Central Limit Theorems

for

Multicolour Urn Models

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

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

von
Noela Müller
aus Frankfurt am Main

Frankfurt am Main (2017)
(D 30)

vom Fachbereich Informatik und Mathematik der

Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Andreas Bernig

Gutachter: Prof. Dr. Ralph Neininger

Prof. Dr. Rudolf Grübel

Datum der Disputation: 22.09.2017

Acknowledgements

I wish to express my sincere gratitude to my advisor Prof. Dr. Ralph Neininger, whose lectures first introduced me to probability theory. During the course of my studies, he wrote uncountably many letters of support at various occasions. When I started my thesis under his supervision, he provided me with an interesting and multifaceted topic and drew my attention to various related publications and directions of research. Finally, I would like to thank him for the joint work on cyclic urns and the opportunity to attend several inspiring conferences and workshops.

For their helpful comments and criticisms during the course of my PhD, I am ever grateful to Dr. Kevin Leckey and Dr. Henning Sulzbach, whose comments lead to a considerable improvement of the results that are collected in the present thesis. I am deeply grateful for their help and patience.

Additionally, I would like to thank my PhD colleague Andrea Kuntschik for various discussions and moral support throughout the last three years.

I also wish to acknowledge my academic and professional indebtedness to all my teachers in the field of probability theory during the last eight years: Prof. Dr. Götz Kersting, Prof. Dr. Nicola Kistler, Prof. Dr. Christoph Kühn and Prof. Dr. Anton Wakolbinger. I have also learned a lot in discussions with former fellow students to whom I express my most heartfelt thanks: James Bell, Steffen Eibelshäuser, Pooya Vahidi Ferdowsi, Tobias Kapetanopoulos, Kathrin Skubch, Adam Zsolt Wagner, Ilias Zadik and Claudius Zibrowius.

Finally, I am very grateful for financial and non-material support during my studies by the *Studienstiftung des deutschen Volkes* as well as for the *DAAD* programme "Strategische Partnerschaften" for enabling a three-week research visit at the University of Birmingham.

Notation

We use the following sets in their common designation in course of the thesis:

- $\mathbb{N} := \{1, 2, \dots\}$ is the set of natural numbers,
- $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ are the natural numbers with zero,
- $\mathbb{Z} := \{\dots, -1, 0, -1, 2, \dots\}$ is the set of integers,
- $\mathbb{Z}_- := \{0, -1, -2, \dots\}$ is the set of non-positive integers,
- $\mathbb{R}$ is the set of real numbers,
- $\mathbb{C}$ is the set of complex numbers.

For a complex number $z \in \mathbb{C}$, we denote by $\mathfrak{R}(z)$, $\mathfrak{I}(z)$ and $|z|$ its real part, imaginary part and complex modulus, respectively. Moreover, $\bar{z}$ denotes the complex conjugate of $z \in \mathbb{C}$. For a complex vector $v \in \mathbb{C}^q$ and $i \in \{1, \dots, q\}$, we denote by $v^{(i)}$ its i^{th} component and by v^t and v^* its transpose and conjugate transpose, respectively. Further let $|v|$ denote its L^1-norm. We equip $\mathbb{C}^q$ with the standard inner product $\langle \cdot, \cdot \rangle$, where $\langle u, v \rangle := u^* v$.
For $x \in \mathbb{R}^d$, we denote by $\|x\|$ the standard Euclidean norm of x, and for $B \in \mathbb{R}^{d \times d}$, $\|B\|_{\text{op}}$ denotes the corresponding operator norm. For random variables X and $p \geq 1$, we denote by $\|X\|_p$ the L_p-norm of X.
We denote by $\text{Id}_{\mathbb{C}^q}$ or $\text{Id}_{\mathbb{R}^q}$ the $q \times q$ identity matrix. Furthermore, for $a, b \in \mathbb{Z}$, $a \mid b$ and $a \nmid b$ are short for a divides (resp. does not divide) b.
The common probability distributions that arise in this thesis are the following:

- For $x \in \{1, \dots, q\}$, δ_x is the Dirac measure in x.
- For $p \in (0, 1)$, $\text{Bern}(p)$ denotes the Bernoulli distribution with success parameter p.
- For $p \in (0, 1)$ and $n \in \mathbb{N}$, let $\text{Bin}(n, p)$ denote the Binomial distribution with parameters n and p.
- $\text{unif}[0, 1]$ denotes the uniform distribution on the interval $[0, 1]$.
- $\mathcal{N}(\mu, \Sigma)$ denotes the (multivariate) normal distribution with mean μ and covariance matrix Σ, where Σ is a symmetric positive semi-definite matrix.
- $\text{Beta}(\alpha, \beta)$ denotes the beta distribution with parameters $\alpha, \beta > 0$.
- For $K \geq 2$, $\text{Dir}(\alpha_1, \dots, \alpha_K)$ denotes the Dirichlet distribution with parameters $\alpha_1, \dots, \alpha_K > 0$.

Furthermore, for $A \subset \{1, \dots, q\}$, we set $\delta_A := \sum_{x \in A} \delta_x$. By $\mathcal{L}(X)$, the distribution of a random variable X is denoted. Finally, almost sure convergence, convergence in probability and convergence in distribution are denoted as $\xrightarrow{\text{a.s.}}$, $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{\mathcal{L}}$, respectively.

We use Bachmann-Landau symbols in asymptotic statements. In particular, if $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ are two complex-valued sequences, then

$$\begin{aligned}
a_n = O(b_n) &\iff \text{There exist } n_0 \text{ and } C > 0 \text{ such that } |a_n| \leq C|b_n| \text{ for all } n \geq n_0,\\
a_n = o(b_n) &\iff |b_n| > 0 \text{ for all but finitely many } n \text{ and } \frac{a_n}{b_n} \to 0, \quad n \to \infty,\\
a_n \sim b_n &\iff |b_n| > 0 \text{ for all but finitely many } n \text{ and } \frac{a_n}{b_n} \to 1, \quad n \to \infty.
\end{aligned}$$

Finally, $\Gamma : \mathbb{C} \setminus \mathbb{Z}_- \to \mathbb{C}$,

$$\Gamma(z) := \frac{1}{z} \prod_{n=1}^{\infty} \frac{\left(1 + \frac{1}{n}\right)^z}{1 + \frac{z}{n}}$$

denotes the complex Gamma function.

Contents

Foreword

Urn models are simple examples for random growth processes that involve various competing types. In the study of these schemes, one is generally interested in the impact of the specific form of interaction on the allocation of elements to the types. Depending on their reciprocal action, effects of cancellation and self-reinforcement become apparent in the long run of the system. For some urn models, the influencing is of a smoothing nature and the asymptotic allocation to the types is close to being a result of independent and identically distributed growth events. On the contrary, for others, almost sure *random* tendencies or logarithmically periodic terms emerge in the second growth order. The present thesis is devoted to the derivation of central limit theorems in the latter case. For urns of this kind, we use a "non-classical" normalisation to derive asymptotic joint normality of the types. This normalisation takes random tendencies and phases into account and consequently involves random centering and, also, possibly random scaling. Following [34], the term central limit theorem *analogue* is sometimes used to refer to this approach. In this foreword, we first motivate and state the main results of the current thesis and then develop the basic idea of these results along the lines of a simple example.

Introduction

By a *generalised Pólya-Eggenberger urn scheme* we understand the following model: Consider the discrete-time evolution of a container (the urn) of infinite capacity that encloses balls from q different types. At time 0, there are $X_0^{(j)} \in \mathbb{N}_0$ balls of type j in the urn, for $j \in \{1, \dots, q\}$. Immediately before time $n \in \mathbb{N}$, a ball is drawn uniformly at random and independently of the previous draws. If the ball drawn is of type $i \in \{1, \dots, q\}$, we add $\Delta_i^{(j)} \in \mathbb{Z}$ balls of type j for $j \in \{1, \dots, q\}$ to the urn, if $\Delta_i^{(j)} \geq 0$, or remove $\Delta_i^{(j)}$ balls of colour j from the urn, if $\Delta_i^{(j)} < 0$.

More generally, for each $n \in \mathbb{N}_0$ and $j \in \{1, \dots, q\}$, let $X_n^{(j)}$ denote the number of balls of type j in the urn after n draws. The vector

$$X_n = \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \\ \vdots \\ X_n^{(q)} \end{pmatrix} \quad \in \mathbb{N}_0^q$$

is called the *urn composition* at time n. We assume $X_0 \in \mathbb{N}_0^q$ to be deterministic, and one is typically interested in the evolution of the urn composition X_n over time.

Furthermore, let $R := (\Delta_1, \dots, \Delta_q) \in \mathbb{Z}^{q \times q}$ be the matrix with columns $\Delta_1, \dots, \Delta_q$. As before, the j^{th} component of the vector Δ_i is given by the integer $\Delta_i^{(j)}$. The increment vector

Δ_i describes the change in the urn composition if a ball of colour i is drawn, and we assume the vectors $\Delta_1, \ldots, \Delta_q \in \mathbb{Z}^q$ to be deterministic. R is called the *generating matrix* of the process. Note that R is the matrix transpose of the so-called replacement matrix.
Thus, q, X_0 and R are the parameters of the process, whose dynamics are fully described by R and X_0. Later on, we also refer to a type as a colour, which usually is an element of $\{1, \ldots, q\}$. Note that the urn composition evolves according to a discrete-time Markov process, whose main feature is that the conditional transition probability, which contains the randomness of the evolution of the process, is *linear* in the composition X_n.

Historical overview and applications

Generalised Pólya-Eggenberger urn schemes have been an object of ongoing research interest for about 90 years. They owe their name to a study by Eggenberger and Pólya [20] from 1923, where the authors introduce the now well-known Pólya urn in order to model contagious diseases. Starting from there, the original model has been generalised with regard to several aspects and urns of a related kind can nowadays be found in many different contexts. The following account of work on urn models does not claim to be complete; it rather serves as a basis for the results presented in the current thesis. For a more detailed history of Pólya urns, we refer the reader to the textbooks [55] and [46].
In the beginning, the generalisation of the original model was limited to other urn schemes involving two colours. Friedman [26] studies characteristic functions of models with general balanced and symmetric 2×2-matrices. The asymptotics of the numbers of balls of the two types in his model are further examined in Freedman [25]. Bagchi and Pal [4] remove the symmetry condition and deduce a Gaussian limit law for a large class of models whose matrices have constant column sum. Their result on central limit theorems is generalised to urns with an arbitrary number of colours in Smythe [74]. Smythe's approach is similar to the one in chapter 3, even though he assumes less general hypotheses. Gouet [28–30] derives results on the strong asymptotics of proportions for a large class of multicolour models via non-constructive martingale techniques that arise from the linearity in the transition mechanism. He also deduces a functional limit theorem for the two-colour case. Pouyanne [67] develops an operator approach for the determination of moments. This study was recently extended in [45]. Of great importance are also the works of Athreya, Karlin and Ney [2,3] who establish an embedding of urn processes into continuous-time multitype branching processes. Janson's seminal work [40] is based on this technique which makes it possible to study urns without constant column sum. In [40], a functional limit theorem for multitype continuous time Markov branching processes is derived that leads to strong limit results as well as central limit theorems for generalised Pólya urns and even explicit formulas for the asymptotic variances and covariances. Extended Pólya-Eggenberger urns with two and three types are also studied via a purely analytic approach in [24]. Neininger and Knape [47] analyse the asymptotics of urn models with an arbitrary number of types via an embedding of their evolution into random rooted trees. This technique enables the study of urns by means of the contraction method. Finally, this thesis draws on the central limit theorems and almost sure asymptotics derived in [4, 40, 47, 67, 74] for urn models whose generating matrices have constant column sum. The extension of the term "CLT" allows to relax assumptions on the spectrum of the matrix R and to synthesise the cases of asymptotic normality and almost sure asymptotics under a common viewpoint. Methodically, we use techniques from [34, 47, 67, 74].

This listing gives an impression of the wide range of methods that have been employed to study Pólya urn models. To summarise, among these are methods from enumerative combinatorics [4] and analytic combinatorics [24], embedding into continuous time branching processes [2, 40], martingale techniques [29, 30] and the contraction method [47].

Applications. Even wider than the range of methods that have been adopted in the study of urn models is their scope of applications, which ranges over a variety of areas such as biology, ecology, statistics, computer science, random network theory and finance. Some examples of their appearance in the literature are given below:
To begin with, urns can be found in biology-related contexts, as in the study of the spread of epidemics [20]. There are famous urns in population genetics [9, 38] as well. Urns are used as treatment allocation schemes in comparative clinical studies [78]. A statistical study of wage disparity between men and women via urns for regression analysis is given in [52]. Furthermore, urns are closely related to growth mechanisms of various random trees. This connection encompasses data structures as m-ary search trees or b-trees, see [1,4,11,13], scale-free trees in the study of random networks [76] as well as stable trees [27]. Pólya urns have also been used to study fringe structures in various random trees [1,36,37]. Aldous [1] derives a result on the convergence of the empirical distributions of an adaptive process that simulates quasi-stationary distributions of Markov chains by means of a correspondence with an urn scheme. Moreover, urn models come up in problems with a background from finance [50]. Finally, they also arise in the analysis of reinforced random walks [64]. Further applications of urn models can be found in [46, 55].

Results

Much of the versatility of and lasting interest in urn models is owed to their flexibility. In line of this, the exact formulation of further assumptions on the model, whose general form was described in the introduction, varies widely and depends as well on the research interest as on the methods used by the author. The main subject of this thesis is the growth behaviour of the urn composition as $n \to \infty$. We therefore assume that the number of balls contained in the urn increases by a constant amount r at each step, and also, that the replacement rules guarantee that the process is well-defined for all $n \in \mathbb{N}_0$. The exact formulation of all assumptions is given in (A1) - (A5) in the beginning of section 1.2. In the following statement of the results, we assume that (A1) - (A5) hold.
It is known since the 1960s [2], that the asymptotic behaviour of the urn composition is directly connected to the spectrum of the matrix R. For example, Gouet [30] shows that the proportions $X_n/(rn)$ of balls of the different types converge almost surely as the number of draws tends to infinity. Their limit can be a random or a deterministic vector. Whether it is random or deterministic depends on the multiplicity of the largest eigenvalue r of R (and on the initial composition of the urn). Secondly and similarly, it is also by now common knowledge in the field of generalised Pólya-Eggenberger urns that the nature of the second order asymptotics of the urn composition depends on the number of eigenvalues with real parts greater than $r/2$, where r is as above (eigenvalues are counted according to their algebraic multiplicity). More precisely, it has been successfully shown by various techniques, that if this number is exactly one, the rescaled urn composition vector is asymptotically normally distributed, see [40, 67, 74]. If, on the other hand, there is more than one eigenvalue with real part greater than $r/2$, no central limit theorems for the sequence $(X_n)_{n\in\mathbb{N}_0}$ are known.
The main contribution of the present thesis consists in the observation that it is always pos-

sible to normalise the urn composition in a way that leads to distributional convergence, see Theorem 1.2.5. The exact form of the central limit theorem analogue depends, as indicated, on the number of eigenvalues whose real parts are greater than $r/2$. As previously mentioned, for matrices with simple largest eigenvalue $r > 0$ and real parts of all other eigenvalues bounded above by $r/2$, the correct normalisation is the classical one. This corresponds to known results. On the other hand, if r is simple and there are other eigenvalues with real parts greater than $\frac{r}{2}$, Theorem 1.2.5 states that the composition vector has to be centered by a random vector and scaled by $\sqrt{n}$ in order to obtain convergence to a normal distribution. Eventually, if r is a multiple eigenvalue, we obtain weak convergence to a mixed normal distribution after centering by a random vector and scaling by $\sqrt{n}$. Consequently, the relaxation of the term "central limit theorem" leads to a unified perspective for the fluctuation of urn composition vectors.

The probably most interesting aspect of this result is a central limit theorem for urn schemes in which periodicities arise, as in the last example in the following section or chapter 2. The proof is based on a spectral decomposition of the process and uses martingale techniques. Additionally, an alternative proof strategy by means of the contraction method is developed for an example, namely the cyclic urn. This strategy is of independent interest, as it does not use underlying martingale structures, which makes it more flexible. Furthermore, the random centering variables in the central limit theorems are typically only described implicitly as martingale limits. In the case of the cyclic urn however, we can explicitly construct these variables from a sequence of i.i.d. uniform variables via an embedding into the random binary search tree (BST). What is more, a closer look at the construction reveals that the martingale limits are functions of the Doob-Martin limit of the associated BST chain.

An easy introduction to the topic

In this section, a more descriptive approach is taken to motivate the results of this thesis. The expert reader is advised to skip this section.

When asked about metaphors for chance, most people will think of devices from games, as for example coins, dice, cards or spinners. These mechanisms are of value in games because they provide a good way to generate "independent" outcomes each time they are used. Implicitly, they come along with a certain notion of genuinely random choice. This idea of the "behaviour" of chance prevails in many people's perception: Random events are subject to the rule "You win a few, you lose a few".

However, in this work, we are confronted with a different action of chance which is more like in the following line of thought: Many factors have an influence on our lives, and as most of them are out of our reach, we can view them as random. First of all, there is the place of our birth. If you are born into a wealthy nation, you are more likely to survive childhood and to be sent to school. Second, the fact that you went to school highly increases your chances to attend university and subsequently to get a job that earns you enough money to meet your needs, and so on. Success or failure in life are not the result of a sequence of independent events, but rather self-reinforcing. Plainly, everything is somehow random, but highly sensitive to the direction taken in early life.

At this point, urn models come into play. They provide simple to describe, still classical random experiments that additionally possess a certain kind of memory. This property gives rise to a more advanced correlation structure and long-term dependencies as illustrated in the preceding paragraph. Of course, urn models are still too simple to describe real world phenomena in an accurate way, but in order to capture features of randomly growing structures, they have proved as useful mathematical models with applications in many areas. Furthermore, despite the simple way to describe them, they give rise to interesting phenomena.

SLLN, CLT and LIL

In order to put the characteristics that we study by means of urn models into a broader perspective, let us briefly recall the "key limit trio" [34] for sequences of *independent* random variables: the strong law of large numbers (SLLN), the central limit theorem (CLT) and the law of the iterated logarithm (LIL). For independent variables, many aspects of their asymptotic behaviour are well-understood.

