Rates of Convergence for Balanced, Irreducible Pólya Urns with Two Colours

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

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

von
Andrea Mechtilde Lisbeth Kuntschik
aus Mainz am Rhein

Frankfurt 2017
(D30)

vom Fachbereich 12 der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Andreas Bernig

Gutachter: Prof. Dr. Ralph Neininger, Johann Wolfgang Goethe-Universität

Ao.Univ.Prof. Dr. Alois Panholzer, Technische Universität Wien

Datum der Disputation: 25. August 2017

Abstract

Urn models are a classical tool to visualise problems in the field of discrete probability. They do not only serve to realise basic discrete distributions but also are used to model dynamically evolving systems. Pólya urn models are among the most popular and flexible of such urn schemes and therefore have a variety of applications in such as epidemiology, bioscience, and computer science.

Pólya urn models serve to describe stochastic processes that evolve in discrete time. Whenever a process evolves in time, it is a classical question to enquire about its long-term behaviour, concerning qualitative and quantitative aspects.

This thesis deals with balanced, irreducible Pólya urn schemes with two colours, say black and white. For this class of Pólya urn schemes limit theorems for the number of black balls after n steps are known, whereas there is a lack of rates of convergence. This thesis focusses upon upper bounds for the rate of convergence in these limit theorems.

It is known that depending on the ratio of the eigenvalues of the replacement matrix, two regimes of limit laws occur: almost sure convergence to a non-degenerate random variable whose distribution depends on the initial composition of the urn and that is known to be not normally distributed and weak convergence to the normal distribution. In this thesis upper bounds on the rates of convergence in both the *non-normal limit case* and the *normal limit case* are given. Parts of the results confirm a conjecture of Svante Janson.

List of Figures

Figures 1.1, 3.1, 3.2, 3.3, 7.1 and 7.2 were created with *Keynote* by *Apple Inc.*

Figures 1.2, 1.3, 1.4, 1.5, B.1, B.2, B.3 and B.4 were created with free software environment for statistical computing and graphics R.

Notation and Preliminary Remarks

In general the notation from Knape and Neininger [27] has been adopted as well as notation common in the field of the contraction method.

For a random variable X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ distributed according to ν, which will be abbreviated by $X \sim \nu$, its distribution is denoted by $\mathcal{L}(X) := \nu$. Moreover, $\mathbb{E}[X] := \int x \,\mathrm{d}\nu(x)$ signifies its mean and $\operatorname{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$ its variance. The fact $\mathcal{L}(X) = \mathcal{L}(Y)$ for two random variables X and Y is abbreviated by $X \stackrel{\mathrm{d}}{=} Y$. Convergence in distribution is denoted by $\stackrel{\mathrm{d}}{\longrightarrow}$.

$\mathcal{N}(\mu, \sigma^2)$ signifies the normal distribution with mean $\mu \in \mathbb{R}$ and variance σ^2, $\sigma > 0$. $\operatorname{Bin}(n, p)$ denotes the binomial distribution with success probability $p \in [0, 1]$ and n trials and $\operatorname{Ber}(p)$ signifies the Bernoulli distribution with success probability $p \in [0, 1]$. For any subset $\mathcal{A} \subset \Omega$ of a set of events Ω its complement is denoted by $\mathcal{A}^c$. The abbreviation *a.s.* means *almost sure* convergence.

Big O notation is used to describe limiting behaviour: Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be two sequences in $\mathbb{R}$. We write, as $n \to \infty$,

- $a_n = O(b_n) :\Leftrightarrow \exists C > 0, n_0 \in \mathbb{N}\, \forall n \geq n_0 : |a_n| \leq C\,|b_n|$;
- $a_n = \Theta(b_n) :\Leftrightarrow a_n = O(b_n)$ and $b_n = O(a_n)$;
- $a_n = o(b_n) :\Leftrightarrow \lim_{n\to\infty} \frac{a_n}{b_n} = 0$.

The Gamma function is denoted by $\Gamma(x) := \int_0^\infty y^{x-1} e^y \mathrm{d}y$, $x > 0$.

Twice, a proof will be interrupted; this will be indicated by ⧄. If a proof for a result is only sketched, this will be indicated by ⊠ instead of □ at the end.

To ease the flow of reading the narrative perspective varies between passive voice and first-person view using the first person plural noun "we".

In proofs, remarks and explanations regarding individual steps are usually ensued thereafter.

Contents

1. Introduction

In probability theory, urn models figure prominently when the question arises how to visualise a problem subject to randomness. An urn usually contains balls of different colours and the problem at hand somehow can be brought in line with drawings from this urn. Most naturally, the imagination of repeated drawings from an urn leads to the desire to capture a stochastic process by means of an urn model.
Pólya urn models are among the most popular and flexible of such urn models. They have a variety of applications and therefore are of great interest in probability theory.

A Pólya urn scheme serves to describe a stochastic process evolving in discrete time steps. At any time, the urn contains balls of different colours. One step of this process is defined as follows: One ball is drawn from the urn uniformly at random and then returned to the urn together with new balls. The rules on how to add new balls to the urn are given by the replacement matrix whose rows correspond to the colour of the drawn ball and whose columns indicate how many balls of which colour to add. Negative entries of the replacement matrix lead to the removal of balls from the urn. So, the replacement matrix $(R_{ij})_{1\leq i,j\leq q}$, for a total of q colours, carries the information to add (or remove) R_{ij} balls of colour j on drawing a ball of colour i. The steps are iterated independently.

What is known today as Pólya urn scheme seems to be first mentioned in 1906 by Markov in his seminal paper [38] as well as in [37]. Markov introduces the nowadays so-called classical Pólya urn with replacement matrix $\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$ as an example for a Markov chain. However, it is named after the the work of Pólya and Eggenberger [14] from 1923; Pólya and Eggenberger generalise Markov's urn such that a total of s balls of the same colour as the drawn ball is added to the urn. Pólya and Eggenberger use this urn model to describe series of interlinked events modelling an effect known as "the rich get richer" (their slogan in [14] is "Chancenvermehrung durch Erfolg"). Their motivation to design such a framework was the common assumption of independent events, which they did not approve of when modelling real life events. In their original paper they used it as a model for contagion processes.

Bernstein [5] and Friedman [18] extend the urn model such that additionally $t \geq 1$ balls of the opposite colour than the drawn ball are added in every step, hence the scheme is balanced, i.e., in every step the same number of balls is added, and the replacement matrix is still symmetric. Bagchi and Pal [3] break the symmetry and coin the notion of *tenable* urn models, that is, the scheme is balanced (with possibly negative entries in the replacement matrix) and designed in such a way that the process cannot get stuck.

Since the introduction of Pólya urns in the beginning of the last century, Pólya urn models have been generalised, comprehensively studied and referred to as the *generalised Pólya-Eggenberger urn scheme*, or, more conveniently, *Pólya urns*. Johnson and Kotz [25] provide a general overview of urn models together with historical information where Chapter 4 covers Pólya urn schemes; a more recent work is Mahmoud's monograph on Pólya urns [34] where a lot of applications, especially in computer science with respect to data structures and bioscience, are discussed; Pemantle [46] gives a survey on random processes with reinforcement where Pólya urns serve as a starting point. In addition, an extensive overview of the literature on Pólya urns can be found in Janson [22], Flajolet et al. [16], Pouyanne [47] as well as Kuba and Sulzbach [28]. As already indicated, Pólya urns cover a broad range of applications; they are prototypical for any process underlying some sort of enrichment dynamics: Populations, epidemics, data storage, reinforcement processes.

Of course, such a Pólya urn scheme can be rather general: The matrix does not necessarily exhibit any sort of structure, there can be an arbitrary number of colours, an arbitrary initial composition of the urn, and replacements could be random. Hence, in order to get a grip on the evolution of the urn process, there are two natural conditions that are usually required to be satisfied by a Pólya urn scheme:

1. *Balancedness*: In every step of the urn the same number of balls is added. This number is called the *balance*.
2. *Irreducibility*: Regardless of the initial composition of the urn a ball of any colour can be observed in the evolution of the urn with positive probability.

Balanced, irreducible Pólya urn schemes mostly find applications in computer science within the scope of data storage.

Naturally, the asymptotic behaviour of Pólya urn models has been of great interest and therefore has been studied comprehensively: With these restrictions on the structure of the Pólya urn scheme, the long-term behaviour of this process in terms of the number of balls of a specific colour is fully understood with respect to limit theorems, see for example Bernstein [5] and Savkevich [54], Bernstein [4], Freedman [17], Blackwell and Kendall [7], Athreya and Karlin [2], Athreya [1], Bagchi and Pal [3], and Smythe [55]. A comprehensive study of limit theorems (also covering the results on limit theorems of all beforehand mentioned works) is Janson [22].

Depending on the "shape" of the replacement matrix different limit laws arise. However, so far almost completely unanswered is the question of the speed or rate of convergence in those limit theorems. *The aim of this thesis is to thoroughly address this question with respect to the class of balanced, irreducible Pólya urn schemes with two colours, say black and white, by providing upper bounds on the rate of convergence.*

There are two settings that this thesis encompasses:

Balanced Irreducible Two-Colour Pólya Urns

(**Det R**)

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ with } a, d \in \mathbb{N}_0 \cup \{-1\} \text{ and } b, c \in \mathbb{N},$$

such that $a + b = c + d =: K - 1 \geq 1$ (balancedness)

and $bc > 0$ (irreducibility).

The first row indicates how to add balls to the urn when a black ball is drawn and the second row is evoked when a white ball is drawn: Hence, in setting **Det R** a black balls and b white balls are added to the urn if a black ball is drawn; otherwise c black and d white balls are added. Drawn balls can be removed, but the removal of other balls than the drawn one is not allowed. Moreover, note that the assumption of irreducibility reduces to the condition $bc > 0$ in the case of two-colour Pólya urns. The assumption of irreducibility cannot be omitted; the asymptotic behaviour of two-colour Pólya urns that are not irreducible (but triangular) is fundamentally different, cf. Janson [23]. The assumption of balancedness is important for the recursive approach from Knape and Neininger [27] that is displayed in Chapter 3. In the course of this thesis we refer to this setting as **Det R** as the replacement itself is deterministic; in contrast to the next setting studied where the replacement does not only depend on the colour of the drawn ball but also is subject to randomness.

Randomised Play-the-Winner Rule

(**Rand R**)

$$\bar{R} = \begin{pmatrix} C_\alpha & 1 - C_\alpha \\ 1 - C_\beta & C_\beta \end{pmatrix} \text{ with } C_\alpha \sim \mathrm{Ber}(\alpha), C_\beta \sim \mathrm{Ber}(\beta),$$

$\alpha, \beta \in (0, 1)$.

In setting **Rand R** the replacement matrix is random. This urn comes with two coins, the coin C_α that is tossed if a black ball is drawn from the urn and the coin C_β that rules the replacement if a white ball is drawn. A successful coin toss results in putting the drawn ball back into the urn together with a new ball of the same colour; otherwise the drawn ball is returned to the urn and a new ball of the other colour is added. It is associated with the design of clinical trials where two treatments are tested against each other and referred to as the "randomised play-the-winner rule", cf. Wei and Durham [59] and Wei [58]. The two treatments are represented by the colours and the coin acts as the outcome of the treatment (beneficial or not). The motivation behind this scheme is an ethical one: The desire to assign the better treatment to more test persons during the course of the trial.
Obviously, the balance is one and the scheme is irreducible as long as both α and β are strictly greater than zero and strictly less than one.

In the entire thesis, we refer to these two urns as **Det R** and **Rand R**. The quantity of interest is the number of black balls after n steps, denoted by B_n.

Asymptotic Behaviour of the Number of Black Balls

The asymptotic behaviour of the normalised number of black balls — of both urns **Det R** and **Rand R** — is closely related to the replacement matrix: In setting **Det R**, it depends on the ratio of smallest to largest eigenvalue of the replacement matrix R which is given by $\lambda = \frac{a-c}{a+b}$. In setting **Rand R**, the ratio of smallest to largest eigenvalue of the matrix that contains the expectations of the entries of $\bar{R}$ determines the asymptotic behaviour; it is given by $\lambda = \alpha + \beta - 1$.

There are two regimes of limit laws: If $\lambda > \frac{1}{2}$, let the normalised number of black balls be given by $X_n := \frac{B_n - \mathbb{E}[B_n]}{n^\lambda}$ for $n \geq 1$. Then X_n converges almost surely two a non-degenerate random variable whose distribution depends on the replacement matrix and the initial composition of the urn. Note that it is not normally distributed; on the other hand, if $\lambda \leq \frac{1}{2}$ (and $\lambda \neq 0$ in setting **Det R**), let the normalised number of black balls be given by $\hat{X}_n := \frac{B_n - \mathbb{E}[B_n]}{\sqrt{\mathrm{Var}(B_n)}}$ for $n \geq 2$ (note that $\mathrm{Var}\,(B_1) = 0$ in setting **Det R** for monochromatic initial compositions). Then $\hat{X}_n$ converges in distribution to the standard normal distribution $\mathcal{N}\,(0,1)$. The first ones to observe this dichotomy are Bernstein [5] and Savkevich [54] in the case of symmetric, balanced, irreducible two-colour Pólya urn schemes. This was followed by Freedman [17], Athreya and Karlin [2], Bagchi and Pal [3], Gouet [20] for setting **Det R**, and Smythe and Rosenberger [56] for setting **Rand R**, and Janson [22, Theorems 3.22, 3.23, 3.24] covering both settings. For properties of the non-normal limit law, see Chauvin et al. [9] and Kuba and Sulzbach [28]. These two regimes are referred to as the *non-normal limit case* and the *normal limit case.* [1]

Remark (Intuition: Two regimes of limit laws)**.** This remark serves to develop a sense for the two regimes of limit laws: In setting **Det R** the ratio of smallest to largest eigenvalue is given by $\lambda = \frac{a-c}{a+b}$. The parameter λ lies in the interval $\left[-\frac{K+1}{K-1}, \frac{K-3}{K-1}\right]$, where $K - 1 := a + b$. Hence, "small" values of λ are close to -1 and "large" ones close to 1.

Now, consider Friedman's urn and the classical Pólya urn: Friedman's urn is ruled by the replacement matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. The classical Pólya urn is ruled by the replacement matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. In Friedman's urn, the ratio of the eigenvalues is -1. In the classical Pólya urn this ratio is 1.

These two urns serve to make us understand the phase transition from weak (normal) limits for $\lambda \leq \frac{1}{2}$ to almost sure (non-normal) limits for $\lambda > \frac{1}{2}$.

In Friedman's urn, in every step the drawn ball is returned to the urn together with a new ball of the opposite colour. Therefore, the proportions of the two colours even out: Whenever one colour is dominating, it is more likely to add balls of the other colour. This finally leads to a normal limit law for the normalised number of black balls, cf. Freedman [17].

In Pólya's urn, life is not that fair. In every step a ball of the same colour as the drawn ball is added. Hence, the proportion of the drawn colour is reinforced. As soon as one colour dominates, this colour is likely to dominate forever, it will grow stronger and stronger. The beginning of the process is critical to its long-term evolution. All in all, these dynamics lead to an almost sure limit depending on the initial composition of the urn, cf. Athreya [1].

Results

The upper bounds on the distance between the normalised number of black balls and its respective limits that are presented in this thesis are completely novel: The results are derived by bringing the problem into the sphere of the contraction method. To this end, the evolution of the urn process is captured recursively. Due to the approach via the contraction method, the rates are derived in different metrics depending on whether the non-normal limit case or the normal limit case is studied.

To state our main result, let ℓ_p denote the Wasserstein metric, ϱ the Kolmogorov-Smirnov distance, and ζ_3 the Zolotarev metric. Formal definitions of these metrics are provided later in Chapter 2.

[1] Note that the non-normal limit case is also referred to as *large* or *large-index urn* whereas the normal limit case is also referred to as *small* or *small-index urn* in the literature, cf. Chauvin et al. [9] as well as Kuba and Sulzbach [28].

Given a balanced, irreducible, two-colour Pólya urn scheme with replacement matrix $R = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ where the integer entries satisfy $a, d \geq -1$ and $b, c > 0$, i.e., a Pólya urn scheme in setting **Det R**. Let $\lambda := \frac{a-c}{a+b}$. Consider the following normalisations of the number of black balls B_n

$$X_n := \frac{B_n - \mathbb{E}\left[B_n\right]}{n^\lambda}, \qquad n \geq 1, \quad \text{in the non-normal limit case } \lambda > \frac{1}{2};$$
$$\hat{X}_n := \frac{B_n - \mathbb{E}\left[B_n\right]}{\sqrt{\operatorname{Var}\left(B_n\right)}}, \qquad n \geq 2, \quad \text{in the normal limit case } \lambda \leq \frac{1}{2}.$$

Let X_R^0 denote the almost sure limit of X_n and $\mathcal{N}(0,1)$ the standard normal distribution and let $\varepsilon > 0$.
Then, as $n \to \infty$,

in the non-normal limit case where $\lambda > \frac{1}{2}$,

$$\ell_p\left(X_n, X_R^0\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right),$$
$$\varrho\left(X_n, X_R^0\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right);$$

in the normal limit case where $\lambda \leq \frac{1}{2}$,

$$\zeta_3\left(\hat{X}_n, \mathcal{N}(0,1)\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda-\frac{1}{2}\right)}\right), & \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon}\right), & \lambda \leq \frac{1}{3}, \lambda \neq 0. \end{cases}$$

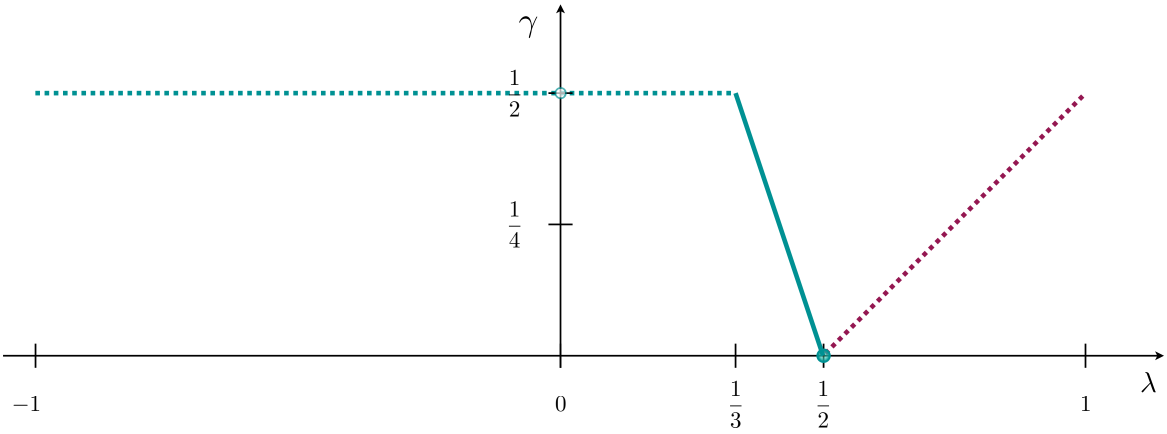
Figures 1.1, 1.2 and 1.4 as well as additional simulations in Appendix B illustrate these rates of convergence. Analogous statements with $\lambda = \alpha + \beta - 1$ hold for the number of black balls in setting **Rand R**.

These results summarise the statements of Theorem 7.4 that covers the non-normal limit case and Theorem 7.6 that covers the normal limit case.

A weaker statement of the results was previously published in the 2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO), see Kuntschik and Neininger [30]; note that since then the rate in the case $\lambda = \frac{1}{2}$ could be improved from $(\ln(n))^{-\frac{1}{2}}$ to $(\ln(n))^{-\frac{3}{2}}$.

The rate $n^{3\left(\lambda-\frac{1}{2}\right)}$ for $\lambda \in \left(\frac{1}{3}, \frac{1}{2}\right)$ confirms a conjecture of Svante Janson stated in [22, Remark 4.7]. Janson did not mention in what metric he expects this rate to hold. Moreover, he begins his remark with the note that his methods give no information on the rate of convergence.

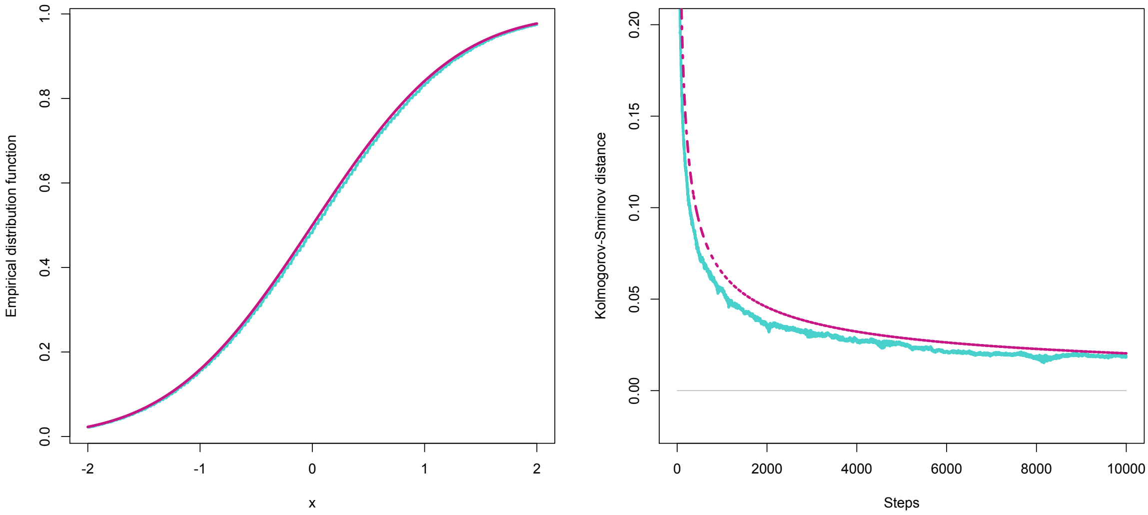
Figure 1.1.: The behaviour of the exponents of the rates stated above is shown.



The negative exponent denoted by γ is shown (hence, $n^{-\gamma}$ yields the order of the upper bound for the respective value of λ). The dashed lines indicate that the exponent also depends on ε and therefore is below the depicted line (but gets arbitrarily close to this line).
The maroon coloured line gives the exponent in the non-normal limit case. The teal coloured line covers the normal limit case. Moreover, for $\lambda = \frac{1}{2}$ there is no polynomial bound for the rate. This is indicated by a teal coloured circle at $(1/2, 0)$. Note that for $\lambda = 0$ in setting **Det R** the evolution of the urn process is deterministic and in setting **Rand R** degenerates to the situation of the classical Central Limit Theorem; this is highlighted by another circle at $(0, 1/2)$.

Methods to Study Pólya Urns

Athreya and Karlin [2] are motivated by the fact that Pólya urn schemes can be represented by Markov chains and study the asymptotic behaviour of Pólya urn schemes via embedding the urn process into continuous time Markov branching processes (and lay the foundation thereof); this approach is a popular tool for analysing urn processes since the resulting continuous time process is better to handle than the discrete time process. The question of asymptotic normality was first tackled by Bernstein [5] and by Bagchi and Pal [3] using the method of moments. Moreover, discrete time martingale techniques were applied to derive limit theorems, see for example Gouet [20]. Janson [22] perfects the approach from Athreya and Karlin. It seems (and Janson declares so in [22] concerning the method of embedding into a continuous time branching process) that these methods do not allow to derive rates of convergence in the respective limit theorems easily. Flajolet et al. [16] present an analytic approach that produces rates of convergence for a subclass, namely urns with subtraction. Until the work of Knape and Neininger [27], the methods the asymptotic behaviour of Pólya urn



(a) Empirical Distribution Function (b) Rate of Convergence

Figure 1.2.: Simulation of 10^4 steps on the basis of 10^5 samples of a Pólya urn with replacement matrix $\left(\begin{smallmatrix}1 & 4\\ 3 & 2\end{smallmatrix}\right)$, hence $\lambda = -\frac{2}{5}$, with one initial black ball.
Figure 1.2a shows the empirical distribution function of the normalised number of black balls (turquoise) compared to the distribution function of the standard normal distribution (magenta).
Figure 1.2b shows the uniform distance between the empirical distribution function and the distribution function of the standard normal distribution (turquoise), i.e., a simulation of the Kolmogorov-Smirnov distance between the normalised number of black balls and the standard normal distribution, compared to a rate of order $n^{-\frac{1}{2}}$ (magenta).

schemes was studied with did not seem to easily permit the derivation of rates of convergence in general settings.

A method that is able to yield both limit theorems and rates of convergence is the contraction method. Knape and Neininger [27] introduced a recursive approach to capture the evolution of the urn process that makes Pólya urns accessible to the contraction method. We exploit this approach in order to derive upper bounds on the rate of convergence in the respective limit theorems. In the non-normal limit case our procedure is inspired by Fill and Janson [15], where rates of convergence for the search algorithm Quicksort are derived. Neininger and Rüschendorf [42] study the contraction method with degenerate limit equations and serve as a model in the normal limit case. We extend both methods to systems of distributional recursions as to the derivation of rates of convergence for Pólya urns.

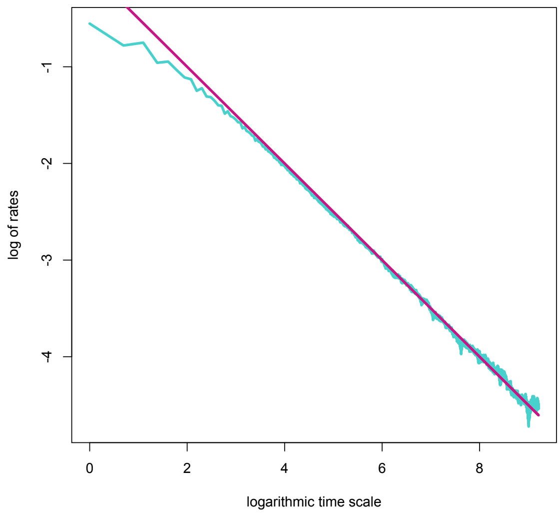


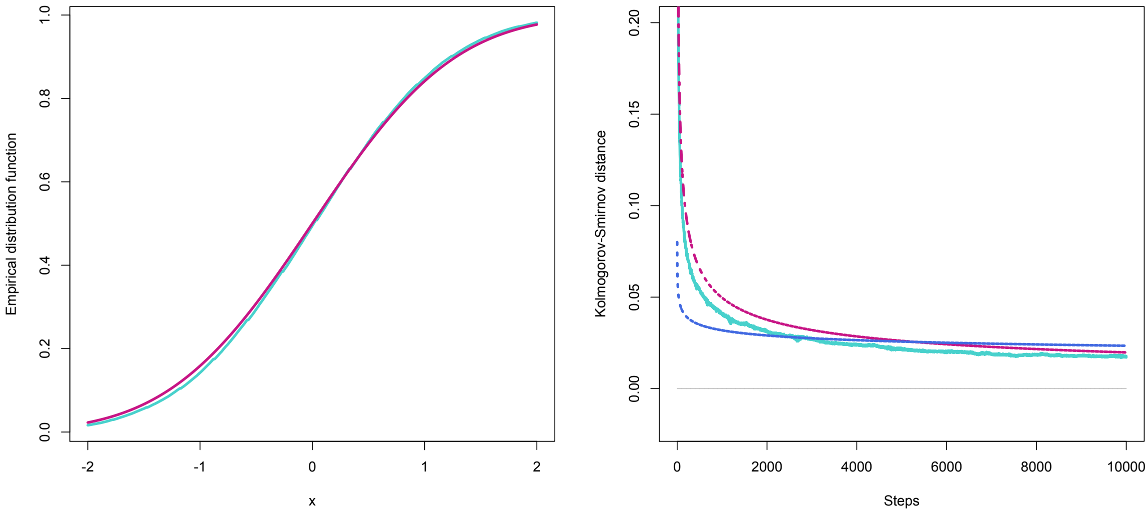
Figure 1.3.: Log log plot, belonging to the simulation of Figure 1.2.

The figure shows the logarithm of the simulated rate (turquoise) and the logarithm of the rate $n^{-\frac{1}{2}}$ (magenta). Hence, the magenta coloured line is a line with slope $-\frac{1}{2}$.

Rates of Convergence for Pólya Urns

For Pólya urn schemes governed by certain replacement matrices, not necessarily fitting our setting, rates of convergence for limit theorems are already known: Hwang [21] shows rates of convergence through a refined method of moments in the Kolmogorov-Smirnov distance for the asymptotic normality of the space requirement of m-ary search trees that are related to Pólya urn schemes. The analytic approach on asymptotic normality for urns with subtraction in Flajolet et al. [16] provides rates of convergence in the Kolmogorov-Smirnov distance as a by-product. Goldstein and Reinert [19] derive a bound on the rate of convergence in the Wasserstein distance for the classical Pólya urn, i.e., two colours and balanced diagonal replacement matrix, by applying Stein's method to a characterising equation for the Beta distribution. Likewise, in Peköz et al. [45] rates of convergence of optimal order in the Kolmogorov-Smirnov distance for a class of time inhomogeneous Pólya urn schemes are derived that cover a balanced, triangular (and therefore not irreducible) two-colour Pólya urn scheme.

It seems that these methods are not easily applicable to the class of balanced, irreducible Pólya urns that are treated in this thesis. Note that there are no results concerning rates of convergence in the non-normal limit case of balanced, irreducible Pólya urn schemes. So far



(a) Empirical Distribution Function (b) Rate of Convergence

Figure 1.4.: Simulation of 10^4 steps on the basis of 10^5 samples of a Pólya urn with replacement matrix $\left(\begin{smallmatrix} 20 & 10 \\ 9 & 21 \end{smallmatrix}\right)$, hence $\lambda = \frac{11}{30}$, with one initial black ball.
Figure 1.4a shows the empirical distribution function of the normalised number of black balls (turquoise) compared to the distribution function of the standard normal distribution (magenta).
Figure 1.4b shows the uniform distance between the empirical distribution function and the distribution function of the normal distribution (turquoise), i.e., a simulation of the Kolmogorov-Smirnov distance between the normalised number of black balls and the standard normal distribution, compared to a rate of order $n^{3\left(\lambda-\frac{1}{2}\right)}$ (magenta) and to a rate of order $n^{\lambda-\frac{1}{2}}$ (royal blue).

there has been no general result on rates of convergence for the class of balanced, irreducible Pólya urn schemes with two colours.

Outline

Chapter 2 summarises all technical information needed in the proofs of this thesis.

The core of the proofs is the recursive approach to the evolution of the urn process that was first introduced by Knape and Neininger [27]. Chapter 3 recalls this approach and adds information on the behaviour of the therein occurring quantities with regard to rates of convergence.

The recursive approach enables us to state a system of distributional recursions for the number of black balls when the urn initially contains a single ball. This system of distributional

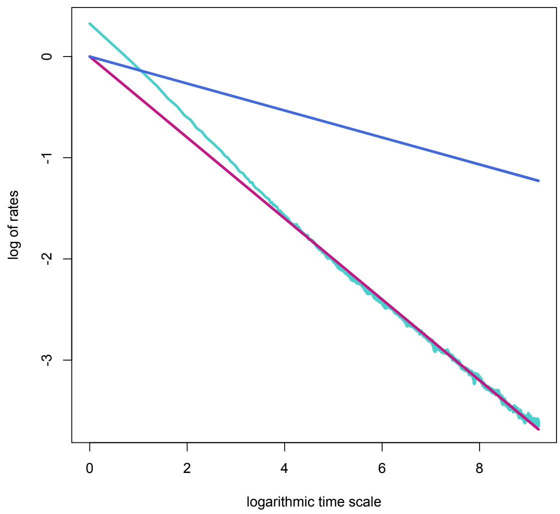


Figure 1.5.: Log log plot, belonging to the simulation of Figure 1.4.

The figure shows the logarithm of the simulated rate (turquoise) and the logarithm of the rate $n^{3\left(\lambda-\frac{1}{2}\right)}$ (magenta) and the logarithm of the rate $n^{\lambda-\frac{1}{2}}$ (royal blue). Hence, the magenta coloured line is a line with slope $3\left(\lambda-\frac{1}{2}\right)$ and the royal blue coloured line is a line with slope $\lambda-\frac{1}{2}$.

recursions describing the two *base cases* links Pólya urns to the contraction method.

In Chapter 4, the contraction method that originates in Rösler's treatment of the sorting algorithm Quicksort [50] is introduced and the general setting of systems of distributional recursions with respect to rates of convergence is explained.

Subsequently, upper bounds for rates of convergence are derived in Chapters 5 and 6 as well as in Chapter 7: Firstly, rates of convergence are derived for the two base cases, i.e., urns that initially contain a single ball — black or white. Chapter 5 treats setting **Det R**, whereas Chapter 6 studies setting **Rand R** and is to be understood complementary to Chapter 5. For that reason, Chapter 6 should be read in parallel to Chapter 5. Chapter 5 is the key to the reasoning performed in Chapter 6 and Chapter 7: Therefore, proofs in Chapter 5 are performed in full detail, whereas Chapter 6 serves to highlight the differences of the two urn settings considered. In Chapter 7 upper bounds for rates of convergence are derived for urns with an arbitrary initial composition of the urn on the basis of the results for the base cases of Chapter 5 (and Chapter 6).

Finally, a discussion of methods and results together with concluding remarks is given in Chapter 8.

In the appendix some technical lemmata are stated that are needed in the course of this thesis. Moreover, mean and variance for the number of black balls are given. Furthermore, the R code underlying the simulations is given together with additional simulations.

2. Technical Preliminaries and Metrics

This chapter provides all information on the spaces and metrics needed to derive rates of convergence in the context of the contraction method. At first, the spaces of probability measures that will be of interest are introduced. Subsequently, the metrics and some of their properties are outlined. The Wasserstein metrics make the start; they are needed in the case of non-normal limits. They are followed by the Kolmogorov-Smirnov distance. The Zolotarev metrics form the end; they are used in the case of normal limits. The stated properties are accompanied by remarks on where they are of use.

Let $\mathcal{M}$ denote the space of probability measures on the real line $\mathbb{R}$ with respect to the Borel σ-algebra. The subspaces with p-th moment, fixed mean, and fixed variance, respectively, are denoted by

$$\mathcal{M}_p := \{\mathcal{L}(X) \in \mathcal{M} : \mathbb{E}[|X^p|] < \infty\}, p \geq 1,$$
$$\mathcal{M}_p(\mu) := \{\mathcal{L}(X) \in \mathcal{M}_p : \mathbb{E}[X] = \mu\}, p \geq 1, \mu \in \mathbb{R},$$
$$\mathcal{M}_p\left(\mu, \sigma^2\right) := \left\{\mathcal{L}(X) \in \mathcal{M}_p(\mu) : \mathrm{Var}(X) = \sigma^2\right\}, p \geq 2, \mu \in \mathbb{R}, \sigma > 0.$$

Besides that, for $d \in \mathbb{N}$, let

$$\mathcal{M}^{\times d} := \underbrace{\mathcal{M} \times \ldots \times \mathcal{M}}_{d \text{ times}}, \qquad \text{and analogously, } (\mathcal{M}_p(\mu))^{\times d}, \left(\mathcal{M}_p\left(\mu, \sigma^2\right)\right)^{\times d}$$

denote the d-fold Cartesian product of $\mathcal{M}$, $\mathcal{M}_p(\mu)$, $\mathcal{M}_p(\mu, \sigma^2)$, respectively.

The Wasserstein Distance

Let $p \geq 1$ and $\nu, \rho \in \mathcal{M}_p$. The *Wasserstein distance* ℓ_p between ν and ρ is defined by

$$\ell_p(\nu, \rho) := \inf\left\{\|V - W\|_p \,|\mathcal{L}(V) = \nu, \mathcal{L}(W) = \rho\right\}$$

where $\|V - W\|_p := (\mathbb{E}[|V - W|^p])^{\frac{1}{p}}$ denotes the p-norm. A pair of random variables (V', W') with $\mathcal{L}(V') = \nu$ and $\mathcal{L}(W') = \rho$ is called a coupling of ν and ρ. A coupling (V', W') of ν and ρ is called an *optimal ℓ_p-coupling* of ν and ρ if $\ell_p(\nu, \rho) = \|V' - W'\|_p$. Note that for probability measures on the real line the coupling does not depend on p. Hence, such a pair of random variables will simply be referred to as an *optimal coupling*.

For probability measures on the real line optimal couplings can easily be constructed with the help of the inverse of the distribution functions and a random variable U uniformly distributed on $[0, 1]$: Let F and G denote the (right-continuous) distribution functions of ν and ρ. Let $F^{-1}(x) := \inf \{y | F(y) \geq x\}$ for $x \in [0, 1]$ (where we set $\inf \emptyset = \infty$) be the (left-continuous) inverse of the distribution function F, and let G^{-1}, analogously defined, be the inverse of G. Then, $\ell_p(\nu, \rho) = \|V^\star - W^\star\|_p$ with $V^\star := F^{-1}(U)$ and $W^\star := G^{-1}(U)$, i.e., the pair $(V^\star, W^\star)$ is an optimal coupling of ν and ρ, see Major [36].

The Wasserstein metric ℓ_p is also referred to as *minimal L_p metric* mirroring the idea of extracting a *simple* metric from the L_p-distances, i.e., a metric whose value is determined only by the marginals of the random variables inserted.

Convergence in the Wasserstein metric ℓ_p implies weak convergence (more precisely, it is equivalent to weak convergence plus convergence of the absolute p-th moment), see Bickel and Freedman [6, Lemma 8.3].

The Wasserstein distance measures the L_p-distance of a pair of random variables with given marginal distributions whose joint distribution minimises the L_p-distance. Note that the L_p-distance of any pair of random variables with the respective marginal distributions gives an upper bound for the Wasserstein distance ℓ_p of these distributions.

To estimate L_p-distances, the Marcinkiewicz-Zygmund inequality as stated in Chow, Teicher [10] will be of use:

Theorem 2.1 (Marcinkiewicz-Zygmund inequality)**.** *If $X_n, n \geq 1$, are independent random variables with mean 0, then for every $p \geq 1$ there exist positive constants A_p, B_p depending only on p for which*

$$A_p \left\| \left(\sum_{j=1}^{n} X_j^2 \right)^{\frac{1}{2}} \right\|_p \leq \left\| \sum_{j=1}^{n} X_j \right\|_p \leq B_p \left\| \left(\sum_{j=1}^{n} X_j^2 \right)^{\frac{1}{2}} \right\|_p .$$

The proof can be found in Chow, Teicher [10, Section 10.3, Theorem 2, p. 386].

Remark 2.2 (Remark on optimal couplings)**.** To understand the impact of optimal couplings consider the following choices of couplings of the uniform distribution on $[0, 1]$ and the uniform distribution on $\left\{\frac{j}{n} | 0 \leq j \leq n-1\right\}$: Let ν be uniformly distributed on the unit interval $[0, 1]$ and ν_n uniformly distributed on the set $\left\{\frac{j}{n} | 0 \leq j \leq n-1\right\}$, $n \geq 1$. Let U be a random variable distributed according to ν. Now, consider three possible constructions of a coupling of ν and ν_n:

$$U_n^{(1)} := \frac{\lfloor nU \rfloor}{n}, \quad U_n^{(2)} := \frac{\mathrm{Bin}\,(n-1, U)}{n}, \quad U_n^{(3)} := \frac{\lfloor n\,(1-U) \rfloor}{n},$$

where $\mathrm{Bin}\,(n-1, U)$ denotes a binomially distributed random variable whose success probability is to be determined by the random variable U. It is easy to check that $\mathcal{L}\left(U_n^{(1)}\right) = \mathcal{L}\left(U_n^{(3)}\right) = \nu_n$. For $U_n^{(2)}$, it holds, with $k \in \{0, \ldots, n-1\}$, with use of relationship between Gamma function and Beta function given by $\int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ with $x, y > 0$,

$$\begin{aligned}\mathbb{P}\left(\mathrm{Bin}\,(n, U) = k\right) &= \int_0^1 \binom{n-1}{k} u^k (1-u)^{n-1-k}\,\mathrm{d}u \\ &= \binom{n-1}{k} \frac{\Gamma(k+1)\,\Gamma(n-1-k+1)}{\Gamma(n+1)} = \frac{1}{n}.\end{aligned}$$

Let $p \geq 1$. In Lemma 3.7 we will see that

$$\left\| U_n^{(1)} - U \right\|_p = \Theta\left(\frac{1}{n}\right).$$

Applying the Marcinkiewicz-Zygmund inequality, Theorem 2.1, we obtain

$$\left\| U_n^{(2)} - U \right\|_p = \frac{1}{n-1}\left(\int_0^1 \mathbb{E}\left[\left|\mathrm{Bin}\,(n-1, u) - (n-1)\,u\right|^p\right]\mathrm{d}u\right)^{\frac{1}{p}} = \Theta\left(\frac{1}{\sqrt{n}}\right).$$

Finally, we have

$$\left\| U_n^{(3)} - U \right\|_p = \Theta(1)$$

due to

$$\begin{aligned}&\int_0^1 \left|\frac{\lfloor n\,(1-u) \rfloor}{n} - u\right|^p \mathrm{d}u \\ =&\frac{1}{n^p}\int_0^1 \left|\lfloor n\,(1-u) \rfloor - n\,(1-u) + n\,(1-u) - nu\right|^p \mathrm{d}u = \Theta(1),\end{aligned}$$

since $\left|\lfloor n\,(1-u) \rfloor - n\,(1-u)\right| \in [0, 1]$ and $\left|n\,(1-u) - nu\right| = \Theta(n)$.

It is easy to check via the way explained above on how to construct optimal couplings that the pairs $\left(U_n^{(1)}, U\right)$ for $n \in \mathbb{N}$ are optimal couplings of ν and ν_n. Hence, for the Wasserstein distance between ν and ν_n it holds

$$\ell_p(\nu_n, \nu) = \Theta\left(\frac{1}{n}\right).$$

Depending on the choice of the coupling, we observe two different rates of convergence and one case where convergence does not even occur. Hence, the knowledge of the existence of optimal couplings can be essential.

The following lemma is a summary of Major [36, Theorem (8.1)] and Bickel and Freedman [6, Lemma 8.2]:

Lemma 2.3. *Let $p \geq 1$. For any family $\mathcal{P} \subset \mathcal{M}_p$ there exist random variables V_ν defined on the same probability space with $\mathcal{L}(V_\nu) = \nu$ for $\nu \in \mathcal{P}$ such that for any $\nu, \rho \in \mathcal{P}$ the pair (V_ν, V_ρ) is an optimal coupling of ν and ρ.*

Lemma 2.4 (Monotonicity of the Wasserstein metrics)**.** *Let $1 \leq q \leq p < \infty$ as well as $\mathcal{L}(X)$ and $\mathcal{L}(Y) \in \mathcal{M}_p$. Then,*

$$\ell_q(X, Y) \leq \ell_p(X, Y).$$

Proof. Let $1 \leq q < p < \infty$ and set $r := \frac{p}{q}$ as well as $s := \frac{p}{p-q}$. Obviously, we have $\frac{1}{r} + \frac{1}{s} = 1$ and Hölder's inequality yields for a random variable Z with $\mathbb{E}[|Z|^p] < \infty$

$$\|Z\|_q^q = \||Z|^q\|_1 \leq \||Z|^q\|_r \|1\|_s = \|Z\|_p^q \text{ and therefore } \|Z\|_q \leq \|Z\|_p.$$

Let $X^\star \stackrel{\mathrm{d}}{=} X$ and $Y^\star \stackrel{\mathrm{d}}{=} Y$ such that the pair $(X^\star, Y^\star)$ is an optimal coupling of $\mathcal{L}(X)$ and $\mathcal{L}(Y)$. Then, we have

$$\ell_q(X, Y) = \|X^\star - Y^\star\|_q \leq \|X^\star - Y^\star\|_p = \ell_p(X, Y).$$

□

For $\nu = (\nu_1, \ldots, \nu_d), \rho = (\rho_1, \ldots, \rho_d) \in (\mathcal{M}_p)^{\times d}$, the Wasserstein metric is extended to $(\mathcal{M}_p)^{\times d}$ via the *maximal Wasserstein metric*

$$\ell_p^\vee(\nu, \rho) := \max_{1 \leq j \leq d} \ell_p(\nu_j, \rho_j).$$

We sloppily write $\ell_p(V, W) := \ell_p(\mathcal{L}(V), \mathcal{L}(W))$ for random variables V and W as well as $\ell_p^\vee((V_1, \ldots, V_d), (W_1, \ldots, W_d)) := \ell_p^\vee((\mathcal{L}(V_1), \ldots, \mathcal{L}(V_d)), (\mathcal{L}(W_1), \ldots, \mathcal{L}(W_d)))$.

Later on, when applying the reasoning of the contraction method, completeness of the metric spaces at hand will be of use.

Lemma 2.5. *Let $1 \leq p < \infty$, $\mu \in \mathbb{R}$ and $d \in \mathbb{N}$. The pairs $(\mathcal{M}_p, \ell_p)$, $(\mathcal{M}_p(\mu), \ell_p)$ and $\left((\mathcal{M}_p(\mu))^{\times d}, \ell_p^{\vee}\right)$ are complete metric spaces.*

Proof. From Bickel and Freedman [6, Lemma 8.1] we have that $(\mathcal{M}_p, \ell_p)$ is a metric space; hence, $(\mathcal{M}_p(\mu), \ell_p)$ is a metric space, too. Identity of indiscernibles, symmetry and the triangle inequality for $\ell_p^{\vee}$ follow directly from these properties of ℓ_p yielding that $\left((\mathcal{M}_p(\mu))^{\times d}, \ell_p^{\vee}\right)$ is a metric space.

Completeness: Consider a Cauchy sequence $(\nu_n)_{n\in\mathbb{N}}$ in $(\mathcal{M}_p, \ell_p)$. Due to Lemma 2.3 there is a sequence of random variables $(V_n)_{n\in\mathbb{N}}$, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that the pair (V_n, V_m) is an optimal coupling of ν_n and ν_m, for all $m, n \geq 1$. Due to $\|V_n - V_m\|_p = \ell_p(\nu_n, \nu_m)$, $(V_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $(L_p(\Omega, \mathcal{A}, \mathbb{P}))$. The Riesz-Fischer Theorem yields the existence of a limit $V \in L_p(\Omega, \mathcal{A}, \mathbb{P})$ with $\|V_n - V\|_p \to 0$ as $n \to \infty$. Denoting $\nu := \mathcal{L}(V)$, we observe $\ell_p(\nu_n, \nu) \leq \|V_n - V\|_p \to 0$. Therefore, the completeness of $(\mathcal{M}_p, \ell_p)$ follows.

For $\mathcal{M}_p(\mu)$ is a closed subset of $\mathcal{M}_p$, $(\mathcal{M}_p(\mu), \ell_p)$ is a complete metric space as well. The completeness of $\left((\mathcal{M}_p(\mu))^{\times d}, \ell_p^{\vee}\right)$ directly follows. □

These and other information that goes beyond what is needed here on the Wasserstein metrics can be found in Cambanis et al. [8], Major [36], Bickel and Freedman [6], and Rachev [49].

The first appearance of what is now called contraction method in Rösler [50] and [51] relies on the Wasserstein metric ℓ_2. Our proofs for upper bounds on rates of convergence in the non-normal limit case rely on the Wasserstein metric ℓ_2, as well, and then are transferred to the Wasserstein metrics ℓ_p. In that context, we make use of the existence of optimal couplings. Additionally, in setting **Rand R** we make use of optimal couplings when it comes to capturing the asymptotic behaviour of the subtree sizes, see Section 3.2.

The Kolmogorov-Smirnov Distance

The *Kolmogorov-Smirnov distance*, also known as *uniform distance*, of two probability measures $\mu, \nu \in \mathcal{M}$ with distribution functions F_μ and F_ν, respectively, is given by

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

The Kolmogorov-Smirnov distance measures the maximal pointwise distance between the distribution functions of the given distributions and therefore gives a vivid impression of what it means when two distributions are close to each other. It is not surprising that this distance is of the greatest interest not only when it comes to applications.

Recall the following correlation between the Wasserstein distances and the Kolmogorov-Smirnov distance from Fill and Janson [15]:

Lemma 2.6 ([15, Lemma 5]). *Assume, $\mu, \nu \in \mathcal{M}$ such that μ is absolutely continuous with a bounded density function f. Let $M := \sup_{x \in \mathbb{R}} |f(x)|$ and $1 \leq p < \infty$. Then,*

$$\varrho(\mu, \nu) \leq (p+1)^{\frac{1}{p+1}} \left(M \ell_p(\mu, \nu)\right)^{\frac{p}{p+1}}.$$

In the case of non-normal limits, this lemma will serve to convey the rates obtained in the maximal Wasserstein metrics $\ell_p^{\vee}$ to the *maximal Kolmogorov-Smirnov distance* $\varrho^{\vee}$ defined via

$$\varrho^{\vee}(\mu, \nu) := \max_{1 \leq j \leq d} \varrho(\mu_j, \nu_j),$$

where $\mu = (\mu_1, \ldots, \mu_d), \nu = (\nu_1, \ldots, \nu_d) \in \mathcal{M}^{\times d}$ with $d \in \mathbb{N}$.

As before, whenever we plug in a random variable (a random vector) into the (maximal) Kolmogorov-Smirnov distance it is to be understood on the level of distributions.

The Zolotarev Metric

The *Zolotarev metric* ζ_s for $s > 0$ is defined by

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

for random variables X and Y and where

$$s = m + \alpha, \quad \text{with } m \in \mathbb{N}_0 \text{ and } 0 < \alpha \leq 1, \text{ and}$$
$$\mathcal{F}_s := \left\{ f \in \mathcal{C}^m(\mathbb{R}) : \left| f^{(m)}(x) - f^{(m)}(y) \right| \leq |x - y|^{\alpha} \right\}$$

with $\mathcal{C}^m(\mathbb{R})$ denoting the set of all m-times differentiable functions.

One could think of the Zolotarev metric ζ_s as the worst amount of cost you have to pay going from distribution $\mathcal{L}(X)$ to distribution $\mathcal{L}(Y)$ where $s = m + \alpha$ tells you what sort of costs to take into account: The costs are given by the set of m-times continuously differentiable functions such that the m-th derivative is α-Hölder continuous.

The Zolotarev metric was introduced by Zolotarev in [60] and [61]. A comprehensive presentation of properties of the Zolotarev metric can be found in Rachev [49]. The following properties are crucial for our reasoning: With $s = m + \alpha > 0$,

1. If random variables X and Y satisfy $\mathbb{E}\left[X^j\right] = \mathbb{E}\left[Y^j\right]$ for $j = 1, \ldots, m$ and $\mathbb{E}\left[|X|^s\right]$, $\mathbb{E}\left[|Y|^s\right] < \infty$, then we have $\zeta_s(X, Y) < \infty$.

2. The Zolotarev metric ζ_s is $(s, +)$-ideal, that is $\zeta_s(X+Z, Y+Z) \leq \zeta_s(X, Y)$ for random variables X, Y satisfying the first property and for a random variable Z independent of (X, Y), and it holds $\zeta_s(cX, cY) = |c|^s \zeta_s(X, Y)$ for $c \in \mathbb{R} \setminus \{0\}$. This implies for $X_1, \ldots, X_n$ independent and $Y_1, \ldots, Y_n$ independent, such that X_j and Y_j satisfy the conditions of the first property for $j = 1, \ldots, n$,

$$\zeta_s\left(\sum_{j=1}^n X_j, \sum_{j=1}^n Y_j\right) \leq \sum_{j=1}^n \zeta_s(X_j, Y_j).$$

From Drmota et al. [13, Theorem 5.1], we have

3. Convergence in ζ_s implies weak convergence.

4. For $1 < s \leq 2$, the pair $(\mathcal{M}_s(\mu), \zeta_s)$ with $\mu \in \mathbb{R}$ is a complete metric space, for $2 < s \leq 3$, the pair $\left(\mathcal{M}_s\left(\mu, \sigma^2\right), \zeta_s\right)$ with $\mu \in \mathbb{R}$ and $\sigma > 0$ is a complete metric space.

The Wasserstein metric and the Zolotarev metric connect in the following way, from Drmota et al. [13]:

Lemma 2.7 ([13, Lemma 5.7]). *Let $1 < s \leq 3$.*

If $1 < s \leq 2$, let $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s(\mu)$ for some $\mu \in \mathbb{R}$.

If $2 < s \leq 3$, let $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s\left(\mu, \sigma^2\right)$ for some $\mu \in \mathbb{R}$, $\sigma > 0$.

Then,

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

On the product spaces $(\mathcal{M}_s(\mu))^{\times d}$, if $1 < s \leq 2$ and $\left(\mathcal{M}_s\left(\mu, \sigma^2\right)\right)^{\times d}$, if $2 < s \leq 3$, we work with the *maximal Zolotarev metric* $\zeta_s^\vee$ defined by

$$\zeta_s^\vee\left(\left(X_1, \ldots, X_d\right), \left(Y_1, \ldots, Y_d\right)\right) := \max_{1 \leq j \leq d} \zeta_s(X_j, Y_j)$$

where $(\mathcal{L}(X_1,), \ldots, \mathcal{L}(X_d))$, $(\mathcal{L}(Y_1,), \ldots, \mathcal{L}(Y_d)) \in (\mathcal{M}_s(\mu))^{\times d}$, for $1 < s \leq 2$, and $(\mathcal{L}(X_1), \ldots, \mathcal{L}(X_d))$, $(\mathcal{L}(Y_1,), \ldots, \mathcal{L}(Y_d)) \in \left(\mathcal{M}_s\left(\mu, \sigma^2\right)\right)^{\times d}$, for $2 < s \leq 3$. Due to the fourth property, the pairs $\left((\mathcal{M}_s(\mu))^{\times d}, \zeta_s^\vee\right)$ for $1 < s \leq 2$ and $\left(\left(\mathcal{M}_s\left(\mu, \sigma^2\right)\right)^{\times d}, \zeta_s^\vee\right)$ for $2 < s \leq 3$ are complete metric spaces.

As already indicated in the definition of the Zolotarev metric and done so in this paragraph, out of convenience, we do not hesitate to plug in random variables however keeping in mind that they only represent their respective distributions.

More information on the Zolotarev metric can be found in Zolotarev [60] and [61], Rachev [49]. Rösler and Rüschendorf [53], Rachev and Rüschendorf [48], and especially Neininger and Rüschendorf [41] as well as [42] present the role of the Zolotarev metric in the context of the contraction method.

The Zolotarev metric comes into play when we deal with the normal-limit cases. We work with $\zeta_3^\vee$ as it seems to be the most convenient choice in order to derive rates of convergence.

3. Recursive Approach: Pólya Urns and Trees

This chapter is dedicated to the recursive understanding of the evolution of the urn process:

At first, the recursive approach of Knape and Neininger [27, Section 2] is recalled. It relies on a combinatorial discrete-time embedding of the evolution of the urn process into a random rooted tree growing simultaneously with the urn. Decomposing this tree at the root leads to a system of distributional recursions for the number of black balls after n steps when starting with a single ball, i.e., the base cases B_n^{b} and B_n^{w}. These distributional recursions make the evolution of the urn process accessible to the contraction method. Therefore, the recursive approach of Knape and Neininger [27] is indispensable for us.

The growth of the subtrees of the root will turn out to be crucial. It is related to another Pólya urn scheme. In order to study the growth of the subtrees, this Pólya urn scheme will be studied as parenthesis. Then, the growth of the subtrees is studied with respect to bounds for the rate of convergence as needed later on.

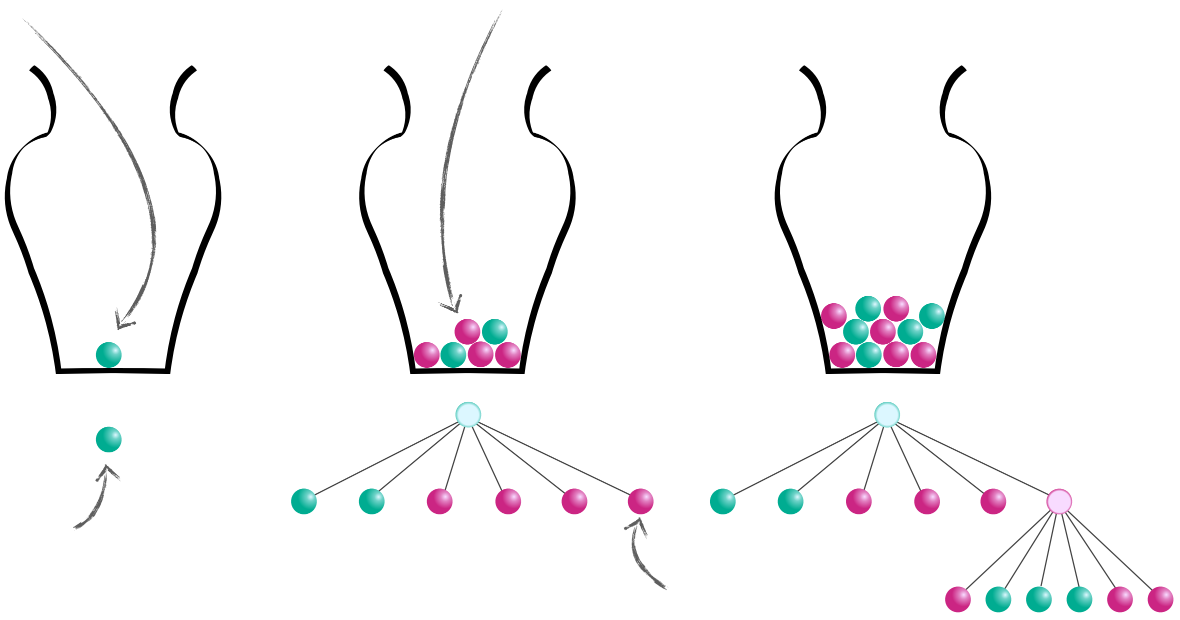
3.1. The Hidden Tree-Structure of the Evolution of the Urn Process

For it is our goal to apply the reasoning of the contraction method to Pólya urns, a recursive description of the evolution of the urn process is needed. The recursive approach that is recalled here (through the example of setting **Det R**) was first introduced by Knape and Neininger [27] and shortly thereafter by Chauvin et al. [9] in 2013. It is based on the observation of a hidden tree-structure within the evolution of the urn; every step of the urn process will be encoded by a tree whose leaves represent the balls in the urn. This tree structure reveals a self-similarity that is used to describe the number of black balls recursively.

Initially, there is one ball in the urn. This starting point is encoded by a tree that consists of one node only that obviously is a leaf. In the first step, this ball is drawn and returned to the urn together with $K-1 \geq 1$ new balls. In terms of the tree, a new level is opened up, consisting of a copy of the root node and $K-1$ new nodes. All of these K nodes are children of the root node and, hence, are the leaves of the tree. After the first step, the tree consists of $K+1$ nodes, of which K are leaves and represent the balls in the urn.

In order to make the leaves of the tree correspond to the balls in the urn, the nodes of the tree are assigned the colours of the balls they represent: The root node is assigned the same colour as the initial ball and passes its colour on to its copy emerging from the first step. The $K-1$ new leaves must be coloured according to the replacement matrix, i.e., a black and b white leaves will arise if the root is black and c black and d white leaves if the root is white.

Figure 3.1.: Evolution of a Pólya urn with black and white balls together with its associated tree:



Two (possible) steps of a Pólya urn governed by the replacement matrix $\left(\begin{smallmatrix}1 & 4\\ 3 & 2\end{smallmatrix}\right)$ and containing a single black ball in the beginning are depicted. Below the urn the respective associated tree is shown. The arrows indicate which ball is drawn and from which leaf the subtree emerges, respectively. The leaves of the associated tree correspond to the balls in the urn. For the sake of clarity, those nodes that do not represent balls in the urn wear faded colours.

From now on, this procedure is iterated and the tree evolves simultaneously with the urn, see Figures 3.1 and 3.2. It is called the *associated tree*. Obviously, we deal with two kinds of associated trees: One tree emerging from a black root, another emerging from a white root. To distinguish between these two, we call the former b-associated tree and the latter w-associated tree.

In general, drawing a ball from the urn is identified with picking the corresponding leaf from the tree. Returning this ball together with $K-1$ new balls according to the replacement matrix is represented by the emergence of a subtree rooted at the picked leaf, consisting of a copy of it and $K-1$ new leaves that are assigned colours as prescribed by the replacement matrix, too.

Every associated tree can be decomposed at the root into K subtrees. A b-associated tree gives rise to $a+1$ subtrees rooted in black and b subtrees rooted in white; analogously, a w-associated tree consists of c subtrees rooted in black and $d+1$ subtrees rooted in white. For that reason the number of black balls can be captured recursively: The number of black balls in the urn and the number of black leaves of the associated tree coincide and can be described as the sum of black leaves of the subtrees of the root.

As we deal with balanced urn schemes, every picked leaf spawns K subtrees regardless of its colour. By $I^{(n)} := \left(I_1^{(n)}, \dots, I_K^{(n)}\right)$ we denote a random vector whose components $I_r^{(n)}$ describe how often a leaf from the r-th subtree of the root was picked within the first n steps. The first step serves to give rise to the K subtrees, hence, $I^{(1)} = 0$ and $I^{(2)}$ is vector with all components equal to zero except for one component that equals 1. Exactly $n-1$ draws will occur among the subtrees, i.e., $\sum_{r=1}^{K} I_r^{(n)} = n-1$, and the distribution of the marginals of $I^{(n)}$ is given by, for $j \in \{0, \dots, n-1\}$ (where we set $\prod_{i=0}^{-1} x_i := 1$),

$$\mathbb{P}\left(I_r^{(n)} = j\right) = \binom{n-1}{j} \frac{\left(\prod_{i=0}^{j-1} (1 + (K-1)\, i)\right) \left(\prod_{i=0}^{n-2-j} (K - 1 + (K-1)\, i)\right)}{\prod_{i=0}^{n-2} (K + (K-1)\, i)}.$$

Moreover, $I^{(n)}$ provides information on the growth of the subtrees: The r-th subtree has $I_r^{(n)}(K-1)+1$ leaves. Therefore, we refer to $I_r^{(n)}$ as the size of the r-th subtree and to $I^{(n)}$ as the vector of the subtree sizes.

At last, we observe that the K subtrees of the root conditioned on $I^{(n)}$ behave independently and their distribution equals that of b- and w-associated trees, respectively, possessing the respective numbers of leaves.

With these observations at hand we are able to state a recursion for the number of black balls in the two base cases, that is when the urn initially contains one ball. Let the number of black balls after n steps when starting with a black ball be denoted by B_n^{b}. It is denoted by B_n^{w} when the initial ball is white.

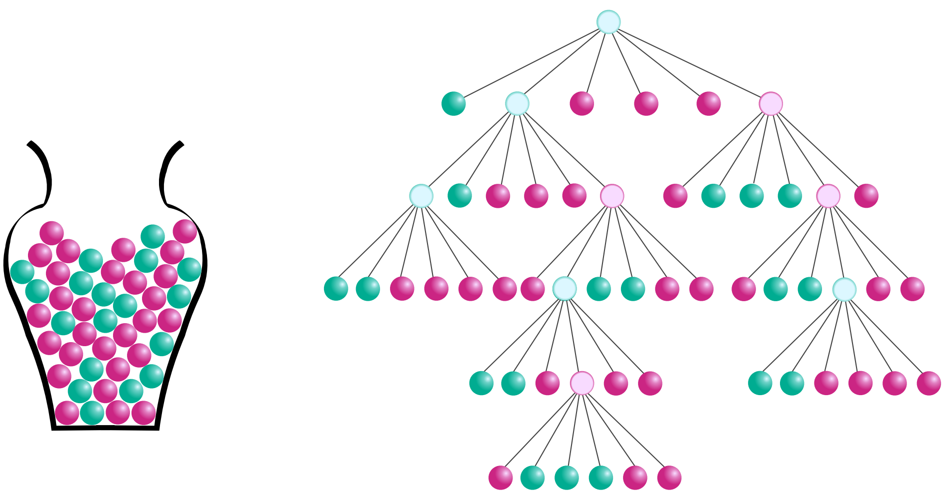
Thus, in setting **Det R**, we obtain the following distributional recursion for the number of black balls after n steps subject to the colour of the initial ball with $B_0^{\mathrm{b}} := 1$ and $B_0^{\mathrm{w}} := 0$

and for $n \geq 1$:

$$B_n^{\mathrm{b}} \overset{\mathrm{d}}{=} \sum_{r=1}^{a+1} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=a+2}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)},$$
$$B_n^{\mathrm{w}} \overset{\mathrm{d}}{=} \sum_{r=1}^{c} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=c+1}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)} \tag{3.1}$$

with $B_j^{\mathrm{b},(r)} \overset{\mathrm{d}}{=} B_j^{\mathrm{b}}$, $B_j^{\mathrm{w},(r)} \overset{\mathrm{d}}{=} B_j^{\mathrm{w}}$ for $r = 1, \dots, K$ and $0 \leq j \leq n$ such that $\left(B_j^{\mathrm{b},(1)}\right)_{0 \leq j \leq n}, \dots,$ $\left(B_j^{\mathrm{b},(K)}\right)_{0 \leq j \leq n}, \left(B_j^{\mathrm{w},(1)}\right)_{0 \leq j \leq n}, \dots, \left(B_j^{\mathrm{b},(K)}\right)_{0 \leq j \leq n}, I^{(n)}$ are independent.

Figure 3.2.: A realisation of the Pólya urn from Figure 3.1 after eight draws together with its associated tree.



According to the situation of Figure 3.1 the urn contains black and white balls. Initially, it contained one black ball and the replacement matrix is given by $\left(\begin{smallmatrix} 1 & 4 \\ 3 & 2 \end{smallmatrix}\right)$.

In setting **Rand R** we obtain with $B_0^{\mathrm{b}} := 1$ and $B_0^{\mathrm{w}} := 0$ and $I_n := I_1^{(n)}$, $J_n := I_2^{(n)} = n - 1 - I_n$,

$$B_n^{\mathrm{b}} \overset{\mathrm{d}}{=} B_{I_n}^{\mathrm{b},(1)} + C_\alpha B_{J_n}^{\mathrm{b},(2)} + (1 - C_\alpha)\, B_{J_n}^{\mathrm{w}},$$
$$B_n^{\mathrm{w}} \overset{\mathrm{d}}{=} B_{I_n}^{\mathrm{w},(1)} + (1 - C_\beta)\, B_{J_n}^{\mathrm{b}} + C_\beta B_{J_n}^{\mathrm{w},(2)} \tag{3.2}$$

with corresponding conditions on distributions and independence as in (3.1).

These recursions lay the foundations for determining rates of convergence via an induction. From systems (3.1) and (3.2) we can infer that the growth of the subtrees plays a key role when working with the recursive approach. Due to balancedness the vector of subtree sizes $I^{(n)}$ captures the growth of the subtrees. Hence, the next section will be dedicated to studying the subtrees.

Remark 3.1 (Remark on Tenable Urn Models)**.** Tenable urn schemes, in the notion of Bagchi and Pal [3], are balanced, irreducible two-colour Pólya urn schemes that allow the removal of balls other than the drawn one under certain divisibility conditions on the entries of the replacement matrix. If we remove balls other than the drawn one from the urn, then the subtrees of the associated tree are no longer independent conditioned on their sizes. Though, the independent behaviour of the subtrees given their sizes is crucial for our distributional recursions. As long as only the drawn ball is removed, one step of the urn happens in exactly one of the K subtrees of the root of the associated tree and does not interfere with any of the other subtrees and the trees behave independently given their sizes. Hence, in this thesis we concentrate on schemes where only the drawn ball may be removed.

3.2. The Behaviour of the Subtree Sizes

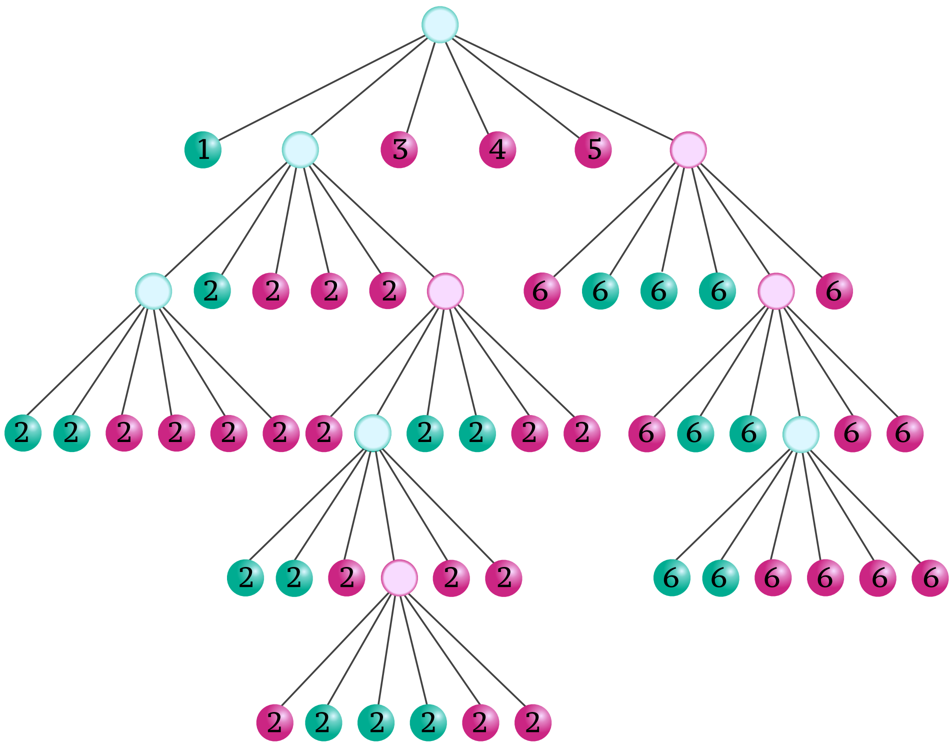
For the purpose of determining rates of convergence for the normalised numbers of black balls in setting **Det R** and **Rand R**, respectively, the behaviour of the random vector $I^{(n)} = \left(I_1^{(n)}, \dots, I_K^{(n)}\right)$ of subtree sizes that arises in the recursive approach in systems (3.1) and (3.2) has to be examined.

