_1\right), \qquad 0 \leq t \leq \left(\frac{m}{2\varrho}\right)^n - 1,$$

where $\alpha_1 = -m \ln \mathbb{P}(Z_1 = m) > 0$ and c_1 approaches 1 as t increases.

Karp and Zhang (1995) gave furthermore a right tail bound for the generation size

$$G_n^{[i]} = \sum_{j=1}^{d} Z_n^{[i]}(j)$$

of some supercritical multitype Galton-Watson processes:

Theorem 3.1.4 (Karp and Zhang (1995)) *Let $\mathbf{Z}_n^{[i]}$ be a positive regular supercritical d-type Galton-Watson process with finite maximum family size $m = \sup\{k \mid \exists\, 1 \leq j \leq d : \mathbb{P}(G_1^{[j]} = k) > 0\}$. If $\varrho > \sqrt{m}$ then for any $t > 0$*

$$\mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} > t\right) \leq \exp\left(-\alpha t^2\right),$$

where $\alpha > 0$ is a constant depending on m and ϱ.

A generalization of Theorem 3.1.2 on supercritical multitype Galton-Watson processes is given by Jones (2004). He studied the tails of random variable $W^{[i]}$, given in Theorem 3.1.1. This theorem yields that

$$\lim_{n\to\infty} \frac{G_n^{[i]}}{\varrho^n} = G^{[i]} \quad \text{a.s.},$$

for some nonnegative random variable $G^{[i]}$ and furthermore that $G^{[i]}$ and $W^{[i]}$ can be easily transferred into each other by

$$W^{[i]} \sum_{j=1}^{d} v_j = G^{[i]}.$$

We will state Jones' result in terms of $G^{[i]}$. Therefore we have to introduce some more notations. Denote $\mathbb{N}_0^{d\times d}$ the set of all $(d \times d)$-matrices, having $\mathbb{N}_0$-valued entries, the i-th row of a matrix $B \in \mathbb{N}_0^{d\times d}$ by $B_{i,\cdot}$. For $\mathbf{Z}_n^{[i]}$, ϱ, $\mathbf{u}$, $\mathbf{v}$ as in Theorem 3.1.1, furthermore $\mathbf{Z}_n^{[i]}$ having a finite maximum family size, and $\mathbf{x} = (x_1, \ldots, x_d)$, $\mathbf{x}' = (x_1', \ldots, x_d')$ we define

$$\begin{aligned}
J^{[i]} &= \left\{\mathbf{x} \in \mathbb{N}_0^d \,\middle|\, \mathbb{P}\left[\mathbf{X}^{[i]} = \mathbf{x}\right] > 0\right\}, \\
U^{[i]} &= \left\{\mathbf{x} \in J^{[i]} \,\middle|\, \forall \mathbf{x}' \neq \mathbf{x}\, \exists j : x_j > x_j'\right\}, \\
U &= \left\{B \in \mathbb{N}_0^{d\times d} \,\middle|\, B_{i,\cdot} \in U^{[i]}\right\}, \\
\mathcal{U}_1(\mathbf{x}) &= \max_{B\in U} B\mathbf{x}, \\
\mathcal{U}_n(\mathbf{x}) &= \mathcal{U}_1(\mathcal{U}_{n-1}(\mathbf{x})), \quad n > 1, \\
\lambda &= \sup_{x_i \geq 0 : \|\mathbf{x}\|=1} \lim_{n\to\infty} \|\mathcal{U}_n(\mathbf{x})\|^{1/n}
\end{aligned}$$

Theorem 3.1.5 (Jones (2004)) *Let $G_n^{[i]}$ be the generation size of a positive regular, supercritical d-type Galton-Watson process with finite maximum family size and $G^{[i]} = \lim_n (G_n^{[i]})/\varrho^n$ a.s. Assume that $\varrho < \lambda$, that it exists an up to a scale factor unique $\mathbf{w} = (w_1, \ldots, w_d)$, with $w_i > 0$, $1 \le i \le d$, and $\mathcal{U}(\mathbf{w}) = \lambda \mathbf{w}$, besides an unique $C \in U$ with $\mathcal{U}(\mathbf{w}) = C\mathbf{w}$ and that $\lim_{n\to\infty} \lambda^{-n} C^n$ exists. Then there is a continuous, multiplicatively periodic function $F : (0,\infty) \to \mathbb{R}_+^d$, $F(t) = (F^{[1]}(t), \ldots, F^{[d]}(t))$, with period λ/ϱ, such that for $\tilde{\kappa} = 1 + 1/(\log\lambda/\log\varrho - 1)$ and every $1 \le i \le d$*

$$\mathbb{P}\left(G^{[i]} - \mathbb{E}\, G^{[i]} > t\right) = \exp\left(-t^{\tilde{\kappa}}\left(F^{[i]}(t + u_i) + o(1)\right)\right), \quad t \to \infty.$$

In this chapter, we will give upper tail bounds on the generation size of supercritical multitype Galton-Watson processes with finite maximum family sizes. We will prove under other conditions than in Theorem 3.1.4 tail bounds $\exp(-\text{const}\ t^{\kappa})$, for $2 \le \kappa < \kappa^*$, where $\kappa^* > 2$ is a constant, depending on offspring distribution μ. In the next section, necessary notations and main result of this chapter are stated. Since it might be difficult to calculate κ^* for a given process, in section 3.4 we give a lower bound on exponent κ^*, which is easily expressed in terms of μ. We do not claim positive regularity of the process in our main result, thus this result is leading to upper tail bounds for the the generation size of Galton-Watson processes with immigration, as is explained in section 3.6. In section 3.5 we will explain that the exponent in Theorem 1.2.6 cannot be improved, as claimed in the first chapter.

3.2 A Tail Bound for the Generation Size of Supercritical Multitype Galton Watson Processes

Approach for our analysis are recursive descriptions of $G_n^{[1]}, \ldots, G_n^{[d]}$. To exemplify these recursive descriptions, let $G_n^{[1]}$, $G_n^{[2]}$ be the generation sizes of the 2-type Galton-Watson process, Karp and Zhang constructed, in order to analyze Snir's algorithm applied on a binary Boolean decision tree of height $2n$ (cf. section 1.2). In subsection 1.2.4 it is explained that the random bivariate sequence $(\mathbf{G}_n)_{n\ge 0}$, given by $\mathbf{G}_0 = (1,1)$ and

$$\mathbf{G}_n \stackrel{d}{=} \mathbf{G}_{n-1}^{(1)} + \mathbf{G}_{n-1}^{(2)} + \begin{bmatrix} B_1 B_2 & 0 \\ 1 - B_2 & 0 \end{bmatrix} \mathbf{G}_{n-1}^{(3)} + \begin{bmatrix} 0 & B_1 \\ B_1 & 0 \end{bmatrix} \mathbf{G}_{n-1}^{(4)}, \quad n \ge 1 \qquad (3.1)$$

where $\mathbf{G}_{n-1}^{(1)}, \ldots, \mathbf{G}_{n-1}^{(4)}$, B_1, B_2 are independent, B_1, B_2 are Bernoulli-(1/2) distributed and $\mathcal{L}(\mathbf{G}_{n-1}^{(1)}) = \cdots = \mathcal{L}(\mathbf{G}_{n-1}^{(4)}) = \mathcal{L}(\mathbf{G}_{n-1})$, has marginals distributed as $G_n^{[1]}$ and $G_n^{[2]}$, respectively. So, one can say that (3.1) is a recursive description of $G_n^{[1]}$ and $G_n^{[2]}$. But there are also other recursive descriptions of $G_n^{[1]}$ and $G_n^{[2]}$. E.g. $\tilde{\mathbf{G}}_n$ given by $\tilde{\mathbf{G}}_0 = (1,1)$ and

$$\tilde{\mathbf{G}}_n \stackrel{d}{=} \tilde{\mathbf{G}}_{n/4}^{(1)} + \tilde{\mathbf{G}}_{n/4}^{(2)} + \begin{bmatrix} B_1 B_2 & 0 \\ 0 & 0 \end{bmatrix} \tilde{\mathbf{G}}_{n/4}^{(3)} + \begin{bmatrix} 0 & B_1 \\ B_3 & 0 \end{bmatrix} \tilde{\mathbf{G}}_{n/4}^{(4)}, + \begin{bmatrix} 0 & 0 \\ B_4 & 0 \end{bmatrix} \tilde{\mathbf{G}}_{n/4}^{(5)},$$

for $n \geq 1$, where $\tilde{\mathbf{G}}_{n-1}^{(1)}, \ldots, \tilde{\mathbf{G}}_{n-1}^{(5)}$, $B_1, \ldots, B_4$ are independent, $B_1, \ldots, B_4$ are Bernoulli-(1/2) distributed and $\mathcal{L}(\tilde{\mathbf{G}}_{n-1}^{(1)}) = \cdots = \mathcal{L}(\tilde{\mathbf{G}}_{n-1}^{(5)}) = \mathcal{L}(\tilde{\mathbf{G}}_{n-1})$ has marginals distributed as $G_n^{[1]}$ and $G_n^{[2]}$, respectively. Here, the marginals are furthermore independent. It can easily be verified that our approach would only yield a weaker version of Theorem 1.2.6 if we would use $\tilde{\mathbf{G}}_n$ instead of $\mathbf{G}_n$ for our analysis. Namely, tail bound $\exp(-\text{const } t^\kappa)$ with $2 \leq \kappa < \ln(5)/\ln(5/4^\alpha) \doteq 2.85$ instead of $2 \leq \kappa < 1/(1-\alpha) \doteq 4.06$. So, for how large exponents κ we can prove a tail bound $\exp(-\text{const } t^\kappa)$, depends on which recursive description we use.

Theorem 3.2.1 yields for the tails of generation sizes $G_n^{[i]}$ of multitype Galton-Watson processes with $\varrho > 1$ bounds $\exp(-\text{const } t^\kappa)$, for $2 \leq \kappa < \kappa^*$, where $\kappa^* > 2$ has to be specified. Due to our observation above, we specify how various recursive descriptions of $G_n^{[1]}, \ldots, G_n^{[d]}$ can look like and which upper bounds they yield on exponents κ, in order to get a possibly large κ^*.

Each recursive description of $G_n^{[1]}, \ldots, G_n^{[d]}$ is determined by a vector of random matrices. E.g. the two above discussed sequences are determined by

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} B_1 B_2 & 0 \\ 1 - B_2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & B_1 \\ B_1 & 0 \end{bmatrix} \right) \tag{3.2}$$

and

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} B_1 B_2 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & B_1 \\ B_3 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ B_4 & 0 \end{bmatrix} \right),$$

respectively.

Next, we will state Theorem 3.2.1, where κ^* is expressed in terms of such sequences of random matrices. Therefore we have to introduce some more notations. In section 3.3 we will show that these sequences determine recursive descriptions of the generation sizes.

In cases, where we want to emphasize that the process has offspring distribution $\mu = (\mu^{[1]}, \ldots, \mu^{[d]})$, we write $\mathbf{Z}_n^{[i]}[\mu]$, $G_n^{[i]}[\mu]$ and $\mathbf{X}^{[i]}[\mu]$ instead of $\mathbf{Z}_n^{[i]}$, $G_n^{[i]}$ and $\mathbf{X}^{[i]}$

respectively. Denote

$$\varrho[\mu] = \inf \left\{ \varrho > 0 \,\middle|\, \sup_{n \geq 0} \frac{\mathbb{E}\, G_n^{[i]}[\mu]}{\varrho^n} < \infty \quad \forall\, 1 \leq i \leq d \right\}.$$

It is $\varrho[\mu] < \infty$ for every bounded μ and we assume throughout this chapter that $\sup_{n\geq 0} \mathbb{E}\, G_n^{[i]}[\mu]/\varrho[\mu]^n < \infty$ and $\varrho[\mu] > 1$. Karlin (1966) proved for positve regular Galton-Watson processes $\mathbf{Z}_n^{[i]}$, with the largest eigenvalue ϱ of mean matrix M, that $\mathbb{E}\,\mathbf{Z}_n^{[i]}/\varrho^n \to c' > 0$, as $n \to \infty$. Hence, if the process is positive regular, $\varrho[\mu] > 1$ is the largest eigenvalue of the mean matrix and the process is supercritical. Recall that $\mathbb{N}_0^{d\times d}$ is the set of all $(d \times d)$-matrices with $\mathbb{N}_0$-valued entries. For an $\mathbb{N}_0^{d\times d}$-valued random variable A, respectively $A \in \mathbb{N}_0^{d\times d}$, denote by $A_{i,j}$ the (i,j)-th component and as before by $A_{i,\cdot}$ the i-th row of A. For a random vector $\mathbf{A}$ with m $\mathbb{N}_0^{d\times d}$-valued components denote by $A^{(r)}$ its r-th component, $1 \leq r \leq m$, and $\mathbf{A}_{i,\cdot} = (A_{i,\cdot}^{(r)})_{1\leq r\leq m}$. Assume that the underlying probability space $(\Omega, F, \mathbb{P})$ is sufficiently large and denote

$$\begin{aligned}
\mathcal{A}_1 &= \left\{ A \in \mathbb{N}_0^{d\times d} \,\middle|\, A_{i,j} = \mathbf{1}[E_{i,j}],\ E_{i,j} \in F,\ \max_{1\leq i\leq d} \sum_{j=1}^d A_{i,j} \leq 1 \text{ a.s.} \right\}, \\
\mathcal{A}_1^k &= \left\{ \mathbf{A} = \left(A^{(1)}, \ldots, A^{(k)}\right) \,\middle|\, A^{(r)} \in \mathcal{A}_1,\ 1 \leq r \leq k \right\}, \\
\mathcal{A}_2 &= \left\{ A \in \mathbb{N}_0^{d\times d} \,\middle|\, A_{i,j} = \mathbf{1}[E_{i,j}],\ E_{i,j} \in F,\ \max_{1\leq j\leq d} \sum_{i=1}^d A_{i,j} \leq 1 \text{ a.s.} \right\}, \\
\mathcal{A}_2^k &= \left\{ \mathbf{A} = \left(A^{(1)}, \ldots, A^{(k)}\right) \,\middle|\, A^{(r)} \in \mathcal{A}_2,\ 1 \leq r \leq k,\ \mathbf{A}_{1,\cdot}, \ldots, \mathbf{A}_{d,\cdot} \text{ independent } \right\}, \\
\mathcal{A}_\ell^* &= \bigcup_{k\geq 1} \mathcal{A}_\ell^k \quad \text{for } \ell = 1, 2, \\
\mathcal{A}^* &= \mathcal{A}_1^* \cup \mathcal{A}_2^*.
\end{aligned}$$