Let X be a real-valued, integrable random variable and $X_1, X_2, \ldots$ a sequence of independent and identically distributed (i.i.d.) copies of X. In this setting, the Kolmogorov strong law of large numbers states that the empirical averages of the sequence converge to the theoretical average $\mathbb{E}[X]$ almost surely, that is,

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{\text{a.s.}} \mathbb{E}[X]$$

as $n \to \infty$, where $\xrightarrow{\text{a.s.}}$ is used to denote almost sure convergence. This result is universal in the sense that it does not depend on the actual distribution of the variable X that comprises the system, but only on its mean. In particular, the almost sure limit is deterministic and thus independent of the actual realisation of the sequence $(X_n)_{n\geq 1}$ when performing the experiment. Knowledge of the first, say, 100 variables, will not lead you to change your predictions about the asymptotics of the empirical average, as individual contributions will even out in the long run.

In a next step, both the CLT and the LIL can be regarded as rate results for the almost sure convergence of the SLLN. To formulate both results, we assume that X and $(X_n)_{n\geq 1}$ are as above and furthermore that X has finite, positive variance $\mathrm{Var}(X)$. Let $S_n := \sum_{i=1}^{n} X_i$. First of all, already a simple application of Chebyshev's inequality indicates that the "typical" size of the deviation $|S_n - \mathbb{E}[X]n|$ of the sum from its first order approximation is rather $O(\sqrt{n})$ than the original order $O(n)$: A "square root cancellation" takes place. The central limit theorem now makes this observation more precise by stating that if $0 < \mathrm{Var}(X) < \infty$,

$$\sqrt{\frac{n}{\mathrm{Var}(X)}}\left(\frac{S_n}{n} - \mathbb{E}[X]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

as $n \to \infty$, where $\xrightarrow{\mathcal{L}}$ is used to denote convergence in distribution and $\mathcal{N}(0,1)$ is the normal distribution with mean 0 and variance 1. This yields a "weak" rate result. Secondly, the law of the iterated logarithm provides a "strong" rate result by saying that there is a first order expansion

$$\frac{S_n}{n} - \mathbb{E}[X] = Z_n\sqrt{\frac{2\mathrm{Var}(X)\log\log(n)}{n}}$$

for some sequence $(Z_n)_{n\geq 1}$, where $\limsup_{n\to\infty} Z_n = 1$ almost surely and $\liminf_{n\to\infty} Z_n = -1$ almost surely. As a consequence, the random variables $\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}[X]\right)$ neither converge in probability nor almost surely; and the same is true for any subsequence.
In both results, again, we find universality, as the appearance of the normal distribution does not depend on the actual distribution of the underlying experiment that, again, only enters through its mean $\mathbb{E}[X]$ and variance $\mathrm{Var}(X)$. Thus, theoretical results about the normal distribution are applicable to a broad class of situations, yielding the basis for its importance. Again, knowledge of the first 100 variables will not lead you to change your predictions about the long-term behaviour of the system.

This is the picture provided by SLLN, CLT and LIL in the setting of independent random variables. In the following, we will be concerned with the guise of CLTs in the context of urn models.

A first example

In urn models, the chances to draw a particular ball immediately before time $n+1$ depend on the current composition of the urn. So, at a first glance, the sequence of added balls is far from being independent. Still, our interest is in the derivation of results in the spirit of the SLLN and the CLT. That is, an almost sure convergence result for the urn composition, followed by a weak rate result. In order to illustrate similarities and differences to the independent case, consider the following three varieties of an introductory example.

Our (slightly unrealistic) example is located at a poll site: Assume that an election takes place, and that there are only two candidates to vote for. One voter arrives at the poll site at a time, and each of them has five votes that he may distribute among the candidates. Now, for some reason, the government wants to ensure that the current voter is confronted with his choice for a final time, and upon his arrival, he is forced to watch a short speech of exactly one of the candidates. The choice among the candidates is random, but such that each candidate is chosen by the proportion of votes he has already received.
Suppose that both candidates are quite bad and that it is a torture to listen to any of them. Correspondingly, the reaction of a voter to the speech shown is such that he instantly changes his mind and puts four votes on the other candidate. In case that the candidate he did not listen to is also bad, he reserves one vote for the candidate he just listened to. This can be modelled by an urn process with black (votes for candidate 1) and white (votes for candidate 2) balls and matrix

$$R_1 := \begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}.$$

Here, entry (i, j) indicates how many votes are given to candidate i if candidate j was in the video.
Suppose that, in the beginning, it is equally likely to be shown a video of both candidates. Reformulating, at time zero, there are one black and one white ball in the urn. We are interested in the asymptotic proportion of votes for each candidate (SLLN). General results on urn models as well as symmetry imply that asymptotically, both candidates will receive 50% of the votes. Consequently, in order to decide the winner of the election, the behaviour of the second order term is important. Let S_n denote the number of black balls in the urn

after n draws. For this urn model, general results imply that, as $n \to \infty$,

$$\frac{2\sqrt{55}}{3}\sqrt{n}\left(\frac{S_n}{5n} - \frac{1}{2}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)\,.$$

Again, the normal distribution arises, even though the asymptotic variance is smaller than if the number of black balls in the urn were the result of the number of heads in a sequence of fair coin tosses. The winner of the election is thus drawn according to a normal distribution, and knowledge of the first draws is of no advantage in the prediction of the final outcome.

In a next step, consider a slight modification of our first example. Suppose that the model is as described above, except that now, both candidates are good ones. The voters adapt their behaviour by giving four votes to the candidate they just listened to, and one vote to the other one, accounting to the fact that they have no impression of this person. This new rule corresponds to an urn with matrix

$$R_2 := \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}.$$

Again, general results on urn models imply that asymptotically, both candidates will receive 50% of the votes. So far, there is no difference between the urns.
However, it turns out that $\sqrt{n}\left(\frac{S_n}{5n} - \frac{1}{2}\right)$ does not converge in distribution. Moreover, it even tends to infinity almost surely. On the other hand, it is known that multiplication by a smaller factor yields almost sure convergence

$$n^{2/5}\left(\frac{S_n}{5n} - \frac{1}{2}\right) \xrightarrow{\text{a.s.}} W$$

as $n \to \infty$. Here, W is a non-degenerate random variable that depends on the initial configuration of the urn. It follows that, in this example, there is a random almost sure drift that dominates the second order term, and that the first draws have a significant influence on the outcome of the election. Grübel [31] calls this effect "persisting randomness".
Having found second order almost sure asymptotics, the line of action taken in the present thesis is to also substract them from the first order approximation and ask about asymptotic normality of

$$\sqrt{n}\left(\frac{S_n}{5n} - \frac{1}{2} - n^{-2/5}W\right).$$

The question whether a weak limit exists is answered in Theorem 1.2.5 in this thesis.

We conclude this mini sequence of examples with a third urn that illustrates yet another form of almost sure asymptotics that may be hidden behind a first oder approximation. For this case, we need three types of balls.
Again, consider our well-known voter scheme, but now with three good candidates and voters that are only willing to give away one vote to one of the candidates whose videos they did not watch. This yields an urn process with matrix

$$R_3 := \begin{pmatrix} 4 & 0 & 1 \\ 1 & 4 & 0 \\ 0 & 1 & 4 \end{pmatrix},$$

where we assume that, at the beginning, there is one ball of each type in the urn. Here, too, after the poll site has been open for a long time, the total number of votes will be roughly evenly split among the three candidates. As in the previous example, $\sqrt{n}\left(\frac{S_n}{5n}-\frac{1}{3}\right)$ does not converge but tends to infinity almost surely. However, a closer look at the votes for candidate 1 reveals that this is not only a result of too large scaling: In fact, there is no real α such that $n^\alpha\left(\frac{S_n}{5n}-\frac{1}{3}\right)$ converges to a non-degenerate limit. Instead, the general theory of urn models implies that there is a complex-valued, non-degenerate random variable V that depends on the initial configuration such that

$$\frac{S_n}{5n}-\frac{1}{3}-n^{-1/10}\Re\left(e^{i\frac{\sqrt{3}}{10}\log(n)}V\right)\xrightarrow{\text{a.s.}}0$$

as $n\to\infty$, where $\Re(z)$ denotes the real part of a complex number z. That is, $n^{1/10}\left(\frac{S_n}{5n}-\frac{1}{3}\right)$ is asymptotically approximated by the logarithmically *oscillating* sequence $\Re\left(e^{i\frac{\sqrt{3}}{10}\log(n)}V\right)$. There exist analogous approximations for the other two types.

This is an example for the appearance of periodic phenomena in urn models, which also arise in various other discrete structures that are related to algorithms. Here, the initial state does not give a direction to the second order asymptotics that is in favour of one of the three candidates. Instead, it defines the phase shift and the amplitude of the oscillations subject to which the candidates take their lead. Based on the asymptotic expansion seen so far, we are, once more, attracted by the following question: Is there a CLT for

$$\sqrt{n}\left(\frac{S_n}{5n}-\frac{1}{3}-n^{-1/10}\Re\left(e^{i\frac{\sqrt{3}}{10}\log(n)}V\right)\right)?$$

Again, this question is answered in Theorem 1.2.5 in the current thesis.

Organisation of the thesis

Chapter 1 is devoted to the description of urn models. For a better understanding, the presentation of the main result is split into two steps: First, it is explained and stated for urns with two types only and a lot of examples are given. After the exposition of this comparatively easy case, we elaborate on the general model, which includes the specification of notation and assumptions as well as the precise statement of the main result.

Chapter 2 deals with the cyclic urn. This particular model is used to further illuminate the result and also to propose a proof strategy that is is not used in the proof of the main theorem, namely the contraction method. Here, a proof that is independent of the proof of Theorem 1.2.5 is given, even though it also uses the spectral decomposition. Moreover, the random variables that arise in the centering of the CLT are constructed from a sequence of i.i.d. uniform random variables via a connection to the random binary search tree.

An explanation of the proof idea and the proof of Theorem 1.2.5 itself are provided in chapter 3. To this end, first of all, properties of the spectral decomposition of the urn process are discussed. Afterwards, the proof of the main theorem is carried out in detail.

Finally, chapter 4 shall convey an idea of some applications of the theorem. However, as calculations of eigenvectors can get quite involved, we do not carry out the computations. We conclude the thesis with a discussion of some further open questions.

1 Generalised Pólya-Eggenberger Urns

1.1 Urns with two types

Historically motivated, we start the study of urn processes with models that only involve two types. To further ease the understanding of the topic, examples are given throughout the text.

1.1.1 Introduction

The urns of this chapter contain two colours, called colour 1 and colour 2. Each time a ball of colour 1 is drawn from the urn, it is put back to the urn together with $a_{1,1}$ balls of colour 1 and $a_{2,1}$ balls of colour 2. Similarly, each time a ball of colour 2 is drawn, it is put back to the urn together with $a_{1,2}$ balls of colour 1 and $a_{2,2}$ balls of colour 2. If one of the integers $a_{1,1}, a_{1,2}, a_{2,1}, a_{2,2}$ is negative, this number of balls is removed from the urn. Recall that the replacement policy of a two-colour urn is concisely described by a 2×2-generating matrix R of the form

$$R := \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \in \mathbb{Z}^{2\times 2}. \tag{1.1}$$

Our focus is on the urn composition after a large number of draws. As in the foreword, we denote the number of type 1 balls after $n \in \mathbb{N}$ draws from the urn as $X_n^{(1)}$ and the number of type 2 balls as $X_n^{(2)}$. Further let

$$Y_n := X_n - \mathbb{E}[X_n] \tag{1.2}$$

denote the centered composition vector.

Example 1 (Pólya Urn)**.** We have already mentioned the famous Pólya urn at various points. It was introduced by the Hungarian mathematician George Pólya in [20] as a model of contagion and is further studied in [65]. The model in its most basic form is an urn scheme with two types and generating matrix

$$R_{\text{Pólya}} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Here, one can imagine a population with two kinds of diseases that uniformly comes into contact with external elements to which their disease is transmitted. In this sense, the balls pass on their disease, and a process of contagion unfolds. Dumas, Flajolet and Puyhaubert [24] also call it the "purely autistic urn", as there is no interaction between the two types.
Now, it is an interesting question how the initial proportions of diseases of both kinds evolve over time. In Theorem 1.1.1, we will see that the urn composition at time n is highly sensitive to what happened in the first draws and varies widely. In fact, the Pólya urn provides a simple

model of a stochastic process in which initial imbalances do not even out, but rather magnify over time. ♢

So far, the definition of the model is too general to exclude pathological behaviour and does not even ensure that the process is well-defined at all times. Throughout this thesis, we additionally work under the following set of assumptions (here formulated for urns with two types):

(A1) R is diagonisable over $\mathbb{C}$.

(A2) R has constant positive column sum $a_{1,1} + a_{2,1} = a_{1,2} + a_{2,2} =: r \geq 1$.

(A3) $a_{2,1}, a_{1,2} \geq 0$. Whenever $a_{1,1} < 0$, $a_{1,1}$ divides $X_0^{(1)}$ and $a_{1,2}$. Whenever $a_{2,2} < 0$, $a_{2,2}$ divides $X_0^{(2)}$ and $a_{2,1}$.

(A4) $a_{1,1} \neq a_{2,1}$.

(A5) The initial composition of the urn is such that for $j \in \{1, 2\}$, there exists $n \in \mathbb{N}_0$ with $\mathbb{P}\left(X_n^{(j)} > 0\right) > 0$.

Assumption (A1) admits simpler proofs. Under assumption (A2), the total number of balls grows at a linear rate of increase, independent of the particular history of draws. The number r is also called the *balance* of the urn. It is necessary for the use of the particular martingale methods used in the proofs of the present work, for example. Correspondingly, in this thesis, we will only be concerned with balanced urns. (For a theory of non-balanced urns, see [40]. Kotz, Mahmoud and Robert [48] also study general urns with two types, non-negative integer valued entries and deterministic replacement, but do not assume the generating matrix to be balanced.) To continue, (A3) is needed to ensure that the process is well-defined for all $n \in \mathbb{N}$ and does not have to stop because it is asked for an impossible removal of balls. Finally, (A4) and (A5) serve the purpose to exclude a deterministic evolution of the urn. It has to be mentioned that all matrices R that satisfy (A2) to (A4) are diagonisable. (A1) is only listed here to prepare the general case.
Note however that the exact formulation of the setting is not unified in the literature on urn models and that it has to be checked whether two models are comparable.

Example 2 (2-3 tree urn)**.** 2-3 trees are data structures that were introduced in the beginning of the 1970s by Rudolf Bayer and Edward M. McCreight [5, 6]. They are an example for the so-called B-trees (see chapter 4), and a special feature of these search trees is that all their leaves are at the same depth. The evolution of the external nodes of the two possible different types can be studied via an urn process with generating matrix

$$R_{2-3} := \begin{pmatrix} -2 & 4 \\ 3 & -3 \end{pmatrix}.$$

Note that, after an admissible choice of initial configuration, this urn model fits into our setting. Bagchi and Pal [4] apply their results to get asymptotics for the mean and the variance of the number of type 1 balls after a large number of draws. By means of these asymptotics, they derive upper and lower bounds for the number of internal nodes in the tree. ♢

Example 3 (Ehrenfest Urn)**.** This model was introduced by mathematician Tatiana Ehrenfest and physicist Paul Ehrenfest in 1907 in order to illustrate their answer to two objections to Boltzmann's H-Theorem. In [21], they consider N labelled particles which are distributed among two chambers. At each time step, one of the labels is chosen uniformly at random and the corresponding particle changes its chamber. If one is only interested in the number of particles in each chamber, this migration process can be modelled by an urn with matrix

$$R_{\text{Ehren}} := \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

However, this example is not of the type considered in the current text, as it does not satisfy (A2) and the number of particles does not increase with time, but stays constant. ♢

Let us now introduce a term that plays an important role in the study of urn models and takes a particular simple form for two-colour urns. We call a two-colour urn *irreducible*, if $a_{1,2}a_{2,1} > 0$. Many important examples satisfy this assumption, which is made by the majority of studies on urn models. On the other hand, two-colour urns with $a_{1,2}a_{2,1} = 0$ are called *reducible* or *triangular*.
It has been known for a long time that irreducible and reducible urns exhibit a different behaviour and that reducible urns are somehow less accessible. For a thorough study of triangular urns (that are not assumed to be balanced), see [42]. Note that in this thesis, there is no assumption on irreducibility. In the current section, we will assume that all reducible urns are of the form

$$\begin{pmatrix} a_{1,1} & 0 \\ a_{2,1} & r \end{pmatrix},$$

i.e., $a_{1,2} = 0$ and $a_{2,1} \geq 0$. Dumas, Flajolet and Puyhaubert [24] provide a witty interpretation of triangular urns as a model for evolving species: Imagine that there are two populations, for example apes and humans, which reproduce and may give birth at each time n. The individual that gives birth at time n is chosen uniformly among all individuals. Each time an individual is chosen, it produces the same number of r offspring. However, humans can only give birth to humans, while there might also be $a_{2,1}$ humans among the offspring of apes as a result of evolution. The authors make the concession that this might be a too optimistic model of reality.

Variants of the model. In the literature on urn models, there are many variants of the model presented here and also extensions in different directions. Of course, one always has to provide assumptions to ensure tenability of the urn (except for the classical sampling urns) and to exclude deterministic behaviour. Furthermore, in the majority of results, balancedness is assumed and restrictions are put on the spectrum of the generating matrix. More precisely, it is assumed that the largest eigenvalue is real and simple (this can be assured by allowing only non-negative off-diagonal entries and assuming irreducibility and balance, for example) and that the real parts of all other eigenvalues are at most half this largest eigenvalue.
However, some variants of the model include real-valued instead of integer-valued replacement or negative off-diagonal entries under another sufficient tenability-condition. It is also possible to introduce a second instance of randomness by choosing the increments $(a_{1,1}, a_{2,1})^t$ and $(a_{1,2}, a_{2,2})^t$ randomly, according to some fixed probability distributions μ and ν, respectively.

Another recently studied model deals with multiple drawings rather than a single draw from the urn (with or without replacement), see [49]. Note that in particular, the results do not depend on the assumption that $a_{1,1}, a_{2,1}, a_{1,2}, a_{2,2}$ are integers. This convention only serves the purpose to use a language involving balls. For a discussion of different hypotheses, we refer the reader to [40].

1.1.2 Convergence of proportions

In the hold of (A1) - (A5), the convergence of the proportions $\frac{X_n^{(1)}}{rn+|X_0|}$ and $\frac{X_n^{(2)}}{rn+|X_0|}$ is well-understood. There are basically two cases of a different character: The first is the Pólya urn, while all other urns in the given setting can be summarised to a second category.

We start with the Pólya urn. The following results are a "classical part of the oral tradition" [25], and can be found in [25], for example. Note that the model below is slightly generalised in comparison to example 1.