First it is observed that the vector of subtree sizes is connected to another well-known Pólya urn scheme with a diagonal replacement matrix:

We assign another label $j \in \{1, \dots, K\}$ to all the leaves of the associated tree specifying the subtree they have emerged from; i.e., a leaf which belongs to the j-th subtree is assigned label j, see Figure 3.3.

Obviously, whenever we draw a leaf of the j-th subtree, $K - 1$ new leaves belonging to the same subtree will appear. Hence, after the first step of the original urn, the behaviour of the number of draws from one subtree is governed by a Pólya urn with K colours that initially contains one ball of each colour where a drawn ball is returned to the urn together with $K-1$ new balls of the same colour. The number of draws from one subtree corresponds to how often balls of one colour were drawn in the new urn.

Figure 3.3.: Studying the growth of the subtrees via another Pólya urn scheme.



Given the associated tree as depicted in Figure 3.2, every leaf is assigned the number of the subtree it belongs to.

The following parenthesis will answer the question of how fast this quantity approaches its limit. Thereby, the long-term behaviour of the vector of rescaled subtree sizes is studied. Finally, a way to couple this vector and its limit is found that enables us to estimate the L_p-distance between both of them.

Parenthesis: Rate of Convergence in the Case of a Diagonal Replacement Matrix

Remark. This parenthesis is a standalone paragraph of this thesis. Herein, we will derive a rate of convergence for Pólya urn schemes with diagonal replacement matrices. To obtain the desired rates of convergence for the urn schemes posed in **Det R** and **Rand R**, we will make use of the results of this parenthesis.

By $\mathcal{C} \geq 2$ we denote a fixed number of colours and consider a Pólya urn that initially contains one ball of each colour $1, \ldots, \mathcal{C}$ and where one step of the urn process is defined as follows: One ball is drawn uniformly at random and is returned to the urn together with $r \geq 1$ new balls of the same colour.

Let $S_n := \left(S_n^{(1)}, \ldots, S_n^{(\mathcal{C})}\right)$ be a vector with entries $S_n^{(j)}$ that denote how often a ball of colour j was drawn within the first n steps of the urn, $j \in \{1, \ldots, \mathcal{C}\}$. Let $X_n := \left(X_n^{(1)}, \ldots, X_n^{(\mathcal{C})}\right)$, $n \geq 0$, be the composition of the urn after n steps, i.e., the urn contains $X_n^{(j)}$ balls of colour j after n steps, where $X_0 = (1, \ldots, 1)$. Let $Y_n := \left(Y_n^{(1)}, \ldots, Y_n^{(\mathcal{C})}\right)$, $n \geq 1$, be the outcome of the n-th draw, i.e., $Y_n^{(j)} = 1$ and $Y_n^{(i)} = 0$, for all $i \neq j$, if a ball of colour j was drawn, i and $j \in \{1, \ldots, \mathcal{C}\}$. Obviously, the quantity $S_n^{(j)}$ for $j \in \{1, \ldots, \mathcal{C}\}$ can be written in terms of $Y_i^{(j)}$: $S_n^{(j)} = \sum_{i=1}^n Y_i^{(j)}$. And likewise $X_n^{(j)}$ can be written in terms of $S_n^{(j)}$: $X_n^{(j)} = 1 + rS_n^{(j)}$ for $n \in \mathbb{N}$ and $j \in \{1, \ldots, \mathcal{C}\}$.

In Athreya [1] the limit of $\frac{1}{n} S_n := \left(\frac{S_n^{(1)}}{n}, \ldots, \frac{S_n^{(\mathcal{C})}}{n}\right)$ is implicitly determined. We state

Lemma 3.2. *The sequence* $\frac{1}{n} S_n$ *converges almost surely to a Dirichlet-distributed random vector* $D := (D_1, \ldots, D_{\mathcal{C}})$ *with all* $\mathcal{C}$ *parameters equal to* $\frac{1}{r}$.

Proof. From Athreya [1, Theorem 1 and Corollary 1] it is known that a random vector D exists such that $\frac{X_n}{\mathcal{C}+nr} \to D$ (a.s.), $n \to \infty$, where the distribution of the random vector $D = (D_1, \ldots, D_{\mathcal{C}})$ is given by the Dirichlet distribution with all parameters equal to $\frac{1}{r}$. Furthermore, for $j \in \{1, \ldots, \mathcal{C}\}$,

$$\left| \frac{1}{n} S_n^{(j)} - \frac{1}{\mathcal{C} + rn} X_n^{(j)} \right| = \left| \frac{\mathcal{C} S_n^{(j)} - n}{n\,(\mathcal{C} + rn)} \right| \leq \frac{\mathcal{C} + 1}{\mathcal{C} + rn} \to 0 \quad \text{a.s.} \quad (n \to \infty)\,.$$

It follows

$$\frac{1}{n} S_n \to D \quad \text{a.s.} \quad (n \to \infty)\,.$$

□

At the end of this paragraph, more information on the Dirichlet distribution will be given.

Since we are interested in rates of convergence, knowing the limiting behaviour is not enough: We want to know how fast this limit is approached. Therefore, we will check that de Finetti's Theorem can be applied to the sequence $(Y_n)_n$, yielding that Y_n conditioned appropriately exhibits a "Bernoulli-like" behaviour which transfers to S_n as a "multinomial-like" behaviour. Finally, we apply the Marcinkiewicz-Zygmund inequality to this result.

Theorem 3.3. *Let $p \geq 1$ and let $D := (D_1, \ldots, D_{\mathcal{C}})$ denote the almost sure limit of $\frac{1}{n} S_n$ from Lemma 3.2. Then there exist random vectors $\bar{S}_n := \left(\bar{S}_n^{(1)}, \ldots, \bar{S}_n^{(\mathcal{C})}\right)$ with $\mathcal{L}\left(\bar{S}_n\right) = \mathcal{L}\left(S_n\right)$ and $\bar{D} := \left(\bar{D}_1, \ldots, \bar{D}_{\mathcal{C}}\right)$ with $\mathcal{L}\left(\bar{D}\right) = \mathcal{L}\left(D\right)$ such that, as $n \to \infty$,*

$$\max_{j=1,\ldots,\mathcal{C}} \left\| \frac{1}{n} \bar{S}_n^{(j)} - \bar{D}_j \right\|_p = O\left(n^{-\frac{1}{2}}\right). \tag{3.3}$$

Proof. First of all, we check that the sequence $(Y_n)_{n \geq 1}$ is exchangeable in order to subsequently apply de Finetti's Theorem. Therefore, let $\mathbb{1}_i$ be the i-th unit vector, for $i \in \{1, \ldots, \mathcal{C}\}$.

Then, we have

$$\mathbb{P}\left(Y_n = \mathbb{1}_i \,|Y_1, \ldots, Y_{n-1}\right) = \frac{1 + S_{n-1}^{(i)} r}{\mathcal{C} + (n-1) r}.$$

Hence, with $y_1, \ldots, y_n \in \{1, \ldots, \mathcal{C}\}$ and $s_k^{(j)} := \sum_{i=1}^k \mathbb{1}_{y_i}^{(j)}$:

$$\begin{aligned}
&\mathbb{P}\left(\bigcap_{i=1}^n \{Y_i = \mathbb{1}_{y_i}\}\right) \\
=&\mathbb{P}\left(Y_n = \mathbb{1}_{y_n} \,|Y_1 = \mathbb{1}_{y_1}, \ldots, Y_{n-1} = \mathbb{1}_{y_{n-1}}\right) \mathbb{P}\left(\bigcap_{i=1}^{n-1} \{Y_i = \mathbb{1}_{y_i}\}\right) \\
=&\frac{1 + s_{n-1}^{(y_n)} r}{\mathcal{C} + (n-1) r} \mathbb{P}\left(\bigcap_{i=1}^{n-1} \{Y_i = \mathbb{1}_{y_i}\}\right) = \ldots = \frac{1 + s_{n-1}^{(y_n)} r}{\mathcal{C} + (n-1) r} \cdot \frac{1 + s_{n-2}^{(y_{n-1})} r}{\mathcal{C} + (n-2) r} \cdot \ldots \cdot \frac{1}{\mathcal{C}}
\end{aligned}$$

$$= \prod_{j=0}^{n-1} \frac{1}{\mathcal{C} + jr} \cdot \prod_{i=1}^{\mathcal{C}} \prod_{j=0}^{s_n^{(i)}-1} (1 + jr). \tag{3.4}$$

For an integer $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$ we denote by $n!_b^a := \prod_{j=0}^n (a + jb)$ the shifted multifactorial. Now, we can rewrite (3.4) as

$$\mathbb{P}\left(\bigcap_{i=1}^n \{Y_i = \mathbb{1}_{y_i}\}\right) = \frac{\prod_{i=1}^{\mathcal{C}} \left(s_n^{(i)} - 1\right)!_r^1}{(n-1)!_r^{\mathcal{C}}}.$$

Observing that the probability does not depend on any ordering of the outcomes, we obtain that $(Y_n)_{n \in \mathbb{N}}$ is an exchangeable sequence of random variables.

De Finetti's Theorem, cf. Klenke [26, Theorem 12.26], yields that there exists a random probability measure Ξ on the vertices of the unit simplex $\{\mathbb{1}_1, \ldots, \mathbb{1}_{\mathcal{C}}\}$ such that conditioned on Ξ the sequence of random variables $(Y_n)_{n\in\mathbb{N}}$ is independent and identically distributed with $\mathcal{L}(Y_i|\Xi) = \Xi$ almost surely.

In particular, we know that $\mathbb{P}(Y_1 = \mathbb{1}_j|\Xi) = \Xi(\{\mathbb{1}_j\})$ almost surely for $j \in \{1, \ldots, \mathcal{C}\}$, and, secondly, we have that $S_n^{(j)}$ given Ξ has the binomial distribution with parameters n and $\Xi(\{\mathbb{1}_j\})$. Furthermore, we can think of Ξ as a random vector $(\Xi(\mathbb{1}_1), \ldots, \Xi(\mathbb{1}_\mathcal{C}))$, since Ξ is uniquely determined by its point masses assigned to the $\mathcal{C}$ vertices of the unit $(\mathcal{C}-1)$-simplex. This random vector $(\Xi(\mathbb{1}_1), \ldots, \Xi(\mathbb{1}_\mathcal{C}))$ can be identified as the limit D of $\frac{1}{n}S_n$:

Conditioned on $\{\Xi = \xi\}$, the random variables $Y_1, Y_2, \ldots$ are identically and independently distributed according to the probability measure $\xi = (\xi(\mathbb{1}_1), \ldots, \xi(\mathbb{1}_\mathcal{C}))$ on the vertices of the standard $(\mathcal{C}-1)$-simplex. Hence, conditioned on $\{\Xi = \xi\}$, we have $\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n Y_i \stackrel{n\to\infty}{\longrightarrow} (\xi(\mathbb{1}_1), \ldots, \xi(\mathbb{1}_\mathcal{C}))$ almost surely by the strong law of large numbers. Unconditioned, we obtain $\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n Y_i \stackrel{\text{a.s.}}{\longrightarrow} (\Xi(\mathbb{1}_1), \ldots, \Xi(\mathbb{1}_\mathcal{C}))$ as $n \to \infty$. Therefore, we conclude with Lemma 3.2 that the limit D and the random measure Ξ arising from de Finetti's Theorem coincide.

Hence, we have that Y_1 conditioned on Ξ behaves as follows:

$$Y_1 = \begin{cases} \mathbb{1}_1, & \text{with probability } \Xi(\{\mathbb{1}_1\}), \\ \mathbb{1}_2, & \text{with probability } \Xi(\{\mathbb{1}_2\}), \\ \vdots & \vdots \\ \mathbb{1}_\mathcal{C}, & \text{with probability } \Xi(\{\mathbb{1}_\mathcal{C}\}). \end{cases}$$

By that, we obtain that the random vector $\left(S_n^{(1)}, \ldots, S_n^{(\mathcal{C})}\right)$ conditioned on D has the multinomial distribution with parameters n and $(\Xi(\{\mathbb{1}_1\}), \ldots, \Xi(\{\mathbb{1}_\mathcal{C}\}))$.

Finally, we determine the order of the distance between $\frac{1}{n}S_n$ and its limit Ξ via the Marcinkiewicz-Zygmund inequality:

Conditioned on Ξ, the quantity $S_n^{(j)} - n\Xi(\{\mathbb{1}_j\}) = \sum_{i=1}^n \left(Y_i^{(j)} - \Xi(\{\mathbb{1}_j\})\right)$ coincides with a sum of n independent identically distributed random variables, which evaluate to $1 - \Xi(\{\mathbb{1}_j\})$ with probability $\Xi(\{\mathbb{1}_j\})$ and to $-\Xi(\{\mathbb{1}_j\})$ with probability $1 - \Xi(\{\mathbb{1}_j\})$, hence, with mean 0. Obviously, every squared summand is bounded by 1.

Via conditioning on Ξ, we can fit the distance we are interested in into the setting of the Marcinkiewicz-Zygmund inequality, Theorem 2.1, and apply the right hand-side thereof, yielding: Let $j \in \{1, \ldots, \mathcal{C}\}$. Then,

$$
\begin{aligned}
\left\| \frac{1}{n} S_n^{(j)} - \Xi\left(\{\mathbb{1}_j\}\right) \right\|_p &= \frac{1}{n} \left(\mathbb{E}\left[\left| S_n^{(j)} - n\Xi\left(\{\mathbb{1}_j\}\right) \right|^p \right] \right)^{\frac{1}{p}} \\
= \quad & \frac{1}{n} \left(\int \mathbb{E}\left[\left| S_n^{(j)} - n\Xi\left(\{\mathbb{1}_j\}\right) \right|^p \middle| \Xi\left(\{\mathbb{1}_j\}\right) = x \right] \mathrm{d}\mathbb{P}_{\Xi(\{\mathbb{1}_j\})}(x) \right)^{\frac{1}{p}} \\
\leq \quad & \frac{1}{n} \left(\int (B_p)^p \, n^{\frac{p}{2}} \mathrm{d}\mathbb{P}_{\Xi(\{\mathbb{1}_j\})}(x) \right)^{\frac{1}{p}} \\
= \quad & \frac{1}{n} \left((B_p)^p \, n^{\frac{p}{2}} \right)^{\frac{1}{p}} = B_p \frac{1}{\sqrt{n}}
\end{aligned}
$$

with a suitable constant $B_p > 0$. This completes the proof. □

The Dirichlet Distribution

For general information on Dirichlet distributions see, for example, Johnson and Kotz [25, Section 2.7.6]. For our purpose, we only deal with the Dirichlet distribution of order N with all parameters equal to $\frac{1}{s}$ with $s \geq 1$. Its density function f in $x = (x_1, \ldots, x_N)$ is given by

$$
f(x) = \begin{cases} \dfrac{\left(\Gamma\left(\frac{1}{s}\right)\right)^N}{\Gamma\left(\frac{N}{s}\right)} \displaystyle\prod_{i=1}^{N} x_i^{\frac{1}{s}-1}, & x \in \mathcal{U}, \\ 0, & \text{elsewhere,} \end{cases}
$$

with $\mathcal{U} := \left\{ (x_1, \ldots, x_N) \in (0,1)^N : \sum_{i=1}^N x_i = 1 \right\}$ denoting the standard $(N-1)$-simplex.

Then, the marginals are Beta-distributed with parameters $\left(\frac{1}{s}, \frac{N-1}{s}\right)$ with density function

$$
g(x) = \begin{cases} \dfrac{\Gamma\left(\frac{N}{s}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{N-1}{s}\right)} x^{\frac{1}{s}-1} (1-x)^{\frac{N-1}{s}-1}, & x \in (0,1), \\ 0, & \text{elsewhere.} \end{cases}
$$

For general information on Beta distributions see, for example, Johnson and Kotz [25, Section 2.2.4].

Let $D = (D_1, \ldots, D_N)$ be Dirichlet-distributed of order N with all parameters equal to $\frac{1}{s}$. Note that

$$
\sum_{r=1}^{N} D_r = 1 \quad \text{almost surely,} \tag{3.5}
$$

and, as $D_r > 0$, we have

$$\sum_{r=1}^{K} \mathbb{E}\left[D_r^{\psi}\right] < 1 \Leftrightarrow \psi > 1, \tag{3.6}$$

which will be needed in the course of the calculations in order to identify contractive behaviour (see (4.13)) in the proofs of Chapters 5, 6 and 7.

Back to the Subtree Sizes

The recursions at hand, (3.1) and (3.2), are on the level of distributions. Therefore, we take the liberty to choose the random vector of subtree sizes suitably coupled to the almost sure limit of the rescaled subtree sizes. Note that this procedure, i.e., choosing the vector that represents the size of the subgroups, here subtrees, of the recursion such that it yields an appropriate coupling together with the limit of the rescaled subgroup sizes, is one of the key elements in the context of the contraction method.

Observing that the vector of subtree sizes $I^{(n)}$ and the vector S_n from the parenthesis are connected via $\mathcal{L}\left(I^{(n)}\right) = \mathcal{L}\left(S_{n-1}\right)$ for $\mathcal{C} = K$ and $r = K-1$ we choose $I^{(n)}$ according to Theorem 3.3 such that the L_p-distance between $\frac{I^{(n)}}{n-1}$ and the limit D satisfies (3.3).

Lemma 3.4. *The vector of the rescaled subtree sizes* $\frac{I^{(n)}}{n} := \left(\frac{I_1^{(n)}}{n}, \dots, \frac{I_K^{(n)}}{n}\right)$ *converges almost surely to a Dirichlet-distributed random vector with all parameters equal to* $\frac{1}{K-1}$ *denoted by* $D = (D_1, \dots, D_K)$*, as* $n \to \infty$*. Furthermore, let* $I^{(n)}$ *and* D *be coupled according to Theorem 3.3 and* $p \geq 1$*. Then, for all* $r = 1, \dots, K$*, as* $n \to \infty$*,*

$$\left\| \frac{I_r^{(n)}}{n} - D_r \right\|_p = O\left(n^{-\frac{1}{2}}\right).$$

Proof. As $\left|\frac{I_r^{(n)}}{n} - \frac{I_r^{(n)}}{n-1}\right| < \frac{1}{n}$, this is an immediate consequence of Lemma 3.2 and Theorem 3.3. □

Corollary 3.5. *Let* $p \geq 2$ *and* $\psi \in (0,1)$*. Then, for* $r = 1, \dots, K$*, as* $n \to \infty$*,*

$$\left\| \left(\frac{I_r^{(n)}}{n}\right)^{\psi} - D_r^{\psi} \right\|_p = O\left(n^{-\frac{\psi}{2}}\right).$$

Proof. In the following calculation, the first inequality is due to (A.1) as a consequence of Lemma A.1 and the second inequality follows by Jensen's inequality

$$\begin{aligned}\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\psi}-D_r^{\psi}\right\|_p &= \left(\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\psi}-D_r^{\psi}\right|^p\right]\right)^{\frac{1}{p}} \leq \left(\mathbb{E}\left[\left|\frac{I_r^{(n)}}{n}-D_r\right|^{p\psi}\right]\right)^{\frac{1}{p}} \\ &\leq \left(\mathbb{E}\left[\left|\frac{I_r^{(n)}}{n}-D_r\right|^{p}\right]\right)^{\frac{\psi}{p}} = \left\|\frac{I_r^{(n)}}{n}-D_r\right\|_p^{\psi} \\ &\leq \left(\frac{A_p}{\sqrt{n}}\right)^{\psi} \leq \max\{A_p,1\}\, n^{-\frac{\psi}{2}}.\end{aligned}$$

with Lemma 3.4 and a suitable constant $A_p>0$. □

Remark 3.6. Janson showed that the correct rate for the quantities in Lemma 3.4 in all Wasserstein distances is of order $\frac{1}{n}$. So far, these results were not published.

When there are only two subtrees, i.e., $K=2$, as in the setting **Rand R**, we write $I_n := I_1^{(n)}$ and $J_n := I_2^{(n)}$. Then, both I_n and J_n are uniformly distributed on $\{0,\dots,n-1\}$. According to Lemma 3.4, the limit is Dirichlet-distributed with both parameters equal to 1, which means

$$\left(\frac{I_n}{n},\frac{J_n}{n}\right)\to(U,1-U), \quad n\to\infty, \quad \text{a.s. and in } L_p,\ p\geq 1,$$

with U uniformly distributed on $[0,1]$. Moreover, we are able to couple the subtree sizes to the limit explicitly: We pick I_n such that the pair (I_n,U) is an optimal coupling; hence, we set $I_n := \lfloor nU\rfloor$ (and therefore have $J_n = n-1-\lfloor nU\rfloor$), see Remark 2.2.

Lemma 3.7 (Twin of Lemma 3.4). *Let $p\geq 1$, U be uniformly distributed on $[0,1]$ and set $I_n := \lfloor nU\rfloor$. Then,*

$$\left\|\frac{I_n}{n}-U\right\|_p = \left(\frac{1}{p+1}\right)^{\frac{1}{p}}\frac{1}{n} = \left\|\frac{J_n}{n}-(1-U)\right\|_p.$$

Proof. We have

$$\begin{aligned}\left\|\frac{I_n}{n}-U\right\|_p^p &= \int_0^1\left|\frac{\lfloor nu\rfloor}{n}-u\right|^p \mathrm{d}u = \sum_{j=0}^{n-1}\int_{\frac{j}{n}}^{\frac{j+1}{n}}\left|\frac{j}{n}-u\right|^p \mathrm{d}u = \sum_{j=0}^{n-1}\int_{\frac{j}{n}}^{\frac{j+1}{n}}\left(u-\frac{j}{n}\right)^p \mathrm{d}u \\ &= \sum_{j=0}^{n-1}\left[\frac{1}{p+1}\left(u-\frac{j}{n}\right)^{p+1}\right]_{\frac{j}{n}}^{\frac{j+1}{n}} = \sum_{j=0}^{n-1}\frac{1}{p+1}\left(\frac{1}{n}\right)^{p+1} = \frac{1}{p+1}\left(\frac{1}{n}\right)^{p}.\end{aligned}$$

For symmetry reasons, the same approach applies to $\left\|\frac{J_n}{n} - (1-U)\right\|_p^p$. The assertion follows. $\square$

Lemma 3.8 (Twin of Corollary 3.5). *Let $p \geq 1$, $\psi \in (0,1)$, U be uniformly distributed on $[0,1]$ and set $I_n := \lfloor nU \rfloor$. Then, for $n \geq 2$,*

$$\left\| \left(\frac{I_n}{n}\right)^{\psi} - U^{\psi} \right\|_p, \left\| \left(\frac{J_n}{n}\right)^{\psi} - (1-U)^{\psi} \right\|_p \leq \begin{cases} \left(2 + \frac{1}{p(\psi-1)+1}\right)^{\frac{1}{p}} n^{-1}, & \psi > \frac{p-1}{p}, \\ \frac{(3\ln(n))^{\frac{1}{p}}}{n}, & \psi = \frac{p-1}{p}, \\ \left(1 + \frac{p(\psi-1)}{p(\psi-1)+1}\right)^{\frac{1}{p}} n^{-\left(\psi + \frac{1}{p}\right)}, & \psi < \frac{p-1}{p}. \end{cases}$$

Proof. We start with

$$\begin{aligned} \text{(3.7)} \qquad \mathbb{E}\left[\left|\left(\frac{I_n}{n}\right)^{\psi} - U^{\psi}\right|^p\right] &= \mathbb{E}\left[\left|\left(\frac{\lfloor nU \rfloor}{n}\right)^{\psi} - U^{\psi}\right|^p\right] = \int_0^1 \left|\left(\frac{\lfloor nu \rfloor}{n}\right)^{\psi} - u^{\psi}\right|^p \mathrm{d}u \\ &= \sum_{j=0}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left|\left(\frac{j}{n}\right)^{\psi} - u^{\psi}\right|^p \mathrm{d}u \\ &= \int_0^{\frac{1}{n}} u^{p\psi} \mathrm{d}u + \sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left|\left(\frac{j}{n}\right)^{\psi} - u^{\psi}\right|^p \mathrm{d}u. \end{aligned}$$

Obviously,

$$\text{(3.8)} \qquad \int_0^{\frac{1}{n}} u^{p\psi} \mathrm{d}u = \frac{1}{p\psi+1} n^{-(p\psi+1)} \leq n^{-(p\psi+1)}.$$

To simplify the second part, we first study $\left|\left(\frac{j}{n}\right)^{\psi} - u^{\psi}\right|$: Observe that the first derivative $x \mapsto \psi x^{\psi-1}$ of $x \mapsto x^{\psi}$ is strictly decreasing. According to the Mean value Theorem, for $u \in \left[\frac{j}{n}, \frac{j+1}{n}\right]$ there is $\xi \in \left(\frac{j}{n}, \frac{j+1}{n}\right)$ such that

$$\left|\left(\frac{j}{n}\right)^{\psi} - u^{\psi}\right| = \left|\psi \xi^{\psi-1}\left(\frac{j}{n} - u\right)\right| \leq \left|\psi \left(\frac{j}{n}\right)^{\psi-1} \frac{1}{n}\right| \leq \psi j^{\psi-1} n^{-\psi}.$$

Therefore, we have

$$\sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n}\right)^{\psi} - u^{\psi} \right|^{p} \mathrm{d}u \leq \sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \psi j^{\psi-1} n^{-\psi} \right|^{p} \mathrm{d}u$$

$$= \psi^{p} n^{-p\psi} \sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} j^{p(\psi-1)} \mathrm{d}u = \psi^{p} n^{-p\psi-1} \sum_{j=1}^{n-1} j^{p(\psi-1)}. \tag{3.9}$$

Now, the behaviour of $\sum_{j=1}^{n-1} j^{p(\psi-1)}$ is to be determined:

In the case $p(\psi-1) \neq -1$, that is $\psi \neq \frac{p-1}{p}$, the sum can be estimated by the integral

$$\sum_{j=1}^{n-1} j^{p(\psi-1)} = 1 + \sum_{j=2}^{n-1} j^{p(\psi-1)} \leq 1 + \int_{1}^{n} x^{p(\psi-1)} \mathrm{d}x \tag{3.10}$$

$$= 1 + \frac{1}{p(\psi-1)+1} \left(n^{p(\psi-1)+1} - 1 \right).$$

In the subcase $p(\psi-1) > -1$, that is $\psi > \frac{p-1}{p}$, combining (3.9) and (3.10) yields

$$\sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n}\right)^{\psi} - u^{\psi} \right|^{p} \mathrm{d}u \leq \psi^{p} n^{-p\psi-1} \left(1 + \frac{1}{p(\psi-1)+1} n^{p(\psi-1)+1} \right).$$

Adding (3.8) to the above expression, we have

$$\mathbb{E}\left[\left| \left(\frac{I_n}{n}\right)^{\psi} - U^{\psi} \right|^{p} \right] \leq n^{-(p\psi+1)} + \psi^{p} n^{-p\psi-1} \left(1 + \frac{1}{p(\psi-1)+1} n^{p(\psi-1)+1} \right)$$

$$\leq 2n^{-p\psi-1} + \frac{1}{p(\psi-1)+1} n^{-p}$$

$$\leq \left(2 + \frac{1}{p(\psi-1)+1} \right) n^{-p}. \tag{3.11}$$

In the other subcase $p(\psi-1) < -1$, that is $\psi < \frac{p-1}{p}$, (3.9) and (3.10) lead to

$$\sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n}\right)^{\psi} - u^{\psi} \right|^{p} \mathrm{d}u \leq \frac{p(\psi-1)}{p(\psi-1)+1} n^{-p\psi-1}.$$

Hence, with (3.8), we have

$$\mathbb{E}\left[\left| \left(\frac{I_n}{n}\right)^{\psi} - U^{\psi} \right|^{p} \right] \leq n^{-(p\psi+1)} + \frac{p(\psi-1)}{p(\psi-1)+1} n^{-p\psi-1}$$

$$= \left(1 + \frac{p(\psi-1)}{p(\psi-1)+1}\right) n^{-(p\psi+1)}. \tag{3.12}$$

Finally, in the other case $p(\psi - 1) = -1$, that is $\psi = \frac{p-1}{p}$, logarithmic terms enter the calculation:

$$\sum_{j=1}^{n-1} j^{p(\psi-1)} = \sum_{j=1}^{n-1} j^{-1} \leq 1 + \sum_{j=2}^{n-1} \int_{j-1}^{j} \frac{1}{x} \mathrm{d}x = 1 + \ln(n-1),$$

yielding together with (3.7), (3.8) and (3.9)

$$\mathbb{E}\left[\left|\left(\frac{I_n}{n}\right)^{\frac{p-1}{p}} - U^{\frac{p-1}{p}}\right|\right] \leq n^{-p} + \left(\frac{p-1}{p}\right)^p n^{-p}(\ln(n-1)+1) \leq 3\frac{\ln(n)}{n^p}. \tag{3.13}$$

Combining each of (3.11), (3.12), (3.13), we have

$$\mathbb{E}\left[\left|\left(\frac{I_n}{n}\right)^{\psi} - U^{\psi}\right|^p\right] \leq \begin{cases} \left(2 + \frac{1}{p(\psi-1)+1}\right) n^{-p}, & \psi > \frac{p-1}{p}, \\ 3\frac{\ln(n)}{n^p}, & \psi = \frac{p-1}{p}, \\ \left(1 + \frac{p(\psi-1)}{p(\psi-1)+1}\right) n^{-(p\psi+1)}, & \psi < \frac{p-1}{p}. \end{cases}$$

Dealing with $\left\|\left(\frac{J_n}{n}\right)^{\psi} - (1-U)^{\psi}\right\|$ accordingly, the assertion follows. □

Remark 3.9. As already stated in Remark 2.2, Lemma 3.7 implies, for $p \geq 1$,

$$\ell_p\left(\frac{I_n}{n}, U\right) = \left(\frac{1}{p+1}\right)^{\frac{1}{p}} \frac{1}{n}.$$

The choice of $I_n := \lfloor nU \rfloor$ enabled us to compute the distances of interest in Lemma 3.7 and Lemma 3.8, respectively, significantly better than in the corresponding situation of Lemma 3.4 and Corollary 3.5 — yielding the same rate as Janson's, noted in Remark 3.6.

4. The Contraction Method

This chapter serves to display the reasoning of the contraction method as applied in this thesis. A full account on capabilities and applications of the contraction method is by no means given. However, the reader is referred to the respective literature.

The contraction method is usually applied to derive limit theorems for quantities that can be described by a distributional recursive equation. It makes use of probability metrics and thereby entails the possibility of determining rates of convergence.

As this thesis deals with systems of distributional recursions the following paragraph is tailored to this setting after a rather general recap.

Application of the contraction method in this context involves the derivation of a system of fixed-point equations. Theorems on existence and uniqueness of solutions thereof are recalled from Knape and Neininger [27]. Finally, the basic idea of how rates of convergence are derived in this thesis is sketched.

4.1. A System of Distributional Recursions: Existence and Uniqueness of a Fixed-Point

Now that we captured the number of black balls recursively, see systems (3.1) and (3.2), the contraction method comes into play. The contraction method dates back to Rösler's treatment of Quicksort in [50] and seminal work [51], where the contraction method in the context of the Wasserstein metric ℓ_2 is developed. Since then the contraction method has been widely studied and extensively developed: In Rachev and Rüschendorf [48] and Rösler and Rüschendorf [53] the Zolotarev metric was introduced to the contraction method. Thereby, the range of problems to be studied via the contraction method opened up significantly due to the fact that certain sorts of distributional recursions cannot be tackled with Wasserstein metrics, cf. Neininger and Rüschendorf [41] as well as Neininger and Rüschendorf [42]. The contraction method became a powerful tool in the analysis of recursive algorithms and recursive structures in general, cf. Cramer and Rüschendorf [11], Rösler [52], Neininger [39], Janson and Neininger

[24], Drmota et al. [13] as well as Neininger and Sulzbach [43]. In Knape and Neininger [27] the range of problems to be treated with the contraction method was extended to Pólya urn schemes.

In general when working with the contraction method one deals with a quantity that can be described recursively on the level of distributions and the following steps are carried out:

1. At first, the quantity of interest has to be normalised appropriately. Usually, it is centred around its mean and scaled by (the order of) the reciprocal of the standard deviation. Hence, the first two moments are to be studied beforehand.

2. The recurrence for the quantity of interest leads to a recurrence for the normalised quantity.

3. From the shape of the recurrence for the normalised quantity a fixed-point equation as limiting equation can be guessed. Then, this fixed-point equation can be considered a self-map of the space of probability measures.

4. Next, one aims to prove existence and uniqueness of a solution of the fixed-point equation via the Banach fixed-point Theorem applied to the associated self-map. Therefore, metric and subspace need to be reasonably chosen in order to deal with a complete metric space.

5. Finally, convergence in distribution of the sequence of the normalised quantity to the fixed-point is established via studying the distance between sequence and limit in the chosen metric.

The goal is to obtain a limit theorem for the normalised quantity from these five steps. Therefore, the metric is to be chosen such that convergence in this metric implies weak convergence.

In the setting of Pólya urns, we work with a system of recurrences and consequently obtain a system of fixed-point equations. The operator associated with this system of fixed-point equations is a self-map of the Cartesian product of the space of probability measures with itself.

At this point, the reasoning for the fourth step of the contraction method in the case of a system of recursions is briefly sketched in general, i.e., how to choose the subspace and metric appropriately in order to obtain existence and uniqueness of a fixed-point for a system of distributional recursions.

Consider $\mathfrak{T}$ (real-valued) random variables $Y_n^{[1]}, \ldots, Y_n^{[\mathfrak{T}]}$ such that the corresponding normalised quantities $X_n^{[1]}, \ldots, X_n^{[\mathfrak{T}]}$ satisfy a system of distributional recursions where the normalised quantities are defined as follows: Let $s_{[k]} : \mathbb{N} \to \mathbb{R}^+$, $j \mapsto s_{[k]}(j)$ such that $s_{[k]}(j) = \Theta\left(\sqrt{\operatorname{Var}\left(Y_j^{[k]}\right)}\right)$ and $e_{[k]} : \mathbb{N} \to \mathbb{R}$, $j \mapsto \mathbb{E}\left[Y_j^{[k]}\right]$ for $k = 1, \ldots, \mathfrak{T}$. Then, $Y_n^{[k]}$ is to be centred around its mean and scaled by (a term of the order of) its standard deviation, hence

$$X_n^{[k]} := \frac{Y_n^{[k]} - \mathbb{E}\left[Y_n^{[k]}\right]}{s_{[k]}(n)}$$

for $n \geq n_0$ for some large enough n_0. We call $X_j^{[k]}$ a quantity of type k of size j, where $k \in \{1, \ldots, \mathfrak{T}\}$ and $j \in \mathbb{N}_0$. The fact that the normalised quantities are to satisfy a system of distributional recursions manifests as follows: Any quantity $X_n^{[k]}$ of type k and size n can be decomposed (in distribution) into $\mathfrak{B}$ contributions coming from the normalised quantities where each of them belongs to one of the $\mathfrak{T}$ types and are of a (random) size strictly smaller than n. We denote by $J^{(n)} := \left(J_1^{(n)}, \ldots, J_{\mathfrak{B}}^{(n)}\right)$ the vector containing these sizes.

For example, in both settings of Pólya urn schemes that we study, the types are given by the number of colours, so $\mathfrak{T} = 2$. In setting **Det R**, the number of black balls at time n is determined by K contributions coming from the K subtrees of the associated tree, i.e., $\mathfrak{B} = K$. In setting **Rand R**, the contributions come from two subtrees, so $\mathfrak{B} = 2$, even if the type of one of the two subtrees is random.

By $\pi := (\pi(k, r))$, $1 \leq k \leq \mathfrak{T}$, $1 \leq r \leq \mathfrak{B}$, we denote a $\mathfrak{T} \times \mathfrak{B}$-matrix with all entries $\pi(k, r) \in \{1, \ldots, \mathfrak{T}\}$. The matrix π carries the information on the type of these contributions; the rows of π indicate to which type each of the $\mathfrak{B}$ contributions belongs. Then, the system of distributional recursions looks as follows:

$$\begin{aligned} X_n^{[1]} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{\mathfrak{B}} A_r^{(n),[\pi(1,r)]} X_{J_r^{(n)}}^{[\pi(1,r)],(r)} + t_{[1]}^{(n)}, \\ &\vdots \\ X_n^{[\mathfrak{T}]} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{\mathfrak{B}} A_r^{(n),[\pi(\mathfrak{T},r)]} X_{J_r^{(n)}}^{[\pi(\mathfrak{T},r)],(r)} + t_{[\mathfrak{T}]}^{(n)}, \end{aligned} \tag{4.1}$$

where for $k = 1, \ldots, \mathfrak{T}$ and $r = 1, \ldots, \mathfrak{B}$

$$A_r^{(n),[\pi(k,r)]} := \frac{s_{[\pi(k,r)]}\left(J_r^{(n)}\right)}{s_{[k]}(n)},$$

$$t_{[k]}^{(n)} := \frac{1}{s_{[k]}(n)} \left(-e_{[k]}(n) + \sum_{r=1}^{\mathfrak{B}} e_{[\pi(k,r)]}\left(J_r^{(n)}\right)\right),$$

where $X_j^{[k],(r)}$ is an independent copy of $X_j^{[k]}$, $0 \le j \le n-1$, $1 \le k \le \mathfrak{T}$ such that the families $\left(X_j^{[k],(r)}\right)_{0\le j\le n-1}$ for $k = 1,\dots,\mathfrak{T}$, $r = 1,\dots,\mathfrak{B}$ are independent and also independent of $J^{(n)}$.

Given that the coefficients and the components of $J^{(n)}$ grow with $n \to \infty$ and $A_r^{(n),[k]}$ as well as the toll terms $t_{[k]}^{(n)}$ converge in a reasonable sense to some $A^{[k]}$ and $t_{[k]}$, respectively, the recursions of the above system suggest that a limit $\left(X^{[1]},\dots,X^{[\mathfrak{T}]}\right)$ of $\left(X_n^{[1]},\dots,X_n^{[\mathfrak{T}]}\right)$ satisfies the following system of fixed-point equations:

$$
\begin{aligned}
X^{[1]} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{\mathfrak{B}} A^{[\pi(1,r)]} X^{[\pi(1,r)],(r)} + t_{[1]},\\
&\ \vdots \\
X^{[\mathfrak{T}]} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{\mathfrak{B}} A^{[\pi(\mathfrak{T},r)]} X^{[\pi(\mathfrak{T},r)],(r)} + t_{[\mathfrak{T}]}
\end{aligned}
\tag{4.2}
$$

with similar conditions on distributions and independence as in system (4.1).

The system of fixed-point equations (4.2) translates into a selfmap $\mathcal{T}$ of the Cartesian product of the space of probability measures. Let $A := (A_{kr})$ be a $\mathfrak{T} \times \mathfrak{B}$-matrix of real-valued random variables and $t := (t_1,\dots,t_{\mathfrak{T}})$ be a vector of real-valued random variables. Then, system (4.2) is associated to the following mapping $\mathcal{T}$, with suitable A and t:

$$
\begin{aligned}
&\mathcal{T} : (\mathcal{M})^{\times\mathfrak{T}} \to (\mathcal{M})^{\times\mathfrak{T}}\\
&(\mu_1,\dots,\mu_{\mathfrak{T}}) \mapsto \left(\mathcal{T}_1(\mu_1,\dots,\mu_{\mathfrak{T}}),\dots,\mathcal{T}_{\mathfrak{T}}(\mu_1,\dots,\mu_{\mathfrak{T}})\right), \quad \text{where for } 1,\dots,\mathfrak{T}\\
&\mathcal{T}_k(\mu_1,\dots,\mu_{\mathfrak{T}}) := \mathcal{L}\left(\sum_{r=1}^{\mathfrak{B}} A_{kr} Z_{kr} + t_k\right)
\end{aligned}
\tag{4.3}
$$

with $(A_{k1},\dots,A_{k\mathfrak{B}},t_k), Z_{k1},\dots,Z_{k\mathfrak{B}}$ independent and $\mathcal{L}(Z_{kr}) = \mu_{\pi(k,r)}$ for $k = 1,\dots,\mathfrak{T}$ and $r = 1,\dots,\mathfrak{B}$. We write $\mathcal{T}_k(\mu) := \mathcal{T}_k(\mu_1,\dots,\mu_{\mathfrak{T}})$, $k = 1,\dots,\mathfrak{T}$, and $\mathcal{T}\mu := (\mathcal{T}_1(\mu),\dots,\mathcal{T}_{\mathfrak{T}}(\mu))$. The equation $\mathcal{T}\mu = \mu$ summarises (4.2).

The following two theorems state how to choose a subspace of $(\mathcal{M})^{\times\mathfrak{T}}$ in order to obtain a unique fixed-point of (4.3). The proofs will be carried out in the metric that later will serve to derive rates of convergence. Of course, to prove only existence and uniqueness of a fixed-point other choices of the metrics could do as well.

Both theorems are borrowed from Knape and Neininger [27]. Their proofs in [27] are performed with the use of the Zolotarev metric. We will re-proof Theorem 4.1 in the "classical" way with the maximal Wasserstein distance $\ell_2^{\vee}$ and give a proof of Theorem 4.2 as in [27]. Both proofs later serve to understand why we use different metrics depending on whether we are in the non-normal limit case or in the normal limit case of our urns **Det R** and **Rand R**.

Theorem 4.1 (Theorem 5.1 in [27]). *Let A_{kr} and t_k of (4.3) be L_2-integrable with*

$$\max_{1\leq k\leq \mathfrak{T}} \mathbb{E}\left[\sum_{r=1}^{\mathfrak{B}} (A_{kr})^2\right] < 1 \quad \textit{and} \quad \mathbb{E}\left[t_k\right] = 0 \tag{4.4}$$

for all $k = 1, \ldots, \mathfrak{T}$ and $r = 1, \ldots, \mathfrak{B}$. Then, the restriction of $\mathcal{T}$ to $(\mathcal{M}_2(0))^{\times\mathfrak{T}}$ has a unique fixed-point.

Proof. First of all, $\mu \in (\mathcal{M}_2(0))^{\times\mathfrak{T}} \Rightarrow \mathcal{T}\mu \in (\mathcal{M}_2(0))^{\times\mathfrak{T}}$: By the independence conditions of (4.3) and due to L_2-integrability of all occurring quantities, we have $\mathcal{T}\mu \in (\mathcal{M}_2)^{\times\mathfrak{T}}$. Using $\mathbb{E}[t_k] = 0$, it follows $\mathcal{T}\mu \in (\mathcal{M}_2(0))^{\times\mathfrak{T}}$.

Now, it is shown that $\mathcal{T}$ restricted to $(\mathcal{M}_2(0))^{\times\mathfrak{T}}$ is a contraction with respect to the maximal Wasserstein metric $\ell_2^{\vee}$. As $\left((\mathcal{M}_2(0))^{\times\mathfrak{T}}, \ell_2^{\vee}\right)$ is a complete metric space, the existence and uniqueness of the fixed-point follows with the Banach fixed-point Theorem.

Let $\mu, \nu \in (\mathcal{M}_2(0))^{\times\mathfrak{T}}$. We choose random variables $Y_{k1}, \ldots, Y_{k\mathfrak{B}}$ and $Z_{k1}, \ldots, Z_{k\mathfrak{B}}$ such that for $k = 1, \ldots, \mathfrak{T}$, $r = 1, \ldots, \mathfrak{B}$,

- $\mathcal{L}(Y_{kr}) = \mu_{\pi(k,r)}$ and $\mathcal{L}(Z_{kr}) = \nu_{\pi(k,r)}$;
- the pair (Y_{kr}, Z_{kr}) is an optimal coupling of the laws $\mu_{\pi(k,r)}$ and $\nu_{\pi(k,r)}$;
- Y_{kr} and Z_{ks} are independent for $r \neq s$;
- $(A_{k1}, \ldots, A_{k\mathfrak{B}}, t_k), Y_{k1}, \ldots, Y_{k\mathfrak{B}}$ are independent, and
- $(A_{k1}, \ldots, A_{k\mathfrak{B}}, t_k), Z_{k1}, \ldots, Z_{k\mathfrak{B}}$ are independent.

Then, we have, for all $k = 1, \ldots, \mathfrak{T}$,

$$\mathcal{L}\left(\sum_{r=1}^{\mathfrak{B}} A_{kr}Y_{kr} + t_k\right) = \mathcal{T}_k(\mu), \quad \mathcal{L}\left(\sum_{r=1}^{\mathfrak{B}} A_{kr}Z_{kr} + t_k\right) = \mathcal{T}_k(\nu)$$

and further

$$\begin{aligned}&(\ell_2(\mathcal{T}_k(\mu), \mathcal{T}_k(\nu)))^2\\ \leq &\left\|\sum_{r=1}^{\mathfrak{B}} A_{kr}(Y_{kr} - Z_{kr})\right\|_2^2\end{aligned}$$

$$(4.5) \qquad = \sum_{r=1}^{\mathfrak{B}} \mathbb{E}\left[(A_{kr})^2 (Y_{kr} - Z_{kr})^2\right] + \sum_{r \neq s} \mathbb{E}\left[A_{kr}(Y_{kr} - Z_{kr}) A_{ks}(Y_{ks} - Z_{ks})\right]$$

$$(4.6) \qquad = \sum_{r=1}^{\mathfrak{B}} \mathbb{E}\left[(A_{kr})^2\right] \|Y_{kr} - Z_{kr}\|_2^2 = \sum_{r=1}^{\mathfrak{B}} \mathbb{E}\left[(A_{kr})^2\right] \left(\ell_2\left(\mu_{\pi(k,r)}, \nu_{\pi(k,r)}\right)\right)^2$$

$$\leq \sum_{r=1}^{\mathfrak{B}} \mathbb{E}\left[(A_{kr})^2\right] \left(\ell_2^{\vee}(\mu,\nu)\right)^2 \leq \max_{1 \leq k \leq \mathfrak{T}} \sum_{r=1}^{\mathfrak{B}} \mathbb{E}\left[(A_{kr})^2\right] \left(\ell_2^{\vee}(\mu,\nu)\right)^2.$$

In (4.5) we have $\sum_{r \neq s} \mathbb{E}\left[A_{kr}(Y_{kr} - Z_{kr}) A_{ks}(Y_{ks} - Z_{ks})\right] = 0$ due to independence and the fact $\mathbb{E}[Y_{kr}] = \mathbb{E}[Z_{kr}] = 0$. Line (4.6) follows with independence and the condition on (Y_{kr}, Z_{kr}) being an optimal coupling of $\mu_{\pi(k,r)}$ and $\nu_{\pi(k,r)}$. This leads to

$$\ell_2^{\vee}(\mathcal{T}\mu, \mathcal{T}\nu) \leq \sqrt{\max_{1 \leq k \leq \mathfrak{T}} \sum_{r=1}^{\mathfrak{B}} \mathbb{E}\left[(A_{kr})^2\right]} \ell_2^{\vee}(\mu,\nu).$$

As $\sqrt{\max_{1 \leq k \leq \mathfrak{T}} \sum_{r=1}^{\mathfrak{B}} \mathbb{E}\left[(A_{kr})^2\right]} < 1$ due to (4.4), it follows that the mapping $\mathcal{T}$ restricted to $(\mathcal{M}_2(0))^{\times \mathfrak{T}}$ is a contraction with respect to $\ell_2^{\vee}$ and the assertion follows. □

Note that the result of Theorem 4.1 can also be derived with the help of the maximal Zolotarev distance $\zeta_2^{\vee}$ (as in [27, Theorem 5.1]). We decided to work with the maximal Wasserstein distance $\ell_2^{\vee}$ for it is possible to deduce rates in other metrics on that base. Moreover, due to Lemma 2.7, upper bounds in $\ell_2^{\vee}$ for the rate of convergence imply upper bounds in $\zeta_2^{\vee}$.

Theorem 4.2 (Theorem 5.2 in [27])**.** *Given the situation of* (4.3)*, let A_{kr} be $L_{2+\varepsilon}$-integrable and $t_k = 0$ for $k = 1, \dots, \mathfrak{T}$ and $r = 1, \dots, \mathfrak{B}$ with*

$$(4.7) \qquad \sum_{r=1}^{\mathfrak{B}} (A_{kr})^2 = 1 \quad \textit{almost surely}, \quad \textit{for all } k = 1, \dots, \mathfrak{T}, \textit{ and}$$

$$(4.8) \qquad \min_{1 \leq k \leq \mathfrak{T}} \mathbb{P}\left(\max_{1 \leq r \leq \mathfrak{B}} |A_{kr}| < 1\right) > 0.$$

Then, for all $\sigma^2 > 0$ the unique fixed-point of $\mathcal{T}$ restricted to $\left(\mathcal{M}_{2+\varepsilon}(0,\sigma^2)\right)^{\times \mathfrak{T}}$ is given by $\left(\mathcal{N}(0,\sigma^2), \dots, \mathcal{N}(0,\sigma^2)\right)$.

Proof. The proof is based on the proofs of Theorems 5.1 and 5.2 in [27]. Fix $\varepsilon > 0$ and $\sigma^2 > 0$. Let $\mu \in \left(\mathcal{M}_{2+\varepsilon}(0,\sigma^2)\right)^{\times \mathfrak{T}}$, then $T\mu \in \left(\mathcal{M}_{2+\varepsilon}(0,\sigma^2)\right)^{\times \mathfrak{T}}$: All occurring quantities are $L_{2+\varepsilon}$-integrable, hence, $T\mu$ is, too. Due to independence from (4.3), condition (4.7) and as $t_k = 0$, it follows $T\mu \in \left(\mathcal{M}_{2+\varepsilon}(0,\sigma^2)\right)^{\times \mathfrak{T}}$.

We now show that the restriction of $\mathcal{T}$ to $\left(\mathcal{M}_{2+\varepsilon}(0,\sigma^2)\right)^{\times \mathfrak{T}}$ is a contraction with respect to $\zeta_{2+\varepsilon}^{\vee}$: Let $\varepsilon \in (0,1]$ such that all occurring Zolotarev distances are finite. Let $\mu, \nu \in$

$\left(\mathcal{M}_{2+\varepsilon}\left(0,\sigma^2\right)\right)^{\times\mathfrak{T}}$. Choose random variables Y_{kr} and Z_{kr} such that for $k=1,\ldots,\mathfrak{T}$, $r=1,\ldots,\mathfrak{B}$

- $\mathcal{L}\left(Y_{kr}\right)=\mu_{\pi(k,r)}$ and $\mathcal{L}\left(Z_{kr}\right)=\nu_{\pi(k,r)}$;
- $Y_{k1},\ldots,Y_{k\mathfrak{B}},(A_{k1},\ldots,A_{k\mathfrak{B}})$ are independent, and
- $Z_{k1},\ldots,Z_{k\mathfrak{B}},(A_{k1},\ldots,A_{k\mathfrak{B}})$ are independent;

Then, for $k=1,\ldots,\mathfrak{T}$,

$$\mathcal{L}\left(\sum_{r=1}^{\mathfrak{B}}A_{kr}Y_{kr}\right)=\mathcal{T}_k\left(\mu\right)\text{ as well as }\mathcal{L}\left(\sum_{r=1}^{\mathfrak{B}}A_{kr}Z_{kr}\right)=\mathcal{T}_k\left(\nu\right).$$

By υ we denote the distribution of $(A_{k1},\ldots,A_{k\mathfrak{B}})$ and obtain via conditioning on it and with the independence conditions stated above

$$\begin{aligned}
\zeta_{2+\varepsilon}\left(T_k\left(\mu\right),T_k\left(\nu\right)\right)&=\sup_{f\in\mathcal{F}_{2+\varepsilon}}\left|\mathbb{E}\left[f\left(\sum_{r=1}^{\mathfrak{B}}A_{kr}Y_{kr}\right)-f\left(\sum_{r=1}^{\mathfrak{B}}A_{kr}Z_{kr}\right)\right]\right|\\
&=\sup_{f\in\mathcal{F}_{2+\varepsilon}}\left|\int\mathbb{E}\left[f\left(\sum_{r=1}^{\mathfrak{B}}\alpha_rY_{kr}\right)-f\left(\sum_{r=1}^{\mathfrak{B}}\alpha_rZ_{kr}\right)\right]\mathrm{d}\upsilon\left(\alpha_1,\ldots,\alpha_{\mathfrak{B}}\right)\right|\\
&\le\int\sup_{f\in\mathcal{F}_{2+\varepsilon}}\left|\mathbb{E}\left[f\left(\sum_{r=1}^{\mathfrak{B}}\alpha_rY_{kr}\right)-f\left(\sum_{r=1}^{\mathfrak{B}}\alpha_rZ_{kr}\right)\right]\right|\mathrm{d}\upsilon\left(\alpha_1,\ldots,\alpha_{\mathfrak{B}}\right) \qquad (4.9)\\
&=\int\zeta_{2+\varepsilon}\left(\sum_{r=1}^{\mathfrak{B}}\alpha_rY_{kr},\sum_{r=1}^{\mathfrak{B}}\alpha_rZ_{kr}\right)\mathrm{d}\upsilon\left(\alpha_1,\ldots,\alpha_{\mathfrak{B}}\right)\\
&\le\int\sum_{r=1}^{\mathfrak{B}}\left|\alpha_r\right|^{2+\varepsilon}\zeta_{2+\varepsilon}\left(Y_{kr},Z_{kr}\right)\mathrm{d}\upsilon\left(\alpha_1,\ldots,\alpha_{\mathfrak{B}}\right) \qquad (4.10)\\
&\le\sum_{r=1}^{\mathfrak{B}}\mathbb{E}\left[\left|A_{kr}\right|^{2+\varepsilon}\right]\zeta^{\vee}_{2+\varepsilon}\left(\mu,\nu\right)\le\max_{1\le k\le\mathfrak{T}}\sum_{r=1}^{\mathfrak{B}}\mathbb{E}\left[\left|A_{kr}\right|^{2+\varepsilon}\right]\zeta^{\vee}_{2+\varepsilon}\left(\mu,\nu\right),
\end{aligned}$$

where (4.9) follows with monotonicity and Jensen's inequality and in (4.10) it was used that $\zeta_{2+\varepsilon}$ is $(2+\varepsilon,+)$-ideal. This yields

$$\zeta^{\vee}_{2+\varepsilon}\left(\mathcal{T}\mu,\mathcal{T}\nu\right)\le\max_{1\le k\le\mathfrak{T}}\sum_{r=1}^{\mathfrak{B}}\mathbb{E}\left[\left|A_{kr}\right|^{2+\varepsilon}\right]\zeta^{\vee}_{2+\varepsilon}\left(\mu,\nu\right).$$

From (4.7) and (4.8), we have $\sum_{r=1}^{\mathfrak{B}}\mathbb{E}\left[\left|A_{kr}\right|^{2+\varepsilon}\right]<1$ for all $k=1,\ldots,\mathfrak{T}$ and therefore, existence and uniqueness of a fixed-point of $\mathcal{T}$ restricted to $\left(\mathcal{M}_{2+\varepsilon}\left(0,\sigma^2\right)\right)^{\times\mathfrak{T}}$ follows with the Banach fixed-point Theorem, since this space endowed with the Zolotarev metric $\zeta^{\vee}_{2+\varepsilon}$ is a complete metric space.

It is left to understand that this fixed-point is given by the normal distribution: We study the characteristic function of $\mathcal{T}_k(\mu^\star)$ with $\mu^\star := (\mathcal{N}(0,\sigma^2), \ldots, \mathcal{N}(0,\sigma^2))$, for $k = 1, \ldots, \mathfrak{T}$. Let $X_1, \ldots, X_{\mathfrak{B}} \sim \mathcal{N}(0,\sigma^2)$ be independent and independent of $(A_{k1}, \ldots, A_{k\mathfrak{B}})$. Then, for $t \in \mathbb{R}$, we know that the characteristic function of $\mathcal{N}(0,\sigma^2)$ is given by

$$\mathbb{E}\left[\exp\left(itX_1\right)\right] = \exp\left(-\frac{t^2\sigma^2}{2}\right).$$

For the characteristic function of $\mathcal{T}_k(\mu^\star)$ we obtain due to independence and with property (4.7)

$$\begin{aligned}
\mathbb{E}\left[\exp\left(it\sum_{r=1}^{\mathfrak{B}} A_{kr}X_r\right)\right] &= \int \mathbb{E}\left[\exp\left(it\sum_{r=1}^{\mathfrak{B}} \alpha_r X_r\right)\right] \mathrm{d}\upsilon\left(\alpha_1, \ldots, \alpha_{\mathfrak{B}}\right) \\
&= \int \prod_{r=1}^{\mathfrak{B}} \mathbb{E}\left[\exp\left(it\alpha_r X_r\right)\right] \mathrm{d}\upsilon\left(\alpha_1, \ldots, \alpha_{\mathfrak{B}}\right) \\
&= \int \prod_{r=1}^{\mathfrak{B}} \exp\left(-\frac{t^2\alpha_r^2\sigma^2}{2}\right) \mathrm{d}\upsilon\left(\alpha_1, \ldots, \alpha_{\mathfrak{B}}\right) \\
&= \int \exp\left(-\frac{t^2\sigma^2}{2}\sum_{r=1}^{\mathfrak{B}} \alpha_r^2\right) \mathrm{d}\upsilon\left(\alpha_1, \ldots, \alpha_{\mathfrak{B}}\right) \\
&= \mathbb{E}\left[\exp\left(-\frac{t^2\sigma^2}{2}\sum_{r=1}^{\mathfrak{B}} \left(A_{kr}\right)^2\right)\right] \\
&= \mathbb{E}\left[\exp\left(-\frac{t^2\sigma^2}{2}\right)\right].
\end{aligned}$$

It follows that $\mathcal{T}_k(\mu^\star) = \mathcal{N}(0,\sigma^2)$ and $\mathcal{T}\mu^\star = \mu^\star$. □

Whenever we derive rates of convergence in the normal limit cases, we work with the maximal Zolotarev metric $\zeta_3^\vee$. This choice seems to fit our reasoning better, see Remark 8.3.

Moreover, in the next paragraph a short discussion of the capabilities of Wasserstein and Zolotarev metrics in the context of the contraction method with respect to our situation follows.

In Chapter 5, we shall see that the behaviour of the variance of the number of black balls will determine whether we are in the situation of Theorem 4.1 or of Theorem 4.2. That is because the scaling factor used for normalising is given by the standard deviation and this scaling factor forms the shape of the coefficients of the limiting equation. Hence, the variance is crucial for us. However, the behaviour of the variances is closely related to the ratio of the eigenvalues of the replacement matrix. Thus, the regimes of non-normal and normal limiting behaviour of the normalised number of black balls are divided up according to the ranges of the ratio of the eigenvalues.

A Note on the Metrics

As already mentioned, the shape of the coefficients of the limiting equation is crucial. This paragraph serves to clarify the choices of different metrics in Theorems 4.1 and 4.2.

The origin of the contraction method makes use of Wasserstein distances, most notably ℓ_2. However, it turned out that there are distributional fixed-point equations that do not fit into the contraction setting when choosing ℓ_2: Let X and Y be independent random variables with $X \stackrel{d}{=} Y$ and consider the distributional fixed-point equation

$$X \stackrel{d}{=} \frac{1}{\sqrt{2}} X + \frac{1}{\sqrt{2}} Y.$$

We try to apply Theorem 4.1 and observe that the squared coefficients sum up to one and, thus, Theorem 4.1 cannot be applied. Even worse, it is easy to check that by the convolution property of the normal distribution any normal distribution $\mathcal{N}\left(0, \sigma^2\right)$ with $\sigma > 0$ solves this distributional fixed-point equation. Therefore, there is no unique fixed-point in $\mathcal{M}_2\left(0\right)$ (or $\mathcal{M}_p\left(0\right)$). Whenever the normal distribution appears as solution of such a fixed-point equation, the same problem arises. Hence, there is no way to choose a metric such that a modified version of Theorem 4.1 could yield a unique fixed-point in $\mathcal{M}_2\left(0\right)$ (or $\mathcal{M}_p\left(0\right)$) endowed with the respective Wasserstein metric.

Thus, the subspace has to be shrunk such that there is a unique fixed-point and a metric has to be chosen such that the coefficients turn the associated mapping into a contraction: That is where the Zolotarev metric appeared as knight in shining armour. Working with the metric spaces $\left(\mathcal{M}_{2+\varepsilon}\left(0, \sigma^2\right), \zeta_{2+\varepsilon}\right)$ analogously to Theorem 4.2 extended the contraction method to such fixed-point equations. The benefit of the Zolotarev distances ζ_s is that they are $(s, +)$-ideal forcing the coefficients to sum up to a number less than one when equipped with exponent $s \geq 2$. Unfortunately, this flexibility comes at a price: The higher s of ζ_s is chosen, the more knowledge is needed about the moments of the quantities in play.

There are further restrictions to the Wasserstein setting that are not discussed here as they do not contribute to the reasoning in this thesis. These problems and more details about the benefits of when to choose which metric can be found in Rachev and Rüschendorf [48], Rösler and Rüschendorf [53], Neininger and Rüschendorf [41] as well as Neininger and Rüschendorf [42].

In the non-normal limit case, the Wasserstein distance ℓ_2 works out and opens the door to transferring the obtained rates not only to ℓ_p but also to the Kolmogorov-Smirnov distance ϱ. Hence, we stick to the original approach. In the normal limit case, Wasserstein metrics do not pave the road to success and therefore we use the Zolotarev metric ζ_3.

4.2. Rates of Convergence via the Contraction Method

The fifth step of the contraction method, establishing convergence to the fixed-point, grants access to the opportunity of performing explicit estimates such that establishing weak convergence goes together with (upper bounds of) rates of convergence in the metric that is chosen in the forth step.

Naturally, exploiting the fifth step of the contraction method for the sake of rates already has found its place in literature: Cramer [12] and Cramer and Rüschendorf [11] did so in the context of the analysis of recursive algorithms. Rates of convergence for Quicksort were derived by Neininger and Rüschendorf [40] in the Zolotarev metric ζ_3 as well as by Fill and Janson [15] in the Wasserstein metrics and the Kolmogorov-Smirnov distance (based on the "big bang" of the contraction method in Rösler's study of Quicksort in [50]). Also in Mahmoud and Neininger [35] for distances in random binary search trees and in Neininger and Rüschendorf [42] in the case of degenerate limit equations, rates of convergence in the Zolotarev metric ζ_3 were stated.

Before going into details in settings **Det R** and **Rand R**, the idea all proofs have in common is sketched:

By $\mathfrak{D}$ we denote some metric on the space of probability measures, by $\mathfrak{D}^\vee$ the corresponding maximal metric on the Cartesian product of that space. We plug in random variables but keep in mind that only their distributions matter. We assume that convergence in $\mathfrak{D}$ implies weak convergence and that steps 1 to 4 of the contraction method already lie behind us.

We denote by $\mathbf{X}_n := \left(X_n^{[1]}, \ldots, X_n^{[\mathfrak{T}]}\right)$ the sequence with distributional recursions (4.1) and its limit by $\mathbf{X} := \left(X^{[1]}, \ldots, X^{[\mathfrak{T}]}\right)$ (that is unique in an appropriately chosen subspace).

The distance between the j-th member of the sequence to its limit is abbreviated by $\mathfrak{d}(j) := \mathfrak{D}^\vee(\mathbf{X}_j, \mathbf{X}) := \max_{1\le k\le \mathfrak{T}} \mathfrak{D}\left(X_j^{[k]}, X^{[k]}\right)$.

Our aim is to determine the law that governs the decrease of $\mathfrak{d}(n) = \mathfrak{D}^\vee(\mathbf{X}_n, \mathbf{X})$ in terms of some positive decreasing function $r : \mathbb{N} \to \mathbb{R}_0^+$ with $r(n) \stackrel{n\to\infty}{\longrightarrow} 0$:

$$\mathfrak{d}(n) = O(r(n)).$$

So, the function $r(n)$ captures the behaviour of the rate (for example, think of $r(n) = n^{\gamma}$ with some negative exponent γ).

The first step in order to prove this is to bound the distance between $X_n^{[k]}$ to its limit $X^{[k]}$, for $k = 1, \ldots, \mathfrak{T}$, in the following way:

$$\mathfrak{D}\left(X_n^{[k]}, X^{[k]}\right) \leq \mathbb{E}\left[\sum_{r=1}^{\mathfrak{B}} \left(A_r^{(n),[\pi(k,r)]}\right)^s \mathfrak{d}\left(J_r^{(n)}\right)\right] + \mathfrak{e}_n^{[k]} \tag{4.11}$$

where $s \geq 1$ is some exponent that stems from the metric $\mathfrak{D}$ and $\mathfrak{e}_n^{[k]}$ is some error term converging to zero.

If we can do so, we can bound $\mathfrak{d}(n)$ in terms of the coefficients occurring in the system of distributional recurrences (4.1) and the distances $\mathfrak{d}(j)$ with $j \in \{0, \ldots, n-1\}$ plus an error term $\max_{1\leq k\leq \mathfrak{T}} \mathfrak{e}_n^{[k]}$.

Describing the distance $\mathfrak{d}(n)$ with the help of the distances of the sub-sizes $J^{(n)}$ enables us to derive rates of convergence via induction: As induction hypothesis we set

$$\mathfrak{d}(j) \leq \mathcal{C}r(j) \text{ for } j = n_*, \ldots, n-1 \tag{4.12}$$

with some constant $\mathcal{C} > 0$ to be determined later and with some large enough $n_* \in \mathbb{N}$.

Next, we firstly aim to confirm that the error term tends to zero "fast" enough, i.e., for all $k = 1, \ldots, \mathfrak{T}$ it holds $\mathfrak{e}_n^{[k]} \leq \mathcal{B}r(n)$ with some constant $\mathcal{B} > 0$ independent of $\mathcal{C}$ from (4.12). Secondly, we plug the induction hypothesis (4.12) into (4.11) and want to exploit the fact that the mapping (4.3) associated to the system of fixed-point equations (4.2) is a contraction. We expect the coefficients of system (4.1) to "contract", too. That manifests in confirming

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{r=1}^{\mathfrak{B}} A_r^{(n),[\pi(k,r)]} \frac{r\left(J_r^{(n)}\right)}{r(n)}\right] < 1. \tag{4.13}$$

This property will be referred to as *contractive behaviour* of the coefficients from now on.

If done so, this yields that there is $\delta \in (0,1)$ such that for n sufficiently large we have

$$\mathfrak{D}\left(X_n^{[k]}, X^{[k]}\right) \leq (1-\delta)\mathcal{C}r(n) + \mathcal{B}r(n), \quad k = 1, \ldots, \mathfrak{T}.$$

Hence, we are able to choose $\mathcal{C}$ such that

$$\mathfrak{D}^{\vee}(\mathbf{X}_n, \mathbf{X}) \leq \mathcal{C}r(n),$$

and thereby confirm that the distances between sequence and limit decrease with a rate of order $r(n)$. This reasoning yields a limit theorem for $\mathbf{X}_n$ together with an upper bound for the rate of convergence.

One important aspect, that is hidden in the above notation, is the role of the variance: In the first step of the contraction method, the quantity of interest is centred around its mean and usually scaled by (a term of the order of) its standard deviation. Hence, the coefficients $A_r^{(n),[\pi(k,r)]}$ occurring in the distributional recursion for the normalised quantity usually are closely related to the standard deviation. Therefore, in order to prove that the distance between sequence and limit converges to zero in the chosen metric, it usually is necessary to know the behaviour of the variance. When it comes to determining rates of convergence, it is necessary to be able to know the shape of the variance in detail.

5. Rates of Convergence for a Two-Colour Pólya Urn with Deterministic Replacement

In this chapter, upper bounds for rates of convergence in setting **Det R** for the base cases are determined, i.e., the number of black balls after n steps when starting with a single ball.

At first, the details of setting **Det R** are stated and some general information on the quantities of interest as well as some enlightening insights are given. This introductory part is followed by the treatment of the non-normal limit case. The normal limit case concludes this chapter.

Recall the situation of setting **Det R**.

Balanced Irreducible Two-Colour Pólya Urns

(**Det R**)

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ with } a, d \in \mathbb{N}_0 \cup \{-1\} \text{ and } b, c \in \mathbb{N},$$

such that $a + b = c + d =: K - 1 \geq 1$ (balancedness)

and $bc > 0$ (irreducibility).

The ratio of smallest to largest eigenvalue is given by $\lambda := \frac{a-c}{a+b}$ and determines whether the normalised number of black balls admits a normal or a non-normal limiting behaviour; first, the non-normal limit case, where $\lambda > \frac{1}{2}$, is studied, then the normal limit case, where $\lambda \leq \frac{1}{2}$. According to the observations from Chapter 4, in the non-normal limit case, the Wasserstein metrics and the Kolmogorov-Smirnov distance serve to measure the distance between the sequence and its limit, whereas the normal limit case is treated with the Zolotarev distance ζ_3.