Furthermore, let $\mathcal{O}$ be the set of all offspring distributions of d-type Galton-Watson processes, i.e.

$$\mathcal{O} = \left\{ \left(\mu^{[1]}, \ldots, \mu^{[d]}\right) \,\middle|\, \mu^{[i]} \text{ is a probability measure on } \mathbb{N}_0^d,\ 1 \leq i \leq d \right\}$$

and for $\mu \in \mathcal{O}$ let $\mu_k = (\mu_k^{[1]}, \ldots, \mu_k^{[d]}) \in \mathcal{O}$ be given by

$$\mu_k^{[i]} = \mathcal{L}\left(\mathbf{Z}_k^{[i]}[\mu]\right), \quad 1 \leq i \leq d.$$

Let $T : \mathcal{A}^* \to \mathcal{O}$ be defined by

$$T(\mathbf{A}) = \left(\mathcal{L}\left(\sum_{r\geq 1} A_{1,\cdot}^{(r)} \right), \ldots, \mathcal{L}\left(\sum_{r\geq 1} A_{d,\cdot}^{(r)} \right) \right).$$

T is well defined, since vector $\mathbf{A} \in \mathcal{A}^*$ has finitely many components, thus the sums on the right hand side are finite. Even more, offspring distribution $T(\mathbf{A})$ is bounded for every $\mathbf{A} \in \mathcal{A}^*$ and also $T^{-1}(\mu)$ is nonempty for every bounded μ, as will be explained in section 3.4. We define $q : \mathcal{A}^* \to \mathbb{R}_+ \cup \{\infty\}$ by

$$q(\mathbf{A}) = \inf \left\{ q \in \mathbb{R}_+ \middle| \operatorname{ess\,sup} \sum_{r \geq 1} \left\| A^{(r)T} A^{(r)} \right\|_{\mathrm{op}}^{q/2} < \varrho[T(\mathbf{A})]^q \right\} \vee 1 \, ,$$

respectively $q(\mathbf{A}) = \infty$, if the infimum does not exist and $q^* : \mathcal{O} \to \mathbb{R}_+ \cup \{\infty\}$ by

$$q^*(\mu) = \inf \left\{ q(\mathbf{A}) \middle| \exists\, k \in \mathbb{N} : \mathbf{A} \in T^{-1}\left(\mu_k\right) \right\} . \tag{3.3}$$

Note that if $\mathbf{A} \in \mathcal{A}_1^*$ then matrix $A^{(r)T} A^{(r)}$ has diagonal entries $\sum_i A_{i,1}^{(r)}, \ldots, \sum_i A_{i,d}^{(r)}$ and all other entries 0. Hence

$$\left\| A^{(r)T} A^{(r)} \right\|_{\mathrm{op}} = \max_{1 \leq i \leq d} \sum_{j=1}^{d} A_{i,j}^{(r)} \tag{3.4}$$

and for $\mathbf{A} \in \mathcal{A}_2^*$ it is

$$\left\| A^{(r)T} A^{(r)} \right\|_{\mathrm{op}} = \left\| A^{(r)} A^{(r)T} \right\|_{\mathrm{op}} = \max_{1 \leq j \leq d} \sum_{i=1}^{d} A_{i,j}^{(r)} .$$

So, one can calculate $q(\mathbf{A})$, easily. Setting $1/0 = \infty$ we get the following tail bound result:

Theorem 3.2.1 *Let $G_n^{[i]} = G_n^{[i]}[\mu]$ be the generation size of a d-type Galton-Watson process with finite maximum family size and $\varrho = \varrho[\mu] > 1$. If $q^* = q^*(\mu) < 2$, then for every $2 \leq \kappa < \kappa^* = 1 + 1/(q^* - 1)$ there exists an $L_\kappa > 0$ such that for every $n \geq 0$ and $1 \leq i \leq d$*

$$\begin{aligned}
\mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} > t \right) &\leq \exp(-L_\kappa t^\kappa) \quad t > 0, \\
\mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} < -t \right) &\leq \exp(-L_\kappa t^\kappa) \quad 0 < t \leq \frac{\mathbb{E}\, G_n^{[i]}}{\varrho^n}.
\end{aligned}$$

This theorem will be proved in section 3.7 by applying Chernoff's bounding technique on Proposition 3.3.3, which is the crucial result for this tail bound and which is stated in the next section.

3.3 Recursive Descriptions

Next, we discuss how recursive descriptions of $G_n^{[1]}, \ldots, G_n^{[d]}$ can look like. This might be useful to get an idea why every $\mathbf{A} \in \mathcal{A}^*$ with $T(\mathbf{A}) = \mu$ determines a recursive description of $G_n^{[1]}, \ldots, G_n^{[d]}$ as stated in Propostion 3.3.1. Corollary 3.3.2 is a generalization which is easily derived from Proposition 3.3.1. Corollary 3.3.2 enables us to prove Theorem 3.2.1 via Proposition 3.3.3. The proofs of these results are given in section 3.7.

The individuals of a Galton-Watson process at time 1 propagate independent of each other. Hence, $\mathcal{L}(G_n^{[i]})$, $1 \leq i \leq d$, can be described recursively by

$$G_n^{[i]} \overset{d}{=} \sum_{r=1}^{X^{[i]}(1)} G_{n-1}^{[1],(r)} + \cdots + \sum_{r=1}^{X^{[i]}(d)} G_{n-1}^{[d],(r)}, \tag{3.5}$$

where $\mathbf{X}^{[i]}$, $G_{n-1}^{[j],(r)}$, $1 \leq j \leq d$, $r \geq 1$, are independent and $\mathcal{L}(G_{n-1}^{[j],(r)}) = \mathcal{L}(G_{n-1}^{[j]})$ and $G_0^{[i]} = 1$ a.s. This equation can be rephrased to

$$G_n^{[i]} \overset{d}{=} \sum_{r\geq 1} \mathbf{1}[X^{[i]}(1) \geq r]\, G_{n-1}^{[1],(r)} + \cdots + \sum_{r\geq 1} \mathbf{1}[X^{[i]}(d) \geq r]\, G_{n-1}^{[d],(r)}$$

and since we observe equality in distribution we have for all Bernoulli-distributed random variables $B_{i,j}^{(r)}$, with $\{B_{i,j}^{(r)} | 1 \leq j \leq d,\, r \geq 1\}$ independent of $G_{n-1}^{[j],(r)}$, $1 \leq j \leq d$, $r \geq 1$, which satisfy

$$\left(X^{[i]}(1), \ldots, X^{[i]}(d)\right) \overset{d}{=} \left(\sum_{r\geq 1} B_{i,1}^{(r)}, \ldots, \sum_{r\geq 1} B_{i,d}^{(r)}\right) \tag{3.6}$$

likewise

$$G_n^{[i]} \overset{d}{=} \sum_{r\geq 1} B_{i,1}^{(r)}\, G_{n-1}^{[1],(r)} + \cdots + \sum_{r\geq 1} B_{i,d}^{(r)}\, G_{n-1}^{[d],(r)}.$$

In particular, if $\mu = T(\mathbf{A})$ for some $\mathbf{A} \in \mathcal{A}^*$, this observation can be restated to

$$G_n^{[i]}[\mu] \overset{d}{=} \sum_{r\geq 1} A_{i,1}^{(r)}\, G_{n-1}^{[1],(r)}[\mu] + \cdots + \sum_{r\geq 1} A_{i,d}^{(r)}\, G_{n-1}^{[d],(r)}[\mu], \tag{3.7}$$

for $\mathbf{A}$, $G_{n-1}^{[j],(r)}[\mu]$, $1 \leq j \leq d$, $r \geq 1$, independent, $\mathcal{L}(G_{n-1}^{[j],(r)}[\mu]) = \mathcal{L}(G_{n-1}^{[j]}[\mu])$. The right hand side of (3.7) is the i-th component of

$$\sum_{r\geq 1} A^{(r)} \left(G_{n-1}^{[1],(r)}[\mu], \ldots, G_{n-1}^{[d],(r)}[\mu]\right)^T.$$

Even more, the following proposition yields that $G_n^{[1]}[\mu], \ldots, G_n^{[d]}[\mu]$ can be described recursively via $\mathbf{A}$:

Proposition 3.3.1 *Let $\mathbf{A} \in \mathcal{A}^*$, $T(\mathbf{A}) = \mu$ and $\mathbf{G}_n$, $n \geq 0$, be a d-dimensional random vector, with distribution $\mathcal{L}(\mathbf{G}_n)$ given by $\mathbf{G}_0 = (1, \ldots, 1)$ and*

$$\mathbf{G}_n \stackrel{d}{=} \sum_{r \geq 1} A^{(r)} \, \mathbf{G}_{n-1}^{(r)} \,, \tag{3.8}$$

for $\mathbf{A}$, $\mathbf{G}_{n-1}^{(r)}$, $r \geq 1$, independent and $\mathcal{L}(\mathbf{G}_{n-1}^{(r)}) = \mathcal{L}(\mathbf{G}_{n-1})$. Then $\mathbf{G}_n$ has marginals distributed as $G_n^{[1]}[\mu], \ldots, G_n^{[d]}[\mu]$ respectively. If $\mathbf{A} \in \mathcal{A}_2^*$, then the marginals of $\mathbf{G}_n$ are furthermore independent.

Instead of recursive descriptions of $G_n^{[1]}[\mu], \ldots, G_n^{[d]}[\mu]$, recursive descriptions of $G_{nk+\ell}^{[1]}, \ldots, G_{nk+\ell}^{[d]}$, for fixed $k \in \mathbb{N}$ and $0 \leq \ell < k$ can be used for the analysis. In this more general case we have analogously to (3.5)

$$G_{nk+\ell}^{[i]} \stackrel{d}{=} \sum_{r=1}^{Z_k^{[i]}(1)} G_{(n-1)k+\ell}^{[1],(r)} + \cdots + \sum_{r=1}^{Z_k^{[i]}(d)} G_{(n-1)k+\ell}^{[d],(r)} \,, \tag{3.9}$$

for $\mathbf{Z}_k^{[i]}$, $G_{(n-1)k+\ell}^{[j],(r)}$, $1 \leq j \leq d$, $r \geq 1$, independent and $\mathcal{L}(G_{(n-1)k+\ell}^{[j],(r)}) = \mathcal{L}(G_{(n-1)k+\ell}^{[j]})$. Here, we used that the individuals at time k propagate independent of each other. Accordingly, we get the following result derived from this distributional equation:

Corollary 3.3.2 *Given $\mu \in \mathcal{O}$ bounded, $k \in \mathbb{N}$, $\ell \in \mathbb{N}_0$, let $\mathbf{A} \in \mathcal{A}^*$, $T(\mathbf{A}) = \mu_k$ and $\mathbf{G}_n$, $n \geq 0$, be a d-dimensional random vector, where $\mathbf{G}_0$ has independent components, distributed as $G_1^{[i]}[\mu_\ell]$, $1 \leq i \leq d$, and*

$$\mathbf{G}_n \stackrel{d}{=} \sum_{r \geq 1} A^{(r)} \, \mathbf{G}_{n-1}^{(r)} \,, \tag{3.10}$$

for $\mathbf{A}$, $\mathbf{G}_{n-1}^{(r)}$, $r \geq 1$, independent and $\mathcal{L}(\mathbf{G}_{n-1}^{(r)}) = \mathcal{L}(\mathbf{G}_{n-1})$. Then $\mathbf{G}_n$ has marginals distributed as $G_{nk+\ell}^{[1]}[\mu], \ldots, G_{nk+\ell}^{[d]}[\mu]$ respectively. If $\mathbf{A} \in \mathcal{A}_2^*$, then the marginals of $\mathbf{G}_n$ are furthermore independent.