Theorem 1.1.1 (Pólya urn). *Consider a Pólya urn scheme with generating matrix*

$$\begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}$$

and assume that $X_0^{(1)}, X_0^{(2)} > 0$. *Then, as* $n \to \infty$,

$$\begin{pmatrix} \frac{X_n^{(1)}}{rn+|X_0|} \\ \frac{X_n^{(2)}}{rn+|X_0|} \end{pmatrix} \xrightarrow{a.s.} \begin{pmatrix} Z \\ 1-Z \end{pmatrix}, \tag{1.3}$$

where $\mathcal{L}(Z) = Beta\left(\frac{X_0^{(1)}}{r}, \frac{X_0^{(2)}}{r}\right)$. *Furthermore, the tail* σ*-field of* $\left(\left(X_n^{(1)}, X_n^{(2)}\right)\right)_{n\geq 0}$ *is equivalent to the* σ*-field spanned by* Z. *Given this* σ*-field, the outcomes of the single draws from the urn form an i.i.d. sequence, with each outcome being type* 1 *with probability* Z, *and type* 2 *with probability* $1-Z$.

Discussion. In particular, if both $X_0^{(1)}, X_0^{(2)} > 0$, the limit in the theorem is random and depends on the particular sequence of draws. As there is no "smoothing" between the colours, the dynamics of the urn are such that initial imbalances rather magnify over time than even out: For example, if $X_0^{(1)}$ is much bigger than $X_0^{(2)}$, it is more likely to draw a type 1 ball in the first step and to further increase the number of type 1 balls. On the other hand, if $X_0^{(1)}$ and $X_0^{(2)}$ are roughly of the same size, but r is comparatively big, then the first few draws will turn the proportion of balls in favour of one colour, which consequently is more likely to be drawn in the following steps and to produce even more balls of its own kind. To summarise, the fact that the urn's asymptotics most notably depend on the initial configuration and the first draw, slightly less on the second draw, even less on the third draw, and so on, leads to instability, or, randomness, in the long-term behavior of the system.

We now turn to the second case. The result in the form stated is due to Gouet [28], who derives strong asymptotics for the composition vector of two-colour urn models with

$\max\{a_{2,1}, a_{1,2}\} > 0$.

Theorem 1.1.2. *For all Pólya-Eggenberger urn schemes with replacement matrix* R *such that* $\max\{a_{2,1}, a_{1,2}\} > 0$ *and (A1)-(A5) hold,*

$$\frac{X_n}{rn + |X_0|} \xrightarrow{a.s.} \frac{1}{a_{2,1} + a_{1,2}} \begin{pmatrix} a_{1,2} \\ a_{2,1} \end{pmatrix}, \qquad n \to \infty. \tag{1.4}$$

Note that in contrast to the asymptotics of the Pólya urn, the ball proportions in all other models in our setting converge to deterministic limits. Furthermore, in triangular urn schemes, the proportion of type 1 balls tends to zero almost surely, irrespective of the initial configuration. This observation concludes the exposition of the two-type case.

1.1.3 Central limit theorem

In a next step, we turn to central limit theorems for two-colour urn models under assumptions (A1) to (A5). As previously mentioned, it is well-known that the *spectral properties* of the generating matrix R (or, equivalently, its transpose R^t, the so-called replacement matrix) determine the urn's probabilistic behaviour to a large extent. Therefore, we now define the relevant quantities for two-colour urns, even though this will be repeated in a more general setting later.
To begin with, the generating matrix R has two *real* eigenvalues, which are $a_{1,1} + a_{1,2} = r$ and $a_{1,1} - a_{2,1} \leq r$. For all urns different from the Pólya urn, these are distinct. Flajolet et al. [24] call $a_{2,1} - a_{1,1} = -(a_{1,1} - a_{2,1})$ the dissimilarity index and introduce three categories:

(1) *Altruistic urns* correspond to $a_{1,1} - a_{2,1} < 0$. Both colours get more reinforcement from the other colour than from their own colour.

(2) *Neutral urns* correspond to $a_{1,1} - a_{2,1} = 0$. The composition of the urn at any instance is completely determined and a linear function of time. (A4) excludes this case from the present setting.

(3) *Selfish Urns* correspond to $a_{1,1} - a_{2,1} > 0$. Both colours get more reinforcement from their own colour than from the other colour.

It turns out that their distinction between altruistic and selfish urns at $a_{1,1} - a_{2,1} = 0$ is not the one at which the transition from asymptotic normality to almost sure drift occurs (which is $a_{1,1} - a_{2,1} = \frac{a_{1,1}+a_{1,2}}{2}$), but it provides a vivid way to think about the underlying parameter.

The spectral decomposition of the urn process into projections onto its eigenspaces will play a central role in this thesis. There are dual bases $\{u_1^t, u_2^t\}$ and $\{v_1, v_2\}$ of left, respectively right, eigenvectors associated to the two eigenvalues. If $\max\{a_{1,2}, a_{2,1}\} > 0$, we choose left and right eigenvectors u_1, u_2 and v_1, v_2 corresponding to r and $a_{1,1} - a_{2,1}$, respectively, as

$$u_1 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad u_2 := \begin{pmatrix} a_{2,1} \\ -a_{1,2} \end{pmatrix}, \quad v_1 := \frac{1}{a_{1,2} + a_{2,1}} \begin{pmatrix} a_{1,2} \\ a_{2,1} \end{pmatrix}, \quad v_2 := \frac{1}{a_{1,2} + a_{2,1}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{1.5}$$

If $a_{1,2} = a_{2,1} = 0$, we choose

$$u_1 := v_1 =: \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u_2 := v_2 =: \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

With this choice, each vector $x = (x_1, x_2)^t \in \mathbb{R}^2$ can be written as

$$x = (u_1^t x)v_1 + (u_2^t x)v_2.$$

Again, the Pólya urn with $a_{1,2} = a_{2,1} = 0$ has to be treated separately. But if we apply the decomposition to an urn process $(X_n)_{n\geq 0}$ with $\max\{a_{1,2}, a_{2,1}\} > 0$, we see that for all $n \geq 0$, the first coefficient $u_1^t X_n = rn + |X_0|$ is deterministic. Thus,

$$X_n - \mathbb{E}[X_n] = u_2^t(X_n - \mathbb{E}[X_n])v_2$$

and all the randomness of the evolution is contained in the projection $u_2^t X_n$. This projection measures the order of magnitude of the process along the direction of v_2. According to Theorem 1.1.2, it is smaller than linear. But how big is it?
It is this second order term which is of importance in the study of CLTs. It has been known since the 1960s that the basic parameter in the derivation of CLTs for urn models is the eigenvalue ratio $\frac{a_{1,1}-a_{2,1}}{r}$ and that there is a qualitative difference between models with $\frac{a_{1,1}-a_{2,1}}{r} \leq \frac{1}{2}$ and with $\frac{a_{1,1}-a_{2,1}}{r} > \frac{1}{2}$. Because the process is not deterministic, $\frac{a_{1,1}-a_{2,1}}{r} \neq 0$, and as $a_{1,2}, a_{2,1} \geq 0$, $\frac{a_{1,1}-a_{2,1}}{r} \leq 1$. On the other hand, unbounded negative values of $\frac{a_{1,1}-a_{2,1}}{r}$ are possible. Returning to our question from the previous paragraph, it turns out that if $\frac{1}{2} < \frac{a_{1,1}-a_{2,1}}{r} < 1$, $u_2^t(X_n - \mathbb{E}[X_n])$ converges almost surely and in L^2 as $n \to \infty$ (see Lemma 3.1.2). We may thus define Ξ as the almost sure limit

$$\Xi := \lim_{n\to\infty} \left(n^{-\frac{a_{1,1}-a_{2,1}}{r}} u_2^t \left(X_n - \mathbb{E}[X_n]\right) \right). \tag{1.6}$$

For more information on Ξ, see [51, 57] or chapter 3. The proceeding in the formulation of a CLT for two-colour urns is now exactly as described in the introduction: For $\frac{a_{1,1}-a_{2,1}}{r} \leq \frac{1}{2}$, there is no almost sure limit Ξ and we are in the classical normalisation of a CLT. However, if $\frac{a_{1,1}-a_{2,1}}{r} > \frac{1}{2}$, a random centering has to be taken into account.
Before we formulate the CLT, let us make the formal introduction of the eigenvalue ratio in the case $a_{1,1} - a_{2,1} > 0$ more plausible by the following simple trick. Write

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = \begin{pmatrix} a_{2,1} & a_{2,1} \\ a_{1,2} & a_{1,2} \end{pmatrix} + \begin{pmatrix} a_{1,1} - a_{2,1} & 0 \\ 0 & a_{1,1} - a_{2,1} \end{pmatrix}.$$

The first matrix in the decomposition can be interpreted as the invariable drift by which the urn composition increases in each step, while the Pólya-like matrix accounts for the balls that deviate from this drift in one of the possible two colours. In this matrix, the amount is fixed, but the coordinate direction is random. We now see that the eigenvalue ratio gives the asymptotic proportion of balls deviating from the deterministic direction: As $a_{1,2} + a_{2,1} = (a_{1,1} + a_{1,2}) - (a_{1,1} - a_{2,1})$, the total increase $a_{1,2} + a_{2,1}$ in the first part is greater or equal than the increase $a_{1,1} - a_{2,1}$ in the second part if and only if $\frac{a_{1,1}-a_{2,1}}{a_{1,1}+a_{1,2}} \leq \frac{1}{2}$. If, on the contrary, $\frac{a_{1,1}-a_{2,1}}{a_{1,1}+a_{1,2}} > \frac{1}{2}$, then asymptotically, more than half of the balls in the urn are a product of the deviation, and it becomes self-reinforcing to a significant amount, leading to an almost sure term.
We now state the CLT for two-colour urns, which is a special case of Theorem 1.2.5. To this

end, for $\max\{a_{1,2}, a_{2,1}\} > 0$, set

$$\Sigma_{2-\mathrm{col}} := \frac{a_{1,2}a_{2,1}r(a_{1,1}-a_{2,1})^2}{|2(a_{1,1}-a_{2,1})-r|}v_2v_2^t = \frac{a_{1,2}a_{2,1}r(a_{1,1}-a_{2,1})^2}{(a_{1,2}+a_{2,1})^2|2(a_{1,1}-a_{2,1})-r|}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}. \quad (1.7)$$

Proposition 1.1.3. *Let* $(X_n)_{n\geq 0}$ *be a generalised Pólya-Eggenberger urn process with matrix (1.1) that satisfies conditions (A1)-(A5).*

(i) If $a_{1,2} = a_{2,1} = 0$*, as* $n \to \infty$*,*

$$\frac{1}{\sqrt{r^2 nZ(1-Z)}}\left(X_n - nr\begin{pmatrix} Z\\ 1-Z\end{pmatrix}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right), \quad (1.8)$$

where Z *has Beta*$\left(\frac{X_0^{(1)}}{r}, \frac{X_0^{(2)}}{r}\right)$*-distribution and is as in (1.3).*

(ii) If $\max\{a_{1,2}, a_{2,1}\} > 0$ *and* $2(a_{1,1}-a_{2,1}) > r$*, as* $n \to \infty$*,*

$$\frac{1}{\sqrt{n}}\left(X_n - \mathbb{E}[X_n] - n^{\frac{a_{1,1}-a_{2,1}}{r}}\Xi\frac{1}{a_{1,2}+a_{2,1}}\begin{pmatrix}1\\ -1\end{pmatrix}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \Sigma_{2-col}\right). \quad (1.9)$$

(iii) If $\max\{a_{1,2}, a_{2,1}\} > 0$ *and* $2(a_{1,1}-a_{2,1}) = r$*, as* $n \to \infty$*,*

$$\frac{1}{\sqrt{n\log(n)}}(X_n - \mathbb{E}[X_n]) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \frac{a_{1,2}a_{2,1}(a_{1,1}-a_{2,1})^2}{(a_{1,2}+a_{2,1})^2}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right). \quad (1.10)$$

(iv) If $\max\{a_{1,2}, a_{2,1}\} > 0$ *and* $2(a_{1,1}-a_{2,1}) < r$*, as* $n \to \infty$*,*

$$\frac{1}{\sqrt{n}}(X_n - \mathbb{E}[X_n]) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \Sigma_{2-col}\right). \quad (1.11)$$

Remark 1 (A short note on what is known in Proposition 1.1.3)**.** Part (i) is basically a result of Hall and Heyde [34], p. 80. Parts (iii) and (iv) in Proposition 1.1.3 are well-known and covered in [74] and [40], for example. Proofs of the almost sure or distributional convergence of $n^{-a_{1,1}/a_{2,1}}X_n^{(1)}$ can be found in [8, 40]. However, the central limit theorems in Proposition 1.1.3 (ii) and in Proposition 1.1.4 appear to be new.

Remark 2. Note that only in case (i), where the asymptotic proportions of balls are random, a random centering *and* scaling are necessary. In (ii), only the centering is random, as the second largest eigenvalue is bigger than the critical value $\frac{r}{2}$. In the multicolour case, we will observe the same pattern.
Also note that in cases (ii)-(iv), the asymptotic probability to draw a ball of type 1 is $a_{2,1}/(a_{1,2}+a_{2,1})$ (say, for $a_{2,1} > 0$). However, the asymptotic variance is $\frac{a_{1,2}a_{2,1}r(a_{1,1}-a_{2,1})^2}{(a_{1,2}+a_{2,1})^2|2(a_{1,1}-a_{2,1})-r|}$ or $\frac{a_{1,2}a_{2,1}(a_{1,1}-a_{2,1})^2}{(a_{1,2}+a_{2,1})^2}$ rather than $a_{1,2}a_{2,1}/(a_{1,2}+a_{2,1})^2$, and we see a difference between the

urn and a coin-tossing game with success probability $a_{2,1}/(a_{1,2}+a_{2,1})$. This is also different from the behaviour in (i), where the sequence of draws of type 1 balls can be regarded as a sequence of coin tosses with success probability Z, conditionally on Z.

Note that the variance of the weak limit in Theorem 1.2.5 is zero for triangular urns, i.e. if $a_{1,2}a_{2,1}=0$. In this case, the theorem yields a degenerate limit. This is due to the fact that the $\sqrt{n}$-scaling is too strong, as $X_n - \mathbb{E}[X_n]$ is governed by the behaviour of $X_n^{(1)}$, which is of sublinear size $n^{\frac{a_{1,1}}{a_{2,2}}}$ in this case. This is an observation that also holds in the multicolour case: If R is reducible, the asymptotic covariance matrix Σ is typically more singular. The following theorem gives the correct scaling for triangular two-colour urns. Its proof is given in chapter 3.

Proposition 1.1.4. *Let $(X_n)_{n\geq 0}$ be a generalised Pólya-Eggenberger urn process with triangular matrix (i.e. $a_{1,2}=0$) that satisfies conditions (A1)-(A5) and additionally, $a_{1,1}>0$. Then as $n\to\infty$,*

$$\frac{1}{\sqrt{n^{\frac{a_{1,1}}{a_{2,2}}}}}\left(X_n - \mathbb{E}[X_n] - n^{\frac{a_{1,1}}{a_{2,2}}}\Xi v_2\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \frac{a_{1,1}X_0^{(1)}\Gamma\left(\frac{|X_0|}{a_{2,2}}\right)}{\Gamma\left(\frac{|X_0|+a_{1,1}}{a_{2,2}}\right)}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right). \tag{1.12}$$

Remark 3. The case $a_{1,1}=0$ is excluded by (A4) and leads to a deterministic evolution of the urn without fluctuation. Furthermore, the case $a_{1,1}<0$ is also immediate, as the balls of colour 1 die out almost surely.

1.1.4 Examples

Below, we consider two examples which both illustrate parts (ii)-(iv) of Proposition 1.1.3. Each of the models is defined by a parameter, leading to a phase transition from (ii) to (iv).

Example 4 (Friedman urn)**.** This urn model, which we have already met in the introduction, is a natural generalisation of Pólya's model. Flajolet et al. [24] describe it as a "simple matrix with many hidden treasures". In its most general form, the generating matrix of the model is given by

$$R_{\text{Friedman}} = \begin{pmatrix} a_{1,1} & a_{1,2}\\ a_{1,2} & a_{1,1}\end{pmatrix},$$

where $a_{1,1}, a_{1,2}\in\mathbb{Z}$, $a_{1,2}\geq 0$, and $a_{1,1}+a_{1,2}=r>0$. If $a_{1,1}=0$ or $a_{1,1}$ is much smaller than $a_{1,2}$, the Friedman urn can be viewed as a model of a propaganda campaign in which candidates are so bad that the persons who listen to them are convinced to vote for the opposite candidate (compare the examples in the introduction).
Even though the model includes the Pólya urn, its behaviour for $a_{1,2}>0$ is very different. Indeed, any large enough deviation of the urn's composition from proportions $(1/2, 1/2)$ tends to correct itself and bring the system back to equilibrium, and accordingly, the composition of the urn is concentrated around its mean value.
The eigenvalues of the symmetric matrix R_{Friedman} are $a_{1,1}+a_{1,2}$ and $a_{1,1}-a_{1,2}$. For $a_{1,2}=0$, we get the original Pólya urn which yields the first case of Proposition 1.1.3. For $a_{1,2}>0$, $a_{1,1}+a_{1,2}$ is a simple eigenvalue and thus no random scaling arises in the limit theorem. Here,

the eigenvectors in (1.5) take the particular form

$$u_1 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad u_2 := \begin{pmatrix} a_{1,2} \\ -a_{1,2} \end{pmatrix}, \quad v_1 := \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 := \frac{1}{2a_{1,2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

For $a_{1,2} > 0$, the behaviour of the urn ranges over the three cases (ii) to (iv) in Proposition 1.1.3: If $a_{1,1} < 3a_{1,2}$, then $\frac{a_{1,1}-a_{1,2}}{a_{1,1}+a_{1,2}} < \frac{1}{2}$, and we are in case (iv) of Proposition 1.1.3. This yields

$$\frac{X_n - \mathbb{E}[X_n]}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{(a_{1,1}+a_{1,2})(a_{1,1}-a_{1,2})^2}{4(3a_{1,2}-a_{1,1})}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right)$$

as $n \to \infty$.
If $a_{1,1} = 3a_{1,2}$, then $\frac{a_{1,1}-a_{1,2}}{a_{1,1}+a_{1,2}} = \frac{1}{2}$, and case (iii) leads to

$$\frac{X_n - \mathbb{E}[X_n]}{\sqrt{n\log(n)}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{(a_{1,1}-a_{1,2})^2}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right),$$

as $n \to \infty$. Here, we have recovered the results of example 3.27 in [40].
If $a_{1,1} > 3a_{1,2}$, then $\frac{a_{1,1}-a_{1,2}}{a_{1,1}+a_{1,2}} > \frac{1}{2}$, and so

$$\frac{X_n - \mathbb{E}[X_n] - n^{\frac{a_{1,1}-a_{1,2}}{a_{1,1}+a_{1,2}}}\Xi v_2}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{(a_{1,1}+a_{1,2})(a_{1,1}-a_{1,2})^2}{4(a_{1,1}-3a_{1,2})}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right)$$

as $n \to \infty$. As above, Ξ_2 is the almost sure limit of $\left(bn^{-\frac{a_{1,1}-a_{1,2}}{a_{1,1}+a_{1,2}}}\left(Y_n^{(1)} - Y_n^{(2)}\right)\right)_{n\geq 1}$. ◊

Example 5 (Generalised Binomial Distribution)**.** Drezner and Farnum [18] introduce the following generalisation of the Binomial distribution as the number of successes in n independent, identically distributed Bernoulli experiments: As usual, let $0 < p < 1$ be a fixed success probability in a Bernoulli experiment. Additionally, consider a fixed correlation parameter $0 \leq \vartheta < 1$. We call $(B_n)_{n\geq 1}$ an *adaptive* Bernoulli process with parameters p and ϑ, if $\mathcal{L}(B_1) = \text{Bern}(p)$ and for $n \in \mathbb{N}$,

$$\mathbb{P}(B_{n+1} = 1 | B_1, \ldots, B_n) = (1-\vartheta)p + \vartheta\frac{S_n}{n}.$$

In the above, $S_n := \sum_{j=1}^{n} B_j$ is the relative frequency of successes among the first n experiments. It is clear from this definition that the model is conceived such that a big number of past successes (more precisely, $S_n > np$) leads to an increase of the probability of future successes. The size of this bias is parametrised by ϑ. For example, for all $n \in \mathbb{N}$, the Binomial distribution with parameters n and p can be regained as the distribution of B_n with the choice $\vartheta = 0$.