By means of the recursive approach exhibited in Chapter 3, the following distributional recursions hold for the number of black balls, recall (3.1), with $B_0^{\mathrm{b}} := 1$ and $B_0^{\mathrm{w}} := 0$ and for $n \geq 1$:

$$
\begin{aligned}
B_n^{\mathrm{b}} &\overset{\mathrm{d}}{=} \sum_{r=1}^{a+1} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=a+2}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)}, \\
B_n^{\mathrm{w}} &\overset{\mathrm{d}}{=} \sum_{r=1}^{c} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=c+1}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)}
\end{aligned}
\tag{5.1}
$$

with $B_j^{\mathrm{b},(r)} \overset{\mathrm{d}}{=} B_j^{\mathrm{b}}$, $B_j^{\mathrm{w},(r)} \overset{\mathrm{d}}{=} B_j^{\mathrm{w}}$ for $r = 1, \dots, K$ and $0 \leq j \leq n-1$ such that $\left(B_j^{\mathrm{b},(1)}\right)_{0\leq j \leq n}$, $\dots$, $\left(B_j^{\mathrm{b},(K)}\right)_{0\leq j \leq n}$, $\left(B_j^{\mathrm{w},(1)}\right)_{0\leq j \leq n}$, $\dots$, $\left(B_j^{\mathrm{b},(K)}\right)_{0\leq j \leq n}$, $I^{(n)}$ are independent.

To normalise these quantities, we are going to need the mean and the variance of the number of black balls. They will be derived in the following two lemmata.

Lemma 5.1. *For the means of the number of black balls* $\mathbb{E}\left[B_n^{\mathrm{b}}\right]$ *and* $\mathbb{E}\left[B_n^{\mathrm{w}}\right]$*, depending on* λ *it holds, as* $n \to \infty$*:*

i) $\lambda > 0$*:*

$$
\begin{aligned}
\mathbb{E}\left[B_n^{\mathrm{b}}\right] &= \frac{c\,(a+b)}{b+c}\, n + \frac{b\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\frac{1+a-c}{a+b}\right)}\, n^{\lambda} + O\left(1\right), \\
\mathbb{E}\left[B_n^{\mathrm{w}}\right] &= \frac{c\,(a+b)}{b+c}\, n - \frac{c\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\frac{1+a-c}{a+b}\right)}\, n^{\lambda} + O\left(1\right).
\end{aligned}
$$

ii) $\lambda < 0$*:*

$$
\begin{aligned}
\mathbb{E}\left[B_n^{\mathrm{b}}\right] &= \frac{c\,(a+b)}{b+c}\, n + O\left(1\right), \\
\mathbb{E}\left[B_n^{\mathrm{w}}\right] &= \frac{c\,(a+b)}{b+c}\, n + O\left(1\right).
\end{aligned}
$$

Proof. The exact expectations were derived in the proof of Bagchi and Pal [3, Lemma 1]. For the above stated asymptotic expansions, see Appendix A.2. □

Lemma 5.2. *For the variances of the number of black balls* $\mathrm{Var}\left(B_n^{\mathrm{b}}\right)$ *and* $\mathrm{Var}\left(B_n^{\mathrm{w}}\right)$*, depending on* λ *it holds, as* $n \to \infty$*:*

i) $\lambda = \frac{1}{2}$*:*

$$
\begin{aligned}
\mathrm{Var}\left(B_n^{\mathrm{b}}\right) &= bc\, n \ln\left(n\right) + O\left(n\right), \\
\mathrm{Var}\left(B_n^{\mathrm{w}}\right) &= bc\, n \ln\left(n\right) + O\left(n\right).
\end{aligned}
$$

ii) $0 < \lambda < \frac{1}{2}$:

$$\mathrm{Var}\left(B_n^{\mathrm{b}}\right) = \frac{(a+b)(a-c)^2 bc}{(a+b-2(a-c))(b+c)^2} n + O\left(n^{2\lambda}\right),$$
$$\mathrm{Var}\left(B_n^{\mathrm{w}}\right) = \frac{(a+b)(a-c)^2 bc}{(a+b-2(a-c))(b+c)^2} n + O\left(n^{2\lambda}\right).$$

iii) $\lambda < 0$:

$$\mathrm{Var}\left(B_n^{\mathrm{b}}\right) = \frac{(a+b)(a-c)^2 bc}{(a+b-2(a-c))(b+c)^2} n + O(1),$$
$$\mathrm{Var}\left(B_n^{\mathrm{w}}\right) = \frac{(a+b)(a-c)^2 bc}{(a+b-2(a-c))(b+c)^2} n + O(1).$$

Proof. For the exact computation, see the proof of Bagchi and Pal [3, Lemma 2]. To derive the asymptotic expansions, compare Appendix A.2. □

For $\lambda > \frac{1}{2}$ the variance is of the order $n^{2\lambda}$. This can be seen either from the calculations in Appendix A.2 or from the asymptotic expansion of B_n stated in Remark 5.5.

Notation 5.3. Henceforth, we abbreviate

$$c_{\mathrm{b}} := \frac{c(a+b)}{b+c}, \quad d_{\mathrm{b}} := \frac{b\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\Gamma\left(\frac{1+a-c}{a+b}\right)}, \quad d_{\mathrm{w}} := -\frac{c\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\Gamma\left(\frac{1+a-c}{a+b}\right)},$$
$$f_{\mathrm{b}} := \frac{(a+b)(a-c)^2 bc}{(a+b-2(a-c))(b+c)^2}.$$

Remark 5.4. The mean of the j-th quantities B_j^{b} and B_j^{w} is abbreviated by $\mu_{\mathrm{b}}(j)$ and $\mu_{\mathrm{w}}(j)$, respectively, as well as their standard deviations by $\sigma_{\mathrm{b}}(j)$ and $\sigma_{\mathrm{w}}(j)$, and their variances by $\sigma_{\mathrm{b}}^2(j)$ and $\sigma_{\mathrm{w}}^2(j)$, respectively. On studying rates of convergence in the setting **Det R**, we may use these abbreviations with random argument $I_r^{(n)}$. However, when e.g. $\mu_{\mathrm{b}}\left(I_r^{(n)}\right)$ appears, it does **not** equal $\mathbb{E}\left[B^{\mathrm{b}}_{I_r^{(n)}}\right]$. The quantity $\mu_{\mathrm{b}}\left(I_r^{(n)}\right)$ is a random variable mapping $I_r^{(n)}$ to the mean of the randomly chosen $I_r^{(n)}$-th quantity, i.e., we pick one of the n quantities $\mu_{\mathrm{b}}(j)$, $j = 0, \ldots, n-1$, according to the law of $I_r^{(n)}$. In contrast, the quantity $\mathbb{E}\left[B^{\mathrm{b}}_{I_r^{(n)}}\right]$ is not subjected to randomness anymore, it is a number. Of course, this applies to all the abbreviations defined here.

Remark 5.5 (Phase transition at $\lambda = \frac{1}{2}$)**.** The following representation of the number of black balls after n steps for $\lambda > \frac{1}{2}$ together with the shape of the variance from Lemma 5.2

helps to understand why the transition from weak normal to almost sure non-normal limits happens at $\lambda = \frac{1}{2}$.

The results of Janson [22] and Pouyanne [47] imply the following decomposition of the number of black balls after n steps when λ is strictly greater than $\frac{1}{2}$, with a suitable random variable B related to the almost sure limit of the normalised number of black balls, see Chauvin et al. [9]:

$$B_n \quad = \quad \tfrac{c(a+b)}{b+c} n \quad + \quad \tfrac{a+b}{b+c} n^\lambda B \quad + \quad o\left(n^\lambda\right), \quad (n \to \infty) \text{ a.s. and in } L_p.$$

Obviously, the behaviour of B_n is ruled by a linear drift accompanied by random fluctuations of order n^λ that capture the impact of the beginning of the urn process. The last term, i.e., $o\left(n^\lambda\right)$, hides the "normal" noise of randomness that contributes of order $\sqrt{n}$. As long as $\lambda > \frac{1}{2}$, the impact of the beginning of the urn process rules out the "normal" randomness. When λ hits $\frac{1}{2}$, this impact is no longer strong enough to do so and the "normal" randomness is rampant, resulting in the normal distribution as weak limit.

5.1. Non-Normal Limit Case: $\lambda > \frac{1}{2}$

The aim of this section is to derive an upper bound for the rate of convergence in the maximal Kolmogorov-Smirnov distance. On that account, at first a rate in the maximal Wasserstein distance $\ell_2^\vee$ is established via induction. After that, this rate is transferred to maximal Wasserstein distances $\ell_p^\vee$, $p \geq 1$, via another induction, using the previously derived $\ell_2^\vee$-rate as induction base. Finally, we will make use of Lemma 2.6 and convey this rate to the maximal Kolmogorov-Smirnov distance.

In the non-normal limit case, it is sufficient to scale the centred quantity by the order of the standard deviation as will be clear when the recursions for the normalised quantity are established and are considered in the light of Theorem 4.1. If $\lambda > \frac{1}{2}$, the variance is of order $n^{2\lambda}$, see Chapter A.2 of the appendix.

Let $\mathscr{X}_0 := 0 =: \mathscr{Y}_0$ and for $n \geq 1$

$$\mathscr{X}_n := \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{n^\lambda}, \quad \mathscr{Y}_n := \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{n^\lambda}. \tag{5.2}$$

The distributional recursions from system (5.1) are transferred to the normalised number of black balls after n steps, for $n \geq 1$:

$$\begin{aligned}
\mathscr{X}_n &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{X}_{I_r^{(n)}}^{(r)} + \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{Y}_{I_r^{(n)}}^{(r)} + b_{\mathrm{b}}\left(I^{(n)}\right), \\
\mathscr{Y}_n &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{X}_{I_r^{(n)}}^{(r)} + \sum_{r=c+1}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{Y}_{I_r^{(n)}}^{(r)} + b_{\mathrm{w}}\left(I^{(n)}\right)
\end{aligned} \tag{5.3}$$

with toll terms $b_{\mathrm{b}}\left(I^{(n)}\right)$ and $b_{\mathrm{w}}\left(I^{(n)}\right)$, which expand via Lemma 5.1.i) and Notation 5.3,

$$\begin{aligned}
b_{\mathrm{b}}\left(I^{(n)}\right) &:= n^{-\lambda}\left(\sum_{r=1}^{a+1} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{b}}(n)\right) \\
&= d_{\mathrm{b}}\left(-1 + \sum_{r=1}^{a+1}\left(\frac{I_r^{(n)}}{n}\right)^\lambda\right) + d_{\mathrm{w}} \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda + O\left(n^{-\lambda}\right), \\
b_{\mathrm{w}}\left(I^{(n)}\right) &:= n^{-\lambda}\left(\sum_{r=1}^{c} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=c+1}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{b}}(n)\right) \\
&= d_{\mathrm{b}} \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^\lambda + d_{\mathrm{w}}\left(-1 + \sum_{r=c+1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^\lambda\right) + O\left(n^{-\lambda}\right),
\end{aligned} \tag{5.4}$$

where $\mathscr{X}_j^{(r)} \stackrel{\mathrm{d}}{=} \mathscr{X}_j$, $\mathscr{Y}_j^{(r)} \stackrel{\mathrm{d}}{=} \mathscr{Y}_j$ for $r = 1, \dots, K$ and $0 \leq j \leq n-1$ and $\left(\mathscr{X}_j^{(1)}\right)_{0\leq j\leq n-1}, \dots,$ $\left(\mathscr{X}_j^{(K)}\right)_{0\leq j\leq n-1}$, $\left(\mathscr{Y}_j^{(1)}\right)_{0\leq j\leq n-1}, \dots,$ $\left(\mathscr{Y}_j^{(K)}\right)_{0\leq j\leq n-1}$, $I^{(n)}$ are independent.

Based on the system of recurrences (5.3) and the asymptotic behaviour of the rescaled subtree sizes $\frac{I_r^{(n)}}{n}$ studied in Lemma 3.4, we expect the following recursions to hold for a possible limit $(\mathscr{X}, \mathscr{Y})$ of $(\mathscr{X}_n, \mathscr{Y}_n)_{n\in\mathbb{N}}$, formally letting $n \to \infty$:

$$
\begin{aligned}
\mathscr{X} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} D_r^{\lambda} \mathscr{X}^{(r)} + \sum_{r=a+2}^{K} D_r^{\lambda} \mathscr{Y}^{(r)} + b_{\mathrm{b}}, \\
\mathscr{Y} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} D_r^{\lambda} \mathscr{X}^{(r)} + \sum_{r=c+1}^{K} D_r^{\lambda} \mathscr{Y}^{(r)} + b_{\mathrm{w}}
\end{aligned}
\tag{5.5}
$$

with toll terms

$$
\begin{aligned}
b_{\mathrm{b}} &:= d_{\mathrm{b}} \left(-1 + \sum_{r=1}^{a+1} D_r^{\lambda}\right) + d_{\mathrm{w}} \sum_{r=a+2}^{K} D_r^{\lambda}, \\
b_{\mathrm{w}} &:= d_{\mathrm{b}} \sum_{r=1}^{c} D_r^{\lambda} + d_{\mathrm{w}} \left(-1 + \sum_{r=c+1}^{K} D_r^{\lambda}\right)
\end{aligned}
\tag{5.6}
$$

with independent copies $\mathscr{X}^{(r)}$ of $\mathscr{X}$, $\mathscr{Y}^{(r)}$ of $\mathscr{Y}$, $r = 1, \dots, K$, and a Dirichlet-distributed random vector $(D_1, \dots, D_K)$ with all parameters equal to $\frac{1}{K-1}$ such that $\mathscr{X}^{(1)}, \dots, \mathscr{X}^{(K)}$, $\mathscr{Y}^{(1)}, \dots, \mathscr{Y}^{(K)}$, and $(D_1, \dots, D_K)$ are independent.

Moreover, we choose $I^{(n)} = \left(I_1^{(n)}, \dots, I_K^{(n)}\right)$ and $D = (D_1, \dots, D_k)$ to be coupled in such a way that Lemma 3.4 holds.

The system of distributional fixed-point equations given by (5.5) can be considered a self-map of $\mathcal{M} \times \mathcal{M}$ as explained in Chapter 4. We easily see that it fits the situation stated in Theorem 4.1 (we have $\sum_{r=1}^{K} \mathbb{E}\left[D_r^{2\lambda}\right] < 1$ by (3.6) and $\mathbb{E}[b_{\mathrm{b}}] = \mathbb{E}[b_{\mathrm{w}}] = 0$ by construction). Thus, we know that there is a unique solution of the system (5.5) in the Cartesian product of the space of centred probability measures with finite second moment $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$. We will denote this solution by $(\mathcal{L}(\Lambda_{\mathrm{b}}), \mathcal{L}(\Lambda_{\mathrm{w}}))$.

The following theorem summarises the results of this section and gives upper bounds for the rates of convergence for the normalised number of black balls in setting **Det R** in the non-normal limit case.

Theorem 5.6. *Given a Pólya urn scheme characterised by* ***Det R*** *with* $\lambda := \frac{a-c}{a+b} > \frac{1}{2}$ *and normalised numbers of black balls* $\mathscr{X}_n$ *and* $\mathscr{Y}_n$ *defined in* (5.2), *let* $p \geq 1$ *and* $\varepsilon > 0$. *Then, as*

$n \to \infty$,

$$\ell_p^{\vee}\left(\left(\mathscr{X}_n, \mathscr{Y}_n\right), \left(\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}}\right)\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right),$$
$$\varrho^{\vee}\left(\left(\mathscr{X}_n, \mathscr{Y}_n\right), \left(\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}}\right)\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right),$$

where $\left(\mathcal{L}\left(\Lambda_{\mathrm{b}}\right), \mathcal{L}\left(\Lambda_{\mathrm{w}}\right)\right)$ *denotes the unique fixed-point of the system* (5.5) *in* $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$.

The proof of Theorem 5.6 consists of three steps: At first, a rate in the maximal Wasserstein distance $\ell_2^{\vee}$ is derived in Proposition 5.8. Then, we proceed to extend the rates to the maximal Wasserstein distances ℓ_p, $p \geq 1$, in Proposition 5.13. And finally, the rate is transferred to the maximal Kolmogorov-Smirnov distance $\varrho^{\vee}$ in Proposition 5.14.

In order to derive rates of convergence in the Wasserstein distances, we make use of the existence of optimal couplings. Since this is a central assumption to our proofs it is prepended as the following remark.

Remark 5.7 (Optimal couplings)**.** A main ingredient brought about by using Wasserstein distances is the question of optimal couplings: In order to switch from $\|\cdot\|_2$ to the Wasserstein distance ℓ_2 (and later from $\|\cdot\|_p$ to Wasserstein distances ℓ_p, $p \geq 2$), we choose, according to Lemma 2.3, the random variables $\mathscr{X}_0, \ldots, \mathscr{X}_n, \Lambda_{\mathrm{b}}$ to form a set of optimal couplings of the distributions $\mathcal{L}(\mathscr{X}_0), \ldots, \mathcal{L}(\mathscr{X}_n), \mathcal{L}(\Lambda_{\mathrm{b}})$ and likewise the random variables $\mathscr{Y}_0, \ldots, \mathscr{Y}_n, \Lambda_{\mathrm{w}}$ to form a set of optimal couplings of the distributions $\mathcal{L}(\mathscr{Y}_0), \ldots, \mathcal{L}(\mathscr{Y}_n), \mathcal{L}(\Lambda_{\mathrm{w}})$. Furthermore, we pick K independent copies of these random variables, where the r-th copy of $\mathscr{X}_j$ is denoted by $\mathscr{X}_j^{(r)}$, of $\mathscr{Y}_j$ by $\mathscr{Y}_j^{(r)}$, $j = 0, \ldots, n$ and, finally, the r-th copy of Λ_{b} by $\mathscr{X}^{(r)}$ just as the r-th copy of Λ_{w} by $\mathscr{Y}^{(r)}$, $r = 1, \ldots, K$. Moreover, we choose all of these random variables to be independent of $\left(I^{(n)}, D\right)$. Hence, this implies that

(5.7) $\quad \mathscr{X}_j^{(r)}$ and $\mathscr{X}^{(r)}$ as well as $\mathscr{Y}_j^{(r)}$ and $\mathscr{Y}^{(r)}$ are optimal couplings of the respective laws for $r = 1, \ldots, K$ and $0 \leq j \leq n$,

(5.8) $\quad \left(I^{(n)}, D\right), \left(\mathscr{X}_j^{(r)}\right)_{0 \leq j \leq n}, \left(\mathscr{Y}_j^{(r)}\right)_{0 \leq j \leq n}, r = 1, \ldots, K$ are independent and

(5.9) $\quad \left(I^{(n)}, D\right), \mathscr{X}^{(r)}, \mathscr{Y}^{(r)}, r = 1, \ldots, K,$ are independent.

Note that

$$\begin{aligned} \Lambda_{\mathrm{b}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} D_r^{\lambda} \mathscr{X}^{(r)} + \sum_{r=a+2}^{K} D_r^{\lambda} \mathscr{Y}^{(r)} + b_{\mathrm{b}}, \\ \Lambda_{\mathrm{w}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} D_r^{\lambda} \mathscr{X}^{(r)} + \sum_{r=c+1}^{K} D_r^{\lambda} \mathscr{Y}^{(r)} + b_{\mathrm{w}} \end{aligned} \tag{5.10}$$

with toll terms b_{b} and b_{w} defined in (5.6).

Establishing a Rate in the Wasserstein Distance $\ell_2^\vee$

As a first step towards Theorem 5.6, an upper bound for the rate in the Wasserstein distance $\ell_2^\vee$ is established:

Proposition 5.8 (Rate of Convergence in $\ell_2^\vee$)**.** *Consider a Pólya urn scheme characterised by* ***Det R*** *where* $\lambda = \frac{a-c}{a+b} > \frac{1}{2}$*. Let* $\mathscr{X}_n$ *and* $\mathscr{Y}_n$ *as in* (5.2)*. Furthermore, let* $(\mathcal{L}(\Lambda_\mathrm{b}), \mathcal{L}(\Lambda_\mathrm{w}))$ *denote the unique solution of system* (5.5) *in* $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$ *and let* $\varepsilon > 0$*. Then, as* $n \to \infty$*,*

$$\ell_2^\vee\left((\mathscr{X}_n, \mathscr{Y}_n), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right).$$

To prove Proposition 5.8, we describe the distance $\ell_2^\vee\left((\mathscr{X}_n, \mathscr{Y}_n), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right)$ recursively in terms of the distances $\ell_2^\vee\left((\mathscr{X}_j, \mathscr{Y}_j), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right)$ with $j = 0, \ldots, n-1$. Then, we give estimates for the occurring terms of this recursive description, that do not contribute to the contractive behaviour. Finally, we confirm the rate given in Proposition 5.8 via induction.

Remark 5.9. In the course of this paragraph we will abbreviate the j-th distance and the squared j-th distance by

$$\Delta(j) := \ell_2^\vee\left((\mathscr{X}_j, \mathscr{Y}_j), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right), \quad \Delta^2(j) := \left(\ell_2^\vee\left((\mathscr{X}_j, \mathscr{Y}_j), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right)\right)^2.$$

Furthermore, both will appear with random argument $I_r^{(n)} \in \{0, \ldots, n-1\}$, which means that we pick one of the n distances $\Delta(j)$, $j = 0, \ldots, n-1$ according to the law of $I_r^{(n)}$. As to avoid confusion, it is emphasised that $\Delta\left(I_r^{(n)}\right)$ (that is still a random variable) does *not* equal $\ell_2^\vee\left(\left(\mathscr{X}_{I_r^{(n)}}, \mathscr{Y}_{I_r^{(n)}}\right), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right)$ (that is a number and does not appear in any calculation of this thesis).

We deal with the distances of the two single components accordingly and abbreviate

$$\Delta_\mathrm{b}(j) := \ell_2(\mathscr{X}_n, \Lambda_\mathrm{b}), \quad \Delta_\mathrm{b}^2(j) := (\ell_2(\mathscr{X}_n, \Lambda_\mathrm{b}))^2,$$
$$\Delta_\mathrm{w}(j) := \ell_2(\mathscr{Y}_n, \Lambda_\mathrm{w}), \quad \Delta_\mathrm{w}^2(j) := (\ell_2(\mathscr{Y}_n, \Lambda_\mathrm{w}))^2.$$

Lemma 5.10 (Recursive description of $\Delta^2(n)$)**.** *In the situation of Proposition 5.8 with assumptions as in Remark 5.7, it holds, for* $n \in \mathbb{N}$*,*

$$\begin{aligned}\Delta^2(n) \quad \leq \quad & \sum_{r=1}^{K}\left(\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda}\Delta^2\left(I_r^{(n)}\right)\right] + L^2\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^\lambda\right\|_2^2\right.\\ & \left. + 2L\,\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^\lambda\right)\right|\Delta\left(I_r^{(n)}\right)\right]\right)\end{aligned}$$

$$+ \max\left\{\left\|b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}}\right\|_2^2, \left\|b_{\mathrm{w}}\left(I^{(n)}\right) - b_{\mathrm{w}}\right\|_2^2\right\}$$

with $L := \max\left\{\|\Lambda_{\mathrm{b}}\|_2, \|\Lambda_{\mathrm{w}}\|_2\right\}$.

Proof. Without loss of generality, all calculations will be carried out for the first components $\mathscr{X}_n$ and Λ_{b}, respectively. They work out completely analogously for the second components $\mathscr{Y}_n$ and Λ_{w}. In the first step, we make use of (5.10),

$$\begin{aligned}
&\Delta_{\mathrm{b}}^2(n) = \left(\ell_2\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right)\right)^2 \\
\leq\ & \mathbb{E}\left[\left|\sum_{r=1}^{a+1}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{X}^{(r)}\right) + \sum_{r=a+2}^{K}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{Y}^{(r)}\right)\right.\right. \\
&\qquad \left.\left. + b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}}\right|^2\right] \\
=\ & \mathbb{E}\left[\left|\sum_{r=1}^{a+1}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{X}^{(r)}\right) + \sum_{r=a+2}^{K}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{Y}^{(r)}\right)\right|^2\right] \\
&+ \mathbb{E}\left[\left|b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}}\right|^2\right] \\
&+ 2\,\mathbb{E}\left[\left(\sum_{r=1}^{a+1}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{X}^{(r)}\right) + \sum_{r=a+2}^{K}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{Y}^{(r)}\right)\right)\right. \\
&\qquad \left.\cdot\left(b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}}\right)\right] \qquad (5.11) \\
=\ & \mathbb{E}\left[\left|\sum_{r=1}^{a+1}\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{X}^{(r)}\right|^2\right] + \mathbb{E}\left[\left|\sum_{r=a+2}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{Y}^{(r)}\right|^2\right] \\
&+ 2\,\mathbb{E}\left[\sum_{r=1}^{a+1}\sum_{s=a+2}^{K}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{X}^{(r)}\right)\left(\left(\frac{I_s^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(s)}_{I_s^{(n)}} - D_s^{\lambda}\mathscr{Y}^{(s)}\right)\right] \qquad (5.12) \\
&+ \left\|b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}}\right\|_2^2 \\
=\ & \sum_{r=1}^{a+1}\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{X}^{(r)}\right|^2\right] \\
&+ \sum_{r,s=1,\, r\neq s}^{a+1}\mathbb{E}\left[\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda}\mathscr{X}^{(r)}\right)\left(\left(\frac{I_s^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(s)}_{I_s^{(n)}} - D_s^{\lambda}\mathscr{X}^{(s)}\right)\right] \qquad (5.13)
\end{aligned}$$

$$+\sum_{r=a+2}^{K}\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(r)}_{I_r^{(n)}}-D_r^{\lambda}\mathscr{Y}^{(r)}\right|^2\right]$$

$$(5.14)\qquad +\sum_{r,s=a+2,\,r\neq s}^{K}\mathbb{E}\left[\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(r)}_{I_r^{(n)}}-D_r^{\lambda}\mathscr{Y}^{(r)}\right)\left(\left(\frac{I_s^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(s)}_{I_s^{(n)}}-D_s^{\lambda}\mathscr{Y}^{(s)}\right)\right]$$

$$+\left\|b_{\mathrm{b}}\left(I^{(n)}\right)-b_{\mathrm{b}}\right\|_2^2$$

$$(5.15)\qquad =\sum_{r=1}^{a+1}\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}}-D_r^{\lambda}\mathscr{X}^{(r)}\right\|_2^2+\sum_{r=a+2}^{K}\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{Y}^{(r)}_{I_r^{(n)}}-D_r^{\lambda}\mathscr{Y}^{(r)}\right\|_2^2$$

$$+\left\|b_{\mathrm{b}}\left(I^{(n)}\right)-b_{\mathrm{b}}\right\|_2^2.$$

The terms in (5.11), (5.12), (5.13) and (5.14) evaluate to 0 due to independence conditioned on $\left(I^{(n)},D\right)$, as stated in (5.8) and (5.9), and $\mathbb{E}\left[\mathscr{X}_j^{(r)}\right]=\mathbb{E}\left[\mathscr{Y}_j^{(r)}\right]=\mathbb{E}\left[\mathscr{X}^{(r)}\right]=\mathbb{E}\left[\mathscr{Y}^{(r)}\right]=0$.

We proceed by inserting a copy of the limit multiplied by the rescaled subtree size into the first part of (5.15):

$$(5.16)\qquad \left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}}-D_r^{\lambda}\mathscr{X}^{(r)}\right\|_2^2$$

$$=\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}_{I_r^{(n)}}-\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}+\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\mathscr{X}^{(r)}-D_r^{\lambda}\mathscr{X}^{(r)}\right\|_2^2$$

$$=\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left(\mathscr{X}^{(r)}_{I_r^{(n)}}-\mathscr{X}^{(r)}\right)+\mathscr{X}^{(r)}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}-D_r^{\lambda}\right)\right\|_2^2$$

$$=\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left(\mathscr{X}^{(r)}_{I_r^{(n)}}-\mathscr{X}^{(r)}\right)\right|^2\right]+\mathbb{E}\left[\left|\mathscr{X}^{(r)}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}-D_r^{\lambda}\right)\right|^2\right]$$

$$+2\,\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left(\mathscr{X}^{(r)}_{I_r^{(n)}}-\mathscr{X}^{(r)}\right)\mathscr{X}^{(r)}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}-D_r^{\lambda}\right)\right].$$

We treat the three summands separately: The first one yields

$$\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left(\mathscr{X}^{(r)}_{I_r^{(n)}}-\mathscr{X}^{(r)}\right)\right|^2\right]$$

$$=\mathbb{E}\left[\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left(\mathscr{X}^{(r)}_{I_r^{(n)}}-\mathscr{X}^{(r)}\right)\right|^2\,\middle|\,I_r^{(n)}\right]\right]$$

$$\begin{aligned}
&= \sum_{j=0}^{n-1} \mathbb{P}\left(I_r^{(n)} = j\right) \mathbb{E}\left[\left(\frac{j}{n}\right)^{2\lambda} \left|\mathscr{X}_j^{(r)} - \mathscr{X}^{(r)}\right|^2 \middle| I_r^{(n)} = j\right] \\
(5.17) \quad &= \sum_{j=0}^{n-1} \mathbb{P}\left(I_r^{(n)} = j\right) \left(\frac{j}{n}\right)^{2\lambda} \mathbb{E}\left[\left|\mathscr{X}_j^{(r)} - \mathscr{X}^{(r)}\right|^2\right] \\
(5.18) \quad &= \sum_{j=0}^{n-1} \mathbb{P}\left(I_r^{(n)} = j\right) \left(\frac{j}{n}\right)^{2\lambda} \Delta_{\mathrm{b}}^2(j) \\
&= \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta_{\mathrm{b}}^2\left(I_r^{(n)}\right)\right],
\end{aligned}$$

where (5.17) holds due to our independence assumptions in (5.8) and (5.9) and (5.18) holds due to our assumption on optimal couplings in (5.7).

For the second summand, we have because of the independence condition in (5.9)

$$\mathbb{E}\left[\left|\mathscr{X}^{(r)}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda}\right)\right|^2\right] = \|\Lambda_{\mathrm{b}}\|_2^2 \left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda}\right\|_2^2.$$

Finally, for the third summand

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \left(\mathscr{X}_{I_r^{(n)}}^{(r)} - \mathscr{X}^{(r)}\right) \mathscr{X}^{(r)} \left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda}\right)\right] \\
&\leq \mathbb{E}\left[\left|\frac{I_r^{(n)}}{n}\right|^{\lambda} \left|\mathscr{X}_{I_r^{(n)}}^{(r)} - \mathscr{X}^{(r)}\right| \left|\mathscr{X}^{(r)}\right| \left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda}\right|\right] \\
&= \sum_{j=0}^{n-1} \mathbb{P}\left(I_r^{(n)} = j\right) \mathbb{E}\left[\left|\left(\frac{j}{n}\right)^{\lambda} \left(\left(\frac{j}{n}\right)^{\lambda} - D_r^{\lambda}\right)\right| \left|\mathscr{X}^{(r)}\right| \left|\mathscr{X}_j^{(r)} - \mathscr{X}^{(r)}\right| \middle| I_r^{(n)} = j\right] \\
&= \sum_{j=0}^{n-1} \mathbb{P}\left(I_r^{(n)} = j\right) \left(\frac{j}{n}\right)^{\lambda} \mathbb{E}\left[\left|\left(\frac{j}{n}\right)^{\lambda} - D_r^{\lambda}\right| \middle| I_r^{(n)} = j\right] \mathbb{E}\left[\left|\mathscr{X}^{(r)}\right| \left|\mathscr{X}_j^{(r)} - \mathscr{X}^{(r)}\right|\right] \\
(5.19) \quad &\leq \sum_{j=0}^{n-1} \mathbb{P}\left(I_r^{(n)} = j\right) \left(\frac{j}{n}\right)^{\lambda} \mathbb{E}\left[\left|\left(\frac{j}{n}\right)^{\lambda} - D_r^{\lambda}\right| \middle| I_r^{(n)} = j\right] \|\Lambda_{\mathrm{b}}\|_2 \left\|\mathscr{X}_j^{(r)} - \mathscr{X}^{(r)}\right\|_2 \\
(5.20) \quad &= \|\Lambda_{\mathrm{b}}\|_2 \sum_{j=0}^{n-1} \mathbb{P}\left(I_r^{(n)} = j\right) \left(\frac{j}{n}\right)^{\lambda} \mathbb{E}\left[\left|\left(\frac{j}{n}\right)^{\lambda} - D_r^{\lambda}\right| \middle| I_r^{(n)} = j\right] \Delta_{\mathrm{b}}(j) \\
&= \|\Lambda_{\mathrm{b}}\|_2 \, \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Delta_{\mathrm{b}}\left(I_r^{(n)}\right) \left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda}\right|\right]
\end{aligned}$$

using the Cauchy-Schwarz inequality at (5.19), observing that $\Delta_{\mathrm{b}}\left(I_r^{(n)}\right)$ is $\sigma\left(I_r^{(n)}\right)$-measurable in the last step and with the condition on optimal couplings of (5.7) in (5.20).

Hence, we estimate for (5.16):

$$\begin{aligned}(5.21)\qquad &\left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^{\lambda} \mathscr{X}^{(r)} \right\|_2^2 \\ \leq\ & \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta_{\mathrm{b}}^2\left(I_r^{(n)}\right)\right] + \left\|\Lambda_{\mathrm{b}}\right\|_2^2 \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_2^2 \\ &+2\left\|\Lambda_{\mathrm{b}}\right\|_2 \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Delta_{\mathrm{b}}\left(I_r^{(n)}\right) \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| \right].\end{aligned}$$

The term $\left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \mathscr{Y}^{(r)}_{I_r^{(n)}} - D_r^{\lambda} \mathscr{Y}^{(r)} \right\|_2^2$ occurring in (5.15) can be treated analogously. This observation together with the estimate stated in (5.21) and letting $L := \max\left\{\left\|\Lambda_{\mathrm{b}}\right\|_2, \left\|\Lambda_{\mathrm{w}}\right\|_2\right\}$ leads to the following estimate for $\Delta_{\mathrm{b}}^2(n)$, according to (5.15),

$$\begin{aligned}\Delta_{\mathrm{b}}^2(n) \ \leq\ & \sum_{r=1}^{K}\left(\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta^2\left(I_r^{(n)}\right)\right] + L^2 \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_2^2 \right. \\ & \left. + 2L\,\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left(\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda}\right)\right| \Delta\left(I_r^{(n)}\right)\right]\right) + \left\| b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}} \right\|_2^2 .\end{aligned}$$

Pursuing the same steps, we obtain the same recursive description as an estimate for $\Delta_{\mathrm{w}}^2(n)$ and the assertion follows. $\square$

Glancing at Lemma 5.10, we see that the behaviour of the sizes of the subtrees of the associated tree, in the form of $I_r^{(n)}$, plays an important role. Hence, we will make use of the results of Section 3.2. For the first time, we will do so in the following corollary dealing with the behaviour of the toll terms.

Lemma 5.11. *For the toll terms $b_{\mathrm{b}}\left(I^{(n)}\right)$ and $b_{\mathrm{w}}\left(I^{(n)}\right)$ defined in* (5.4) *compared to b_{b} and b_{w} defined in* (5.6)*, we have, as $n \to \infty$,*

$$\max\left\{\left\| b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}} \right\|_2, \left\| b_{\mathrm{w}}\left(I^{(n)}\right) - b_{\mathrm{w}} \right\|_2\right\} = O\left(n^{-\frac{\lambda}{2}}\right).$$

Proof. Recalling Corollary 3.5 with $p = 2$ and $\psi = \lambda$ in the penultimate step, we obtain

$$\left\| b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}} \right\|_2$$

$$\begin{aligned}
&= \left\| d_{\mathrm{b}} \sum_{r=1}^{a+1} \left(\left(\frac{I_r^{(n)}}{n} \right)^{\lambda} - D_r^{\lambda} \right) + d_{\mathrm{w}} \sum_{r=a+2}^{K} \left(\left(\frac{I_r^{(n)}}{n} \right)^{\lambda} - D_r^{\lambda} \right) + O\left(n^{-\lambda}\right) \right\|_2 \\
&\leq d_{\mathrm{b}} \sum_{r=1}^{a+1} \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\lambda} - D_r^{\lambda} \right\|_2 + |d_{\mathrm{w}}| \sum_{r=a+2}^{K} \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\lambda} - D_r^{\lambda} \right\|_2 + O\left(n^{-\lambda}\right) \\
&\leq d_{\mathrm{b}} \sum_{r=1}^{a+1} C n^{-\frac{\lambda}{2}} + |d_{\mathrm{w}}| \sum_{r=a+2}^{K} C n^{-\frac{\lambda}{2}} + O\left(n^{-\lambda}\right) = O\left(n^{-\frac{\lambda}{2}}\right)
\end{aligned}$$

and likewise for $\left\| b_{\mathrm{w}}\left(I^{(n)}\right) - b_{\mathrm{w}} \right\|_2$. □

We finally confirm Proposition 5.8 via induction.

Proof of Proposition 5.8. Let $\varepsilon > 0$ and set as induction hypothesis:

$$\exists C > 0 \ \forall j \in \{1, \ldots, n-1\} : \ \Delta(j) \leq C j^{-\lambda + \frac{1}{2} + \varepsilon}. \tag{5.22}$$

Note that $\Delta(0) < \infty$ does not contribute as it is multiplied by 0, whenever it appears in the computation. We will have a closer look at the terms occurring in the recursive description of Lemma 5.10, given by

$$\begin{aligned}
\Delta^2(n) \leq \sum_{r=1}^{K} \Bigg(& \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n} \right)^{2\lambda} \Delta^2\left(I_r^{(n)}\right) \right] + L^2 \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\lambda} - D_r^{\lambda} \right\|_2^2 \\
& + 2L\, \mathbb{E}\left[\left| \left(\frac{I_r^{(n)}}{n} \right)^{\lambda} \left(\left(\frac{I_r^{(n)}}{n} \right)^{\lambda} - D_r^{\lambda} \right) \right| \Delta\left(I_r^{(n)}\right) \right] \Bigg) \\
& + \max\left\{ \left\| b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}} \right\|_2^2, \left\| b_{\mathrm{w}}\left(I^{(n)}\right) - b_{\mathrm{w}} \right\|_2^2 \right\}.
\end{aligned} \tag{5.23}$$

Firstly, we have

$$\begin{aligned}
& L^2 \sum_{r=1}^{K} \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\lambda} - D_r^{\lambda} \right\|_2^2 + \max\left\{ \left\| b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}} \right\|_2^2, \left\| b_{\mathrm{w}}\left(I^{(n)}\right) - b_{\mathrm{w}} \right\|_2^2 \right\} \\
&= O\left(n^{-\lambda}\right)
\end{aligned} \tag{5.24}$$

by Corollary 3.5, with $p = 2$, $\psi = \lambda$, and Lemma 5.11.

Secondly, we estimate, plugging in the induction hypothesis (5.22),

$$
\begin{aligned}
(5.25)\qquad & 2L\sum_{r=1}^{K}\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}-D_r^{\lambda}\right|\Delta\left(I_r^{(n)}\right)\right]\\
&\leq\; 2L\,C\,n^{-\lambda}\sum_{r=1}^{K}\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}-D_r^{\lambda}\right|\left(I_r^{(n)}\right)^{\frac{1}{2}+\varepsilon}\right]\\
(5.26)\qquad &\leq\; 2L\,C\,n^{-\lambda}\sum_{r=1}^{K}\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}-D_r^{\lambda}\right\|_2\left\|\left(I_r^{(n)}\right)^{\frac{1}{2}+\varepsilon}\right\|_2\\
(5.27)\qquad &\leq\; C\,A\,n^{-\lambda}\,n^{-\frac{\lambda}{2}}\,n^{\frac{1}{2}+\varepsilon}=C\,n^{-2\lambda+1+2\varepsilon}\,A\,n^{-\frac{1}{2}-\varepsilon+\frac{\lambda}{2}}\\
(5.28)\qquad &\leq\; C\,n^{-2\lambda+1+2\varepsilon}A\,n^{-\varepsilon}
\end{aligned}
$$

with the Cauchy-Schwarz inequality in (5.26), Corollary 3.5 ($p=2$, $\psi=\lambda$) and a suitable constant $A>0$ in (5.27) and $\frac{\lambda}{2}-\frac{1}{2}<0$ in (5.28).

Plugging (5.24) and (5.28) in (5.23) and using the induction hypothesis (5.22) we obtain with a suitable constant $B>0$

$$
\begin{aligned}
\Delta^2(n) &\leq \sum_{r=1}^{K}\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda}\Delta^2\left(I_r^{(n)}\right)\right]+Cn^{-2\lambda+1+2\varepsilon}An^{-\varepsilon}+Bn^{-\lambda}\\
&\leq C^2n^{-2\lambda}\sum_{r=1}^{K}\mathbb{E}\left[\left(I_r^{(n)}\right)^{2\lambda-2\lambda+1+2\varepsilon}\right]+Cn^{-2\lambda+1+2\varepsilon}An^{-\varepsilon}+Bn^{-\lambda}\\
&\leq C^2n^{-2\lambda+1+2\varepsilon}\sum_{r=1}^{K}\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{1+2\varepsilon}\right]+Cn^{-2\lambda+1+2\varepsilon}An^{-\varepsilon}+Bn^{-\lambda}\\
(5.29)\qquad &\leq (1-\delta)\,C^2n^{-2\lambda+1+2\varepsilon}+Cn^{-2\lambda+1+2\varepsilon}An^{-\varepsilon}+Bn^{-\lambda}\\
(5.30)\qquad &\leq (1-(\delta-\delta'))\,C^2n^{-2\lambda+1+2\varepsilon}+Bn_0^{-1+\lambda-2\varepsilon}n^{-2\lambda+1+2\varepsilon}.
\end{aligned}
$$

To establish (5.29) we use Lemma 3.4 combined with property (3.6) and conclude that

$$\sum_{r=1}^{K}\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{1+2\varepsilon}\right]\to\xi<1,\quad n\to\infty.$$

Hence, there is $0<\delta<1$ such that

$$\sum_{r=1}^{K}\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{1+2\varepsilon}\right]\leq 1-\delta$$

for n large enough. For (5.30), we choose $0 < \delta' < \delta$ such that for n large enough it holds

$$An^{-\varepsilon} < C\delta'.$$

Furthermore, we fix $n_0 \in \mathbb{N}$ such that all estimates hold for $n \geq n_0$.

Hence, the steps in (5.29) and (5.30) constitute the contractive behaviour, compare (4.13), that we need to conduct the induction, as they enable us to choose C such that

$$C^2 \geq \Delta^2(0) \vee \max\left\{\Delta^2(j)\, j^{2\lambda-1-2\varepsilon} | j = 1, \ldots, n_0\right\} \vee \frac{Bn_0^{-1+\lambda-2\varepsilon}}{\delta - \delta'}.$$

By this, we obtain

$$\Delta^2(n) \leq C^2 n^{-2\lambda+1+2\varepsilon},$$

concluding the proof of Proposition 5.8. □

Rates of Convergence in the Wasserstein Distances $\ell_p^\vee$, $p \geq 1$

This paragraph serves to transfer the rate of Proposition 5.8 to Wasserstein distances $\ell_p^\vee$ with $p \geq 1$. The transfer will be based upon the following lemma which is an easy extension of Fill and Janson [15, Lemma 3.2] that allows us to conduct an induction on p for deriving rates in $\ell_p^\vee$.

Lemma 5.12. *Let $Z_1, \ldots, Z_{K+1}$, $K \geq 2$ be independent random variables and $p \geq 2$ be integer. Then,*

$$\mathbb{E}\left[\left|\sum_{i=1}^{K+1} Z_i\right|^p\right] \leq \sum_{i=1}^{K} \mathbb{E}\left[|Z_i|^p\right] + \left(\sum_{i=1}^{K} \|Z_i\|_{p-1} + \|Z_{K+1}\|_p\right)^p.$$

Proof. Expansion of the product yields, due to the Multinomial Theorem,

$$\begin{aligned} \mathbb{E}\left[\left|\sum_{i=1}^{K+1} Z_i\right|^p\right] &\leq \mathbb{E}\left[\left(\sum_{i=1}^{K+1} |Z_i|\right)^p\right] \\ &= \mathbb{E}\left[\sum_{x_1+\ldots+x_{K+1}=p} \binom{p}{x_1, \ldots, x_{K+1}} \prod_{i=1}^{K+1} |Z_i|^{x_i}\right] \\ &= \sum_{x_1+\cdots+x_{K+1}=p} \binom{p}{x_1, \ldots, x_{K+1}} \prod_{i=1}^{K+1} \mathbb{E}\left[|Z_i|^{x_i}\right] \qquad (5.31) \end{aligned}$$

$$
\begin{aligned}
(5.32) \qquad &\leq \sum_{i=1}^{K} \mathbb{E}\left[|Z_i|^p\right] + \sum_{\substack{x_1+\ldots+x_{K+1}=p \\ x_1,\ldots,x_K\leq p-1}} \binom{p}{x_1,\ldots,x_{K+1}} \prod_{i=1}^{K} \|Z_i\|_{p-1}^{x_i} \|Z_{K+1}\|_p^{x_{K+1}} \\
&\leq \sum_{i=1}^{K} \mathbb{E}\left[|Z_i|^p\right] + \sum_{x_1+\ldots+x_{K+1}=p} \binom{p}{x_1,\ldots,x_{K+1}} \prod_{i=1}^{K} \|Z_i\|_{p-1}^{x_i} \|Z_{K+1}\|_p^{x_{K+1}} \\
&= \sum_{i=1}^{K} \mathbb{E}\left[|Z_i|^p\right] + \left(\sum_{i=1}^{K} \|Z_i\|_{p-1} + \|Z_{K+1}\|_p\right)^p
\end{aligned}
$$

using independence in (5.31), and $\mathbb{E}\left[|Z_i|^{x_i}\right] = \|Z_i\|_{x_i}^{x_i} \leq \|Z_i\|_{p-1}^{x_i}$ for $i \in \{1,\ldots,K\}$ and $x_i \leq p$ as well as $\mathbb{E}\left[|Z_{K+1}|^{x_{K+1}}\right] = \|Z_{K+1}\|_{x_{K+1}}^{x_{K+1}} \leq \|Z_{K+1}\|_p^{x_{K+1}}$ in (5.32). □

Now, we extend the statement of Proposition 5.8 providing a rate in all maximal Wasserstein distances $\ell_p^\vee$, $p \geq 1$:

Proposition 5.13. *Consider a Pólya urn scheme characterised by* ***Det R*** *where* $\lambda = \frac{a-c}{a+b} > \frac{1}{2}$. *Let* $\mathscr{X}_n$ *and* $\mathscr{Y}_n$ *be as in* (5.2). *Furthermore, let* $(\mathcal{L}(\Lambda_\mathrm{b}), \mathcal{L}(\Lambda_\mathrm{w}))$ *denote the unique fixed-point of system* (5.5) *in* $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$. *Let* $p \geq 1$ *and* $\varepsilon > 0$, *then, as* $n \to \infty$,

$$\ell_p^\vee\left((\mathscr{X}_n, \mathscr{Y}_n), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right).$$

Proof. Lemma 5.12 enables us to estimate $\ell_p^\vee\left((\mathscr{X}_n, \mathscr{Y}_n), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right)$ recursively in terms of both n and p. To prove this proposition we first conduct an induction on p. Hence, p is integer in our calculations. We use Proposition 5.8 as base case, i.e., $p = 2$. Then, the proof is concluded by another induction on n in every step $p \to p+1$. Due to monotonicity of the Wasserstein distances (i.e., $\ell_q \leq \ell_p$ for $q \leq p$, see Lemma 2.4), we subsequently conclude that the result holds for any $p \geq 1$.

Still, we set conditions on independence and optimal couplings as in Remark 5.7. According to Remark 5.9 we abbreviate

$$\Delta_q(j) := \ell_q^\vee\left((\mathscr{X}_j, \mathscr{Y}_j), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right)$$

and it is to be understood in the same way as in Remark 5.9 when we write $\Delta_q\left(I_r^{(n)}\right)$ with random argument $I_r^{(n)}$.

We decompose the distance $\ell_p\left((\mathscr{X}_n, \mathscr{Y}_n), (\Lambda_\mathrm{b}, \Lambda_\mathrm{w})\right)$ into $K+1$ summands in order to apply Lemma 5.12. Again, all calculations will be executed for the first component $\mathscr{X}_n$ since calculations for the second component $\mathscr{Y}_n$ work out analogously. We set as induction hypothesis

$$\Delta_q(j) \leq C_q j^{-\lambda+\frac{1}{2}+\varepsilon} \qquad \text{for } q = 2,\ldots,p-1,\ j = 1,\ldots,n-1,$$

(again, note that $\Delta_q(0)$ does not contribute) and, keeping (5.10) in mind, begin with

$$\begin{aligned}
&\ell_p(\mathscr{X}_n, \Lambda_\mathrm{b}) \\
&\leq \left\| \left\{ \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n} \right)^\lambda \mathscr{X}^{(r)}_{I_r^{(n)}} - D_r^\lambda \mathscr{X}^{(r)} \right\} + \left\{ \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n} \right)^\lambda \mathscr{Y}^{(r)}_{I_r^{(n)}} - D_r^\lambda \mathscr{Y}^{(r)} \right\} + b_\mathrm{b}\left(I^{(n)}\right) - b_\mathrm{b} \right\|_p \\
&=: \left\| \sum_{r=1}^{a+1} V_r + \sum_{r=a+2}^{K} V_r + V_{K+1} \right\|_p .
\end{aligned}$$

The summands $V_1, \ldots, V_{K+1}$ are independent conditioned on the random vector $\left(D, I^{(n)}\right)$; hence, conditioned on $\left(D, I^{(n)}\right)$ we can apply Lemma 5.12.

For the sake of transparency, we write $Z^{(r)} = \begin{cases} \mathscr{X}^{(r)}, & r = 1, \ldots, a+1, \\ \mathscr{Y}^{(r)}, & r = a+2, \ldots, K, \end{cases}$ and we deal analogously with $Z^{(r)}_{I_r^{(n)}}$ (representing $\mathscr{X}^{(r)}_{I_r^{(n)}}$ and $\mathscr{Y}^{(r)}_{I_r^{(n)}}$ accordingly). Furthermore, note that we have from Kuba and Sulzbach [28, Theorem 2] that Λ_b and Λ_w are Subgaussian and therefore $M_p := \max\left\{ \|\Lambda_\mathrm{b}\|_p, \|\Lambda_\mathrm{w}\|_p \right\} < \infty$, $p \geq 2$ (note that this also yields $\Delta_p(0) < \infty$).

We abbreviate $\underline{x} := (x_1, \ldots, x_K)$ and $\underline{i} = (i_1, \ldots, i_K)$ and condition on $\left(D, I^{(n)}\right)$. Let $\underline{b_\mathrm{b}}$ denote the constant version of b_b that arises by conditioning on $\left(D, I^{(n)}\right)$. Then, we have, applying Lemma 5.12 in the final step,

$$\begin{aligned}
\left\| \sum_{r=1}^{a+1} V_r + \sum_{r=a+2}^{K} V_r + V_{K+1} \right\|_p^p &= \mathbb{E}\left[\left| \sum_{r=1}^{a+1} V_r + \sum_{r=a+2}^{K} V_r + V_{K+1} \right|^p \right] \\
&= \mathbb{E}\left[\left| \left\{ \sum_{r=1}^{K} \left(\frac{I_r^{(n)}}{n} \right)^\lambda Z^{(r)}_{I_r^{(n)}} - D_r^\lambda Z^{(r)} \right\} + \left(b_\mathrm{b}\left(I^{(n)}\right) - b_\mathrm{b} \right) \right|^p \right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left| \left\{ \sum_{r=1}^{K} \left(\frac{I_r^{(n)}}{n} \right)^\lambda Z^{(r)}_{I_r^{(n)}} - D_r^\lambda Z^{(r)} \right\} + \left(b_\mathrm{b}\left(I^{(n)}\right) - b_\mathrm{b} \right) \right|^p \,\middle|\, \left(D, I^{(n)}\right) \right] \right] \\
&= \int \mathbb{E}\left[\left| \left\{ \sum_{r=1}^{K} \left(\frac{i_r}{n} \right)^\lambda Z^{(r)}_{i_r} - x_r^\lambda Z^{(r)} \right\} + \left(b_\mathrm{b}(\underline{i}) - \underline{b_\mathrm{b}} \right) \right|^p \right] \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}(\underline{x}, \underline{i}) \qquad (5.33) \\
&\leq \int \left\{ \sum_{r=1}^{K} \mathbb{E}\left[\left| \left(\frac{i_r}{n} \right)^\lambda Z^{(r)}_{i_r} - x_r^\lambda Z^{(r)} \right|^p \right] \right. \\
&\qquad \left. + \left(\sum_{r=1}^{K} \left\| \left(\frac{i_r}{n} \right)^\lambda Z^{(r)}_{i_r} - x_r^\lambda Z^{(r)} \right\|_{p-1} + \left\| b_\mathrm{b}(\underline{i}) - \underline{b_\mathrm{b}} \right\|_p \right)^p \right\} \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}(\underline{x}, \underline{i})
\end{aligned}$$

where in (5.33) the equality holds since $Z_1^{(r)}, \ldots, Z_{n-1}^{(r)}, Z^{(r)},\ r = 1, \ldots, K$ and $\left(D, I^{(n)}\right)$ are independent.

Let $q \geq 2$, we continue with

$$\begin{aligned}
\left\| \left(\frac{i_r}{n}\right)^{\lambda} Z_{i_r}^{(r)} - x_r^{\lambda} Z^{(r)} \right\|_q &= \left\| \left(\frac{i_r}{n}\right)^{\lambda} Z_{i_r}^{(r)} - \left(\frac{i_r}{n}\right)^{\lambda} Z^{(r)} + \left(\frac{i_r}{n}\right)^{\lambda} Z^{(r)} - x_r^{\lambda} Z^{(r)} \right\|_q \\
&= \left\| \left(\frac{i_r}{n}\right)^{\lambda} \left(Z_{i_r}^{(r)} - Z^{(r)}\right) + Z^{(r)} \left(\left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda}\right) \right\|_q \\
&\leq \left(\frac{i_r}{n}\right)^{\lambda} \left\| Z_{i_r}^{(r)} - Z^{(r)} \right\|_q + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| \left\| Z^{(r)} \right\|_q \\
&\leq \left(\frac{i_r}{n}\right)^{\lambda} \Delta_q\left(i_r\right) + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_q
\end{aligned}$$

using assumption (5.7) in the last step.

We note that

$$\begin{aligned}
\left\| b_{\mathrm{b}}\left(\underline{i}\right) - \underline{b_{\mathrm{b}}} \right\|_p &= \left| d_{\mathrm{b}} \sum_{r=1}^{a+1} \left(\left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda}\right) + d_{\mathrm{w}} \sum_{r=a+1}^{K} \left(\left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda}\right) + O\left(n^{-\lambda}\right) \right| \\
&\leq A_1 \sum_{r=1}^{K} \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| + O\left(n^{-\lambda}\right)
\end{aligned}$$

with a suitable constant $A_1 > 0$.

Hence, we have with the induction hypothesis in the last step

$$\begin{aligned}
&\left(\ell_p\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right)\right)^p \\
\leq\ & \int \Bigg\{ \sum_{r=1}^{K} \mathbb{E}\left[\left| \left(\frac{i_r}{n}\right)^{\lambda} Z_{i_r}^{(r)} - x_r^{\lambda} Z^{(r)} \right|^p \right] \\
&\quad + \left(\sum_{r=1}^{K} \left\| \left(\frac{i_r}{n}\right)^{\lambda} Z_{i_r}^{(r)} - x_r^{\lambda} Z^{(r)} \right\|_{p-1} + \left\| b_{\mathrm{b}}\left(\underline{i}\right) - \underline{b_{\mathrm{b}}} \right\|_p \right)^p \Bigg\} \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}\left(\underline{x}, \underline{i}\right) \\
\leq\ & \int \Bigg\{ \sum_{r=1}^{K} \left(\left(\frac{i_r}{n}\right)^{\lambda} \Delta_p\left(i_r\right) + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_p \right)^p \\
&\quad + \left(\sum_{r=1}^{K} \left[\left(\frac{i_r}{n}\right)^{\lambda} \Delta_{p-1}\left(i_r\right) + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_{p-1} \right] \right. \\
&\qquad \left. + A_1 \sum_{r=1}^{K} \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| + O\left(n^{-\lambda}\right) \right)^p \Bigg\} \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}\left(\underline{x}, \underline{i}\right)
\end{aligned}$$

$$\begin{aligned}
\leq \quad & \int \Bigg\{ \sum_{r=1}^{K} \left(\left(\frac{i_r}{n}\right)^{\lambda} \Delta_p\left(i_r\right) + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_p \right)^p \\
& \quad + \Bigg(\sum_{r=1}^{K} \left[\left(\frac{i_r}{n}\right)^{\lambda} C_{p-1} i_r^{-\lambda+\frac{1}{2}+\varepsilon} + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_{p-1} \right] \\
& \qquad + A_1 \sum_{r=1}^{K} \left| \left(\frac{i_r}{n}\right)^{\lambda} - x^{\lambda} \right| + O\left(n^{-\lambda}\right) \Bigg)^p \Bigg\} \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}\left(\underline{x}, \underline{i}\right).
\end{aligned}$$

Finally, we set as induction hypothesis for the induction on n

$$\Delta_p\left(j\right) \leq C_p j^{-\lambda+\frac{1}{2}+\varepsilon}, \quad j = 1, \ldots, n-1$$

and proceed

$$\begin{aligned}
& \left(\ell_p\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right)\right)^p \\
\leq \quad & \int \Bigg\{ \sum_{r=1}^{K} \left(\left(\frac{i_r}{n}\right)^{\lambda} \Delta_p\left(i_r\right) + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_p \right)^p \\
& \quad + \Bigg(\sum_{r=1}^{K} \left[\left(\frac{i_r}{n}\right)^{\lambda} C_{p-1} i_r^{-\lambda+\frac{1}{2}+\varepsilon} + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_{p-1} \right] \\
& \qquad + A_1 \sum_{r=1}^{K} \left| \left(\frac{i_r}{n}\right)^{\lambda} - x^{\lambda} \right| + O\left(n^{-\lambda}\right) \Bigg)^p \Bigg\} \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}\left(\underline{x}, \underline{i}\right) \\
\leq \quad & \int \sum_{r=1}^{K} \left(\left(\frac{i_r}{n}\right)^{\lambda} C_p i_r^{-\lambda+\frac{1}{2}+\varepsilon} + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_p \right)^p \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}\left(\underline{x}, \underline{i}\right) \\
& + \int \Bigg(\sum_{r=1}^{K} \left[\left(\frac{i_r}{n}\right)^{\lambda} C_{p-1} i_r^{-\lambda+\frac{1}{2}+\varepsilon} + \left| \left(\frac{i_r}{n}\right)^{\lambda} - x_r^{\lambda} \right| M_{p-1} \right] \\
& \qquad + A_1 \sum_{r=1}^{K} \left| \left(\frac{i_r}{n}\right)^{\lambda} - x^{\lambda} \right| + O\left(n^{-\lambda}\right) \Bigg)^p \mathrm{d}\mathbb{P}_{\left(D, I^{(n)}\right)}\left(\underline{x}, \underline{i}\right) \\
= \quad & \sum_{r=1}^{K} \mathbb{E}\left[\left(\frac{C_p}{n^{\lambda}} \left(I_r^{(n)}\right)^{\frac{1}{2}+\varepsilon} + \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| M_p \right)^p \right] \qquad (5.34) \\
& + \mathbb{E}\Bigg[\Bigg(\sum_{r=1}^{K} \left\{ \frac{C_{p-1}}{n^{\lambda}} \left(I_r^{(n)}\right)^{\frac{1}{2}+\varepsilon} + \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| M_{p-1} \right\} \\
& \qquad + A_1 \sum_{r=1}^{K} \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| + O\left(n^{-\lambda}\right) \Bigg)^p \Bigg].
\end{aligned}$$

For the first part of (5.34), we have a closer look on the p-th root of the summands:

$$\begin{aligned}
&\left\| C_p n^{-\lambda} \left(I_r^{(n)}\right)^{\frac{1}{2}+\varepsilon} + \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| M_p \right\|_p \\
\leq\ & C_p n^{-\lambda+\frac{1}{2}+\varepsilon} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_p \\
\leq\ & C_p n^{-\lambda+\frac{1}{2}+\varepsilon} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p A_2 n^{-\frac{\lambda}{2}} && (5.35) \\
\leq\ & \left(C_p \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p A_2 \right) n^{-\lambda+\frac{1}{2}+\varepsilon} && (5.36)
\end{aligned}$$

with Corollary 3.5 and a suitable constant $A_2 > 0$ in (5.35), and since $\frac{1}{2} < \lambda < 1$ yields $-\lambda + \frac{1}{2} + \varepsilon > -\frac{\lambda}{2}$ in (5.36).

Adding up, we obtain for the first part of (5.34):

$$\begin{aligned}
&\sum_{r=1}^{K} \mathbb{E}\left[\left(\frac{C_p}{n^{\lambda}} \left(I_r^{(n)}\right)^{\frac{1}{2}+\varepsilon} + \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| M_p \right)^p \right] \\
\leq\ & \sum_{r=1}^{K} \left(C_p \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p A_2 \right)^p n^{(-\lambda+\frac{1}{2}+\varepsilon)p}.
\end{aligned}$$

For the second part of (5.34), we have, again with Corollary 3.5, used twice, and another suitable constant $A_3 > 0$

$$\begin{aligned}
&\left\| \sum_{r=1}^{K} \left\{ C_{p-1} n^{-\lambda} \left(I_r^{(n)}\right)^{\frac{1}{2}+\varepsilon} + \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| M_{p-1} \right\} \right. \\
&\quad \left. + A_1 \sum_{r=1}^{K} \left| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right| + O\left(n^{-\lambda}\right) \right\|_p \\
\leq & \sum_{r=1}^{K} \left\{ C_{p-1} n^{-\lambda+\frac{1}{2}+\varepsilon} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_{p-1} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_p \right\} \\
&+ A_1 \sum_{r=1}^{K} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_p + O\left(n^{-\lambda}\right)
\end{aligned}$$

$$\leq \sum_{r=1}^{K} C_{p-1} n^{-\lambda+\frac{1}{2}+\varepsilon} \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p + A_3 n^{-\frac{\lambda}{2}}$$
$$\leq (KC_{p-1} + A_3)\, n^{-\lambda+\frac{1}{2}+\varepsilon}.$$

Inserting these estimates for (5.34), we have

$$(\ell_p(\mathscr{X}_n, \Lambda_{\mathrm{b}}))^p \leq n^{(-\lambda+\frac{1}{2}+\varepsilon)p} \left[\sum_{r=1}^{K} \left(C_p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p A_2 \right)^p + (KC_{p-1} + A_3)^p \right]. \tag{5.37}$$

Finally, we want to know if there is C_p such that

$$\sum_{r=1}^{K} \left(C_p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p A_2 \right)^p + (KC_{p-1} + A_3)^p \leq C_p^p.$$

Therefore, observe that

$$\left(C_p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p A_2 \right)^p = \sum_{j=0}^{p} \binom{p}{j} \left(C_p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p \right)^j (M_p A_2)^{p-j}$$
$$= \left(C_p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p \right)^p + \sum_{j=0}^{p-1} \binom{p}{j} C_p^j \underbrace{\left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^j}_{\leq 1} (M_p A_2)^{p-j}$$
$$\leq C_p^p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^p + \sum_{j=0}^{p-1} \binom{p}{j} C_p^j (M_p A_2)^{p-j}$$
$$\leq C_p^p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^p + C_p^{p-1} (M_p A_2)^{p-1} \sum_{j=0}^{p-1} \binom{p}{j} \tag{5.38}$$
$$\leq C_p^p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^p + C_p^{p-1} (M_p A_2)^{p-1}\, 2^p$$

with sufficiently large C_p and A_2 in (5.38). Hence, we estimate further

$$\sum_{r=1}^{K} \left(C_p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p + M_p A_2 \right)^p + (KC_{p-1} + A_3)^p$$

$$\leq \sum_{r=1}^{K} \left(C_p^p \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^p + C_p^{p-1} \left(M_p A_2 \right)^{p-1} 2^p \right) + \left(K C_{p-1} + A_3 \right)^p$$

$$\leq C_p^p \sum_{r=1}^{K} \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^p + A_4 \leq (1-\delta)\, C_p^p + A_4 \tag{5.39}$$

with $A_4 > K C_p^{p-1} \left(M_p A_2 \right)^{p-1} 2^p + \left(K C_{p-1} + A_3 \right)^p$. Observing that we have, as $n \to \infty$,

$$\sum_{r=1}^{K} \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^p = \sum_{r=1}^{K} \mathbb{E} \left[\left| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right|^p \right] = \sum_{r=1}^{K} \mathbb{E} \left[\left(\frac{I_r^{(n)}}{n} \right)^{p\left(\frac{1}{2}+\varepsilon\right)} \right] \to \xi < 1 \tag{5.40}$$

due to Lemma 3.4 and property (3.6), with $p\left(\frac{1}{2}+\varepsilon\right) > 1$ since $p \geq 2$. This yields that there is $0 < \delta < 1$ such that for all n large enough we have $\sum_{r=1}^{K} \left\| \left(\frac{I_r^{(n)}}{n} \right)^{\frac{1}{2}+\varepsilon} \right\|_p^p \leq 1 - \delta$.

We now fix $n_0 \in \mathbb{N}$ such that all estimates hold for $n \geq n_0$. Hence, choosing

$$C_p \geq \Delta_p(0) \vee \max\left\{ \Delta_p(j)\, j^{\lambda - \frac{1}{2} - \varepsilon} \middle| j = 1, \ldots, n_0 \right\} \vee \sqrt[p]{\frac{A_4}{\delta}}$$

we finally obtain

$$\ell_p\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right) \leq C_p n^{-\lambda + \frac{1}{2} + \varepsilon}.$$

Obviously, the same reasoning remains valid for $\ell_p\left(\mathscr{Y}_n, \Lambda_{\mathrm{w}}\right)$. Hence, Proposition 5.13 follows. □

Transferring the Rate of Convergence to the Kolmogorov-Smirnov Distance

As final step, the rate obtained in Wasserstein distances in Proposition 5.13 is transferred to the Kolmogorov-Smirnov distance.

Proposition 5.14. *Consider a Pólya urn scheme characterised by **Det R** where* $\lambda = \frac{a-c}{a+b} > \frac{1}{2}$. *Let* $\mathscr{X}_n$ *and* $\mathscr{Y}_n$ *as in* (5.2). *Furthermore, let* $\left(\mathcal{L}(\Lambda_{\mathrm{b}}), \mathcal{L}(\Lambda_{\mathrm{w}})\right)$ *denote the unique solution of system* (5.5) *in* $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$ *and let* $\varepsilon > 0$, *Then, as* $n \to \infty$,

$$\varrho^{\vee}\left(\left(\mathscr{X}_n, \mathscr{Y}_n\right), \left(\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}}\right)\right) = O\left(n^{-\lambda + \frac{1}{2} + \varepsilon}\right).$$

Proof. The rates obtained in the previous section in the Wasserstein metrics $\ell_p^{\vee}$ are transferred to the Kolmogorov-Smirnov distance via Lemma 2.6. From Kuba and Sulzbach [28, Theorem

2] we have that both Λ_{b} and Λ_{w} admit a bounded density $f_{\Lambda_{\mathrm{b}}}$, $f_{\Lambda_{\mathrm{w}}}$, respectively, on $(-\infty, \infty)$. Now, Lemma 2.6 serves to bound the Kolmogorov-Smirnov distance with the help of the Wasserstein metric ℓ_p: Let $0 < \varepsilon' < \varepsilon$. Due to Proposition 5.13, for $p \geq 1$ there is $C_p > 0$ such that

$$\begin{aligned}
\varrho\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right) &\leq (p+1)^{\frac{1}{p+1}} \left(\sup_{x \in \mathbb{R}} \left| f_{\Lambda_{\mathrm{b}}}(x) \right| \ell_p\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right) \right)^{\frac{p}{p+1}} \\
&\leq (p+1)^{\frac{1}{p+1}} \left(\sup_{x \in \mathbb{R}} \left| f_{\Lambda_{\mathrm{b}}}(x) \right| C_p n^{-\lambda + \frac{1}{2} + \varepsilon'} \right)^{\frac{p}{p+1}} \\
&\leq (p+1)^{\frac{1}{p+1}} \left(\sup_{x \in \mathbb{R}} \left| f_{\Lambda_{\mathrm{b}}}(x) \right| C_p \right)^{\frac{p}{p+1}} n^{\left(-\lambda + \frac{1}{2} + \varepsilon'\right)\frac{p}{p+1}} \\
&= (p+1)^{\frac{1}{p+1}} \left(\sup_{x \in \mathbb{R}} \left| f_{\Lambda_{\mathrm{b}}}(x) \right| C_p \right)^{\frac{p}{p+1}} n^{-\lambda + \frac{1}{2} + \varepsilon' + \frac{1}{p+1}\left(\lambda - \frac{1}{2} - \varepsilon'\right)}.
\end{aligned}$$

We now choose p large enough such that

$$\varepsilon' + \frac{1}{p+1}\left(\lambda - \frac{1}{2} - \varepsilon'\right) < \varepsilon \Leftrightarrow p > \frac{\lambda - \frac{1}{2} - \varepsilon}{\varepsilon - \varepsilon'}.$$

As $(p+1)^{\frac{1}{p+1}} \to 1$, $(p \to \infty)$ there exists a constant $C_{\mathrm{KS}} > 0$ such that

$$\varrho\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right) \leq C_{\mathrm{KS}} n^{-\lambda + \frac{1}{2} + \varepsilon}.$$

The same reasoning applies to $\varrho\left(\mathscr{Y}_n, \Lambda_{\mathrm{w}}\right)$ and the assertion follows. □

5.2. Normal Limit Case: $\lambda \leq \frac{1}{2}$

In the normal limit case, upper bounds for rates of convergence for the normalised number of black balls are derived in the Zolotarev metric ζ_3. This is due to the fact that our proofs are based on the contraction method, and Wasserstein distances are not able to detect a contraction when the normal distribution appears as limit. Recall the situation of p. 49 et seq., Notation 5.3 and Remark 5.4.