For a normalized version of $\mathbf{G}_n$ a bound on the moment generating function is obtained:

Proposition 3.3.3 *Given $\mu \in \mathcal{O}$ bounded, $k \in \mathbb{N}$, $\ell \in \mathbb{N}_0$, let $\mathbf{G}_n$, $\mathbf{A}$ be as in Corollary 3.3.2, $\varrho = \varrho[\mu] > 1$ and denote $\mathbf{Y}_n := (\mathbf{G}_n - \mathbb{E}\,\mathbf{G}_n)/\varrho^{nk+\ell}$. Then for* *every* $1 < q \leq 2$ *satisfying* $\operatorname{ess\,sup} \varrho^{-kq} \sum_{r \geq 1} \|A^{(r)T} A^{(r)}\|_{\mathrm{op}}^{q/2} < 1$, *there is a constant* $K_q > 0$ *with*

$$\mathbb{E}\ \exp\langle \mathbf{s}, \mathbf{Y}_n \rangle \leq \exp(K_q \|\mathbf{s}\|^q) \tag{3.11}$$

for all $\mathbf{s} \in \mathbb{R}^d$ *and* $n \geq 0$.

For fixed $1 < q \leq 2$ this proposition yields the tail bound inequalities stated in Theorem 3.2.1 for $\kappa = 1 + 1/(q-1)$, as long as it exists $k \in \mathbb{N}$, $\mathbf{A} \in \mathcal{A}^*$ with $T(\mathbf{A}) = \mu_k$ and $\operatorname{ess\,sup} \varrho^{-kq} \sum_{r \geq 1} \|A^{(r)T} A^{(r)}\|_{\mathrm{op}}^{q/2} < 1$. This is verified in the proof Theorem 3.2.1 (see p. 74), where also constant L_κ is specified. This explains the construction of $q^*(\mu)$ (see (3.3)) and hence upper bound κ^* (see Theorem 3.2.1).

3.4 Lower bound on κ^*

Corollary 3.4.2 stated below is a weaker tail bound result than Theorem 3.2.1, i.e. $1 + 1/(\ln(\Delta)/\ln(\varrho) - 1) < \kappa^*$. The use of this result is that this upper bound on the exponents is much easier yielded by the offspring distribution (see (3.12)) than κ^*.

If offspring distribution $\mu \in \mathcal{O}$ is bounded, then

$$\left\{ B_{i,j}^{(r)} = \mathbf{1}[X^{[i]}(j) \geq 1] \middle| \, 1 \leq r \leq \operatorname{ess\,sup} X^{[i]}(j) < \infty, 1 \leq i,j \leq d \right\},$$

with $\mathbf{X}^{[1]}, \ldots, \mathbf{X}^{[d]}$ independent, is a finite set of Bernoulli-distributed random variables, satisfying (3.6). Given this set, it obviously exists a finite sequence $\mathbf{A} = (A^{(s)})_{s \geq 1}$ of random matrices where for every $s \geq 1$ there are $1 \leq i,j \leq d$, $r \geq 1$, with $A_{i,j}^{(s)} = B_{i,j}^{(r)}$ and $A_{i',j'}^{(s)} = 0$ for $(i',j') \neq (i,j)$. Hence it is $\mathbf{A} \in \mathcal{A}^*$ and $T(\mathbf{A}) = \mu$. Thus $T^{-1}(\mu) \neq \emptyset$ for every bounded μ. Even more, there is a vector $\mathbf{A} \in T^{-1}(\mu)$ of which we know the length precisely and which yields an upper bound on $q^*(\mu)$:

Proposition 3.4.1 *Given a d-type Galton-Watson process with finite maximum family size,* $\mathbf{X}^{[i]} = \mathbf{X}^{[i]}[\mu]$, *denote*

$$\Delta = \Delta(\mu) = \max_{1 \leq i \leq d} \sum_{j=1}^{d} \operatorname{ess\,sup} X^{[i]}(j) \ \vee \ \max_{1 \leq j \leq d} \sum_{i=1}^{d} \operatorname{ess\,sup} X^{[i]}(j). \tag{3.12}$$

Then it exists $\mathbf{A} = (A^{(1)}, \ldots, A^{(\Delta)}) \in \mathcal{A}_1^* \cap \mathcal{A}_2^* \subseteq \mathcal{A}^*$ *with* $T(A) = \mu$.

If $\mathbf{A} = (A^{(1)}, \ldots, A^{(\Delta)}) \in \mathcal{A}_1^* \cap \mathcal{A}_2^*$, then (3.4) and the definition of $\mathcal{A}_1^*$ and $\mathcal{A}_2^*$ yield $\|A^{(r)T}A^{(r)}\|_{\text{op}} \in \{0,1\}$ a.s. for $1 \leq r \leq \Delta$ (which is equivalent to $A^{(r)}$ having in every row and in every column at most on nonzero entry a.s.). Thus

$$\sum_{r=1}^{\Delta} \left\| A^{(r)T} A^{(r)} \right\|_{\text{op}} \leq \Delta \quad \text{a.s.}$$

So, this proposition leads to:

Corollary 3.4.2 *Let μ, $G_n^{[i]}$, ϱ be as in Theorem 3.2.1 and Δ as in Proposition 3.4.1. If $\log\Delta/\log\varrho < 2$, then for every $2 \leq \kappa < 1 + 1/(\log\Delta/\log\varrho - 1)$ there exists an $L_\kappa > 0$ such that for every $n \geq 0$ and $1 \leq i \leq d$*

$$\begin{aligned} \mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} > t\right) &\leq \exp(-L_\kappa t^\kappa) \quad t > 0, \\ \mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} < -t\right) &\leq \exp(-L_\kappa t^\kappa) \quad 0 < t \leq \frac{\mathbb{E}\, G_n^{[i]}}{\varrho^n}. \end{aligned}$$

This corollary is proved analogously to Theorem 3.2.1 with Proposition 3.3.3 using Chernoff's bounding technique. The difference is that in the proof of Corollary 3.4.2 Proposition 3.3.3 is applied on $\mathbf{A} \in \mathcal{A}_1^* \cap \mathcal{A}_2^*$ with $T(\mathbf{A}) = \mu$ and not arbitrary $\mathbf{A} \in \mathcal{A}^*$ with $T(\mathbf{A}) = \mu_k$, $k \in \mathbb{N}$. The existence of such an $\mathbf{A}$ is guarantueed by Proposition 3.4.1.

3.5 Relation to other works and a note on Karp and Zhang's process

For singletype Galton-Watson processes with finite maximum family size we have $G_n^{[1]} = Z_n$ and only one recursive description: $Z_0 = 1$ and

$$Z_n \stackrel{d}{=} Z_{n-1}^{(1)} + \cdots + Z_{n-1}^{(Z_1)},$$

with $Z_1, Z_{n-1}^{(1)}, Z_{n-1}^{(2)}, \ldots$ independent and $\mathcal{L}(Z_{n-1}^{(1)}) = \mathcal{L}(Z_{n-1}^{(2)}) = \cdots = \mathcal{L}(Z_{n-1})$. And for $G_{nk+\ell}^{[1]} = Z_{nk+\ell}$, $k \geq 1$ and $\ell \geq 0$ fix, we have only recursive description

$$Z_{nk+\ell} \stackrel{d}{=} Z_{(n-1)k+\ell}^{(1)} + \cdots + Z_{(n-1)k+\ell}^{(Z_k)},$$

with $Z_k, Z_{(n-1)k+\ell}^{(1)}, Z_{(n-1)k+\ell}^{(2)}, \ldots$ independent and $\mathcal{L}(Z_{(n-1)k+\ell}^{(1)}) = \mathcal{L}(Z_{(n-1)k+\ell}^{(2)}) = \cdots = \mathcal{L}(Z_{(n-1)k+\ell})$. For $\varrho = \mathbb{E}\, Z_1$, $m = \operatorname{ess\,sup} Z_1$ it is $\mathbb{E}\, Z_k = \varrho^k$ and $\operatorname{ess\,sup} Z_k =$

m^k. Hence we get

$$q^*(\mu) = \inf\{q \in \mathbb{R}_+ | m < \varrho^q\} = \ln(m)/\ln(\varrho)$$

and

$$\begin{aligned} \mathbb{P}\left(\frac{Z_n - \mathbb{E}\, Z_n}{\varrho^n} > t\right) &\leq \exp(-L_\kappa t^\kappa) \quad t > 0, \\ \mathbb{P}\left(\frac{Z_n - \mathbb{E}\, Z_n}{\varrho^n} < -t\right) &\leq \exp(-L_\kappa t^\kappa) \quad 0 < t \leq 1. \end{aligned}$$

for constant $L_\kappa > 0$ and $2 \leq \kappa < \kappa^* = 1 + 1/(\ln(m)/\ln(\varrho) - 1)$. Note that $1 + 1/(\ln(m)/\ln(\varrho) - 1)$ is the exponent, arising in Theorems 3.1.2 and 3.1.3.

Since in particular $(G_n^{[i]} - \mathbb{E}\, G_n^{[i]})/\varrho^n \to (G^{[i]} - \mathbb{E}\, G^{[i]})$ in distribution, as $n \to \infty$, we have for all fix t, where function $t \mapsto \mathbb{P}((G^{[i]} - \mathbb{E}\, G[i]) > t)$ is continuous that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} > t\right) = \mathbb{P}\left(G^{[i]} - \mathbb{E}\, G^{[i]} > t\right).$$

Hence Theorem 3.1.5 implies for such t, for $\tilde{\kappa}$ and $F^{[i]}$ as in that theorem and some positive constants $\widetilde{L}_0$, $\widetilde{L}_1$

$$\exp\left(-\widetilde{L}_0 t^{\tilde{\kappa}}\right) \leq \lim_{n\to\infty} \mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} > t\right) \leq \exp\left(-\widetilde{L}_1 t^{\tilde{\kappa}}\right). \tag{3.13}$$

This is true, because the functions $F^{[i]}$ are continuous and multiplicatively periodic by assumption, hence bounded. The first inequality is even more true for all fix t, as can be shown by a.s. convergence of $(G_n^{[i]} - \mathbb{E}\, G_n^{[i]})/\varrho^n$ and Fatou's Lemma. Thus, whenever a multitype Galton-Watson process fulfills the conditions of Theorem 3.2.1 and Theorem 3.1.5, we have

$$\kappa^* \leq \tilde{\kappa}. \tag{3.14}$$

Then, an advantage of Theorem 3.1.5 over Theorem 3.2.1 is that it yields the exact exponent $\tilde{\kappa}$ and not an upper bound κ^*. On the other hand an advantage of Theorem 3.2.1 over Theorem 3.1.5 is that it yields upper tail bounds for all $n \geq 0$, not only for the limit. In particular, Theorem 3.2.1 yields an estimate for $\mathbb{P}((G_n^{[i]} - \mathbb{E}\, G_n^{[i]})/\varrho^n > t)$, where t depends on n.

Karp and Zhang's process is a positive regular, supercritical 2-type Galton-Watson process with finite maximum family size, which fulfills the conditions of both theorems: Let μ be its offspring distribution (defined in subsection 1.2.3) and $\mathbf{A}$ be the sequence of random matrices given by (3.2). In section 1.2 we derived elaborately, what can be summed up in this chapter's notation as follows: It is $\varrho = \varrho(\mu) = (17+\sqrt{33})/8 \doteq 2.84$, $T(\mathbf{A}) = \mu$ and $q(\mathbf{A}) = \ln(4)/\ln(\varrho) \doteq 1.33$. Hence $q^*(\mu) \leq q(\mathbf{A}) < 2$ and Theorem 3.2.1 is applicable. On the other hand one can calculate easily

$$\begin{aligned} J^{[1]} &= \{(2,0),(2,1),(3,1)\}, \quad U^{[1]} = \{(3,1)\}, \\ J^{[2]} &= \{(0,2),(1,2),(2,2)\}, \quad U^{[2]} = \{(2,2)\}, \\ U &= \left\{ \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \right\}. \end{aligned}$$

Cf. p. 58 for definition of these quantities. Since $U = \{C\}$, λ is the largest eigenvalue of C, hence $\lambda = 4$, the corresponding eigenspace is $\{t(1,1)|t \in \mathbb{R}\}$, $\varrho < \lambda$ and $\lim_{n\to\infty} \lambda^{-n}C^n$ exists. Thus Jones' tail bound result is applicable and yields

$$\mathbb{P}\left(Y^{[i]} > t\right) = \exp\left(-t^{\tilde{\kappa}}\left(F^{[i]}(t+u_i) + o(1)\right)\right), \quad t \to \infty,$$

where $F(t) = (F^{[1]}(t), \ldots, F^{[d]}(t))$ has period $\lambda/\varrho = 32/(17+\sqrt{33})$ and exponent

$$\tilde{\kappa} = 1 + \frac{1}{\log 4/\log \varrho - 1} = 1 + \frac{1}{q(\mathbf{A}) - 1}.$$

Hence with (3.14) and $q^*(\mu) \leq q(\mathbf{A})$ we have

$$\kappa^* \leq \tilde{\kappa} = 1 + \frac{1}{q(\mathbf{A}) - 1} \leq \kappa^*$$

and thus $q(\mathbf{A}) = q^*(\mu)$. This yields that we used a best possible recursive description of $G_n^{[1]}$ and $G_n^{[2]}$ for our analysis and even more $\kappa^* = \tilde{\kappa}$. Hence the first inequality of (3.13) implies

$$\forall\, \kappa > \kappa^*, L > 0\, \exists\, t', n' : \mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} > t\right) > \exp(-Lt^{\kappa}) \quad \forall\, t > t', n > n'.$$

So, we cannot improve upon upper bound κ^*, as already stated in subsection 1.2.2.