In [18] it is shown that for all $n \geq 1$, $\mathbb{E}[S_n] = np$ and

$$\text{Var}(S_n) \sim \begin{cases} \frac{np(1-p)}{1-2\vartheta}, & \vartheta < \frac{1}{2}, \\ p(1-p)n\log(n), & \vartheta = \frac{1}{2}, \\ \frac{n^{2\vartheta}p(1-p)}{(2\vartheta-1)\Gamma(2\vartheta)}, & \vartheta > \frac{1}{2} \end{cases}$$

as $n \to \infty$. This expansion of the variance strongly resembles the phase transitions in our urn models.
Note that by Chebyshev's inequality, $\frac{S_n}{n}$ converges in probability to p as $n \to \infty$, irrespective of the value of ϑ. Heyde [35] studies the rate of convergence of $\frac{S_n}{n}$ to p via martingale techniques and derives the following theorem:

Theorem 1.1.5 (Heyde). *(i) If $\vartheta < \frac{1}{2}$, then $n^{-1/2}(S_n - np) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{p(1-p)}{1-2\vartheta}\right)$ as $n \to \infty$.*

(ii) If $\vartheta = \frac{1}{2}$, then $(n \log(n))^{-1/2}(S_n - np) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1-p))$ as $n \to \infty$.

(iii) If $\vartheta > \frac{1}{2}$, then $n^{-\vartheta}(S_n - np) \xrightarrow{a.s.} W$ as $n \to \infty$, where W is a non-degenerate, centered random variable with

$$\mathbb{E}\left[W^2\right] = \frac{p(1-p)}{(2\vartheta - 1)\Gamma(2\vartheta)}.$$

Now, this result can be recovered and complemented by Proposition 1.1.3. Rarivoarimanana [69] observes that the correlated Bernoulli process can, in fact, be translated into an urn model. This urn is not integer-valued, but as noted before, this is of no importance.
Given an an *adaptive* Bernoulli process $(B_n)_{n\geq 1}$, set $X_0 := (p, 1-p)^t$ and for $n \geq 1$,

$$X_n := \begin{pmatrix} n(1-\vartheta)p + \vartheta S_n \\ n(1-\vartheta)(1-p) + \vartheta(n - S_n) \end{pmatrix}.$$

Then $(X_n)_{n\geq 0}$ is a two-colour urn model with generating matrix

$$R_{\text{Gen.Bin.}} = \begin{pmatrix} (1-\vartheta)p + \vartheta & (1-\vartheta)p \\ (1-\vartheta)(1-p) & (1-\vartheta)(1-p) + \vartheta \end{pmatrix}.$$

Conversely, $(S_n)_{n\geq 1} = \left(\frac{1}{\vartheta}\left(X_n^{(1)} - n(1-\vartheta)p\right)\right)_{n\geq 1}$.

$R_{\text{Gen.Bin.}}$ has eigenvalues 1 and $\vartheta < 1$. In this example,

$$u_1 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad u_2 := (1-\vartheta)\begin{pmatrix} p \\ -(1-p) \end{pmatrix}, \quad v_1 := \begin{pmatrix} 1-p \\ p \end{pmatrix}, \quad v_2 := \frac{1}{1-\vartheta}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We only consider the case $\vartheta > \frac{1}{2}$, which is the most interesting part from our point of view. However, the other two cases follow analogously. Proposition 1.1.3 implies that for $\vartheta > \frac{1}{2}$,

$$\frac{X_n - \mathbb{E}[X_n] - n^{\vartheta}\Xi v_2}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\vartheta^2 p(1-p)}{2\vartheta - 1}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right)$$

as $n \to \infty$. Here, Ξ is the almost sure limit of $n^{-\vartheta}u_2^t(X_n - \mathbb{E}[X_n]) = n^{-\vartheta}(1-\vartheta)\left(X_n^{(1)} - \mathbb{E}\left[X_n^{(1)}\right]\right)$.
In particular, $\vartheta(1-\vartheta)\Xi = W$.
This in turn implies for the correlated Bernoulli model that as $n \to \infty$,

$$\frac{S_n - np - n^{\vartheta}W}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{p(1-p)}{2\vartheta - 1}\right).$$

Once more, note the symmetry of the cases (i) in Heyde's theorem and the above. ♢

1.2 Multicolour urn models

In this section, we turn our attention to urn models that involve more than two colours. To this end, recall the definition of a generalised Pólya-Eggenberger urn scheme with $q \geq 2$ colours from the beginning of the introduction. At this earlier point, we did not go into detail about our assumptions and their implications. This omission has to be rectified in the present section.

Note that the dynamics of the urn composition process $(X_n)_{n\geq 0}$ are fully described by its initial value X_0 and the generating matrix R. Our basic assumptions on R and X_0 are the following:

(A1) R is diagonisable over $\mathbb{C}$.

(A2) R has constant column sum $r \geq 1$.

(A3) $R_{i,j} \geq 0$ for $i \neq j$ and if $R_{i,i} < 0$, then $|R_{i,i}|$ divides $X_0^{(i)}$ and $R_{i,j}$ for all $1 \leq j \leq q$.

(A4) No two columns of R are identical.

(A5) The initial composition of the urn is such that for all colours j, there exists $n \in \mathbb{N}_0$ with $\mathbb{P}\left(X_n^{(j)} > 0\right) > 0$.

(A1) to (A5) are satisfied in most applications. As in the two-colour case, (A2) guarantees a steady linear growth and (A3) ensures tenability of the urn, i.e., that the process does not get stuck due to a lack of balls. (A4) and (A5) prevent an easy reduction to an urn model with less colours.

1.2.1 The ordering of types

Before we go into detail about the spectral properties and the eigenvalues of R, let us first introduce some consistency to the ordering of the colours. For more structural overview, we write $i \to j$ and say that colour i *leads* to colour j if, starting with one ball of colour i, we have $\mathbb{P}\left(X_n^{(j)} > 0\right) > 0$ for some $n \in \mathbb{N}_0$. Equivalently, $(R^n)_{j,i} > 0$. We say that i and j *communicate* and write $i \leftrightarrow j$ if $i \to j$ and $j \to i$. The equivalence relation $\leftrightarrow$ partitions the set $\{1, \ldots, q\}$ of colours into d equivalence classes $\mathcal{C}_1, \ldots, \mathcal{C}_d$. If $d = 1$, the process is called *irreducible* (this generalises the definition for $q = 2$ in the last section). Otherwise, it is called *reducible.*
The case when the replacement matrix is irreducible is important and includes many applications. Limit theorems for the irreducible case have been given by many authors, see for example [2], [3], [4] or [40] and the references therein. It has been known for a long time that other phenomena arise in the reducible case. Here, we do not assume irreducibility. But somehow, the picture given by Theorem 1.2.5 is too rough if applied to some reducible cases.

Example 6 (Triangular 3×3 Urns)**.** Consider a generating matrix of the form

$$R_{\text{Triangular}} = \begin{pmatrix} a_{1,1} & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 \\ r - a_{1,1} - a_{2,1} & r - a_{2,2} & r \end{pmatrix},$$

where $a_{1,1}a_{2,1}a_{2,2} \neq 0$. It is easy to see that the corresponding urn model is reducible, as there are three equivalence classes, each consisting of a single type. Similar to triangular two-colour urns, this urn can be regarded as a simple model for evolving species. However, note that it is not diagonisable for $a_{1,1} = a_{2,2}$. $\diamondsuit$

Furthermore, we write $\mathcal{C}_i \to \mathcal{C}_j$ and say that class $\mathcal{C}_i$ leads to class $\mathcal{C}_j$ if some (then all) colours in $\mathcal{C}_i$ lead to some (then all) colours in $\mathcal{C}_j$. We call a class *dominant* if it does not lead to any other class except possibly itself.
With this, the classes $\mathcal{C}_1, \dots, \mathcal{C}_d$ decompose into three different categories. Class $\mathcal{C}_i$ is of category 1 if it is dominant and there is no $j \neq i$ with $\mathcal{C}_j \to \mathcal{C}_i$. Any dominant class which is not of category 1 is of category 2. All non dominant classes are of category 3. Note that (A4) implies that in particular, we start with at least one ball from each class of category 1. Subsequently, we may assume that the classes are ordered as follows: with $a, b, c \geq 0$ and $a + b + c = d$, classes $\mathcal{C}_1, \dots, \mathcal{C}_a$ are of category 1, classes $\mathcal{C}_{a+1}, \dots, \mathcal{C}_{a+c}$ are of category 2, and the remaining classes $\mathcal{C}_{a+c+1}, \dots, \mathcal{C}_d$ are of the third category. $\mathcal{C}_{a+c+1}, \dots, \mathcal{C}_d$ are ordered such that $\mathcal{C}_i \to \mathcal{C}_j$ implies $i \leq j$. Note that $a + c \geq 1$. Under this ordering of the colours, the matrix R has the following lower triangular block structure:

$$R = \begin{pmatrix} T_{1,1} & & & & & & & & \\ & \ddots & & & & & & & \\ & & T_{a,a} & & & & & & \\ & & & P_{1,1} & & & & & \\ & & & * & \ddots & & & & \\ & & & * & * & P_{b,b} & & & \\ & & & * & * & * & Q_{1,1} & & \\ & & & * & * & * & & \ddots & \\ & & & * & * & * & & & Q_{c,c} \end{pmatrix}.$$

Here, blocks $T_{1,1}, \dots, T_{a,a}$ correspond to category 1 classes $\mathcal{C}_1, \dots, \mathcal{C}_a$, blocks $Q_{1,1}, \dots, Q_{c,c}$ to category 2 classes $\mathcal{C}_{a+1}, \dots, \mathcal{C}_{a+c}$ and blocks $P_{1,1}, \dots, P_{b,b}$ to category 3 classes $\mathcal{C}_{a+c+1}, \dots, \mathcal{C}_d$. The middle part corresponding to category 3 classes is a lower triangular block matrix in which beneath each of the blocks $P_{1,1}, \dots, P_{b,b}$, there is at least one positive entry. Similarly, to the left of each block $Q_{1,1}, \dots, Q_{c,c}$, there is at least one positive entry. Otherwise, the ordering would be a different one.

1.2.2 The spectrum of R

As the transitions of $(X_n)_{n \geq 0}$ are governed by the matrix R and multiplication by R is just multiplication by a scalar if restricted to each of its eigenspaces, the spectral decomposition of $(X_n)_{n \geq 0}$ with respect to the eigenspaces of R should give a good way of finding out about components of different sizes in the asymptotics of the process. In this sense, the choice of a convenient basis of $\mathbb{C}^q$ is crucial in the following.
Indeed, with respect to central limit theorems, it is well known that irreducible, balanced and tenable Pólya urn processes undergo a phase transition that is caused by the spectral properties of their generating matrix. This change can also be seen in the variance of the urn composition vector. More precisely, their asymptotics depend on the relation of the real parts of the largest and the second largest eigenvalue. Call these σ_1 and σ_2, respectively. When $\sigma_2 < \sigma_1/2$ or $\sigma_2 = \sigma_1/2$, the urn often is called "strictly small" or "small" and there are Gaussian fluctuations. On the other hand, if $\sigma_2 > \sigma_1/2$, the urn is called "large" and either

has a second order convergent term (if the second largest eigenvalue is real) or an almost sure oscillating behaviour (if the second largest eigenvalue is complex). This observation is already made in [3]. Some simulations of this phenomenon in the context of B-urns can be found in [11].
In this section, we will see that for large urns, the CLT in fact depends on *all* eigenvalues that have real parts greater than $\sigma_1/2$.

Matrices with non-negative off-diagonal entries are called *Metzler-Leontief matrices*. All diagonal blocks in R are irreducible Metzler-Leontief matrices. In the following, we will use certain spectral properties of irreducible Metzler-Leontief matrices, see [73] or [30]:

Theorem 1.2.1. *Let $B = (B_{i,j})$ be an irreducible Metzler-Leontief matrix. Then, there exists a dominant eigenvalue τ of B such that*

(i) τ *is real, has multiplicity* 1 *and the associated left and right eigenvectors are positive;*

(ii) $\tau > \Re(\lambda)$ *where* $\lambda \neq \tau$ *is any other eigenvalue of* B*;*

(iii) $\min_j \sum_i B_{i,j} \leq \tau \leq \max_j \sum_i B_{i,j}$*;*

(iv) If there exist a non-negative vector x *and a real number* ρ *such that* $Bx \leq \rho x$*, then* $\rho \geq \tau$*;* $\rho = \tau$ *if and only if* $Bx = \rho x$*;*

(v) $\sum_i B_{i,j} = 1$ *for all* j *implies* $\tau = 1$*; and*

(vi) $\sum_i B_{i,j} \leq 1$ *for all* j *with at least one strict inequality implies* $\tau < 1$.

Note that because R is diagonisable, so are all blocks on its diagonal. We say that an eigenvalue λ_k belongs to class $\mathcal{C}_m$ if it is an eigenvalue of the restriction of R to $\mathcal{C}_m$. These properties imply the following: As the columns of $T_{1,1}, \ldots, T_{a,a}$, $Q_{1,1}, \ldots, Q_{c,c}$ sum to r, and in each of $P_{1,1}, \ldots, P_{b,b}$, there is a column that sums to less than r, we can order the q eigenvalues of R by $r = \lambda_1 = \ldots = \lambda_{a+c} > \Re(\lambda_{a+c+1}) \geq \ldots \geq \Re(\lambda_q)$. Non-dominant eigenvalues with equal real parts are ordered by decreasing size of imaginary parts. If eigenvalue λ has multiplicity $m > 1$, λ is repeated m times in this list.
Eigenvectors of the blocks can be extended to eigenvectors of R in the following way:

1. If λ is an eigenvalue with multiplicity m of $T_{i,i}$ for some $1 \leq i \leq a$, then there exist m corresponding left (and right) eigenvectors which are zero on every colour outside $T_{i,i}$.

2. If λ is an eigenvalue with multiplicity m of $P_{i,i}$ for some $1 \leq i \leq b$, then there exist m corresponding left eigenvectors which are zero on colours in category 1 and category 2 classes and m right eigenvectors which are zero on category 1 classes.

3. Similarly, if λ is an eigenvalue with multiplicity m of $Q_{i,i}$ for some $1 \leq i \leq c$, then there exist m corresponding left eigenvectors which are zero on all colours in category 1 blocks, in category 2 blocks $Q_{j,j}$ for $j \in \{1, \ldots, c\} \setminus \{i\}$ and in category 3 blocks $P_{j,j}$ that do not lead to $Q_{i,i}$. There exist m corresponding right eigenvectors that are zero on every colour outside $Q_{i,i}$.

We choose dual bases $\{u_1^t, \ldots, u_q^t\}$ and $\{v_1, \ldots, v_q\}$ of left and right eigenvectors of R, respectively, such that for each $k \in \{1, \ldots q\}$, u_k (v_k) is a left (right) eigenvector to λ_k. These are chosen to satisfy the above and the following additional properties. Here, we use the convention that a row vector x^t can be identified with the linear map it induces on $\mathbb{C}^q$. It is easily seen that R is irreducible if and only if $v_1^{(i)} > 0$ for every $i \in \{1, \dots, q\}$.

Furthermore, we assume that if λ is a complex eigenvalue with left and right eigenvectors u_k, v_k, then the eigenvectors corresponding to $\bar{\lambda}$ are $\bar{u}_k, \bar{v}_k$. If λ is a real eigenvalue, then both u_k, v_k are chosen real. For $A \subseteq \{1, \dots, q\}$ and $v \in \mathbb{C}^q$, let v_A be the q-dimensional vector defined by $v_A^{(i)} = v^{(i)} \cdot \delta_A(i)$, $i \in \{1, \dots, q\}$. Let $\mathbf{1}$ denote the q-dimensional all ones vector. We further assume that both left and right eigenvectors to r are real and are of the following form: $u_i := \mathbf{1}_{\mathcal{C}_i}$ for $i = 1, \dots, a$. Further, we can choose the remaining eigenvectors $u_{a+1}, \dots, u_{a+c}$ orthogonal in such a way that $u_{a+s} = \mathbf{1}_{\mathcal{C}_{a+b+s}} + v_s$, where v_s is only nonzero on colour classes of category 3 leading to $\mathcal{C}_{a+b+s}$.
If R is irreducible, we choose $u_1 = (1, \dots, 1)^t = \mathbf{1}$ to be the only eigenvector corresponding to the eigenvalue r.