In order to obtain the standard normal distribution as limit of the normalised quantities, the reciprocal of the standard deviation of the number of black balls serves as scaling factor. We consider the normalised numbers of black balls given by $\hat{\mathscr{X}}_0 := 0 =: \hat{\mathscr{Y}}_0$, $\hat{\mathscr{X}}_1 := 0 =: \hat{\mathscr{Y}}_1$ (note that $\mathrm{Var}\left(B_1^{\mathrm{b}}\right) = \mathrm{Var}\left(B_1^{\mathrm{w}}\right) = 0$) and, for $n \geq 2$,

$$(5.41) \qquad \hat{\mathscr{X}}_n := \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{\sqrt{\mathrm{Var}\left(B_n^{\mathrm{b}}\right)}}, \quad \hat{\mathscr{Y}}_n := \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{\sqrt{\mathrm{Var}\left(B_n^{\mathrm{w}}\right)}}.$$

Due to the recursive description of the number of black balls from (5.1) we have the following system of distributional recursions for the normalised quantities for $n \geq 2$:

$$(5.42) \qquad \begin{aligned} \hat{\mathscr{X}}_n &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} \hat{\mathscr{X}}^{(r)}_{I_r^{(n)}} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} \hat{\mathscr{Y}}^{(r)}_{I_r^{(n)}} + t_{\mathrm{b}}\left(I^{(n)}\right), \\ \hat{\mathscr{Y}}_n &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} \hat{\mathscr{X}}^{(r)}_{I_r^{(n)}} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} \hat{\mathscr{Y}}^{(r)}_{I_r^{(n)}} + t_{\mathrm{w}}\left(I^{(n)}\right) \end{aligned}$$

with

$$\begin{aligned} t_{\mathrm{b}}\left(I^{(n)}\right) &:= \frac{1}{\sigma_{\mathrm{b}}(n)} \left(\sum_{r=1}^{a+1} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{b}}(n) \right), \\ t_{\mathrm{w}}\left(I^{(n)}\right) &:= \frac{1}{\sigma_{\mathrm{w}}(n)} \left(\sum_{r=1}^{c} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=c+1}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{w}}(n) \right), \end{aligned}$$

where $\hat{\mathscr{X}}_j^{(r)} \stackrel{\mathrm{d}}{=} \hat{\mathscr{X}}_j$, $\hat{\mathscr{Y}}_j^{(r)} \stackrel{\mathrm{d}}{=} \hat{\mathscr{Y}}_j$ for $r = 1, \ldots, K$ and $0 \leq j \leq n$ such that $\left(\hat{\mathscr{X}}_j^{(1)}\right)_{0 \leq j \leq n}, \ldots, \left(\hat{\mathscr{X}}_j^{(K)}\right)_{0 \leq j \leq n}, \left(\hat{\mathscr{Y}}_j^{(1)}\right)_{0 \leq j \leq n}, \ldots, \left(\hat{\mathscr{Y}}_j^{(K)}\right)_{0 \leq j \leq n}, I^{(n)}$ are independent.

Letting formally $n \to \infty$ in (5.42) we obtain, with the asymptotic behaviour of the rescaled subtree sizes $\frac{I_r^{(n)}}{n}$, that are hidden in the ratio of the standard deviations, the following system of fixed-point equations that a limit $\left(\hat{\mathscr{X}}, \hat{\mathscr{Y}}\right)$ of $\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right)_{n\in\mathbb{N}}$ should satisfy:

$$\begin{aligned} \hat{\mathscr{X}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} \sqrt{D_r}\hat{\mathscr{X}}^{(r)} + \sum_{r=a+2}^{K} \sqrt{D_r}\hat{\mathscr{Y}}^{(r)}, \\ \hat{\mathscr{Y}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} \sqrt{D_r}\hat{\mathscr{X}}^{(r)} + \sum_{r=c+1}^{K} \sqrt{D_r}\hat{\mathscr{Y}}^{(r)} \end{aligned} \tag{5.43}$$

with independent copies $\hat{\mathscr{X}}^{(r)}$ of $\hat{\mathscr{X}}$ and independent copies $\hat{\mathscr{Y}}^{(r)}$ of $\hat{\mathscr{Y}}$, $r = 1, \ldots, K$ and a Dirichlet-distributed random vector $D = (D_1, \ldots, D_K)$ with all parameters equal to $\frac{1}{K-1}$ such that $\hat{\mathscr{X}}^{(1)}, \ldots, \hat{\mathscr{X}}^{(K)}, \hat{\mathscr{Y}}^{(1)}, \ldots, \hat{\mathscr{Y}}^{(K)}$, and D are independent.

Obviously, regarding the system of distributional fixed-point equations given by (5.43) as self-map of $\mathcal{M} \times \mathcal{M}$ puts us in the situation of Theorem 4.2 (we have $\sum_{r=1}^{K} \left(\sqrt{D_r}\right)^2 = 1$ almost surely, compare (3.5)). Since we scaled by the reciprocal of the exact standard deviation, the unique solution of system (5.43) in $\mathcal{M}_3(0,1) \times \mathcal{M}_3(0,1)$ is given by $(\mathcal{N}(0,1), \mathcal{N}(0,1))$.

Theorem 5.15. *Given a Pólya urn scheme characterised by* ***Det R*** *with* $\lambda = \frac{a-c}{a+b} \leq \frac{1}{2}$, *let* $\hat{\mathscr{X}}_n$ *and* $\hat{\mathscr{Y}}_n$ *as in* (5.41), *respectively, and* $\varepsilon > 0$. *Then, as* $n \to \infty$,

$$\zeta_3^{\vee}\left(\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right), (\mathcal{N}(0,1), \mathcal{N}(0,1))\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda-\frac{1}{2}\right)\vee\left(-\frac{1}{2}+\varepsilon\right)}\right), & \lambda < \frac{1}{2},\ \lambda \neq 0. \end{cases}$$

Remark 5.16. This result confirms a conjecture of Svante Janson stated in [22, Remark 4.7] for $\lambda \in \left(\frac{1}{3}, \frac{1}{2}\right)$. For $\lambda \leq \frac{1}{3}$ Janson expects the rate to be of order $n^{-\frac{1}{2}}$. Unfortunately, it seems that our methods do not enable us to prove the conjectured rate for $\lambda \leq \frac{1}{3}$; this is caused by the need to attain the contractive behaviour of the coefficients, see Remark 8.1.

Proof of Theorem 5.15. The distance of $\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right)$ to its limit $(\mathcal{N}(0,1), \mathcal{N}(0,1))$ is measured in the maximal Zolotarev distance. We conduct all reasoning for the first component $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$. Calculations for the second component $\zeta_3\left(\hat{\mathscr{Y}}_n, \mathcal{N}(0,1)\right)$ pass off analogously.

To prove Theorem 5.15, we first introduce accompanying sequences $\mathscr{Q}_n^{\mathrm{b}}$ and $\mathscr{Q}_n^{\mathrm{w}}$ that combine the recursive descriptions of $\hat{\mathscr{X}}_n$ and $\hat{\mathscr{Y}}_n$ with their limiting distribution, the standard normal distribution. Let, for $n \geq 2$,

$$\mathscr{Q}_n^{\mathrm{b}} := \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} N_r + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} N_r + t_{\mathrm{b}}\left(I^{(n)}\right), \tag{5.44}$$

$$\mathscr{Q}_n^{\mathrm{w}} := \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} N_r + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} N_r + t_{\mathrm{w}}\left(I^{(n)}\right) \tag{5.45}$$

with $N_1, \dots, N_K, I^{(n)}$ independent and $N_1, \dots, N_K$ standard normally distributed. Due to $\mathbb{E}\left[\hat{\mathscr{X}_n}\right] = \mathbb{E}\left[\mathscr{Q}_n^{\mathrm{b}}\right] = \mathbb{E}\left[\hat{\mathscr{Y}_n}\right] = \mathbb{E}\left[\mathscr{Q}_n^{\mathrm{w}}\right] = 0$, $\mathrm{Var}\left(\hat{\mathscr{X}_n}\right) = \mathrm{Var}\left(\mathscr{Q}_n^{\mathrm{b}}\right) = \mathrm{Var}\left(\hat{\mathscr{Y}_n}\right) = \mathrm{Var}\left(\mathscr{Q}_n^{\mathrm{w}}\right) = 1$ and $\mathbb{E}\left[\left|\hat{\mathscr{X}_n}\right|^3\right] < \infty$, $\mathbb{E}\left[\left|\mathscr{Q}_n^{\mathrm{b}}\right|^3\right] < \infty$, $\mathbb{E}\left[\left|\hat{\mathscr{Y}_n}\right|^3\right] < \infty$, $\mathbb{E}\left[\left|\mathscr{Q}_n^{\mathrm{w}}\right|^3\right] < \infty$ for $n \geq 2$, both distances $\zeta_3\left(\hat{\mathscr{X}_n}, \mathscr{Q}_n^{\mathrm{b}}\right)$ and $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$, as well as $\zeta_3\left(\hat{\mathscr{Y}_n}, \mathscr{Q}_n^{\mathrm{w}}\right)$ and $\zeta_3\left(\mathscr{Q}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right)$ are finite for all $n \geq 2$.

We then bound the distance between $\hat{\mathscr{X}_n}$ and the normal distribution by the distances of both to the accompanying sequence:

$$\zeta_3\left(\hat{\mathscr{X}_n}, \mathcal{N}(0,1)\right) \leq \zeta_3\left(\hat{\mathscr{X}_n}, \mathscr{Q}_n^{\mathrm{b}}\right) + \zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right). \tag{5.46}$$

The proof consists of three consecutive phases:

Step 1 First of all, the distance from the sequence to the accompanying sequence at stage n will be dealt with by giving a recursive estimate that describes $\zeta_3\left(\hat{\mathscr{X}_n}, \mathscr{Q}_n^{\mathrm{b}}\right)$ in terms of the distances between the sequence and its limit up to time $n-1$. Therefore, the j-th distance of the sequence to its limit will be denoted by

$$\hat{\Delta}(j) := \zeta_3^{\vee}\left(\left(\hat{\mathscr{X}_j}, \hat{\mathscr{Y}_j}\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right).$$

It will also appear with random argument $I_r^{(n)}$, which means that one of the n distances $\hat{\Delta}(0)$, $\dots$, $\hat{\Delta}(n-1)$ is picked randomly according to the law of $I_r^{(n)}$ (see Remark 5.9).

Step 2 The distance between accompanying sequence and limit, $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$, will be treated with fundamentally different methods and the structure of the Zolotarev distance will be crucial.

Step 3 Finally, we will assemble these pieces and derive an upper bound for the rate of convergence via induction. Thereto, the latter distance of the right-hand side of (5.46) will feed us with information on how fast a rate could possibly be, whereas the former one will serve to adjust the speed and decide for the induction hypothesis.

We shall see that the variance, more precisely, the asymptotic expansion of the variance, is crucial. Lemma 5.2 yields the occurrence of three different kinds of behaviour depending on λ. It is necessary to have a tight upper bound for the second order term of the variance in order to derive rates of convergence.

As the treatment of the three steps, **Step 1**, **Step 2** and **Step 3**, is rather lengthy, every step is taken in the shape of interim results. These finally lead to the proof of this theorem.

Still, the proof should be easy to follow, hence, the order of these interim results is kept as if it was to be read as one single proof.

We interrupt the Proof of Theorem 5.15 and treat the three steps separately, leading to interim results that finally constitute the proof. ◪

Ad Step 1: Recursive Description of $\zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right)$ and $\zeta_3\left(\hat{\mathscr{Y}}_n, \mathscr{Q}_n^{\mathrm{w}}\right)$

Proposition 5.17. *In the situation of Theorem 5.15 with $\mathscr{Q}_n^{\mathrm{b}}$ and $\mathscr{Q}_n^{\mathrm{w}}$ defined in* (5.44) *and* (5.45)*, respectively, it holds, for all $n \ge 2$,*

$$\zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right) \le \mathbb{E}\left[\sum_{r=1}^{a+1}\left(\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K}\left(\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right)\right]$$

as well as

$$\zeta_3\left(\hat{\mathscr{Y}}_n, \mathscr{Q}_n^{\mathrm{w}}\right) \le \mathbb{E}\left[\sum_{r=1}^{c}\left(\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right) + \sum_{r=c+1}^{K}\left(\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right)\right].$$

Proof. Given $I^{(n)}$, all occurring random variables are independent. Denoting the distribution of $I^{(n)}$ by υ and conditioning on the random vector $I^{(n)}$ (we abbreviate $\underline{i} = (i_1, \ldots, i_K)$), we have:

$$\begin{aligned}
&\zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right)\\
&= \zeta_3\Bigg(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{X}}_{I_r^{(n)}}^{(r)} + \sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{Y}}_{I_r^{(n)}}^{(r)} + t_{\mathrm{b}}\left(I^{(n)}\right),\\
&\qquad \sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}N_r + \sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}N_r + t_{\mathrm{b}}\left(I^{(n)}\right)\Bigg)\\
&= \sup_{f\in\mathcal{F}_3}\Bigg|\mathbb{E}\Bigg[f\left(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{X}}_{I_r^{(n)}}^{(r)} + \sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{Y}}_{I_r^{(n)}}^{(r)} + t_{\mathrm{b}}\left(I^{(n)}\right)\right)\\
&\qquad -f\left(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}N_r + \sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}N_r + t_{\mathrm{b}}\left(I^{(n)}\right)\right)\Bigg]\Bigg|\\
&= \sup_{f\in\mathcal{F}_3}\Bigg|\int \mathbb{E}\Bigg[f\left(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{X}}_{i_r}^{(r)} + \sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{Y}}_{i_r}^{(r)} + t_{\mathrm{b}}(\underline{i})\right)\\
&\qquad -f\left(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}N_r + \sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}N_r + t_{\mathrm{b}}(\underline{i})\right)\Bigg]\mathrm{d}\upsilon(\underline{i})\Bigg|
\end{aligned}$$

$$
\begin{aligned}
&\le \int \sup_{f\in\mathcal{F}_3}\left|\mathbb{E}\left[f\left(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{X}}_{i_r}^{(r)}+\sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{Y}}_{i_r}^{(r)}+t_{\mathrm{b}}(\underline{i})\right)\right.\right.\\
&\qquad\left.\left.-f\left(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}N_r+\sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}N_r+t_{\mathrm{b}}(\underline{i})\right)\right]\right|\mathrm{d}\upsilon(\underline{i})\\
&=\int\zeta_3\left(\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{X}}_{i_r}^{(r)}+\sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}\hat{\mathscr{Y}}_{i_r}^{(r)}+t_{\mathrm{b}}(\underline{i}),\right.\\
&\qquad\left.\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}N_r+\sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}N_r+t_{\mathrm{b}}(\underline{i})\right)\mathrm{d}\upsilon(\underline{i})
\end{aligned}
$$

$$
\le\int\sum_{r=1}^{a+1}\left(\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}\right)^3\zeta_3\left(\hat{\mathscr{X}}_{i_r}^{(r)},N_r\right)+\sum_{r=a+2}^{K}\left(\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}\right)^3\zeta_3\left(\hat{\mathscr{Y}}_{i_r}^{(r)},N_r\right)\mathrm{d}\upsilon(\underline{i}) \tag{5.47}
$$

$$
\begin{aligned}
&\le\int\sum_{r=1}^{a+1}\left(\frac{\sigma_{\mathrm{b}}(i_r)}{\sigma_{\mathrm{b}}(n)}\right)^3\hat{\Delta}(i_r)+\sum_{r=a+2}^{K}\left(\frac{\sigma_{\mathrm{w}}(i_r)}{\sigma_{\mathrm{b}}(n)}\right)^3\hat{\Delta}(i_r)\,\mathrm{d}\upsilon(\underline{i})\\
&=\mathbb{E}\left[\sum_{r=1}^{a+1}\left(\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3\hat{\Delta}\left(I_r^{(n)}\right)+\sum_{r=a+2}^{K}\left(\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3\hat{\Delta}\left(I_r^{(n)}\right)\right],
\end{aligned}
$$

where we used that ζ_3 is $(3,+)$-ideal in (5.47).

For $\zeta_3\left(\hat{\mathscr{Y}}_n,\mathscr{Q}_n^{\mathrm{w}}\right)$ the same procedure leads to the recursive estimate claimed above. □

Later on in **Step 3**, when we proceed with this recursive estimate of $\zeta_3\left(\hat{\mathscr{X}}_n,\mathscr{Q}_n^{\mathrm{b}}\right)$, the handling of the ratios $(\sigma_{\mathrm{b}}(j)/\sigma_{\mathrm{b}}(n))^3$ and $(\sigma_{\mathrm{w}}(j)/\sigma_{\mathrm{b}}(n))^3$ will turn out to be crucial. There, the asymptotic expansions of the variance given in Lemma 5.2 will come into play. Beforehand in the next paragraph, we study $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}},\mathcal{N}(0,1)\right)$, the distance between accompanying sequence and the normal distribution.

Ad Step 2: How fast can we get?

There are at least two reasonable ways how to study the quantity $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}},\mathcal{N}(0,1)\right)$. The first way is to bound the Zolotarev distance ζ_3 by the Wasserstein distance ℓ_3 as in Lemma 2.7. This seems to be the most common approach for estimating the distance between accompanying sequence and limit when working with the contraction method in terms of the Zolotarev metrics; e.g., Knape and Neininger did so in their proofs of [27, Theorems 6.1, 6.2, 6.4 and 6.6] in the Pólya urn setting. It was also used by Neininger and Rüschendorf to determine rates of convergence for Quicksort in [40] and by Mahmoud and Neininger to derive rates of convergence for the distribution of distances in random binary search trees in [35]. Hence, it seemed reasonable to give it a try: With the approach suggested by Lemma 2.7, we obtained a rate of order $O\left(n^{\left(\lambda-\frac{1}{2}\right)\vee\left(-\frac{1}{4}\right)}\right)$, $\lambda < \frac{1}{2}$, $\lambda \neq 0$, for $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$, leading to the same upper bound for $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$, cf. Remark 6.16.

This rate obtained via Lemma 2.7 did not look reasonable in any way. There was no reason to expect that the rate would abide at $n^{-\frac{1}{4}}$ with λ decreasing from $\frac{1}{4}$. Moreover, it was not consistent with rates obtained in the setting of urns with subtraction by Flajolet et al. [16] and rates in the context of m-ary search trees, that are related to Pólya urn schemes, in Hwang [21], that both serve as a basis for Janson's conjecture [22, Remark 4.7].

From today's perspective, one problem is that the rate obtained in Lemma 3.4 is not optimal, and as a consequence the rate in Corollary 3.5 is not either. We now know from Janson, see Remark 3.6, that the correct rate for the quantities in Lemma 3.4 (in Wasserstein distances) is of order n^{-1}. With his result and the approach suggested by Lemma 2.7, we would obtain a rate of order $O\left(n^{\left(\lambda-\frac{1}{2}\right)\vee\left(-\frac{1}{2}+\varepsilon\right)}\right)$ for $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$. Except for the ε that we seem to lose in the exponent due to our methods, this looks better than the rate before, because it seems reasonable that the rate should abide at $n^{-\frac{1}{2}}$ since this is the order of the reciprocal of the standard deviation. In addition, it agrees with the results from Flajolet et al. [16]. But still, it does not match the results of Hwang [21] and Janson's conjecture.

To illustrate the approach suggested by Lemma 2.7, it will be elaborated in the context of setting **Rand R**, since Lemmata 3.7 and 3.8 yield better estimates for the growth of the rescaled subtrees than Lemma 3.4 and Corollary 3.5. In Remark 6.16 we explain the problems arising in more detail. Additionally, the simulations presented in Chapter 1 and in Appendix B serve to compare the rates available via Lemma 2.7 and the rates stated in Theorem 5.15.

Hence, we tried another approach to estimate $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ inspired by the treatment of degenerate limit equations by Neininger and Rüschendorf [42, Proof of Theorem 2.1]: It mainly relies on the convolution property of the normal distribution and on applying Taylor's Theorem to the test functions of the Zolotarev distance leading to the following result:

Proposition 5.18. *In the situation of Theorem 5.15 with $\mathscr{Q}_n^{\mathrm{b}}$ and $\mathscr{Q}_n^{\mathrm{w}}$ as in* (5.44) *and* (5.45)*, it holds, as n $\to \infty$,*

$$\begin{aligned}
\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) &= O\left(\left\|\sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - 1\right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3\right), \\
\zeta_3\left(\mathscr{Q}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right) &= O\left(\left\|\sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} - 1\right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\|t_{\mathrm{w}}\left(I^{(n)}\right)\right\|_3^3\right).
\end{aligned}$$

Proof. First, we recall the convolution property of the normal distribution: For independent random variables V and W with $\mathcal{L}(V) = \mathcal{N}(v, w^2)$ and $\mathcal{L}(W) = \mathcal{N}(x, y^2)$ where $v, x \in \mathbb{R}$, $w, y \in \mathbb{R}_+$ it holds $\mathcal{L}(W+V) = \mathcal{N}(v+x, w^2+y^2)$. If $v = x = 0$ this yields $V + W \stackrel{\mathrm{d}}{=} \sqrt{w^2+y^2}N$ with standard normally distributed N.

Due to the convolution property of the normal distribution, the accompanying sequences $\mathscr{Q}_n^{\mathrm{b}}$ and $\mathscr{Q}_n^{\mathrm{w}}$ of (5.44) and (5.45) can be rewritten in a distributional sense as

$$\begin{aligned}
\mathscr{Q}_n^{\mathrm{b}} &\stackrel{\mathrm{d}}{=} \left(\sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}\right)^{\frac{1}{2}} N + t_{\mathrm{b}}\left(I^{(n)}\right), \\
\mathscr{Q}_n^{\mathrm{w}} &\stackrel{\mathrm{d}}{=} \left(\sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)}\right)^{\frac{1}{2}} N + t_{\mathrm{w}}\left(I^{(n)}\right)
\end{aligned}$$

with standard normally distributed N independent of $I^{(n)}$.

Abbreviating the coefficients of the normal distribution as

$$\begin{aligned}
G_n^{\mathrm{b}} &:= \left(\sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}\right)^{\frac{1}{2}} \quad \text{and} \\
G_n^{\mathrm{w}} &:= \left(\sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)}\right)^{\frac{1}{2}},
\end{aligned}$$

we use the distributional rewritings from above to split the coefficients G_n^{b} and G_n^{w} using the convolution property of the normal distribution again with the help of

$$\begin{aligned}
A_{\mathrm{b}} &:= \left\{G_n^{\mathrm{b}} \geq 1\right\}, \quad \Delta_n^{\mathrm{b}} := \sqrt{\left|\left(G_n^{\mathrm{b}}\right)^2 - 1\right|}, \\
A_{\mathrm{w}} &:= \left\{G_n^{\mathrm{w}} \geq 1\right\}, \quad \Delta_n^{\mathrm{w}} := \sqrt{\left|\left(G_n^{\mathrm{w}}\right)^2 - 1\right|}.
\end{aligned}$$

Thereby, we have with standard normally distributed N' independent of N and $I^{(n)}$

$$\begin{aligned}
\mathscr{Q}_n^{\mathrm{b}} &\stackrel{\mathrm{d}}{=} \mathbb{1}_{A_{\mathrm{b}}}\left(N + \Delta_n^{\mathrm{b}} N' + t_{\mathrm{b}}\left(I^{(n)}\right)\right) + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}} N + t_{\mathrm{b}}\left(I^{(n)}\right)\right) =: \check{Q}_n^{\mathrm{b}}, \\
\mathscr{Q}_n^{\mathrm{w}} &\stackrel{\mathrm{d}}{=} \mathbb{1}_{A_{\mathrm{w}}}\left(N + \Delta_n^{\mathrm{w}} N' + t_{\mathrm{w}}\left(I^{(n)}\right)\right) + \mathbb{1}_{(A_{\mathrm{w}})^c}\left(G_n^{\mathrm{w}} N + t_{\mathrm{w}}\left(I^{(n)}\right)\right) =: \check{Q}_n^{\mathrm{w}}.
\end{aligned} \tag{5.48}$$

We will now dive into calculations making explicit use of the test functions of the Zolotarev distance. All calculations will be conducted for $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ (and the above defined quantities related to $\mathscr{Q}_n^{\mathrm{b}}$). However, the same reasoning holds for $\zeta_3\left(\mathscr{Q}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right)$.

Inspired by (5.48), we decompose the normal distribution as follows:

$$
\begin{aligned}
(5.49) \qquad N &\stackrel{\mathrm{d}}{=} \mathbb{1}_{A_{\mathrm{b}}} N + \mathbb{1}_{(A_{\mathrm{b}})^c} \left(G_n^{\mathrm{b}} N + \Delta_n^{\mathrm{b}} N' \right) =: \check{N}, \\
N &\stackrel{\mathrm{d}}{=} \mathbb{1}_{A_{\mathrm{w}}} N + \mathbb{1}_{(A_{\mathrm{w}})^c} \left(G_n^{\mathrm{w}} N + \Delta_n^{\mathrm{w}} N' \right),
\end{aligned}
$$

where the latter decomposition refers to the calculations to be conducted when estimating $\zeta_3 \left(\mathscr{Q}_n^{\mathrm{w}}, \mathcal{N}(0,1) \right)$.

Obviously, the quantities A_{b}, A_{w} and $\check{N}$ also depend on n, and the latter should carry a "b". For the sake of readability in the course of the following calculations, those sub-/superscripts were omitted.

The Zolotarev distance is a simple metric; hence, we have $\zeta_3 \left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1) \right) = \zeta_3 \left(\check{Q}_n^{\mathrm{b}}, \check{N} \right)$ and by definition that is

$$
\zeta_3 \left(\check{Q}_n^{\mathrm{b}}, \check{N} \right) = \sup_{f \in \mathcal{F}_3} \left| \mathbb{E} \left[f \left(\check{Q}_n^{\mathrm{b}} \right) - f \left(\check{N} \right) \right] \right| .
$$

We will study this quantity by having a closer look at the test functions given by $\mathcal{F}_3$. Therefore, we investigate their Taylor series expansions.

Firstly, let $f \in \mathcal{F}_3$ and let V and W be random variables with $\mathbb{E}[V] = \mathbb{E}[W]$, $\mathbb{E}\left[V^2\right] = \mathbb{E}\left[W^2\right]$ and $\mathbb{E}\left[V^3\right], \mathbb{E}\left[W^3\right] < \infty$. Then, we observe that g with $g(x) := f(x) - \frac{x^2}{2} f''(0)$ is twice continuous differentiable and Hölder continuous with Hölder constant 1. To put it in another way, we have that $g \in \mathcal{F}_3$, and $\mathbb{E}[f(V) - f(W)] = \mathbb{E}[g(V) - g(W)]$. Hence, w.l.o.g., we can assume $f''(0) = 0$. Taylor's Theorem accompanied by the Peano Form of the Remainder applied to f yields at $a \in \mathbb{R}$

$$
f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f''(\xi) - f''(a)}{2}(x-a)^2
$$

for some ξ between a and x. Writing $R(x,a) := \frac{f''(\xi)-f''(a)}{2}(x-a)^2$ for the remainder, we have, expanding f at N,

$$
\begin{aligned}
(5.50) \qquad & f\left(\check{Q}_n^{\mathrm{b}}\right) - f\left(\check{N}\right) \\
= \; & f'(N)\left(\check{Q}_n^{\mathrm{b}} - \check{N}\right) + \frac{f''(N)}{2}\left(\left(\check{Q}_n^{\mathrm{b}} - N\right)^2 - \left(\check{N} - N\right)^2\right) + R\left(\check{Q}_n^{\mathrm{b}}, N\right) - R\left(\check{N}, N\right).
\end{aligned}
$$

The above equation (5.50) is the key to the analysis of the distance $\zeta_3 \left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1) \right)$. In the following calculations, the three terms of the right-hand side of (5.50) will be studied separately. We will see that the first of them does not contribute at all and that the second term will be dominated by the estimates of the remainders — especially the remainder comprising the modified accompanying sequence $\check{Q}_n^{\mathrm{b}}$ will be crucial.

Plugging in (5.48) and (5.49), we obtain for the first term of the right-hand side of (5.50)

$$\begin{aligned}
& f'(N)\left(\check{Q}_n^{\mathrm{b}} - \check{N}\right) \\
= \; & f'(N)\left[\mathbb{1}_{A_{\mathrm{b}}}\left(N + \Delta_n^{\mathrm{b}} N' + t_{\mathrm{b}}\left(I^{(n)}\right)\right) + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}} N + t_{\mathrm{b}}\left(I^{(n)}\right)\right)\right. \\
& \qquad \left. - \mathbb{1}_{A_{\mathrm{b}}} N - \mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}} N + \Delta_n^{\mathrm{b}} N'\right)\right] \\
= \; & f'(N)\left[\mathbb{1}_{A_{\mathrm{b}}}\left(\Delta_n^{\mathrm{b}} N' + t_{\mathrm{b}}\left(I^{(n)}\right)\right) + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(t_{\mathrm{b}}\left(I^{(n)}\right) - \Delta_n^{\mathrm{b}} N'\right)\right] \\
= \; & f'(N)\left(t_{\mathrm{b}}\left(I^{(n)}\right) + \Delta_n^{\mathrm{b}} N'\left(\mathbb{1}_{A_{\mathrm{b}}} - \mathbb{1}_{(A_{\mathrm{b}})^c}\right)\right),
\end{aligned}$$

yielding

$$\begin{aligned}
& \mathbb{E}\left[f'(N)\left(\check{Q}_n^{\mathrm{b}} - \check{N}\right)\right] \\
= \; & \mathbb{E}\left[f'(N)\right]\left(\mathbb{E}\left[t_{\mathrm{b}}\left(I^{(n)}\right)\right] + \mathbb{E}\left[\Delta_n^{\mathrm{b}}\left(\mathbb{1}_{A_{\mathrm{b}}} - \mathbb{1}_{(A_{\mathrm{b}})^c}\right)\right]\mathbb{E}\left[N'\right]\right) = 0,
\end{aligned} \tag{5.51}$$

where we used that N' is standard normally distributed, hence $\mathbb{E}[N'] = 0$, as well as $\mathbb{E}\left[t_{\mathrm{b}}\left(I^{(n)}\right)\right] = 0$ due to $\mathbb{E}\left[\hat{\mathscr{X}}_n\right] = 0$.

For the second term of the right-hand side of (5.50), we obtain with (5.48) and (5.49)

$$\begin{aligned}
& \left(\check{Q}_n^{\mathrm{b}} - N\right)^2 - \left(\check{N} - N\right)^2 \\
= \; & \left[\mathbb{1}_{A_{\mathrm{b}}}\left(N + \Delta_n^{\mathrm{b}} N' + t_{\mathrm{b}}\left(I^{(n)}\right)\right) + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}} N + t_{\mathrm{b}}\left(I^{(n)}\right)\right) - N\right]^2 \\
& - \left[\mathbb{1}_{A_{\mathrm{b}}} N + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}} N + \Delta_n^{\mathrm{b}} N'\right) - N\right]^2 \\
= \; & \left[\mathbb{1}_{A_{\mathrm{b}}}\left(\Delta_n^{\mathrm{b}} N' + t_{\mathrm{b}}\left(I^{(n)}\right)\right) + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(N\left(G_n^{\mathrm{b}} - 1\right) + t_{\mathrm{b}}\left(I^{(n)}\right)\right)\right]^2 \\
& - \left[\mathbb{1}_{(A_{\mathrm{b}})^c}\left(N\left(G_n^{\mathrm{b}} - 1\right) + \Delta_n^{\mathrm{b}} N'\right)\right]^2 \\
= \; & \mathbb{1}_{A_{\mathrm{b}}}\left(\Delta_n^{\mathrm{b}} N' + t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2 + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(N\left(G_n^{\mathrm{b}} - 1\right) + t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2 \qquad (5.52) \\
& - \mathbb{1}_{(A_{\mathrm{b}})^c}\left(N\left(G_n^{\mathrm{b}} - 1\right) + \Delta_n^{\mathrm{b}} N'\right)^2 \\
= \; & \mathbb{1}_{A_{\mathrm{b}}}\left(\left(\Delta_n^{\mathrm{b}} N'\right)^2 + 2\Delta_n^{\mathrm{b}} N' t_{\mathrm{b}}\left(I^{(n)}\right) + \left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2\right) \\
& + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(N^2\left(G_n^{\mathrm{b}} - 1\right)^2 + 2N\left(G_n^{\mathrm{b}} - 1\right) t_{\mathrm{b}}\left(I^{(n)}\right) + \left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2\right) \\
& - \mathbb{1}_{(A_{\mathrm{b}})^c}\left(N^2\left(G_n^{\mathrm{b}} - 1\right)^2 + 2N\left(G_n^{\mathrm{b}} - 1\right)\Delta_n^{\mathrm{b}} N' + \left(\Delta_n^{\mathrm{b}} N'\right)^2\right) \\
= \; & \left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2 + \left(\Delta_n^{\mathrm{b}} N'\right)^2\left(\mathbb{1}_{A_{\mathrm{b}}} - \mathbb{1}_{(A_{\mathrm{b}})^c}\right)
\end{aligned}$$

$$+\mathbb{1}_{A_{\mathrm{b}}} 2\Delta_n^{\mathrm{b}} N' t_{\mathrm{b}}\left(I^{(n)}\right)+\mathbb{1}_{(A_{\mathrm{b}})^c} 2N\left(G_n^{\mathrm{b}}-1\right)\left(t_{\mathrm{b}}\left(I^{(n)}\right)-\Delta_n^{\mathrm{b}} N'\right)$$

using in (5.52) that mixed terms do not contribute because the product of the indicators make them vanish.

For the moment, we have for the second term of the right-hand side of (5.50)

$$\begin{aligned}
& \mathbb{E}\left[\frac{f''(N)}{2}\left(\left(\check{Q}_n^{\mathrm{b}}-N\right)^2-\left(\check{N}-N\right)^2\right)\right] \\
= \ & \mathbb{E}\left[\frac{f''(N)}{2}\left(\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2+\mathbb{1}_{(A_{\mathrm{b}})^c} 2N\left(G_n^{\mathrm{b}}-1\right)\left(t_{\mathrm{b}}\left(I^{(n)}\right)-\Delta_n^{\mathrm{b}} N'\right)\right)\right] \\
& +\mathbb{E}\left[\frac{f''(N)}{2}\left(\Delta_n^{\mathrm{b}}\right)^2\left(\mathbb{1}_{A_{\mathrm{b}}}-\mathbb{1}_{(A_{\mathrm{b}})^c}\right)\right]\mathbb{E}\left[\left(N'\right)^2\right] \\
& +\mathbb{E}\left[f''(N)\, \mathbb{1}_{A_{\mathrm{b}}} \Delta_n^{\mathrm{b}} t_{\mathrm{b}}\left(I^{(n)}\right)\right]\mathbb{E}\left[N'\right] \\
= \ & \mathbb{E}\left[\frac{f''(N)}{2}\left(\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2+\left(\Delta_n^{\mathrm{b}}\right)^2\left(\mathbb{1}_{A_{\mathrm{b}}}-\mathbb{1}_{(A_{\mathrm{b}})^c}\right)\right)\right] \\
& +\mathbb{E}\left[f''(N)\, \mathbb{1}_{(A_{\mathrm{b}})^c} N\left(G_n^{\mathrm{b}}-1\right) t_{\mathrm{b}}\left(I^{(n)}\right)\right] \\
& -\mathbb{E}\left[f''(N)\, \mathbb{1}_{(A_{\mathrm{b}})^c} N\left(G_n^{\mathrm{b}}-1\right) \Delta_n^{\mathrm{b}}\right]\mathbb{E}\left[N'\right] \\
= \ & \mathbb{E}\left[\frac{f''(N)}{2}\left(\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2+\left(\Delta_n^{\mathrm{b}}\right)^2\left(\mathbb{1}_{A_{\mathrm{b}}}-\mathbb{1}_{(A_{\mathrm{b}})^c}\right)\right)\right] \\
& +\mathbb{E}\left[f''(N)\, \mathbb{1}_{(A_{\mathrm{b}})^c} N\left(G_n^{\mathrm{b}}-1\right) t_{\mathrm{b}}\left(I^{(n)}\right)\right]
\end{aligned}$$

using $\mathbb{E}\left[\left(N'\right)^2\right]=1$ and $\mathbb{E}\left[N'\right]=0$.

Now, we observe

1. $\left(\Delta_n^{\mathrm{b}}\right)^2\left(\mathbb{1}_{A_{\mathrm{b}}}-\mathbb{1}_{(A_{\mathrm{b}})^c}\right)=\left|\left(G_n^{\mathrm{b}}\right)^2-1\right|\left(\mathbb{1}_{\{G_n^{\mathrm{b}}\geq 1\}}-\mathbb{1}_{\{G_n^{\mathrm{b}}<1\}}\right)=\left(G_n^{\mathrm{b}}\right)^2-1,$

2. $\mathbb{E}\left[\left(G_n^{\mathrm{b}}\right)^2+\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2\right]=1$, since

$$1=\operatorname{Var}\left(\mathscr{Q}_n^{\mathrm{b}}\right)=\mathbb{E}\left[\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}+\sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}+\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2\right]$$

due to the independence of $N^{(1)},\ldots,\ N^{(K)}$ and $I^{(n)}$ and $\mathbb{E}\left[N^{(r)}\right]=0$ as well as $\operatorname{Var}\left(N^{(r)}\right)=1$ for $r=1,\ldots,K$.

Hence, we have

$$\mathbb{E}\left[\frac{f''(N)}{2}\left(\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2+\left(\Delta_n^{\mathrm{b}}\right)^2\left(\mathbb{1}_{A_{\mathrm{b}}}-\mathbb{1}_{(A_{\mathrm{b}})^c}\right)\right)\right]$$

$$= \mathbb{E}\left[\frac{f''(N)}{2}\right]\mathbb{E}\left[\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2+\left(G_n^{\mathrm{b}}\right)^2-1\right]=0$$

and obtain for the second term of the right-hand side of (5.50)

$$\mathbb{E}\left[\frac{f''(N)}{2}\left(\left(\check{Q}_n^{\mathrm{b}}-N\right)^2-\left(\check{N}-N\right)^2\right)\right]=\mathbb{E}\left[f''(N)N\right]\mathbb{E}\left[\mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}}-1\right)t_{\mathrm{b}}\left(I^{(n)}\right)\right].$$

With Lemma A.2, we have

$$\begin{aligned}
&\left|\mathbb{E}\left[\frac{f''(N)}{2}\left(\left(\check{Q}_n^{\mathrm{b}}-N\right)^2-\left(\check{N}-N\right)^2\right)\right]\right|\\
\leq\ &\left|\mathbb{E}\left[\mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}}-1\right)t_{\mathrm{b}}\left(I^{(n)}\right)\right]\right|\\
\leq\ &\mathbb{E}\left[\left|G_n^{\mathrm{b}}-1\right|\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|\right]\\
\leq\ &\left\|G_n^{\mathrm{b}}-1\right\|_{\frac{3}{2}}\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3 \qquad (5.53)
\end{aligned}$$

with Hölder's inequality.

Finally, we have to turn our attention to the third and fourth term of (5.50): We have to find estimates for the remainders. As $f \in \mathcal{F}_3$, we have $|f''(x)-f''(y)| \leq |x-y|$ and estimate

$$\left|\frac{f''(\xi)-f''(a)}{2}(x-a)^2\right| \leq \frac{|\xi-a|}{2}(x-a)^2 \leq \frac{|x-a|^3}{2}.$$

We start with the remainder comprising N and the modified accompanying sequence $\check{Q}_n^{\mathrm{b}}$:

$$\begin{aligned}
&\left|R\left(\check{Q}_n^{\mathrm{b}},N\right)\right|\\
\leq\ &\frac{1}{2}\left|\check{Q}_n^{\mathrm{b}}-N\right|^3\\
=\ &\frac{1}{2}\left|\mathbb{1}_{A_{\mathrm{b}}}\left(N+\Delta_n^{\mathrm{b}}N'+t_{\mathrm{b}}\left(I^{(n)}\right)\right)+\mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}}N-t_{\mathrm{b}}\left(I^{(n)}\right)\right)-N\right|^3\\
\leq\ &\left|\mathbb{1}_{A_{\mathrm{b}}}\left(\Delta_n^{\mathrm{b}}N'+t_{\mathrm{b}}\left(I^{(n)}\right)\right)+\mathbb{1}_{(A_{\mathrm{b}})^c}\left(N\left(G_n^{\mathrm{b}}-1\right)+t_{\mathrm{b}}\left(I^{(n)}\right)\right)\right|^3\\
\leq\ &\mathbb{1}_{A_{\mathrm{b}}}\left(\left|\Delta_n^{\mathrm{b}}N'\right|+\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|\right)^3+\mathbb{1}_{(A_{\mathrm{b}})^c}\left(\left|N\left(G_n^{\mathrm{b}}-1\right)\right|+\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|\right)^3\\
=\ &\mathbb{1}_{A_{\mathrm{b}}}\left(\left|\Delta_n^{\mathrm{b}}N'\right|^3+3\left|\Delta_n^{\mathrm{b}}N'\right|^2\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|+3\left|\Delta_n^{\mathrm{b}}N'\right|\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^2+\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^3\right)\\
&+\mathbb{1}_{(A_{\mathrm{b}})^c}\left(\left|N\left(G_n^{\mathrm{b}}-1\right)\right|^3+3\left|N\left(G_n^{\mathrm{b}}-1\right)\right|^2\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|\right.\\
&\qquad\left.+3\left|N\left(G_n^{\mathrm{b}}-1\right)\right|\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^2+\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^3\right)\\
\leq\ &\mathbb{1}_{A_{\mathrm{b}}}\left(\left|\Delta_n^{\mathrm{b}}N'\right|^3+3\left|\Delta_n^{\mathrm{b}}N'\right|^2\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|+3\left|\Delta_n^{\mathrm{b}}N'\right|\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^2+\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^3\right)\\
&+\mathbb{1}_{(A_{\mathrm{b}})^c}\left(\left|\Delta_n^{\mathrm{b}}N\right|^3+3\left|\Delta_n^{\mathrm{b}}N\right|^2\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|+3\left|\Delta_n^{\mathrm{b}}N\right|\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^2+\left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^3\right),
\end{aligned}$$

where we used in the last step $\left|G_n^{\mathrm{b}} - 1\right| = \left|\sqrt{\left(G_n^{\mathrm{b}}\right)^2} - \sqrt{1}\right| \leq \sqrt{\left|\left(G_n^{\mathrm{b}}\right)^2 - 1\right|} = \Delta_n^{\mathrm{b}}$, as $\left|\sqrt{x} - \sqrt{y}\right| \leq \sqrt{|x-y|}$, for $x, y \geq 0$, see (A.1).

This yields

$$\begin{aligned}
& \mathbb{E}\left[\left|R\left(\check{Q}_n^{\mathrm{b}}, N\right)\right|\right] \\
\leq\ & \|N\|_3^3 \left\|\Delta_n^{\mathrm{b}}\right\|_3^3 + \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3 \\
& +3\,\mathbb{E}\left[\left|\Delta_n^{\mathrm{b}}\right|^2 \left|t_{\mathrm{b}}\left(I^{(n)}\right)\right| \left(|N'|^2 + |N|^2\right)\right] + 3\,\mathbb{E}\left[\left|\Delta_n^{\mathrm{b}}\right| \left|t_{\mathrm{b}}\left(I^{(n)}\right)\right|^2 \left(|N'| + |N|\right)\right] \\
\leq\ & \|N\|_3^3 \left\|\Delta_n^{\mathrm{b}}\right\|_3^3 + \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3 \\
& +6 \left\|\left(\Delta_n^{\mathrm{b}}\right)^2\right\|_{\frac{3}{2}} \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3 + 6\,\mathbb{E}\left[|N|\right] \left\|\Delta_n^{\mathrm{b}}\right\|_3 \left\|\left(t_{\mathrm{b}}\left(I^{(n)}\right)\right)^2\right\|_{\frac{3}{2}} \\
\leq\ & \|N\|_3^3 \left\|\Delta_n^{\mathrm{b}}\right\|_3^3 + \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3 \\
& +6 \left\|\Delta_n^{\mathrm{b}}\right\|_3^2 \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3 + 6 \left\|\Delta_n^{\mathrm{b}}\right\|_3 \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^2 \qquad (5.54) \\
\leq\ & \|N\|_3^3 \left\|\Delta_n^{\mathrm{b}}\right\|_3^3 + \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3 + 12\left(\left\|\Delta_n^{\mathrm{b}}\right\|_3^3 + \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3\right) \\
=\ & O\left(\left\|\Delta_n^{\mathrm{b}}\right\|_3^3 + \left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3\right) \qquad (5.55)
\end{aligned}$$

with $\mathbb{E}\left[|N|\right] < 1$ due to $0 < \mathrm{Var}\left(|N|\right) = \mathbb{E}\left[N^2\right] - \mathbb{E}\left[|N|\right]^2$ in (5.54). For the remainder comprising the decomposed normal distribution $\check{N}$ and N, we have

$$\begin{aligned}
\left|R\left(\check{N}, N\right)\right| \leq & \frac{1}{2}\left|\check{N} - N\right|^3 = \frac{1}{2}\left|\mathbb{1}_{A_{\mathrm{b}}} N + \mathbb{1}_{(A_{\mathrm{b}})^c}\left(G_n^{\mathrm{b}} N + \Delta_n^{\mathrm{b}} N'\right) - N\right|^3 \\
= & \frac{1}{2}\left|\mathbb{1}_{(A_{\mathrm{b}})^c}\left(N\left(G_n^{\mathrm{b}} - 1\right) + \Delta_n^{\mathrm{b}} N'\right)\right|^3 \\
\leq & \frac{1}{2}\left|\mathbb{1}_{(A_{\mathrm{b}})^c}\left(|N|\left|G_n^{\mathrm{b}} - 1\right| + \left|\Delta_n^{\mathrm{b}}\right| |N'|\right)\right|^3 \\
\leq & \frac{1}{2}\left|\mathbb{1}_{(A_{\mathrm{b}})^c}\left(|N|\left|\left(G_n^{\mathrm{b}}\right)^2 - 1\right| + \left|\Delta_n^{\mathrm{b}}\right| |N'|\right)\right|^3 \qquad (5.56) \\
\leq & \frac{1}{2}\left(\left|N\Delta_n^{\mathrm{b}}\right| + \left|N'\Delta_n^{\mathrm{b}}\right|\right) \leq \left|\Delta_n^{\mathrm{b}}\right|^3 \left(|N| + |N'|\right)^3.
\end{aligned}$$

The following observation leads to (5.56): $|a-1| \leq \left|a^2 - 1\right|$ for $a \geq 0$, since

1. for $a = 0, a = 1$ it is obvious,
2. for $a > 1$, we have $0 < a - 1 \leq a^2 - 1$, and
3. for $0 < a < 1$, it holds $|a-1| = 1 - a \leq 1 - a^2 = \left|a^2 - 1\right|$.

Hence, we obtain

$$\mathbb{E}\left[\left|R\left(\check{N},N\right)\right|\right] \leq \mathbb{E}\left[\left(|N|+|N'|\right)^3\right]\left\|\Delta_n^{\mathrm{b}}\right\|_3^3 = O\left(\left\|\Delta_n^{\mathrm{b}}\right\|_3^3\right). \tag{5.57}$$

Plugging in the definition of Δ_n^{b}, we have:

$$\begin{aligned}\left\|\Delta_n^{\mathrm{b}}\right\|_3 &= \left(\mathbb{E}\left[\left|\left(G_n^{\mathrm{b}}\right)^2-1\right|^{\frac{3}{2}}\right]\right)^{\frac{1}{3}} = \left\|\left(G_n^{\mathrm{b}}\right)^2-1\right\|_{\frac{3}{2}}^{\frac{1}{2}} \\ &= \left\|\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}+\sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}-1\right\|_{\frac{3}{2}}^{\frac{1}{2}}.\end{aligned} \tag{5.58}$$

Reassembling (5.51), (5.53), (5.55), (5.57), and (5.58) to estimate (5.50), we obtain

$$\begin{aligned}&\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}},\mathcal{N}(0,1)\right) = \zeta_3\left(\check{Q}_n^{\mathrm{b}},\check{N}\right) \\ \leq\ &\sup_{f\in\mathcal{F}_3}\left|\mathbb{E}\left[f'(N)\left(\check{Q}_n^{\mathrm{b}}-\check{N}\right)+\frac{f''(N)}{2}\left(\left(\check{Q}_n^{\mathrm{b}}-N\right)^2-\left(\check{N}-N\right)^2\right)\right.\right. \\ &\qquad\left.\left.+R\left(\check{Q}_n^{\mathrm{b}},N\right)-R\left(\check{N},N\right)\right]\right| \\ \leq\ &\left\|G_n^{\mathrm{b}}-1\right\|_{\frac{3}{2}}\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3+O\left(\left\|\Delta_n^{\mathrm{b}}\right\|_3^3+\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3\right) \\ \leq\ &\left\|\left(G_n^{\mathrm{b}}\right)^2-1\right\|_{\frac{3}{2}}\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3+O\left(\left\|\left(G_n^{\mathrm{b}}\right)^2-1\right\|_{\frac{3}{2}}^{\frac{3}{2}}+\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3\right) \\ =\ &O\left(\left\|\left(G_n^{\mathrm{b}}\right)^2-1\right\|_{\frac{3}{2}}^{\frac{3}{2}}+\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3\right) \qquad (5.59)\\ =\ &O\left(\left\|\sum_{r=1}^{a+1}\frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}+\sum_{r=a+2}^{K}\frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)}-1\right\|_{\frac{3}{2}}^{\frac{3}{2}}+\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3^3\right),\end{aligned}$$

where in (5.59) the following observation was applied to the null sequences $\left\|\left(G_n^{\mathrm{b}}\right)^2-1\right\|_{\frac{3}{2}}$ and $\left\|t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3$: Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ both be positive sequences with $a_n, b_n \to 0$ as $n\to\infty$. Then, it holds

$$a_n b_n \leq a_n^{\frac{3}{2}} + b_n^3$$

for all n large enough (such that $a_n, b_n \in (0,1)$), because:

1. If $a_n \geq b_n$, then $a_n b_n \leq a_n^2 \leq a_n^{\frac{3}{2}}$.

2. If $a_n < b_n$, two subcases occur:

a) If $b_n^2 < a_n < b_n$, then $a_n b_n \leq a_n a_n^{\frac{1}{2}} = a_n^{\frac{3}{2}}$.

b) If $a_n \leq b_n^2$, then $a_n b_n \leq b_n^2 b_n = b_n^3$.

Hence, we have

$$\left\| \left(G_n^{\mathrm{b}}\right)^2 - 1 \right\|_{\frac{3}{2}} \left\| t_{\mathrm{b}}\left(I^{(n)}\right) \right\|_3 \leq O\left(\left\| \left(G_n^{\mathrm{b}}\right)^2 - 1 \right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\| t_{\mathrm{b}}\left(I^{(n)}\right) \right\|_3^3 \right).$$

All calculations for the distance $\zeta_3\left(\mathscr{Q}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right)$ pass off analogously. Finally, the assertion of Proposition 5.18 follows. □

In a next step, we have to deal with the toll terms $t_{\mathrm{b}}\left(I^{(n)}\right)$ and $t_{\mathrm{w}}\left(I^{(n)}\right)$ as well as the behaviour of the ratios of the variances that both arise from normalising in (5.42) in order to apply Proposition 5.18. For both, we will make use of the asymptotic expansions in Lemma 5.1 and Lemma 5.2.

We start with the toll terms of our recursion $t_{\mathrm{b}}\left(I^{(n)}\right)$ and $t_{\mathrm{w}}\left(I^{(n)}\right)$.

Lemma 5.19. *For the toll terms $t_{\mathrm{b}}\left(I^{(n)}\right)$ and $t_{\mathrm{w}}\left(I^{(n)}\right)$, defined via (5.42), we have, as $n \to \infty$,*

$$\max\left\{ \left\| t_{\mathrm{b}}\left(I^{(n)}\right) \right\|_3, \left\| t_{\mathrm{w}}\left(I^{(n)}\right) \right\|_3 \right\} = \begin{cases} O\left((\ln(n))^{-\frac{1}{2}} \right), & \text{if } \lambda = \frac{1}{2}, \\ O\left(n^{\lambda - \frac{1}{2}} \right), & \text{if } 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}} \right), & \text{if } \lambda < 0. \end{cases}$$

Proof. All calculations will be conducted for $t_{\mathrm{b}}\left(I^{(n)}\right)$ as calculations for $t_{\mathrm{w}}\left(I^{(n)}\right)$ run accordingly. We recall $t_{\mathrm{b}}\left(I^{(n)}\right) = \frac{1}{\sigma_{\mathrm{b}}(n)} \left(\sum_{r=1}^{a+1} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{b}}(n) \right)$. Hence, both Lemma 5.1 and Lemma 5.2 come into play. Therefore, we have to distinguish between the three cases $\lambda = \frac{1}{2}$, $0 < \lambda < \frac{1}{2}$ and $\lambda < 0$.

Case $\lambda = \frac{1}{2}$:
From Lemma 5.1.i) with Notation 5.3, we have $\mu_{\mathrm{b}}(j) = c_{\mathrm{b}} n + d_{\mathrm{b}} n^{\frac{1}{2}} + O(1)$. We know from Lemma 5.2.i) that the standard deviation is of order $\sqrt{n \ln(n)}$. We easily see that the linear terms cancel out up to a constant in $t_{\mathrm{b}}\left(I^{(n)}\right)$. The second order terms of the expectation contribute of order $n^{\frac{1}{2}}$:

$$t_{\mathrm{b}}\left(I^{(n)}\right) = \frac{1}{\sigma_{\mathrm{b}}(n)} \left(\sum_{r=1}^{a+1} d_{\mathrm{b}} \left(I_r^{(n)}\right)^{\frac{1}{2}} + \sum_{r=a+2}^{K} d_{\mathrm{w}} \left(I_r^{(n)}\right)^{\frac{1}{2}} - d_{\mathrm{b}} n^{\frac{1}{2}} + O(1) \right).$$

Plugging in the asymptotic expansion for the standard deviation, we finally obtain, making use of $0 \leq \frac{I_r^{(n)}}{n} \leq 1$,

$$\begin{aligned}\left\| t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3 &\leq \frac{n^{\frac{1}{2}}}{\sigma_{\mathrm{b}}(n)}\left(\sum_{r=1}^{a+1}|d_{\mathrm{b}}|\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\frac{1}{2}}\right\|_3 + \sum_{r=a+1}^{K}|d_{\mathrm{w}}|\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\frac{1}{2}}\right\|_3 + |d_{\mathrm{b}}|\right)\\ &\quad + O\left(\frac{1}{\sqrt{n\ln(n)}}\right)\\ &= O\left((\ln(n))^{-\frac{1}{2}}\right).\end{aligned}$$

Case $0 < \lambda < \frac{1}{2}$:
Again from Lemma 5.1.i) with Notation 5.3, we have $\mu_{\mathrm{b}}(j) = c_{\mathrm{b}}n + d_{\mathrm{b}}n^{\lambda} + O(1)$. From Lemma 5.2.ii), we have that the standard deviation is of order $\sqrt{n}$. As in the previous case, linear terms of the expectations cancel out up to a constant, we are left with terms of order n^{λ} from the second order term of the expectations and obtain

$$t_{\mathrm{b}}\left(I^{(n)}\right) = \frac{1}{\sigma_{\mathrm{b}}(n)}\left(\sum_{r=1}^{a+1} d_{\mathrm{b}}\left(I_r^{(n)}\right)^{\lambda} + \sum_{r=a+2}^{K} d_{\mathrm{w}}\left(I_r^{(n)}\right)^{\lambda} - d_{\mathrm{b}}n^{\lambda} + O(1)\right)$$

and therefore, with $0 \leq \frac{I_r^{(n)}}{n} \leq 1$,

$$\begin{aligned}&\left\| t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3\\ \leq\ & \frac{n^{\lambda}}{\sigma_{\mathrm{b}}(n)}\left(\sum_{r=1}^{a+1}|d_{\mathrm{b}}|\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\right\|_3 + \sum_{r=a+2}^{K}|d_{\mathrm{w}}|\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\right\|_3 + |d_{\mathrm{b}}|\right) + O\left(n^{-\frac{1}{2}}\right)\\ =\ & O\left(n^{\lambda-\frac{1}{2}}\right).\end{aligned}$$

Case $\lambda < 0$:
Finally, we have, from Lemma 5.1.ii), $\mu_{\mathrm{b}}(j) = c_{\mathrm{b}}j + O(1)$. From Lemma 5.2.iii) we know that the standard deviation is of order $\sqrt{n}$. As before, linear terms of the expectations cancel out and only terms of constant order remain. It immediately follows

$$\left\| t_{\mathrm{b}}\left(I^{(n)}\right)\right\|_3 = O\left(n^{-\frac{1}{2}}\right).$$

□

Lemma 5.20. *In the situation of Theorem 5.15, we have for the quantities appearing in the estimates for* $\zeta_3^\vee\left(\left(\mathscr{Q}_n^{\mathrm{b}}, \mathscr{Q}_n^{\mathrm{w}}\right), (\mathcal{N}(0,1), \mathcal{N}(0,1))\right)$ *of Proposition 5.18, as* $n \to \infty$,

$$\left\| \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - 1 \right\|_{\frac{3}{2}} = \begin{cases} O\left((\ln(n))^{-1}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{2\left(\lambda - \frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-1}\right), & \lambda < 0, \end{cases}$$

as well as

$$\left\| \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} - 1 \right\|_{\frac{3}{2}} = \begin{cases} O\left((\ln(n))^{-1}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{2\left(\lambda - \frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-1}\right), & \lambda < 0. \end{cases}$$

Proof. By plugging in the asymptotic expansions for the variance given in Lemma 5.2, we observe that the ratio of the variances $\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)/\sigma_{\mathrm{b}}^2(n)$ comes out to be roughly the proportion of leaves belonging to the r-th subtree to all leaves of the associated tree. The limit of these quantities also appears hidden in $1 = \sum_{r=1}^K D_r$. Hence, the undistorted subtree sizes that satisfy $\sum_{r=1}^K I_r^{(n)} = n - 1$ are brought into play:

$$\begin{aligned} &\left\| \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - 1 \right\|_{\frac{3}{2}} \\ = &\left\| \sum_{r=1}^{K} \frac{I_r^{(n)}}{n} - 1 + \left(\sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right) + \left(\sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right) \right\|_{\frac{3}{2}} \\ = &\left\| \frac{n-1}{n} - 1 + \left(\sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right) + \left(\sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right) \right\|_{\frac{3}{2}} \\ \leq &\ \frac{1}{n} + \sum_{r=1}^{a+1} \left\| \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} + \sum_{r=a+2}^{K} \left\| \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} . \end{aligned}$$

Comparing the ratios of the variances to the rescaled subtree sizes, we again have three cases to deal with due to the different behaviours of the variance depending on the range of λ. Note that $\sigma_{\mathrm{b}}(j) = \sigma_{\mathrm{w}}(j) = 0$ for $j = 1, 2$. Hence, when plugging in the asymptotic expansions of Lemma 5.2 we add an indicator for the event $\left\{I_r^{(n)} \geq 2\right\}$.

Case $\lambda = \frac{1}{2}$:
Using Lemma 5.2.i) for the variances, we estimate

$$\begin{aligned}
& \sum_{r=1}^{a+1} \left\| \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} + \sum_{r=a+2}^{K} \left\| \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} \\
\leq \; & \sum_{r=1}^{K} \left\| \frac{I_r^{(n)} \ln\left(I_r^{(n)}\right) + O(n)}{n \ln(n) + O(n)} \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} \\
\leq \; & \sum_{r=1}^{K} \left\| \frac{n I_r^{(n)} \ln\left(I_r^{(n)}\right) - n I_r^{(n)} \ln(n) + O\left(n^2\right)}{n\left(n \ln(n) + O(n)\right)} \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}} \right\|_{\frac{3}{2}} + O\left(\frac{1}{n}\right) \\
\leq \; & (1 + o(1)) \frac{1}{\ln(n)} \sum_{r=1}^{K} \left\| \frac{I_r^{(n)}}{n} \ln\left(\frac{I_r^{(n)}}{n}\right) \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}} \right\|_{\frac{3}{2}} + O\left(\frac{1}{\ln(n)}\right) \\
= \; & O\left(\frac{1}{\ln(n)}\right),
\end{aligned}$$

where we used that $|x \ln(x)| \leq \frac{1}{e}$ for $x \in [0,1]$ with $0 \ln(0) := 0$ as continuous extension of $x \mapsto x \ln(x)$ for $x \geq 0$. We obtain

$$\left\| \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - 1 \right\|_{\frac{3}{2}} = O\left(\frac{1}{\ln(n)}\right).$$

Case $0 < \lambda < \frac{1}{2}$:
With Lemma 5.2.ii) for the variances, we estimate

$$\begin{aligned}
& \sum_{r=1}^{a+1} \left\| \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} + \sum_{r=a+2}^{K} \left\| \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} \\
\leq \; & \sum_{r=1}^{K} \left\| \frac{I_r^{(n)} + O\left(n^{2\lambda}\right)}{n + O\left(n^{2\lambda}\right)} \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} \\
\leq \; & \sum_{r=1}^{K} \left\| \frac{n I_r^{(n)} - n I_r^{(n)} + O\left(n^{2\lambda+1}\right)}{n\left(n + O\left(n^{2\lambda}\right)\right)} \right\|_{\frac{3}{2}} + O\left(\frac{1}{n}\right) \\
= \; & O\left(n^{2\left(\lambda - \frac{1}{2}\right)}\right).
\end{aligned}$$

We obtain

$$\left\| \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - 1 \right\|_{\frac{3}{2}} = O\left(n^{2\left(\lambda-\frac{1}{2}\right)}\right).$$

Case $\lambda < 0$:
With Lemma 5.2.iii), we finally can estimate

$$\begin{aligned}
&\sum_{r=1}^{a+1} \left\| \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} + \sum_{r=a+2}^{K} \left\| \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} \\
\leq\ & \sum_{r=1}^{K} \left\| \frac{I_r^{(n)} + O(1)}{n + O(1)} \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}} - \frac{I_r^{(n)}}{n} \right\|_{\frac{3}{2}} \\
\leq\ & \sum_{r=1}^{K} \left\| \frac{nI_r^{(n)} - nI_r^{(n)} + O(n)}{n(n + O(1))} \right\|_{\frac{3}{2}} + O\left(\frac{1}{n}\right) \\
=\ & O\left(n^{-1}\right),
\end{aligned}$$

yielding for negative λ

$$\left\| \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}^2(n)} - 1 \right\|_{\frac{3}{2}} = O\left(\frac{1}{n}\right).$$

Obviously, calculations for

$$\left\| \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}^2\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}^2(n)} - 1 \right\|_{\frac{3}{2}}$$

work in exactly the same way and lead to the same bounds. Thus, the assertion follows. □

We plug the results of Lemma 5.19 and Lemma 5.20 into Proposition 5.18 and obtain an upper bound for the distance between accompanying sequence and limit:

Corollary 5.21. *In the situation of Theorem 5.15 and Proposition 5.18, it holds, as $n \to \infty$,*

$$\zeta_3^{\vee}\left(\left(\mathscr{Q}_n^{\mathrm{b}}, \mathscr{Q}_n^{\mathrm{w}}\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda-\frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{3}{2}}\right), & \lambda < 0. \end{cases}$$

Ad Step 3: Merging Previous Results

Now, we want to reap the fruits of the former labour, **Step 1** and **Step 2**, and prove an upper bound for $\zeta_3^{\vee}\left(\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right), \left(\mathcal{N}\left(0,1\right), \mathcal{N}\left(0,1\right)\right)\right)$. This will be accomplished via induction on the basis of Proposition 5.17 together with Proposition 5.18 and Corollary 5.21, respectively.

In Proposition 5.17 that describes $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}\left(0,1\right)\right)$ recursively in terms of the distances $\zeta_3\left(\hat{\mathscr{X}}_j, \mathcal{N}\left(0,1\right)\right) = \hat{\Delta}\left(j\right)$ with $j \in 0, \ldots, n-1$ and therefore is the core to derive a rate inductively, the ratios $\left(\sigma_{\mathrm{b}}(j)/\sigma_{\mathrm{b}}(n)\right)^3$ and $\left(\sigma_{\mathrm{w}}(j)/\sigma_{\mathrm{b}}(n)\right)^3$ are crucial. So, again, the asymptotic expansions for the variance in Lemma 5.2 play an important role and depending on the range of λ three different cases for the induction will arise. The ratios $\left(\sigma_{\mathrm{b}}(j)/\sigma_{\mathrm{b}}(n)\right)^3$ and $\left(\sigma_{\mathrm{w}}(j)/\sigma_{\mathrm{b}}(n)\right)^3$ are treated with the help of Lemma A.3 of the appendix. Moreover, note that $\hat{\Delta}\left(0\right)$ and $\hat{\Delta}\left(1\right)$ do not contribute in any of the following calculations as they are accompanied by the factors $\sigma_{\mathrm{b}}\left(j\right) = 0$ or $\sigma_{\mathrm{w}}\left(j\right) = 0$ with $j = 0, 1$, respectively.

The following corollaries present the rates for the cases $\lambda = \frac{1}{2}$, $0 < \lambda < \frac{1}{2}$, and $\lambda < 0$ concluding the proof of Theorem 5.15.

The constants $A, A', A'', B, C, D > 0$, and $\delta, \delta', \xi \in (0,1)$ as well as $n_0 \in \mathbb{N}$ occur several times in Corollaries 5.22, 5.23 and 5.24 with potentially different meanings but during one and the same induction their meaning does not change — they are to be thought of as "local variables". Whenever they appear they are positive and chosen such that the estimates hold for sufficiently large n, i.e., $n \geq n_0$.

Corollary 5.22. *In the situation of Theorem 5.15, if $\lambda = \frac{1}{2}$, it holds, as $n \to \infty$,*

$$\zeta_3^{\vee}\left(\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right), \left(\mathcal{N}\left(0,1\right), \mathcal{N}\left(0,1\right)\right)\right) = O\left(\left(\ln\left(n\right)\right)^{-\frac{3}{2}}\right).$$

Proof. We start with the ratios $\left(\sigma_{\mathrm{b}}(j)/\sigma_{\mathrm{b}}(n)\right)^3$ and $\left(\sigma_{\mathrm{w}}(j)/\sigma_{\mathrm{b}}(n)\right)^3$: From Lemma 5.2.i), we have $\sigma_{\mathrm{b}}^2\left(j\right), \sigma_{\mathrm{w}}^2\left(j\right) \leq bc\, j \ln\left(j\right) + Aj$ for $j \geq 2$ with a suitable constant $A > 0$. Applying Lemma A.3 as done in the appendix in (A.2), we obtain, with constants $A', A'' > 0$,

$$\begin{aligned}\left(\frac{\sigma_{\mathrm{b}}\left(j\right)}{\sigma_{\mathrm{b}}\left(n\right)}\right)^3 &= \left(\frac{bcj\ln\left(j\right) + Aj}{bcn\ln\left(n\right) + O\left(n\right)}\right)^{\frac{3}{2}} = \left(\frac{j\ln\left(j\right) + A'j}{n\ln\left(n\right) + O\left(n\right)}\right)^{\frac{3}{2}} \\ &\leq \frac{\left(j\ln\left(j\right)\right)^{\frac{3}{2}} + A''j^{\frac{3}{2}}\left(\ln\left(j\right)\right)^{\frac{1}{2}}}{\left(n\ln\left(n\right) + O\left(n\right)\right)^{\frac{3}{2}}}.\end{aligned}$$

Obviously, we handle the ratio $\left(\sigma_{\mathrm{w}}(j)/\sigma_{\mathrm{b}}(n)\right)^3$ the same way.