3.6 Galton-Watson processes with Immigration

Let $\mu^{[1]}, \ldots, \mu^{[d]}, \nu$ be probability distributions on $\mathbb{N}_0^d$, $\mu = (\mu^{[1]}, \ldots, \mu^{[d]})$. A d-type Galton-Watson process with immigration $\mathbf{Z}_n^{[i]}[\mu,\nu] = (Z_n^{[i]}[\mu,\nu](1), \ldots, Z_n^{[i]}[\mu,\nu](d))$,

$n \geq 0$, is a Markov chain on $\mathbb{N}_0^d$ with $\mathbf{Z}_0^{[i]}[\mu,\nu] = \mathbf{e}_i$ and

$$\mathbf{Z}_n^{[i]}[\mu,\nu] = \sum_{r=1}^{Z_{n-1}^{[i]}(1)} \mathbf{X}^{[1],(r,n)} + \cdots + \sum_{r=1}^{Z_{n-1}^{[i]}(d)} \mathbf{X}^{[d],(r,n)} + \mathbf{V}^{(n)}, \quad n \geq 1,$$

where $(Z_{n-1}^{[i]}(1),\ldots,Z_{n-1}^{[i]}(d)) = \mathbf{Z}_{n-1}^{[i]}[\mu,\nu]$, $\mathbf{X}^{[j],(r,n)}$, $\mathbf{V}^{(n)}$, $1 \leq j \leq d$, $r \geq 1$, $n \geq 1$, are independent $\mathcal{L}(\mathbf{X}^{[j],(r,n)}) = \mu^{[j]}$ and $\mathcal{L}(\mathbf{V}^{(n)}) = \nu$. So, $\mathbf{Z}_n^{[i]}[\mu,\nu] = (Z_n^{[i]}[\mu,\nu](1),\ldots,Z_n^{[i]}[\mu,\nu](d))$ can be interpreted as the vector of numbers of individuals of type $1,\ldots,d$ at time n. At time 0 there is only a single type-i individual and during a time step every type-k individual, $1 \leq k \leq d$, splits idenpendently of the other individuals, into a random number of individuals of any type, according to distribution $\mu^{[k]}$ and additionally, independently a random number of individuals of any type immigrate according to immigration distribution ν. As before, μ is called offspring distribution and furthermore ν is called immigration distribution.

A d-type Galton-Watson process with immigration can be characterized by a $(d+1)$-type Galton-Watson process (without immigration) as follows: Individuals of type $1,\ldots,d$ split as before and additionally there is a type-$(d+1)$ individual, which splits in every time unit into a single type-$(d+1)$ individual and individuals of type $1,\ldots,d$ according to the immigration-distribution. I.e. the immigration in every generation is generated by a type-$(d+1)$ individual. Formally, for $\tilde{\mu} = (\tilde{\mu}^{[1]},\ldots,\tilde{\mu}^{[d+1]})$ defined by the product measures

$$\begin{aligned} \tilde{\mu}^{[i]} &= \mu^{[i]} \otimes \delta_0, \quad \text{for } 1 \leq i \leq d \\ \tilde{\mu}^{[d+1]} &= \nu \otimes \delta_1. \end{aligned} \tag{3.15}$$

we have

$$\left(Z_n^{[i]}[\mu,\nu](1),\ldots,Z_n^{[i]}[\mu,\nu](d),0\right) \stackrel{d}{=} \mathbf{Z}_n^{[i]}[\tilde{\mu}] + \mathbf{Z}_n^{[d+1]}[\tilde{\mu}] - \mathbf{e}_{d+1}, \tag{3.16}$$

for $\mathbf{Z}_n^{[i]}[\tilde{\mu}]$, $\mathbf{Z}_n^{[d+1]}[\tilde{\mu}]$ independent. Based on this relationship for

$$\tilde{q}(\tilde{\mu}) = \inf\left\{ q(\mathbf{A}) \middle| \exists\, k \in \mathbb{N} : \mathbf{A} \in T^{-1}(\tilde{\mu}_k) \cap \mathcal{A}_2^* \right\}.$$

we get from Proposition 3.3.3 the following tail bound result for generation size

$$G_n^{[i]}[\mu,\nu] = \sum_{j=1}^{d} Z_n^{[i]}(j)[\mu,\nu].$$

Theorem 3.6.1 *Let $G_n^{[i]}[\mu,\nu]$ be the generation size of a d-type Galton-Watson process with immigration, where μ and ν are bounded, $\varrho[\mu] > 1$ and $\tilde{\mu} = (\tilde{\mu}^{[1]}, \ldots, \tilde{\mu}^{[d+1]})$ defined by (3.15). If $\tilde{q}(\tilde{\mu}) < 2$, then for every $2 \leq \kappa < \tilde{\kappa}(\tilde{\mu}) = 1 + 1/(\tilde{q}(\tilde{\mu}) - 1)$ there exists an $L_\kappa > 0$ such that for every $n \geq 0$ and $1 \leq i \leq d$*

$$\begin{aligned}
\mathbb{P}\left(\frac{G_n^{[i]}[\mu,\nu] - \mathbb{E}\, G_n^{[i]}[\mu,\nu]}{\varrho[\mu]^n} > t\right) &\leq \exp(-L_\kappa t^\kappa) \quad t > 0,\\
\mathbb{P}\left(\frac{G_n^{[i]}[\mu,\nu] - \mathbb{E}\, G_n^{[i]}[\mu,\nu]}{\varrho[\mu]^n} < -t\right) &\leq \exp(-L_\kappa t^\kappa) \quad 0 < t \leq \frac{\mathbb{E}\, G_n^{[i]}}{\varrho^n}.
\end{aligned}$$

$\mathbf{Z}_n^{[i]}[\tilde{\mu}]$ is not a positive regular process, as will be shown in the proof of this corollary (cf. p. 78). Hence we can use the idea of charactarizing a $\mathbf{Z}_n^{[i]}[\mu,\nu]$ by $\mathbf{Z}_n^{[i]}[\tilde{\mu}]$ just, because we do not claim positive regularity in Proposition 3.3.3.

3.7 Proofs

Proof of Proposition 3.3.1: The assertion is proved by induction on $n \geq 0$: For $n = 0$ the assertion is true by definition of $\mathbf{G}_0$. Now, if the i-th component of $\mathbf{G}_{n-1}$ is distributed as $G_{n-1}^{[i]}[\mu]$, for every $1 \leq i \leq d$, then distributional equation (3.8) yields that the i-th component of $\mathbf{G}_n$ is distributed as

$$\sum_{r\geq 1} A_{i,1}^{(r)}\, \hat{G}_{n-1}^{[1],(r)} + \cdots + \sum_{r\geq 1} A_{i,d}^{(r)}\, \hat{G}_{n-1}^{[d],(r)}\,, \tag{3.17}$$

with $\mathcal{L}(\hat{G}_{n-1}^{[j],(r)}) = \mathcal{L}(G_{n-1}^{[j]}[\mu])$ and $\mathbf{A}$, $\{\hat{G}_{n-1}^{[j],(r)} \,|\, 1 \leq j \leq d\}$, $r \geq 1$, independent. According to (3.7) term (3.17) would be distributed as $G_n^{[i]}[\mu]$, if furthermore $\hat{G}_{n-1}^{[j],(r)}$,$1 \leq j \leq d$ were independent. Indeed it suffices to show that $\{G_{n-1}^{[j],(r)} \,|\, A_{i,j}^{(r)} = 1,\, 1 \leq j \leq d\}$ consists of independent random variables for a.e. realization of $\mathbf{A}$, because this yields that in (3.17) almost all $\hat{G}_{n-1}^{[j],(r)}$, which are not multiplied by 0, are independent. If $\mathbf{A} \in \mathcal{A}_1^*$ then $\#\{A_{i,j}^{(r)} = 1 \,|\, 1 \leq j \leq d\} \leq 1$ a.s. and hence the assertion is true.
For $\mathbf{A} \in \mathcal{A}_2^*$, we have to prove in addition inductively that $\mathbf{G}_n$ has independent components: For $n = 0$ this is obvious. Assume now that $\mathbf{G}_{n-1}$ has components, which are independent and distributed as $G_{n-1}^{[1]}[\mu], \ldots, G_{n-1}^{[d]}[\mu]$, respectively. Then, equation (3.8) yields that the i-th component of $\mathbf{G}_n$ is distributed as (3.17), where now moreover, because of independent components, $\mathbf{A}$, $\hat{G}_{n-1}^{[j],(r)}$, $1 \leq j \leq d$, $r \geq 1$ are independent. As explained above, this proves that the i-th component

of $\mathbf{G}_n$ is distributed as $G_n^{[i]}[\mu]$. $\mathbf{A}_{1,\cdot}, \dots, \mathbf{A}_{d,\cdot}$ are independent for $\mathbf{A} \in \mathcal{A}_2^*$ and $A_{i,j}^{(r)} \hat{G}_{n-1}^{[j],(r)} = \hat{G}_{n-1}^{[j],(r)}$ implies $A_{i',j}^{(r)} \hat{G}_{n-1}^{[j],(r)} = 0$ a.s., for $i \neq i'$, $1 \leq i, j \leq d$, $r \geq 1$, since $\sum_i A_{i,j} \leq 1$ a.s. for $1 \leq j \leq d$. Hence (3.17) for $1 \leq i \leq d$ yields that $\mathbf{G}_n$ has independent components. This completes the proof. ■

Proof of Corollary 3.3.2: Corollary 3.3.2 is proved with distributional recursion (3.9) by using the same argumentation, as given in section 3.3 and the previous proof, in order to derive Proposition 3.3.1 from equation (3.5). ■

Proof of Proposition 3.3.3: The distributional recurrence (3.10) for $\mathbf{G}_n$ implies the relation

$$\mathbf{Y}_n \stackrel{d}{=} \sum_{r \geq 1} (A^{(r)}/\varrho^k) \mathbf{Y}_{n-1}^{(r)} + \mathbf{b}_n, \quad n \geq 1, \tag{3.18}$$

with $\mathbf{Y}_{n-1}^{(r)} = (\mathbf{G}_{n-1}^{(r)} - \mathbb{E}\, \mathbf{G}_{n-1}^{(r)})/\varrho^{(n-1)k+\ell}$, and

$$\mathbf{b}_n = \frac{1}{\varrho^{nk+\ell}} \mathbb{E}\, \mathbf{G}_{n-1} \sum_{r \geq 1} A^{(r)} - \frac{1}{\varrho^{nk+\ell}} \mathbb{E} \mathbf{G}_n.$$

We prove the assertion by induction on n. For $n = 0$ it suffices to prove the assertion for $q = 1$ and $q = 2$, since $1 < q \leq 2$ by assumption. By Corollary 3.3.2 the i-th component of $\mathbf{Y}_0$, denoted by $Y(i)$, is distributed as $(G_\ell^{[i]} - \mathbb{E}\, G_\ell^{[i]})/\varrho^\ell$ and hence $\operatorname{ess\,sup} \|\mathbf{Y}_0\| =: c_1 < \infty$ and $\operatorname{ess\,sup} \max_i |Y(i)| =: c_2 < \infty$, since μ is bounded. Thus we have for every $\mathbf{s} = (s_1, \dots, s_d) \in \mathbb{R}^d$

$$\mathbb{E} \exp\langle \mathbf{s}, \mathbf{Y}_0 \rangle \leq \mathbb{E} \exp(\|\mathbf{s}\| \|\mathbf{Y}_0\|) \leq \exp\left(c_1 \|\mathbf{s}\|^1\right).$$