Having fixed the particular choice of eigenvectors, we turn to the spectral decomposition of $\mathbb{C}^q$ relative to R. Let $\pi_k : \mathbb{C}^q \to \mathbb{C}$ be the linear map defined by

$$\pi_k(v) := u_k^t v. \tag{1.13}$$

Then $\pi_k(v)$ is the coeffcient of the vector v_k in the representation of v with respect to the eigenvector basis $\{v_1, \dots, v_q\}$, i.e.

$$v = \sum_{k=1}^{q} \pi_k(v) v_k. \tag{1.14}$$

In the following, we will apply this decomposition to $(Y_n)_{n \geq 0}$ and study its single components, where

$$Y_n := X_n - \mathbb{E}[X_n], \quad \mathcal{F}_n := \sigma(X_0, \dots, X_n). \tag{1.15}$$

Let $\mathrm{Id}_{\mathbb{C}^q}$ denote the $q \times q$-identity matrix.

1.2.3 Convergence of proportions

Due to the balance condition (A2), there are $rn + |X_0|$ balls in the urn at time $n \geq 0$, regardless of the particular outcome of the first n draws. Accordingly, it is natural to consider the proportions of balls of the various types. This subsection resumes observations on $\frac{X_n}{rn+|X_0|}$ that were already made in Section 1.1. We briefly recall the results of the preceding section: In urn schemes with two types only that satisfy (A1)-(A5), the proportions of balls of colours 1 and 2 are almost surely convergent. Whether the almost sure limit is random or deterministic depends on the multiplicity of the largest eigenvalue.
Now, there are very similar results in the multicolour case. We present them stepwise, starting with urn models that admit a single greatest eigenvalue. Theorems 3.1 and 3.5 in [43] give the desired result in this case.

Theorem 1.2.2 (Janson)**.** *Assume (A1)-(A5) and additionally that* $\Re(\lambda_2) < \lambda_1 = r$. *Then, as* $n \to \infty$

$$\frac{X_n}{n} \longrightarrow r \cdot v_1$$

almost surely and in L^2*. Furthermore*

$$\mathbb{E}[X_n] = r \cdot n v_1 + o(n),$$

and if the urn is strictly small, then $\mathbb{E}[X_n] = r \cdot n v_1 + o\left(n^{\frac{1}{2}}\right)$*, as* $n \to \infty$*.*

So again, if the largest eigenvalue is simple, the proportions converge to deterministic numbers. These limits are all positive if and only if the generating matrix is irreducible.
In a next step, we consider the case where r is allowed to be a multiple eigenvalue, but there is only one class of category 2 (i.e., $c = 1$). This case is treated in Theorem 3.1 of [30].

Theorem 1.2.3 (Gouet)**.** *Suppose that (A1)-(A5) hold and that* $c = 1$*. Then, as* $n \to \infty$*,*

$$\frac{X_n}{rn + |X_0|} \to \sum_{i=1}^{a} D^{(i)} v_i + D^{(a+1)} v_{a+1} =: V$$

almost surely, where

$$\mathcal{L}\left(\left(D^{(1)}, \dots, D^{(a)}, D^{(a+1)}\right)^t\right) = \mathit{Dir}\left(\frac{|(X_0)_{\mathcal{C}_1}|}{r}, \dots, \frac{|(X_0)_{\mathcal{C}_a}|}{r}, \frac{|X_0|}{r} - \sum_{j=1}^{a} \frac{|(X_0)_{\mathcal{C}_j}|}{r}\right).$$

For the same result without identification of the Dirichlet components, see Theorem 3.5 in [67]. Hence, the limit is random, if the algebraic (= geometric, in our case) multiplicity of r is greater than one.
Finally, the case where $c \geq 2$ is more involved, as components $\mathcal{C}_{a+1}, \dots, \mathcal{C}_{a+c}$ are interconnected via category 3 classes. However, Theorem 1.2.3 makes it plausible that as $n \to \infty$,

$$\frac{X_n}{rn + |X_0|} \to \sum_{i=1}^{a} D^{(i)} v_i + D^{(a+1)} (\Gamma_{a+1} v_{a+1} + \dots + \Gamma_{a+c} v_{a+c}) =: V$$

almost surely, where $\left(D^{(1)}, \dots, D^{(a)}, D^{(a+1)}\right)^t$ is Dirichlet distributed with parameters as in Theorem 1.2.3. $\Gamma_{a+1}, \dots, \Gamma_{a+c}$ are random variables that sum to 1 almost surely and are independent of the Dirichlet random vector. Their distribution is an interesting question in its own right, but not answered in this text, because they arise as martingale limits. A proof of this result can easily be obtained along the lines of the proofs given in chapter 3, and we omit the details. In particular, Theorem 3.1.1 states that $\Gamma_{a+1}, \dots, \Gamma_{a+c}$ are non-deterministic.
In the following, the letter V is used to denote the random vector that arises as the almost sure limit of the proportions $\frac{X_n}{rn+|X_0|}$. Note that it is zero in all category 3 components.

Interpretation. Theorem 1.2.3 and its extension yield an interpretation of the urn's dynamics as a superposition of the dynamics of a classical Pólya urn and several irreducible urns. First, $\mathcal{C}_1, \dots, \mathcal{C}_a$ and $\mathcal{C}_{a+1} \cup \dots \cup \mathcal{C}_d$ are isolated and can thus, on a higher level, be regarded as balls in a Pólya urn. Consequently, the asymptotic proportions among these supercolours are Dirichlet distributed. On an intermediate level, the random variables $\Gamma_{a+1}, \dots, \Gamma_{a+c}$ are the asymptotic proportions of the non-isolated dominant classes inside supercolour $\mathcal{C}_{a+1} \cup \dots \cup \mathcal{C}_d$. Finally, on a local level, inside a particular dominant component, the asymptotic proportions of balls are deterministic and given by the components of the right eigenvector corresponding to the class.

1.2.4 Central limit theorem

Before we state the main theorem, let us give a brief impression of known results on variances and weak convergence in the case where r is a simple eigenvalue and $\Re(\lambda_2) < r/2$. It was mentioned at various points that the variance is of a different character depending on whether $\Re(\lambda_2) < r/2$, $\Re(\lambda_2) = r/2$ or $\Re(\lambda_2) > r/2$. To be more precise, Lemma 6.2 in [43] is essentially the following.

Lemma 1.2.1 (Janson). *For $n \geq 2$,*

$$\mathrm{Cov}(X_n) = \begin{cases} O(n), & \Re(\lambda_2) < \frac{r}{2}, \\ O(n\log(n)), & \Re(\lambda_2) = \frac{r}{2}, \\ O\left(n^{2\frac{\Re(\lambda_2)}{r}}\right), & \Re(\lambda_2) > \frac{r}{2}. \end{cases}$$

In particular, if $\Re(\lambda_2) < r$, then

$$\mathrm{Cov}(X_n) = o(n^2).$$

It is also well-known that the distribution of the urn composition is asymptotically normal for *small* Pólya-Eggenberger urns as in the present setting. Recall that an urn is called "small" if the second largest real part of an eigenvalue is at most half the largest eigenvalue.
Janson [43] provides asymptotics for the mean and the covariance matrix of small urns. He shows that after appropriate normalisation, the mean and covariance matrix converge to the mean and variance of the limiting normal distribution, and also includes a result on non-degeneracy of the limit. As his set of conditions is quite general, Theorems 3.2 and 3.3 in [43] can be translated to the present setting.

Theorem 1.2.4 (Janson). *Assume that (A1)-(A5) hold and that the urn is strictly small, i.e. $\Re(\lambda_2) < \frac{r}{2}$. Then as $n \to \infty$,*

$$\mathrm{Var}\left(\frac{X_n - \mathbb{E}[X_n]}{\sqrt{n}}\right) \longrightarrow \sum_{i=2}^{q}\sum_{j=2}^{q} \frac{\lambda_i\lambda_j}{1 - \frac{\lambda_i+\lambda_j}{r}} \left(\sum_{\ell=1}^{q} v_1^{(\ell)} u_i^{(\ell)} u_j^{(\ell)}\right) v_i v_j^t.$$

The asymptotic covariance matrix is always singular.
If (A1)-(A5) hold and $\Re(\lambda_2) = \frac{r}{2}$, then as $n \to \infty$,

$$\mathrm{Var}\left(\frac{X_n - \mathbb{E}[X_n]}{\sqrt{n\log(n)}}\right) \longrightarrow \sum_{i:\Re(\lambda_i)=\frac{r}{2}} |\lambda_i|^2 \langle v_1, |u_i|^2\rangle v_i v_i^*.$$

Note that in the second case, the rank of the asymptotic covariance matrix is comparatively low and equal to the number of eigenvalues with real part equal to $r/2$.

To formulate the main result of the thesis, set

$$M := (\Re(v_1), -\Im(v_1), \Re(v_2), -\Im(v_2), \ldots, \Re(v_q), -\Im(v_q)) \in \mathbb{R}^{q\times 2q}.$$

With this definition, the purpose of the current thesis is to prove and illustrate the following result.

Theorem 1.2.5. *In the setting above, suppose that (A1)-(A5) hold. Let* $p := \max\{k : \Re(\lambda_k)/r > 1/2\}$.

1. *Suppose that for all* $k \in \{1, \dots, q\}$, $\Re(\lambda_k) \neq r/2$ *for all* λ_k *that belong to a dominant class and that in total, there are at least two dominant types. Then there exist complex-valued mean zero random variables* $\Xi_1, \dots, \Xi_p$ *such that*

$$\frac{1}{\sqrt{n}}\left(Y_n - \sum_{k=1}^{p} n^{\frac{\lambda_k}{r}} \Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_V) \tag{1.16}$$

as $n \to \infty$. *Here,* $\mathcal{N}$ *has a non-degenerate, centered multivariate Gaussian mixture distribution with mixture components* $V^{(1)}, \dots, V^{(q)}$ *and covariance matrix*

$$A_V := M\Sigma_V M^t,$$

where Σ_V *is defined in (3.7) to (3.9) below Theorem 3.2.1. Furthermore,* $(A_V)_{i,i} > 0$ *almost surely for dominant colours* i, *whereas* $(A_V)_{i,i} = 0$ *almost surely for non dominant colours* i.

2. *Suppose that there is some* $k \in \{1, \dots, q\}$ *such that* $\Re(\lambda_k) = r/2$ *and that* λ_k *belongs to a dominant class. Then there exist complex-valued mean zero random variables* $\Xi_1, \dots, \Xi_p$ *such that*

$$\frac{1}{\sqrt{n\log(n)}}\left(Y_n - \sum_{k=1}^{p} n^{\frac{\lambda_k}{r}} \Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_V) \tag{1.17}$$

as $n \to \infty$. *Here* $\mathcal{N}$ *has a non-degenerate, centered multivariate Gaussian mixture distribution with with mixture components* $V^{(1)}, \dots, V^{(q)}$ *and covariance matrix*

$$A_V := M\Sigma_V M^t,$$

where Σ_V *is defined in (3.10) and (3.11) below Theorem 3.2.1.* $(A_V)_{i,i} > 0$ *almost surely for dominant colours* i *that belong to the irreducible classes of eigenvalues with real part* $r/2$, *whereas* $(A_V)_{i,i} = 0$ *almost surely for all other colours.*

Remark 4. Theorem 1.2.5 covers the probably most accessible case of urns with a simple largest eigenvalue (or irreducible R, which has the same consequence). For these urns, typically one of the following two kinds of results is derived: In the first kind, asymptotic normality of the rescaled composition vector is proved (e.g., $m-$ary search tree for $m \leq 26$). This is the case where any eigenvalue λ_k different from r satisfies $\Re(\lambda_k) \leq r/2$, which is well-known and treated in [40] and [74]. In the second kind, one shows that the urn exhibits some almost sure oscillating behaviour that is not caused by the normalisation (e.g., $m-$ary search tree for $m \geq 27$, see [13]). Theorem 1.2.5 states that in the latter case, a CLT can always be derived, too, if the oscillating term is also used for the centering of the composition vector (possibly together with some lower order oscillating terms as well). It also covers the classical case of

the Pólya urn, where the almost sure limit is random.
Note that Theorem 1.2.5 says nothing particularly interesting about non-dominant colours. As we have seen in Proposition 1.1.4 in the case of urns with two types, a different scaling is necessary for these colours, as they are not drawn sufficiently often from the urn. This question is not pursued in this thesis.

Remark 5. The work which inspired the above extended view on CLTs is [61], where a CLT for the number of key comparisons in Quicksort under the random permutation model is derived. An alternative proof of the result was later given in [33] and extended to a law of the iterated logarithm in [75].

Remark 6. Special reducible multicolour urn models with three and four types are studied by Bose, Dasgupta and Maulik in [8]. The authors analyse the asymptotic behaviour of linear combinations of the types by an approach that is similar to the methods used in this thesis. In particular, certain linear combinations of the balls of different colours are shown to have limiting distributions which are variance mixtures of normal distributions.

We close this chapter with a brief summary of the results and the involved dynamics. Recall that for urns with two colours, we had real eigenvalues $r = \lambda_1 \geq \lambda_2$ and used this spectrum to distinguish between the cases $\lambda_2 < \frac{r}{2}$ (*strictly small* urn), $\lambda_2 = \frac{r}{2}$ (*small, but not strictly small* urn) and $\lambda_2 > \frac{r}{2}$ (*large* urn). In the first two cases, one can obtain asymptotic normality by the "classical" scaling, with an additional log-factor in the asymptotic variance in the second case. However, in the third case, an additional random term has to be substracted, and if $\lambda_2 = \lambda_1$, also the scaling is random. In the triangular case, the eigenvalues are simply the diagonal elements $a_{1,1}$ and r. Subject to our balance condition, we only consider the case where $a_{1,1} < r$ for urns other than the Pólya urn, in which we have a similar behaviour to the irreducible case. There are several differences when $a_{1,1} > r$, see [42].
For multicolour urns with irreducible replacement matrices, it is also well-known that the type of asymptotics depends on the relation between the eigenvalues of the replacement matrix. See [3] and [40]. Again, we have the distinction between *small* urns, in which case asymptotic normality is known to hold, and *large* urns. In the latter case, there are almost sure limits or oscillations for wich no simple description as in the normal case is known. Pouyanne [67] proves a limit theorem for balanced and large urns, which contains this almost sure result as well as convergence in L^p for any $p \geq 1$, and thus convergence of all moments.
Another observation that we have already made in the case of irreducible small urns with two types only is that the weak limit is independent of the initial state. Consequently, it is not influenced by the outcomes of any finite fixed set of draws, but rather determined by the large quantity of negligible draws in the late evolution of the process. These "smoothing" effects lead to the emergence of the normal distribution. On the other hand, for large urns, there are almost sure, non-degenerate limits. In this case, imbalances caused by the initial distribution or the first draws create a drift that reinforces itself at a sufficient pace to have an impact on the asymptotics of the urn composition. See [40] for a discussion of these effects.

2 Cyclic Urns

Before Theorem 1.2.5 is proved in the general setting, let us exploit its statement for an assessable example: the cyclic urn. This urn model can be defined for each number of colours $q \geq 2$, which makes it a decent example to illustrate various phenomena: On the one hand, the cyclic urn is subject to a phase transition with respect to the number of colours q and thus may be used to explain different types of limiting behaviour for urns within the same model. On the other hand, it is simple enough to allow for explicit calculations. Furthermore, the proof presented here is of its own interest. This is due to the fact that the proof given below is based on the *contraction method*, while the proof of the main theorem relies on martingale techniques. Finally a connection between the cyclic urn process and the random binary search tree (BST) is used to construct the random variables $\Xi_1, \ldots, \Xi_p$ in Theorem 1.2.5 explicitly as functions of the BST chain's limit in its Doob-Martin boundary.

The proceeding and techniques presented below are based on an approach from [61], where a central limit theorem analogue is derived for the number of key comparisons of the Quicksort algorithm. The adaption of these techniques for periodicities in urn models might also be of use in the analysis of periodic phenomena in other discrete structures, where no martingale structure is at hand. In particular, the oscillating behaviour often is formulated in a distributional sense only, and one is confronted with the question of the existence of an almost sure periodic approximation in the first place.

The current chapter is joint work with Prof. Dr. Ralph Neininger; the results presented below were announced in the extended abstract [59]. A preprint including proofs is available under `https://arxiv.org/abs/1612.08930`, and large parts of this chapter are a close adaption to the content of the mentioned work.

2.1 Introduction

Let $q \geq 2$ be a fixed number of colours. In this chapter, we deviate from our former agreement that a colour is an element of $\{1, \ldots, q\}$ and adjust the definition to the specific nature of the cyclic urn. That is, we define the set of types to be $\{0, \ldots, q-1\}$ in the context of cyclic urns. This way of speaking turns out to be much more convenient, but is of course extraneous to the results.

A cyclic urn with $q \geq 2$ colours is given by the following model: The evolution starts with a deterministic initial configuration, say one ball of colour 0 at time 0. At each integer time $n \geq 1$, a ball is drawn from the urn uniformly at random and independently of the previous draws. If its type is j, it is returned to the urn together with a new ball of type $j+1 \mod q$.

This replacement policy is visualised by the generating matrix

$$R_{Cyc} := \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 1 \\ 1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & 1 & 0 \end{pmatrix} \in \mathbb{R}^{q\times q}, \tag{2.1}$$

where $(R_{Cyc})_{i,j}$ is the number of balls of type $i-1$ that are added to the urn after drawing a ball of colour $j-1$ for all $i,j \in \{1,\dots,q\}$. Note that R_{Cyc} is an irreducible matrix that satisfies assumptions (A1)-(A4).

The cyclic urn can be seen as a generalisation of Friedman's urn (example 4), which is exhaustively studied in [25]. We have noted several times that a nice way to think about this type of urn is as a model of a propaganda campaign, in which all q candidates are so bad that the persons who listen to them are immediately convinced to vote for the next candidate on the list of candidates.

Recall that $X_n \in \mathbb{R}^q$ denotes the urn composition vector, whose i^{th} coordinate is given by the number of balls of type $i-1$ after n draws from the urn. The asymptotic distributional behaviour of the sequence $(X_n)_{n\geq 0}$ up to second order expansions is identified in [39–41] and [66, 67]; see also [25]. Cyclic urns appear in [39], example 6.3, as well; a curious emergence of cyclic urns is observed in [41], where Janson uses the model to study congruence classes of depths in random recursive trees.