Corollary 5.21 suggests a rate of order $\left(\ln\left(n\right)\right)^{-\frac{3}{2}}$. Hence, we set the induction hypothesis as

$$\hat{\Delta}\left(j\right) \leq C\left(\ln\left(j\right)\right)^{-\frac{3}{2}}, \quad j = 2, \ldots, n-1.$$

Finally, the induction is performed: The distance $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$ is estimated with the help of Proposition 5.17 and Corollary 5.21. The induction hypothesis does not cover the distances $\hat{\Delta}(0)$ and $\hat{\Delta}(1)$. Hence, the contributions of the event $\left\{I_r^{(n)} < 2\right\}$ have to be treated separately. Luckily, we observe that $\sigma_{\mathrm{b}}(j) = \sigma_{\mathrm{w}}(j) = 0$ for $j = 0, 1$. Thus, $\hat{\Delta}(j)$ with $j = 0, 1$ do not contribute. Therefore, the indicator for $\left\{I_r^{(n)} \geq 2\right\}$ can be added. Thus, it holds, with a suitable constant $B > 0$,

$$
\begin{aligned}
&\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right) \leq \zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right) + \zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \\
\leq\ & \mathbb{E}\left[\sum_{r=1}^{a+1}\left(\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K}\left(\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right)\right] \\
&+ \zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \\
\leq\ & \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)} \ln\left(I_r^{(n)}\right)}{n \ln(n) + O(n)}\right)^{\frac{3}{2}} \hat{\Delta}\left(I_r^{(n)}\right) \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}}\right] \\
&+ \mathbb{E}\left[\sum_{r=1}^{K}\frac{A''\left(I_r^{(n)}\right)^{\frac{3}{2}}\left(\ln\left(I_r^{(n)}\right)\right)^{\frac{1}{2}}}{\left(n \ln(n) + O(n)\right)^{\frac{3}{2}}} \hat{\Delta}\left(I_r^{(n)}\right) \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}}\right] + B\left(\ln(n)\right)^{-\frac{3}{2}} \\
\leq\ & \frac{(1 + o(1))\, C}{(n \ln(n))^{\frac{3}{2}}} \mathbb{E}\left[\sum_{r=1}^{K}\left(I_r^{(n)}\right)^{\frac{3}{2}} \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}}\right] \\
&+ \frac{(1 + o(1))\, C A''}{(n \ln(n))^{\frac{3}{2}}} \mathbb{E}\left[\left(I_r^{(n)}\right)^{\frac{3}{2}}\left(\ln\left(I_r^{(n)}\right)\right)^{-1} \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}}\right] + B\left(\ln(n)\right)^{-\frac{3}{2}} \\
\leq\ & (1 + o(1))\ C\left(\ln(n)\right)^{-\frac{3}{2}}\ \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\frac{3}{2}}\right] \\
&+ C\left(\ln(n)\right)^{-\frac{3}{2}}\ (1 + o(1))\, A''\, n^{-\frac{3}{2}}\, \mathbb{E}\left[\sum_{r=1}^{K}\left(I_r^{(n)}\right)^{\frac{3}{2}}\left(\ln\left(I_r^{(n)}\right)\right)^{-1} \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}}\right] \qquad (5.60) \\
&+ B\left(\ln(n)\right)^{-\frac{3}{2}} \\
\leq\ & (1 - \delta)\, C\left(\ln(n)\right)^{-\frac{3}{2}} + C\left(\ln(n)\right)^{-\frac{3}{2}}\ D\left(\ln(n)\right)^{-1} + B\left(\ln(n)\right)^{-\frac{3}{2}} \qquad (5.61) \\
\leq\ & \left(1 - (\delta - \delta')\right) C\left(\ln(n)\right)^{-\frac{3}{2}} + B\left(\ln(n)\right)^{-\frac{3}{2}} \qquad (5.62)
\end{aligned}
$$

with the following observations:

In (5.60) we observed

$$\begin{aligned}
&\mathbb{E}\left[\sum_{r=1}^{K}\left(I_r^{(n)}\right)^{\frac{3}{2}}\left(\ln\left(I_r^{(n)}\right)\right)^{-1}\mathbb{1}_{\left\{I_r^{(n)}\geq 2\right\}}\right]\\
=\;&\mathbb{E}\left[\sum_{r=1}^{K}\left(I_r^{(n)}\right)^{\frac{3}{2}}\left(\ln\left(I_r^{(n)}\right)\right)^{-1}\left(\mathbb{1}_{\left\{2\leq I_r^{(n)}\leq\left\lfloor\frac{n}{\ln(n)}\right\rfloor\right\}}+\mathbb{1}_{\left\{\left\lceil\frac{n}{\ln(n)}\right\rceil\leq I_r^{(n)}\leq n-1\right\}}\right)\right]\\
\leq\;&\frac{1}{\ln(2)}\sum_{r=1}^{K}\left\lfloor\frac{n}{\ln(n)}\right\rfloor^{\frac{3}{2}}+\frac{1}{\ln\left(\left\lceil\frac{n}{\ln(n)}\right\rceil\right)}\mathbb{E}\left[\sum_{r=1}^{K}\left(I_r^{(n)}\right)^{\frac{3}{2}}\right]\\
\leq\;&\frac{K}{\ln(2)}\left(\frac{n}{\ln(n)}\right)^{\frac{3}{2}}+\frac{n^{\frac{3}{2}}}{\ln(n)}\left(1+o(1)\right)\underbrace{\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\frac{3}{2}}\right]}_{\leq 1}\\
=\;&O\left(\frac{n^{\frac{3}{2}}}{\ln(n)}\right).
\end{aligned}$$

Hence, the constant D is chosen such that it satisfies for n large enough

$$\left(1+o(1)\right)A''n^{-\frac{3}{2}}\mathbb{E}\left[\sum_{r=1}^{K}\left(I_r^{(n)}\right)^{\frac{3}{2}}\left(\ln\left(I_r^{(n)}\right)\right)^{-1}\mathbb{1}_{\left\{I_r^{(n)}\geq 2\right\}}\right]\leq D\left(\ln(n)\right)^{-1}.$$

Equations (5.61) and (5.62) constitute the contractive behaviour, compare (4.13): Due to Lemma 3.4 combined with property (3.6), we know that

$$0<\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\frac{3}{2}}\right]\to\xi<1,\quad n\to\infty.$$

Hence, there is $0<\delta<1$ such that we have in (5.61)

$$\left(1+o(1)\right)\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\frac{3}{2}}\right]\leq 1-\delta$$

for n sufficiently large.

In addition to that, to obtain (5.61) we choose n large enough such that

$$D\left(\ln(n)\right)^{-1}\leq\delta'<\delta.$$

We now fix $n_0\in\mathbb{N}$ such that all these estimates hold for $n\geq n_0$ and choose

$$C\geq\max\left\{\hat{\Delta}(j)\left(\ln(j)\right)^{\frac{3}{2}}\,|j=2,\ldots,n_0-1\right\}\vee\frac{B}{\delta-\delta'}.$$

Then, we obtain

$$\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right) \leq C(\ln(n))^{-\frac{3}{2}} \quad \text{for } n \geq 2.$$

Treating the distance $\zeta_3\left(\hat{\mathscr{Y}}_n, \mathcal{N}(0,1)\right)$ accordingly, the assertion follows. □

For $\lambda \in \left(0, \frac{1}{2}\right)$, Corollary 5.21 makes us expect two "sorts" of rates: In Corollary 5.21, we see that the term $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ presents an accelerating rate whose exponent is linearly decreasing when λ decreases from $\frac{1}{2}$ to 0 and then abides with further decreasing λ. Moreover, for $\lambda < \frac{1}{3}$ this rate is faster than the order of the reciprocal of the standard deviation of the number of black balls (i.e., faster than $n^{-\frac{1}{2}}$). Hence, we expect that λ will appear in the exponent of the upper bound for $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$ until λ hits $\frac{1}{3}$ coming from $\frac{1}{2}$. Then, for $\lambda \leq \frac{1}{3}$, it is reasonable to expect the rate to stay constant and not to be faster than the reciprocal of the standard deviation of the number of black balls that is of order $n^{-\frac{1}{2}}$.

Unfortunately, we are not able to transfer a rate of order $n^{-\frac{1}{2}}$ to $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$ for $\lambda \leq \frac{1}{3}$ due to the fact that the induction with this rate does not yield the desired contractive behaviour, see Remark 8.1. Therefore, we have to diminish the rate to $n^{-\frac{1}{2}+\varepsilon}$ with arbitrarily small $\varepsilon > 0$.

Due to the slightly different behaviour of the variance for negative λ this case is treated later in Corollary 5.24; however, Corollary 5.23 and Corollary 5.24 could have been combined.

Corollary 5.23. *Let $\varepsilon > 0$. In the situation of Theorem 5.15, if $0 < \lambda < \frac{1}{2}$, it holds, as $n \to \infty$,*

$$\zeta_3^{\vee}\left(\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right) = \begin{cases} O\left(n^{3\left(\lambda-\frac{1}{2}\right)}\right), & \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon}\right), & 0 < \lambda \leq \frac{1}{3}. \end{cases}$$

Proof. From Lemma 5.2.ii), we have $\sigma_{\mathrm{b}}^2(j), \sigma_{\mathrm{w}}^2(j) \leq \frac{(a+b)bc(a-c)^2}{(a+b-2(a-c))(b+c)^2} j + Aj^{2\lambda}$ for $j \geq 0$ with a suitable constant $A > 0$. Applying Lemma A.3 as done in the appendix in (A.3) leads to

$$\left(\frac{\sigma_{\mathrm{b}}(j)}{\sigma_{\mathrm{b}}(n)}\right)^3 \leq \left(\frac{f_{\mathrm{b}} j + Aj^{2\lambda}}{f_{\mathrm{b}} n + O(n^{2\lambda})}\right)^{\frac{3}{2}} = \left(\frac{j + A' j^{2\lambda}}{n + O(n^{2\lambda})}\right)^{\frac{3}{2}} \leq \frac{j^{\frac{3}{2}} + A'' j^{2\lambda+\frac{1}{2}}}{\left(n + O(n^{2\lambda})\right)^{\frac{3}{2}}}$$

with suitable constants $A', A'' > 0$.

The cases $\frac{1}{3} < \lambda < \frac{1}{2}$ and $0 < \lambda < \frac{1}{3}$ are treated separately:

We start with $\frac{1}{3} < \lambda < \frac{1}{2}$ and set as induction hypothesis

$$\hat{\Delta}(j) \le C\, j^{3\left(\lambda - \frac{1}{2}\right)}, \quad j = 2, \ldots, n-1.$$

Evoking Proposition 5.17 and Corollary 5.21, we have with a suitable constant $B > 0$,

$$\begin{aligned}
&\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right) \le \zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right) + \zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \\
\le\; & \mathbb{E}\left[\sum_{r=1}^{a+1}\left(\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K}\left(\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right)\right] \\
& + \zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \\
\le\; & \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n + O\left(n^{2\lambda}\right)}\right)^{\frac{3}{2}} \hat{\Delta}\left(I_r^{(n)}\right) \mathbb{1}_{\left\{I_r^{(n)} \ge 2\right\}}\right] \\
& + \mathbb{E}\left[\sum_{r=1}^{K}\frac{A''\left(I_r^{(n)}\right)^{2\lambda+\frac{1}{2}}}{\left(n + O\left(n^{2\lambda}\right)\right)^{\frac{3}{2}}} \hat{\Delta}\left(I_r^{(n)}\right) \mathbb{1}_{\left\{I_r^{(n)} \ge 2\right\}}\right] + Bn^{3\left(\lambda - \frac{1}{2}\right)} \\
\le\; & \frac{1 + o(1)}{n^{\frac{3}{2}}}\, \mathbb{E}\left[\sum_{r=1}^{K} C\left(I_r^{(n)}\right)^{3\lambda} + CA''\left(I_r^{(n)}\right)^{5\lambda - 1}\right] + Bn^{3\left(\lambda - \frac{1}{2}\right)} \\
\le\; & (1 + o(1))\, Cn^{3\left(\lambda - \frac{1}{2}\right)}\, \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{3\lambda}\right] \\
& + (1 + o(1))\, n^{5\left(\lambda - \frac{1}{2}\right)} CA'' \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{5\lambda - 1}\right] + Bn^{3\left(\lambda - \frac{1}{2}\right)} \\
\le\; & (1 - \delta)\, Cn^{3\left(\lambda - \frac{1}{2}\right)} + DCn^{5\left(\lambda - \frac{1}{2}\right)} + Bn^{3\left(\lambda - \frac{1}{2}\right)} \qquad (5.63) \\
\le\; & \left(1 - (\delta - \delta')\right) Cn^{3\left(\lambda - \frac{1}{2}\right)} + Bn^{3\left(\lambda - \frac{1}{2}\right)} \qquad (5.64)
\end{aligned}$$

with the following observations:
In (5.63) we argue (as in (5.61)): With Lemma 3.4 and property (3.6), observing $3\lambda > 1$, we have

$$0 < \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{3\lambda}\right] \to \xi < 1, \quad n \to \infty,$$

so we can find $0 < \delta < 1$ such that

$$(1 + o(1))\, \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{3\lambda}\right] \le 1 - \delta$$

for n large enough. The constant D is chosen such that

$$D \geq (1 + o(1)) A'' \mathbb{E} \left[\sum_{r=1}^{K} \left(\frac{I_r^{(n)}}{n} \right)^{5\lambda - 1} \right]$$

because $\mathbb{E} \left[\sum_{r=1}^{K} \left(\frac{I_r^{(n)}}{n} \right)^{5\lambda-1} \right] \leq K$ due to $5\lambda - 1 > \frac{2}{3}$ and $\frac{I_r^{(n)}}{n} \leq 1$.

In (5.64), we choose $0 < \delta' < \delta$ and n large enough such that

$$0 < D n^{2\left(\lambda - \frac{1}{2}\right)} < \delta' < \delta.$$

In a nutshell: In (5.63) and (5.64) we exploit the contractive behaviour enabling us to complete the induction. We fix $n_0 \in \mathbb{N}$ such that all previous estimates hold for $n \geq n_0$ and choose

$$C \geq \max \left\{ \hat{\Delta}(j) \, j^{3\left(\frac{1}{2} - \lambda\right)} | j = 2, \ldots, n_0 - 1 \right\} \vee \frac{B}{\delta - \delta'}$$

and obtain, for $n \geq 2$,

$$\zeta_3 \left(\mathscr{X}_n, \mathcal{N}(0,1) \right) \leq C \, n^{3\left(\lambda - \frac{1}{2}\right)}.$$

We now turn to $0 < \lambda \leq \frac{1}{3}$. Let $\varepsilon > 0$ and set as induction hypothesis

$$\hat{\Delta}(j) \leq C \, j^{-\frac{1}{2} + \varepsilon}, \quad j = 2, \ldots, n-1.$$

As before, Proposition 5.17 and Corollary 5.21 yield, with suitable $B > 0$

$$\begin{aligned}
& \zeta_3 \left(\hat{\mathscr{X}_n}, \mathcal{N}(0,1) \right) \leq \zeta_3 \left(\hat{\mathscr{X}_n}, \mathscr{Q}_n^{\mathrm{b}} \right) + \zeta_3 \left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1) \right) \\
\leq \; & \mathbb{E} \left[\sum_{r=1}^{a+1} \left(\frac{\sigma_{\mathrm{b}} \left(I_r^{(n)} \right)}{\sigma_{\mathrm{b}}(n)} \right)^3 \hat{\Delta} \left(I_r^{(n)} \right) + \sum_{r=a+2}^{K} \left(\frac{\sigma_{\mathrm{w}} \left(I_r^{(n)} \right)}{\sigma_{\mathrm{b}}(n)} \right)^3 \hat{\Delta} \left(I_r^{(n)} \right) \right] \\
& + \zeta_3 \left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1) \right) \\
\leq \; & \mathbb{E} \left[\sum_{r=1}^{K} \left(\frac{I_r^{(n)}}{n + O(n^{2\lambda})} \right)^{\frac{3}{2}} \hat{\Delta} \left(I_r^{(n)} \right) \mathbb{1}_{\left\{ I_r^{(n)} \geq 2 \right\}} \right] \\
& + \mathbb{E} \left[\sum_{r=1}^{K} \frac{A'' \left(I_r^{(n)} \right)^{2\lambda + \frac{1}{2}}}{\left(n + O(n^{2\lambda})\right)^{\frac{3}{2}}} \hat{\Delta} \left(I_r^{(n)} \right) \mathbb{1}_{\left\{ I_r^{(n)} \geq 2 \right\}} \right] + B n^{3\left(\lambda - \frac{1}{2}\right)} \\
\leq \; & \frac{1 + o(1)}{n^{\frac{3}{2}}} \mathbb{E} \left[\sum_{r=1}^{K} C \left(I_r^{(n)} \right)^{1+\varepsilon} + C A'' \left(I_r^{(n)} \right)^{2\lambda + \varepsilon} \right] + B n^{3\left(\lambda - \frac{1}{2}\right)}
\end{aligned}$$

$$\begin{aligned}
&\leq (1+o(1))\, n^{-\frac{1}{2}+\varepsilon} C\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{1+\varepsilon}\right] \\
&\quad + (1+o(1))\, CA'' n^{-\frac{3}{2}+2\lambda+\varepsilon}\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda+\varepsilon}\right] + Bn^{3\left(\lambda-\frac{1}{2}\right)} \\
&\leq (1-\delta)\, Cn^{-\frac{1}{2}+\varepsilon} + CDn^{-\frac{3}{2}+2\lambda+\varepsilon} + Bn^{3\left(\lambda-\frac{1}{2}\right)} \qquad (5.65)\\
&\leq (1-(\delta-\delta'))\, Cn^{-\frac{1}{2}+\varepsilon} + Bn_0^{3\lambda-1-\varepsilon} n^{-\frac{1}{2}+\varepsilon}. \qquad (5.66)
\end{aligned}$$

In (5.65), as before, we use Lemma 3.4 and property (3.6) and have

$$0 < \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{1+\varepsilon}\right] \to \xi < 1, \quad n \to \infty,$$

and infer that there is $0 < \delta < 1$ such that

$$(1+o(1))\, \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{1+\varepsilon}\right] \leq 1-\delta$$

for sufficiently large n. We denote by D a constant that satisfies

$$D \geq (1+o(1))\, A''\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda+\varepsilon}\right]$$

for sufficiently large n (note that $\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda+\varepsilon}\right] \leq K$ due to $2\lambda+\varepsilon > 0$ and $\frac{I_r^{(n)}}{n} \leq 1$).

Then, in (5.66), we take $0 < \delta' < \delta$ such that for n large enough

$$0 < Dn^{2\left(\lambda-\frac{1}{2}\right)} < \delta' < \delta$$

and we choose $n_0 \in \mathbb{N}$ such that all preceding estimates hold for $n \geq n_0$.

We finally choose

$$C \geq \max\left\{\hat{\Delta}(j)\, j^{\frac{1}{2}-\varepsilon} \middle| j = 2, \ldots, n_0 - 1\right\} \vee \frac{Bn_0^{3\lambda-1-\varepsilon}}{\delta-\delta'}$$

leading to

$$\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right) \leq C\, n^{-\frac{1}{2}+\varepsilon} \quad \text{for } n \geq 2.$$

Treating $\zeta_3\left(\hat{\mathscr{Y}}_n, \mathcal{N}(0,1)\right)$ likewise yields the assertion. □

For negative λ, Corollary 5.21 yields a rate for $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ that is faster than the reciprocal of the standard deviation of the number of black balls. Hence, we do not expect to be able to transfer this rate to $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$. We expect the correct rate for $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$ to be of the order of the reciprocal of the standard deviation, i.e., $n^{-\frac{1}{2}}$, but, as before for $0 < \lambda \leq \frac{1}{3}$, the best we can get (with our methods) is $n^{-\frac{1}{2}+\varepsilon}$.

Corollary 5.24. *Let $\varepsilon > 0$. In the situation of Theorem 5.15, if $\lambda < 0$, it holds, as $n \to \infty$,*

$$\zeta_3^{\vee}\left(\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right) = O\left(n^{-\frac{1}{2}+\varepsilon}\right).$$

Proof. For the variances, we have with Lemma 5.2.iii) $\sigma_{\mathrm{b}}^2(j), \sigma_{\mathrm{w}}^2(j) \leq f_{\mathrm{b}} j + A$ for $j \geq 0$ with a suitable constant $A > 0$. Hence, for the ratios $\frac{\sigma_{\mathrm{b}}(j)}{\sigma_{\mathrm{b}}(n)}$ and $\frac{\sigma_{\mathrm{w}}(j)}{\sigma_{\mathrm{b}}(n)}$ occurring in Proposition 5.17, we estimate with Lemma A.3 applied as in (A.4) with suitable constants $A', A'' > 0$

$$\left(\frac{\sigma_{\mathrm{b}}(j)}{\sigma_{\mathrm{b}}(n)}\right)^3 \leq \left(\frac{f_{\mathrm{b}} j + A}{f_{\mathrm{b}} n + O(1)}\right)^{\frac{3}{2}} = \left(\frac{j + A'}{n + O(1)}\right)^{\frac{3}{2}} \leq \frac{j^{\frac{3}{2}} + A'' j^{\frac{1}{2}}}{(n + O(1))^{\frac{3}{2}}}.$$

We set as induction hypothesis with an arbitrary $\varepsilon > 0$

$$\hat{\Delta}(j) \leq C\, j^{-\frac{1}{2}+\varepsilon}, \quad j = 2, \ldots, n-1.$$

The induction in this case passes off analogously to the induction in the case $0 < \lambda \leq \frac{1}{3}$ with slight variations due to the estimate for the variances. Plugging that into Proposition 5.17 and making use of Corollary 5.21, we obtain, with a suitable constant $B > 0$,

$$\begin{aligned}
&\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right) \leq \zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right) + \zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \\
\leq\ & \mathbb{E}\left[\sum_{r=1}^{a+1} \left(\frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K} \left(\frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)}\right)^3 \hat{\Delta}\left(I_r^{(n)}\right)\right] \\
&+ \zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \\
\leq\ & \mathbb{E}\left[\sum_{r=1}^{K} \left(\frac{I_r^{(n)}}{n + O(1)}\right)^{\frac{3}{2}} \hat{\Delta}\left(I_r^{(n)}\right) \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}}\right] \qquad (5.67) \\
&+ \mathbb{E}\left[\sum_{r=1}^{K} \frac{A''\left(I_r^{(n)}\right)^{\frac{1}{2}}}{(n + O(1))^{\frac{3}{2}}} \hat{\Delta}\left(I_r^{(n)}\right) \mathbb{1}_{\left\{I_r^{(n)} \geq 2\right\}}\right] \\
&+ B n^{-\frac{3}{2}} \\
\leq\ & \frac{1 + o(1)}{n^{\frac{3}{2}}} C \mathbb{E}\left[\sum_{r=1}^{K} \left(\left(I_r^{(n)}\right)^{\frac{3}{2}} + \left(I_r^{(n)}\right)^{\frac{1}{2}}\right) \hat{\Delta}\left(I_r^{(n)}\right)\right] + B n^{-\frac{3}{2}}
\end{aligned}$$

$$\begin{aligned}
&= (1+o(1))\, n^{-\frac{1}{2}+\varepsilon} C\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{1+\varepsilon}\right] \\
&\quad + (1+o(1))\, n^{-\frac{3}{2}+\varepsilon} A'' C\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\varepsilon}\right] + Bn^{-\frac{3}{2}} \\
\text{(5.68)} \quad &\leq (1-\delta)\, Cn^{-\frac{1}{2}+\varepsilon} + CDn^{-\frac{3}{2}+\varepsilon} + Bn^{-\frac{3}{2}} \\
\text{(5.69)} \quad &\leq (1-(\delta-\delta'))\, Cn^{-\frac{1}{2}+\varepsilon} + Bn_0^{-(1+\varepsilon)} n^{-\frac{1}{2}+\varepsilon}.
\end{aligned}$$

In (5.68), the reasoning is exactly the same as for (5.65) yielding that there is $0 < \delta < 1$ such that

$$(1+o(1))\, \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{1+\varepsilon}\right] \leq 1-\delta$$

for n sufficiently large. Furthermore, we used a constant D satisfying

$$D \geq (1+o(1))\, A''\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\varepsilon}\right]$$

for sufficiently large n, noting $\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\varepsilon}\right] \leq K$ due to $\varepsilon > 0$ and $\frac{I_r^{(n)}}{n} \leq 1$.

In (5.69), we choose $0 < \delta' < \delta$ such that for n large enough

$$0 < Dn^{-1} < \delta' < \delta$$

and fix $n_0 \in \mathbb{N}$ such that all estimates hold for $n \geq n_0$. Finally, we choose

$$C \geq \max\left\{\hat{\Delta}(j)\, j^{\frac{1}{2}-\varepsilon} | j = 2, \ldots, n_0 - 1\right\} \vee \frac{B}{n_0^{1+\varepsilon}(\delta-\delta')}$$

and therefore obtain, for $n \geq 2$,

$$\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right) \leq C\, n^{-\frac{1}{2}+\varepsilon}.$$

The same reasoning applies to $\zeta_3\left(\hat{\mathscr{Y}}_n, \mathcal{N}(0,1)\right)$, completing the proof. □

Resumption of the Proof of Theorem 5.15. We bounded the distance $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$ via the triangle inequality by $\zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right)$ and $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ and studied these distances in order to derive rates for $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$ via

Step 1 that yields a recursive description of $\zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right)$ and $\zeta_3\left(\hat{\mathscr{Y}}_n, \mathcal{N}(0,1)\right)$ in Proposition 5.17;

Step 2 where we derive upper bounds for $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ and $\zeta_3\left(\mathscr{Q}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right)$, respectively, in Corollary 5.21 on the basis of Proposition 5.18; and finally,

Step 3 that combines the results of **Step 1** and **Step 2** in an induction deriving upper bounds for the rate of $\zeta_3^{\vee}\left(\left(\hat{\mathscr{X}}_n, \hat{\mathscr{Y}}_n\right), \mathcal{N}(0,1)\right)$ in Corollaries 5.22, 5.23 and 5.24.

This concludes the proof of Theorem 5.15. □

6. Rates of Convergence for a Two-Colour Pólya Urn with Random Replacement

In this chapter, the Pólya urn characterised by setting **Rand R** is studied with respect to rates of convergence. This chapter should be read in parallel with Chapter 5 since the same reasoning as before is conducted.

In Chapter 5, determining upper bounds for rates of convergence could mostly be broken down to a thorough treatment of the coefficients of the distributional recursions (5.3) and (5.42) for the normalised numbers of black balls. The coefficients are indeed closely related to the subtree sizes that enter our reasoning via the recursive approach detailed in Chapter 3. In setting **Det R**, the behaviour of the subtree sizes had not been captured in the best possible way, cf. Remark 3.6; whereas in setting **Rand R**, the rescaled subtree sizes and their limit can easily be coupled optimally. Hence, this is done and we want to outline the differences.

Note that setting **Rand R** is the simplest example of a balanced and irreducible two-colour Pólya urn scheme where the balance equals one — therefore, we study the setting **Rand R** in order to understand the problems of setting **Det R**.

At first, the details of setting **Rand R** are stated. Subsequently, the non-normal limit case is sketched. The treatment of the normal limit case follows.

Recall setting **Rand R** from the introduction:

Randomised Play-the-Winner Rule

(**Rand R**)
$$\bar{R} = \begin{pmatrix} C_\alpha & 1 - C_\alpha \\ 1 - C_\beta & C_\beta \end{pmatrix} \text{ with } C_\alpha \sim \mathrm{Ber}\,(\alpha)\,, C_\beta \sim \mathrm{Ber}\,(\beta)\,,$$
$$\alpha, \beta \in (0,1)\,.$$

This urn comes with two coins, a "black" and a "white" coin. In every step, one ball is drawn uniformly at random from the urn. The colour of the drawn ball determines which coin is tossed. Then, the outcome of the coin toss determines the colour of the ball that is added to the urn when the drawn ball is returned to the urn.

Let $\tilde{R} := \left(\begin{smallmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{smallmatrix}\right)$ be the matrix that contains the means of the entries of the replacement matrix $\bar{R}$. Then, the crucial parameter determining the regime of asymptotic behaviour of the normalised number of black balls is given by the ratio of the eigenvalues of $\tilde{R}$, i.e., $\lambda := \alpha + \beta - 1$. As in Chapter 5, the non-normal limit case, i.e., $\lambda > \frac{1}{2}$, is studied first, followed by the normal limit case, i.e., $\lambda \leq \frac{1}{2}$.

The parameter λ takes values in the interval $(-1, 1)$, where the case $\lambda = 0$ is special: Unlike the situation in setting **Det R**, the behaviour of the number of black balls is not deterministic. When $\lambda = 0$ the two coins our urn comes with have the same success probability, so there is only one coin, and, therefore, the number of black balls is the number of initial black balls plus a sum of independent ($\mathrm{Ber}\,(\alpha)$-distributed) random variables, leading to the setting of the classical Central Limit Theorem. The case $\lambda = 0$ is treated in Remark 6.23.

In terms of the recursive approach displayed in Chapter 3, we decompose the number of black balls at time n recursively into the contributions of black leaves of the two subtrees of the root of the associated tree. For the second subtree, the coin toss in the first step of the urn decides whether it is a b-associated or a w-associated tree. By I_n we denote the number of draws from the first subtree which belongs to the original root node of the associated tree (the initial ball of the urn) and by J_n the number of draws from the second subtree that belongs to the ball added in the first step.

Then, the recursive approach from Section 3.2 yields the following distributional recursions, recall (3.2), for the number of black balls after n steps, beginning with a black or a white ball, respectively, with $B_0^{\mathrm{b}} := 1$ and $B_0^{\mathrm{w}} := 0$:

$$(6.1) \qquad \begin{aligned} B_n^{\mathrm{b}} &\stackrel{\mathrm{d}}{=} B_{I_n}^{\mathrm{b},(1)} + C_\alpha B_{J_n}^{\mathrm{b},(2)} + (1 - C_\alpha)\, B_{J_n}^{\mathrm{w}}, \\ B_n^{\mathrm{w}} &\stackrel{\mathrm{d}}{=} B_{I_n}^{\mathrm{w},(1)} + (1 - C_\beta)\, B_{J_n}^{\mathrm{b}} + C_\beta B_{J_n}^{\mathrm{w},(2)}, \end{aligned}$$

where $B_j^{\mathrm{b},(r)} \stackrel{\mathrm{d}}{=} B_j^{\mathrm{b}}$ and $B_j^{\mathrm{w},(r)} \stackrel{\mathrm{d}}{=} B_j^{\mathrm{w}}$ for $r = 1, 2$ and $j = 0, \ldots, n-1$, such that $\left(B_j^{\mathrm{b},(r)}\right)_{j \geq 1}$, $\left(B_j^{\mathrm{w},(r)}\right)_{j \geq 1}$, $\left(B_j^{\mathrm{b}}\right)_{j \geq 1}$, $\left(B_j^{\mathrm{w}}\right)_{j \geq 1}$, C_α, C_β, and $I^{(n)} = (I_n, J_n)$ are independent.

Note that I_n and J_n are uniformly distributed on $\{0, \ldots, n-1\}$ with $I_n + J_n = n - 1$. From Lemma 3.4 we know the behaviour of the rescaled subtree sizes:

$$\left(\frac{I_n}{n}, \frac{J_n}{n}\right) \to (U, 1 - U), \quad n \to \infty, \quad \text{a.s. and in } L_p,\ p \geq 1,$$

with U uniformly distributed on $[0,1]$. As all our recurrences take place on the level of distributions we take the liberty to choose

$$I_n := \lfloor nU \rfloor \text{, and accordingly, } J_n = n - 1 - \lfloor nU \rfloor . \tag{6.2}$$

Therefore, we have Lemma 3.7 and Lemma 3.8 at hand when it comes to handling the behaviour of the rescaled subtree sizes compared to their limits.

Lemma 6.1. *Depending on λ, the following holds for the means of the number of black balls $\mathbb{E}\left[B_n^{\mathrm{b}}\right]$ and $\mathbb{E}\left[B_n^{\mathrm{w}}\right]$, as $n \to \infty$:*

i) $\lambda > 0$*:*

$$\mathbb{E}\left[B_n^{\mathrm{b}}\right] = \frac{1-\beta}{1-\lambda} n + \frac{1-\alpha}{1-\lambda} \frac{1}{\Gamma(\lambda+1)} n^{\lambda} + O(1),$$
$$\mathbb{E}\left[B_n^{\mathrm{w}}\right] = \frac{1-\beta}{1-\lambda} n - \frac{1-\beta}{1-\lambda} \frac{1}{\Gamma(\lambda+1)} n^{\lambda} + O(1).$$

ii) $\lambda < 0$*:*

$$\mathbb{E}\left[B_n^{\mathrm{b}}\right] = \frac{1-\beta}{1-\lambda} n + O(1),$$
$$\mathbb{E}\left[B_n^{\mathrm{w}}\right] = \frac{1-\beta}{1-\lambda} n + O(1).$$

Proof. See Chapter A.2 of the appendix. □

Lemma 6.2. *Depending on λ, for the variances of the number of black balls* $\mathrm{Var}\left(B_n^{\mathrm{b}}\right)$ *and* $\mathrm{Var}\left(B_n^{\mathrm{w}}\right)$*, it holds, as $n \to \infty$:*

i) $\lambda = \frac{1}{2}$*:*

$$\mathrm{Var}\left(B_n^{\mathrm{b}}\right) = 4(1-\alpha)(1-\beta) n \ln(n) + O(n),$$
$$\mathrm{Var}\left(B_n^{\mathrm{w}}\right) = 4(1-\alpha)(1-\beta) n \ln(n) + O(n).$$

ii) $0 < \lambda < \frac{1}{2}$*:*

$$\mathrm{Var}\left(B_n^{\mathrm{b}}\right) = \frac{(1-\alpha)(1-\beta)}{(1-2\lambda)(1-\lambda)^2} n + O\left(n^{2\lambda}\right),$$
$$\mathrm{Var}\left(B_n^{\mathrm{w}}\right) = \frac{(1-\alpha)(1-\beta)}{(1-2\lambda)(1-\lambda)^2} n + O\left(n^{2\lambda}\right).$$

iii) $\lambda < 0$*:*

$$\operatorname{Var}\left(B_n^{\mathrm{b}}\right) = \frac{(1-\alpha)(1-\beta)}{(1-2\lambda)(1-\lambda)^2} n + O(1),$$
$$\operatorname{Var}\left(B_n^{\mathrm{w}}\right) = \frac{(1-\alpha)(1-\beta)}{(1-2\lambda)(1-\lambda)^2} n + O(1).$$

Proof. See Chapter A.2 of the appendix. □

Notation 6.3. We use the following abbreviations

$$\tilde{c}_{\mathrm{b}} = \frac{1-\beta}{1-\lambda}, \quad \tilde{d}_{\mathrm{b}} = \frac{1-\alpha}{1-\lambda}\frac{1}{\Gamma(\lambda+1)}, \quad \tilde{d}_{\mathrm{w}} = -\frac{1-\beta}{1-\lambda}\frac{1}{\Gamma(\lambda+1)}.$$

Furthermore, the mean of the j-th quantities B_j^{b} and B_j^{w} is abbreviated by $\tilde{\mu}_{\mathrm{b}}(j)$ and $\tilde{\mu}_{\mathrm{w}}(j)$, respectively, as well as their standard deviations by $\tilde{\sigma}_{\mathrm{b}}(j)$ and $\tilde{\sigma}_{\mathrm{w}}(j)$, and their variances by $\tilde{\sigma}_{\mathrm{b}}^2(j)$ and $\tilde{\sigma}_{\mathrm{w}}^2(j)$, respectively. As before, we use these abbreviations with random arguments I_n and J_n. Compare Remark 5.4.

6.1. Non-Normal Limit Case: $\lambda > \frac{1}{2}$

In this section, all results are analogous to the results of Section 5.1. Their proofs are very briefly sketched, since the choice (6.2) for the subtree sizes does not influence the results; the tighter estimates for the asymptotic behaviour of the rescaled subtree sizes do not yield any improvement compared to the non-normal limit case in setting **Det R**. The details of the proofs were already checked in [29].

As before in the non-normal limit case of setting **Det R**, the number of black balls after n steps is centred around the expectation and scaled by the order of the standard deviation. Let $\mathcal{X}_0 := 0 =: \mathcal{Y}_0$ and, for $n \geq 1$,

$$\begin{aligned} \mathcal{X}_n :=& \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{n^\lambda}, \\ \mathcal{Y}_n :=& \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{n^\lambda}. \end{aligned} \tag{6.3}$$

The distributional recursions of (6.1) lead to, $n \geq 1$,

$$\begin{aligned} \mathcal{X}_n &\stackrel{\mathrm{d}}{=} \left(\frac{I_n}{n}\right)^\lambda \mathcal{X}_{I_n}^{(1)} + C_\alpha \left(\frac{J_n}{n}\right)^\lambda \mathcal{X}_{J_n}^{(2)} + (1 - C_\alpha) \left(\frac{J_n}{n}\right)^\lambda \mathcal{Y}_{J_n} + \tilde{b}_{\mathrm{b}}(I_n), \\ \mathcal{Y}_n &\stackrel{\mathrm{d}}{=} \left(\frac{I_n}{n}\right)^\lambda \mathcal{Y}_{I_n}^{(1)} + C_\beta \left(\frac{J_n}{n}\right)^\lambda \mathcal{Y}_{J_n}^{(2)} + (1 - C_\beta) \left(\frac{J_n}{n}\right)^\lambda \mathcal{X}_{J_n} + \tilde{b}_{\mathrm{w}}(I_n) \end{aligned} \tag{6.4}$$

with toll terms (and with Lemma 6.1.i))

$$\begin{aligned} \tilde{b}_{\mathrm{b}}(I_n) :=& \frac{1}{n^\lambda}\left(\tilde{\mu}_{\mathrm{b}}(I_n) + C_\alpha \tilde{\mu}_{\mathrm{b}}(J_n) + (1 - C_\alpha)\tilde{\mu}_{\mathrm{w}}(J_n) - \tilde{\mu}_{\mathrm{b}}(n)\right) \\ =& \tilde{d}_{\mathrm{b}}\left(\left(\frac{I_n}{n}\right)^\lambda + C_\alpha \left(\frac{J_n}{n}\right)^\lambda - 1\right) + \tilde{d}_{\mathrm{w}}(1 - C_\alpha)\left(\frac{J_n}{n}\right)^\lambda + O\left(n^{-\lambda}\right), \\ \tilde{b}_{\mathrm{w}}(I_n) :=& \frac{1}{n^\lambda}\left(\tilde{\mu}_{\mathrm{w}}(I_n) + (1 - C_\beta)\tilde{\mu}_{\mathrm{b}}(J_n) + C_\beta \tilde{\mu}_{\mathrm{w}}(J_n) - \tilde{\mu}_{\mathrm{w}}(n)\right) \\ =& \tilde{d}_{\mathrm{w}}\left(\left(\frac{I_n}{n}\right)^\lambda + C_\beta \left(\frac{J_n}{n}\right)^\lambda - 1\right) + \tilde{d}_{\mathrm{b}}(1 - C_\beta)\left(\frac{J_n}{n}\right)^\lambda + O\left(n^{-\lambda}\right), \end{aligned} \tag{6.5}$$

where $\mathcal{X}_j^{(1)} \stackrel{\mathrm{d}}{=} \mathcal{X}_j \stackrel{\mathrm{d}}{=} \mathcal{X}_j^{(2)}$ and $\mathcal{Y}_j^{(1)} \stackrel{\mathrm{d}}{=} \mathcal{Y}_j \stackrel{\mathrm{d}}{=} \mathcal{Y}_j^{(2)}$ for $j = 0, \ldots, n-1$, such that $(\mathcal{X}_j)_{0 \leq j \leq n-1}$, $\left(\mathcal{X}_j^{(1)}\right)_{0 \leq j \leq n-1}$, $\left(\mathcal{X}_j^{(2)}\right)_{0 \leq j \leq n-1}$, $(\mathcal{Y}_j)_{0 \leq j \leq n-1}$, $\left(\mathcal{Y}_j^{(1)}\right)_{0 \leq j \leq n-1}$, $\left(\mathcal{Y}_j^{(2)}\right)_{0 \leq j \leq n-1}$, I_n, C_α, and C_β are independent.

Formally letting $n \to \infty$, we expect the following system of distributional fixed-point equations to hold for a limit $(\mathcal{X},\mathcal{Y})$ of $(\mathcal{X}_n,\mathcal{Y}_n)_{n\in\mathbb{N}}$:

$$\begin{aligned}\mathcal{X} &\stackrel{\mathrm{d}}{=} U^\lambda \mathcal{X}^{(1)} + C_\alpha (1-U)^\lambda \mathcal{X}^{(2)} + (1-C_\alpha)(1-U)^\lambda \mathcal{Y} + \tilde{b}_{\mathrm{b}},\\ \mathcal{Y} &\stackrel{\mathrm{d}}{=} U^\lambda \mathcal{Y}^{(1)} + C_\beta (1-U)^\lambda \mathcal{Y}^{(2)} + (1-C_\beta)(1-U)^\lambda \mathcal{X} + \tilde{b}_{\mathrm{w}}\end{aligned} \tag{6.6}$$

with toll terms

$$\begin{aligned}\tilde{b}_{\mathrm{b}} &:= \tilde{d}_{\mathrm{b}}\left(U^\lambda + C_\alpha(1-U)^\lambda - 1\right) + \tilde{d}_{\mathrm{w}}(1-C_\alpha)(1-U)^\lambda,\\ \tilde{b}_{\mathrm{w}} &:= \tilde{d}_{\mathrm{w}}\left(U^\lambda + C_\beta(1-U)^\lambda - 1\right) + \tilde{d}_{\mathrm{b}}(1-C_\beta)(1-U)^\lambda,\end{aligned} \tag{6.7}$$

where $\mathcal{X}^{(1)} \stackrel{\mathrm{d}}{=} \mathcal{X} \stackrel{\mathrm{d}}{=} \mathcal{X}^{(1)}$ and $\mathcal{Y}^{(1)} \stackrel{\mathrm{d}}{=} \mathcal{Y} \stackrel{\mathrm{d}}{=} \mathcal{Y}^{(1)}$ as well as $\mathcal{X}$, $\mathcal{X}^{(1)}$, $\mathcal{X}^{(2)}$, $\mathcal{Y}$, $\mathcal{Y}^{(1)}$, $\mathcal{Y}^{(2)}$, U, C_α, and C_β are independent.

Theorem 4.1 applies to the system of fixed-point equations given by (6.6). Hence, there is a unique solution of system (6.6) in the Cartesian product of the space of centred probability measures with finite second moment $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$ and it will be denoted by $\left(\mathcal{L}\left(\tilde{\Lambda}_{\mathrm{b}}\right), \mathcal{L}\left(\tilde{\Lambda}_{\mathrm{w}}\right)\right)$.

As before in the non-normal limit case of setting **Det R**, we bound the rate of convergence in all Wasserstein distances and in the Kolmogorov-Smirnov distance:

Theorem 6.4 (Twin of Theorem 5.6)**.** *Given a Pólya urn scheme characterised by* ***Rand R*** *with $\lambda := \alpha + \beta - 1 > \frac{1}{2}$, let $\mathcal{X}_n$ and $\mathcal{Y}_n$ be defined as in* (6.3) *and let $\left(\mathcal{L}\left(\tilde{\Lambda}_{\mathrm{b}}\right), \mathcal{L}\left(\tilde{\Lambda}_{\mathrm{w}}\right)\right)$ denote the unique solution of system* (6.6) *in $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$. Let $p \geq 1$ and $\varepsilon > 0$. Then, as $n \to \infty$,*

$$\begin{aligned}\ell_p^\vee\left((\mathcal{X}_n,\mathcal{Y}_n), \left(\tilde{\Lambda}_{\mathrm{b}}, \tilde{\Lambda}_{\mathrm{w}}\right)\right) &= O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right),\\ \varrho^\vee\left((\mathcal{X}_n,\mathcal{Y}_n), \left(\tilde{\Lambda}_{\mathrm{b}}, \tilde{\Lambda}_{\mathrm{w}}\right)\right) &= O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right).\end{aligned}$$

Theorem 6.4 follows from Proposition 6.9 and Proposition 6.10.

Remark 6.5. In the same manner as Remark 5.7 we make use of optimal couplings. We assume conditions (5.7), (5.8) and (5.9) — stated in the setting **Det R** — to hold analogously for the respective quantities of setting **Rand R**.

As in the proof of Theorem 5.6, we will state certain interim results that will constitute the proof of Theorem 6.4. The reasoning to prove these interim results and thereby Theorem 6.4 remains the same as for Theorem 5.6. Hence, the proofs are not repeated in detail.

Proposition 6.6 (Twin of Proposition 5.8)**.** *Consider a Pólya urn scheme characterised by* ***Rand R*** *where* $\lambda := \alpha+\beta-1 > \frac{1}{2}$*. Let* $\mathcal{X}_n$ *and* $\mathcal{Y}_n$ *be defined as in* (6.3)*, let* $\left(\mathcal{L}\left(\hat{\Lambda}_{\mathrm{b}}\right), \mathcal{L}\left(\hat{\Lambda}_{\mathrm{w}}\right)\right)$ *denote the unique solution of system* (6.6) *in* $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$ *and* $\varepsilon > 0$*. Then, as* $n \to \infty$*,*

$$\ell_2^{\vee}\left(\left(\mathcal{X}_n, \mathcal{Y}_n\right), \left(\tilde{\Lambda}_{\mathrm{b}}, \tilde{\Lambda}_{\mathrm{w}}\right)\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right).$$

As Proposition 5.8, Proposition 6.6 is proven via induction. The following three lemmata constitute the proof of Proposition 6.6: Lemma 6.7 figures as backdrop for the induction. Lemma 3.8 with $p = 2$ and $\psi = \lambda$ (hence, $\psi > \frac{p-1}{p}$) as well as Lemma 6.8 serve as preparatory tools in order to finally complete the induction.

Let $\mathscr{D}(j) := \ell_2^{\vee}\left(\left(\mathcal{X}_j, \mathcal{Y}_j\right), \left(\tilde{\Lambda}_{\mathrm{b}}, \tilde{\Lambda}_{\mathrm{w}}\right)\right)$ as well as $\mathscr{D}^2(j) := \left(\ell_2^{\vee}\left(\left(\mathcal{X}_j, \mathcal{Y}_j\right), \left(\tilde{\Lambda}_{\mathrm{b}}, \tilde{\Lambda}_{\mathrm{w}}\right)\right)\right)^2$, $j \in \mathbb{N}$. The quantities $\mathscr{D}(j)$ and $\mathscr{D}^2(j)$ are treated according to Remark 5.9.

Lemma 6.7 (Twin of Lemma 5.10)**.** *In the situation of Proposition 6.6, it holds, for* $n \geq 1$*,*

$$\begin{aligned}
\mathscr{D}^2(n) \quad \leq \quad & 2\,\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda} \mathscr{D}^2(I_n)\right] + 2\tilde{L}^2 \left\| \left(\frac{I_n}{n}\right)^{\lambda} - U^{\lambda} \right\|_2^2 \\
& + 4\tilde{L}\,\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{\lambda} \left| \left(\frac{I_n}{n}\right)^{\lambda} - U^{\lambda} \right| \mathscr{D}(I_n)\right] \\
& + \max\left\{ \left\| \tilde{b}_{\mathrm{b}}(I_n) - \tilde{b}_{\mathrm{b}} \right\|_2^2, \left\| \tilde{b}_{\mathrm{w}}(I_n) - \tilde{b}_{\mathrm{w}} \right\|_2^2 \right\}
\end{aligned}$$

with $\tilde{L} := \max\left\{\left\|\tilde{\Lambda}_{\mathrm{b}}\right\|_2, \left\|\tilde{\Lambda}_{\mathrm{w}}\right\|_2\right\}$*.*

Sketch of Proof. The estimates go through exactly as in the proof of Lemma 5.10. ⊠

Lemma 6.8 (Twin of Lemma 5.11)**.** *For the toll terms* $\tilde{b}_{\mathrm{b}}(I_n)$ *and* $\tilde{b}_{\mathrm{w}}(I_n)$ *defined in* (6.5) *compared to* $\tilde{b}_{\mathrm{b}}$ *and* $\tilde{b}_{\mathrm{w}}$ *defined in* (6.7)*, it holds, as* $n \to \infty$*,*

$$\max\left\{ \left\| \tilde{b}_{\mathrm{b}}(I_n) - \tilde{b}_{\mathrm{b}} \right\|_p, \left\| \tilde{b}_{\mathrm{w}}(I_n) - \tilde{b}_{\mathrm{w}} \right\|_p \right\} = O\left(n^{-\lambda}\right).$$

Sketch of Proof. Due to Lemma 3.8, we can find a tighter estimate (tighter than the estimate obtained in Lemma 5.11) for the behaviour of the toll terms, by comparing with the proof of Lemma 5.11 and inserting the result of Lemma 3.8 with $p = 2$ and $\psi = \lambda$ (hence, $\psi > \frac{p-1}{p}$) in the right places. Note that in this case due to Lemma 3.8 the exponent of the rate is twice the exponent of the rate in the situation of Lemma 5.11. ⊠

Using Proposition 6.6 as base case, we derive rates of convergence in all Wasserstein distances via induction on p (and n, again).

Proposition 6.9 (Twin of Proposition 5.13)**.** *Given a Pólya urn scheme characterised by* ***Rand R*** *where* $\lambda := \alpha + \beta - 1 > \frac{1}{2}$*, let* $\mathcal{X}_n$ *and* $\mathcal{Y}_n$ *be defined as in* (6.3)*. Furthermore, let* $\left(\mathcal{L}\left(\tilde{\Lambda}_{\mathrm{b}}\right), \mathcal{L}\left(\tilde{\Lambda}_{\mathrm{w}}\right)\right)$ *denote the unique solution of system* (6.6) *in* $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$*. Let* $p \geq 1$ *and* $\varepsilon > 0$*. Then, as* $n \to \infty$*,*

$$\ell_p^{\vee}\left(\left(\mathcal{X}_n, \mathcal{Y}_n\right), \left(\tilde{\Lambda}_b, \tilde{\Lambda}_w\right)\right) = O\left(n^{-\lambda + \frac{1}{2} + \varepsilon}\right).$$

Sketch of Proof. Rewriting $\ell_p\left(\mathcal{X}_n, \tilde{\Lambda}_{\mathrm{b}}\right)$ by means of Lemma 5.12 we proceed analogously to the proof of Proposition 5.13. ⊠

Finally, we transfer the rates obtained in Wasserstein distances to the Kolmogorov-Smirnov distance:

Proposition 6.10 (Twin of Proposition 5.14)**.** *Consider a Pólya urn scheme characterised by* ***Rand R*** *where* $\lambda := \alpha + \beta - 1 > \frac{1}{2}$ *and let* $\mathcal{X}_n$ *and* $\mathcal{Y}_n$ *be defined as in* (6.3)*. Let* $\left(\mathcal{L}\left(\tilde{\Lambda}_{\mathrm{b}}\right), \mathcal{L}\left(\tilde{\Lambda}_{\mathrm{w}}\right)\right)$ *denote the unique solution of system* (6.6) *in* $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$ *and let* $\varepsilon > 0$*. Then, as* $n \to \infty$*,*

$$\varrho^{\vee}\left(\left(\mathcal{X}_n, \mathcal{Y}_n\right), \left(\tilde{\Lambda}_b, \tilde{\Lambda}_w\right)\right) = O\left(n^{-\lambda + \frac{1}{2} + \varepsilon}\right).$$

Sketch of Proof. As in the proof of Proposition 5.14, we take advantage of Lemma 2.7. The boundedness of the densities of $\tilde{\Lambda}_{\mathrm{b}}$ and $\tilde{\Lambda}_{\mathrm{w}}$ is proven in Leckey [33, Theorem 3.4]. All further steps of the proof are analogous to the proof of Proposition 5.14. ⊠

6.2. Normal Limit Case: $\lambda \le \frac{1}{2}$

In the normal limit case, the tighter estimates available in Lemma 3.7 and Lemma 3.8 make a difference in the approach suggested by Lemma 2.7 for estimating the distance between accompanying sequence and limit, compare introductory part of **Step 2**, p. 76 et seq. in Chapter 5. Recall the situation of p. 101 as well as Notation 6.3.

Therefore, the approach via Lemma 2.7 will be displayed in detail, whereas the approach based on Proposition 5.18 will be sketched briefly.

In order to obtain the standard normal distribution as unique fixed-point of the limiting equation for the normalised number of black balls, it is centred around its mean and scaled by the standard deviation. Let $\hat{\mathcal{X}}_0 := 0 =: \hat{\mathcal{Y}}_0$ and, for $n \ge 1$,

$$\hat{\mathcal{X}}_n := \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{\sqrt{\operatorname{Var}\left(B_n^{\mathrm{b}}\right)}}, \quad \hat{\mathcal{Y}}_n := \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{\sqrt{\operatorname{Var}\left(B_n^{\mathrm{w}}\right)}}. \tag{6.8}$$

This transfers to the distributional recurrence obtained in (6.1) as

$$\begin{aligned} \hat{\mathcal{X}}_n &\stackrel{\mathrm{d}}{=} \frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \hat{\mathcal{X}}_{I_n}^{(1)} + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \hat{\mathcal{X}}_{J_n}^{(2)} + (1 - C_\alpha) \frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \hat{\mathcal{Y}}_{J_n} + \hat{t}_{\mathrm{b}}(I_n), \\ \hat{\mathcal{Y}}_n &\stackrel{\mathrm{d}}{=} \frac{\tilde{\sigma}_{\mathrm{w}}(I_n)}{\tilde{\sigma}_{\mathrm{w}}(n)} \hat{\mathcal{Y}}_{I_n}^{(1)} + C_\beta \frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{w}}(n)} \hat{\mathcal{Y}}_{J_n}^{(2)} + (1 - C_\beta) \frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{w}}(n)} \hat{\mathcal{X}}_{J_n} + \hat{t}_{\mathrm{w}}(I_n) \end{aligned} \tag{6.9}$$

with

$$\begin{aligned} \hat{t}_{\mathrm{b}}(I_n) &:= \frac{1}{\tilde{\sigma}_{\mathrm{b}}(n)} \left(\tilde{\mu}_{\mathrm{b}}(I_n) + C_\alpha \tilde{\mu}_{\mathrm{b}}(J_n) + (1 - C_\alpha)\tilde{\mu}_{\mathrm{w}}(J_n) - \tilde{\mu}_{\mathrm{b}}(n)\right), \\ \hat{t}_{\mathrm{w}}(I_n) &:= \frac{1}{\tilde{\sigma}_{\mathrm{w}}(n)} \left(\tilde{\mu}_{\mathrm{w}}(I_n) + C_\beta \tilde{\mu}_{\mathrm{w}}(J_n) + (1 - C_\beta)\tilde{\mu}_{\mathrm{b}}(J_n) - \tilde{\mu}_{\mathrm{w}}(n)\right) \end{aligned} \tag{6.10}$$

with similar conditions on distributions and independence as in (6.1). We keep the choice $I_n := \lfloor nU \rfloor$ of (6.2) with U uniformly distributed on $[0,1]$. Formally letting $n \to \infty$, we expect a limit $\left(\mathcal{L}\left(\hat{\mathcal{X}}\right), \mathcal{L}\left(\hat{\mathcal{Y}}\right)\right)$ of $\left(\mathcal{L}\left(\hat{\mathcal{X}}_n\right), \mathcal{L}\left(\hat{\mathcal{Y}}_n\right)\right)_{n \in \mathbb{N}}$ due to Lemma 3.4 to satisfy the following distributional fixed-point equation

$$\begin{aligned} \hat{\mathcal{X}} &\stackrel{\mathrm{d}}{=} \sqrt{U}\hat{\mathcal{X}}^{(1)} + C_\alpha \sqrt{1-U}\hat{\mathcal{X}}^{(2)} + (1 - C_\alpha)\sqrt{1-U}\hat{\mathcal{Y}}, \\ \hat{\mathcal{Y}} &\stackrel{\mathrm{d}}{=} \sqrt{U}\hat{\mathcal{Y}}^{(1)} + C_\beta \sqrt{1-U}\hat{\mathcal{Y}}^{(2)} + (1 - C_\beta)\sqrt{1-U}\hat{\mathcal{X}}, \end{aligned} \tag{6.11}$$

where U is uniformly distributed on $[0,1]$ and $\hat{\mathcal{X}}^{(1)} \stackrel{\mathrm{d}}{=} \hat{\mathcal{X}} \stackrel{\mathrm{d}}{=} \hat{\mathcal{X}}^{(2)}$ as well as $\hat{\mathcal{Y}}^{(1)} \stackrel{\mathrm{d}}{=} \hat{\mathcal{Y}} \stackrel{\mathrm{d}}{=} \hat{\mathcal{Y}}^{(2)}$ such that $\hat{\mathcal{X}}$, $\hat{\mathcal{X}}^{(1)}$, $\hat{\mathcal{X}}^{(2)}$, $\hat{\mathcal{Y}}$, $\hat{\mathcal{Y}}^{(1)}$, $\hat{\mathcal{Y}}^{(2)}$, C_α, C_β, and U are independent.

Associating the system (6.11) with a self-map as in Chapter 4 puts us in the situation of Theorem 4.2. Hence, we know that the unique solution of system (6.11) in $\mathcal{M}_3(0,1) \times \mathcal{M}_3(0,1)$ is given by $(\mathcal{N}(0,1), \mathcal{N}(0,1))$.

Theorem 6.11 (Twin of Theorem 5.15)**.** *Given a Pólya urn scheme characterised by* ***Rand R*** *with* $\lambda := \alpha + \beta - 1 \leq \frac{1}{2}$*, let* $\hat{\mathcal{X}}_n$ *and* $\hat{\mathcal{Y}}_n$ *be as in* (6.8) *and let* $\varepsilon > 0$*. Then, as* $n \to \infty$*,*

$$\zeta_3^{\vee}\left(\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Y}}_n\right), (\mathcal{N}(0,1), \mathcal{N}(0,1))\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \text{if } \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda - \frac{1}{2}\right)}\right), & \text{if } \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon}\right), & \text{if } -1 < \lambda \leq \frac{1}{3}. \end{cases}$$

As before, in setting **Det R**, we measure the distance of $\hat{\mathcal{X}}_n$ to its limit $\mathcal{N}(0,1)$ in the Zolotarev distance by comparing both with a suitable accompanying sequence. The comparison of the limit $\mathcal{N}(0,1)$ and the accompanying sequence is crucial and determines the "best" rate that could be obtained, when finally proving a rate for $\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right)$.

In the introductory part of **Step 2**, p. 76 et seq., it was explained that there are at least two ways of estimating the Zolotarev distance between the accompanying sequence and the standard normal distribution: Firstly, Lemma 2.7 that finds an upper bound in the ℓ_3-Wasserstein metric; secondly, the approach introduced in Proposition 5.18.

The approach via Lemma 2.7 can be considered the more common way, hence it served as first choice in tackling the distance between accompanying sequence and limit. The most unsatisfactory part in bounding via ℓ_3 was that Corollary 3.5 was not strong enough to deliver reasonable estimates in setting **Det R**. Now, we are better off: We are able to couple the rescaled subtree sizes to their limits explicitly, see (6.2). This enables us to give more precise estimates in the study of the behaviour of the rescaled subtree sizes, see Lemmata 3.7 and 3.8, than displayed in Lemma 3.4 and Corollary 3.5. Thus, we will demonstrate both approaches to highlight the differences to setting **Det R**.

For the rates stated in Theorem 6.11 in the cases with positive λ, i.e., $\lambda = \frac{1}{2}$ and $0 < \lambda < \frac{1}{2}$, the approach of Proposition 5.18 for the distance between accompanying sequence and limit is needed. For negative λ, both approaches, i.e., the approach via Lemma 2.7 and the analogue of Proposition 5.18, yield the rate of Theorem 6.11. Note that in setting **Det R** the ℓ_3-estimates tracing back to Lemma 2.7 were not even satisfactory for negative λ.

Proof of Theorem 6.11. The itinerary of our proof will be as in the proof of Theorem 5.15: We want to find estimates for $\zeta_3^{\vee}\left(\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Y}}_n\right), (\mathcal{N}(0,1), \mathcal{N}(0,1))\right)$. At first, we will squeeze in the accompanying sequences $\hat{\mathcal{Q}}_n^{\mathrm{b}}$ and $\hat{\mathcal{Q}}_n^{\mathrm{w}}$ that act as linkage between the sequences $\hat{\mathcal{X}}_n$ and $\hat{\mathcal{Y}}_n$ and their limit $\mathcal{N}(0,1)$. Let the accompanying sequences be given by

$$\hat{\mathcal{Q}}_n^{\mathrm{b}} := \frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} N_1 + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} N_2 + (1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} N_3 + \hat{t}_{\mathrm{b}}(I_n),\, n \geq 1, \tag{6.12}$$

$$\hat{\mathcal{Q}}_n^{\mathrm{w}} := \frac{\tilde{\sigma}_{\mathrm{w}}(I_n)}{\tilde{\sigma}_{\mathrm{w}}(n)} N_1 + C_\beta \frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{w}}(n)} N_2 + (1-C_\beta)\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{w}}(n)} N_3 + \hat{t}_{\mathrm{w}}(I_n),\, n \geq 1, \tag{6.13}$$

with $N_1, N_2, N_3 \sim \mathcal{N}(0,1)$ independent. Note that for $n \geq 1$, $\mathbb{E}\left[\hat{\mathcal{X}}_n\right] = \mathbb{E}\left[\hat{\mathcal{Y}}_n\right] = \mathbb{E}\left[\hat{\mathcal{Q}}_n^{\mathrm{b}}\right] = \mathbb{E}\left[\hat{\mathcal{Q}}_n^{\mathrm{w}}\right] = 0$ and $\mathrm{Var}\left(\hat{\mathcal{X}}_n\right) = \mathrm{Var}\left(\hat{\mathcal{Y}}_n\right) = \mathrm{Var}\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}\right) = \mathrm{Var}\left(\hat{\mathcal{Q}}_n^{\mathrm{w}}\right) = 1$ as well as $\left\|\hat{\mathcal{X}}_n\right\|_3$, $\left\|\hat{\mathcal{Y}}_n\right\|_3$, $\left\|\hat{\mathcal{Q}}_n^{\mathrm{b}}\right\|_3$, $\left\|\hat{\mathcal{Q}}_n^{\mathrm{w}}\right\|_3 < \infty$. Therefore, both distances $\zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right)$ and $\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ as well as $\zeta_3\left(\hat{\mathcal{Y}}_n, \hat{\mathcal{Q}}_n^{\mathrm{w}}\right)$ and $\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right)$ are finite.

We perform all reasoning for the first component $\hat{\mathcal{X}}_n$ as proofs work out analogously for the second component $\hat{\mathcal{Y}}_n$. Hence, we will derive an estimate for $\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right)$. The proof of Theorem 6.11 breaks down to three steps (as in the proof of Theorem 5.15):

Step 1* Denoting by $\hat{\mathscr{D}}(j) := \zeta_3^{\vee}\left(\left(\hat{\mathcal{X}}_j, \hat{\mathcal{Y}}_j\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right)$ the maximal Zolotarev distance between the j-th member of the sequence and its limit, we derive a recursive estimate of $\zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right)$ in terms of $\hat{\mathscr{D}}(j)$ with $j = 0, \ldots, n-1$ (when we write $\hat{\mathscr{D}}(I_n)$ we mean the distance of the randomly chosen I_n-th quantity to the limit, it does *not* equal $\zeta_3^{\vee}\left(\left(\hat{\mathcal{X}}_{I_n}, \hat{\mathcal{Y}}_{I_n}\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right)$, see Remark 5.9). **Step 1*** is analogous to **Step 1** in setting **Det R**.

Step 2* Next, we want to appoint a candidate for the rate of convergence by studying $\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$. Before we sketch the analogue of Proposition 5.18, we demonstrate how to bound $\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ in ℓ_3 in order to illustrate the problem arising from Corollary 3.5 in setting **Det R**.

Step 3* Finally, we will put all these estimates together and proof a rate of convergence via induction.

Firstly, by the triangle inequality the distance of interest is bounded by

$$\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right) \leq \zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right) + \zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right). \tag{6.14}$$

As in the proof of Theorem 5.15 in setting **Det R**, we treat the three steps **Step 1***, **Step 2*** and **Step 3*** separately. We sever the proof of Theorem 6.11 and engage with the three steps that will finally conclude this proof. ⧄

Ad Step 1*: Recursive Description of $\zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right)$ and $\zeta_3\left(\hat{\mathcal{Y}}_n, \hat{\mathcal{Q}}_n^{\mathrm{w}}\right)$

Proposition 6.12 (Twin of Proposition 5.17)**.** *In the situation of Theorem 6.11 with $\hat{\mathcal{Q}}_n^{\mathrm{b}}$ and $\hat{\mathcal{Q}}_n^{\mathrm{w}}$ defined as in* (6.12) *and* (6.13)*, it holds, for $n \in \mathbb{N}$,*

$$\begin{aligned}
&\zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right)\\
\leq\ & \mathbb{E}\left[\left(\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3 \hat{\mathscr{D}}(I_n) + \left(\alpha\left(\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3 + (1-\alpha)\left(\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3\right)\hat{\mathscr{D}}(J_n)\right],\\
&\zeta_3\left(\hat{\mathcal{Y}}_n, \hat{\mathcal{Q}}_n^{\mathrm{w}}\right)\\
\leq\ & \mathbb{E}\left[\left(\frac{\tilde{\sigma}_{\mathrm{w}}(I_n)}{\tilde{\sigma}_{\mathrm{w}}(n)}\right)^3 \hat{\mathscr{D}}(I_n) + \left((1-\beta)\left(\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{w}}(n)}\right)^3 + \beta\left(\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{w}}(n)}\right)^3\right)\hat{\mathscr{D}}(J_n)\right].
\end{aligned}$$

Proof. Keep in mind that I_n is uniformly distributed on $\{0, \ldots, n-1\}$ and $I_n + J_n = n - 1$. We condition on I_n and obtain due to independence

$$\begin{aligned}
&\zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right) = \sup_{f\in\mathcal{F}_3}\left|\mathbb{E}\left[f\left(\hat{\mathcal{X}}_n\right) - f\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}\right)\right]\right|\\
=\ & \sup_{f\in\mathcal{F}_3}\left|\mathbb{E}\left[\mathbb{E}\left[f\left(\hat{\mathcal{X}}_n\right) - f\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}\right)|I_n\right]\right]\right|\\
=\ & \sup_{f\in\mathcal{F}_3}\left|\sum_{j=0}^{n-1}\frac{1}{n}\mathbb{E}\left[f\left(\hat{\mathcal{X}}_n\right) - f\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}\right)|I_n = j\right]\right|\\
\leq\ & \sum_{j=0}^{n-1}\frac{1}{n}\sup_{f\in\mathcal{F}_3}\left|\mathbb{E}\left[f\left(\hat{\mathcal{X}}_n\right) - f\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}\right)|I_n = j\right]\right|\\
=\ & \sum_{j=0}^{n-1}\frac{1}{n}\zeta_3\left(\frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\hat{\mathcal{X}}_j^{(1)} + C_\alpha\frac{\tilde{\sigma}_{\mathrm{b}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\hat{\mathcal{X}}_{n-1-j}^{(2)}\right.\\
&\qquad + (1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\hat{\mathcal{Y}}_{n-1-j} + \hat{t}_{\mathrm{b}}(j),\\
&\qquad \frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_1 + C_\alpha\frac{\tilde{\sigma}_{\mathrm{b}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_2\\
&\qquad \left. + (1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_3 + \hat{t}_{\mathrm{b}}(j)\right)\\
\leq\ & \sum_{j=0}^{n-1}\frac{1}{n}\left[\zeta_3\left(\frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\hat{\mathcal{X}}_j^{(1)}, \frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_1\right)\right.\\
&\quad + \zeta_3\left(C_\alpha\frac{\tilde{\sigma}_{\mathrm{b}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\hat{\mathcal{X}}_{n-1-j}^{(2)}, C_\alpha\frac{\tilde{\sigma}_{\mathrm{b}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_2\right)\\
&\quad \left. + \zeta_3\left((1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\hat{\mathcal{Y}}_{n-1-j}, (1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_3\right) + 0\right]
\end{aligned} \tag{6.15}$$

$$
\begin{aligned}
(6.16) \quad &\leq \sum_{j=0}^{n-1} \frac{1}{n} \Bigg[\left(\frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \zeta_3 \left(\hat{\mathscr{X}}_j^{(1)}, N_1 \right) + \left(\frac{\tilde{\sigma}_{\mathrm{b}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \zeta_3 \left(C_\alpha \hat{\mathscr{X}}_{n-1-j}^{(2)}, C_\alpha N_2 \right) \\
&\qquad + \left(\frac{\tilde{\sigma}_{\mathrm{w}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \zeta_3 \left((1-C_\alpha) \hat{\mathscr{Y}}_{n-1-j}, (1-C_\alpha) N_3 \right) \Bigg] \\
(6.17) \quad &= \sum_{j=0}^{n-1} \frac{1}{n} \Bigg[\left(\frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \zeta_3 \left(\hat{\mathscr{X}}_j^{(1)}, N_1 \right) + \alpha \left(\frac{\tilde{\sigma}_{\mathrm{b}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \zeta_3 \left(\hat{\mathscr{X}}_{n-1-j}^{(2)}, N_2 \right) \\
&\qquad + (1-\alpha) \left(\frac{\tilde{\sigma}_{\mathrm{w}}(n-1-j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \zeta_3 \left(\hat{\mathscr{Y}}_{n-1-j}, N_3 \right) \Bigg] \\
&\leq \mathbb{E} \left[\left(\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \hat{\mathscr{D}}(I_n) + \alpha \left(\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \hat{\mathscr{D}}(J_n) + (1-\alpha) \left(\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \hat{\mathscr{D}}(J_n) \right],
\end{aligned}
$$

where we used that ζ_3 is $(3,+)$-ideal in (6.15) and (6.16). In (6.17) we conditioned on the events $\{C_\alpha = 1\}$ and $\{C_\alpha = 0\}$. □

Ad Step 2*: Determining a Candidate for the Rate of Convergence

For studying the Zolotarev distance between accompanying sequence and limit, we first illustrate the approach suggested by Lemma 2.7 ("option 1"). Afterwards we sketch the analogue of Proposition 5.18 as "option 2".