By induction on d we get $\mathbb{E}\, |W_1 \dots W_d| \leq \prod_{i=1}^d \left(\mathbb{E}\, |W_i|^d\right)^{1/d}$ for random variables $W_1 \dots W_d$ with finite d-th moment:

$$\begin{aligned}
\mathbb{E} \left| \prod_{i=1}^d W_i \right| &\leq \left(\mathbb{E} \prod_{i=1}^{d-1} |W_i|^{d/(d-1)} \right)^{(d-1)/d} \left(\mathbb{E}\, |W_d|^d \right)^{1/d} \\
&\leq \left(\prod_{i=1}^{d-1} \left(\mathbb{E} \left(|W_i|^{d/(d-1)} \right)^{d-1} \right)^{1/(d-1)} \right)^{(d-1)/d} \left(\mathbb{E}\, |W_d|^d \right)^{1/d} \\
&= \prod_{i=1}^d \left(\mathbb{E}\, |W_i|^d \right)^{1/d},
\end{aligned}$$

using Hölder's inequality for the first and induction hypothesis for $|W_1|^{d/(d-1)}, \ldots, |W_{d-1}|^{d/(d-1)}$ for the second inequality. This yields

$$\begin{aligned}\mathbb{E}\,\exp\langle \mathbf{s}, \mathbf{Y}_0\rangle &= \mathbb{E}\prod_{i=1}^{d}\exp(s_i Y(i)) \le \prod_{i=1}^{d}(\,\mathbb{E}\,\exp(s_i d Y(i)))^{1/d} \\ &\le \prod_{i=1}^{d}\left(\exp\left(s_i^2\frac{d^2c_2^2}{2}\right)\right)^{1/d} = \exp\left(\frac{d\,c_2^2}{2}\|s\|^2\right),\end{aligned}$$

where the second inequality is given by Hoeffding's inequality (see Lemma 2.3.4). This proves the induction hypothesis for $K_q = K = c_1 \vee d\,c_2^2/2$, for every $1 < q \le 2$. Assume the assertion is true for $n-1$. Then, conditioning on $(\mathbf{b}_n, \mathbf{A})$, denoting the distribution of this vector by σ_n, and using the induction hypothesis, we obtain

$$\begin{aligned}&\mathbb{E}\,\exp\langle \mathbf{s}, \mathbf{Y}_n\rangle \\ &= \int \exp\langle \mathbf{s}, \mathbf{b}_n\rangle \prod_{r\ge 1}\mathbb{E}\,\exp\langle \mathbf{s}, (a^{(r)}/\varrho^k)\mathbf{Y}_{n-1}\rangle d\sigma_n(\mathbf{b}_n, a^{(1)}, a^{(2)}, \ldots) \\ &\le \int \exp\langle \mathbf{s}, \mathbf{b}_n\rangle \prod_{r\ge 1}\exp(K_q(\|(a^{(r)})^T\mathbf{s}\|/\varrho^k)^q) d\sigma_n(\mathbf{b}_n, a^{(1)}, a^{(2)}, \ldots) \\ &\le \int \exp\left(\langle \mathbf{s}, \mathbf{b}_n\rangle + K_q\|\mathbf{s}\|^q\sum_{r\ge 1}(\|a^{(r)T}a(r)\|_{\text{op}}^{1/2}/\varrho^k)^q\right) d\sigma_n(\mathbf{b}_n, a^{(1)}, a^{(2)}, \ldots) \\ &= \mathbb{E}\,\exp(\langle \mathbf{s}, \mathbf{b}_n\rangle + K_q\|\mathbf{s}\|^q U)\exp(K_q\|\mathbf{s}\|^q),\end{aligned}$$

with $U := \sum_{r\ge 1}\left(\|A^{(r)T}A^{(r)}\|_{\text{op}}^{1/2}/\varrho^k\right)^q - 1$. Hence, the proof is completed by showing

$$\sup_{k\ge 1}\mathbb{E}\,\exp(\langle \mathbf{s}, \mathbf{b}_n\rangle + K_q\|\mathbf{s}\|^q U) \le 1,$$

for some appropriate K_q. We denote $\xi = -\operatorname{ess\,sup} U$ and have $\xi > 0$ by assumption.

Small $\|\mathbf{s}\|$: First we consider small $\|\mathbf{s}\|$ with $\|\mathbf{s}\| \le c/\sup_{n\ge 0}\|\mathbf{b}_n\|_{2,\infty}$ for some $c > 0$, where $\|\mathbf{b}_n\|_{2,\infty} = \|\,\|\mathbf{b}_n\|\,\|_\infty$, Note that $\sup_{n\ge 0}\|\mathbf{b}_n\|_{2,\infty} < \infty$, since $\sup_n \mathbb{E}\,G_{nk+\ell}^{[i]}/\varrho^{nk+\ell} < \infty$. For these small $\|\mathbf{s}\|$ we have

$$\mathbb{E}\,\exp((\langle \mathbf{s}, \mathbf{b}_n\rangle + K_q\|\mathbf{s}\|^q U) \le \exp(-K_q\|\mathbf{s}\|^q\xi)\mathbb{E}\,\exp\langle \mathbf{s}, \mathbf{b}_n\rangle$$

and, with $\mathbb{E}\langle \mathbf{s}, \mathbf{b}_n\rangle = 0$,

$$\begin{aligned}
\mathbb{E}\exp\langle \mathbf{s}, \mathbf{b}_n\rangle &= \mathbb{E}\left[1 + \langle \mathbf{s}, \mathbf{b}_n\rangle + \sum_{m=2}^{\infty} \frac{\langle \mathbf{s}, \mathbf{b}_n\rangle^m}{m!}\right] \\
&= 1 + \mathbb{E}\langle \mathbf{s}, \mathbf{b}_n\rangle^2 \sum_{m=2}^{\infty} \frac{\langle \mathbf{s}, \mathbf{b}_n\rangle^{m-2}}{m!} \\
&\leq 1 + \|\mathbf{s}\|^2 \mathbb{E}\|\mathbf{b}_n\|^2 \sum_{m=2}^{\infty} \frac{c^{m-2}}{m!} \\
&= 1 + \|\mathbf{s}\|^2 \mathbb{E}\|\mathbf{b}_n\|^2 \frac{e^c - 1 - c}{c^2}.
\end{aligned}$$

Using $\exp(-K_q\|\mathbf{s}\|^q\xi) \leq 1/(1 + K_q\|\mathbf{s}\|^q\xi)$ and with $\Psi(c) = (e^c - 1 - c)/c^2$ we obtain

$$\mathbb{E}\exp(\langle \mathbf{s}, \mathbf{b}_n\rangle + K_q\|\mathbf{s}\|^q U) \leq \frac{1 + \|\mathbf{s}\|^2\mathbb{E}\|\mathbf{b}_n\|^2\Psi(c)}{1 + K_q\|\mathbf{s}\|^q\xi}.$$

Hence, we have to choose K_q with

$$K_q \geq \frac{\|\mathbf{s}\|^{2-q}\Psi(c)}{\xi} \sup_{n\geq 0} \mathbb{E}\|\mathbf{b}_n\|^2.$$

The right hand side is increasing in $\|\mathbf{s}\|$ for $q \leq 2$, so with $\|\mathbf{s}\| \leq c/\sup_{n\geq 0}\|\mathbf{b}_n\|_{2,\infty}$ a possible choice is

$$K_q = \frac{\sup_{n\geq 0}\mathbb{E}\|\mathbf{b}_n\|^2}{\sup_{n\geq 0}\|\mathbf{b}_n\|_{2,\infty}^{2-q}} \frac{\Psi_q(c)}{\xi} \vee K, \tag{3.19}$$

with $\Psi_q(c) = (e^c - 1 - c)/c^q$.

Large $\|\mathbf{s}\|$: For general $\mathbf{s} \in \mathbb{R}^d$ we have

$$\langle \mathbf{s}, \mathbf{b}_n\rangle + K_q\|\mathbf{s}\|^q U \leq \|\mathbf{s}\|\|\mathbf{b}_n\| - \|\mathbf{s}\|^q K_q\xi \leq \|\mathbf{s}\|\|\mathbf{b}_n\|_{2,\infty} - \|\mathbf{s}\|^q K_q\xi,$$

and this is less than zero if

$$\|\mathbf{s}\|^{q-1} \geq \frac{\sup_{n\geq 0}\|\mathbf{b}_n\|_{2,\infty}}{K_q\xi} = \frac{\sup_{n\geq 0}\|\mathbf{b}_n\|_{2,\infty}^{3-q}}{\sup_{n\geq 0}\mathbb{E}\|\mathbf{b}_n\|^2\Psi_q(c)}.$$

If $\|\mathbf{s}\|$ satisfies the latter inequality we call it large. Thus, for large $\|\mathbf{s}\|$ we have $\sup_{n\geq 0}\mathbb{E}\exp(\langle \mathbf{s}, \mathbf{b}_n\rangle + K_q\|\mathbf{s}\|^q U) \leq 1$.

In order to overlap the regions for small and large $\|\mathbf{s}\|$ we need

$$\Psi_1(c) \geq \frac{\sup_{n\geq 0}\|\mathbf{b}_n\|_{2,\infty}^2}{\sup_{n\geq 0}\mathbb{E}\|\mathbf{b}_n\|^2}.$$

Assume w.l.o.g. that $(\mathbf{G}_n)_{n\in\mathbb{N}_0}$ is not a.s. deterministic. This yields $\mathbb{P}(\mathbf{b}_n = (0,\dots,0)) < 1, n \geq 1$, because $\mathbf{b}_n = (0,\dots,0)$ a.s. implies $\mathbb{E}\,\mathbf{G}_n = (1/\varrho^n)\sum_{r\geq 1} A^{(r)}\,\mathbb{E}\,\mathbf{G}_{n-1}$. Hence, it is $\sup_{n\geq 1}\mathbb{E}\,\|\mathbf{b}_n\|^2 > 0$ and so the right hand side of the latter display is finite. Because $\lim_{c\to\infty}\Psi_1(c) = \infty$ there exists a c for which the inequality is true and the proof is completed. ■

Proof of Theorem 3.2.1: For every bounded μ it can be verified easily that $\varrho[\mu_k] = \varrho[\mu]^k$. Thus by definition for every $q > q^*(\mu)$ there is a $k \in \mathbb{N}$, $\mathbf{A} \in \mathcal{A}^*$ with $T(\mathbf{A}) = \mu_k$ and

$$\operatorname{ess\,sup} \varrho^{-kq} \sum_{r\geq 1} \|A^{(r)T} A^{(r)}\|_{\mathrm{op}}^{q/2} < 1.$$

For fixed $0 \leq \ell \leq k-1$ let $\mathbf{Y}_n$ be as in Proposition 3.3.3, $Y_n^{[i]} = (G_n^{[i]} - \mathbb{E}\,G_n^{[i]})/\varrho^n$ and $\mathbf{e}_i \in \mathbb{R}^d$ the i-th unit vector. Then Corollary 3.3.2 and Proposition 3.3.3 yield for $K_q(\ell) = K_q$

$$\begin{aligned}
\mathbb{P}\left(\frac{G_{nk+\ell}^{[i]} - \mathbb{E}\,G_{nk+\ell}^{[i]}}{\varrho^{nk+\ell}} > t\right) &= \mathbb{P}(\exp(uY_{nk+\ell}^{[i]}) > \exp(ut))) \\
&\leq \mathbb{E}\,\exp(uY_{nk+\ell}^{[i]} - ut) \\
&= \mathbb{E}\,\exp(\langle u\mathbf{e}_i, \mathbf{Y}_n\rangle - ut) \\
&\leq \exp(K_q(\ell)\,u^q - ut).
\end{aligned}$$

Minimizing over $u > 0$ we obtain the bound

$$\mathbb{P}\left(\frac{G_{nk+\ell}^{[i]} - \mathbb{E}\,G_{nk+\ell}^{[i]}}{\varrho^{nk+\ell}} > t\right) \leq \exp(-L_\kappa(\ell)t^\kappa), \quad n \geq 0$$

for $\kappa = 1 + 1/(q-1)$ and $L_\kappa(\ell) = K_{\kappa/(\kappa-1)}^{1-\kappa}(\ell)\,(\kappa-1)^{\kappa-1}/\kappa^\kappa$. Hence it is

$$\mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\,G_n^{[i]}}{\varrho^n} > t\right) \leq \exp(-L_\kappa t^\kappa), \quad n \geq 0$$

for $L_\kappa = \min_{0\leq\ell<k} L_\kappa(\ell)$. The same bound applies to the left tail. This completes the proof. ■

Proof of Proposition 3.4.1: For better legibility, we denote for any $(d\times d)$-matrix its (i,j)-th component by $M(i,j)$. For deterministic matrix $M \in \mathbb{N}_0^{d\times d}$ we define furthermore

$$\begin{aligned}
r(M,i) &= \sum_{j=1}^{d} M(i,j), \qquad \text{for } 1 \le i \le d,\\
c(M,j) &= \sum_{i=1}^{d} M(i,j), \qquad \text{for } 1 \le j \le d\\
s(M) &= \max_{1\le i\le d} r(M,i) \vee \max_{1\le j\le d} c(M,j).
\end{aligned}$$

So, by definition it is

$$s\left(\left[\operatorname{ess\,sup} X^{[i]}(j)\right]_{1\le i,j\le d}\right) = \Delta.$$

We are going to prove the following assertion:

Lemma 3.7.1 *Every deterministic matrix* $0 \neq M \in \mathbb{N}_0^{d\times d}$ *can be partitioned into two matrices* $M', B \in \mathbb{N}_0^{d\times d}$*, i.e.*

$$M' + S = M,$$

with

$$s(M') = s(M) - 1$$

and in every row and in every column of S *there is at most one entry* 1 *and the other* $d-1$ *entries are* 0*.*

Hence, in particular $[\operatorname{ess\,sup} X^{[i]}(j)]_{1\le i,j\le d}$ can be sucessively partitioned into matrices $S^{(1)}, \dots, S^{(\Delta)} \in \mathbb{N}_0^{d\times d}$, i.e.