We now introduce some notation that is needed to formulate the main results. Let $e_1,\dots,e_q$ be the standard unit vectors in $\mathbb{R}^q$, such that $X_0 = e_1$. For fixed $q \geq 2$, set $\omega := \exp(\frac{2\pi i}{q})$. The eigenvalues of R_{Cyc} are the q^{th} roots of unity $\omega^0, \omega, \dots, \omega^{q-1}$. This circular ordering of the eigenvalues is probably most natural, but we follow the notation introduced in chapter 1 and order them by decreasing real parts instead:

$$\lambda_1 = 1, \lambda_2 = \omega, \lambda_3 = \omega^{-1}, \dots, \lambda_q = \omega^{\lceil \frac{q}{2} \rceil}. \tag{2.2}$$

Further, we set

$$\sigma_k := \Re(\lambda_k), \qquad \mu_k := \Im(\lambda_k).$$

Dual bases of right and left eigenvectors $v_1,\dots,v_q \in \mathbb{R}^q$, $u_1,\dots,u_q \in \mathbb{R}^q$ for $\lambda_1,\dots,\lambda_q$, respectively, are given by

$$v_k := \frac{1}{q}\left(1, \lambda_k^{-1}, \lambda_k^{-2}, \dots, \lambda_k^{-(q-1)}\right)^t \in \mathbb{C}^q, \tag{2.3}$$

$$u_k := \left(1, \lambda_k, \lambda_k^2, \dots, \lambda_k^{(q-1)}\right)^t \in \mathbb{C}^q, \quad 1 \leq k \leq q. \tag{2.4}$$

The asymptotic behaviour of $(X_n)_{n\geq 0}$ is as follows. First, it is proved in [30, 40] (see also Theorem 1.2.3), that for all $q \geq 2$ and initial configurations with at least one ball in the urn

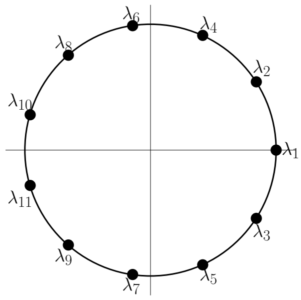

Figure 2.1: Eigenvalues of the cyclic urn for $q = 11$

at time 0,

$$\frac{X_n}{n + |X_0|} \xrightarrow{\text{a.s.}} \nu_1,$$

as $n \to \infty$. That is, the proportions of types in the urn are asymptotically spread evenly among the colours, and ν_1 gives the direction of an almost sure drift.
In second order, cyclic urns leave more room for surprises. Let us have a look at the variance of X_n. Note that as q increases, the eigenvalue with second largest real part and positive imaginary part $\lambda_2 = \omega$ approaches $\lambda_1 = 1$ "from the left" (in terms of real parts) and gets arbitrarily close to it. Lemma 1.2.1 now implies that the variance of X_n depends on whether $\sigma_2 = \cos\left(\frac{2\pi}{q}\right) > \frac{1}{2}$, $\sigma_2 = \frac{1}{2}$ or $\sigma_2 < \frac{1}{2}$. As σ_2 is monotonously increasing in q, there is a phase change in the model at $q = 7$, which is the first q with $\sigma_2 > \frac{1}{2}$.
The same threshold marks the border for the validity of CLTs, as the variance of the Gaussian limit also increases in q: General results from [40, 67, 74] imply that the normalised composition vector X_n converges in distribution to a multivariate normal distribution for $2 \leq q \leq 6$. For $q \geq 7$ however, the situation changes, as the amplitude of the fluctuation becomes too pronounced: no classically standardised version of X_n converges weakly to a non-degenerate limit law. Instead, there exists a complex-valued random variable Ξ_2, which depends on q and X_0, such that as $n \to \infty$,

$$\frac{X_n - \mathbb{E}[X_n]}{n^{\sigma_2}} - 2\Re\left(n^{i\mu_2}\Xi_2\nu_2\right) \xrightarrow{\text{a.s.}} 0. \tag{2.5}$$

In other words, there are *infinitely many* subsequences $\left((X_{n_m} - \mathbb{E}[X_{n_m}])/n_m^{\sigma_2}\right)_{m\geq 1}$ that converge weakly to different limit laws, as $m \to \infty$. The plane spanned by ν_2 and ν_3 in $\mathbb{C}^q$ determines the subspace in which the almost sure periodic deviation from the drift is located.
The convergence in (2.5) has been analysed by means of different techniques. These include an embedding into continuous time multitype branching processes, stochastic fixed point arguments and martingale techniques, see [40, 47, 66]. The first two approaches yield a weak formulation of (2.5), but provide information about the distribution of Ξ_2.
Summarised in the table below, we have the following limit trichotomy for the cyclic urn

model:

q	approximation of X_n	mode of approximation
$2,\dots,5$	$X_n = nv_1 + \sqrt{n}\cdot\mathcal{N}$	in distribution
6	$X_n = nv_1 + \sqrt{n\log(n)}\cdot\mathcal{N}$	in distribution
$7,8,\dots$	$X_n = nv_1 + \Re\left(n^{\lambda_2}\left(\Xi_2 + \frac{1}{\Gamma(1+\omega)}\right)\cdot v_2\right) + o(n^{\sigma_2})$	almost surely

However, the question whether a rescaled version of X_n, centered by $\mathbb{E}[X_n] + 2\Re\left(n^{\omega}\Xi_2 v_2\right)$, obeys a central limit theorem, is not answered in the literature.

The formulation of Theorem 1.2.5 for cyclic urns depends on the number of eigenvalues that have real parts greater than $1/2$. There are $2\lfloor\frac{q-1}{6}\rfloor$ of them, and we distinguish the cases $6 \mid q$ and $6 \nmid q$ as follows.

Theorem 2.1.1. *Let $q \geq 2$ with $6 \nmid q$ and set $p := 2\lfloor\frac{q-1}{6}\rfloor$. Then, there exist complex-valued random variables $\Xi_1,\dots,\Xi_p$ such that, as $n \to \infty$,*

$$\frac{1}{\sqrt{n}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k}\Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma^{(q)}\right).$$

The covariance matrix $\Sigma^{(q)}$ has rank $q-1$ and is given by

$$\Sigma^{(q)} := \sum_{k=2}^{q} \frac{1}{|2\sigma_k - 1|} v_k v_k^*.$$

Note that $\Sigma^{(q)}$ cannot have full rank q, as all row sums are r and the projection of X_n onto the subspace spanned by v_1 is deterministically $(n+1)v_1$. There is no randomness along this direction.
When $6 \mid q$, the normalisation requires an additional $\sqrt{\log n}$ factor and the rank of the covariance matrix is reduced to 2:

Theorem 2.1.2. *Let $q \geq 2$ with $6 \mid q$ and set $p := 2\lfloor\frac{q-1}{6}\rfloor$. Then, there exist complex-valued random variables $\Xi_1,\dots,\Xi_p$ such that, as $n \to \infty$,*

$$\frac{1}{\sqrt{n\log n}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k}\Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma^{(q)}\right).$$

The covariance matrix $\Sigma^{(q)}$ has rank 2 and is given by

$$\Sigma^{(q)} = v_{q/3}v_{q/3}^* + v_{q/3+1}v_{q/3+1}^*.$$

The convergences in Theorems 2.1.1 and 2.1.2 also hold for all moments. For an expansion of $\mathbb{E}[X_n]$, see (2.6).
Note that the sum $\sum_{k=1}^{p}$ in Theorem 2.1.1 is empty for $2 \leq q \leq 5$, as well as in Theorem 2.1.2 for $q = 6$. Hence, our theorems reduce to the central limit laws of Janson [39–41] for $2 \leq q \leq 6$. The covariance matrices for $q \leq 6$ are calculated explicitly in [41], according to

the formulas in [39].
Theorems 2.1.1 and 2.1.2 imply that, in fact, there are ongoing phase changes in the cyclic urn model. At each value $6k+1$ for integer k, a new pair of complex conjugated eigenvalues with real parts greater than the mark $1/2$ emerges. This pair gives rise to almost sure periodicities in the urn composition that are larger than $\sqrt{n}$. Consequently, two additional terms have to be added to the centering of the urn composition vector to obtain asymptotic normality. This random centering can be viewed as an asymptotic expansion of the random variables X_n.

2.1.1 Oscillations

Above, we claimed that the oscillations for $q \geq 7$ are genuine and not a result of inappropriate scaling. This can be seen along the lines of [41], Section 6, for example: (2.5) implies that, along any subsequence such that $\mu_2 \log(n_m) \mod 2\pi \to \gamma$ as $m \to \infty$ for some $\gamma \in [0, 2\pi]$,

$$\frac{X_n^{(j)} - \frac{n}{q}}{n^{\sigma_2}} \xrightarrow{\mathcal{L}} \frac{2}{q}\Re\left(e^{i\gamma - 2\pi(j-1)/q}\left(\Xi_2 + \frac{1}{\Gamma(1+\omega)}\right)\right),$$

jointly in $j = 1, \ldots, q$. Now assume that there is some normalisation that yields weak convergence $a_n(X_n^{(1)} - b_n) \xrightarrow{\mathcal{L}} V$, for some non-degenerate random variable V and real constants $a_n > 0$ and b_n. Janson shows, using convergence of the subsequences above, that this implies that either

(i) $\Xi_2 \stackrel{\mathcal{L}}{=} aW$ for some real random variable W and a complex constant a, or

(ii) $\mathbb{E}\left[\Xi_2^3\right] = 0$.

Now for all q, $\mathbb{E}\left[\Xi_2^3\right] \neq 0$ (see Remark 6.3 in [41]). The first possibility can also be excluded, as it is shown in [51], that Ξ_2 has a Lebesgue density on $\mathbb{C}$. It follows that there really are oscillations, even with different normalisations.
However, this can also be seen by using the almost sure decomposition of the composition vector into its spectral components, which are of different sizes for eigenvalues that are not complex conjugated. This decomposition is explained in section 2.1.2.

We expect that results analogous to Theorems 2.1.1 and 2.1.2 hold for other random combinatorial structures in which periodicities have been observed. These structures include other urn models with almost sure random periodic behavior, see Janson [40, Theorem 3.24], the size of random m-ary search trees [15], the number of leaves in random d-dimensional (point) quadtrees [14], secondary cost measures of quicksort with median-of-$(2t+1)$ [16], the size of random fragmentation trees [44] or the probability that there is a single winner in the leader election algorithm [10].

2.1.2 Components of the process

In this subsection, we prepare the proof of Theorems 2.1.1 and 2.1.2 with some notation and observations that are repeated in a more general setting in chapter 3. For cyclic urns, the spectral decomposition (1.14) assumes a particularly simple form, as all eigenvalues are distinct, v_1 is deterministic and a multiple of $(1, \ldots, 1)^t$ and the right eigenvectors form an orthonormal basis of $\mathbb{C}^q$. The joint asymptotics of the projections are identified in Proposition 2.1.3. The proposition directly implies Theorems 2.1.1 and 2.1.2 and also explains the occurrence of the random centering as a sum of terms of the form $\Re(n^{\lambda_k}\Xi_k v_k)$ as well as the normal fluctuation in Theorems 2.1.1 and 2.1.2. All proofs and calculations are deferred to Section 2.2.

Recall that $q \geq 2$ is fixed and $X_0 = e_1$. An exact asymptotic expression for the mean value of X_n is given in [47], Lemma 6.7. This expression implies the expansion, as $n \to \infty$,

$$\mathbb{E}\left[X_n\right] = nv_1 + \sum_{k=2}^{p} n^{\lambda_k} \frac{1}{\Gamma(1+\lambda_k)} v_k + \begin{cases} o(\sqrt{n}), & \text{if } 6 \nmid q, \\ O(\sqrt{n}), & \text{if } 6 \mid q. \end{cases} \tag{2.6}$$

The sum in the expansion of the mean already indicates that there are more oscillating terms than (2.5) suggests at a first glance. Moreover, for $q \geq 7$, Lemma 1.2.1 states that the variances and covariances of X_n are of order $n^{2\sigma_2}$. We will recover this result in section 2.2; this also explains the normalisation $n^{-\sigma_2}(X_n - nv_1)$ in (2.5).

Using (1.13) and (1.14), we arrive at the spectral decomposition

$$X_n = \sum_{k=1}^{q} \pi_k(X_n) v_k$$

of X_n. Let $(\mathcal{F}_n)_{n\geq 0}$ denote the canonical filtration of the urn process as defined in (1.15). The reason why (1.14) is a particularly good decomposition lies in the form of the conditional expectation, which is

$$\mathbb{E}\left[X_{n+1} \,|\, \mathcal{F}_n\right] = \sum_{k=1}^{q} \frac{X_n^{(k)}}{n+1}(X_n + R_{\mathrm{Cyc}} e_k) = \left(\mathrm{Id}_{\mathbb{C}^q} + \frac{R_{\mathrm{Cyc}}}{n+1}\right) X_n, \qquad n \geq 0. \tag{2.7}$$

In other words, the conditional expectation is a linear function of the current state. This in turn implies that each coefficient in the projections $\pi_1(X_n)v_1, \ldots, \pi_q(X_n)v_q$ can be rescaled in order to yield a complex-valued martingale. In the following, we will study the projections $\pi_1(X_n)v_1, \ldots, \pi_q(X_n)v_q$ separately via martingale techniques and then analyse their joint behaviour by means of the contraction method. The eigenspace decomposition in combination with martingale techniques is implicit in the work of Smythe [74] and more explicit in the proof of Theorem 3.5 in [67] for certain projections. We adopt it for all eigenspace projections.

There are two types of projections: If $\sigma_k > \frac{1}{2}$ (that is, $k \leq p$), the corresponding projection $\pi_k(\cdot)v_k$ is called *large*. Large components of X_n give, after scaling, rise to almost sure limits, and their magnitudes are *larger* than $\sqrt{n}$. The emergence of these almost sure limits is the reason why there is no direct central limit theorem for $q > 6$. However, when considered individually, there is a CLT analogue for the fluctuation of each large martingale about its limit, which we will use in the CLT analogue for $(X_n)_{n\geq 0}$. In contrast, projections $\pi_k(\cdot)v_k$ with $\sigma_k \leq \frac{1}{2}$ are often called *small* and yield non-convergent martingales.
Because each projection has a complex-conjugated counterpart, oscillations as in (2.5) arise. In (2.5) only the largest oscillating term is considered. On the other hand, the small components $\pi_k(X_n)v_k$ oscillate, too, but their *small* oscillation is overcast by the general $\sqrt{n}$ noise.

Because the projections are typically complex-valued, we jointly consider their real and

imaginary parts in the following. Note also that it is sufficient to consider projections whose eigenvalues have non-negative imaginary parts; that is, we restrict to projections corresponding to $\sigma_1, \sigma_2, \sigma_4, \dots, \sigma_{\lfloor \frac{q}{2} \rfloor}$.
Moreover, note that $u_1^t(X_n) = X_n^{(1)} + \dots X_n^{(q)}$ is a particularly simple projection, as it counts the number of balls in the urn at time n, which is deterministically $n+1$. Consequently, for all $n \geq 0$,

$$u_1^t(X_n - \mathbb{E}[X_n]) = 0,$$

and the corresponding projection will not be considered in the following.
The behaviour of the small projections $\pi_k(X_n)$ is determined in [40, 57]. If $k \in \mathbb{N}$ is such that $\sigma_{2k} < \frac{1}{2}$ and $\sigma_{2k} \neq -1$, the results in [40, 57] imply that

$$P_{n,k} := \frac{1}{\sqrt{n}} \begin{pmatrix} \Re(u_{2k}^t(X_n - \mathbb{E}[X_n])) \\ \Im(u_{2k}^t(X_n - \mathbb{E}[X_n])) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\mathrm{Id}_{\mathbb{R}^2}}{1 - 2\sigma_{2k}}\right). \tag{2.8}$$

If q is even, the imaginary part of the projection associated with $\sigma_q = -1$ is zero and $P_{n,q/2} := n^{-1/2} u_q^t(X_n - \mathbb{E}[X_n]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/3)$. For $q = 2$, the last mentioned result is also proved in [25, Theorem 5.1].
If, on the other hand, $6 \mid q$, there is a pair of eigenvalues with real parts $\sigma_{q/3} = \sigma_{q/3+1} = \frac{1}{2}$. In this case the scaling requires an additional $\sqrt{\log n}$ factor, and, again, it follows with [40, 57], that

$$P_{n,k} := \frac{1}{\sqrt{n \log n}} \begin{pmatrix} \Re(u_{q/3}^t(X_n - \mathbb{E}[X_n])) \\ \Im(u_{q/3}^t(R_n - \mathbb{E}[X_n])) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{2}\mathrm{Id}_{\mathbb{R}^2}\right). \tag{2.9}$$

In contrast, we set

$$P_{n,k} := \frac{1}{\sqrt{n}} \begin{pmatrix} \Re\left(u_{2k}^t(X_n - \mathbb{E}[X_n]) - n^{\lambda_{2k}} \Xi_{2k}\right) \\ \Im\left(u_{2k}^t(X_n - \mathbb{E}[X_n]) - n^{\lambda_{2k}} \Xi_{2k}\right) \end{pmatrix} \tag{2.10}$$

for large projections $\pi_k(\cdot)v_k$, $k \in \{1, \dots, q\}$. Here, the complex-valued random variable Ξ_{2k} is defined as a martingale limit in (2.13), Section 2.2.1. In this section, we also prove the convergence of the variances and covariances of all $P_{n,k}$.

Now, the random two-dimensional vectors $P_{n,k}$ are defined for all $1 \leq k \leq \lfloor \frac{q}{2} \rfloor$. Note that components (2.8) and (2.9) describe the normalised fluctuations of projections whose associated martingales are not convergent. In contrast, in the case of projections that can be rescaled to convergent martingales, components (2.10) describe their normalised fluctuations about an almost sure approximation coming from the martingale limit. As a main contribution to the known results on cyclic urns, we show that $P_{n,1}, \dots, P_{n,\lfloor \frac{q}{2} \rfloor}$ are *jointly* asymptotically normal and even asymptotically independent. This is the content of the following proposition, which directly implies Theorems 2.1.1 and 2.1.2 via an application of the Cramér-Wold device.

Proposition 2.1.3. *Assume that $6 \mid q$, and set $Z_n := (P_{n,1}, \dots, P_{n,q/2}) \in \mathbb{R}^{q-1}$ with components as defined in (2.8)–(2.10). Then, as $n \to \infty$,*

$$Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, M_q),$$

where

$$M_q := \frac{1}{2}\mathrm{diag}\left(\frac{\mathrm{Id}_{\mathbb{R}^2}}{|2\sigma_2 - 1|}, \dots, \frac{\mathrm{Id}_{\mathbb{R}^2}}{|2\sigma_{p-1} - 1|}, \mathrm{Id}_{\mathbb{R}^2}, \frac{\mathrm{Id}_{\mathbb{R}^2}}{|2\sigma_{p+3} - 1|}, \dots, \frac{\mathrm{Id}_{\mathbb{R}^2}}{|2\sigma_{q-2} - 1|}, \frac{2}{3}\right). \tag{2.11}$$

For general clarification purposes, we only give a complete proof of Proposition 2.1.3 for $6 \mid q$. This case is more complex as there are two components with scaling $\sqrt{n\log(n)}$ instead of $\sqrt{n}$. However, it is immediate from the derived asymptotics of the projections and the proof technique, that the proposition holds for all q.