Option 1: Bounding by the ℓ_3-Wasserstein distance

To begin with, we will give estimates obtained via Lemma 2.7 in distinction from **Step 2** in Chapter 5 in setting **Det R**.

Proposition 6.13. *In the situation of Theorem 6.11 with $\hat{\mathscr{Q}}_n^{\mathrm{b}}$ and $\hat{\mathscr{Q}}_n^{\mathrm{w}}$ as in* (6.12) *and* (6.13)*, it holds, as $n \to \infty$,*

$$
\zeta_3^{\vee} \left(\left(\hat{\mathscr{Q}}_n^{\mathrm{b}}, \hat{\mathscr{Q}}_n^{\mathrm{w}} \right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1) \right) \right) = \begin{cases} O\left((\ln(n))^{-\frac{1}{2}} \right), & \lambda = \frac{1}{2}, \\ O\left(n^{\lambda - \frac{1}{2}} \right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}} \right), & -1 < \lambda < 0. \end{cases}
$$

Proof. To get hold of the latter distance of the right-hand side of (6.14) we use the link between Zolotarev and Wasserstein distance, Lemma 2.7:

$$
\zeta_3 \left(\hat{\mathscr{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1) \right) \leq \left(\mathbb{E} \left[\left| \hat{\mathscr{Q}}_n^{\mathrm{b}} \right|^3 \right]^{\frac{2}{3}} + \mathbb{E} \left[|N|^{\frac{2}{3}} \right] \right) \ell_3 \left(\hat{\mathscr{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1) \right),
$$

where $N := \sqrt{U}N_1 + C_\alpha\sqrt{1-U}N_2 + (1-C_\alpha)\sqrt{1-U}N_3$ is standard normally distributed with N_1, N_2, N_3 standard normally distributed, independent of all other quantities.

With (subsequent) Lemma 6.14, the quantity $\mathbb{E}\left[\left|\hat{\mathcal{Q}}_n^{\mathrm{b}}\right|^3\right]^{\frac{2}{3}}$ is uniformly bounded in n:

$$\begin{aligned} & \left\|\hat{\mathcal{Q}}_n^{\mathrm{b}}\right\|_3 \\ \leq\ & \left\|\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_1\right\|_3 + \left\|C_\alpha\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_2\right\|_3 + \left\|(1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_3\right\|_3 + \left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3 \\ \leq\ & 3\left\|N_1\right\|_3 + o(1). \end{aligned}$$

Hence, the expression $\mathbb{E}\left[\left|\hat{\mathcal{Q}}_n^{\mathrm{b}}\right|^3\right]^{\frac{2}{3}} + \mathbb{E}\left[|N|^{\frac{2}{3}}\right]$ is uniformly bounded in n and we can proceed to $\ell_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$:

$$\begin{aligned} & \ell_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \leq \left\|\hat{\mathcal{Q}}_n^{\mathrm{b}} - N\right\|_3 \\ =\ & \left\|\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_1 + C_\alpha\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_2 + (1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}N_3 + \hat{t}_{\mathrm{b}}(I_n)\right. \\ & \left. -\sqrt{U}N_1 - C_\alpha\sqrt{1-U}N_2 - (1-C_\alpha)\sqrt{1-U}N_3\right\|_3 \\ \leq\ & \left\|N_1\right\|_3\left\|\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{U}\right\|_3 + \left\|N_2\right\|_3\alpha^{\frac{1}{3}}\left\|\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{1-U}\right\|_3 \\ & + \left\|N_3\right\|_3(1-\alpha)^{\frac{1}{3}}\left\|\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{1-U}\right\|_3 + \left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3. \end{aligned}$$

The behaviour of $\left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3$ is studied in Lemma 6.14. Estimates for the terms $\left\|\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{U}\right\|_3$, $\left\|\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{1-U}\right\|_3$ and $\left\|\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{1-U}\right\|_3$ are given in Lemma 6.15. Lemmata 6.14 and 6.15 yield, for symmetry reasons,

$$\begin{aligned} & \ell_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \\ \leq\ & \left\|N_1\right\|_3\left\|\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{U}\right\|_3 + \left\|N_2\right\|_3\alpha^{\frac{1}{3}}\left\|\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{1-U}\right\|_3 \\ & + \left\|N_3\right\|_3(1-\alpha)^{\frac{1}{3}}\left\|\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{1-U}\right\|_3 + \left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3 \\ =\ & \begin{cases} O\left((\ln(n))^{-\frac{1}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{\lambda-\frac{1}{2}}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}}\right), & -1 < \lambda < 0. \end{cases} \end{aligned}$$

The assertion follows. □

Lemma 6.14 (Twin of Lemma 5.19)**.** *For the toll terms $\hat{t}_{\mathrm{b}}(I_n)$ and $\hat{t}_{\mathrm{w}}(I_n)$ defined in* (6.10)*, it holds, as $n \to \infty$,*

$$\max\left\{\left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3, \left\|\hat{t}_{\mathrm{w}}(I_n)\right\|_3\right\} = \begin{cases} O\left((\ln(n))^{-\frac{1}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{\lambda-\frac{1}{2}}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}}\right), & -1 < \lambda < 0. \end{cases}$$

Proof. The proof is analogous to the proof of Lemma 5.19. For the sake of completeness and readability, the full proof is given below. Due to the different behaviours depending on the range of λ, we consider the three cases $\lambda = \frac{1}{2}$, $0 < \lambda < \frac{1}{2}$ and $\lambda < 0$ separately.

Case $\lambda = \frac{1}{2}$:
Making use of Lemma 6.1.i) and Lemma 6.2.i) for asymptotic expansions of expectations and standard deviations, respectively, we obtain

$$\begin{aligned} & \hat{t}_{\mathrm{b}}(I_n) \\ = \; & \frac{\tilde{\mu}_{\mathrm{b}}(I_n) + C_\alpha \tilde{\mu}_{\mathrm{b}}(J_n) + (1 - C_\alpha)\tilde{\mu}_{\mathrm{w}}(J_n) - \tilde{\mu}_{\mathrm{b}}(n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \\ = \; & \frac{2(1-\beta)I_n + C_\alpha 2(1-\beta)J_n + (1-C_\alpha)2(1-\beta)J_n - 2(1-\beta)n}{\tilde{\sigma}_{\mathrm{b}}(n)} && (6.18) \\ & + \frac{2}{\Gamma\left(\frac{3}{2}\right)} \frac{(1-\alpha)I_n^{\frac{1}{2}} + C_\alpha(1-\alpha)J_n^{\frac{1}{2}} - (1-C_\alpha)(1-\beta)J_n^{\frac{1}{2}} - (1-\alpha)n^{\frac{1}{2}}}{\tilde{\sigma}_{\mathrm{b}}(n)} && (6.19) \\ & + \frac{1}{\tilde{\sigma}_{\mathrm{b}}(n)} O(1), \end{aligned}$$

where the linear terms in the nominator of (6.18) cancel out and a term of constant order (remembering $I_n + J_n = n - 1$) is left. We obtain, with $\frac{I_n}{n}, \frac{J_n}{n} < 1$ in (6.19), and the fact that the standard deviation is of order $\sqrt{n \ln(n)}$,

$$\begin{aligned} & \left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3 \\ \leq \; & \frac{2n^{\frac{1}{2}}}{\Gamma\left(\frac{3}{2}\right)\tilde{\sigma}_{\mathrm{b}}(n)} \left(\left\|\left(\frac{I_n}{n}\right)^{\frac{1}{2}}\right\|_3 + 2\left\|\left(\frac{J_n}{n}\right)^{\frac{1}{2}}\right\|_3 + 1\right) + O\left((n\ln(n))^{-\frac{1}{2}}\right) \\ = \; & O\left((\ln(n))^{-\frac{1}{2}}\right). \end{aligned}$$

Case $0 < \lambda < \frac{1}{2}$:
We use the asymptotic expansions from Lemma 6.1.i) for the expectations together with

abbreviations of Notation 6.3 and from Lemma 6.2.ii) for the standard deviation. Then,

$$\begin{aligned}\hat{t}_{\mathrm{b}}(I_n) &= \frac{\tilde{\mu}_{\mathrm{b}}(I_n) + C_\alpha \tilde{\mu}_{\mathrm{b}}(J_n) + (1-C_\alpha)\tilde{\mu}_{\mathrm{w}}(J_n) - \tilde{\mu}_{\mathrm{b}}(n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \\ &= \frac{\tilde{c}_{\mathrm{b}} I_n + C_\alpha \tilde{c}_{\mathrm{b}} J_n + (1-C_\alpha)\tilde{c}_{\mathrm{b}} J_n - \tilde{c}_{\mathrm{b}} n}{\tilde{\sigma}_{\mathrm{b}}(n)} \\ &\quad + \frac{\tilde{d}_{\mathrm{b}} I_n^\lambda + C_\alpha \tilde{d}_{\mathrm{b}} J_n^\lambda - (1-C_\alpha)\tilde{d}_{\mathrm{w}} J_n^\lambda - \tilde{d}_{\mathrm{b}} n^\lambda}{\tilde{\sigma}_{\mathrm{b}}(n)} \\ &\quad + \frac{1}{\tilde{\sigma}_{\mathrm{b}}(n)} O(1).\end{aligned}$$

The standard deviation is of order $\sqrt{n}$. Hence, we estimate according to the previous case:

$$\begin{aligned}&\left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3 \\ \le\ & \frac{n^\lambda}{(1-\lambda)\Gamma(1+\lambda)\tilde{\sigma}_{\mathrm{b}}(n)}\left(\left\|\left(\frac{I_n}{n}\right)^\lambda\right\|_3 + 2\left\|\left(\frac{J_n}{n}\right)^\lambda\right\|_3 + 1\right) + O\left(n^{-\frac{1}{2}}\right) \\ =\ & O\left(n^{\lambda-\frac{1}{2}}\right).\end{aligned}$$

Case $-1 < \lambda < 0$:
Finally, we use Lemma 6.1.ii) and Lemma 6.2.iii) and obtain, similarly as above,

$$\begin{aligned}\hat{t}_{\mathrm{b}}(I_n) &= \frac{\tilde{\mu}_{\mathrm{b}}(I_n) + C_\alpha \tilde{\mu}_{\mathrm{b}}(J_n) + (1-C_\alpha)\tilde{\mu}_{\mathrm{w}}(J_n) - \tilde{\mu}_{\mathrm{b}}(n)}{\tilde{\sigma}_{\mathrm{b}}(n)} \\ &= \frac{\tilde{c}_{\mathrm{b}} I_n + C_\alpha \tilde{c}_{\mathrm{b}} J_n + (1-C_\alpha)\tilde{c}_{\mathrm{b}} J_n - \tilde{c}_{\mathrm{b}} n + O(1)}{\tilde{\sigma}_{\mathrm{b}}(n)} \\ &= \frac{1}{\tilde{\sigma}_{\mathrm{b}}(n)} O(1) = O\left(n^{-\frac{1}{2}}\right),\end{aligned}$$

yielding

$$\left\|\hat{t}_{\mathrm{b}}(I_n)\right\|_3 = O\left(n^{-\frac{1}{2}}\right).$$

Likewise treatment of $\left\|\hat{t}_{\mathrm{w}}(I_n)\right\|_3$ yields the assertion. □

Lemma 6.15. *In the situation of Theorem 6.11 with I_n and U as in* (6.2)*, it holds, as $n \to \infty$,*

$$\max\left\{\left\|\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{U}\right\|_3, \left\|\frac{\tilde{\sigma}_{\mathrm{w}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{U}\right\|_3\right\} = \begin{cases} O\left((\ln(n))^{-\frac{1}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{\lambda-\frac{1}{2}}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}}\right), & -1 < \lambda < 0. \end{cases}$$

Proof. In order to handle $\left\| \frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{U} \right\|_3$, we squeeze in the square root of the rescaled subtree sizes:

$$\left\| \frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{U} \right\|_3 \leq \left\| \sqrt{\frac{I_n}{n}} - \sqrt{U} \right\|_3 + \left\| \frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{\frac{I_n}{n}} \right\|_3 . \tag{6.20}$$

Obviously, the L_3-distance between the square root of the rescaled subtree sizes and the square root of its limit is central.

For the first part, we have according to Lemma 3.8 with $\psi = \frac{1}{2}$ and $p = 3$ (hence, $\psi < \frac{p-1}{p}$)

$$\left\| \sqrt{\frac{I_n}{n}} - \sqrt{U} \right\|_3 = O\left(n^{-\frac{5}{6}}\right). \tag{6.21}$$

For the second part of the right-hand side of (6.20), we have three cases depending on the ranges of λ caused by the different asymptotic expansions in the variance, see Lemma 6.2.

For $\lambda = \frac{1}{2}$, we use the asymptotic expansion of the variance from Lemma 6.2.i). Note that $\tilde{\sigma}_{\mathrm{b}}(0) = \tilde{\sigma}_{\mathrm{w}}(0) = 0$ as well as $\tilde{\sigma}_{\mathrm{b}}(1) = \sqrt{\alpha(1-\alpha)}$ and $\tilde{\sigma}_{\mathrm{w}}(1) = \sqrt{\beta(1-\beta)}$. Hence, "small" $j = 0, 1$ are not covered by the asymptotic expansions given in Lemma 6.2.i). Therefore, we insert the asymptotic expansions together with a suitable indicator and "hide" the "small" j in $O\left(n^{-\frac{1}{2}}\right)$. We obtain

$$\begin{aligned}
\left\| \frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{\frac{I_n}{n}} \right\|_3 &\leq \left\| \mathbb{1}_{\{I_n \geq 1\}} \sqrt{\frac{I_n \ln(I_n) + O(n)}{n \ln(n) + O(n)}} - \sqrt{\frac{I_n}{n}} \right\|_3 + O\left(n^{-\frac{1}{2}}\right) \\
&\leq \left\| \mathbb{1}_{\{I_n \geq 1\}} \sqrt{\frac{\left| n I_n \ln\left(\frac{I_n}{n}\right) + O\left(n^2\right) \right|}{n\left(n \ln(n) + O(n)\right)}} \right\|_3 + O\left(n^{-\frac{1}{2}}\right) \\
&\leq \frac{1}{\sqrt{n^2 \ln(n)}} \left\| \mathbb{1}_{\{I_n \geq 1\}} \sqrt{n^2 \left| \frac{I_n}{n} \ln\left(\frac{I_n}{n}\right) \right|} \right\|_3 + O\left((\ln(n))^{-\frac{1}{2}}\right) \\
&\leq \frac{1}{e\sqrt{\ln(n)}} + O\left((\ln(n))^{-\frac{1}{2}}\right) = O\left((\ln(n))^{-\frac{1}{2}}\right) \qquad (6.22)
\end{aligned}$$

using that the function $x \to |x \ln(x)|$, continuously extended by $0 \ln(0) := 0$, is bounded by $\frac{1}{e}$ for $x \in [0, 1]$ in (6.22).

In the case $0 < \lambda < \frac{1}{2}$, the expansion of the variance in Lemma 6.2.ii) yields

$$\left\| \frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)} - \sqrt{\frac{I_n}{n}} \right\|_3 = \left\| \sqrt{\frac{I_n + O\left(n^{2\lambda}\right)}{n + O\left(n^{2\lambda}\right)}} - \sqrt{\frac{I_n}{n}} \right\|_3$$

$$
\begin{aligned}
&= \left\| \frac{\sqrt{n\left(I_n + O\left(n^{2\lambda}\right)\right)} - \sqrt{I_n\left(n + O\left(n^{2\lambda}\right)\right)}}{\sqrt{n\left(n + O\left(n^{2\lambda}\right)\right)}} \right\|_3 \\
&\leq \left\| \frac{\sqrt{n\left(I_n + O\left(n^{2\lambda}\right)\right) - I_n\left(n + O\left(n^{2\lambda}\right)\right)}}{\sqrt{n\left(n + O\left(n^{2\lambda}\right)\right)}} \right\|_3 \\
&\leq \sqrt{\frac{O\left(n^{2\lambda+1}\right)}{n\left(n + O\left(n^{2\lambda}\right)\right)}} = O\left(n^{\lambda - \frac{1}{2}}\right).
\end{aligned}
$$

Finally, when $\lambda < 0$, we have from Lemma 6.2.iii)

$$
\left\| \frac{\tilde{\sigma}_{\mathrm{b}}\left(I_n\right)}{\tilde{\sigma}_{\mathrm{b}}\left(n\right)} - \sqrt{\frac{I_n}{n}} \right\|_3 = \left\| \sqrt{\frac{I_n + O\left(1\right)}{n + O\left(1\right)}} - \sqrt{\frac{I_n}{n}} \right\|_3 = O\left(n^{-\frac{1}{2}}\right).
$$

Combining these three estimates together with (6.21) in (6.20), the assertion follows. □

Remark 6.16. Using Lemma 2.7 in setting **Det R** leads to estimates corresponding to those of Proposition 6.13, Lemma 6.14 as well as Lemma 6.15. As **Det R**-version of Proposition 6.13 we would obtain the following result for the distance of the accompanying sequences $\mathscr{Q}_n^{\mathrm{b}}$ and $\mathscr{Q}_n^{\mathrm{w}}$ to the normal distribution that looks amiss:

$$
\zeta_3^{\vee}\left(\left(\mathscr{Q}_n^{\mathrm{b}}, \mathscr{Q}_n^{\mathrm{w}}\right), \left(\mathcal{N}\left(0,1\right), \mathcal{N}\left(0,1\right)\right)\right) = \begin{cases} O\left(\left(\ln\left(n\right)\right)^{-1}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{\lambda - \frac{1}{2}}\right), & \frac{1}{4} \leq \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{4}}\right), & \lambda < \frac{1}{4},\ \lambda \neq 0. \end{cases}
$$

The exponent of the rate stops changing with λ at $\lambda = \frac{1}{4}$. The "fastest" rate we would obtain is $n^{-\frac{1}{4}}$ for $\lambda \leq \frac{1}{4}$, $\lambda \neq 0$. In the situation of Proposition 6.13, the exponent of the rate is a linear decreasing function of λ for $\lambda \in \left(0, \frac{1}{2}\right)$ and abides at the order of the reciprocal of the standard deviation of the number of black balls for negative λ.

The vital difference here is the treatment of (6.20) in the proof of Lemma 6.15: We are able to analyse $\left\| \sqrt{\frac{I_n}{n}} - \sqrt{U} \right\|_3$ more precisely due to the explicit connection of the rescaled subtree sizes and its limit via $I_n = \lfloor nU \rfloor$. We have done so in Lemma 3.8. In setting **Det R**, we have to retreat to Corollary 3.5 that is a consequence of Lemma 3.4. But already in Lemma 3.4, we were not able to couple the rescaled subtree sizes and its Dirichlet-distributed limit accordingly. Hence, our estimates are rather blunt.

Nevertheless, we need to consider the approach of Proposition 5.18 for **Step 2*** to accelerate the rate for $0 < \lambda \leq \frac{1}{2}$.

Option 2: Analogue of Proposition 5.18

Using the same methods as in the proof of Proposition 5.18, we obtain the following

Proposition 6.17 (Twin of Proposition 5.18)**.** *In the situation of Theorem 6.11 with* $\hat{\mathcal{Q}}_n^{\mathrm{b}}$ *and* $\hat{\mathcal{Q}}_n^{\mathrm{w}}$ *defined in* (6.12) *and* (6.13)*, it holds, as* $n \to \infty$*,*

$$\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) = O\left(\left\|\frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + (1 - C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - 1\right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\|\hat{t}_{\mathrm{b}}(n)\right\|_3^3\right)$$

as well as

$$\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right) = O\left(\left\|\frac{\tilde{\sigma}_{\mathrm{w}}^2(I_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + (1 - C_\beta)\frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + C_\beta \frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} - 1\right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\|\hat{t}_{\mathrm{w}}(n)\right\|_3^3\right).$$

Sketch of Proof. The reasoning that is to be conducted in order to prove Proposition 6.17 is the same as in the proof of Proposition 5.18. Hence, the calculations thereof are not repeated but only the quantities in use that need to be adjusted are given. Plugging in the quantities defined here in the calculations of the proof of Proposition 5.18 leads to the assertion of Proposition 6.17.

Making use of the convolution property of the normal distribution, we have

$$\begin{aligned}
\hat{\mathcal{Q}}_n^{\mathrm{b}} &\overset{\mathrm{d}}{=} \left(\frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + (1 - C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)}\right)^{\frac{1}{2}} N + \hat{t}_{\mathrm{b}}(I_n) \\
&\overset{\mathrm{d}}{=} \mathbb{1}_{\tilde{A}_{\mathrm{b}}^c}\left(\tilde{G}_n^{\mathrm{b}} N + \hat{t}_{\mathrm{b}}(I_n)\right) + \mathbb{1}_{\tilde{A}_{\mathrm{b}}}\left(N + \tilde{\Delta}_n^{\mathrm{b}} N' + \hat{t}_{\mathrm{b}}(I_n)\right) \\
&=: \bar{Q}_n^{\mathrm{b}}
\end{aligned}$$

with

$$\begin{aligned}
\tilde{G}_n^{\mathrm{b}} &:= \left(\frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + (1 - C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)}\right)^{\frac{1}{2}}, \\
\tilde{A}_{\mathrm{b}} &:= \left\{\tilde{G}_n^{\mathrm{b}} \geq 1\right\}, \\
\tilde{\Delta}_n^{\mathrm{b}} &:= \sqrt{\left|\left(\tilde{G}_n^{\mathrm{b}}\right) - 1\right|}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\mathcal{Q}}_n^{\mathrm{w}} &\overset{\mathrm{d}}{=} \left(\frac{\tilde{\sigma}_{\mathrm{w}}^2(I_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + (1 - C_\beta)\frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + C_\beta \frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)}\right)^{\frac{1}{2}} N + \hat{t}_{\mathrm{w}}(I_n) \\
&\overset{\mathrm{d}}{=} \mathbb{1}_{\tilde{A}_{\mathrm{b}}^c}\left(\tilde{G}_n^{\mathrm{b}} N + \hat{t}_{\mathrm{b}}(I_n)\right) + \mathbb{1}_{\tilde{A}_{\mathrm{b}}}\left(N + \tilde{\Delta}_n^{\mathrm{b}} N' + \hat{t}_{\mathrm{b}}(I_n)\right)
\end{aligned}$$

$$=: \bar{Q}_n^{\mathrm{w}}$$

with

$$\begin{aligned}
\tilde{G}_n^{\mathrm{w}} &:= \left(\frac{\tilde{\sigma}_{\mathrm{w}}^2 (I_n)}{\tilde{\sigma}_{\mathrm{w}}^2 (n)} + (1 - C_\beta) \frac{\tilde{\sigma}_{\mathrm{b}}^2 (J_n)}{\tilde{\sigma}_{\mathrm{w}}^2 (n)} + C_\beta \frac{\tilde{\sigma}_{\mathrm{w}}^2 (J_n)}{\tilde{\sigma}_{\mathrm{w}}^2 (n)} \right)^{\frac{1}{2}}, \\
\tilde{A}_{\mathrm{w}} &:= \left\{ \tilde{G}_n^{\mathrm{w}} \geq 1 \right\}, \\
\tilde{\Delta}_n^{\mathrm{w}} &:= \sqrt{\left| \left(\tilde{G}_n^{\mathrm{w}} \right) - 1 \right|}.
\end{aligned}$$

The modified accompanying sequences $\bar{Q}_n^{\mathrm{b}}$ and $\bar{Q}_n^{\mathrm{w}}$ are then compared, in terms of the Zolotarev distance ζ_3, with

$$\begin{aligned}
\bar{N}^{\mathrm{b}} &:= \mathbb{1}_{A_{\mathrm{b}}} N + \mathbb{1}_{A_{\mathrm{b}}^c} \left(\tilde{G}_n^{\mathrm{b}} + \tilde{\Delta}_n^{\mathrm{b}} N' \right) \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1) \text{ and} \\
\bar{N}^{\mathrm{w}} &:= \mathbb{1}_{A_{\mathrm{w}}} N + \mathbb{1}_{A_{\mathrm{w}}^c} \left(\tilde{G}_n^{\mathrm{w}} + \tilde{\Delta}_n^{\mathrm{w}} N' \right) \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1),
\end{aligned}$$

respectively.

The difference $f\left(\bar{Q}_n^{\mathrm{b}}\right) - f\left(\bar{N}^{\mathrm{b}}\right)$, that arises in the Zolotarev distance $\zeta_3\left(\bar{Q}_n^{\mathrm{b}}, \bar{N}^{\mathrm{b}}\right)$ with $f \in \mathcal{F}_3$, is treated with Taylor expansion at N in completely the same way as in pp. 79–85 of the proof of Proposition 5.18. This yields

$$\begin{aligned}
\zeta_3\left(\bar{Q}_n^{\mathrm{b}}, \bar{N}^{\mathrm{b}}\right) &\leq O\left(\left\| \left(\tilde{G}_n^{\mathrm{b}} \right)^2 - 1 \right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\| \hat{t}_{\mathrm{b}}(n) \right\|_3^3 \right) \quad \text{as well as} \\
\zeta_3\left(\bar{Q}_n^{\mathrm{w}}, \bar{N}^{\mathrm{w}}\right) &\leq O\left(\left\| \left(\tilde{G}_n^{\mathrm{w}} \right)^2 - 1 \right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\| \hat{t}_{\mathrm{w}}(n) \right\|_3^3 \right),
\end{aligned}$$

leading to the claim of Proposition 6.17. ⊠

Lemma 6.18 (Twin of Lemma 5.20)**.** *In the situation of Theorem 6.11, we have for the quantities appearing in the estimates for* $\zeta_3^{\vee}\left(\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \hat{\mathcal{Q}}_n^{\mathrm{w}}\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right)$ *of Proposition 6.17, as* $n \to \infty$,

$$\left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + (1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - 1 \right\|_{\frac{3}{2}} = \begin{cases} O\left((\ln(n))^{-1}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{2\left(\lambda - \frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-1}\right), & \lambda < 0, \end{cases}$$

as well as

$$\left\| \frac{\tilde{\sigma}_{\mathrm{w}}^2(I_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + (1-C_\beta) \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + C_\beta\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} - 1 \right\|_{\frac{3}{2}} = \begin{cases} O\left((\ln(n))^{-1}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{2\left(\lambda - \frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-1}\right), & \lambda < 0. \end{cases}$$

Proof. The situation is analogous to the situation of Lemma 5.20: We add the undistorted rescaled subtree sizes $\frac{I_n}{n}$ and $\frac{J_n}{n}$ and subtract $\frac{n-1}{n}$. Then,

$$\begin{aligned} &\left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + (1-C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - 1 \right\|_{\frac{3}{2}} \\ =& \left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{I_n}{n} + C_\alpha \left(\frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{J_n}{n}\right) + (1-C_\alpha)\left(\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{J_n}{n}\right) + \frac{n-1}{n} - 1 \right\|_{\frac{3}{2}} \\ \leq& \left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{I_n}{n} \right\|_{\frac{3}{2}} + \left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{J_n}{n} \right\|_{\frac{3}{2}} + \left\| \frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{J_n}{n} \right\|_{\frac{3}{2}} + \frac{1}{n}. \end{aligned}$$

Obviously, for $\left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{I_n}{n} \right\|_{\frac{3}{2}}$ we have to go into the three cases $\lambda = \frac{1}{2}$, $0 < \lambda < \frac{1}{2}$ and $\lambda < 0$ in consequence of the different behaviour of the variances, see Lemma 6.2: The treatment of this term differs from the situation of Lemma 5.20 only in the constants that arise in the asymptotic expansions of the variances. Hence, with the reasoning of Lemma 5.20 in the three subcases, we have:

$$\begin{aligned} \lambda = \frac{1}{2}: &\quad \left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{I_n}{n} \right\|_{\frac{3}{2}} = O\left((\ln(n))^{-1}\right), \\ 0 < \lambda < \frac{1}{2}: &\quad \left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{I_n}{n} \right\|_{\frac{3}{2}} = O\left(n^{2\left(\lambda - \frac{1}{2}\right)}\right), \\ \lambda < 0: &\quad \left\| \frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{I_n}{n} \right\|_{\frac{3}{2}} = O\left(n^{-1}\right). \end{aligned}$$

The same reasoning applies to $\left\|\frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{J_n}{n}\right\|_{\frac{3}{2}}$ and $\left\|\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - \frac{J_n}{n}\right\|_{\frac{3}{2}}$. Hence, we obtain

$$\left\|\frac{\tilde{\sigma}_{\mathrm{b}}^2(I_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + C_\alpha \frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} + (1 - C_\alpha)\frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{b}}^2(n)} - 1\right\|_{\frac{3}{2}} = \begin{cases} O\left((\ln(n))^{-1}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{2\left(\lambda - \frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-1}\right), & \lambda < 0. \end{cases}$$

Treating $\left\|\frac{\tilde{\sigma}_{\mathrm{w}}^2(I_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + (1 - C_\beta)\frac{\tilde{\sigma}_{\mathrm{b}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} + C_\beta \frac{\tilde{\sigma}_{\mathrm{w}}^2(J_n)}{\tilde{\sigma}_{\mathrm{w}}^2(n)} - 1\right\|_{\frac{3}{2}}$ accordingly, the assertion follows. □

Corollary 6.19 (Twin of Corollary 5.21). *In the situation of Theorem 6.11 and Proposition 6.17, it holds, as $n \to \infty$,*

$$\zeta_3^{\vee}\left(\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \hat{\mathcal{Q}}_n^{\mathrm{w}}\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda - \frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{3}{2}}\right), & \lambda < 0. \end{cases}$$

Proof. It follows immediately from Proposition 6.17 together with Lemma 6.14 and Lemma 6.18. □

Remark 6.20. In Corollary 6.19, the exponents of all the rates are accelerated by a factor 3 in contrast to the rates obtained in Proposition 6.13.

Ad Step 3*: Link of Interim Stages — Induction

An upper bound for the rate of convergence on the basis of Proposition 6.12 together with Proposition 6.13 and Corollary 6.19, respectively, is derived via induction — analogously to **Step 3** in setting **Det R**.

The following Corollaries 6.21 and 6.22 are analogues of Corollaries 5.22, 5.23 and 5.24. Hence, some of the estimates are not carried out in full detail. Still, crucial or short parts of the calculations are written in such a manner that there is no need to consult the calculations of Corollaries 5.22, 5.23 and 5.24.

In Corollaries 6.21 and 6.22 the constants $A, B, C, D > 0$, and $\xi, \delta, \delta' \in (0,1)$ as well as an integer called $n_0 \in \mathbb{N}$ occur several times with potentially different meanings but during one and the same induction their meaning does not change. There are three inductions; one for each of the cases: $\lambda = \frac{1}{2}$, $\frac{1}{3} < \lambda < \frac{1}{2}$, and $\lambda \leq \frac{1}{3}$.

Corollary 6.21 (Twin of Corollary 5.22)**.** *In the situation of Theorem 6.11, if $\lambda = \frac{1}{2}$, it holds, as $n \to \infty$,*

$$\zeta_3^{\vee}\left(\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Y}}_n\right), (\mathcal{N}(0,1), \mathcal{N}(0,1))\right) = O\left((\ln(n))^{-\frac{3}{2}}\right).$$

Proof. From Lemma 6.2.i) and Lemma A.3 applied as in (A.2), we have with a suitable constant $A > 0$

$$\left(\frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3, \left(\frac{\tilde{\sigma}_{\mathrm{w}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3 \leq \frac{(j\ln(j))^{\frac{3}{2}} + Aj^{\frac{3}{2}}(\ln(j))^{\frac{1}{2}}}{(n\ln(n) + O(n))^{\frac{3}{2}}}.$$

Proposition 6.13 suggests a rate of order $(\ln(n))^{-\frac{1}{2}}$, whereas Corollary 6.19 suggests a rate of order $(\ln(n))^{-\frac{3}{2}}$. Of course, we try to transfer the faster of the two rates to the distances $\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right)$ and set as induction hypothesis

$$\hat{\mathscr{D}}(j) \leq C(\ln(j))^{-\frac{3}{2}} \quad \text{for } j = 2, \ldots, n-1.$$

On plugging in the induction hypothesis, we have to treat the contributions on the events $\{I_n < 2\}$ and $\{J_n < 2\}$ separately. Therefore, we add indicators and hide the contributions not covered by these indicators in $O\left(\frac{1}{n}\right)$. Evoking Proposition 6.12 and Corollary 6.19 and plugging in the induction hypothesis in (6.23), we have with a suitable constant $B > 0$

$$\begin{aligned}
&\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right) \leq \zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right) + \zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)\\
&\leq \mathbb{E}\left[\left(\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3 \hat{\mathscr{D}}(I_n) + \left(\alpha\left(\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3 + (1-\alpha)\left(\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3\right)\hat{\mathscr{D}}(J_n)\right]\\
&\quad + \zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)\\
&\leq \mathbb{E}\left[\left(\frac{I_n\ln(I_n)}{n\ln(n)+O(n)}\right)^{\frac{3}{2}}\hat{\mathscr{D}}(I_n)\mathbb{1}_{\{I_n\geq 2\}} + \left(\frac{J_n\ln(J_n)}{n\ln(n)+O(n)}\right)^{\frac{3}{2}}\hat{\mathscr{D}}(J_n)\mathbb{1}_{\{J_n\geq 2\}}\right]\\
&\quad + \mathbb{E}\left[\frac{AI_n^{\frac{3}{2}}(\ln(I_n))^{\frac{1}{2}}}{(n\ln(n)+O(n))^{\frac{3}{2}}}\hat{\mathscr{D}}(I_n)\mathbb{1}_{\{I_n\geq 2\}} + \frac{AJ_n^{\frac{3}{2}}(\ln(J_n))^{\frac{1}{2}}}{(n\ln(n)+O(n))^{\frac{3}{2}}}\hat{\mathscr{D}}(J_n)\mathbb{1}_{\{J_n\geq 2\}}\right]\\
&\quad + \zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) + O\left(\frac{1}{n}\right)\\
\text{(6.23)}\quad &\leq \frac{1+o(1)}{(n\ln(n))^{\frac{3}{2}}}\left(C\mathbb{E}\left[I_n^{\frac{3}{2}} + J_n^{\frac{3}{2}}\right] + AC\mathbb{E}\left[\frac{I_n^{\frac{3}{2}}}{\ln(I_n)}\mathbb{1}_{\{I_n\geq 2\}} + \frac{J_n^{\frac{3}{2}}}{\ln(J_n)}\mathbb{1}_{\{J_n\geq 2\}}\right]\right)\\
&\quad + B(\ln(n))^{-\frac{3}{2}}\\
&\leq (1+o(1))\,C(\ln(n))^{-\frac{3}{2}}\,\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{\frac{3}{2}} + \left(\frac{J_n}{n}\right)^{\frac{3}{2}}\right]
\end{aligned}$$

$$
\begin{aligned}
&+C\frac{1+o(1)}{(n\ln(n))^{\frac{3}{2}}}A\cdot 2\sum_{j=2}^{n-1}\frac{1}{n}\frac{j^{\frac{3}{2}}}{\ln(j)}+B(\ln(n))^{-\frac{3}{2}}\\
\text{(6.24)}\quad&\leq(1-\delta)C(\ln(n))^{-\frac{3}{2}}+Cn^{-\frac{5}{2}}(\ln(n))^{-\frac{3}{2}}D\frac{n^{\frac{5}{2}}}{\ln(n)}+B(\ln(n))^{-\frac{3}{2}}\\
&=(1-\delta)C(\ln(n))^{-\frac{3}{2}}+C(\ln(n))^{-\frac{3}{2}}D(\ln(n))^{-1}+B(\ln(n))^{-\frac{3}{2}}\\
\text{(6.25)}\quad&\leq(1-(\delta-\delta'))C(\ln(n))^{-\frac{3}{2}}+B(\ln(n))^{-\frac{3}{2}}.
\end{aligned}
$$

In (6.24) we used that by Corollary 3.5 and property (3.6) it holds

$$\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{\frac{3}{2}}+\left(\frac{J_n}{n}\right)^{\frac{3}{2}}\right]\to\xi<1,\quad n\to\infty.$$

So, there is $0<\delta<1$ such that for n sufficiently large

$$\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{\frac{3}{2}}+\left(\frac{J_n}{n}\right)^{\frac{3}{2}}\right]<1-\delta.$$

Furthermore, we used in (6.24) the following estimate for the sum $\sum_{j=2}^{n-1}\frac{j^{\frac{3}{2}}}{\ln(j)}$:

$$
\begin{aligned}
\sum_{j=2}^{n-1}\frac{j^{\frac{3}{2}}}{\ln(j)}&\leq\sum_{j=2}^{\left\lfloor\frac{n}{\ln(n)}\right\rfloor}\frac{j^{\frac{3}{2}}}{\ln(j)}+\sum_{j=\left\lceil\frac{n}{\ln(n)}\right\rceil}^{n-1}\frac{j^{\frac{3}{2}}}{\ln(j)}\\
&\leq\frac{E}{\ln(2)}\left\lfloor\frac{n}{\ln(n)}\right\rfloor^{\frac{5}{2}}+\frac{F}{\ln\left(\left\lceil\frac{n}{\ln(n)}\right\rceil\right)}n^{\frac{5}{2}}\\
&\leq E'\left(\frac{n}{\ln(n)}\right)^{\frac{5}{2}}+F(1+o(1))\frac{n^{\frac{5}{2}}}{\ln(n)}=O\left(\frac{n^{\frac{5}{2}}}{\ln(n)}\right)
\end{aligned}
$$

with suitable constants $E,E',F>0$. Hence, we can choose a constant D such that

$$D\frac{n^{\frac{5}{2}}}{\ln(n)}\geq 2A(1+o(1))\sum_{j=2}^{n-1}\frac{j^{\frac{3}{2}}}{\ln(j)}$$

for n large enough.

In (6.25), we chose $0<\delta'<\delta$ such that for n large enough

$$D(\ln(n))^{-1}<\delta'.$$

Finally, let $n_0\in\mathbb{N}$ be fixed such that above esimates hold for $n\geq n_0$ and choose

$$C\leq\hat{\mathscr{D}}(1)\vee\max\left\{\hat{\mathscr{D}}(j)(\ln(j))^{\frac{3}{2}}\,|j=2,\dots,n_0-1\right\}\vee\frac{B}{\delta-\delta'}.$$

Then, we have

$$\zeta_3 \left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1) \right) \leq C (\ln(n))^{-\frac{3}{2}} \quad \text{for } n \geq 2.$$

The same reasoning applied to $\zeta_3 \left(\hat{\mathcal{Y}}_n, \mathcal{N}(0,1) \right)$ yields the assertion. □

For $0 < \lambda < \frac{1}{2}$, Corollary 6.19 suggests a rate of order $n^{3\left(\lambda - \frac{1}{2}\right)}$. Supposing this rate, the contractive behaviour (see Remark 8.1) needed to conduct the induction, is satisfied for $\frac{1}{3} < \lambda < \frac{1}{2}$. For $\lambda \leq \frac{1}{3}$ we expect the rate to be of the order $n^{-\frac{1}{2}}$, the reciprocal of the standard deviation, but can only prove a rate of order $n^{-\frac{1}{2}+\varepsilon}$ for arbitrary $\varepsilon > 0$. Note that for negative λ the less elaborate estimates leading to Proposition 6.13 are sufficient, in contrary to the estimates we are able to find in the setting **Det R** (sketched in Remark 6.16).

In Corollaries 5.23 and 5.24, estimates for the cases $0 < \lambda \leq \frac{1}{2}$ and $\lambda < 0$ only differ in the second order term of the variance. The necessary steps are nearly identical. Thus, in the following corollary we do not distinguish between these two regimes.

Corollary 6.22 (Twin of Corollaries 5.23 and 5.24)**.** *In the situation of Theorem 6.11, if $\lambda < \frac{1}{2}$, $\lambda \neq 0$, let $\varepsilon > 0$. Then, as $n \to \infty$,*

$$\zeta_3^{\vee} \left(\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Y}}_n \right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1) \right) \right) = \begin{cases} O\left(n^{3\left(\lambda - \frac{1}{2}\right)} \right), & \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon} \right), & \lambda \leq \frac{1}{3},\ \lambda \neq 0. \end{cases}$$

Proof. The ratios $\left(\frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3$ and $\left(\frac{\tilde{\sigma}_{\mathrm{w}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3$ occurring in Proposition 6.12 can be estimated for both $0 < \lambda < \frac{1}{2}$ and $\lambda < 0$ with the help of Lemma 6.2.ii), iii) and Lemma A.3 applied as in (A.3) by, with a suitable constant $A > 0$,

$$\left(\frac{\tilde{\sigma}_{\mathrm{b}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3, \left(\frac{\tilde{\sigma}_{\mathrm{w}}(j)}{\tilde{\sigma}_{\mathrm{b}}(n)} \right)^3 \leq \frac{j^{\frac{3}{2}} + A j^{2\lambda + \frac{1}{2}}}{n + O\left(n^{2\lambda}\right)}. \tag{6.26}$$

This estimate seems to be a little rough for negative λ. However, it does not influence the order of an upper bound for the rate of convergence. Of course, it should be refined when determining constants for the rates.

We begin with the case $\frac{1}{3} < \lambda < \frac{1}{2}$:
We set as induction hypothesis

$$\hat{\mathscr{D}}(j) \leq C j^{3\left(\lambda - \frac{1}{2}\right)}, \quad j = 1, \ldots, n-1.$$

Again, note that $\hat{\mathscr{D}}(0)$ does not contribute. With Proposition 6.12, Corollary 6.19, the estimates above and the induction hypothesis, we have with a suitable constant $B > 0$

$$\begin{aligned}
&\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right) \le \zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right) + \zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right) \le \zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right) + Bn^{3\left(\lambda-\frac{1}{2}\right)} \\
\le&\mathbb{E}\left[\left(\frac{\tilde{\sigma}_{\mathrm{b}}(I_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3 \hat{\mathscr{D}}(I_n) + \left(\alpha\left(\frac{\tilde{\sigma}_{\mathrm{b}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3 + (1-\alpha)\left(\frac{\tilde{\sigma}_{\mathrm{w}}(J_n)}{\tilde{\sigma}_{\mathrm{b}}(n)}\right)^3\right)\hat{\mathscr{D}}(J_n)\right] \\
&+ Bn^{3\left(\lambda-\frac{1}{2}\right)} \\
\le&\mathbb{E}\left[\left(\frac{I_n}{n+O(n^{2\lambda})}\right)^{\frac{3}{2}} \hat{\mathscr{D}}(I_n) + \left(\frac{J_n}{n+O(n^{2\lambda})}\right)^{\frac{3}{2}} \hat{\mathscr{D}}(J_n)\right] \\
&+ A\mathbb{E}\left[\frac{I_n^{2\lambda+\frac{1}{2}}}{(n+O(n^{2\lambda}))^{\frac{3}{2}}}\hat{\mathscr{D}}(I_n) + \frac{J_n^{2\lambda+\frac{1}{2}}}{(n+O(n^{2\lambda}))^{\frac{3}{2}}}\hat{\mathscr{D}}(J_n)\right] + Bn^{3\left(\lambda-\frac{1}{2}\right)} \\
\le&\frac{(1+o(1))\,C}{n^{\frac{3}{2}}}\mathbb{E}\left[I_n^{3\lambda} + J_n^{3\lambda}\right] + \frac{(1+o(1))\,AC}{n^{\frac{3}{2}}}\mathbb{E}\left[I_n^{5\lambda-1} + J_n^{5\lambda-1}\right] + Bn^{3\left(\lambda-\frac{1}{2}\right)}
\end{aligned}$$

$$= (1+o(1))\,Cn^{3\left(\lambda-\frac{1}{2}\right)}\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{3\lambda} + \left(\frac{J_n}{n}\right)^{3\lambda}\right] \tag{6.27}$$

$$+ (1+o(1))\,ACn^{5\left(\lambda-\frac{1}{2}\right)}\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{5\lambda-1} + \left(\frac{J_n}{n}\right)^{5\lambda-1}\right] + Bn^{3\left(\lambda-\frac{1}{2}\right)}$$

$$\le (1-\delta)\,Cn^{3\left(\lambda-\frac{1}{2}\right)} + Cn^{3\left(\lambda-\frac{1}{2}\right)}Dn^{2\left(\lambda-\frac{1}{2}\right)} + Bn^{3\left(\lambda-\frac{1}{2}\right)} \tag{6.28}$$

$$\le (1-(\delta-\delta'))\,Cn^{3\left(\lambda-\frac{1}{2}\right)} + Bn^{3\left(\lambda-\frac{1}{2}\right)}. \tag{6.29}$$

In (6.28): From Corollary 3.5 together with property (3.6) and $3\lambda > 1$ we know that

$$\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{3\lambda} + \left(\frac{J_n}{n}\right)^{3\lambda}\right] \to \xi < 1, \quad n \to \infty.$$

Hence, there is $0 < \delta < 1$ such that

$$(1+o(1))\,\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{3\lambda} + \left(\frac{J_n}{n}\right)^{3\lambda}\right] \le 1-\delta$$

for n sufficiently large. Additionally, we chose a constant D such that

$$D \ge (1+o(1))\,A\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{5\lambda-1} + \left(\frac{J_n}{n}\right)^{5\lambda-1}\right]$$

for n sufficiently large (observe that $\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{5\lambda-1} + \left(\frac{J_n}{n}\right)^{5\lambda-1}\right] \le 2$ due to $5\lambda - 1 > \frac{2}{3}$ and $\frac{I_n}{n}, \frac{J_n}{n} \le 1$). In (6.29), we choose $0 < \delta' < \delta$ such that for n large enough

$$Dn^{2\left(\lambda-\frac{1}{2}\right)} < \delta'.$$

Now, we fix $n_0 \in \mathbb{N}$ such that all of the above estimates hold for $n \geq n_0$ and choose

$$C \geq \max\left\{\hat{\mathscr{D}}(j)\, j^{3\left(\frac{1}{2}-\lambda\right)} | j = 1, \ldots, n_0 - 1\right\} \vee \frac{B}{\delta - \delta'}.$$

Then, we have

$$\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right) \leq C\, n^{3\left(\lambda - \frac{1}{2}\right)} \quad \text{for } n \geq 1.$$

The same holds for $\zeta_3\left(\hat{\mathcal{Y}}_n, \mathcal{N}(0,1)\right)$ (with readjusted C).

We now devote our attention to the case $\lambda \leq \frac{1}{3}$, $\lambda \neq 0$:
Let $\varepsilon > 0$ and set as induction hypothesis

$$\hat{\mathscr{D}}(j) \leq C j^{-\frac{1}{2}+\varepsilon}, \quad j = 1, \ldots, n-1.$$

As before, $\hat{\mathscr{D}}(0)$ does not contribute in the following computation. As the rates suggested by Corollary 6.19 are too fast to be transferred we can now treat the distance $\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ crudely and choose to estimate it by $n^{-\frac{1}{2}}$. In the following calculation, w.l.o.g., we assume $\lambda > 0$. However, the calculations serve, with suitably chosen constants, as upper bound for the case with negative λ. With Proposition 6.12, as before, and a suitable constant $B > 0$ we have

$$\begin{aligned}
&\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right)\\
&\leq \mathbb{E}\left[\left(\frac{I_n}{n + O(n^{2\lambda})}\right)^{\frac{3}{2}} \hat{\mathscr{D}}(I_n) + \left(\frac{J_n}{n + O(n^{2\lambda})}\right)^{\frac{3}{2}} \hat{\mathscr{D}}(J_n)\right]\\
&\quad + A\mathbb{E}\left[\frac{I_n^{2\lambda+\frac{1}{2}}}{(n + O(n^{2\lambda}))^{\frac{3}{2}}} \hat{\mathscr{D}}(I_n) + \frac{J_n^{2\lambda+\frac{1}{2}}}{(n + O(n^{2\lambda}))^{\frac{3}{2}}} \hat{\mathscr{D}}(J_n)\right] + Bn^{-\frac{1}{2}}\\
&\leq \frac{(1 + o(1))\, C}{n^{\frac{3}{2}}} \mathbb{E}\left[I_n^{1+\varepsilon} + J_n^{1+\varepsilon}\right] + \frac{(1 + o(1))\, AC}{n^{\frac{3}{2}}} \mathbb{E}\left[I_n^{2\lambda+\varepsilon} + J_n^{2\lambda+\varepsilon}\right] + Bn^{-\frac{1}{2}}\\
&= (1 + o(1))\, Cn^{-\frac{1}{2}+\varepsilon} \mathbb{E}\left[\left(\frac{I_n}{n}\right)^{1+\varepsilon} + \left(\frac{J_n}{n}\right)^{1+\varepsilon}\right]\\
&\quad + (1 + o(1))\, ACn^{-\frac{3}{2}+2\lambda+\varepsilon} \mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda+\varepsilon} + \left(\frac{J_n}{n}\right)^{2\lambda+\varepsilon}\right] + Bn^{-\frac{1}{2}}\\
&\leq (1 - \delta)\, Cn^{-\frac{1}{2}+\varepsilon} + Cn^{-\frac{1}{2}+\varepsilon} Dn^{-1+2\lambda} + Bn^{-\frac{1}{2}} \qquad (6.30)\\
&\leq (1 - (\delta - \delta'))\, Cn^{-\frac{1}{2}+\varepsilon} + Bn^{-\frac{1}{2}}. \qquad (6.31)
\end{aligned}$$

In (6.30) we use Corollary 3.5 together with property (3.6) and $1 + \varepsilon > 1$ and have

$$\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{1+\varepsilon} + \left(\frac{J_n}{n}\right)^{1+\varepsilon}\right] \to \xi < 1, \quad n \to \infty.$$

Hence, there is $0 < \delta < 1$ such that

$$(1 + o(1))\, \mathbb{E}\left[\left(\frac{I_n}{n}\right)^{1+\varepsilon} + \left(\frac{J_n}{n}\right)^{1+\varepsilon}\right] \leq 1 - \delta$$

for n sufficiently large. Additionally, we choose a constant D such that

$$D \geq (1 + o(1))\, A\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda+\varepsilon} + \left(\frac{J_n}{n}\right)^{2\lambda+\varepsilon}\right]$$

for n sufficiently large (observe that $\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda+\varepsilon} + \left(\frac{J_n}{n}\right)^{2\lambda+\varepsilon}\right] \leq 2$ due to $2\lambda + \varepsilon > 0$ and $\frac{I_n}{n}, \frac{J_n}{n} \leq 1$). In (6.31), we choose $0 < \delta' < \delta$ such that for n large enough

$$Dn^{-1+2\lambda} < \delta'.$$

Then, fix $n_0 \in \mathbb{N}$ such that all previous estimates hold for $n \geq n_0$ and choose

$$C \geq \max\left\{\hat{\mathscr{D}}(j)\, j^{\frac{1}{2}+\varepsilon} | j = 1, \ldots, n_0 - 1\right\} \vee \frac{B}{\delta - \delta'}.$$

Finally, we have

$$\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right) \leq Cn^{-\frac{1}{2}+\varepsilon} \quad \text{for } n \geq 1.$$

The same reasoning holds for $\zeta_3\left(\hat{\mathcal{Y}}_n, \mathcal{N}(0,1)\right)$ and therefore the assertion of Corollary 6.22 follows. □

Resumption of the Proof of Theorem 6.11. Finally, we want to condense the results of the three steps:

Step 1* yields a recursive estimate for $\zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right)$ stated in Proposition 6.12;

Step 2* gives candidates for the rates by identifying upper bounds for $\zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ in Proposition 6.13 and Corollary 6.19, respectively; Finally, in

Step 3* we merge the results of **Step 1*** and **Step 2*** into rates for $\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}(0,1)\right) \leq \zeta_3\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Q}}_n^{\mathrm{b}}\right) + \zeta_3\left(\hat{\mathcal{Q}}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ in Corollary 6.21 and Corollary 6.22.

Hence, we obtain

$$\zeta_3^{\vee}\left(\left(\hat{\mathcal{X}}_n, \hat{\mathcal{Y}}_n\right), \left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda - \frac{1}{2}\right)}\right), & \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon}\right), & \lambda < \frac{1}{3},\ \lambda \neq 0. \end{cases}$$

Finally, this is the assertion of Theorem 6.11. $\square$

Remark 6.23 (The Case $\lambda = 0$)**.** As mentioned in the beginning of this chapter, when $\lambda = 0$, the number of black balls is not driven by the draws from the urn. The “black” and the “white” coin are the same. Therefore, the number of black balls is a sum of independent $\mathrm{Ber}\left(\alpha\right)$-distributed random variables plus initial number of black balls. Studying the asymptotic behaviour of the normalised number of black balls reduces to the situation of the classical Central Limit Theorem:

In every step of the urn, a black ball is added with probability α and a white ball with probability $1-\alpha$. Hence, let $\left(C_j\right)_{j\in\mathbb{N}}$ be a sequence of independent $\mathrm{Ber}\left(\alpha\right)$-distributed random variables. Then, for $n \geq 1$,

$$B_n^{\mathrm{b}} \stackrel{\mathrm{d}}{=} 1 + \sum_{j=1}^{n} C_j, \quad B_n^{\mathrm{w}} \stackrel{\mathrm{d}}{=} 0 + \sum_{j=1}^{n} C_j,$$

and

$$\hat{\mathcal{X}}_n = \frac{B_n^{\mathrm{b}} - (1 - n\alpha)}{\sqrt{n\alpha\left(1-\alpha\right)}}, \quad \hat{\mathcal{Y}}_n = \frac{B_n^{\mathrm{w}} - n\alpha}{\sqrt{n\alpha\left(1-\alpha\right)}}.$$

Now, let $N \sim \mathcal{N}\left(0,1\right)$, $N_i \sim \mathcal{N}\left(0,\alpha\left(1-\alpha\right)\right)$, $i = 1,\ldots,n$, and let these random variables be independent of C_i, $i = 1,\ldots,n$. Then, we have

$$\begin{aligned}
\zeta_3\left(\hat{\mathcal{X}}_n, \mathcal{N}\left(0,1\right)\right) =& \zeta_3\left(\frac{\sum_{i=1}^{n} C_i - n\alpha}{\sqrt{n\alpha\left(1-\alpha\right)}}, \mathcal{N}\left(0,1\right)\right) \\
\leq& \left(n\alpha\left(1-\alpha\right)\right)^{-\frac{3}{2}} \zeta_3\left(\sum_{i=1}^{n} C_i - n\alpha, \sqrt{n\alpha\left(1-\alpha\right)}N\right) \\
=& \left(n\alpha\left(1-\alpha\right)\right)^{-\frac{3}{2}} \zeta_3\left(\sum_{i=1}^{n} \left(C_i - \alpha\right), \sum_{i=1}^{n} N_i\right) \\
\leq& \left(n\alpha\left(1-\alpha\right)\right)^{-\frac{3}{2}} \sum_{i=1}^{n} \zeta_3\left(C_i - \alpha, N_i\right) \\
=& \left(n\alpha\left(1-\alpha\right)\right)^{-\frac{3}{2}} n \underbrace{\zeta_3\left(C_1 - \alpha, N_1\right)}_{<\infty} \\
=& O\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where we used that $\sqrt{n\alpha\left(1-\alpha\right)}N \sim \mathcal{N}\left(0, n\alpha\left(1-\alpha\right)\right)$ with $\sqrt{n\alpha\left(1-\alpha\right)}N \stackrel{\mathrm{d}}{=} \sum_{i=1}^{n} N_i$, due to the convolution property of the normal distribution. Note that all Zolotarev distances are finite, as first and second moments of the involved random variables coincide and third moments obviously exist.

7. Rates of Convergence: Arbitrary Initial Composition of the Urn

In this chapter, the rates obtained in Chapter 5 are proven to also hold for a Pólya urn characterised by setting **Det R** with an arbitrary initial composition. Therefore, the recursive approach of Chapter 3 is extended such that the number of black balls is described by a forest of associated trees. Thereby, the number of black balls after n steps when starting with an arbitrary initial composition can be characterised in distribution by a combination of the base cases. In doing so, the results on upper bounds for rates of convergence stated in Chapter 5 for the base cases are transferred to the number of black balls when starting with an arbitrary initial composition of the urn.

Firstly, details of setting **Det R** are recalled and stated. Then, the recursive approach of Chapter 3 is extended. It serves to determine a distributional characterisation for the number of black balls in terms of the number of black balls when starting with a single ball B_n^{b} and B_n^{w}. Then, divided into the non-normal and the normal limit case, rates of convergence on the basis of Theorem 5.6 and Theorem 5.15 are derived. The non-normal limit case is studied first and the normal limit case concludes this chapter. In both cases, the distance between the number of black balls and its limit can be combined via the distances of the base cases to their limits, with respect to the Wasserstein distances in the non-normal limit case and with respect to the Zolotarev distance in the normal limit case.

Balanced Irreducible Two-Colour Pólya Urns

(**Det R**)

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ with } a, d \in \mathbb{N}_0 \cup \{-1\} \text{ and } b, c \in \mathbb{N},$$

such that $a + b = c + d =: K - 1 \geq 1$ (balancedness)

and $bc > 0$ (irreducibility).

Recall that the ratio of the eigenvalues of the replacement matrix that determines the asymptotic behaviour is given by $\lambda := \frac{a-c}{a+b}$.

The number of black balls after n steps in a Pólya urn scheme driven by setting **Det R** with initial composition $\mathscr{I} := (b_0, w_0)$, i.e., in the beginning the urn contains b_0 black and w_0 white balls, is denoted by B_n.

In order to normalise, information on mean and variance of the number of black balls B_n is needed:

Lemma 7.1 (Mean and Variance of the Number of Black Balls)**.**

1. *For the mean of B_n, it holds, as $n \to \infty$,*

   a) *For $\lambda > 0$:*

$$\mathbb{E}\left[B_n\right] = \frac{c\left(a+b\right)}{b+c}n + \left(b_0 - \frac{c}{b+c}\left(b_0+w_0\right)\right)\frac{\Gamma\left(\frac{b_0+w_0}{a+b}\right)}{\Gamma\left(\frac{b_0+w_0+a-c}{a+b}\right)}n^{\lambda} + O\left(1\right).$$

   b) *For $\lambda < 0$:*

$$\mathbb{E}\left[B_n\right] = \frac{c\left(a+b\right)}{b+c}n + O\left(1\right).$$

2. *For the variance of B_n, it holds, as $n \to \infty$,*

   a) *For $\lambda = \frac{1}{2}$:*

$$\mathrm{Var}\left(B_n\right) = bc\, n \ln\left(n\right) + O\left(n\right).$$

   b) *For $0 < \lambda < \frac{1}{2}$:*

$$\mathrm{Var}\left(B_n\right) = \frac{bc\left(a-c\right)^2}{\left(a+b-2\left(a-c\right)\right)\left(b+c\right)^2}n + O\left(n^{2\lambda}\right).$$

   c) *For $\lambda < 0$:*

$$\mathrm{Var}\left(B_n\right) = \frac{bc\left(a-c\right)^2}{\left(a+b-2\left(a-c\right)\right)\left(b+c\right)^2}n + O\left(1\right).$$

We abbreviate $d' := \left(b_0 - \frac{c}{b+c}\left(b_0+w_0\right)\right)\frac{\Gamma\left(\frac{b_0+w_0}{a+b}\right)}{\Gamma\left(\frac{b_0+w_0+a-c}{a+b}\right)}$ as well as mean and variance of the j-th quantity by $\mu_{\mathscr{I}}\left(j\right) := \mathbb{E}\left[B_j\right]$ and $\sigma^2_{\mathscr{I}}\left(j\right) := \mathrm{Var}\left(B_j\right)$ (for the standard deviation $\sigma_{\mathscr{I}}\left(j\right)$, analogously) and use both with random argument as explained before in Remark 5.4.

Proof. See Appendix A.2. □

7.1. A Forest of Associated Trees

Following the recursive approach of Knape and Neininger [27] that was recalled in Chapter 3, the evolution of this urn process can be captured by a forest of associated trees, see also Chauvin et al. [9, Section 3.1]:

Figure 7.1.: A realisation of a forest of associated trees.

Consider a Pólya urn with black and white balls governed by the replacement matrix $\left(\begin{smallmatrix}1 & 4\\ 3 & 2\end{smallmatrix}\right)$ that initially contains one black ball and two white balls.
On the left the urn at time zero and the urn after one draw are depicted. Both are accompanied by the corresponding forest of associated trees at the respective stage.
Every initial ball gives rise to an associated tree of its colour. In the beginning, all the associated trees consist of the root node of the respective colour only.
After the first step, all but one of the associated trees still consist of a root node only. The other associated tree belongs to the ball that was drawn from the urn. It grows according to the procedure explained in detail in Chapter 3, Section 3.1.

Any initial ball is assigned a b- or w-associated tree depending on its colour. The number of black balls after n steps is written in terms of the contributions among these $b_0 + w_0$ trees. This yields the following distributional representation of the number of black balls after n

steps:

Let $\mathcal{I}_\ell^{(n)}$ denote the number of draws from the ℓ-th associated tree with $\ell = 1, \ldots, b_0 + w_0$, note that $\sum_{\ell=1}^{b_0+w_0} \mathcal{I}_\ell^{(n)} = n$ and the marginals are distributed according to

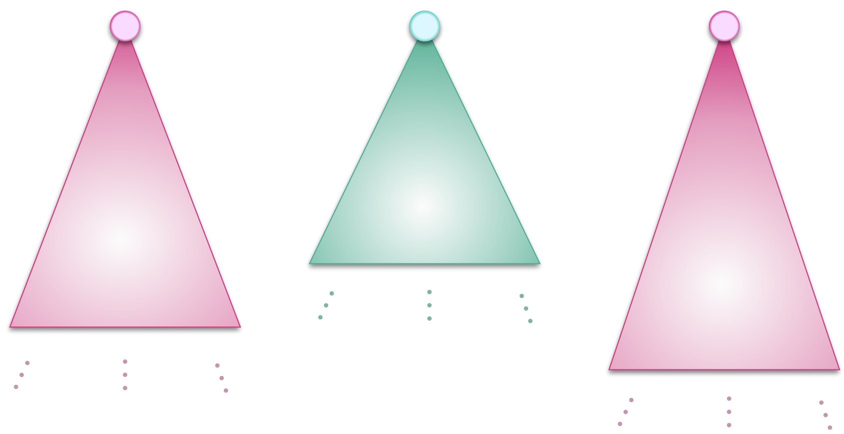
$$\mathbb{P}\left(\mathcal{I}_\ell^{(n)} = j\right) = \binom{n}{j} \frac{\left(\prod_{i=0}^{j-1} (1 + (K-1)\, i)\right) \left(\prod_{i=0}^{n-1-j} (b_0 + w_0 - 1 + (K-1)\, i)\right)}{\prod_{i=0}^{n-1} (b_0 + w_0 + (K-1)\, i)}$$

for $j \in \{0, \ldots, n\}$ (and we set $\prod_{i=0}^{-1} x_i := 1$). The number of black balls after n steps can be characterised via $\mathcal{L}\left(B_n^{\mathrm{b}}\right)$ and $\mathcal{L}\left(B_n^{\mathrm{w}}\right)$, with $B_0 = b_0$ and for $n \geq 1$, as follows:

$$B_n \stackrel{\mathrm{d}}{=} \sum_{\ell=1}^{b_0} B_{\mathcal{I}_\ell^{(n)}}^{\mathrm{b},(\ell)} + \sum_{\ell=b_0+1}^{b_0+w_0} B_{\mathcal{I}_\ell^{(n)}}^{\mathrm{w},(\ell)} \tag{7.1}$$

with $B_j^{\mathrm{b},(\ell)} \stackrel{\mathrm{d}}{=} B_j^{\mathrm{b}}$, $B_j^{\mathrm{w},(\ell)} \stackrel{\mathrm{d}}{=} B_j^{\mathrm{w}}$ for $j = 0, \ldots, n$ and $\ell = 1, \ldots, w_0 + b_0$ such that the families $\left(B_j^{\mathrm{b},(\ell)}\right)_{0 \leq j \leq n}$, $\left(B_j^{\mathrm{w},(\ell)}\right)_{0 \leq j \leq n}$, and $\mathcal{I}^{(n)} := \left(\mathcal{I}_1^{(n)}, \ldots, \mathcal{I}_{b_0+w_0}^{(n)}\right)$ are independent for $\ell = 1, \ldots, b_0 + w_0$.

Figure 7.2.: Evolution of the forest of associated trees: In continuation of the example given in Figure 7.1, the associated trees grow conditioned on their size independently in the passage of time.



The associated trees that correspond to the initial balls in the urn grow analogously to the subtrees of the recursive approach in Chapter 3. Therefore, the results of the parenthesis in Chapter 3, with $\mathcal{C} = b_0 + w_0$ and $r = K - 1$ can be applied to this situation as well: The limit of $\frac{\mathcal{I}^{(n)}}{n}$ is given by a Dirichlet-distributed random vector $\mathcal{D} := (\mathcal{D}_1, \ldots, \mathcal{D}_{b_0+w_0})$ with all $b_0 + w_0$ parameters equal to $\frac{1}{K-1}$, compare Section 3.2. Since characterisation (7.1) is on the level of distributions, as before, we take the liberty to choose the random vector $\mathcal{I}^{(n)}$ suitably coupled to its limit: According to Theorem 3.3, we choose $\mathcal{I}^{(n)}$ and its limit $\mathcal{D}$ such that the L_p-distance between $\frac{\mathcal{I}_\ell^{(n)}}{n}$ and $\mathcal{D}_\ell$ satisfies (3.3). Hence, the growth of the trees of the associated forest can be captured analogously to Lemma 3.4 and Corollary 3.5 which served to describe the behaviour of the rescaled subtree sizes:

Lemma 7.2 (Cousin of Lemma 3.4)**.** *The random vector* $\frac{\mathcal{I}^{(n)}}{n} := \left(\frac{\mathcal{I}_1^{(n)}}{n}, \ldots, \frac{\mathcal{I}_{b_0+w_0}^{(n)}}{n}\right)$ *converges almost surely to a Dirichlet-distributed random vector* $\mathcal{D} := (\mathcal{D}_1, \ldots, \mathcal{D}_{b_0+w_0})$ *with all parameters equal to* $\frac{1}{K-1}$*, as* $n \to \infty$*. Furthermore, let* $\mathcal{I}^{(n)}$ *and* $\mathcal{D}$ *be coupled according to Theorem 3.3 and* $p \geq 1$*. Then, for all* $\ell = 1, \ldots, b_0 + w_0$*, as* $n \to \infty$*,*

$$\left\| \frac{\mathcal{I}_\ell^{(n)}}{n} - \mathcal{D}_\ell \right\|_p = O\left(n^{-\frac{1}{2}}\right).$$

Proof. This follows immediately from Lemma 3.2 and Theorem 3.3. See also the proof of Lemma 3.4. □

Corollary 7.3 (Cousin of Corollary 3.5)**.** *Let* $p \geq 2$ *and* $\psi \in (0,1)$*. Then, for* $\ell = 1, \ldots, b_0 + w_0$*, we have, as* $n \to \infty$*,*

$$\left\| \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\psi - \mathcal{D}_\ell^\psi \right\|_p = O\left(n^{-\frac{\psi}{2}}\right).$$

Proof. Compare the proof of Corollary 3.5. □

7.2. Non-Normal Limit Case: $\lambda > \frac{1}{2}$

In this section, an upper bound for the rate of convergence in the non-normal limit case for the number of black balls when starting the urn process with an arbitrary initial composition is derived. Representation (7.1) is transferred to the normalised quantities. From the distributional representation of the normalised quantities a distributional characterisation for the limiting distribution is deduced.

Since the distribution of the number of black balls is characterised by a convolution of the distributions of the base cases plus a toll term, it is not surprising that also the limit of the normalised number of black balls can be represented in distribution with respect to the limits in the base cases.

With these observations, rates of convergence in the non-normal limit case are derived on the basis of Theorem 5.6. The first step is to derive an upper bound in the Wasserstein distances ℓ_p with $p \geq 1$. As in Chapter 5, the rate can then be transferred to the Kolmogorov-Smirnov distance.

Recall from Section 5.1 the normalised quantities of the base cases B_n^{b} and B_n^{w}, see (5.2), defined via $\mathscr{X}_0 := 0 =: \mathscr{Y}_0$ and, for $n \geq 1$,

$$\mathscr{X}_n := \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{n^\lambda}, \quad \mathscr{Y}_n := \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{n^\lambda}.$$

Accordingly, the normalised number of black balls is defined by $X_0 := 0$ and, for $n \geq 1$,

$$X_n := \frac{B_n - \mathbb{E}\left[B_n\right]}{n^\lambda}. \tag{7.2}$$

The distributional representation of B_n in (7.1) leads to

$$X_n \stackrel{\mathrm{d}}{=} \sum_{\ell=1}^{b_0} \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda X^{\mathrm{b},(\ell)}_{\mathcal{I}_\ell^{(n)}} + \sum_{\ell=b_0+1}^{b_0+w_0} \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda X^{\mathrm{w},(\ell)}_{\mathcal{I}_\ell^{(n)}} + b_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right), \quad n \geq 1, \tag{7.3}$$

where the toll term expands via Lemma 5.1, Lemma 7.1 and Notation 5.3 to

$$\begin{aligned} b_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right) &:= \frac{1}{n^\lambda}\left(\sum_{\ell=1}^{b_0} \mu_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right) + \sum_{\ell=b_0+1}^{b_0+w_0} \mu_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right) - \mu_{\mathscr{I}}(n)\right) \\ &= d_{\mathrm{b}} \sum_{\ell=1}^{b_0} \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda + d_{\mathrm{w}} \sum_{\ell=b_0+1}^{b_0+w_0} \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda - d' + O\left(n^{-\lambda}\right) \end{aligned}$$

with $X_j^{\mathrm{b},(\ell)} \stackrel{\mathrm{d}}{=} \mathscr{X}_j$, $X_j^{\mathrm{w},(\ell)} \stackrel{\mathrm{d}}{=} \mathscr{Y}_j$ for $j = 0, \dots, n$ and $\ell = 1, \dots, b_0 + w_0$ such that $\left(X_j^{\mathrm{b},(\ell)}\right)_{0 \le j \le n}$, $\left(X_j^{\mathrm{w},(\ell)}\right)_{0 \le j \le n}$, $\ell = 1, \dots, b_0 + w_0$, and $\mathcal{I}^{(n)}$ are independent.