$$S^{(1)} + \cdots + S^{(\Delta)} = \left[\operatorname{ess\,sup} X^{[i]}(j)\right]_{1\le i,j\le d},$$

where in every row and in every column of $S^{(r)}$, $1 \le r \le \Delta$, there is at most one entry 1 and the other $d-1$ entries are 0. Given $S^{(1)}, \dots, S^{(\Delta)}$, we define the vector of random matrices $\mathbf{A} = (A^{(1)}, \dots, A^{(\Delta)})$ by

$$A^{(r)}(i,j) = \begin{cases} \mathbf{1}[(X^{[i]}(j)) \ge \sum_{\ell=1}^{r} S^{(\ell)}(i,j)] & \text{if } S^{(r)}(i,j) = 1,\\ 0 & \text{if } S^{(r)}(i,j) = 0,\end{cases}$$

for $1 \le i,j \le d$, $1 \le r \le \Delta$. It is easy to check that $\mathbf{A} \in \mathcal{A}_1^* \cap \mathcal{A}_2^*$ and $T(\mathbf{A}) = \mu$. Thus, the proof Proposition 3.4.1 is completed by proving Lemma 3.7.1. ■

Proof of Lemma 3.7.1: Let $M \in \mathbb{N}_0^{d\times d}$ be fix and cut short $r(M,i) = r(i)$, $c(M,j) = c(j)$ and $s(M) = s$. If it exists a permutation π on $\{1,\dots d\}$ with $M(i,\pi(i)) > 0$, for every row i with $r(i) = s$ and $M(\pi^{-1}(j),j) > 0$ for every column j with $c(j) = s$, then S given by

$$S(i,j) = \begin{cases} 1 & \text{if } j = \pi(i) \text{ and } M(i,j) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

has the properties claimed in Lemma 3.7.1 and furthermore $M' = M - B \in \mathbb{N}_0^{d\times d}$ fulfills

$$s(M') = s - 1,$$

as required. In order to prove the existence of such a permutation π, let $\tilde{\pi}$ be an arbitrary permutation on $\{1,\dots,d\}$. We will show that if $\tilde{\pi}$ is not a possible choice of π, then one can construct π successively from $\tilde{\pi}$: Note that definition of s yields for every set $B \subseteq \{1,\dots,d\}$, with $\ell = \#B$, we have

$$\sum_{j\in B} c(j) \le \ell s \tag{3.20}$$

and

$$\sum_{i\in B} r(i) = \ell s \Leftrightarrow r(i) = s \quad \forall\, i \in B. \tag{3.21}$$

If it exists i_0 with $r(i_0) = s$ and $M(i_0,\tilde{\pi}(i_0)) = 0$, then we define recursively

$$\begin{aligned} B_1 &= \{i_1 |\, M(i_0,\tilde{\pi}(i_1)) > 0\}, \\ B_m &= \{i_m |\, \exists\, i_{m-1} \in B_{m-1} : M(i_{m-1},\tilde{\pi}(i_m)) > 0\}, \qquad \text{for } m > 1. \end{aligned}$$

If for all $1 \le k \le m-1$, $i_k \in B_k$, it is

$$\begin{aligned} r(i_k) &= s, & (3.22) \\ M(i_k,\tilde{\pi}(i_0)) &= 0, & (3.23) \\ M(i_k,\tilde{\pi}(i_k)) &> 0, & (3.24) \end{aligned}$$

then we have $B_{m-1} \subsetneq B_m$: Because of (3.24) it is $B_k \subseteq B_{k+1}$, for all $1 \le k \le m-1$. Assuming $B_{m-1} = B_m$ yields

$$M(i_{m-1},j) > 0 \Rightarrow \tilde{\pi}^{-1}(j) \in B_{m-1} \quad \forall\, i_{m-1} \in B_{m-1} \tag{3.25}$$

and $B_1 \subseteq B_{m-1}$ yields

$$M(i_0,j) > 0 \Rightarrow \tilde{\pi}^{-1}(j) \in B_{m-1}. \tag{3.26}$$

By (3.23) it is $i_0 \notin B_{m-1}$ and hence $\#(B_{m-1} \cup \{i_0\}) = \ell + 1$, for $\ell = \#B_{m-1}$. Thus we have

$$\begin{aligned}
(\ell+1)s &\overset{(3.21),(3.22)}{=} \sum_{i \in B_{m-1} \cup \{i_0\}} r(i) \\
&\overset{(3.25),(3.26)}{=} \sum_{i \in B_{m-1} \cup \{i_0\}} \sum_{j : \tilde{\pi}^{-1}(j) \in B_{m-1}} M(i,j) \\
&\leq \sum_{j : \tilde{\pi}^{-1}(j) \in B_{m-1}} \sum_{i=1}^{d} M(i,j) \\
&\overset{(3.20)}{\leq} \ell s,
\end{aligned}$$

which proves $B_{m-1} \subsetneq B_m$ by contradiction. Since $B_m \subseteq \{1, \ldots, d\}$, for every $m \geq 1$, there must be an $m \geq 1$ with $B_k \subsetneq B_{k+1}$, for $1 \leq k \leq m-1$ and $B_m = B_{m+1}$ or $B_m \nsubseteq B_{m+1}$. Hence it exists $i_m \in B_m$ for which condition (3.22), (3.23) or (3.24) is not valid. B_1 is not empty, since $r(i_0) = s > 0$, thus by definition of $B_1, \ldots, B_m$ there is a sequence $(i_0, \ldots, i_m)$, $i_k \in B_k$, with $M(i_{k-1}, \tilde{\pi}(i_k)) > 0$, for $1 \leq k \leq m$. Given $(i_0, \ldots, i_m)$ we define a permutation $\hat{\pi}$ by

$$\hat{\pi}(i) = \begin{cases} \tilde{\pi}(i_{k+1}) & \text{if } i = i_k, \text{ for } 0 \leq k \leq m-1, \\ \tilde{\pi}(i_0) & \text{if } i = i_m, \\ \tilde{\pi}(i) & \text{otherwise.} \end{cases}$$

Because $M(i_k, \hat{\pi}(i_k)) > 0$, for $0 \leq k \leq m-1$, $M(i, \hat{\pi}(i)) = M(i, \tilde{\pi}(i))$ for $i \notin \{i_0, \ldots, i_m\}$ and $M(i_m, \hat{\pi}(i_m)) > 0$, if $r(i_m) = s$ and $M(i_m, \tilde{\pi}(i_m)) > 0$, since (3.22), (3.23) or (3.24) is not valid for i_m, we have

$$M(i, \tilde{\pi}(i)) > 0 \quad \Rightarrow \quad M(i, \hat{\pi}(i)) > 0 \qquad \forall\, i : r(i) = s. \tag{3.27}$$

Having in addition $M(i_0, \tilde{\pi}(i_0)) = 0$ yields

$$M(\tilde{\pi}^{-1}(j), j) > 0 \quad \Rightarrow \quad M(\hat{\pi}^{-1}(j), j) > 0 \qquad \forall\, 1 \leq j \leq d \tag{3.28}$$

and furthermore with (3.27) that

$$\#\{i |\, M(i, \hat{\pi}(i)) = 0, r(i) = s\} < \#\{i |\, M(i, \tilde{\pi}(i)) = 0, r(i) = s\},$$

whereas (3.28) yields

$$\#\{j |\, M(\hat{\pi}^{-1}(j), j) = 0, c(j) = s\} \leq \#\{j |\, M(\tilde{\pi}^{-1}(j), j) = 0, c(j) = s\}.$$

Hence, successively we get a permutation π' on $\{1,\dots,d\}$ with

$$\#\{i|M(i,\pi'(i))=0, r(i)=s\}=0$$

and

$$\#\{j|\,M(\pi'^{-1}(j),j)=0, c(j)=s\} \leq \#\{j|\,M(\tilde{\pi}^{-1}(j),j)=0, c(j)=s\}.$$

Applying above argumentation on π'^{-1} and M^t, instead of $\tilde{\pi}$ and M, yields that there is a permutation π with $M(i,\pi(i)) > 0$, for all i with $r(i)=s$ and $M(\pi^{-1}(j),j) > 0)$ for all j with $c(j)=s$. This completes the proof. ■

Proof of Corollary 3.4.2: Proposition 3.4.1 yields that it exists $\mathbf{A} = (A^{(1)},\dots,A^{(\Delta)}) \in \mathcal{A}_1^* \cap \mathcal{A}_2^*$ with $T(\mathbf{A})=\mu$ and hence

$$\sum_{r=1}^{\Delta} \left\| A^{(r)T} A^{(r)} \right\|_{\text{op}} \leq \Delta \quad \text{a.s.}$$

by (3.4). So the proof is completed by applying the arguments of the proof of Theorem 3.2.1 on this particular $\mathbf{A}$. ■

Proof of Corollary 3.6.1: Since we want to apply Proposition 3.3.3 on $G_n^{[i]}[\tilde{\mu}]$, we first show $\varrho[\mu]=\varrho[\tilde{\mu}]$: Let $M \in \mathbb{N}_0^{d\times d}$ and $\tilde{M} \in \mathbb{N}_0^{(d+1)\times(d+1)}$ be the mean matrices corresponding to μ and $\tilde{\mu}$ respectively and $\mathcal{L}((V_1,\dots,V_d))=\nu$. Then it is

$$\begin{array}{lcll} \tilde{M}_{i,j} & = & M_{i,j}, & 1 \leq i,j \leq d, \\ \tilde{M}_{i,d+1} & = & 0, & 1 \leq i \leq d, \\ \tilde{M}_{d+1,j} & = & \mathbb{E}\, V_j, & 1 \leq j \leq d, \\ \tilde{M}_{d+1,d+1} & = & 1. & \end{array}$$

Denote $M_{i,j}^{(n)}$, $\tilde{M}_{i,j}^{(n)}$ the (i,j)-th component of M^n, $\tilde{M}^n$, respectively. Then we get by induction on $n \geq 1$

$$\begin{array}{lcll} \tilde{M}_{i,j}^{(n)} & = & M_{i,j}^{(n)}, & 1 \leq i,j \leq d, \\ \tilde{M}_{i,d+1}^{(n)} & = & 0, & 1 \leq i \leq d, \\ \tilde{M}_{d+1,j}^{(n)} & = & \tilde{M}_{d+1,j}^{(n-1)} + \sum_{i=1}^{d} M_{i,j}^{(n-1)}\, \mathbb{E}\, V_i, & 1 \leq j \leq d, \\ \tilde{M}_{d+1,d+1}^{(n)} & = & 1. & \end{array}$$

Thus $\mathbb{E}\, G_n^{[i]}[\mu] = \mathbb{E}\, G_n^{[i]}[\tilde{\mu}]$, for $1 \leq i \leq d$, and hence $\varrho[\mu] \leq \varrho[\tilde{\mu}]$. Furthermore, if $c_3 := \max_i \sup_n \mathbb{E}\, G_n^{[i]}[\mu]/\tilde{\varrho}^n < \infty$ for some $\tilde{\varrho} > 0$ then for

$$c_4 = \frac{c_3 \sum_{i=1}^{d} \mathbb{E}\, V_i}{\tilde{\varrho}-1} \vee c_3 \vee 1 < \infty,$$

it is $\mathbb{E}\, G_n^{[d+1]}[\tilde{\mu}]/\tilde{\varrho}^n \leq c_4$ for every $n \geq 0$, as can be proved inductively: For $n = 0$ this is trivial and assuming $\mathbb{E}\, G_{n-1}^{[d+1]}[\tilde{\mu}]/\tilde{\varrho}^{n-1} \leq c_4$, we get

$$\begin{aligned}
\mathbb{E}\, G_n^{[d+1]}[\tilde{\mu}] &= \sum_{j=1}^{d} \tilde{M}_{i,j}^{(n-1)} + \sum_{j=1}^{d}\sum_{i=1}^{d} M_{i,j}^{(n-1)}\, \mathbb{E}\, V_i + 1 \\
&= \mathbb{E}\, G_{n-1}^{[d+1]}[\tilde{\mu}] + \sum_{i=1}^{d} \mathbb{E}\, G_{n-1}^{[i]}\, \mathbb{E}\, V_i \\
&\leq c_4 \tilde{\varrho}^{n-1} + \sum_{i=1}^{d} c_3 \tilde{\varrho}^{n-1}\, \mathbb{E}\, V_i \\
&= \left(c_4 + c_3 \sum_{i=1}^{d} \mathbb{E}\, V_i \right) \tilde{\varrho}^{n-1} \\
&\leq c_4 \tilde{\varrho}^n,
\end{aligned}$$

which proves $\varrho[\mu] = \varrho[\tilde{\mu}]$.