The proof of Proposition 2.1.3 is organised as follows: We first derive first and second moments as well as mixed moments for all projections in Section 2.2.1. In Section 2.2.2, we use an almost sure recurrence relation for the sequence $(X_n)_{n\geq 0}$ as explained in [47] to derive pointwise decompositions of the complex random variables $\Xi_1, \dots, \Xi_p$. These decompositions imply recurrences for the components of Z_n, see equation (2.19) in Section 2.2.2. Starting from the recurrence for Z_n, we are then able to prove Proposition 2.1.3 by means of stochastic fixed point arguments in the context of the contraction method within the Zolotarev metric ζ_3, see [62] for general reference. More specifically, we draw back to an approach to bound the Zolotarev distance and some estimates from [61] where a related univariate problem was discussed.

2.2 Proof of Theorems 2.1.1 and 2.1.2

We start with estimates for the covariance matrix of the sequence $(Z_n)_{n\geq 0}$.

2.2.1 Convergence of the covariance matrix

Recall that we study the centered process $(X_n - \mathbb{E}[X_n])_{n\geq 0}$ via its spectral decomposition with respect to the orthogonal basis $\{v_k : 1 \leq k \leq q\}$ of the unitary vector space $\mathbb{C}^q$, i.e.

$$X_n - \mathbb{E}[X_n] = \sum_{k=1}^{q} \pi_k\left(X_n - \mathbb{E}[X_n]\right) v_k = \sum_{k=1}^{q} u_k^t\left(X_n - \mathbb{E}[X_n]\right) v_k$$

as in (1.13). The evolution (2.7) of the process implies that for each eigenspace coefficient $u_k^t\left(X_n - \mathbb{E}\left[X_n\right]\right)$, $1 \leq k \leq q$, there is a complex normalisation

$$M_n^{(k)} := \frac{\Gamma(n+1)}{\Gamma(n+1+\lambda_k)} u_k^t\left(X_n - \mathbb{E}\left[X_n\right]\right), \quad n \geq 1, \tag{2.12}$$

that turns it into a centered martingale. We set $M_0^{(k)} := 0$.
The asymptotic behaviour of these eigenspace-martingales is known to depend on λ_k, see [40, 41, 66]: First, for all $k \in \{1, \dots, q\}$ with $\sigma_k > 1/2$, there exists a complex-valued, centered random variable Ξ_k such that, as $n \to \infty$,

$$M_n^{(k)} \xrightarrow{\text{a.s.}} \Xi_k. \tag{2.13}$$

The convergence also holds in L_p for every $p \geq 1$. Note that the random variables $\Xi_1, \dots, \Xi_p$ in (2.13) are identical with $\Xi_1, \dots, \Xi_p$ in (2.10) and in Theorems 2.1.1 and 2.1.2. On the other hand, if $\sigma_k \le 1/2$, $\left(M_n^{(k)}\right)_{n\ge 0}$ is known to converge in distribution to a normal limit law, after proper normalisation.

Our subsequent analysis requires asymptotics of the mean and covariance structure of $u_1^t(X_n), \dots, u_q^t(X_n)$. Exploiting the dynamics of the urn in (2.7), elementary calculations imply the results presented in the following lemma.

Lemma 2.2.1. *For* $k \in \{1, \dots, q\}$,

$$\mathbb{E}\left[u_k^t(X_n)\right] = \sum_{j=1}^{q} (\lambda_k)^{j-1} \mathbb{E}\left[X_n^{(j)}\right] = \begin{cases} \frac{\Gamma(n+1+\lambda_k)}{\Gamma(n+1)\Gamma(1+\lambda_k)}, & \lambda_k \neq -1, \\ 0, & \lambda_k = -1. \end{cases} \tag{2.14}$$

For $k, \ell \in \{1, \dots, q\}$,

$$\begin{aligned} &\mathbb{E}\left[u_k^t(X_n)\, u_\ell^t(X_n)\right] \\ &= \prod_{j=1}^{n} \left(\frac{j+\lambda_k+\lambda_\ell}{j}\right) + \sum_{m=1}^{n} \frac{\lambda_k\lambda_\ell}{m} \prod_{j=1}^{m-1} \left(\frac{j+\lambda_k\lambda_\ell}{j}\right) \prod_{t=m+1}^{n} \left(\frac{t+\lambda_k+\lambda_\ell}{t}\right). \end{aligned} \tag{2.15}$$

Proof. (2.14) immediately follows from (2.7). For (2.15), let $k, \ell \in \{1, \dots, q\}$ and $n \ge 1$ and note that, almost surely,

$$\mathbb{E}\left[u_k^t(X_n) u_\ell^t(X_n) | \mathcal{F}_{n-1}\right] = \left(1 + \frac{\lambda_k+\lambda_\ell}{n}\right) u_k^t(X_{n-1}) u_\ell^t(X_{n-1}) + \frac{\lambda_k\lambda_\ell}{n} (u_k u_\ell)^t(X_{n-1}).$$

Here, we use the abbreviation $(u_k u_\ell)^t(X_{n-1}) := \sum_{j=1}^{q} (\lambda_k\lambda_\ell)^{j-1} X_{n-1}^{(j)}$. □

Remark 1. Equation (2.15) implies that for all k with $\sigma_k < 1/2$, $\mathbb{E}[|u_k^t(X_n - \mathbb{E}[X_n])|^2]$ is of linear order, for all k with $\sigma_k = 1/2$, $\mathbb{E}[|u_k^t(X_n - \mathbb{E}[X_n])|^2]$ is of order $n \log n$ and for all k with $1/2 < \sigma_k < 1$, $\mathbb{E}[|u_k^t(X_n - \mathbb{E}[X_n])|^2]$ is of order $n^{2\sigma_k}$. To make this more visible from (2.15), consider the following case distinctions.

We first consider the real cases $k = \ell = 1$ and $k = \ell = q$ for $2 \mid q$:

$$\mathbb{E}\left[|u_1^t(X_n - \mathbb{E}[X_n])|^2\right] = (n+1)^2 - (n+1)^2 = 0$$

and, if $2 \mid q$,

$$\mathbb{E}\left[|u_q^t(X_n - \mathbb{E}[X_n])|^2\right] = \frac{n+1}{3} - 0 = \frac{n+1}{3}.$$

Now, $\lambda_k + \lambda_\ell = -1$ only if $3 \mid q$ and $\{k, \ell\} = \{2q/3, 2q/3+1\}$. In this case,

$$\mathbb{E}\left[|u_{2q/3}^t(X_n - \mathbb{E}[X_n])|^2\right] = \frac{1}{n}\sum_{j=1}^{n} j - \left|\frac{\Gamma(n+1+\lambda_k)}{\Gamma(n+1)\Gamma(1+\lambda_k)}\right|^2 \sim \frac{n+1}{2}.$$

On the other hand, $\lambda_k\lambda_\ell = \lambda_k + \lambda_\ell$ only if $6 \mid q$ and $\{k, \ell\} = \{q/3, q/3+1\}$. In this case,

$\lambda_k + \lambda_\ell = 1$ and

$$\mathbb{E}\left[|u^t_{q/3}(X_n - \mathbb{E}[X_n])|^2\right] = (n+1)\sum_{j=1}^{n+1}\frac{1}{j} - \left|\frac{\Gamma(n+1+\lambda_k)}{\Gamma(n+1)\Gamma(1+\lambda_k)}\right|^2 \sim n\log n.$$

Thirdly, $\lambda_k + \lambda_\ell = 0$ if and only if $2 \mid q$ and $\sigma_k = -\sigma_\ell$ and $\mu_k = -\mu_\ell$. If $\{k, \ell\} = \{1, q\}$, it is immediate that $\mathbb{E}\left[u^t_k(X_n - \mathbb{E}[X_n])u^t_\ell(X_n - \mathbb{E}[X_n])\right] = 0$. In all other cases,

$$\mathbb{E}\left[u^t_k(X_n - \mathbb{E}[X_n])u^t_\ell(X_n - \mathbb{E}[X_n])\right] \sim \frac{n^{\lambda_k\lambda_\ell}}{\Gamma(1+\lambda_k\lambda_\ell)} - \frac{1}{\Gamma(1+\lambda_k)\Gamma(1+\lambda_\ell)}.$$

Finally, $\lambda_k\lambda_\ell = -1$ if $\sigma_k = -\sigma_\ell$ and $\mu_k = \mu_\ell$ and then,

$$\begin{aligned}
&\mathbb{E}\left[u^t_k(X_n - \mathbb{E}[X_n])u^t_\ell(X_n - \mathbb{E}[X_n])\right] \\
&= \frac{(\lambda_k+\lambda_\ell)\Gamma(n+1+\lambda_k+\lambda_\ell)}{\Gamma(2+\lambda_k+\lambda_\ell)\Gamma(n+1)} - \frac{\Gamma(n+1+\lambda_k)\Gamma(n+1+\lambda_\ell)}{\Gamma(n+1)^2\Gamma(1+\lambda_k)\Gamma(1+\lambda_\ell)} \\
&\sim \left(\frac{\lambda_k+\lambda_\ell}{\Gamma(2+\lambda_k+\lambda_\ell)} - \frac{1}{\Gamma(1+\lambda_k)\Gamma(1+\lambda_\ell)}\right) n^{\lambda_k+\lambda_\ell}.
\end{aligned}$$

In all other cases,

$$\begin{aligned}
&\mathbb{E}\left[u^t_k(X_n - \mathbb{E}[X_n])u^t_\ell(X_n - \mathbb{E}[X_n])\right] \\
&= \frac{1}{\lambda_k\lambda_\ell - \lambda_k - \lambda_\ell}\left(\frac{\Gamma(n+1+\lambda_k\lambda_\ell)}{\Gamma(n+1)\Gamma(\lambda_k\lambda_\ell)} - \frac{\Gamma(n+1+\lambda_k+\lambda_\ell)}{\Gamma(n+1)\Gamma(\lambda_k+\lambda_\ell)}\right) - \frac{\Gamma(n+1+\lambda_k)\Gamma(n+1+\lambda_\ell)}{\Gamma(n+1)^2\Gamma(1+\lambda_k)\Gamma(1+\lambda_\ell)} \\
&\sim \frac{1}{(\lambda_k\lambda_\ell - \lambda_k - \lambda_\ell)\Gamma(\lambda_k\lambda_\ell)} n^{\lambda_k\lambda_\ell} - \left(\frac{1}{(\lambda_k\lambda_\ell - \lambda_k - \lambda_\ell)\Gamma(\lambda_k+\lambda_\ell)} + \frac{1}{\Gamma(1+\lambda_k)\Gamma(1+\lambda_\ell)}\right) n^{\lambda_k+\lambda_\ell}.
\end{aligned}$$

Remark 7. Mixed moments of the corresponding real and imaginary parts can be obtained from (2.15) via the identities

$$\begin{aligned}
\mathbb{E}\left[\Re(u^t_k(X_n))\Re(u^t_\ell(X_n))\right] &= \frac{1}{2}\Re\left(\mathbb{E}\left[u^t_k(X_n)u^t_\ell(X_n)\right] + \mathbb{E}\left[u^t_k(X_n)\overline{u^t_\ell(X_n)}\right]\right), \\
\mathbb{E}\left[\Im(u^t_k(X_n))\Im(u^t_\ell(X_n))\right] &= \frac{1}{2}\Re\left(\mathbb{E}\left[u^t_k(X_n)\overline{u^t_\ell(X_n)}\right] - \mathbb{E}\left[u^t_k(X_n)u^t_\ell(X_n)\right]\right), \\
\mathbb{E}\left[\Re(u^t_k(X_n))\Im(u^t_\ell(X_n))\right] &= \frac{1}{2}\Im\left(\mathbb{E}\left[u^t_k(X_n)u^t_\ell(X_n)\right] + \mathbb{E}\left[\overline{u^t_k(X_n)}u^t_\ell(X_n)\right]\right).
\end{aligned}$$

However, for martingales $\left(M^{(k)}_n\right)_{n\geq 0}$ corresponding to eigenvalues with real parts $\sigma_k > \frac{1}{2}$, we need the L_2-rate of convergence to the limit rather than their variance. This rate is calculated in the following lemma.

Lemma 2.2.2. *For $k \geq 1$ such that $1/2 < \sigma_k < 1$ and Ξ_k as in (2.13), as $n \to \infty$,*

$$\mathbb{E}\left[\left|M^{(k)}_n - \Xi_k\right|^2\right] \sim \frac{1}{2\sigma_k - 1} n^{1-2\sigma_k}$$

and

$$\mathbb{E}\left[\left(M_n^{(k)}-\Xi_k\right)^2\right]\sim\frac{1}{(1-2\lambda_k^{-1})\Gamma(2\lambda_k)}n^{-1}.$$

In particular,

$$\begin{aligned}\mathbb{E}\left[\Re\left(M_n^{(k)}-\Xi_k\right)^2\right]&\sim\frac{1}{2}\frac{1}{2\sigma_k-1}n^{1-2\sigma_k},\\ \mathbb{E}\left[\Im\left(M_n^{(k)}-\Xi_k\right)^2\right]&\sim\frac{1}{2}\frac{1}{2\sigma_k-1}n^{1-2\sigma_k},\\ \mathbb{E}\left[\Re\left(M_n^{(k)}-\Xi_k\right)\Im\left(M_n^{(k)}-\Xi_k\right)\right]&\sim\frac{1}{2}\Im\left(\frac{1}{(1-2\lambda_k^{-1})\Gamma(2\lambda_k)}\right)n^{-1}.\end{aligned}$$

Proof. We show the claim for $\mathbb{E}\left[\left|M_n^{(k)}-\Xi_k\right|^2\right]$ in an exemplary way. Here, we decompose

$$\begin{aligned}&\mathbb{E}\left[\left|M_n^{(k)}-\Xi_k\right|^2\right]\\ &=\sum_{m=n}^{\infty}\mathbb{E}\left[\left|M_m^{(k)}-M_{m+1}^{(k)}\right|^2\right]\\ &=\sum_{m=n}^{\infty}\left|\frac{\Gamma(m+2)}{\Gamma(m+2+\lambda_k)}\right|^2\mathbb{E}\left[\left|u_k^t(X_{m+1}-X_m)-\frac{\lambda_k}{m+1}u_k^t(X_m)\right|^2\right]\\ &=\sum_{m=n}^{\infty}\left|\frac{\Gamma(m+2)}{\Gamma(m+2+\lambda_k)}\right|^2\left(\mathbb{E}\left[\left|u_k^t(X_{m+1}-X_m)\right|^2\right]-\frac{1}{(m+1)^2}\mathbb{E}\left[\left|u_k^t(X_m)\right|^2\right]\right)\\ &=\sum_{m=n}^{\infty}\left|\frac{\Gamma(m+2)}{\Gamma(m+2+\lambda_k)}\right|^2\left(1+\frac{1}{1-2\sigma_k}\frac{1}{(m+1)^2}\left(\frac{\Gamma(m+1+2\sigma_k)}{\Gamma(m+1)\Gamma(2\sigma_k)}-(m+1)\right)\right)\\ &\sim\sum_{m=n}^{\infty}m^{-2\sigma_k}\sim\frac{1}{2\sigma_k-1}n^{1-2\sigma_k}\end{aligned}$$

as $n\to\infty$. □

Taken together, the calculations of this subsection imply that as $n\to\infty$, the covariance matrix of Z_n converges to M_q, which is defined in (2.11).

2.2.2 Decomposition and recursions

In this subsection, we derive a recurrence for the sequence $(Z_n)_{n\geq 0}$. The proceeding is the following: First, we briefly explain how to derive an almost sure recurrence for the sequence $(X_n)_{n\geq 0}$ which then extends to the projections. These recursive representations yield almost sure decompositions of the martingale limits Ξ_k and thus also of the components of Z_n.

Recall that in the current chapter, we always start with one ball of type zero. However, in the recursions below, it will be necessary to start the cyclic urn process with one ball of an arbitrary type $j\in\{0,\dots,q-1\}$. Denote the resulting sequence of composition vectors by

$\left(X_n^{[j]}\right)_{n\geq 0}$. Their distribution satisfies the relation

$$\mathcal{L}\left(\left(X_n^{[j]}\right)_{n\geq 0}\right) = \mathcal{L}\left(\left((R_{\mathrm{Cyc}})^j X_n\right)_{n\geq 0}\right), \quad 0 \leq j \leq q-1. \tag{2.16}$$

Consequently, the martingales obtained from the processes $\left(X_n^{[j]}\right)_{n\geq 0}$, $j \in \{0, \dots, q-1\}$, are related via the distributional identity

$$\mathcal{L}\left(\left(M_n^{(k),[j]}\right)_{n\geq 0}\right) = \mathcal{L}\left(\left(\lambda_k^j M_n^{(k)}\right)_{n\geq 0}\right),$$

where $j \in \{0, \dots, q-1\}$.