As in Chapter 5, we denote the limits of $\mathscr{X}_n$ and $\mathscr{Y}_n$ by Λ_{b} and Λ_{w} that are characterised by system (5.5).

The number of black balls B_n converges in distribution (and almost surely) to a random variable $\Lambda_{\mathscr{I}}$ whose distribution depends on the initial composition of the urn and the replacement matrix, see Janson [22], Chauvin et al. [9]. The distributional representation in (7.3) suggests that this limit satisfies the following distributional equation

$$(7.4) \quad \Lambda_{\mathscr{I}} \stackrel{\mathrm{d}}{=} \sum_{\ell=1}^{b_0} \mathcal{D}_\ell^\lambda \Lambda_{\mathrm{b}}^{(\ell)} + \sum_{\ell=b_0+1}^{b_0+w_0} \mathcal{D}_\ell^\lambda \Lambda_{\mathrm{w}}^{(\ell)} + b_{\mathscr{I}},$$

where

$$b_{\mathscr{I}} := d_b \sum_{\ell=1}^{b_0} \mathcal{D}_\ell^\lambda + d_w \sum_{\ell=b_0+1}^{b_0+w_0} \mathcal{D}_\ell^\lambda - d'$$

with independent copies $\Lambda_{\mathrm{b}}^{(\ell)}$ of Λ_{b} and $\Lambda_{\mathrm{w}}^{(\ell)}$ of Λ_{w} for $\ell = 1, \dots, b_0 + w_0$.

The following theorem extends the results of Theorem 5.6 to Pólya urns with an arbitrary initial composition in the non-normal limit case:

Theorem 7.4. *Let X_n denote the normalised number of black balls of a Pólya urn scheme characterised by **Det R** with arbitrary initial composition $\mathscr{I} = (b_0, w_0)$ as defined in (7.2) and $\lambda := \frac{a-c}{a+b} > \frac{1}{2}$. Furthermore, let $\mathcal{L}(\Lambda_{\mathscr{I}})$ be characterised by (7.4) and $\varepsilon > 0$ as well as $p \ge 1$. Then, as $n \to \infty$,*

$$\ell_p(X_n, \Lambda_{\mathscr{I}}) = O\left(n^{-\lambda + \frac{1}{2} + \varepsilon}\right),$$
$$\varrho(X_n, \Lambda_{\mathscr{I}}) = O\left(n^{-\lambda + \frac{1}{2} + \varepsilon}\right).$$

Proof. Let $\varepsilon > 0$. We choose the quantities arising in (7.3) and (7.4) such that the pairs $\left(X_j^{\mathrm{b},(\ell)}, \Lambda_{\mathrm{b}}^{(\ell)}\right)$ and $\left(X_j^{\mathrm{w},(\ell)}, \Lambda_{\mathrm{w}}^{(\ell)}\right)$ are optimal couplings of the respective distributions and furthermore $\Lambda_{\mathrm{b}}^{(\ell)}$ and $\Lambda_{\mathrm{w}}^{(\ell)}$ are independent of $\mathcal{I}^{(n)}$ for $\ell = 1, \dots, b_0 + w_0$.

We begin with the Wasserstein distances ℓ_p: Let $p \ge 1$ and $\mathscr{M}_p := \max\left\{\|\Lambda_{\mathrm{b}}\|_p, \|\Lambda_{\mathrm{w}}\|_p\right\} < \infty$ (due to Kuba and Sulzbach [28, Theorem 2]). From Theorem 5.6 we know of the existence of a constant $C_p > 0$ such that

$$(7.5) \quad \Delta_p(j) := \ell_p^\vee\left((\mathscr{X}_j, \mathscr{Y}_j), (\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}})\right) \le C_p j^{-\lambda + \frac{1}{2} + \varepsilon} \quad \text{for } j \ge 1,$$

and note that $\Delta_p(0)$ does not contribute in the following computation. With distributional representations (7.3) and (7.4), we have

$$\begin{aligned}
\ell_p\left(X_n,\Lambda_{\mathscr{I}}\right) \leq\ & \Bigg\| \sum_{\ell=1}^{b_0}\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\left(X^{\mathrm{b},(\ell)}_{\mathcal{I}_\ell^{(n)}}-\Lambda_{\mathrm{b}}^{(\ell)}\right)+\sum_{\ell=b_0+1}^{b_0+w_0}\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\left(X^{\mathrm{w},(\ell)}_{\mathcal{I}_\ell^{(n)}}-\Lambda_{\mathrm{w}}^{(\ell)}\right)\\
&+\sum_{\ell=1}^{b_0}\left(\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda-\mathcal{D}_\ell^\lambda\right)\Lambda_{\mathrm{b}}^{(\ell)}+\sum_{\ell=b_0+1}^{b_0+w_0}\left(\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda-\mathcal{D}_\ell^\lambda\right)\Lambda_{\mathrm{w}}^{(\ell)}\\
&+b_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right)-b_{\mathscr{I}}\Bigg\|_p\\
\leq\ & \sum_{\ell=1}^{b_0}\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\left(X^{\mathrm{b},(\ell)}_{\mathcal{I}_\ell^{(n)}}-\Lambda_{\mathrm{b}}^{(\ell)}\right)\right\|_p+\sum_{\ell=b_0+1}^{b_0+w_0}\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\left(X^{\mathrm{w},(\ell)}_{\mathcal{I}_\ell^{(n)}}-\Lambda_{\mathrm{w}}^{(\ell)}\right)\right\|_p \qquad (7.6)\\
&+\sum_{\ell=1}^{b_0+w_0}\mathscr{M}_p\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda-\mathcal{D}_\ell^\lambda\right\|_p+\left\|b_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right)-b_{\mathscr{I}}\right\|_p.
\end{aligned}$$

Via conditioning on $\mathcal{I}^{(n)}$ and due to independence conditioned on $\mathcal{I}^{(n)}$, we observe

$$\begin{aligned}
\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\left(X^{\mathrm{b},(\ell)}_{\mathcal{I}_\ell^{(n)}}-\Lambda_{\mathrm{b}}^{(\ell)}\right)\right\|_p^p &= \sum_{j=0}^{n-1}\mathbb{P}\left(\mathcal{I}_\ell^{(n)}=j\right)\left(\frac{j}{n}\right)^{p\lambda}\mathbb{E}\left[\left|X_j^{\mathrm{b},(\ell)}-\Lambda_{\mathrm{b}}^{(\ell)}\right|^p\right]\\
&\leq \sum_{j=0}^{n-1}\mathbb{P}\left(\mathcal{I}_\ell^{(n)}=j\right)\left(\frac{j}{n}\right)^{p\lambda}\left(\Delta_p(j)\right)^p\\
&= \left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\Delta_p\left(\mathcal{I}_\ell^{(n)}\right)\right\|_p^p,
\end{aligned}$$

where we make use of the fact that the pair $\left(X_j^{\mathrm{b},(\ell)},\Lambda_{\mathrm{b}}^{(\ell)}\right)$ is an optimal coupling. Hence, we have for $\ell=1,\ldots,b_0+w_0$,

$$\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\left(X^{\mathrm{b},(\ell)}_{\mathcal{I}_\ell^{(n)}}-\Lambda_{\mathrm{b}}^{(\ell)}\right)\right\|_p,\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\left(X^{\mathrm{w},(\ell)}_{\mathcal{I}_\ell^{(n)}}-\Lambda_{\mathrm{w}}^{(\ell)}\right)\right\|_p\leq\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda\Delta_p\left(\mathcal{I}_\ell^{(n)}\right)\right\|_p. \qquad (7.7)$$

From Corollary 7.3, we have

$$\sum_{\ell=1}^{b_0+w_0}\mathscr{M}_p\left\|\left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^\lambda-\mathcal{D}_\ell^\lambda\right\|_p=O\left(n^{-\frac{\lambda}{2}}\right). \qquad (7.8)$$

Moreover, repeating the proof of Lemma 5.11 with Corollary 7.3 instead of Corollary 3.5 we obtain

$$\left\| b_{\mathscr{I}}\left(\mathcal{I}_\ell^{(n)}\right) - b_{\mathscr{I}} \right\|_2 = O\left(n^{-\frac{\lambda}{2}}\right). \tag{7.9}$$

Therefore, (7.6) combined with (7.7), (7.8) and (7.9) yields

$$\begin{aligned}
\ell_p\left(X_n, \Lambda_{\mathscr{I}}\right) &\leq \sum_{\ell=1}^{b_0+w_0} \left\| \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^{\lambda} \Delta_p\left(\mathcal{I}_\ell^{(n)}\right) \right\|_p + O\left(n^{-\frac{\lambda}{2}}\right) \\
&\leq \sum_{\ell=1}^{b_0+w_0} C_p n^{\lambda-\frac{1}{2}+\varepsilon} \left\| \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^{\frac{1}{2}+\varepsilon} \right\|_p + O\left(n^{-\frac{\lambda}{2}}\right) &\text{(7.10)} \\
&\leq \left(b_0 + w_0\right) C_p n^{\lambda-\frac{1}{2}+\varepsilon} + O\left(n^{-\frac{\lambda}{2}}\right), &\text{(7.11)}
\end{aligned}$$

where we used (7.5) in (7.10) as well as $\left\| \left(\frac{\mathcal{I}_\ell^{(n)}}{n}\right)^{\frac{1}{2}+\varepsilon} \right\|_p \leq 1$ in (7.11). Hence, we have

$$\ell_p\left(X_n, \Lambda_{\mathscr{I}}\right) = O\left(n^{\lambda-\frac{1}{2}+\varepsilon}\right). \tag{7.12}$$

Finally, we can transfer the rate to the Kolmogorov-Smirnov distance: Based on (7.12), the final step consists of performing the same considerations as in the proof of Proposition 5.14; from Kuba and Sulzbach [28, Theorem 2] we know that the density of $\Lambda_{\mathscr{I}}$ is bounded. Hence, Lemma 2.6 can be applied yielding, as $n \to \infty$,

$$\varrho\left(\hat{X}_n, \Lambda_{\mathscr{I}}\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right).$$

Finally, the assertion follows. □

The proof of Theorem 7.4, keeping representation (7.4) in mind, shows that the limiting distribution $\mathcal{L}\left(\Lambda_{\mathscr{I}}\right)$ is a convolution of the "base cases" $\mathcal{L}\left(\Lambda_{\mathrm{b}}\right)$ and $\mathcal{L}\left(\Lambda_{\mathrm{w}}\right)$, where both appear with the respective number of initial balls, plus a toll term that also takes the initial composition of the urn into account. In a slightly different situation, an analogous observation was already made in Chauvin et al. [9].

Remark 7.5. The proof of Theorem 7.4 shows that the analogous result holds for the number of black balls in setting **Rand R**.

7.3. Normal Limit Case: $\lambda \leq \frac{1}{2}$

This section deals with upper bounds for rates of convergence in the normal limit case for the number of black balls when starting with an arbitrary initial composition. Representation (7.1) leads to a distributional characterisation of the normalised number of black balls in terms of the base cases.

Theorem 5.15 serves as basis for the results stated in this section. The proof is performed with the help of an accompanying sequence that links the sequence and its limit.

Recall from Section 5.2 the normalised number of black balls for the base cases B_n^{b} and B_n^{w}, see (5.41): $\hat{\mathscr{X}}_0 := 0 =: \hat{\mathscr{Y}}_0$, $\hat{\mathscr{X}}_1 := 0 =: \hat{\mathscr{Y}}_1$ and, for $n \geq 2$,

$$\hat{\mathscr{X}}_n := \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{\sqrt{\operatorname{Var}\left(B_n^{\mathrm{b}}\right)}}, \quad \hat{\mathscr{Y}}_n := \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{\sqrt{\operatorname{Var}\left(B_n^{\mathrm{w}}\right)}}.$$

Furthermore, recall the situation of p. 101 et seq. as well as the notation given in Lemma 7.1.

According to the situation of Section 5.2, in order to normalise the number of black balls B_n the exact scaling is used. Let for $n \geq 2$ (note that for monochromatic initial compositions the first step is deterministic and therefore $\operatorname{Var}(B_1) = 0$),

$$(7.13)\quad \hat{X}_n := \frac{B_n - \mathbb{E}\left[B_n\right]}{\sqrt{\operatorname{Var}\left(B_n\right)}}.$$

The distributional representation of B_n in (7.1) transfers to the normalised quantity as

$$(7.14)\quad \hat{X}_n \stackrel{\mathrm{d}}{=} \sum_{\ell=1}^{b_0} \frac{\sigma_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)} \hat{X}^{\mathrm{b},(\ell)}_{\mathcal{I}_\ell^{(n)}} + \sum_{\ell=b_0+1}^{b_0+w_0} \frac{\sigma_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)} \hat{X}^{\mathrm{w},(\ell)}_{\mathcal{I}_\ell^{(n)}} + t_{\mathscr{I}}\left(\mathcal{I}_\ell^{(n)}\right), \quad n \geq 2,$$

where

$$t_{\mathscr{I}}\left(\mathcal{I}_\ell^{(n)}\right) := \frac{1}{\sigma_{\mathscr{I}}(n)} \left(\sum_{\ell=1}^{b_0} \mu_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right) + \sum_{\ell=b_0+1}^{b_0+w_0} \mu_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right) - \mu_{\mathscr{I}}(n) \right)$$

with $\hat{X}_j^{\mathrm{b},(\ell)} \stackrel{\mathrm{d}}{=} \hat{\mathscr{X}}_j$, $\hat{X}_j^{\mathrm{w},(\ell)} \stackrel{\mathrm{d}}{=} \hat{\mathscr{Y}}_j$ for $j = 0, \ldots, n$ and $\ell = 1, \ldots, b_0 + w_0$ such that $\left(\hat{X}_j^{\mathrm{b},(\ell)}\right)_{0 \leq j \leq n}$, $\left(\hat{X}_j^{\mathrm{w},(\ell)}\right)_{0 \leq j \leq n}$, $\ell = 1, \ldots, b_0 + w_0$, and $\mathcal{I}^{(n)}$ are independent.

The normalised quantity $\hat{X}_n$ converges in distribution to the standard normal distribution, cf. Janson [22].

The normal distribution is rewritten inspired by (7.14) as

$$\mathcal{N}(0,1) \stackrel{\mathrm{d}}{=} \sum_{\ell=1}^{b_0} \sqrt{\mathcal{D}_\ell} N_\ell + \sum_{\ell=b_0+1}^{b_0+w_0} \sqrt{\mathcal{D}_\ell} N_\ell$$

with independent and standard normally distributed N_ℓ, $\ell = 1, \ldots, b_0 + w_0$.

Theorem 7.6. *Let $\hat{X}_n$ denote the normalised number of black balls of a Pólya urn scheme characterised by* ***Det R*** *as defined in* (7.13) *with $\lambda := \frac{a-c}{a+b} \leq \frac{1}{2}$ and let $\varepsilon > 0$. Then, as $n \to \infty$,*

$$\zeta_3\left(\hat{X}_n, \mathcal{N}(0,1)\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda - \frac{1}{2}\right)}\right), & \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon}\right), & \lambda \leq \frac{1}{3},\ \lambda \neq 0. \end{cases}$$

Proof. The steps of this proof are very similar to the steps performed in the proof of Theorem 5.15. Therefore, not every step is carried out in detail but the reader is rather referred to the respective part of the proof of Theorem 5.15.

As in the proof of Theorem 5.15, at first the accompanying sequence is introduced. Let Q_n denote the accompanying sequence, defined by

$$Q_n := \sum_{\ell=1}^{b_0} \frac{\sigma_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)} N_\ell + \sum_{\ell=b_0+1}^{b_0+w_0} \frac{\sigma_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)} N_\ell + t_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right), \quad n \geq 2,$$

with independent standard normally distributed N_ℓ, $\ell = 1, \ldots, b_0 + w_0$, also independent of all quantities occurring in (7.14).

With the triangle inequality, we have

$$\zeta_3\left(\hat{X}_n, \mathcal{N}(0,1)\right) \leq \zeta_3\left(\hat{X}_n, Q_n\right) + \zeta_3\left(Q_n, \mathcal{N}(0,1)\right).$$

Note that both distances $\zeta_3\left(\hat{X}_n, Q_n\right)$ and $\zeta_3\left(Q_n, \mathcal{N}(0,1)\right)$ are finite for $n \geq 2$ since $\mathbb{E}\left[\hat{X}_n\right] = \mathbb{E}[Q_n] = 0$, $\operatorname{Var}\left(\hat{X}_n\right) = \operatorname{Var}(Q_n) = 1$ and $\left\|\hat{X}_n\right\|_3 < \infty$ due to $\left\|X_j^{\mathrm{b}}\right\|_3, \left\|X_j^{\mathrm{w}}\right\|_3 < \infty$ for $j \in \mathbb{N}$ as well as $\|Q_n\|_3 < \infty$ due to $\|N_\ell\|_3 < \infty$, $\ell = 1, \ldots, b_0 + w_0$ and the boundedness of the toll term; of course, for $N \sim \mathcal{N}(0,1)$ the same conditions on mean, variance and third moment are satisfied.

Then, the distances $\zeta_3\left(\hat{X}_n, Q_n\right)$ and $\zeta_3\left(Q_n, \mathcal{N}(0,1)\right)$ are studied separately.

As in Chapter 5, let $\hat{\Delta}(j) := \zeta_3^\vee\left(\left(\hat{\mathscr{X}}_j, \hat{\mathscr{Y}}_j\right), (\mathcal{N}(0,1), \mathcal{N}(0,1))\right)$. Let $\varepsilon > 0$, then there is a constant $C > 0$ such that $\hat{\Delta}(j) \leq C r_\lambda(j)$ for $j \geq 2$ where we set $r_\lambda(n) := 0$ for $n = 0, 1$ and

$$r_\lambda(n) := \begin{cases} (\ln(n))^{-\frac{3}{2}}, & \lambda = \frac{1}{2}, \\ n^{3\left(\lambda - \frac{1}{2}\right)}, & \frac{1}{3} < \lambda < \frac{1}{2}, \\ n^{-\frac{1}{2}+\varepsilon}, & \lambda \leq \frac{1}{3},\ \lambda \neq 0. \end{cases}$$

Note that $\hat{\Delta}(0)$ and $\hat{\Delta}(1)$ do not contribute.

Performing the same steps as in Proposition 5.17, one obtains

$$\begin{aligned} &\zeta_3\left(\hat{X}_n, Q_n\right) \\ \leq\ & \mathbb{E}\left[\sum_{\ell=1}^{b_0}\left(\frac{\sigma_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \hat{\Delta}\left(\mathcal{I}_\ell^{(n)}\right) + \sum_{\ell=b_0+1}^{b_0+w_0}\left(\frac{\sigma_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \hat{\Delta}\left(\mathcal{I}_\ell^{(n)}\right)\right]. \end{aligned} \tag{7.15}$$

Due to Lemma 7.1, the same estimates for the ratios of the standard deviations hold as in Corollaries 5.22, 5.23 and 5.24.

The other distance $\zeta_3(Q_n, \mathcal{N}(0,1))$ fits in the setting of Proposition 5.18 with the following observations: With $N \sim \mathcal{N}(0,1)$ independent of $\mathcal{I}^{(n)}$, it holds

$$Q_n \stackrel{\mathrm{d}}{=} \left(\sum_{\ell=1}^{b_0}\frac{\sigma_{\mathrm{b}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)} + \sum_{\ell=b_0+1}^{b_0+w_0}\frac{\sigma_{\mathrm{w}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)}\right)^{\frac{1}{2}} N + t_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right).$$

Let

$$G_n := \left(\sum_{\ell=1}^{b_0}\frac{\sigma_{\mathrm{b}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)} + \sum_{\ell=b_0+1}^{b_0+w_0}\frac{\sigma_{\mathrm{w}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)}\right)^{\frac{1}{2}}$$

and

$$A_n := \{G_n \geq 1\} \quad \text{as well as} \quad \Delta_n := \sqrt{|G_n^2 - 1|}.$$

Then, we define

$$\begin{aligned} \check{Q}_n &:= \mathbb{1}_{A_n}\left(N + \Delta_n N' + t_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right)\right) + \mathbb{1}_{A_n^{\mathrm{c}}}\left(G_n N + t_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right)\right) \stackrel{\mathrm{d}}{=} Q_n, \quad \text{and} \\ \check{N}_n &:= \mathbb{1}_{A_n} N + \mathbb{1}_{A_n^{\mathrm{c}}}\left(G_n + \Delta_n N'\right) \stackrel{\mathrm{d}}{=} N. \end{aligned}$$

Inserting $\check{Q}_n$ and $\check{N}_n$ for $\hat{Q}_n^{\mathrm{b}}$ and $\hat{N}$ in the proof of Proposition 5.18 yields

$$\zeta_3\left(Q_n, \mathcal{N}(0,1)\right) = O\left(\left\|\sum_{\ell=1}^{b_0} \frac{\sigma_{\mathrm{b}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)} + \sum_{\ell=b_0+1}^{b_0+w_0} \frac{\sigma_{\mathrm{w}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)} - 1\right\|_{\frac{3}{2}}^{\frac{3}{2}} + \left\|t_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right)\right\|_3^3\right).$$

Observe that (almost) the same estimates as done in Lemmata 5.19 and 5.20 hold for

$$\left\|t_{\mathscr{I}}\left(\mathcal{I}^{(n)}\right)\right\|_3 \quad \text{and} \quad \left\|\sum_{\ell=1}^{b_0} \frac{\sigma_{\mathrm{b}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)} + \sum_{\ell=b_0+1}^{b_0+w_0} \frac{\sigma_{\mathrm{w}}^2\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}^2(n)} - 1\right\|_{\frac{3}{2}},$$

respectively.

Hence, we have

$$\zeta_3\left(Q_n, \mathcal{N}(0,1)\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda - \frac{1}{2}\right)}\right), & 0 < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{3}{2}}\right), & \lambda < 0. \end{cases} \tag{7.16}$$

Combining (7.15) and (7.16) yields, for $n \geq 2$,

$$\begin{aligned} &\zeta_3\left(\hat{X}_n, \mathcal{N}(0,1)\right) \\ \leq\ & \mathbb{E}\left[\sum_{\ell=1}^{b_0} \left(\frac{\sigma_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \hat{\Delta}\left(\mathcal{I}_\ell^{(n)}\right) + \sum_{\ell=b_0+1}^{b_0+w_0} \left(\frac{\sigma_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \hat{\Delta}\left(\mathcal{I}_\ell^{(n)}\right)\right] \\ &+\zeta_3\left(Q_n, \mathcal{N}(0,1)\right) \\ \leq\ & C r_\lambda(n)\, \mathbb{E}\left[\sum_{\ell=1}^{b_0} \left(\frac{\sigma_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \frac{r_\lambda\left(\mathcal{I}_\ell^{(n)}\right)}{r_\lambda(n)} + \sum_{\ell=b_0+1}^{b_0+w_0} \left(\frac{\sigma_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \frac{r_\lambda\left(\mathcal{I}_\ell^{(n)}\right)}{r_\lambda(n)}\right] \\ &+B r_\lambda(n) \end{aligned}$$

with a suitable constant $B > 0$.

Similar to considerations (5.61) and (5.62) for $\lambda = \frac{1}{2}$, (5.63) and (5.64) for $\frac{1}{3} < \lambda < \frac{1}{2}$, (5.65) and (5.66) for $0 < \lambda \leq \frac{1}{3}$ and, finally, (5.68) and (5.69) for $\lambda < 0$, it holds

$$\mathbb{E}\left[\sum_{\ell=1}^{b_0} \left(\frac{\sigma_{\mathrm{b}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \frac{r_\lambda\left(\mathcal{I}_\ell^{(n)}\right)}{r_\lambda(n)} + \sum_{\ell=b_0+1}^{b_0+w_0} \left(\frac{\sigma_{\mathrm{w}}\left(\mathcal{I}_\ell^{(n)}\right)}{\sigma_{\mathscr{I}}(n)}\right)^3 \frac{r_\lambda\left(\mathcal{I}_\ell^{(n)}\right)}{r_\lambda(n)}\right] = O(1).$$

This yields for $n \geq 2$

$$\zeta_3\left(\hat{X}_n, \mathcal{N}(0,1)\right) = O\left(r_\lambda(n)\right).$$

□

Remark 7.7. The proof of Theorem 7.6 shows that the analogous result holds for the number of black balls in setting **Rand R**.

8. Concluding Remarks and Future Work

In this thesis, for the first time comprehensive results on upper bounds for rates of convergence covering the class of balanced, irreducible two-colour Pólya urn schemes are presented.
As conclusion of this thesis, the following chapter serves to discuss the results and the methods leading to them. To enhance structure and readability, some parts are captured as remark whenever they were referred to earlier in this thesis. Moreover, questions for future research are raised.

The proofs of Chapters 5 and 6 and the reasoning of Chapter 7 suggest that the results of Theorem 7.4 and Theorem 7.6 can be extended to a Pólya urn scheme ruled by a replacement matrix $\mathscr{R}$ that combines the properties of R from setting **Det R** and $\bar{R}$ from setting **Rand R**, i.e.,

$$\mathscr{R} := \begin{pmatrix} \mathscr{R}_{11} & \mathscr{R}_{12} \\ \mathscr{R}_{21} & \mathscr{R}_{22} \end{pmatrix}$$

with random integer entries $\mathscr{R}_{11}$, $\mathscr{R}_{12}$, $\mathscr{R}_{21}$, and $\mathscr{R}_{22}$ such that $\mathscr{R}_{jj}$ takes values in $\mathbb{N}_0 \cup \{-1\}$ for $j = 1, 2$ and $\mathscr{R}_{ij}$ takes values in $\mathbb{N}$ for $i \neq j$ (yielding $\mathscr{R}_{12}\mathscr{R}_{21} > 0$) with $\mathscr{R}_{11} + \mathscr{R}_{12} = \mathscr{R}_{21} + \mathscr{R}_{22} := K - 1$ almost surely.

Furthermore, it is left for future research to extend the results stated in Theorem 5.6 and Theorem 7.4 as well as Theorem 5.15 and Theorem 7.6 to balanced, irreducible Pólya urns with more than two colours and to the class of tenable Pólya urns.

Moreover, the proofs of Chapters 5 and 6 suggest that the results on upper bounds for rates of convergence can be extended to systems of distributional recursions in general. Recall the general setting of Chapter 4, especially the quantities of (4.1) and (4.2). The proofs of Theorems 5.6 and 5.15 suggest theorems of the following form:

Upper Bound for a Rate of Convergence for a System of Distributional Recursions in the Non-Normal Limit Setting from Theorem 4.1
Given that the normalised quantity $\mathbf{X}_n := \left(\left(X_n^{[1]}, \ldots, X_n^{[\mathfrak{T}]}\right)\right)$ as in (4.1) and its limit $\mathbf{X} =$

$\left(\left(X^{[1]}, \ldots, X^{[\mathfrak{T}]}\right)\right)$ as in (4.2) satisfy Theorem 4.1, let $r : \mathbb{N} \to \mathbb{R}_0^+$ be a decreasing sequence with $r(n) \to 0$. If the following conditions are satisfied

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{r=1}^{\mathfrak{B}} \left(A_r^{(n),[\pi(k,r)]}\right)^2 \left(\frac{r\left(J_r^{(n)}\right)}{r(n)}\right)^2\right] < 1,$$
$$\ell_2\left(A_r^{(n),[\pi(k,r)]}, A_{[\pi(k,r)]r}\right) = O\left(r\left(n\right)\right), \quad k = 1, \ldots, \mathfrak{T},\ r = 1, \ldots, \mathfrak{B},$$
$$\ell_2\left(t_{[k]}\left(n\right), t_{[k]}\right) = O\left(r\left(n\right)\right), \quad k = 1, \ldots, \mathfrak{T},\ r = 1, \ldots, \mathfrak{B},$$

then,

$$\ell_2^{\vee}\left(\mathbf{X}_n, \mathbf{X}\right) = O\left(r\left(n\right)\right).$$

Upper Bounds for a Rate of Convergence for a System of Distributional Recursions in the Normal Limit Setting from Theorem 4.2
Given that the normalised quantity $\mathbf{X}_n := \left(\left(X_n^{[1]}, \ldots, X_n^{[\mathfrak{T}]}\right)\right)$ as in (4.1), where the scaling is the exact standard deviation, and its limit $\mathbf{X} = \left(\left(X^{[1]}, \ldots, X^{[\mathfrak{T}]}\right)\right)$ as in (4.2) satisfy Theorem 4.2, let $r : \mathbb{N} \to \mathbb{R}_0^+$ be a decreasing sequence with $r(n) \to 0$. If the following conditions are satisfied

$$\limsup_{n\to\infty} \mathbb{E}\left[\sum_{r=1}^{\mathfrak{B}} \left(A_r^{(n),[\pi(k,r)]}\right)^3 \left(\frac{r\left(J_r^{(n)}\right)}{r(n)}\right)\right] < 1,$$
$$\left(\ell_{\frac{3}{2}}\left(\left(A_r^{(n),[\pi(k,r)]}\right)^2, \left(A_{[\pi(k,r)]r}\right)^2\right)\right)^{\frac{3}{2}} = O(r(n)),\ k = 1, \ldots, \mathfrak{T},\ r = 1, \ldots, \mathfrak{B},$$
$$\left(\ell_3\left(t_{[k]}\left(n\right), t_{[k]}\right)\right)^3 = O\left(r\left(n\right)\right), \quad k = 1, \ldots, \mathfrak{T},\ r = 1, \ldots, \mathfrak{B},$$

then,

$$\zeta_3^{\vee}\left(\mathbf{X}_n, \left(\mathcal{N}\left(0,1\right), \ldots, \mathcal{N}\left(0,1\right)\right)\right) = O\left(r\left(n\right)\right).$$

In contrast to other methods Pólya urns were treated with, the possibility to determine rates of convergence is immanent in the contraction method. It seems that this is a very big advantage of the contraction method. Moreover, note that especially in the non-normal limit case — where the limit distribution is not fully known so far — not much knowledge of the properties of the limit is needed.

The following remark is dedicated to explain how the ε enters the upper bounds for the rates stated in Theorem 5.6 and Theorem 5.15, and likewise in Theorems 6.4 and 6.11.

Remark 8.1 (Contractive Behaviour)**.** For performing the inductions, the contractive behaviour of the coefficients as explained in Section 4.2 is crucial. In some cases, the rate that we expect to hold had to be slowed down artificially in order to work properly as induction hypothesis: For our approach it is vital that the exponents, called θ for the moment, of the ratios $\frac{I_r^{(n)}}{n}$, $r = 1, \ldots, K$ in setting **Det R** (in (5.29) & (5.30), (5.40), (5.63) & (5.64), (5.65) & (5.66) and (5.68) & (5.69)) as well as $\frac{I_n}{n}$ and $\frac{J_n}{n}$ in setting **Rand R** (in (6.28) & (6.29) and (6.30) & (6.31)) are strictly greater than 1. Then, we know from Corollary 3.5 together with property (3.6) that the sum of these ratios in the respective settings converges to a number $\xi \in (0, 1)$:

$$\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\theta}\right] \to \xi, \quad n \to \infty.$$

This contractive behaviour guarantees that we have enough "space" left, in terms of a number $\delta \in (0, 1)$, to hide all the other terms arising — given they behave appropriately, i.e., their rate is "fast enough":

$$\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\theta}\right] \leq 1 - \delta.$$

This is the reason why we have to add an $\varepsilon > 0$ to the exponents of the rates for $\lambda > \frac{1}{2}$ and $\lambda \leq \frac{1}{3}$: In these cases the rate without ε in the exponent would yield $\theta = 1$ implying that the coefficients converge to 1 and not to a number strictly less than 1.

Hence, an interesting question that arises as a consequence of our results is whether the ε appearing in our rates can be dropped or not.

In the next remark, we explain which upper bound for the rate of convergence is derived in the non-normal limit case when working with the Zolotarev distance ζ_2 instead of Wasserstein distances.

Remark 8.2 (Remark on the Non-Normal Limit Case)**.** When transferring the rate from Wasserstein to Kolmogorov-Smirnov distance with the help of Lemma 2.6, we "lose" ε in the exponent of the rate in any case. Here again the question arises whether the ε can be dropped.

As mentioned below Theorem 4.1, instead of $\ell_2^{\vee}$, the Zolotarev metric $\zeta_2^{\vee}$ could be used to derive a rate. Of course, the rate with respect to $\zeta_2^{\vee}$ is bounded by the rate with respect to $\ell_2^{\vee}$ due to Lemma 2.7.

Treating the distance $\zeta_2^{\vee}\left(\left(\mathscr{X}_n, \mathscr{Y}_n\right), \left(\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}}\right)\right)$ as done in Knape and Neininger [27, Proof of Theorem 6.1], we observe that we are able to conduct the induction choosing as induction hypothesis the rate that we expect to be correct. Hence, as $n \to \infty$,

$$\zeta_2^{\vee}\left(\left(\mathscr{X}_n, \mathscr{Y}_n\right), \left(\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}}\right)\right) = O\left(n^{-\lambda+\frac{1}{2}}\right).$$

We give a sketch, compare [27, Section 6.1]: Abbreviating $\Delta^{\zeta}(j) := \zeta_2^{\vee}\left(\left(\mathscr{X}_j, \mathscr{Y}_j\right), \left(\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}}\right)\right)$ and setting as induction hypothesis

$$\Delta^{\zeta}(j) \leq C j^{-\lambda+\frac{1}{2}}, \quad 1 \leq j \leq n-1,$$

we obtain with the help of Lemma 2.7 along the lines of the proofs of [27, Proof of Theorem 6.1]

$$\begin{aligned}
\zeta_2\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right) \leq\ & \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta^{\zeta}\left(I_r^{(n)}\right)\right] \\
& + O\left(\sum_{r=1}^{K}\left\|\left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - \left(D_r\right)^{\lambda}\right\|_2\right) \\
& + O\left(\max\left\{\left\|b_{\mathrm{b}}\left(I^{(n)}\right) - b_{\mathrm{b}}\right\|_2, \left\|b_{\mathrm{w}}\left(I^{(n)}\right) - b_{\mathrm{w}}\right\|_2\right\}\right).
\end{aligned}$$

With Corollary 3.5 and Lemma 5.11, we obtain with a suitable constant $A > 0$,

$$\begin{aligned}
\zeta_2\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right) & \leq \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta^{\zeta}\left(I_r^{(n)}\right)\right] + A n^{-\frac{\lambda}{2}} \\
& \leq C n^{-\lambda+\frac{1}{2}} \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda}\left(\frac{I_r^{(n)}}{n}\right)^{-\lambda+\frac{1}{2}}\right] + A n^{-\frac{\lambda}{2}} \\
& \leq C n^{-\lambda+\frac{1}{2}} \mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\lambda+\frac{1}{2}}\right] + A n^{-\frac{\lambda}{2}}.
\end{aligned}$$

Now, we observe that $\lambda + \frac{1}{2} > 1$ and therefore we have with Corollary 3.5 and property (3.6) that there is $\delta > 0$ such that for sufficiently large n, it holds

$$\mathbb{E}\left[\sum_{r=1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^{\lambda+\frac{1}{2}}\right] \leq 1 - \delta.$$

Finally, we are able to choose C such that

$$\zeta_2\left(\mathscr{X}_n, \Lambda_{\mathrm{b}}\right) \leq C n^{-\lambda+\frac{1}{2}}.$$

Obviously, the same can be performed for the distance $\zeta_2(\mathscr{Y}_n, \Lambda_{\mathrm{w}})$.

We see that we are able to prove the rate that we expect to be correct in the Zolotarev metric $\zeta_2^{\vee}$. However, we are not able to deduce a rate in the Kolmogorov-Smirnov distance from that basis.

A natural question to ask is what meaning a rate of convergence in the Zolotarev distance bears, especially in comparison to the Kolmogorov-Smirnov distance:

When dealing with convergence in distribution and rates thereof, it is reasonable to ask for rates of convergence in the Kolmogorov-Smirnov distance. The Kolmogorov-Smirnov distance seems to capture the distance between two distributions in the most natural way and therefore is most commonly used in applications. In the non-normal limit case we give rates of convergence in this distance. When it comes to the normal limit case we are not able to do so. It still remains an open question what behaviour rates of convergence in the Zolotarev distance, here ζ_3, suggest for rates of convergence in the Kolmogorov-Smirnov distance.

In Cramer and Rüschendorf [11] several examples concerning the complexity of recursive algorithms can be found, where the Zolotarev distance ζ_3 yields the same order for an upper bound as the Kolmorogov-Smirnov distance. In the case of Quicksort, Fill and Janson derive both upper and lower bounds for the rate of convergence in both Wasserstein distances and the Kolmogorov-Smirnov distance in [15]. On the other hand, Neininger and Rüschendorf [40] determine the correct rate in the Zolotarev distance ζ_3, i.e., order of upper and lower bound coincide. Fill and Janson mention that they expect the rate derived in the Zolotarev distance also to be the correct rate in the Kolmogorov-Smirnov distance.

Mahmoud and Neininger derive the correct rate of convergence for the distribution of distances in random binary search trees in the Zolotarev metric ζ_3 in [35] that was confirmed to be the same rate in the Kolmogorov-Smirnov metric by Panholzer [44].

When it comes to Pólya urns, the knowledge of rates of convergence is still very sparse. Flajolet et al. [16] derive a rate for urns with subtraction in the Kolmogorov-Smirnov distance. These urns are not fully covered by our approach as they allow for the removal of other balls than the drawn one. However, they belong to the regime of normal limit laws (as the ratio of the eigenvalues λ is negative). They give a rate of order $n^{-\frac{1}{2}}$, that we would expect to be the correct rate for negative λ, as well. Even more confirming are the rates of convergence in the Kolmogorov-Smirnov distance of Hwang [21] in the context of m-ary search trees that are exactly mirrored by the rates we give (except for the ε) in the normal limit case. Hence, it seems reasonable to expect the rates of Theorem 5.15 to hold also in the Kolmogorov-Smirnov distance.

A question that directly follows is what happens if the Zolotarev distance $\zeta_{2+\varepsilon}$ is used instead of ζ_3:

Remark 8.3. As a consequence of Theorem 4.2 it is natural to ask if the rates of convergence derived in Theorem 5.15 also hold in the Zolotarev metric $\zeta_{2+\varepsilon}$. There seem to be two problems: First of all, it is not clear how to find suitable estimates in the proof of Proposition 5.18 such that a corresponding result holds in $\zeta_{2+\varepsilon}$. Secondly, the inductions performed in Corollaries 5.22, 5.23, and 5.24 in **Step 3** of Section 5.2 do not seem to allow a transfer of the rates stated in Theorem 5.15 but seem to demand for a reduction of speed.

Our results do not give information on lower bounds and therefore we cannot make a statement on the optimality of the bounds. Hwang [21] argues that the order of his rates is optimal and so do Goldstein and Reinert [19]. Flajolet et al. [16] do not discuss the optimality of their rate but point out that their rate is already implied by results from Gouet [20]; however it seems that Gouet's results only imply upper bounds. Peköz et al. [45] give lower and upper bounds that are of the same order and therefore provide optimal rates; the Pólya urns they studied do not belong to the class of balanced and irreducible urns studied in this thesis.

In the light of the recursional distributions (3.1) or (3.2) it might look as if a bivariate approach was to be favoured rather than working with systems of distributional recursions as it is done in this thesis. Knape and Neininger dedicate a section to weighing the pros and cons of a bivariate approach, see [27, Section 7]; they find that the approach via systems of distributional recursions where the quantities B_n^{b} and B_n^{w} are not coupled but only their distributions matter is to be preferred. Note that in fact they introduce the notion of a system of distributional recursions in the setting of the contraction method in [27]. Also, Leckey et al. [32] and Leckey et al. [31] prefer to work with systems of distributional recursions.

A. Appendix

A.1. Technical Lemmata

Lemma A.1. *Let* $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ *be twice differentiable with* $f(0) = 0$*,* $f' > 0$*,* $f'' < 0$ *and* $x, y > 0$*. Then,*

$$|f(x) - f(y)| \leq f(|x - y|).$$

Proof. W.l.o.g. let $x > y > 0$.
Case 1: $x - y \leq y$.
According to the Mean value Theorem, there is some $\xi \in (y, x)$ such that $f(x) - f(y) = f'(\xi)(x - y)$ and some $\zeta \in (0, x - y)$ such that $f(x - y) - f(0) = f'(\zeta)(x - y - 0)$. As $f'' < 0$, f' is strictly decreasing and, therefore, $f'(\zeta) > f'(\xi)$. This yields

$$f(x) - f(y) = f'(\xi)(x - y) \leq f'(\zeta)(x - y) = f(x - y).$$

Case 2: $y < x - y$.
Again, by the Mean value Theorem, we obtain the existence of some $\xi \in (x - y, x)$ such that $f(x) - f(x - y) = f'(\xi) y$ as well as some $\zeta \in (0, y)$ such that $f(y) - f(0) = f'(\zeta) y$. As above, we have $f'(\zeta) > f'(\xi)$ yielding

$$f(x) - f(x - y) = f'(\xi) y \leq f'(\zeta) y = f(y) \Rightarrow f(x) - f(y) \leq f(x - y).$$

The assertion follows. □

Let $\psi \in (0, 1)$, then we have with Lemma A.1

(A.1) $\left|x^\psi - y^\psi\right| \leq |x - y|^\psi$ with $x, y \geq 0$.

Lemma A.2. *Let* X *be standard normally distributed and* $\varphi : \mathbb{R} \to \mathbb{R}$ *be continuous with* $\varphi(0) = 0$ *and* $|\varphi(x) - \varphi(y)| \leq |x - y|$*. Then,*

$$|\mathbb{E}[\varphi(X) X]| \leq 1.$$

Proof.

$$|\mathbb{E}[\varphi(X)X]| \leq \mathbb{E}[|\varphi(X) - \varphi(0)||X|] \leq \mathbb{E}[|X - 0||X|] = \mathbb{E}\left[X^2\right] = 1.$$

□

Lemma A.3. *Let $\varepsilon > 0$ and $\varphi : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ with $\varphi(x) \leq c \cdot x$, where $c > 0$ is a suitable constant. Then,*

$$(x + \varphi(x))^{1+\varepsilon} \leq x^{1+\varepsilon} + (1+\varepsilon)(1+c)^{\varepsilon} x^{\varepsilon}\varphi(x).$$

Proof. Let $f(y) = y^{1+\varepsilon}$ and, then it holds, with the Mean value Theorem and since $f'(y) = (1+\varepsilon)y^{\varepsilon}$ is monotonously increasing, for $0 < z < y$

$$\begin{aligned} f(y) - f(z) = f'(\xi)(y - z) \Leftrightarrow f(y) =& f(z) + f'(\xi)(y - z) \\ \leq& f(z) + f'(y)(y - z) \end{aligned}$$

for some $\xi \in (z, y)$.

Setting $y = x + \varphi(x)$ and $z = x$, this yields

$$\begin{aligned} (x + \varphi(x))^{1+\varepsilon} \leq& x^{1+\varepsilon} + (1+\varepsilon)(x + \varphi(x))^{\varepsilon}\varphi(x) \\ \leq& x^{1+\varepsilon} + (1+\varepsilon)(1+c)^{\varepsilon} x^{\varepsilon}\varphi(x). \end{aligned}$$

□

Lemma A.3 is essential for estimating the ratios of the standard deviations in the recursive estimate of Proposition 5.17 (and Proposition 6.12). Therefore, the lemma has to be applied in Corollaries 5.22, 5.23 and 5.24 (and in the analogous situations in Corollaries 6.21 and 6.22) with $\varepsilon = \frac{1}{2}$ as follows

(A.2) $(j\ln(j) + c_1 j)^{\frac{3}{2}} \leq (j\ln(j))^{\frac{3}{2}} + c_1' j^{\frac{3}{2}} (\ln(j))^{\frac{1}{2}}$,

(A.3) $\left(j + c_2 j^{2\lambda}\right)^{\frac{3}{2}} \leq j^{\frac{3}{2}} + c_2' j^{2\lambda + \frac{1}{2}}$,

(A.4) $(j + c_3)^{\frac{3}{2}} \leq j^{\frac{3}{2}} + c_3' j^{\frac{1}{2}}$,

for $j \in \mathbb{N}$ and suitable constants $c_1, c_2, c_3 > 0$.

A.2. Mean and Variance of the Number of Black Balls

Calculations concerning mean and variance of the number of black balls after n steps are sketched. For a deterministic replacement matrix, Savkevich [54] derived explicit formulas. On the way to showing asymptotic normality for such urns when the ratio of the eigenvalues is less or equal to $\frac{1}{2}$, Bagchi and Pal [3] study mean and variance as well. Their way seems to fit best to derive asymptotic expansions of mean and variance as needed for our proofs.

Throughout this chapter, B_n denotes the number of black balls in the urn after n steps and, to keep calculations readable, t_n denotes the total number of balls in the urn after n steps, $n \geq 0$.

As before, B_n^{b} will denote the number of black balls after n steps when starting with a black ball (hence, $B_0^{\mathrm{b}} = 1$), and B_n^{w} (here, $B_0^{\mathrm{w}} = 0$) when starting with a white ball.

The calculations that are to follow constitute the proofs of Lemma 5.1, Lemma 5.2, Lemma 6.1, Lemma 6.2 as well as Lemma 7.1.

To derive asymptotic expansions for both expectations and variances, the following fact about the ratio of Gamma functions is used, see Tricomi and Erdélyi [57]: For $\theta, \psi \in \mathbb{R}$, it holds, as $n \to \infty$,

$$\text{(A.5)} \quad \frac{\Gamma(n+\theta)}{\Gamma(n+\psi)} = n^{\theta-\psi}\left(1 + O\left(\frac{1}{n}\right)\right), \quad (n \to \infty).$$

Furthermore, we make use of

$$\text{(A.6)} \quad \sum_{j=1}^{n} \frac{1}{j} = \ln(n) + O(1), \qquad \text{as } n \to \infty.$$

Deterministic Replacement Matrix

Setting **Det R**, i.e., a Pólya urn scheme with two colours, balanced and irreducible, deterministic replacement matrix, makes the beginning. Note that in this case $t_n = t_0 + (a+b)\,n$, $n \geq 0$. The ratio of the eigenvalues is denoted by $\lambda := \frac{a-c}{a+b}$. We recollect some of the calculations performed by Bagchi and Pal in [3].

Mean

Bagchi and Pal conduct the following calculations in the proof of [3, Lemma 1] that lead to an exact formula for the mean:

$$\mathbb{E}\left[B_{n+1} | B_n\right] = \left(1 + \frac{a-c}{t_n}\right) B_n + c, \quad n \geq 0.$$

Substituting $D_n := B_n - \frac{c}{b+c} t_n$, $n \geq 0$, we have

$$\mathbb{E}\left[D_{n+1}\right] = \left(1 + \frac{a-c}{t_n}\right) \mathbb{E}\left[D_n\right], \quad n \geq 0,$$

yielding

$$\text{(A.7)} \quad \mathbb{E}\left[D_n\right] = \prod_{j=0}^{n-1} \left(1 + \frac{a-c}{t_j}\right) D_0 = \frac{\Gamma\left(\frac{t_0}{a+b}\right)}{\Gamma\left(\frac{a-c+t_0}{a+b}\right)} \frac{\Gamma\left(n + \frac{a-c+t_0}{a+b}\right)}{\Gamma\left(n + \frac{t_0}{a+b}\right)} D_0.$$

Plugging in $\lambda = \frac{a-c}{a+b}$, we obtain, for $n \geq 1$,

$$\begin{aligned} \mathbb{E}\left[B_n\right] &= \frac{\Gamma\left(\frac{t_0}{a+b}\right)}{\Gamma\left(\lambda + \frac{t_0}{a+b}\right)} \frac{\Gamma\left(n + \lambda + \frac{t_0}{a+b}\right)}{\Gamma\left(n + \frac{t_0}{a+b}\right)} \left(B_0 - \frac{c}{b+c} t_0\right) + \frac{c}{b+c} t_n \\ &= \frac{c\,(a+b)}{b+c} n + \left(B_0 - \frac{c}{b+c} t_0\right) \frac{\Gamma\left(\frac{t_0}{a+b}\right)}{\Gamma\left(\lambda + \frac{t_0}{a+b}\right)} \frac{\Gamma\left(n + \lambda + \frac{t_0}{a+b}\right)}{\Gamma\left(n + \frac{t_0}{a+b}\right)} + \frac{c t_0}{b+c}. \end{aligned}$$

This yields for B_n^{b} and B_n^{w}, $n \geq 1$:

$$\begin{aligned} \mathbb{E}\left[B_n^{\mathrm{b}}\right] &= \frac{c\,(a+b)}{b+c} n + \frac{b\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\lambda + \frac{1}{a+b}\right)} \frac{\Gamma\left(n + \lambda + \frac{1}{a+b}\right)}{\Gamma\left(n + \frac{1}{a+b}\right)} + \frac{c}{b+c} \quad \text{and} \\ \mathbb{E}\left[B_n^{\mathrm{w}}\right] &= \frac{c\,(a+b)}{b+c} n - \frac{c\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\lambda + \frac{1}{a+b}\right)} \frac{\Gamma\left(n + \lambda + \frac{1}{a+b}\right)}{\Gamma\left(n + \frac{1}{a+b}\right)} + \frac{c}{b+c}. \end{aligned}$$

The asymptotic expansions stated in Lemma 5.1 and Lemma 7.1 follow with (A.5).

<u>*Variance*</u>
We recall the proof of [3, Lemma 2.i)] and derive a formula for the variance of B_n. First, Bagchi and Pal observe

$$\text{(A.8)} \quad \operatorname{Var}\left(B_n\right) = \operatorname{Var}\left(D_n\right) = \mathbb{E}\left[D_n^2\right] - \mathbb{E}\left[D_n\right]^2$$

and, hence, derive $\mathbb{E}\left[D_n^2\right]$:

$$\begin{aligned} \mathbb{E}\left[D_{n+1}^2 \middle| D_n\right] &= \left(D_n + a - \frac{c\,(a+b)}{b+c}\right)^2 \left(\frac{c}{b+c} + \frac{D_n}{t_n}\right) \\ &\quad + \left(D_n + c - \frac{c\,(a+b)}{b+c}\right)^2 \left(\frac{b}{b+c} - \frac{D_n}{t_n}\right), \quad n \geq 0, \end{aligned}$$

leading to the following recurrence

$$\text{(A.9)}\quad \mathbb{E}\left[D_{n+1}^2\right] = \left(1+\frac{2(a-c)}{t_n}\right)\mathbb{E}\left[D_n^2\right] + \frac{b-c}{b+c}\frac{(a-c)^2}{t_n}\mathbb{E}\left[D_n\right] + \frac{bc(a-c)^2}{(b+c)^2},\ \ n \geq 0.$$

In the case $\lambda < \frac{1}{2}$, the homogeneous equation

$$\mathbb{E}\left[D_{n+1}^2\right] - \left(1+\frac{2(a-c)}{t_n}\right)\mathbb{E}\left[D_n^2\right] = 0$$

has a solution of the form

$$\begin{aligned} H(0) &= 1, \\ H(n) &= \prod_{j=0}^{n-1}\left(1+\frac{2(a-c)}{t_j}\right) = \frac{\Gamma\left(\frac{t_0}{a+b}\right)}{\Gamma\left(\frac{2(a-c)+t_0}{a+b}\right)}\frac{\Gamma\left(n+\frac{2(a-c)+t_0}{a+b}\right)}{\Gamma\left(n+\frac{t_0}{a+b}\right)}, \quad n \geq 1. \end{aligned}$$

Bagchi and Pal find as particular solution

$$P(n) = \frac{bc}{a+b-2(a-c)}\left(\frac{a-c}{b+c}\right)^2 t_n - \frac{b-c}{b+c}(a-c)\,\mathbb{E}\left[D_n\right], \quad n \geq 0.$$

Hence, the complete solution of the above recurrence is given by

$$\mathbb{E}\left[D_n^2\right] = kH(n) + P(n)$$

with k determined by the initial condition $D_0^2 = \left(B_0 - \frac{c}{b+c}t_0\right)^2$. Hence,

$$\operatorname{Var}(B_n) = kH(n) + P(n) - \left(\mathbb{E}\left[D_n\right]\right)^2,$$

where $H(n)$ contributes terms of order $O\left(n^{\frac{2(a-c)}{a+b}}\right)$, $P(n)$ contributes terms of linear order in the shape of t_n and of order $n^{\frac{a-c}{a+b}}$ in the form of $\mathbb{E}[D_n]$, see (A.7), and, finally, $(\mathbb{E}[D_n])^2$ contributes terms of the same order as $H(n)$ (that do not cancel out). Note that $t_n = t_0 + n(a+b)$; therefore, the initial number of balls only contributes in the form of $O(1)$ to the variance and does not influence the constant that accompanies the linear order. Plugging in $\lambda = \frac{a-c}{a+b}$, this yields with (A.5), as $n \to \infty$,

$$\text{(A.10)}\quad 0 < \lambda < \tfrac{1}{2}: \quad \operatorname{Var}(B_n) = \frac{bc}{a+b-2(a-c)}\left(\frac{a-c}{b+c}\right)^2 n + O\left(n^{2\lambda}\right),$$

$$\text{(A.11)}\quad \lambda < 0: \quad \operatorname{Var}(B_n) = \frac{bc}{a+b-2(a-c)}\left(\frac{a-c}{b+c}\right)^2 n + O(1).$$

If $\lambda = \frac{1}{2}$, we follow the proof of [3, Lemma 2.ii)]. Things look a bit different here, logarithmic terms enter: The recurrence (A.9) becomes

$$\mathbb{E}\left[D_{n+1}^2\right] = \frac{t_{n+1}}{t_n}\mathbb{E}\left[D_n^2\right] + \frac{b^2 - c^2}{t_n}\mathbb{E}\left[D_n\right] + bc,$$

since $2\left(a - c\right) = a + b \Leftrightarrow a - c = b + c$.

The homogeneous equation

$$\mathbb{E}\left[D_{n+1}^2\right] - \frac{t_{n+1}}{t_n}\mathbb{E}\left[D_n\right] = 0$$

has a solution

$$\hat{H}\left(n\right) = t_n, \quad n \geq 0,$$

and Bagchi and Pal find as particular solution via substitution with $\hat{d} = \frac{D_0^2 + (b-c)D_0}{t_0}$

$$\hat{P}\left(n\right) = \left(c - b\right)\mathbb{E}\left[D_n\right] + \hat{d}t_n + t_n bc\sum_{j=1}^{n}\frac{1}{t_j}, \quad n \geq 1.$$

Thus, a complete solution is given by

$$\mathbb{E}\left[D_n^2\right] = \hat{k}\hat{H}\left(n\right) + \hat{P}\left(n\right), \quad n \geq 1,$$

with $\hat{k}$ given by initial conditions $\hat{k} = \frac{\mathbb{E}\left[D_1^2\right] + \hat{P}(1)}{\hat{H}(1)}$. Hence, the variance of the number of black balls is given by

$$\mathrm{Var}\left(B_n\right) = \hat{k}\hat{H}\left(n\right) + \hat{P}\left(n\right) - \left(\mathbb{E}\left[D_n\right]\right)^2,$$

where $\hat{P}\left(n\right)$ contributes terms of order $n\ln\left(n\right)$ as well as linear and logarithmic terms, $\hat{H}(n)$ and $\left(\mathbb{E}\left[D_n\right]\right)^2$ contribute a linear term. The leading order is $n\ln\left(n\right)$ and the initial number of balls t_0 hidden in t_n contributes solely to terms of lower order. Therefore, it holds with (A.6) and (A.5), as $n \to \infty$,

$$\text{(A.12)} \quad \lambda = \frac{1}{2}: \quad \mathrm{Var}\left(B_n\right) = bc\,n\ln\left(n\right) + O\left(n\right).$$

(A.10), (A.11) and (A.12) lead to the statements of Lemma 5.2 and Lemma 7.1.

Randomised Play-the-Winner-Rule

To derive mean and variance for the Pólya urn scheme characterised by setting **Rand R**, we borrow the reasoning of Bagchi and Pal [3] displayed above and adapt it to this Pólya urn scheme. Note that now $t_n = t_0 + n$ and the ratio of the eigenvalues is given by $\lambda := \alpha + \beta - 1$. We stick to the notation of the last paragraph.

Mean
We have

$$\mathbb{E}\left[B_{n+1}|B_n\right] = \left(1 + \frac{\alpha+\beta-1}{t_n}\right) B_n + (1-\beta)\,.$$

We now substitute

$$D_n := B_n - \frac{1-\beta}{2-\alpha-\beta} t_n, \quad n \geq 0,$$

and have

$$\mathbb{E}\left[D_{n+1}\right] = \left(1 + \frac{\alpha+\beta-1}{t_n}\right) \mathbb{E}\left[D_n\right]$$

yielding

$$\text{(A.13)}\quad \begin{aligned} \mathbb{E}\left[D_n\right] &= \prod_{j=0}^{n-1} \left(1 + \frac{\alpha+\beta-1}{t_j}\right) D_0 \\ &= \frac{\Gamma(t_0)}{\Gamma(t_0+\alpha+\beta-1)} \frac{\Gamma(n+\alpha+\beta-1+t_0)}{\Gamma(n+t_0)} D_0. \end{aligned}$$

Plugging in $\lambda = \alpha + \beta - 1$, it holds, $n \geq 1$,

$$\begin{aligned} \mathbb{E}\left[B_n\right] &= \frac{\Gamma(t_0)}{\Gamma(t_0+\lambda)} \frac{\Gamma(n+\lambda+t_0)}{\Gamma(n+t_0)} \left(B_0 - \frac{1-\beta}{1-\lambda} t_0\right) + \frac{1-\beta}{1-\lambda} t_n \\ &= \frac{1-\beta}{1-\lambda} n + \frac{\Gamma(t_0)}{\Gamma(t_0+\lambda)} \frac{\Gamma(n+\lambda+t_0)}{\Gamma(n+t_0)} \left(B_0 - \frac{1-\beta}{1-\lambda} t_0\right) + \frac{1-\beta}{1-\lambda} t_0. \end{aligned}$$

This yields for B_n^{b} and B_n^{w}, $n \geq 1$:

$$\begin{aligned} \mathbb{E}\left[B_n^{\mathrm{b}}\right] &= \frac{1-\beta}{1-\lambda} n + \frac{1-\alpha}{1-\lambda} \frac{1}{\Gamma(\lambda+1)} \frac{\Gamma(n+1+\lambda)}{\Gamma(n+1)} + \frac{1-\beta}{1-\lambda} \quad \text{and} \\ \mathbb{E}\left[B_n^{\mathrm{w}}\right] &= \frac{1-\beta}{1-\lambda} n - \frac{1-\beta}{1-\lambda} \frac{1}{\Gamma(\lambda+1)} \frac{\Gamma(n+1+\lambda)}{\Gamma(n+1)} + \frac{1-\beta}{1-\lambda}. \end{aligned}$$

The asymptotic expansions stated in Lemma 6.1 follow with (A.5).

<u>*Variance*</u>
For the variance it holds

$$\text{(A.14)}\quad \operatorname{Var}(B_n) = \operatorname{Var}(D_n) = \mathbb{E}\left[D_n^2\right] - \left(\mathbb{E}[D_n]\right)^2.$$

We obtain

$$\begin{aligned}\mathbb{E}\left[D_{n+1}^2 \middle| D_n\right] &= \left(D_n + \frac{1-\alpha}{2-\alpha-\beta}\right)^2 \left((\alpha+\beta-1)\frac{D_n}{t_n} + \frac{1-\beta}{2-\alpha-\beta}\right) \\ &\quad + \left(D_n - \frac{1-\beta}{2-\alpha-\beta}\right)^2 \left(-(\alpha+\beta-1)\frac{D_n}{t_n} + \frac{1-\alpha}{2-\alpha-\beta}\right)\end{aligned}$$

yielding the recurrence, with $\lambda = \alpha+\beta-1$,

$$\text{(A.15)}\quad \mathbb{E}\left[D_{n+1}^2\right] = \left(1+\frac{2\lambda}{t_n}\right)\mathbb{E}\left[D_n^2\right] + \frac{\lambda(\beta-\alpha)}{(1-\lambda)\,t_n}\mathbb{E}[D_n] + \frac{(1-\alpha)(1-\beta)}{(1-\lambda)^2}.$$

In the case $\lambda < \frac{1}{2}$, we proceed with the homogeneous equation

$$\mathbb{E}\left[D_{n+1}^2\right] - \left(1+\frac{2(\alpha+\beta-1)}{t_n}\right)\mathbb{E}\left[D_n^2\right] = 0$$

that has a solution

$$\begin{aligned}\tilde{H}(0) &:= 1, \\ \tilde{H}(n) &:= \prod_{j=0}^{n-1}\left(1+\frac{2(\alpha+\beta-1)}{t_j}\right) = \frac{\Gamma(t_0)}{\Gamma(2\lambda+t_0)}\frac{\Gamma(n+2\lambda+t_0)}{\Gamma(n+t_0)}, \quad n \geq 1.\end{aligned}$$

A particular solution is given by

$$\tilde{P}(n) := \frac{(1-\alpha)(1-\beta)}{1-2(\alpha+\beta-1)}\frac{t_n}{(2-\alpha-\beta)^2} - \frac{\beta-\alpha}{2-\alpha-\beta}\mathbb{E}[D_n], \quad n \geq 0.$$

Hence,

$$\text{(A.16)}\quad \mathbb{E}\left[D_n^2\right] = \tilde{k}\tilde{H}(n) + \tilde{P}(n), \quad n \geq 1.$$

Therefore, the variance of B_n is given by

$$\operatorname{Var}(B_n) = \tilde{k}\tilde{H}(n) + \tilde{P}(n) - \left(\mathbb{E}[D_n]\right)^2,$$

where $\tilde{P}(n)$ contributes terms of linear order in the shape of t_n as well as of order n^{λ} in the shape of $\mathbb{E}[D_n]$ and $\tilde{H}(n)$ and $(\mathbb{E}[D_n])^2$ contribute a term of order $n^{2\lambda}$. The initial number of balls t_0 only contributes of constant order. Therefore, it holds with (A.5), as $n \to \infty$,

$$\text{(A.17)} \quad 0 < \lambda < \frac{1}{2}: \qquad \operatorname{Var}(B_n) = \frac{(1-\alpha)(1-\beta)}{(1-2\lambda)(1-\lambda)^2} n + O\left(n^{2\lambda}\right),$$

$$\text{(A.18)} \quad \lambda < 0: \qquad \operatorname{Var}(B_n) = \frac{(1-\alpha)(1-\beta)}{(1-2\lambda)(1-\lambda)^2} n + O(1).$$

When $\lambda = \frac{1}{2}$, again the recurrence (A.15) simplifies to:

$$\text{(A.19)} \quad \mathbb{E}\left[D_{n+1}^2\right] = \left(1 + \frac{1}{t_n}\right)\mathbb{E}\left[D_n^2\right] + \frac{\beta-\alpha}{t_n}\mathbb{E}\left[D_n\right] + 4(1-\alpha)(1-\beta).$$

The corresponding homogeneous equation

$$\mathbb{E}\left[D_{n+1}^2\right] - \left(1 + \frac{1}{t_n}\right)\mathbb{E}\left[D_n^2\right] = 0$$

has a solution of the form

$$\bar{H}(n) = t_n, \quad n \geq 0.$$

In order to derive a particular solution we try, in accordance with the approach of Bagchi and Pal as in [3, Lemma 2.ii)], the substitution

$$\bar{P}(n) = t_n \bar{g}(n) + 2(\alpha-\beta)\mathbb{E}\left[D_n\right], \quad n \geq 0,$$

and obtain from (A.19) the recurrence

$$\bar{g}(n+1) = \bar{g}(n) + \frac{4(1-\alpha)(1-\beta)}{t_{n+1}} = \bar{g}(0) + 4(1-\alpha)(1-\beta)\sum_{j=1}^{n+1}\frac{1}{t_j}$$

with $\bar{g}(0) = \frac{D_0^2 - 2(\alpha-\beta)D_0}{t_0}$. Hence, we obtain for $n \geq 1$

$$\begin{aligned} \text{(A.20)} \quad \mathbb{E}\left[D_n^2\right] &= \bar{k}\bar{H}(n) + \bar{P}(n) \\ &= \bar{k}t_n + t_n\left(\bar{g}(0) + 4(1-\alpha)(1-\beta)\sum_{j=1}^{n}\right) 2(\alpha-\beta)\mathbb{E}\left[D_n\right] \\ &= \bar{k}t_n + \bar{g}(0)t_n + 4(1-\alpha)(1-\beta)t_n\sum_{j=1}^{n}\frac{1}{t_j} + 2(\alpha-\beta)\mathbb{E}\left[D_n\right], \end{aligned}$$

where $\bar{k}$ is to be determined by the initial condition. With

$$\operatorname{Var}(B_n) = \mathbb{E}\left[D_n^2\right] - \left(\mathbb{E}\left[D_n\right]\right)^2$$

the variance is composed as follows: With (A.20) the leading term is of order $n \ln(n)$ followed by linear terms (from (A.20) and $(\mathbb{E}[D_n])^2$). The initial number of balls t_0 only contributes of constant order. Hence, it holds with (A.6) and (A.5), as $n \to \infty$,

$$\text{(A.21)} \quad \operatorname{Var}(B_n) = 4(1-\alpha)(1-\beta)\, n \ln(n) + O(n) .$$

(A.17), (A.18) and (A.21) lead to the statement of Lemma 6.2.