By definition of $\tilde{q}(\tilde{\mu})$, for every $\tilde{q}(\tilde{\mu}) < q \leq 2$ it exist $k \in \mathbb{N}$, $\mathbf{A} \in \mathcal{A}_2^*$ with $T(\mathbf{A}) = \tilde{\mu}_k$ and $\operatorname{ess\,sup} \varrho^{-kq} \sum_{r\geq 1} \|A^{(r)T} A^{(r)}\|_{\mathrm{op}}^{q/2} < 1$. Hence Corollary 3.3.2 and Proposition 3.3.3 yield that vector $\mathbf{Y}_n = (\mathbf{G}_n - \mathbb{E}\,\mathbf{G}_n)/\varrho^n$, where $\mathbf{G}_n$ has independent marginals distributed as $G_{nk+\ell}^{[1]}[\tilde{\mu}], \ldots, G_{nk+\ell}^{[d]}[\tilde{\mu}]$, respectively satisfies

$$\mathbb{E}\ \exp\langle \mathbf{s}, \mathbf{Y}_n \rangle \leq \exp(K_q(\ell)\|\mathbf{s}\|^q)$$

for some constant $K_q(\ell) > 0$. Furthermore (3.16) yields

$$\frac{G_n^{[i]}[\mu,\nu] - \mathbb{E}\, G_n^{[i]}[\mu,\nu]}{\varrho[\mu]^n} \stackrel{d}{=} \frac{G_n^{[i]}[\tilde{\mu}] - \mathbb{E}\, G_n^{[i]}[\tilde{\mu}] + G_n^{[d+1]}[\tilde{\mu}] - \mathbb{E}\, G_n^{[d+1]}[\tilde{\mu}]}{\varrho[\tilde{\mu}]^n},$$

for $G_n^{[i]}[\tilde{\mu}]$, $G_n^{[d+1]}[\tilde{\mu}]$ independent. Thus we have

$$\mathbb{P}\left(\frac{G_n^{[i]}[\mu,\nu] - \mathbb{E}\, G_n^{[i]}[\mu,\nu]}{\varrho[\mu]^n} > t \right) \leq \mathbb{E}\ \exp(\langle u(\mathbf{e}_i + \mathbf{e}_{d+1}), \mathbf{Y}_n \rangle - ut)$$

and get by analog calculations as in proof of Theorem 3.2.1 the assertion for

$$L_\kappa = \min_{0\leq \ell < k} K(\ell)_{\kappa/(\kappa-1)}^{1-\kappa}\, 2^{-\kappa/2}\, \frac{(\kappa-1)^{\kappa-1}}{\kappa^\kappa},$$

which completes the proof. ■

Bibliography

[1] Aldous, D. (1991) The continuum random tree II: an overview. *Stochastic Analysis, Cambridge University Press*, 23–70.

[2] Ali Khan, T. and Neininger, R. (2004) Probabilistic analysis for randomized game tree evaluation. *Mathmatics and Computer Science III. Birkhäuser, Basel*, 163–174.

[3] Ali Khan, T., Devroye, L. and Neininger, R. (2005) A limit law for the root value of minimax trees. *Electronic Communications in Probability 10*, 273–281.

[4] Athreya, K. B. and Ney, P. (1972) Branching processes. *Die Grundlehren der mathematischen Wissenschaften, Bd. 196, Springer-Verlag, New York-Heidelberg.*

[5] Bennett, G. (1962) Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association 57*, 33–45

[6] Biggins J.D. (2004) The growth of iterates of multivariate generating functions. *Preprint, University of Sheffield.*

[7] Devroye, L. (1986) A note on the height of binary search trees. *Journal of Association for Computing Machinery 33*, 498–498.

[8] Devroye, L. (1998) Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic methods for algorithmic discrete mathematics, Algorithms Combin., 16, Springer, Berlin*, 249–314.

[9] Devroye, L. and Kamoun, O. (1996) Random minimax game trees. *Random Discrete Structures, John Wiley, New York*, 55–80.

[10] Entringer, R.C., Meir, A., Moon, J.W., and Székely (1994) The Wiener index of trees from certain families. *Australasian Journal of Combinatorics 10*, 211–224.

[11] Fill, J.A. and Janson, S. (2001) Approximating the limiting quicksort distribution. *Random Structures Algorithms 19*, 376–406.

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

[13] Fill, J.A. and Janson, S. (2002) Appendix to quicksort asymptotics. Unpublished, available from `http://www.math.uu.se/~svante/` or `http://www.mts.jhu.edu/~fill/`

[14] Harris, T. E. (1963) The theory of branching processes. *Die Grundlehren der Mathematischen Wissenschaften, Bd. 119, Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J.*

[15] Hayward, R.B. and McDiarmid, C.J.H. (1996) Large deviations for quicksort. *Journal of Algorithms 21*, 476–507.

[16] Hwang, H.-K. and Neininger, R. (2002) Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM Journal of Computing 31*, 1687–1722.

[17] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *Journal of American Statistical Association 58*, 13–30.

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

[19] Jones, O.D. (2004) Large deviations for supercritical multitype branching processes *Journal of Applied Probability 41*, 703–720.

[20] Karlin, S. (1966) A first course in stochastic processes. *Academic Press, New York.*

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

[22] Kesten, H. and Stigum, B.P. (1966a) A limit theorem for multidimensional Galton-Watson processes. *Annals of Mathematical Statistics 37*, 1211–1223.

[23] Kesten, H. and Stigum, B.P. (1966b) Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Annals of Mathematical Statistics 37*, 1463–1481.

[24] Knuth, D.E. (1997a) The art of computer programming. Volume 1: fundamental algorithms. 3rd ed. *Addison-Wesley.*

[25] Knuth, D.E. (1997b) The art of computer programming. Volume 2: seminumerical algorithms. 3rd ed. *Addison-Wesley.*

[26] Knuth, D.E. (1998) The art of computer programming. Volume 3: sorting and searching. 2nd ed. *Addison-Wesley.*

[27] Knuth, D.E. and Moore, R.W. (1975) An analysis of alpha-beta pruning. *Artificial Intelligence 6*, 293–326.

[28] Ledoux, M. (1995/97) On Talagrand's deviation inequalities for product measures. *ESIAM Probab. Statist 1*, 63–87 (electronic).

[29] Ledoux, M. (1996) Isoperimetrie and Gaussian analysis. *Lectures on probability and statistics (Saint-Flour, 1994). Lecture Notes in Mathematics 1648*, 165–294.

[30] Lugosi, G. (2006) Concentration-of-measure inequalities. Lecture notes, available from `http://www.econ.upf.es/~lugosi/`

[31] Mahmoud, H.M. and Smythe, R.T. (1994) A survey of recursive trees. *Theory of Probability and Mathematical Statistics 51*, 1–29.

[32] Meiners, M. (2006) Über stochastische Minimax-Fixpunktgleichungen. Diploma thesis. *Westfälische Wilhelms-Universität Münster.*

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

[34] Nau, D.S. (1982) An investigation of the causes of pathology in games. *Artificial Intelligence 19*, 257–278.

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

[36] Neininger, R. (2002) The Wiener index of random trees. *Combinatorics, Probability & Computing 11*, 587–597.

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

[38] Newborn, M. (1977) The effiency of the alpha-beta search of trees with branch-depend terminal node scores. *Artificial Intelligence 8*, 137–153.

[39] Pearl, J. (1980) Asymptotic properties of minimax trees in game-searching procedures. *Aritificial Intelligence 14*, 113–126.

[40] Pearl, J. (1984) Heuristics: Intelligent search strategies. *Addison-Wesley*

[41] Rabin, M.O. (1976) Probabilistic Algorithms. *Algorithms and Complexity: New Directions and Recent Results, Traub, J. editor, Academic Press*, 21–39

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

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

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

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

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

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

[48] Talagrand, M. (1995) Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Etudes Sci. Publ. Math. 81*, 73–205.

[49] Wagner, S. (2006) On the average Wiener index of degree-restricted tree. *Australasian Journal of Combinatorics,* to appear

[50] Wiener, H. (1947) Structural determination of paraffin boiling points. *Journal of the American Chemical Society 69*, 17–20.

[51] Zhang, Y. (1987) The Variance of Two Game Tree Algorithms. *Proceedings of the of the eighth annual ACM-SIAM symposium on Discrete algorithms, New Orleans , Lousianna, U.S.*, 268–277.

Zusammenfassung

Ausgangspunkt dieser Dissertation ist die stochastische Analyse rekursiver Algorithmen und Datenstrukturen. Die Analyse von Algorithmen befasst sich mit der Bewertung der Effizienz von Algorithmen. Dabei wird die Komplexität eines Algorithmus als ein Parameter definiert, der die Größen, die am wichtigsten für die Effizienz des Algorithmus sind, widerspiegelt. Meistens ist die Laufzeit eine solche Größe aber auch Speicherplatzbedarf kann eine solche sein. Größen wie Laufzeit und Speicherplatzbedarf eines Algorithmus hängen nicht nur von dem Algorithmus ab, sondern auch von der Eingabe. Somit hängt auch die Komplexität eines Algorithmus von dem Algorithmus und von der Eingabe ab. Folglich reicht es nicht aus, wenn man die Komplexität von zwei oder mehreren Algorithmen, die dasselbe Problem lösen, miteinander vergleichen will, deren Komplexität für nur eine oder wenige Eingaben miteinander zu vergleichen. Andererseits ist es häufig unmöglich die Komplexität von Algorithmen für alle Eingaben miteinander zu vergleichen, da es zu viele Eingaben gibt. Um dieses Dilemma der Analyse von Algorithmen zu überwinden, macht man sich folgende Beobachtung zu Nutze: Die Komplexität der meisten Algorithmen wächst im Großen und Ganzen mit der Länge der Eingabe. Deshalb wird die Komplexität von Algorithmen in Abhängigkeit von ihrer Eingabelänge untersucht. Dabei werden häufig asymptotische Resultate für wachsende Eingabelänge geliefert.

Eine Methode, um die Komplexität von Algorithmen in Abhängigkeit von ihrer Eingabe zu untersuchen, ist die Average-Case-Analyse, die 1963 von D.E. Knuth begründet wurde. Dabei wird eine Verteilung auf der Menge aller Eingaben gleicher Länge definiert und die dadurch determinierte erwartete Komplexität (Average-Case-Komplexität) studiert. Häufig ist dies die uniforme Verteilung, doch, motiviert durch Anwendungen, können auch andere Verteilungen von Interesse sein. Seit den 1980er Jahren wird für solche stochastischen Modelle die Verteilung der Komplexität detaillierter studiert, als nur ihr Erwartungswert. Außerdem wird bisweilen die zufällige Ausgabe des Algorithmus analysiert.

Eine andere wichtige Methode, die in der Informatik häufig verwendet wird, ist die Worst-Case-Analyse. Bei der Worst-Case-Analyse wird die maximale Komplexität untersucht, wobei das Maximum über alle Eingaben gleicher Länge genommen wird. Die maximale Komplexität wird auch Worst-Case-Komplexität genannt und jede Eingabe, die eine Worst-Case-Komplexität liefert, wird Worst-Case-Eingabe genannt. Der Vorteil der Worst-Case-Analyse besteht darin, dass, falls die Worst-Case-Laufzeit eines Algorithmus als klein nachzuweisen ist, folglich die Komplexität des Algorithmus für jede Eingabe klein ist.

Nun gibt es Algorithmen deren Average-Case-Komplexität klein, aber deren Worst-Case-Komlexität groß ist. Ein wichtiges Prinzip der Informatik, das in solchen Situationen oft verwendet wird, ist das Randomisieren von Algorithmen. Dabei wird die Auswertungsreihenfolge des Algorithmus an manchen Stellen randomisiert. Dadurch wird die Komplexität zu jeder fest gegebenen Eingabe zufällig. Neben der erwähnten Modellannahme von zufälligen Eingaben ist dies ein weiterer Aspekt der Informatik, der eine stochastische Analyse motivieren kann. Wir interessieren uns in dieser Arbeit nur für solche randomisierten Algorithmen, die immer ein richtiges Ergebnis liefern. Bei randomisierten Algorithmen wird die (maximale) erwartete Komplexität analysiert. Aber auch andere Charakteristika wie Varianz, Grenzwertsatz, Konvergenzraten und Tailschranken werden studiert. Neben dem Erwartungswert sind obere Schranken für den rechten Tail für die Informatik besonders interessant, da man schlechtes Verhalten des Algorithmus — d.h. Komplexitäten, die wesentlich größer sind als erwartet — mit möglichst großer Wahrscheinlichkeit ausschließen möchte. Ist dies für einen Algorithmus gewährleistet und hat der Algorithmus für jede Eingabe eine gute erwartete Komplexität, so ist es sinnvoll, ihn zu verwenden, selbst wenn seine Worst-Case-Komplexität schlecht ist.

Ein wichtiges Werkzeug für die Analyse von Tailschranken sind stochastische Konzentrationsungleichungen, für die es verschieden Herangehensweisen gibt.

Eine Herangehensweise ist *Chernoff's bounding technique*. Die Idee davon besteht darin, die erzeugende Funktion $\mathbb{E}\exp(sX)$ einer zentrierten Zufallsvariable X von oben abzuschätzen, um mittels der Markov-Ungleichung eine obere Schranke für $\mathbb{P}(|X| > t)$ zu erhalten.

Die Azuma-Ungleichung (s. Azuma (1976)) ist eine Tailschranke für Martingale mit beschränkten Differenzen, die mit Hilfe von *Chernoff's bounding technique* bewiesen wird. Die Azuma-Ungleichung kann selbst wiederum verwendet werden, um $\mathbb{P}(|X| > t)$ abzuschätzen, indem man sich durch X und eine geeignete Filtration ein Doob'sches Martingal definiert und dessen Martingaldifferenzen abschätzt. Diese Herangehensweise wird Martingaldifferenzmethode oder Methode beschränkter Differenzen genannt.

Vgl. McDiarmid (1998) und Lugosi (2006) zur detaillierten Beschreibung dieser und anderer Zugänge zu stochastischen Konzentrationsungleichungen.

In dieser Arbeit werden verschiedene Folgen von multivariaten Zufallsvariablen, die eine rekursive Struktur haben, studiert. Dabei steht die Analyse ihrer Tailschranken im Mittelpunkt. Im ersten und zweiten Kapitel sind die Folgen aus Problemen der stochastischen Analyse von Algorithmen entstanden. Im dritten Kapitel werden superkritische Multityp-Galton-Watson-Prozesse studiert.

Roter Faden dieser Arbeit sind die oberen Tailschranken, die für diese zufälligen Strukturen bewiesen werden und die Methode, mit der wir sie erhalten. In jedem Kapitel werden normalisierte Versionen der multivariaten Zufallsvariablen, bezeichnet als $\mathbf{Y}_n$, $n \geq 1$, mittels *Chernoff's bounding technique* abgeschätzt. Dazu werden die multivariaten erzeugenden Funktionen $\mathbb{E} \exp\langle \mathbf{s}, \mathbf{Y}_n \rangle$ mit Induktion nach n abgeschätzt. Dabei wird ihre rekursive Struktur ausgenutzt. Im Zusammenhang mit Algorithmen wurde dieser Ansatz erstmals von Rösler (1991) für eine univariate rekursive Struktur verwendet. Es stellt sich heraus, dass die schwierigste Aufgabe darin besteht, den Induktionsschritt für $\mathbf{s}$ nah bei $(0, \ldots, 0)$ zu beweisen. Im wesentlichen geschieht dies durch eine Rechnung bezüglich $\mathbf{b}_n$, einem additiven Term, der in der Rekursionsgleichung von $\mathbf{Y}_n$ auftaucht (s. (1.4), (2.3) and (3.18)): Da $\mathbb{E}\, \mathbf{b}_n = (0, \ldots, 0)$ ist, gilt $\mathbb{E} \exp\langle \mathbf{s}, \mathbf{b}_n \rangle = 1 + \mathrm{O}(\|\mathbf{s}\|^2)$, für $\|\mathbf{s}\| \to 0$. Eine explizite Schranke erhalten wir dadurch, dass wir $\mathbb{E} \exp\langle \mathbf{s}, \mathbf{b}_n \rangle$ als Taylorreihe entwickeln. Eine ähnliche Rechnung wurde im Beweis von Bennett's Ungleichung verwendet (s. Bennett(1962)).

Im ersten Kapitel untersuchen wir Minimaxbäume. Im zweiten Abschnitt betreiben wir Worst-Case-Analyse für Snirs randomisierten Algorithmus zum Auswerten Boolscher Entscheidungsbäume. Dazu zeigen wir, dass es immer eine Eingabe $v^\star$ gibt, deren zufällige Komplexität $C(v^\star)$ die Komplexität von jeder anderen Eingabe mit gleicher Länge stochastisch dominiert. Dies rechtferitgt es,

$C(v^\star)$ als Worst-Case-Komplexität zu interpretieren. Für diese zufällige Worst-Case-Komplexität beweisen wir den exakten Erwartungswert, eine Asymptotik für die Varianz, einen Grenzwertsatz mit eindeutig charakterisiertem Grenzwert sowie die folgende Tailschranken:

Theorem 1.2.6 *Für alle $2 \leq \kappa < 1/(1-\alpha) \doteq 4.06$ existiert ein $L > 0$, sodass für jedes $t > 0$ und $n = 2^{2k}$*

$$\mathbb{P}\left(\frac{C(v^\star) - \mathbb{E}\,C(v^\star)}{n^\alpha} > t\right) \leq \exp(-Lt^\kappa)$$

gilt. Ein expliziter Wert von L ist in (1.6) gegeben. Dieselbe Schranke gilt für den linken Tail.

In Abschnitt 3.5 wird gezeigt, dass die obere Schranke $1/(1-\alpha)$ in Theorem 1.2.6 nicht verbesserbar ist. Diese Tailschranken gelten für binäre Entscheidungsbäume. Die Verallgemeinerung von diesem und allen anderen Resultaten für d-näre Entscheidungsbäume steht in Theorem 1.2.7.

In Pearls Modell für d-näre Minimaxbäume der Höhe $2k$ haben alle $n = d^{2k}$ Blätter des Minimaxbaumes unabhängig identisch verteilte Werte, wobei ihre Verteilungsfunktion F_V stetig und streng mononton steigend auf dem Bildbereich von $0 < F_V < 1$ ist. Für den Wert des Minimaxbaumes unter Pearls Modell wird im dritten Abschnitt des ersten Kapitels der folgende Grenzwertsatz bewiesen:

Theorem 1.3.1 *Für $d \geq 2$ sei W_n der Wert des d-nären Minimaxbaumes der Höhe $2k$, $n = d^{2k}$, unter Pearls Modell, q der einzige Fixpunkt von $f(x) = (1-(1-x)^d)^d$ auf $(0,1)$, $\xi = f'(q)$ und $\alpha = \log(\xi)/\log(d^2) \in (0,1)$. Dann gilt*

$$n^\alpha(F_V(W_n) - q) \xrightarrow{\mathcal{L}} W, \quad k \to \infty.$$

Die Zufallsvariablte W hängt nicht von $\mathcal{L}(V)$ ab, hat eine stetige Verteilungsfunktion F_W mit $0 < F_W < 1$, $F_W(0) = q$ und

$$F_W(x) = f\left(F_W(x/\xi)\right), \quad x \in \mathbb{R}.$$

Im zweiten Kapitel untersuchen wir die Tailschranken des Wiener-Index von zufälligen Binärsuchbäumen. Binärsuchbäume sind eine fundamentale Datenstruktur der Informatik zum Verwalten von Listen. Insbesondere besteht eine wohlbekannte Äquivalenz zwischen Binärsuchbäumen und Quicksort. In Abschnitt 2.2 analysieren wir den Wiener-Index mit *Chernoff's bounding technique*, wodurch wir folgende obere Tailschranke beweisen:

Theorem 2.1.1 *Es sei W_n der Wiener-Index des zufälligen Binärsuchbaumes der Ordnung n, $L_0 \doteq 5.0177$ die größte Wurzel von $e^L = 6L$ und $c = (L_0 - 1)/(24L_0^2) \doteq 0.0066$. Dann gilt für jedes $t > 0$ und jedes $n \geq 0$*

$$\mathbb{P}\left(\frac{W_n - w_n}{n^2} > t\right) \leq \begin{cases} \exp(-1/36t^2), & \textit{für } 0 \leq t \leq 9 \\ \exp(-1/96t^2), & \textit{für } 9 < t \leq 48L_0 \doteq 240.848 \\ \exp(-ct^2), & \textit{für } 44L_0 < t \leq 24L_0^2 \doteq 604.256 \\ \exp(-t(\ln(t) - \ln(4e)), & \textit{für } 24L_0^2 < t. \end{cases}$$

Dieselbe Schranke gilt für den linken Tail.

Da $\mathbb{E}\, W_n = 2n^2 \ln(n) + \mathrm{O}(n^2)$ ist (s. Hwang und Neininger (2002)), folgt daraus insbesondere die folgende Schranke für große Abweichungen:

$$\mathbb{P}\left(|W_n - \mathbb{E}\, W_n| > t\, \mathbb{E}\, W_n\right) \leq n^{-2t(\ln(\ln(n)) + \ln(t) - \ln(2e) + o(1))} \quad \forall\, t > 0, \forall\, n \geq 0.$$

Als alternative Herangehensweise analysieren wir die Tails des Wiener-Index in Abschnitt 2.3 mit der Methode beschränkter Differenzen. Dadurch erhalten wir Theorem 2.3.1, dass eine etwas schlechtere Abschätzung der Tails als die zuletzt aufgestellte Ungleichung liefert. Darüber hinaus beweisen wir in Abschnitt 2.4 die folgende untere Schranke für die Tails on W_n:

Theorem 2.1.3 *Für jedes feste $0 < t \leq 1$ gilt*

$$\mathbb{P}\left(|W_n - \mathbb{E}\, W_n| > t\, \mathbb{E}\, W_n\right) \geq \mathbb{P}\left(W_n - \mathbb{E}\, W_n > t\, \mathbb{E}\, W_n\right) \geq n^{-8t\left(\ln(\ln(n)) + O\left(\ln^{(3)}(n)\right)\right)},$$

für $n \to \infty$.

Die Worst-Case-Komplexität aus dem ersten Kapitel kann als die Generationengröße eines superkritischen 2-Typ-Galton-Watsonprozesses dargestellt werden, der von Karp und Zhang (1995) vorgestellt wurde. Im dritten Kapitel wird die Methode, mit der im ersten Kapitel die Tailschranken analysiert wurden, verallgemeinert. Dadurch werden folgende Tailschranken für die Generationengröße von superkritischen Multityp-Galton-Watson-Prozessen bewiesen:

Theorem 3.2.1 *Es sei $G_n^{[i]}$ die Generationengröße eines d-Typ-Galton-Watson-Prozesses mit endlicher maximaler Familiengröße zur Zeit n, der mit einem Typ-i-Individuum startet, und es sei $\varrho > 1$. Falls $q^* < 2$, dann existiert für jedes $2 \leq \kappa < \kappa^* = 1 + 1/(q^* - 1)$ eine Konstante $L_\kappa > 0$, sodass für jedes $n \geq 0$ und $1 \leq i \leq d$ gilt:*

$$\begin{aligned} \mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} > t\right) &\leq \exp(-L_\kappa t^\kappa) \quad t > 0, \\ \mathbb{P}\left(\frac{G_n^{[i]} - \mathbb{E}\, G_n^{[i]}}{\varrho^n} < -t\right) &\leq \exp(-L_\kappa t^\kappa) \quad 0 < t \leq \frac{\mathbb{E}\, G_n^{[i]}}{\varrho^n}. \end{aligned}$$

q^* und ϱ sind Größen, die durch die Nachkommensverteilung festgelegt und in Abschnitt 3.2 definiert sind. Bei positiv regulären Galton-Watson-Prozessen ist $\varrho > 1$ der größte Eigenwert der Mittelwertmatrix. q^* — und somit auch κ^* — ergibt sich auf kompliziertere Weise aus der Nachkommensverteilung. Aus diesem Grund geben wir implizit mit Proposition 3.4.1 eine untere Schranke für κ^* an, die sich unmittelbar aus der Nachkommensverteilung ergibt. Die daraus folgenden Tailschranken sind in Korollar 3.4.2 angeführt.

Da wir für die Analyse nicht voraussetzen müssen, dass der Galton-Watson-Prozess positiv regulär ist, gelingt es uns, darüber hinaus Theorem 3.6.1 zu beweisen, das Tailschranken für superkritische Multityp-Galton-Watson-Prozesse mit Migration liefert (vgl. Abschnitt 3.6).

Lebenslauf

Tämur Ali Khan
Pestalozziplatz 4
60385 Frankfurt am Main

Geburtsdatum: 12. Januar 1976
Geburtsort: Frankfurt am Main
Staatsangehörigkeit: deutsch

1982 - 1992	Grundschule Süd-West in Eschborn, Heinrich-von-Kleist-Gesamtschule in Eschborn
1992 - 1995	Gymnasiale Oberstufe in Schwalbach am Taunus, Allgemeine Hochschulreife
1995 - 1996	Zivildienst im Kinderhort Süd-West in Eschborn
Okt. 1996 - Aug. 2003	Johann Wolfgang Goethe-Universität Frankfurt am Main Okt. 96 - Sep. 97 Studium Lehramt an Grundschulen, Fächer Mathematik (Hauptfach), Deutsch und Sport Okt. 97 - Aug. 03 Studium Mathematik auf Diplom, Nebenfach Informatik: März 99 Vordiplom 'sehr gut', Professoren G. Burde, M. Reichert-Hahn, M. Schmidt-Schauß, A. Wakolbinger Diplomarbeit bei Prof. A. Wakolbinger, Titel „Ulams Problem und Hammersleys Prozess. Eine Ideengeschichte" Gutachter: Professoren G. Kersting, A. Wakolbinger Aug. 03 Diplom 'mit Auszeichnung', Professoren M. Reichert-Hahn, C.-P. Schnorr, A. Wakolbinger
seit Sep. 2003	Doktorand und wissenschaftlicher Angestellter bei Prof. R. Neininger, Nachwuchsgruppe „Probabilistische Analyse rekursiver Algorithmen" im Emmy Noether-Programm der DFG Fachbereich Informatik und Mathematik Institut für Stochastik und Mathematische Informatik