B. R Code for Simulations

B.1. Code

```

############################################################
### Simulation of Kolmogorov-Smirnov distance between normalised number of black balls and standard normal distribution
############################################################

############################################################
polyaurn <- function(ibl, iwh, a,b,c,d, n, m){
# ibl: initial number of black balls
# iwh: initial number of white balls
# a, b: 1st row of replacement matrix
# c, d: 2nd row of replacement matrix
# n: number of steps considered
# m: number of samples/urns

lambda <- (a-c)/(a+b)
if (lambda > 0.5){
print("This Polya urn does not belong to the normal limit case. Lambda follows:")
return(lambda)
}
if (lambda==0){
print("Deterministic evolution, no rate to observe. Lambda follows")
return(lambda)
}
if (a+b != c+d){
```

```
print("Not a balanced scheme. Row sums follow:")
return(c(a+b,c+d))
}

time <- seq(1,n)
null <- rep(0,n)

blb <- rep(ibl, m) #number of black balls
total <- ibl + iwh #total number of balls in the urn
meanbl <- ibl # mean of number of black balls
meantemp <- ibl - ((c/(b+c)) * total) # quantity for
    establishing recursion for mean

weight <- blb/total #vector of length m carrying proportion
     of black balls

kolmo <- rep(0,n) #carries Kolmogorov-Smirnov distance at
    times 1 to n

z <- seq(-2,2, by=0.01)
temp2 <- pnorm(z) #distribution function of standard normal
     distribution evaluated at z

for (i in 1:n){
##########################################################
# generating blb
##########################################################
meantemp <- ( 1 + ((a-c)/total) ) * meantemp

temp <- runif(m) # m samples of uniform distribution on
    [0,1]
result <- temp < weight
for (j in 1:m){
if (result[j]){# black ball was drawn
blb[j] <- blb[j] + a #number of black balls increases by a
}
else{# white ball was drawn
blb[j] <- blb[j] + c #number of black balls increases by c
```

```
}
}

#############################################################
# mean and standard deviation according to formulas derived in thesis
#############################################################

# mean of number of black balls
meanbl <- meantemp + ((c/(b+c)) * total) #mean of number of black balls
#############################################################

# (rough) standard deviation of the number of black balls
if (lambda == 0.5) {
if (i==1){
devblack <- c(1)
}
if (i > 1){
devblack <-  sqrt((b * c * i * log(i)) )
}
}
if (lambda < 0.5) {
devblack <- sqrt( ((a+b)*(a-c)^2* b * c /((a+b-2*(a-c))*(b+c)^2)) * i) # + i^lambda
}

#############################################################
# normalising
#############################################################
normblb <- (blb-meanbl)/devblack # (roughly) normalised number of black balls

#############################################################
# empirical distribution function
#############################################################
tempecdf <- ecdf(normblb) #empricial distribution function of normalised number of black balls
```

```
temp1 <- tempecdf(z) #tempecdf evaluated at z

# Kolmogorov-Smirnov distance
diff <- abs(temp1 - temp2)
kolmo[i] <- max(diff)

##########################################################
# update total and weight
##########################################################
total <- total + a + b
weight <- blb/total
}

##########################################################
# plot rate and compare with theoretical rate from thesis,
# determination of the constant on the basis of the sampled
    data
##########################################################
# ideal: rate expected as consequence of Zolotarev distance
# gamma: negative exponent of the rate as consequence of
   Zolotarev distance
# ideal2: another plausible rate, serves to compare
   simulated exponent
# gamma2: exponent of rate used for comparison
# constant: contains ratio of rate observed in sample to
   expected rate
# choice: constant chosen for plotting

if (lambda == 0.5){
gamma <- 1.5
gamma2 <- 0.5
time1 <- seq(2,n)
ideal <- (log(time1))^(-gamma)
ideal2 <- (log(time1))^(-gamma2)
ideal <- c(1,ideal)
ideal2 <- c(1,ideal2)
}
else {
```

```
if ( (1/3) < lambda && lambda < 0.5 ){
gamma <- -(3 * (lambda - 0.5))
ideal <- time^(-gamma)
gamma2 <- 0.5 - lambda
ideal2 <- time^(-gamma)
}
if (lambda <= (1/3)){
gamma <- 0.5
ideal <- 1/sqrt(time)
}
}

constant <- kolmo/ideal
minconst <- min(constant)
maxconst <- max(constant)
##########################################################
## different choices for constant of ideal rate
##########################################################
#choice <- round(max(constant))+1
#choice <- mean(constant)
#choice <- median(constant)
#choice <- max(constant[n-100:n])
#choice <- mean(constant[n-100:n])
choice <- (1+0.2)*median(constant[n-100:n])

rate <- choice * ideal

##########################################################
##plot of the empirical distribution function of the
    normalised number of black balls in comparison the the
    distribution function of the standard normal distribution

quartz("(Empirical) Distribution Function")
plot(z, temp1
, type = "l", lty = 1, lwd = 3
, col="mediumturquoise"
, xlab = "x", ylab = "Empirical distribution function"
#, cex.lab=0.75, cex.axis=0.75
```

```
)
lines(z,temp2,col="mediumvioletred", lwd = 3)
###########################################################

###########################################################
##plot displaying the rate observed in sample in comparison
    to expected rate

quartz("Rate Compared to Theoretical Bounds")
plot(time, kolmo
, type = "l", lty = 1, lwd = 3
, col="mediumturquoise"
, ylim = c(-0.02,0.2), ylab="Kolmogorov-Smirnov distance"
, xlab="Steps"
#, cex.lab=0.75, cex.axis=0.75
)
lines(time,rate,col="mediumvioletred", lty = 4, lwd = 3)
lines(time,null,col="grey")

if (lambda > 0 && lambda <= 0.5
#lambda == 0.5
){
constant2 <- kolmo/ideal2
minconst2 <- min(constant2)
maxconst2 <- max(constant2)
#choice2 <- mean(kolmo/ideal2)
#choice2 <- median(constant2[n-100:n])
choice2 <- (1+0.2)*median(constant2)

rate2 <- choice2 * ideal2
lines(time,rate2,col="royalblue", lty = 3, lwd = 3)
}

###########################################################

###########################################################
## log log plot of constant

```

```
if (lambda == 0.5){
quartz("Log Log Plot")
plot(log(time),(log(kolmo)-log((2 * choice/(1+0.2)))-2)
, type = "l"
, lty = 1
, lwd = 3
, xlab = "logarithmic time scale", ylab = "log of rates"
, col="mediumturquoise"
#, cex.lab=0.75, cex.axis=0.75
)
}
else{
quartz("Log Log Plot")
plot(log(time),(log(kolmo)-log((choice/(1+0.2))))
, type = "l", lty = 1, lwd = 3
, xlab = "logarithmic time scale", ylab = "log of rates"
, col="mediumturquoise"
#, cex.lab=0.75, cex.axis=0.75
)
lines(log(time),log(time^(-gamma)), col="mediumvioletred",
    lwd = 3)
if (lambda >0 && lambda < 0.5){
lines(log(time), log(time^(-gamma2)), col="royalblue", lwd
    = 3)
}
}

##############################################################

##############################################################
return(c(lambda,gamma, min(kolmo), max(kolmo[100:n])))
##############################################################
}
```

B.2. Input for Simulations in Chapter 1

```
> polyaurn(1,0,1,4,3,2,10000,100000)
[1] -0.40000000  0.50000000  0.01375639  0.16611372
```

```
> polyaurn(1,0,20,10,9,21,10000,100000)
[1] 0.3666667 0.4000000 0.0162350 0.1052983
```

B.3. Additional Simulations

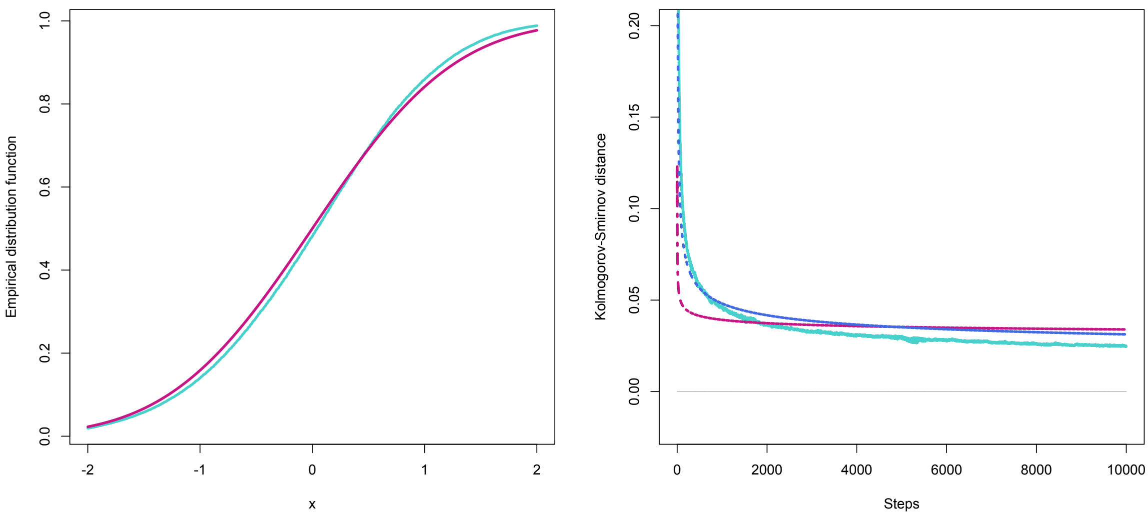
In the following, results of further simulations are displayed.

At first, two simulations for an urn with logarithmic rate are given. Of course, these are to be treated with utmost caution since logarithmic rates are not easily simulated due to the “very slow” decay.

Afterwards, two more simulations follow that are similar to the simulations presented in the Introduction, Chapter 1, but with interchanged values for steps and samples.

Additional simulation of a Pólya urn that leads to logarithmic rates. The simulation performs 10^4 steps on the basis of 10^5 samples.

```
> polyaurn(1,0,5,1,2,4,10000,100000)
[1] 0.50000000 1.50000000 0.02609695 0.11507863
```



(a) Empirical Distribution Function

(b) Rate of Convergence

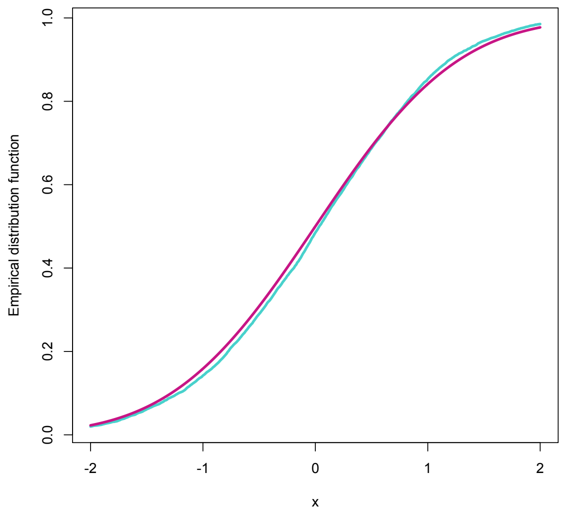
Figure B.1.: Simulation of 10^4 steps on the basis of 10^5 samples of a Pólya urn with replacement matrix $\left(\begin{smallmatrix} 5 & 1 \\ 2 & 4 \end{smallmatrix}\right)$, hence $\lambda = \frac{1}{2}$, with one initial black ball.

Figure B.1a shows the empirical distribution function of the normalised number of black balls (turquoise) compared to the distribution function of the standard normal distribution (magenta).

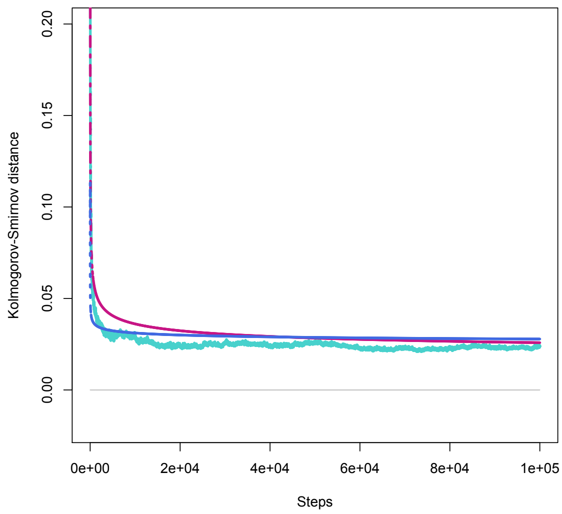
Figure B.1b shows the uniform distance between the empirical distribution function and the distribution function of the standard normal distribution (turquoise), i.e., a simulation of the Kolmogorov-Smirnov distance between the normalised number of black balls and the standard normal distribution, compared to a rate of order $(\ln(n))^{-\frac{3}{2}}$ (magenta) and to a rate of order $(\ln(n))^{-\frac{1}{2}}$ (royal blue).

Additional simulation of a Pólya urn that leads to logarithmic rates. The simulation performs 10^5 steps on the basis of 10^4 samples.

```
> polyaurn(1,0,5,1,2,4,100000,10000)
[1] 0.50000000 1.50000000 0.01835207 0.11501863
```



(a) Empirical Distribution Function

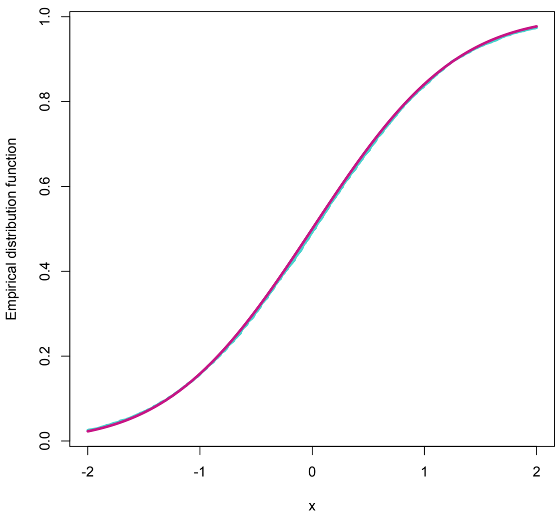


(b) Rate of Convergence

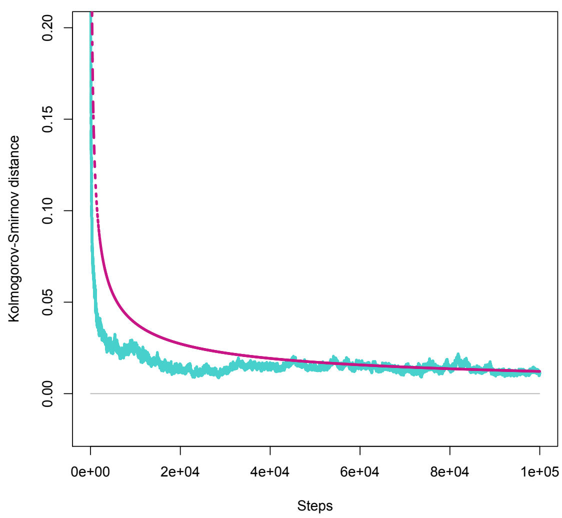
Figure B.2.: Simulation of 10^5 steps on the basis of 10^4 samples of a Pólya urn with replacement matrix $\left(\begin{smallmatrix} 5 & 1 \\ 2 & 4 \end{smallmatrix}\right)$, hence $\lambda = \frac{1}{2}$, with one initial black ball.
For explanations, see Figure B.1.

Additional simulation, similar to simulation for Figure 1.2 and 1.3, but 10^5 steps on the basis of 10^4 samples. See Figures 1.2 and 1.3 for explanations.

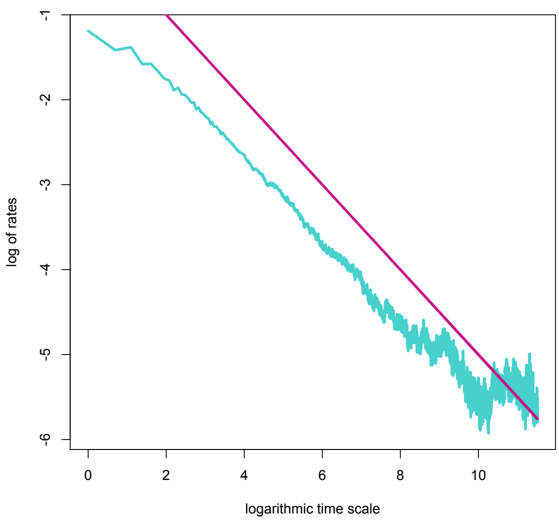
```
> polyaurn(1,0,1,4,3,2,100000,10000)
[1] -0.400000000  0.500000000  0.007982308
   0.175223716
```



(a) Empirical Distribution Function



(b) Rate of Convergence

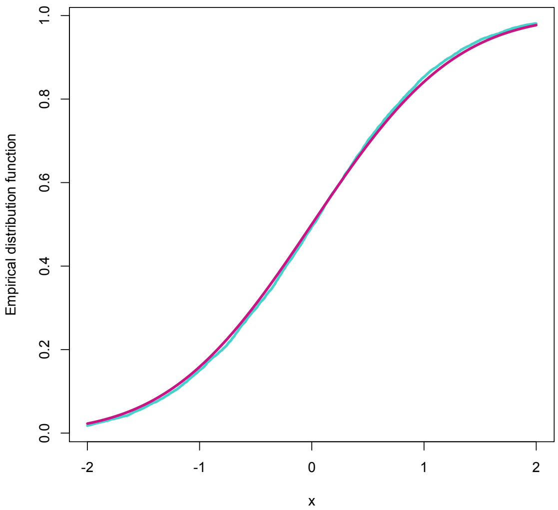


(c) Log log plot

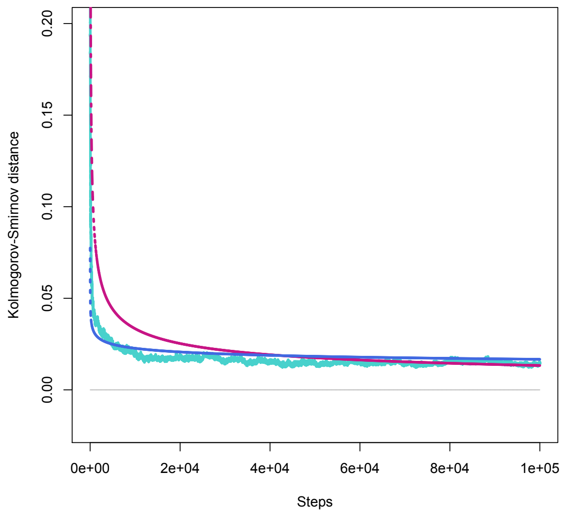
Figure B.3.: Simulation of 10^5 steps on the basis of 10^4 samples of a Pólya urn with replacement matrix $\left(\begin{smallmatrix}1 & 4\\ 3 & 2\end{smallmatrix}\right)$, hence $\lambda = -\frac{2}{5}$, with one initial black ball.

Additional simulation, similar to simulation for Figure 1.4 and 1.5, but 10^5 steps on the basis of 10^4 samples. See Figures 1.4 and 1.5 for explanations.

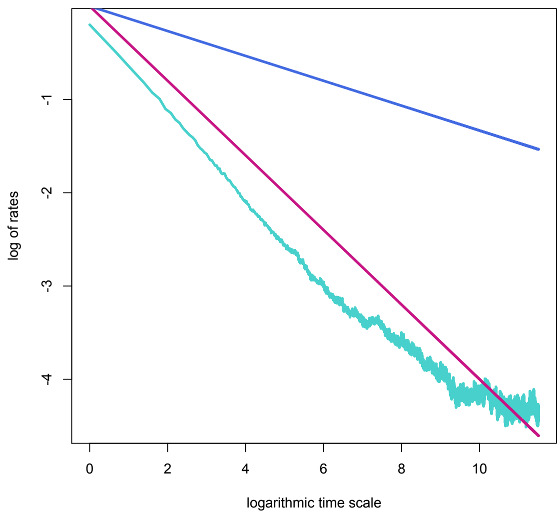
```
> polyaurn(1,0,20,10,9,21,100000,10000)
[1] 0.36666667 0.40000000 0.01032573 0.10382357
```



(a) Empirical Distribution Function



(b) Rate of Convergence



(c) Log log plot

Figure B.4.: Simulation of 10^5 steps on the basis of 10^4 samples of a Pólya urn with replacement matrix $\begin{pmatrix} 20 & 10 \\ 9 & 21 \end{pmatrix}$, hence $\lambda = \frac{11}{30}$, with one initial black ball.

Bibliography

[1] Athreya, K. B. On a Characteristic Property of Polya's Urn. *Studia Scientiarum Mathematicarum Hungarica* **4** (1969), pp. 31–35.

[2] Athreya, K. B. and Karlin, S. Embedding of Urn Schemes into Continuous Time Markov Branching Processes and Related Limit Theorems. *The Annals of Mathematical Statistics* **39**(6) (1968), pp. 1801–1817.

[3] Bagchi, A. and Pal, A. K. Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM Journal on Algebraic Discrete Methods* **6** (1985), pp. 394–405.

[4] Bernstein, S. Nouvelles applications des grandeurs aléatoires presqu'indépendantes. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya* **4**(2) (1940), pp. 137–150.

[5] Bernstein, S. Sur un probleme du schéma des urnes à composition variable (in Russian). *Comptes Rendus de l'Académie des Sciences de l'URSS* **28** (1940), pp. 5–7.

[6] Bickel, P. J. and Freedman, D. A. Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9**(6) (1981), pp. 1196–1217.

[7] Blackwell, D. and Kendall, D. The Martin Boundary for Pólya's Urn Scheme, and an Application to Stochastic Population Growth. *Journal of Applied Probability* **1**(2) (1964), pp. 284–296.

[8] Cambanis, S., Simons, G., and Stout, W. Inequalities for Ek(X, Y) when the marginals are fixed. *Probability Theory and Related Fields* **36**(4) (1976), pp. 285–294.

[9] Chauvin, B., Mailler, C., and Pouyanne, N. Smoothing equations for large Pólya urns. *Journal of Theoretical Probability* **28**(3) (2015), pp. 923–957.

[10] Chow, Y. S. and Teicher, H. *Probability theory: independence, interchangeability, martingales.* third. Springer Texts in Statistics. Springer-Verlag, New York, 1997.

[11] Cramer, M and Rüschendorf, L. Analysis of recursive algorithms by the contraction method. *Athens Conference on Applied Probability and Time Series Analysis* (1996), pp. 18–33.

[12] Cramer, M. Stochastische Analyse rekursiver Algorithmen mit idealen Metriken. PhD thesis. Albert-Ludwigs-Universität Freiburg i. Br., 1995.

[13] Drmota, M., Janson, S., and Neininger, R. A functional limit theorem for the profile of search trees. *The Annals of Applied Probability* **18**(1) (2008), pp. 288–333.

[14] Eggenberger, F. and Pólya, G. Über die Statistik verketteter Vorgänge. *Zeitschrift für Angewandte Mathematik und Mechanik* **3** (1923), pp. 279–289.

[15] Fill, J. A. and Janson, S. Quicksort Asymptotics. *Journal of Algorithms* **44** (2002), pp. 4–28.

[16] Flajolet, P., Gabarró, J., and Pekari, H. Analytic urns. *The Annals of Probability* **33**(3) (2005), pp. 1200–1233.

[17] Freedman, D. A. Bernard Friedman's urn. *The Annals of Mathematical Statistics* **36**(3) (1965), pp. 956–970.

[18] Friedman, B. A simple urn model. *Communications on Pure and Applied Mathematics* **2**(1) (1949), pp. 59–70.

[19] Goldstein, L. and Reinert, G. Stein's method for the Beta distribution and the Polya-Eggenberger Urn. *Journal of Applied Probability* **50**(4) (2013), pp. 1187–1205.

[20] Gouet, R. Martingale functional central limit theorems for a generalized Pólya urn. *The Annals of Probability* **21**(3) (1993), pp. 1624–1639.

[21] Hwang, H.-K. Second Phase Changes in Random m-ary Search Trees and Generalized Quicksort: Convergence Rates. *Annals of Probability* **31**(2) (2003), pp. 609–629.

[22] Janson, S. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications* **110**(2) (2004), pp. 177–245.

[23] Janson, S. Limit theorems for triangular urn schemes. *Probability Theory and Related Fields* **134**(3) (2006), pp. 417–452.

[24] Janson, S. and Neininger, R. The size of random fragmentation trees. *Probability Theory and Related Fields* **142** (2008), pp. 399–442.

[25] Johnson, N. L. and Kotz, S. *Urn models and their application. An approach to modern discrete probability theory.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-London-Sydney, 1977.

[26] Klenke, A. *Probability Theory. A Comprehensive Course.* Second edition. Universitext. Springer-Verlag London, 2014.

[27] Knape, M. and Neininger, R. Pólya urns via the contraction method. *Combinatorics, Probability and Computing* **23**(6) (2014), pp. 1148–1186.

[28] Kuba, M. and Sulzbach, H. On martingale tail sums in affine two-color urn models with multiple drawings. *Journal of Applied Probability* **54**(1) (2017), pp. 96–117.

[29] Kuntschik, A. Konvergenzraten für Pólya-Urnen (in German). MA thesis. `http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/44186`: Johann Wolfgang Goethe-Universität, Frankfurt am Main, 2014.

[30] Kuntschik, A. and Neininger, R. Rates of convergence for balanced irreducible two-color Pólya urns. *2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)* (2017), pp. 94–99.

[31] Leckey, K., Neininger, R., and Sulzbach, H. Process convergence for the complexity of Radix Selection on Markov sources. *arXiv preprint arXiv:1605.02352* (May 2016).

[32] Leckey, K., Neininger, R., and Szpankowski, W. A Limit Theorem for Radix Sort and Tries with Markovian Input. *arXiv preprint arXiv:1505.07321* (May 2015).

[33] Leckey, K. On Densities for Solutions to Stochastic Fixed Point Equations. *arXiv preprint arXiv:1604.05787* (Apr. 2016).

[34] Mahmoud, H. *Pólya urn models*. Texts in Statistical Science Series. CRC press, 2008.

[35] Mahmoud, H. and Neininger, R. Distribution of distances in random binary search trees. *The Annals of Applied Probability* **13**(1) (2003), pp. 253–276.

[36] Major, P. On the Invariance Principle for Sums of Independent Identically Distributed Random Variables. *Journal of Multivariate Analysis* **8**(4) (1978), pp. 487–517.

[37] Markov, A. A. Sur quelques formules limites du calcul des probabilités (in Russian). *Bulletin de l'Académie des Sciences* **11**(3) (1917), pp. 177–186.

[38] Markov, A. A. Extension of the law of large numbers to quantities, depending on each other (1906). Reprint. (in Russian). eng. *Journal Électronique d'Histoire des Probabilités et de la Statistique [electronic only]* **2**(1b) (2006), Article 10, 12 p. url: `http://eudml.org/doc/128778`.

[39] Neininger, R. On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures & Algorithms* **19**(3-4) (2001), pp. 498–524.

[40] Neininger, R. and Rüschendorf, L. Rates of convergence for Quicksort. *Journal of Algorithms* **44**(1) (2002), pp. 52–62.

[41] Neininger, R. and Rüschendorf, L. A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability* **14**(1) (2004), pp. 378–418.

[42] Neininger, R. and Rüschendorf, L. On the contraction method with degenerate limit equation. *The Annals of Probability* **32**(3B) (2004), pp. 2838–2856.

[43] Neininger, R. and Sulzbach, H. On a functional contraction method. *The Annals of Probability* **43**(4) (2015), pp. 1777–1822.

[44] Panholzer, A. The Distribution of the Size of the Ancestor-Tree and of the Induced Spanning Subtree for Random Trees. *Random Structures & Algorithms* **25**(2) (2004), pp. 179–207.

[45] Peköz, E. A., Röllin, A., and Ross, N. Generalized Gamma Approximation with Rates for Urns, Walks and Trees. *The Annals of Probability* **44**(3) (2016), pp. 1776–1816.

[46] Pemantle, R. A survey of random processes with reinforcement. *Probability Surveys* **4** (2007), pp. 1–79.

[47] Pouyanne, N. An algebraic approach to Pólya processes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **44**(2) (2008), pp. 293–323.

[48] Rachev, S. T. and Rüschendorf, L. Probability metrics and recursive algorithms. *Advances in Applied Probability* **27**(3) (1995), pp. 770–799.

[49] Rachev, S. T. *Probability Metrics and the Stability of Stochastic Models*. Vol. 269. John Wiley & Son Ltd, 1991.

[50] Rösler, U. A limit theorem for "Quicksort". *Informatique théorique et applications* **25**(1) (1991), pp. 85–100.

[51] Rösler, U. A fixed point theorem for distributions. *Stochastic Processes and their Applications* **42**(2) (1992), pp. 195–214.

[52] Rösler, U. On the analysis of stochastic divide and conquer algorithms. *Algorithmica* **29**(1) (2001), pp. 238–261.

[53] Rösler, U. and Rüschendorf, L. The contraction method for recursive algorithms. *Algorithmica* **29**(1) (2001), pp. 3–33.

[54] Savkevich, V. Sur le schéma des urnes à composition variable (in Russian). *Comptes Rendus de l'Académie des Sciences de l'URSS* **28** (1940), pp. 8–12.

[55] Smythe, R. T. Central limit theorems for urn models. *Stochastic Processes and their Applications* **65**(1) (1996), pp. 115–137.

[56] Smythe, R. T. and Rosenberger, W. F. Play-the-Winner Designs, Generalized Pólya Urns, and Markov Branching Processes. *Lecture Notes-Monograph Series* (1995), pp. 13–22.

[57] Tricomi, F. G. and Erdélyi, A. The asymptotic expansion of a ratio of gamma functions. *Pacific Journal of Mathematics* **1**(1) (1951), pp. 133–142.

[58] Wei, L. J. The Generalized Polya's Urn Design for Sequential Medical Trials. *The Annals of Statistics* **7**(2) (1979), pp. 291–296.

[59] Wei, L. J. and Durham, S. The Randomized Play-the-Winner Rule in Medical Trials. *Journal of the American Statistical Association* **73**(364) (1978), pp. 840–843.

[60] Zolotarev, V. M. Metric distances in spaces of random variables and their distributions. *Sbornik: Mathematics* **30**(3) (1976), pp. 373–401.

[61] Zolotarev, V. Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory of Probability & Its Applications* **21**(4) (1977), pp. 721–737.

Deutsche Zusammenfassung

Urnenmodelle im Allgemeinen und Pólya-Urnen-Modelle im Speziellen bilden ein wichtiges Mittel in der diskreten Wahrscheinlichkeitstheorie zur Veranschaulichung und Modellierung diskreter Verteilungen. Darüber hinausgehend dienen Urnenmodelle auch zur Modellierung von sich mit dem Verlauf der Zeit entwickelnden Prozessen. Pólya-Urnenmodelle erfreuen sich besonderer Beliebtheit und finden Anwendung u.a. in der Biologie, z.B. Populationsgenetik und Epidemiologie, Sozialwissenschaften und Informatik, etwa im Bereich der Datenstrukturen.

Aufgrund ihrer Bedeutung ist die Untersuchung des Langzeitverhaltens bestimmter Kenngrößen von Pólya-Urnen von großem Interesse und somit ein zentrales Thema in der Stochastik. Wann immer ein Konvergenzverhalten zu beobachten ist, drängt sich die Frage nach Konvergenzraten auf, welche beschreiben, auf welche Art sich der Abstand zwischen Kenngröße und Grenzwert in Abhängigkeit von der Zeit verringert.

In der vorliegenden Doktorarbeit werden im Rahmen von bereits bekannten Grenzwertsätzen obere Schranken der zugehörigen Konvergenzraten hergeleitet.

Im Folgenden werden zunächst die untersuchte Klasse von Pólya-Urnen sowie bekannte Eigenschaften dazu vorgestellt und dann die Resultate formuliert. Im Anschluss daran wird ein rekursiver Zugang zur Entwicklung des Urnenprozesses erläutert, welcher für die Ermittlung von Konvergenzraten mittels der Kontraktionsmethode unerlässlich ist. Darauffolgend wird die Beweisführung skizziert. Die Zusammenfassung wird durch einige Anmerkungen zu den Beweisen und Resultaten abgeschlossen.

Rahmen

Eine Pólya-Urne beschreibt einen stochastischen Prozess, der sich in diskreten Zeitschritten entwickelt. Zu jedem Zeitpunkt befinden sich Kugeln verschiedener Farben in der Urne. Ein Schritt besteht daraus, eine Kugel aus der Urne zu ziehen und zusammen mit neuen Kugeln, deren Anzahl und Zusammensetzung von der Farbe der gezogenen Kugel abhängt, zurück in die Urne zu legen. Üblicherweise werden die Regeln, die das Hinzufügen neuer

Kugeln beschreiben, in einer sogenannten Rücklegematrix zusammengefasst. Die Rücklegematrix zusammen mit der Anfangsbelegung der Urne beschreibt diesen Prozess vollständig.

In der vorliegenden Arbeit werden balancierte, irreduzible Pólya-Urnen mit zwei Farben, schwarz und weiß, betrachtet. Balanciertheit äußert sich darin, dass in jedem Schritt der Urne dieselbe Anzahl von Kugeln hinzugefügt wird. Irreduzibilität bedeutet, dass unabhängig von der Anfangsbelegung der Urne Kugeln aller Farben mit positiver Wahrscheinlichkeit zu beobachten sind.

Es werden zwei Pólya-Urnenmodelle studiert: Zunächst das folgende Urnenmodell, bei dem die Rücklegematrix gegeben ist durch

$$(1) \qquad R = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ mit } a, d \in \mathbb{N}_0 \cup \{-1\} \text{ und } b, c \in \mathbb{N},$$

sodass $a + b = c + d =: K - 1 \geq 1$ (Balanciertheit), sowie $bc > 0$ (Irreduzibilität) gilt. Wird also in einem Schritt eine schwarze Kugel aus der Urne gezogen, so wird diese mit a schwarzen und b weißen Kugeln zusammen in die Urne zurückgelegt. Andernfalls, wenn eine weiße Kugel in einem Schritt aus der Urne gezogen wird, wird diese mit c schwarzen und d weißen Kugeln in die Urne zurückgelegt.

Des weiteren wird ein Urnenmodell untersucht, bei dem die Einträge der Rücklegematrix zufällig sind. Dieses wird im Anschluss an die Zusammenfassung im Abschnitt „Anmerkungen“ genannt. Für dieses Urnenmodell ergeben sich dieselben Resultate wie für das erste. Im weiteren Verlauf der Zusammenfassung beziehen sich alle Aussagen auf das zuerst genannte Urnenmodell.

Es sei B_n die Anzahl schwarzer Kugeln nach n Schritten. Das Konvergenzverhalten dieser Größe ist bereits umfassend im Rahmen von Grenzwertsätzen untersucht worden. Es bezeichne λ das Verhältnis zwischen größtem und kleinstem Eigenwert der Rücklegematrix, d.h. $\lambda := \frac{a-c}{a+b}$. Es ist bekannt, dass die Anzahl schwarzer Kugeln nach n Schritten, passend normalisiert, in Abhängigkeit von λ verschiedene Grenzverhalten zutage bringt: Gilt $\lambda > \frac{1}{2}$, so konvergiert die passend normalisierte Anzahl schwarzer Kugeln nach n Schritten fast sicher gegen eine nicht-triviale Zufallsvariable, deren Verteilung von Anfangsbelegung und Rücklegematrix abhängt. Andernfalls, wenn also $\lambda \leq \frac{1}{2}$ erfüllt ist, konvergiert die passend normalisierte Anzahl schwarzer Kugeln nach n Schritten in Verteilung gegen die Normalverteilung, siehe beispielsweise [22].

In der vorliegenden Arbeit werden obere Schranken für die Konvergenzraten in verschiedenen Wahrscheinlichkeitsmetriken im Rahmen der bekannten Grenzwertsätze hergeleitet. Dazu wird die Entwicklung des Urnenprozesses rekursiv aufgefasst und dementsprechend mithilfe eines Baumes kodiert. Diese Kodierung eröffnet Zugang zu einer Selbstähnlichkeit, welche im Urnenprozess versteckt ist. Mithilfe dieser rekursiven Auffassung des Urnenprozesses kann die Untersuchung der Anzahl schwarzer Kugeln in den Rahmen der Kontraktionsmethode gebracht werden, welche die Verwendung verschiedener Wahrscheinlichkeitsmetriken begründet. Die Kontraktionsmethode weiß die zugrunde liegende Selbstähnlichkeit auszunutzen und ist nicht nur dazu in der Lage, die bisher bekannten Grenzwertsätze herzuleiten, vgl. Knape und Neininger [27], sondern ermöglicht es auch, das Konvergenzverhalten quantitativ zu erfassen und somit obere Schranken für die Konvergenzraten zu ermitteln.

Ergebnisse

Die folgenden drei Metriken werden in den Beweisen verwendet:

Die Wasserstein-Metrik ℓ_p ist gegeben durch

$$\ell_p(V,W) := \ell_p(\mathcal{L}(V), \mathcal{L}(W)) := \inf\left\{\|V'-W'\|_p \,|\, \mathcal{L}(V') = \mathcal{L}(V), \mathcal{L}(W') = \mathcal{L}(W)\right\}$$

für alle $1 \leq p < \infty$ und Zufallsvariablen V und W mit $\|V\|_p, \|W\|_p < \infty$.

Weiter bezeichne F_V die Verteilungsfunktion einer Zufallsvariablen V. Dann ist der Kolmogorov-Smirnov-Abstand gegeben durch

$$\varrho(V,W) := \sup_{x\in\mathbb{R}} |F_V(x) - F_W(x)|.$$

Schließlich ist die Zolotarev-Metrik ζ_s mit $s = 3$ gegeben durch

$$\zeta_3(V,W) := \sup_{f\in\mathcal{F}_3} |\mathbb{E}[f(V) - f(W)]|,$$

wobei $\mathcal{F}_3$ wie folgt definiert ist: $\mathcal{F}_3 := \{f \in C^2(\mathbb{R},\mathbb{R}) : |f''(x) - f''(y)| \leq |x-y|\}$.

Wie bereits erwähnt, hängt das asymptotische Verhalten der (passend normalisierten) Anzahl schwarzer Kugeln nach n Schritten von dem Verhältnis der Eigenwerte der Rücklegematrix λ ab. Der Parameter λ bewegt sich im Intervall $\left[-\frac{K+1}{K-1}, \frac{K-3}{K-1}\right]$.

Wenn $\lambda > \frac{1}{2}$ erfüllt ist, so existiert eine nicht-triviale, nicht-normalverteilte Zufallsvariable X_R^0, deren Verteilung von der Anfangsbelegung der Urne und der Rücklegematrix abhängt, sodass für $n \to \infty$ gilt:

$$X_n := \frac{B_n - \mathbb{E}[B_n]}{n^\lambda} \longrightarrow X_R^0 \quad \text{fast sicher.}$$

Wenn jedoch $\lambda \leq \frac{1}{2}$ erfüllt ist, so konvergiert die standardisierte Anzahl schwarzer Kugeln in Verteilung gegen die Standard-Normalverteilung, d.h. für $n \to \infty$ gilt

$$\hat{X}_n := \frac{B_n - \mathbb{E}[B_n]}{\sqrt{\operatorname{Var}(B_n)}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Es folgen die Ergebnisse in beiden Fällen, die für beliebige Anfangsbelegungen der Urne gelten:

Theorem 1. *Es sei $\varepsilon > 0$ und $1 \leq p < \infty$. Im Falle $\lambda > \frac{1}{2}$ gilt, für $n \to \infty$,*

$$\ell_p\left(X_n, X_R^0\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right),$$

sowie

$$\varrho\left(X_n, X_R^0\right) = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right).$$

Theorem 2. *Es sei $\varepsilon > 0$. Im Falle $\lambda \leq \frac{1}{2}$ gilt, für $n \to \infty$,*

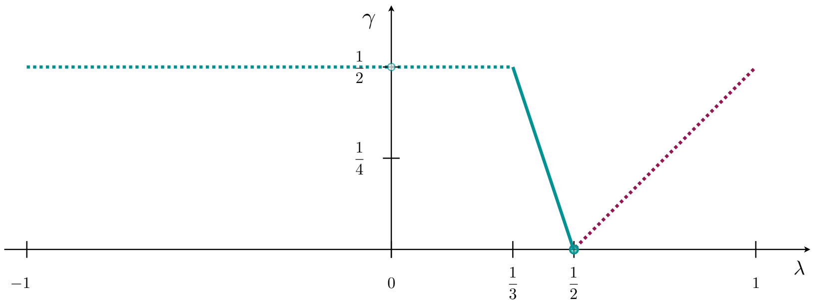
$$\zeta_3\left(\hat{X}_n, \mathcal{N}(0,1)\right) = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda-\frac{1}{2}\right)}\right), & \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon}\right), & \lambda \leq \frac{1}{3}, \lambda \neq 0. \end{cases}$$

Rekursive Beschreibung der Entwicklung der Urne

Die Beweise beruhen auf einem rekursiven Ansatz, welcher im Jahre 2013 von Knape und Neininger [27] entworfen und vorgestellt wurde. Zu jedem Zeitpunkt wird die Urne durch einen Baum kodiert, dessen Blätter den Kugeln in der Urne entsprechen. Dieser Baum wird als *assoziierter Baum* bezeichnet. Auf Grundlage dieser Beobachtung lässt sich für die Anzahl schwarzer Kugeln nach n Schritten eine Verteilungsrekursion entwickeln, welche dann mithilfe der Kontraktionsmethode untersucht werden kann. Knape und Neininger haben dies in [27] bereits im Rahmen von Grenzwertsätzen ausgearbeitet; hier wird nun dieser Ansatz fortgeführt und im Hinblick auf Konvergenzraten ausgenutzt.

Im assoziierten Baum wird ein Schritt der Urne wie folgt realisiert: Das Ziehen einer Kugel aus der Urne entspicht dem zufälligen Auswählen eines Blatts des Baums. Wurde eine schwarze Kugel gezogen, so wird diese mit a schwarzen und b weißen Kugeln zurück in die Urne gelegt. Im Baum wird das gezogene Blatt zu einem inneren Knoten und es erhält K Kinder, von denen $a+1$ schwarz und b weiß gefärbt sind. Analog wird im Falle des Ziehens einer weißen Kugel das entsprechende weiße Blatt im Baum Vorfahr von c schwarzen Blättern und $d+1$ weißen Blättern und selbst zu einem inneren Knoten.

Die Graphik veranschaulicht das Verhalten des negativen Exponenten, welcher hier mit γ bezeichnet wird, in den oben angegebenen oberen Schranken der Konvergenzraten in Theoremen 1 und 2.



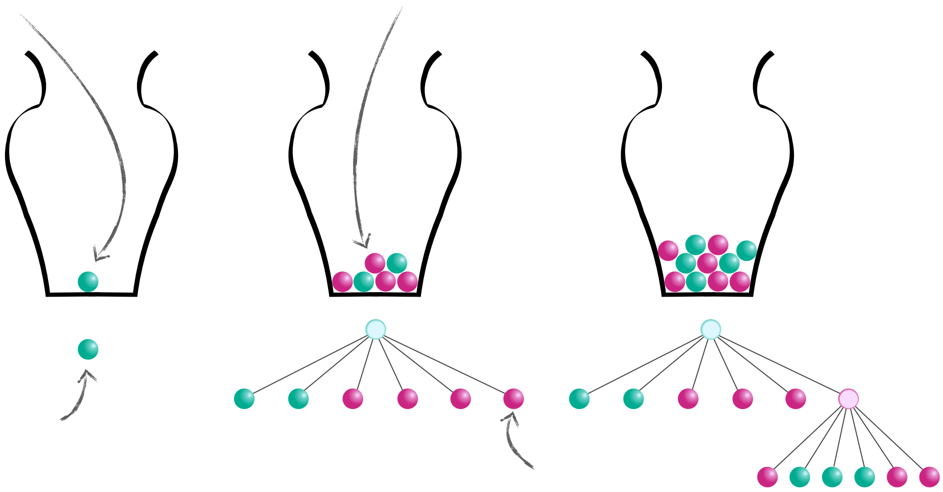
Die gestrichelte Linie soll andeuten, dass sich der Exponent nur bis auf ein beliebiges $\varepsilon > 0$ der Linie nähert; der „wahre" Exponent liegt also (beliebig nah) unterhalb der gestrichelten Linie. Die dunkelviolette Linie stellt den negativen Exponenten im Falle von nicht-normalverteilten Grenzwerten dar, die blaugrüne Linie im Falle des normalverteilten Grenzwerts. Für $\lambda = 1/2$ ist die Ordnung der Rate nicht polynomiell, sondern logarithmisch; dies wird durch einen Kreis bei $(1/2, 0)$ angedeutet. Für $\lambda = 0$ ist das Verhalten der Urne deterministisch; darauf weist ein weiterer Kreis bei $(0, 1/2)$ hin.

Der Baum entwickelt sich also simultan zur Urne. Je nach Farbe der Kugel, die zu Beginn in der Urne liegt, bzw. der Farbe der Wurzel gibt es zwei Typen solcher Bäume: Ist die erste Kugel bzw. die Wurzel schwarz, so wird von einem *b-assoziierten Baum* gesprochen (wegen *black*). Andernfalls wird der Baum als *w-assoziierter Baum* bezeichnet (wegen *white*). Befindet sich zu Beginn mehr als eine Kugel in der Urne, so wird die Entwicklung der Urne durch einen Wald von assoziierten Bäumen erfasst, von denen jeder zu einer der Startkugeln gehört. Nach Konstruktion stimmen die Anzahl der schwarzen Kugeln in der Urne und die Anzahl der schwarzen Blätter im Baum (bzw. in den Bäumen des Waldes) überein.

Zunächst wird davon ausgegangen, dass sich zu Beginn nur eine Kugel in der Urne befindet und die Entwicklung der Urne somit durch einen Baum, keinen Wald, erfasst wird. Es bezeichne $I^{(n)} := \left(I_1^{(n)}, \ldots, I_K^{(n)}\right)$ den Zufallsvektor, dessen r-ter Eintrag die Anzahl der Züge im r-ten Teilbaum der Wurzel innerhalb der ersten n Schritte beschreibt. Es gilt $\sum_{r=1}^{K} I_r^{(n)} = n - 1$, da der erste Zug aus der Urne zum Erzeugen der Teilbäume dient und die darauffolgenden $n - 1$ Züge dann in den Teilbäumen der Wurzel stattfinden. Aufgrund der Balanciertheit der Urne besitzt der r-te Teilbaum der Wurzel nach n Schritten genau $1 + I_r^{(n)}(K - 1)$ Blätter und der Vektor $I^{(n)}$ beschreibt somit die Größe der Teilbäume.

Schließlich ist festzustellen, dass die Teilbäume der Wurzel bedingt auf den Vektor ihrer

Realisierung einer Urne zusammen mit ihrem assoziierten Baum:



Das Bild zeigt die ersten zwei Schritte einer Urne mit schwarzen und weißen Kugeln, die zu Beginn eine schwarze Kugel enthält und deren Dynamik der Rücklegematrix $\left(\begin{smallmatrix}1 & 4\\ 3 & 2\end{smallmatrix}\right)$ folgt. Je Schritt ist der jeweilige assoziierte Baum darunter abgebildet. Die Pfeile sind auf die Kugeln in der Urne bzw. Blätter des Baums gerichtet, die im jeweiligen Schritt gezogen werden. In jedem Schritt entsprechen die Blätter des Baumes den Kugeln in der Urne; zur Klarheit sind Knoten im Baum, die nicht mehr einer Kugel entsprechenden, in verblassender Farbe dargestellt.

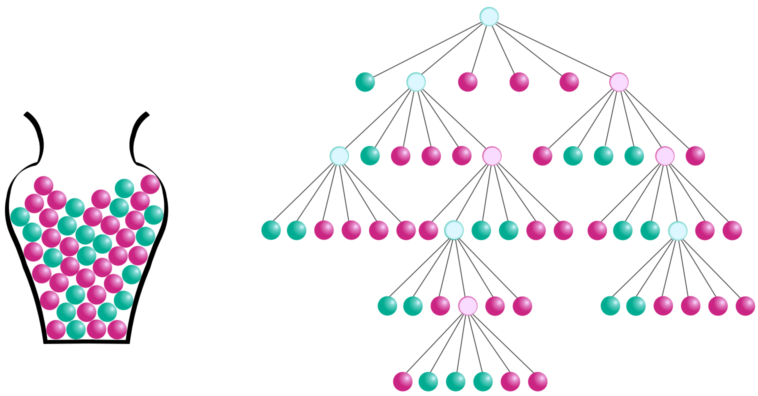
Größen $I^{(n)}$ unabhängig sind und verteilt sind wie b- bzw. w-assoziierte Bäume mit der jeweiligen Anzahl an Blättern.

Mithilfe dieser Eigenschaften und Beobachtungen lässt sich nun die Anzahl schwarzer Kugeln nach n Schritten rekursiv auffassen. Dazu wird der Baum an der Wurzel zerlegt und die Anzahl der schwarzen Kugeln über die Summe der Anzahl schwarzer Blätter je Teilbaum ermittelt. Es bezeichne B_n^{b} die Anzahl schwarzer Kugeln nach n Schritten, wenn die Urne zu Beginn eine schwarze Kugel enthält, und B_n^{w} die Anzahl schwarzer Kugeln nach n Schritten, wenn die Urne zu Beginn eine weiße Kugel enthält. Mit $B_0^{\mathrm{b}} = 1$ und $B_0^{\mathrm{w}} = 0$ ergeben sich die folgenden Verteilungsrekursionen für $n \geq 1$

$$
\begin{aligned}
B_n^{\mathrm{b}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=a+2}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)}, \\
B_n^{\mathrm{w}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=c+1}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)}
\end{aligned} \tag{2}
$$

mit $B_j^{\mathrm{b},(r)} \stackrel{\mathrm{d}}{=} B_j^{\mathrm{b}}$, $B_j^{\mathrm{w},(r)} \stackrel{\mathrm{d}}{=} B_j^{\mathrm{w}}$ für $r = 1, \dots, K$ und $0 \leq j \leq n$, sodass die Größen $\left(B_j^{\mathrm{b},(1)}\right)_{0\leq j\leq n}$, $\dots$, $\left(B_j^{\mathrm{b},(K)}\right)_{0\leq j\leq n}$, $\left(B_j^{\mathrm{w},(1)}\right)_{0\leq j\leq n}$, $\dots$, $\left(B_j^{\mathrm{b},(K)}\right)_{0\leq j\leq n}$, $I^{(n)}$ unabhängig sind.

Das Bild zeigt eine Realisierung der Urne mit schwarzen und weißen Kugeln, die zu Beginn eine schwarze Kugel enthält und deren Dynamik der Rücklegematrix $\left(\begin{smallmatrix} 1 & 4 \\ 3 & 2 \end{smallmatrix}\right)$ folgt, nach acht Zügen aus der Urne.



Neben der Urne ist der assoziierte Baum zu sehen, dessen Blätter den Kugeln in der Urne entsprechen. Wie zuvor sind Knoten des Baums, die nicht mehr einem Blatt entsprechen, in verblassender Farbe dargestellt.

Die Normalisierung der Anzahl schwarzer Kugeln im Rahmen des Systems von Verteilungsrekursionen gegeben in (2) führt zu einem System von Verteilungsrekursionen für die normalisierten Größen. Diese Verteilungsrekursionen für die normalisierten Größen bilden dann den Ausgangspunkt für die Kontraktionsmethode. Aus (2) ist bereits ersichtlich, dass das Verhalten der Teilbaumgrößen beschrieben durch den Zufallsvektor $I^{(n)}$ eine große Rolle spielt. Es lässt sich zeigen, dass der Vektor der relativen Teilbaumgrößen fast sicher gegen einen Dirichlet-verteilten Vektor konvergiert.

Zunächst werden basierend auf diesen Verteilungsrekursionen Konvergenzraten für entsprechende Pólya-Urnen, welche zu Beginn nur eine Kugel enthalten, hergeleitet. Schließlich werden diese Ergebnisse auf Urnen mit einer beliebigen Anfangsbelegung erweitert und führen so zu den bereits genannten Hauptresultaten.

Beweisführung

Zunächst werden Raten für die *Basisfälle*, in denen die Urne mit einer einzigen Kugel gestartet wird, hergeleitet. Schließlich lassen sich die Raten für Urnen mit einer beliebigen Anfangsbelegung daraus „zusammensetzen". Aufgrund der verschiedenen Normalisierungen im Falle eines nicht-normalverteilten Grenzwerts und im Falle des normalverteilten Grenzwerts führt das System gegeben in (2) zu zwei Systemen von Rekursionen, welche im Rahmen der Kontraktionsmethode mit unterschiedlichen Metriken studiert werden müssen. Im Falle eines nicht-normalveteilten Grenzwerts ist es in diesem Rahmen möglich, Raten mit den Wasserstein-Metriken zu bestimmen, aus denen dann eine Rate in der Kolmogorov-Smirnov-Metrik gefolgert werden kann. Im Falle des normalverteilten Grenzwerts liefern die Wasserstein-Metriken keine Kontraktionseigenschaft für die auftretende Fixpunktgleichung, sodass die Zolotarev-Metrik, welche in dieser Hinsicht mehr Spielraum eröffnet, zur Anwendung kommt. Es werden die folgenden Bezeichnungen verwendet: $\mu_{\mathrm{b}}(n) := \mathbb{E}\left[B_n^{\mathrm{b}}\right], \mu_{\mathrm{w}}(n) := \mathbb{E}\left[B_n^{\mathrm{w}}\right], \sigma_{\mathrm{b}}^2(n) := \operatorname{Var}\left(B_n^{\mathrm{b}}\right)$ sowie $\sigma_{\mathrm{w}}^2(n) := \operatorname{Var}\left(B_n^{\mathrm{w}}\right)$; die jeweilige Standardabweichung wird dann mit $\sigma_{\mathrm{b}}(n)$ bzw. $\sigma_{\mathrm{w}}(n)$ bezeichnet.

Der nicht-normalverteilte Fall: $\lambda > \frac{1}{2}$:
Die Anzahl schwarzer Kugeln wird zentriert und mit der Größenordnung der Standardabweichung skaliert: $\mathscr{X}_0 := 0 =: \mathscr{Y}_0$ und für $n \geq 1$

$$\mathscr{X}_n := \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{n^\lambda}, \quad \mathscr{Y}_n := \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{n^\lambda}.$$

Dann ergibt sich das folgende System von Verteilungsrekursionen für die normalisierten Größen:

$$\begin{aligned}
\mathscr{X}_n &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{X}_{I_r^{(n)}}^{(r)} + \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{Y}_{I_r^{(n)}}^{(r)} + b_{\mathrm{b}}\left(I^{(n)}\right), \\
\mathscr{Y}_n &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{X}_{I_r^{(n)}}^{(r)} + \sum_{r=c+1}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda \mathscr{Y}_{I_r^{(n)}}^{(r)} + b_{\mathrm{w}}\left(I^{(n)}\right)
\end{aligned}$$

mit

$$\begin{aligned}
b_{\mathrm{b}}\left(I^{(n)}\right) &:= n^{-\lambda}\left(\sum_{r=1}^{a+1} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{b}}(n)\right), \\
b_{\mathrm{w}}\left(I^{(n)}\right) &:= n^{-\lambda}\left(\sum_{r=1}^{c} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=c+1}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{b}}(n)\right),
\end{aligned}$$

wobei analoge Bedingungen an Verteilung und (bedingte) Unabhängigkeit erfüllt sein sollen wie in (2).

Im Rahmen der Kontraktionsmethode lässt sich der Grenzwert dieser Größen mithilfe des folgenden Systems von Verteilungsrekursionen charakterisieren:

$$\begin{aligned}\mathscr{X} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} D_r^{\lambda} \mathscr{X}^{(r)} + \sum_{r=a+2}^{K} D_r^{\lambda} \mathscr{Y}^{(r)} + b_{\mathrm{b}}, \\ \mathscr{Y} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} D_r^{\lambda} \mathscr{X}^{(r)} + \sum_{r=c+1}^{K} D_r^{\lambda} \mathscr{Y}^{(r)} + b_{\mathrm{w}}\end{aligned} \tag{3}$$

mit

$$\begin{aligned}b_{\mathrm{b}} &:= d_{\mathrm{b}} \left(-1 + \sum_{r=1}^{a+1} D_r^{\lambda}\right) + d_{\mathrm{w}} \sum_{r=a+2}^{K} D_r^{\lambda}, \\ b_{\mathrm{w}} &:= d_{\mathrm{b}} \sum_{r=1}^{c} D_r^{\lambda} + d_{\mathrm{w}} \left(-1 + \sum_{r=c+1}^{K} D_r^{\lambda}\right)\end{aligned}$$

mit unabhängigen Kopien $\mathscr{X}^{(r)}$ von $\mathscr{X}$, $\mathscr{Y}^{(r)}$ von $\mathscr{Y}$, $r = 1, \ldots, K$, einem Dirichlet-verteilten Zufallsvektor $(D_1, \ldots, D_K)$, dessen Parameter alle gleich $\frac{1}{K-1}$ sind, sodass $\mathscr{X}^{(1)}, \ldots, \mathscr{X}^{(K)}$, $\mathscr{Y}^{(1)}, \ldots, \mathscr{Y}^{(K)}$ und $(D_1, \ldots, D_K)$ unabhängig sind.

Unter allen Paaren von zentrierten Verteilungen mit zweitem Moment gibt es einen eindeutigen Fixpunkt, welcher das System gegeben durch (3) löst, der von nun an mit $(\mathcal{L}(\Lambda_{\mathrm{b}}), \mathcal{L}(\Lambda_{\mathrm{w}}))$ bezeichnet wird und die Grenzverteilung von $(\mathscr{X}_n)_{n\geq 1}$ und $(\mathscr{Y}_n)_{n\geq 1}$ liefert, vgl. [27].

Hierfür lässt sich zeigen, dass für alle $\varepsilon > 0$ und $1 \leq p < \infty$ folgendes gilt:

$$\max\{\ell_p(\mathscr{X}_n, \Lambda_{\mathrm{b}}), \ell_p(\mathscr{Y}_n, \Lambda_{\mathrm{w}})\} = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right), \tag{4}$$

$$\max\{\varrho(\mathscr{X}_n, \Lambda_{\mathrm{b}}), \varrho(\mathscr{Y}_n, \Lambda_{\mathrm{w}})\} = O\left(n^{-\lambda+\frac{1}{2}+\varepsilon}\right). \tag{5}$$

Der erste Schritt des Beweises besteht daraus, per Induktion über n eine obere Schranke für den Abstand $\ell_2(\mathscr{X}_n, \Lambda_{\mathrm{b}})$ herzuleiten. Dazu wird dieser Abstand rekursiv mithilfe der Abstände $\ell_2(\mathscr{X}_j, \Lambda_{\mathrm{b}})$ und $\ell_2(\mathscr{Y}_j, \Lambda_{\mathrm{w}})$ mit $j \in \{0, \ldots, n-1\}$ abgeschätzt. Mit der passend gewählten Induktionsvoraussetzung für $\max\{\ell_2(\mathscr{X}_j, \Lambda_{\mathrm{b}}), \ell_2(\mathscr{Y}_j, \Lambda_{\mathrm{w}})\}$ lässt sich die in (4) genannte Rate per Induktion beweisen. Ebenso lässt sich der Abstand $\ell_2(\mathscr{Y}_n, \Lambda_{\mathrm{w}})$ behandeln, sodass man (4) mit $p = 2$ erhält.

Darauf aufbauend wird mit einer Induktion über p und n die in (4) genannte obere Schranke für beliebiges p hergeleitet: Dazu wird der Abstand $\ell_p(\mathscr{X}_n, \Lambda_{\mathrm{b}})$ in Abhängigkeit von den Abständen mit kleiner Indizes, d.h. $\max\{\ell_q(\mathscr{X}_j, \Lambda_{\mathrm{b}}), \ell_q(\mathscr{Y}_j, \Lambda_{\mathrm{w}})\}$ mit $q \leq p$ und $j \leq n-1$, abgeschätzt; gleichermaßen kann mit $\ell_p(\mathscr{Y}_n, \Lambda_{\mathrm{w}})$ verfahren werden.

Mithilfe von (4) und Wissen über Existenz und Eigenschaften der Dichten der Grenzverteilungen $\mathcal{L}(\Lambda_{\mathrm{b}})$ und $\mathcal{L}(\Lambda_{\mathrm{w}})$ lässt sich schließlich eine obere Schranke in der Kolmogorov-Smirnov-Metrik herleiten, welche in (5) formuliert ist.

Um nun Theorem 1 zu erhalten, setzt man die Anzahl schwarzer Kugeln nach n Schritten bei einer Startbelegung mit mehr als einer Kugel mithilfe der Beiträge der einzelnen assoziierten Bäume des Waldes zusammen. Je assoziiertem Baum ist eine obere Schranke für den Abfall der Rate bekannt. Das Vorgehen ist ähnlich zum Vorgehen in den beiden Basisfällen, jedoch ist es nicht möglich, auf eine Verteilungsrekursion für die Anzahl schwarzer Kugeln zurückzugreifen: Man beginnt auch hier mit einer Rate in der Wasserstein-Metrik ℓ_2, verallgemeinert diese auf ℓ_p mit $1 \leq p < \infty$ und überträgt diese Rate schließlich auf die Kolmogorov-Smirnov-Metrik.

Der normalverteilte Fall: $\lambda \leq \frac{1}{2}$
Die Anzahl schwarzer Kugeln nach n Schritten wird zentriert und mit der Standardabweichung zentriert: $\hat{\mathscr{X}}_0 := 0 =: \hat{\mathscr{Y}}_0$, $\hat{\mathscr{X}}_1 := 0 =: \hat{\mathscr{Y}}_1$ und für $n \geq 2$,

$$\hat{\mathscr{X}}_n := \frac{B_n^{\mathrm{b}} - \mathbb{E}\left[B_n^{\mathrm{b}}\right]}{\sqrt{\mathrm{Var}\left(B_n^{\mathrm{b}}\right)}}, \quad \hat{\mathscr{Y}}_n := \frac{B_n^{\mathrm{w}} - \mathbb{E}\left[B_n^{\mathrm{w}}\right]}{\sqrt{\mathrm{Var}\left(B_n^{\mathrm{w}}\right)}}. \tag{6}$$

Es ergibt sich das folgende System von Verteilungsrekursionen für die normalisierte Anzahl schwarzer Kugeln

$$\begin{aligned} \hat{\mathscr{X}}_n &\overset{\mathrm{d}}{=} \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} \hat{\mathscr{X}}_{I_r^{(n)}}^{(r)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} \hat{\mathscr{Y}}_{I_r^{(n)}}^{(r)} + t_{\mathrm{b}}\left(I^{(n)}\right), \\ \hat{\mathscr{Y}}_n &\overset{\mathrm{d}}{=} \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} \hat{\mathscr{X}}_{I_r^{(n)}}^{(r)} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} \hat{\mathscr{Y}}_{I_r^{(n)}}^{(r)} + t_{\mathrm{w}}\left(I^{(n)}\right) \end{aligned} \tag{7}$$

mit

$$\begin{aligned} t_{\mathrm{b}}\left(I^{(n)}\right) &:= \frac{1}{\sigma_{\mathrm{b}}(n)} \left(\sum_{r=1}^{a+1} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{b}}(n)\right), \\ t_{\mathrm{w}}\left(I^{(n)}\right) &:= \frac{1}{\sigma_{\mathrm{w}}(n)} \left(\sum_{r=1}^{c} \mu_{\mathrm{b}}\left(I_r^{(n)}\right) + \sum_{r=c+1}^{K} \mu_{\mathrm{w}}\left(I_r^{(n)}\right) - \mu_{\mathrm{w}}(n)\right), \end{aligned}$$

wobei analoge Bedingungen an Verteilung und (bedingte) Unabhängigkeit erfüllt sein sollen wie in (2).

Im Geiste der Kontraktionsmethode führt dies zum folgenden System von Fixpunktgleichungen für die Grenzverteilungen:

$$
\begin{aligned}
\hat{\mathscr{X}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{a+1} \sqrt{D_r}\hat{\mathscr{X}}^{(r)} + \sum_{r=a+2}^{K} \sqrt{D_r}\hat{\mathscr{Y}}^{(r)}, \\
\hat{\mathscr{Y}} &\stackrel{\mathrm{d}}{=} \sum_{r=1}^{c} \sqrt{D_r}\hat{\mathscr{X}}^{(r)} + \sum_{r=c+1}^{K} \sqrt{D_r}\hat{\mathscr{Y}}^{(r)}
\end{aligned} \tag{8}
$$

mit unabhängigen Kopien $\hat{\mathscr{X}}^{(r)}$ von $\hat{\mathscr{X}}$ und $\hat{\mathscr{Y}}^{(r)}$ von $\hat{\mathscr{Y}}$, $r = 1, \ldots, K$ und einem Dirichlet-verteilten Zufallsvektor $(D_1, \ldots, D_K)$, dessen Parameter alle gleich $\frac{1}{K-1}$ sind, sodass die Größen $\hat{\mathscr{X}}^{(1)}, \ldots, \hat{\mathscr{X}}^{(K)}, \hat{\mathscr{Y}}^{(1)}, \ldots, \hat{\mathscr{Y}}^{(K)}$ und $(D_1, \ldots, D_K)$ unabhängig sind.

Im Raum aller Paare von zentrierten Verteilungen mit Varianz 1 und drittem Moment ist der eindeutige Fixpunkt zum System (8) gegeben durch $(\mathcal{N}(0,1), \mathcal{N}(0,1))$.

Nun lässt sich folgendes zeigen: Für $\varepsilon > 0$ gilt

$$
\max\left\{\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right), \zeta_3\left(\hat{\mathscr{Y}}_n, \mathcal{N}(0,1)\right)\right\} = \begin{cases} O\left((\ln(n))^{-\frac{3}{2}}\right), & \lambda = \frac{1}{2}, \\ O\left(n^{3\left(\lambda-\frac{1}{2}\right)}\right), & \frac{1}{3} < \lambda < \frac{1}{2}, \\ O\left(n^{-\frac{1}{2}+\varepsilon}\right), & \lambda \leq \frac{1}{3},\ \lambda \neq 0. \end{cases} \tag{9}
$$

Auch der Beweis von (9) wird je Bereich von λ in drei Schritten vollzogen. Diese sind jedoch einer völlig anderen Natur als im Fall eines nicht-normalverteilten Grenzwerts.

Wie üblich in der Kontraktionsmethode bei der Verwendung der Zolotarev-Metrik wird eine Folge eingeschoben, die als Bindeglied zwischen ursprünglicher Folge und Grenzwert dient: Für $n \geq 2$ sei

$$
\begin{aligned}
\mathscr{Q}_n^{\mathrm{b}} &:= \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} N_r + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{b}}(n)} N_r + t_{\mathrm{b}}\left(I^{(n)}\right), \\
\mathscr{Q}_n^{\mathrm{w}} &:= \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} N_r + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}\left(I_r^{(n)}\right)}{\sigma_{\mathrm{w}}(n)} N_r + t_{\mathrm{w}}\left(I^{(n)}\right)
\end{aligned}
$$

mit Standard-normalverteilten $N_1, \ldots, N_K$, sodass $N_1, \ldots, N_K, I^{(n)}$ unabhängig sind.

Im ersten Schritt wird $\zeta_3\left(\hat{\mathscr{X}}_n, \mathscr{Q}_n^{\mathrm{b}}\right)$ rekursiv abgeschätzt. Dazu wird dieser Abstand in Abhängigkeit von $\max\left\{\zeta_3\left(\hat{\mathscr{X}}_j, \mathcal{N}(0,1)\right), \zeta_3\left(\hat{\mathscr{Y}}_j, \mathcal{N}(0,1)\right)\right\}$ mit $j \in \{0, \ldots, n-1\}$ unter Verwendung der $(3,+)$-Idealität der Zolotarev-Metrik ζ_3 dargestellt. Als zweites werden die Abstände $\zeta_3\left(\mathscr{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\right)$ und $\zeta_3\left(\mathscr{Q}_n^{\mathrm{w}}, \mathcal{N}(0,1)\right)$ abgeschätzt. Dazu werden die Faltungseigenschaft der Normalverteilung und Symmetrieeigenschaften dieser ausgenutzt, sowie die Struktur der Zolotarev-Metrik genauer untersucht, indem die Testfunktionen aus $\mathcal{F}_3$ mit einer

Taylor-Entwicklung studiert werden. Schließlich dienen die Zwischenergebnisse der ersten beiden Schritte als Abschätzung für den Abstand $\zeta_3\left(\hat{\mathscr{X}}_n, \mathcal{N}(0,1)\right)$ und eine Induktion über n mit passenden Induktionsvoraussetzungen für $\max\left\{\zeta_3\left(\hat{\mathscr{X}}_j, \mathcal{N}(0,1)\right), \zeta_3\left(\hat{\mathscr{Y}}_j, \mathcal{N}(0,1)\right)\right\}$ mit $j \in \{0, \ldots, n-1\}$ je Bereich von λ liefert die Rate aus (9). Der zweite Abstand $\zeta_3\left(\hat{\mathscr{Y}}_n, \mathcal{N}(0,1)\right)$ kann auf die gleiche Weise studiert werden, sodass schließlich die Aussage in (9) folgt.

Für die Anzahl schwarzer Kugeln nach n Schritten bei einer beliebigen Anfangsbelegung mit mehr als einer Kugel wird diese wieder mithilfe der Beiträge der einzelnen assoziierten Bäume erfasst und somit Theorem 2 aus der Kenntnis des Verhaltens der Basisfälle ermittelt.

Anmerkungen

Mit derselben Strategie wird auch eine weitere Urne untersucht, deren Rücklegematrix $\bar{R}$ zufällig ist:

$$\bar{R} = \begin{pmatrix} C_\alpha & 1 - C_\alpha \\ 1 - C_\beta & C_\beta \end{pmatrix} \text{mit } C_\alpha \sim \text{Ber}(\alpha), C_\beta \sim \text{Ber}(\beta), \alpha, \beta \in (0,1).$$

Dieses Urnenmodell findet seinen Ursprung in der Entwicklung von klinischen Studien im Rahmen von adaptiven Designs und wird in diesem Kontext als „Randomised Play-the-Winner Rule“ bezeichnet. Anschaulich gesprochen gehören zu der Urne zwei Münzen – eine „schwarze“ und eine „weiße“ Münze. Wird eine Kugel aus der Urne gezogen, so wird sie zusammen mit einer weiteren Kugel in die Urne zurückgelegt, deren Farbe durch einen Münzwurf entschieden wird: Wird eine schwarze Kugel gezogen, so wird die schwarze Münze geworfen und mit Wahrscheinlichkeit α wird eine schwarze Kugel hinzugefügt, mit Wahrscheinlichkeit $1 - \alpha$ eine weiße. Wird eine weiße Kugel gezogen, so entscheidet der Ausgang des Münzwurfs mit der weißen Münze über die Farbe der neuen Kugel und es wird mit Wahrscheinlichkeit β eine weitere weiße, mit Wahrscheinlichkeit $1 - \beta$ eine schwarze Kugel hinzugefügt.

Auch für diese Urne wurden die in Theorem 1 und Theorem 2 genannten Ergebnisse hergeleitet, wobei der Parameter, der über nicht-normalverteilte und normalverteilte Grenzwerte entscheidet, nun gegeben ist durch $\lambda = \alpha + \beta - 1$.

Eine entscheidende Rolle in den oben vorgestellten Beweisstrategien spielt das Verhalten der Varianz der Anzahl schwarzer Kugeln nach n Schritten. In der Skalierung geht sie in der Form der Standardabweichung bzw. deren Größenordnung in die Rekursionen und somit in die Rechnungen ein. Für die oben vorgestellten Beweisstrategien ist es unumgänglich, den führenden Ordnungsterm der Varianz mit exakter Konstante und eine groß O-Abschätzung für den zweiten Ordnungsterm zu kennen.

Über die Optimalität der hier erzielten Raten kann keine Aussage gemacht werden, da keine unteren Schranken bekannt sind.

Die Dissertation wird durch Simulationen im Falle von normalverteilten Grenzwerten ergänzt.

Theorem 2 bestätigt zum Teil eine Vermutung von Svante Janson in [22]. Er vermutet, dass die Rate von der Größenordnung $O\left(n^{3\left(\lambda-\frac{1}{2}\right)\vee\left(-\frac{1}{2}\right)}\right)$ ist. Es bleibt offen, ob das ε in Theorem 2 (und auch in Theorem 1) nötig ist.