We now use the following simple change of perspective to decompose $(X_n)_{n\geq 0} = \left(X_n^{[0]}\right)_{n\geq 0}$ into two cyclic urn processes: In our original process, there is one ball of colour 0 in the urn at time 0. This ball is drawn with probability one in the first step and put back to the urn, together with a ball of colour 1. Now imagine that the original urn is split into two smaller urns, and we put the ball of type 0 into the first urn and the ball of type 1 into the second urn. Therefore, like in a matryoshka doll, we find two smaller urns within the original urn. The evolution of the cyclic urn process continues as usual, with the only difference that we add the new balls to the same urn as the drawn ball. This yields a natural decomposition of all balls that are added to the urn after time 1 into balls that are descendants of the first type 0 ball and balls that are descendants of the first type 1 ball. (Note that this is just another formulation of the tree-based approach in [47, Section 6.3], see also [12]. This approach is used in section 2.3, where we also explain the BST algorithm, but in order to avoid an additional load of notation, we use urn language in this place.)
In this subdivision of the original urn, let I_n denote the number of draws from the type 0 urn after $n \geq 1$ draws from the original urn. In particular, I_n gives the local time at the first of the two smaller urns and thus, $I_1 = 0$. It is easy to see that $(I_{n+1}+1)_{n\geq 0}$ follows the same dynamics as the number of balls of type 1 in the Pólya urn from example 1. Consequently, at any time $n \geq 1$, the number of descendants of the first ball is uniformly distributed on $\{0, \dots, n-1\}$. In particular, divided by n, it almost surely converges to a random variable U with $\mathcal{L}(U) = \mathrm{unif}(0,1)$. Furthermore, conditionally on $\{U = u\}$, I_n is $\mathrm{Bin}(n-1, u)$-distributed. This implies, with $J_n := n - 1 - I_n$ (the local time in the second urn), the recurrence

$$X_n^{[0]} = X_{I_n}^{[0],\{1\}} + X_{J_n}^{[1],\{2\}} = X_{I_n}^{[0],\{1\}} + R_{\mathrm{Cyc}} X_{J_n}^{[0],\{2\}}, \qquad n \geq 1. \tag{2.17}$$

Here, the sequences $(X_n^{[0],\{1\}})_{n\geq 0}$ and $(X_n^{[1],\{2\}})_{n\geq 0}$ denote the composition vectors of the cyclic urns given by the evolutions of the balls in the two smaller urns (upper indices $[0]$ and $[1]$ denote the initial type, upper indices $\{1\}$ and $\{2\}$ distinguish between the first and the second urn). They are independent of I_n. We have set $(X_n^{[0],\{2\}})_{n\geq 0} := (R_{\mathrm{Cyc}}^t X_n^{[1],\{2\}})_{n\geq 0}$, and note that due to identity (2.16), $(X_n^{[0],\{2\}})_{n\geq 0}$ is a cyclic urn process started with one ball of type 0 at time 0.
Now, applying the transformation and scaling (2.12) which turn X_n into $M_n^{(k)}$ to the left and right hand side of (2.17), letting $n \to \infty$ and using the convergence in (2.13) yields the following almost sure decomposition of the Ξ_k:

Proposition 2.2.1. *For all* $k \geq 1$ *with* $\sigma_k > \frac{1}{2}$ *there exist random variables* $\Xi_k^{\{1\}}$, $\Xi_k^{\{2\}}$ *such that*

$$\Xi_k = U^{\lambda_k}\Xi_k^{\{1\}} + \lambda_k(1-U)^{\lambda_k}\Xi_k^{\{2\}} + g_k(U). \tag{2.18}$$

$U, \Xi_k^{\{1\}}$, $\Xi_k^{\{2\}}$ *are independent and* $\Xi_k^{\{1\}}$ *and* $\Xi_k^{\{2\}}$ *have the same distribution as* Ξ_k*. Here,*

$$g_k(u) := \frac{1}{\Gamma(1+\lambda_k)}\left(u^{\lambda_k} + \lambda_k(1-u)^{\lambda_k} - 1\right).$$

Remark 8. In [41], it is noted that the distribution of Ξ_2 is also determined by the distributional equations

$$\begin{aligned}\mathcal{L}(Z) &= \mathcal{L}\left(W^{\omega}\left(\Xi_2 + \frac{1}{\Gamma(1+\lambda_2)}\right)\right),\\ \mathcal{L}(Z) &= \mathcal{L}\left(U^{\omega}(Z+\omega Z')\right),\end{aligned}$$

where $\mathcal{L}(Z) = \mathcal{L}(Z')$, $\mathcal{L}(W) = \mathrm{Exp}(1)$, $\mathcal{L}(U) = \mathrm{unif}(0,1)$, and W, Ξ_2, U, Z, Z' are independent, together with $\mathbb{E}[Z] = 1$. Here, $\omega = \exp\left(\frac{2\pi i}{q}\right)$ as before.

Finally, we use (2.17) and (2.18) to obtain a decomposition of $(Z_n)_{n\geq 0}$. After some calculations, we see that

$$Z_n = \rho_{I_n}^{-1}\rho_n Z_{I_n}^{\{1\}} + \rho_{J_n}^{-1}\rho_n \mathcal{D} Z_{J_n}^{\{2\}} + \rho_n F_n, \qquad n \geq 1, \tag{2.19}$$

where $\rho_0 := \rho_1 := \mathrm{Id}_{\mathbb{R}^{q-1}}$ and $\rho_k := \frac{1}{\sqrt{k}}\mathrm{diag}\left(1,\dots,1,\frac{1}{\sqrt{\log k}},\frac{1}{\sqrt{\log k}},1,\dots,1\right)$ for $k \geq 2$. The additional factor of $\sqrt{\log k}$ is needed for the real and imaginary part of eigenspace $q/3$ (recall that $\sigma_{q/3} = \frac{1}{2}$). $\mathcal{D}$ is a $(q-1)\times(q-1)$ matrix, which is composed of rotation matrices

$$\mathcal{D} = \begin{pmatrix} \cos\left(\frac{2\pi}{q}\right) & -\sin\left(\frac{2\pi}{q}\right) & & & & \\ \sin\left(\frac{2\pi}{q}\right) & \cos\left(\frac{2\pi}{q}\right) & & & & \\ & & \ddots & & & \\ & & & \cos\left(\frac{2\pi(q/2-1)}{q}\right) & -\sin\left(\frac{2\pi(q/2-1)}{q}\right) & \\ & & & \sin\left(\frac{2\pi(q/2-1)}{q}\right) & \cos\left(\frac{2\pi(q/2-1)}{q}\right) & \\ & & & & & -1 \end{pmatrix}.$$

Lastly, the "error term" F_n is made up of three components: Setting

$$G_{k,n}(\ell) := \frac{\Gamma(\ell+1+\lambda_k)}{\Gamma(\ell+1)\Gamma(1+\lambda_k)} + \lambda_k\frac{\Gamma((n-1-\ell)+1+\lambda_k)}{\Gamma((n-1-\ell)+1)\Gamma(1+\lambda_k)} - \frac{\Gamma(n+1+\lambda_k)}{\Gamma(n+1)\Gamma(1+\lambda_k)} \tag{2.20}$$

for $\ell \in \{0, \dots, n-1\}$, we have $F_n = F_n^{\{1\}} + F_n^{\{2\}}$, where

$$F_n^{\{1\}} := \begin{pmatrix} \Re\left(G_{2,n}(I_n)\right) \\ \Im\left(G_{2,n}(I_n)\right) \\ \vdots \\ \Re\left(G_{p-1,n}(I_n)\right) \\ \Im\left(G_{p-1,n}(I_n)\right) \\ \Re\left(G_{p+1,n}(I_n)\right) \\ \Im\left(G_{p+1,n}(I_n)\right) \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} \Re\left(n^{\lambda_2} g_2(U)\right) \\ \Im\left(n^{\lambda_2} g_2(U)\right) \\ \vdots \\ \Re\left(n^{\lambda_{p-1}} g_{p-1}(U)\right) \\ \Im\left(n^{\lambda_{p-1}} g_{p-1}(U)\right) \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and $F_n^{\{2\}}$ is given by the sum

$$\begin{pmatrix} \Re\left(\left(I_n^{\lambda_2} - (nU)^{\lambda_2}\right)\Xi_2^{\{1\}} + \lambda_2\left(J_n^{\lambda_2} - (n(1-U))^{\lambda_2}\right)\Xi_2^{\{2\}}\right) \\ \Im\left(\left(I_n^{\lambda_2} - (nU)^{\lambda_2}\right)\Xi_2^{\{1\}} + \lambda_2\left(J_n^{\lambda_2} - (n(1-U))^{\lambda_2}\right)\Xi_2^{\{2\}}\right) \\ \vdots \\ \Re\left(\left(I_n^{\lambda_{p-1}} - (nU)^{\lambda_{p-1}}\right)\Xi_{p-1}^{\{1\}} + \lambda_{p-1}\left(J_n^{\lambda_{p-1}} - (n(1-U))^{\lambda_{p-1}}\right)\Xi_{p-1}^{\{2\}}\right) \\ \Im\left(\left(I_n^{\lambda_{p-1}} - (nU)^{\lambda_{p-1}}\right)\Xi_{p-1}^{\{1\}} + \lambda_{p-1}\left(J_n^{\lambda_{p-1}} - (n(1-U))^{\lambda_{p-1}}\right)\Xi_{p-1}^{\{2\}}\right) \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Note that $\mathcal{D} M_q \mathcal{D}^t = M_q$.

2.2.3 Contraction method

By decomposition (2.19), we prepared a proof of Proposition 2.1.3 that is based on the contraction method. The contraction method is an approach for the derivation of weak convergence that originates in the analysis of algorithms and data structures and that recently has proved successful in the context of Pólya urn schemes as well, see [47]. In order to apply the method, in the first place, information about the asymptotics of moments is needed. The basic proceeding then is the following: Starting from a distributional recurrence for a normalised sequence $(V_n)_{n\geq 0}$ of random variables under consideration, a distributional fixed point equation for a potential limit is guessed. Then, the fixed point equation is used to define a contractive (with respect to an appropriate probability metric) measure-valued map that has a unique fixed point. Finally, using the recurrence and the fixed point equation, convergence of the distributions $(\mathcal{L}(V_n))_{n\geq 0}$ with respect to the probability metric is shown, and if this implies weak convergence, we are done.
Note that this approach yields a characterisation of the limit as a fixed point of measure-valued maps in appropriate spaces. However, as in the derivation of limit theorems via martingale arguments, the limit distribution usually is not given explicitly and it can be a hard task to derive some of its properties.
The contraction method had its debut in algorithmic contexts in the analysis of the Quick-

sort algorithm in Rösler [71]. Subsequently, it was independently developed by Rösler [72] and Rachev and Rüschendorf [68]. Until 2001, most applications from computer science were treated using the minimal L_2-metric ℓ_2. Neininger and Rüschendorf [62,63] further developed the contraction method for ideal metrics. The most prominent example for these metrics is the Zolotarev metric, which will also be used in the present proof. Since its introduction, the contraction method has been successfully applied to Quicksort, depths of nodes in random binary search trees, the profile of random binary search trees, the Wiener index of random binary search trees, size of random m-ary search trees, the size and path lengths of random tries, mergesort, randomized game tree evaluation, maxima in right triangles, the size of critical Galton Watson trees and broadcast communication models, for example. For an overview in German, see [60].

Problem Structure. The general setting in which the contraction method is usually applied, is the following: Let $(\tilde{V}_n)_{n\in\mathbb{N}_0}$ be a sequence of $d-$dimensional random variables and $n_0 \in \mathbb{N}$. We assume that the first n_0 variables $\tilde{V}_0, \dots, \tilde{V}_{n_0-1}$ are given initialising random vectors, and that for $n \geq n_0$, the sequence satisfies a distributional recursion of the form

$$\mathcal{L}\left(\tilde{V}_n\right) = \mathcal{L}\left(\sum_{j=1}^{K} \tilde{A}_n^{\{j\}} \tilde{V}^{\{j\}}_{I_n^{(j)}} + \tilde{b}_n\right), \qquad n \geq n_0.$$

In the above, $K \geq 1$ is a number which typically is deterministic, $\tilde{A}_n^{\{1\}}, \dots, \tilde{A}_n^{\{K\}}$ are random $d \times d$-matrices, $\tilde{b}_n$ is a random $d-$dimensional vector and I_n is a vector whose components $I_n^{(j)}$ are elements of $\{0, \dots, n\}$ for $j \in \{1, \dots, K\}$. Furthermore, $(\tilde{V}_n^{\{1\}})_{n\geq 0}, \dots, (\tilde{V}_n^{\{K\}})_{n\geq 0}$ are identically distributed as $(\tilde{V}_n)_{n\geq 0}$ and $(\tilde{A}_n^{\{1\}}, \dots, \tilde{A}_n^{\{K\}}, \tilde{b}_n, I_n)_{n\geq 0}$, $(\tilde{V}_n^{\{1\}})_{n\geq 0}, \dots, (\tilde{V}_n^{\{K\}})_{n\geq 0}$ are independent.
In a next step, the random sequence $(\tilde{V}_n)_{n\geq 0}$ is normalised. That is, we define

$$V_n := C_n^{-1}(\tilde{V}_n - M_n), \quad n \geq 0,$$

for vectors $M_n \in \mathbb{R}^d$ and symmetric, positive definite $d \times d$-matrices C_n. These are usually of the order of the expectation and the covariance matrix of $\tilde{V}_n$, if they exist. Accordingly, there is a similar recursion for $(V_n)_{n\geq 0}$,

$$\mathcal{L}(V_n) = \mathcal{L}\left(\sum_{j=1}^{K} A_n^{\{j\}} V^{\{j\}}_{I_n^{(j)}} + b_n\right), \quad n \geq n_0. \tag{2.21}$$

Here, independence conditions and distributional relations are as in the original recursion, and

$$A_n^{\{j\}} := C_n^{-1/2} \tilde{A}_n^{\{j\}} C^{1/2}_{I_n^{(j)}}, \quad b_n := C_n^{-1/2}\left(\tilde{b}_n - M_n + \sum_{j=1}^{K} \tilde{A}_n^{\{j\}} M_{I_n^{(j)}}\right).$$

Limit equations. If, indeed, $(V_n)_{n\geq 0}$ converges weakly to some random vector V and, additionally, the coefficients in (2.21) converge in an appropriate sense, the right hand side of (2.21) can be regarded as a transformation of a single probability measure on $\mathbb{R}^d$. More precisely, if $A_n^{\{j\}} \to A_j$, $b_n \to b$ as $n \to \infty$ (in some sense), it is plausible from equation (2.21), that the distribution of V satisfies a fixed point equality of the form

$$\mathcal{L}(V) = \mathcal{L}\left(\sum_{j=1}^{K} A^{\{j\}} V^{\{j\}} + b\right), \tag{2.22}$$

where $(A^{\{1\}}, \dots, A^{\{K\}}, b)$, $V^{\{1\}}, \dots, V^{\{K\}}$ are independent and $\mathcal{L}(V^{\{j\}}) = \mathcal{L}(V)$ for $j = 1, \dots, K$. In order to turn this observation into a proof strategy, let $\mathcal{M}$ denote the space of all probability measures on $\mathbb{R}^d$ and let T be a map on probability distributions, defined by

$$T : \mathcal{M} \to \mathcal{M}, \quad \mu \mapsto \mathcal{L}\left(\sum_{j=1}^{K} A^{\{j\}} Z^{\{j\}} + b\right), \tag{2.23}$$

where, again, $(A^{\{1\}}, \dots, A^{\{K\}}, b)$, $Z^{\{1\}}, \dots, Z^{\{K\}}$ are independent and $\mathcal{L}(Z^{\{j\}}) = \mu$ for $j = 1, \dots, K$. If there is a weak limit V of $(V_n)_{n\geq 0}$ with property (2.22), its distribution should be a fixed point of the map T. Moreover, asymptotically, $\mathcal{L}(V_{n+1})$ should be close to $T(\mathcal{L}(V_n))$.
It now stands to reason to think of the setting of Banach's fixed point theorem, and this is exactly what the contraction method formalises, as the name suggests. The first step in the reasoning is to equip a suitable subspace $\mathcal{M}^* \subset \mathcal{M}$ with a complete metric δ, in such a way that the restriction of T to $\mathcal{M}^*$ is a contraction in the metric space $(\mathcal{M}^*, \delta)$. In a second step, convergence of $\mathcal{L}(V_n)$ to the distribution of the fixed point in δ is shown. If the metric implies weak convergence, convergence in distribution follows.
Note that in particular, this proceeding yields a unique fixed point of T in $\mathcal{M}^*$. Without restricting T to an appropriate subspace $\mathcal{M}^* \subset \mathcal{M}$, this usually is not true, as maps as (2.23) often do not have unique fixed points in the space of all probability measures. However, the fixed points that arise in the analysis of algorithms as limits are usually characterised by additional moment conditions.

2.2.4 The Zolotarev metric

As briefly mentioned, in Proposition 2.1.3, we show weak convergence by means of the stronger convergence within the Zolotarev metric. The Zolotarev metric has been studied systematically in the context of distributional recurrences in [62]. In this place, we restrict ourselves to giving definitions of the relevant quantities and properties.
Let X and Y be two $\mathbb{R}^d$-valued random variables. The Zolotarev distance $\zeta_3(X, Y)$ of their distributions is defined as

$$\zeta_3(X, Y) := \zeta_3\left(\mathcal{L}(X), \mathcal{L}(Y)\right) := \sup_{f \in \mathcal{F}_3} |\mathbb{E}[f(X) - f(Y)]|,$$

where

$$\mathcal{F}_3 := \left\{ f \in C^2(\mathbb{R}^d, \mathbb{R}) : \|D^2 f(x) - D^2 f(y)\|_{\mathrm{op}} \leq \|x - y\|, \ \ x, y \in \mathbb{R}^d \right\}.$$

In words, $\mathcal{F}_3$ is the set of all twice continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}$, whose second derivative is 1-Hölder continuous.

We call a pair (X, Y) ζ_3-compatible if the expectation and the covariance matrix of X and Y coincide and if both $\|X\|_3, \|Y\|_3 < \infty$. It is known that this implies the finiteness of the Zolotarev distance, $\zeta_3(X, Y) < \infty$. An important property of ζ_3 is that it is $(3, +)$-ideal:

$$\zeta_3(X + Z, Y + Z) \leq \zeta_3(X, Y), \quad \zeta_3(cX, cY) = c^3 \zeta_3(X, Y)$$

for random vectors X, Y, Z, where Z is independent of X, Y and $c > 0$. For a linear transformation A of $\mathbb{R}^d$, we further have

$$\zeta_3(AX, AY) \leq \|A\|_{op}^3 \zeta_3(X, Y). \tag{2.24}$$

The use of the Zolotarev metric ζ_3 for our purposes requires a slightly modified version of recurrence (2.19), as we need to ensure its finiteness in the first place. Consequently, by the above, we need to find a sequence of random vectors that is sufficiently close to $(Z_n)_{n\geq 0}$ and has fixed covariance matrix M_q to guarantee the finiteness of the corresponding Zolotarev distances ζ_3.

As noted in section 2.2.1, the covariance matrices $(\mathrm{Cov}(Z_n))_{n\geq 0}$ converge componentwise to M_q, and their limit M_q is invertible. Thus, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $\mathrm{Cov}(Z_n)$ is invertible. If we set

$$\Sigma_n := \mathbb{1}_{\{n<n_0\}} \mathrm{Id}_{\mathbb{R}^{q-1}} + \mathbb{1}_{\{n\geq n_0\}} M_q^{1/2} \mathrm{Cov}(Z_n)^{-1/2}, \tag{2.25}$$

Σ_n is invertible for all $n \geq 0$ and we see that $\Sigma_n Z_n$ has covariance matrix M_q for all $n \geq n_0$. Let

$$N_n := \Sigma_n Z_n = A_n^{\{1\}} N_{I_n}^{\{1\}} + A_n^{\{2\}} N_{J_n}^{\{2\}} + b_n, \tag{2.26}$$

where the right hand side is a recursive decomposition of N_n with coefficients

$$A_n^{\{1\}} := \Sigma_n \rho_n \rho_{I_n}^{-1} \Sigma_{I_n}^{-1}, \quad A_n^{\{2\}} := \Sigma_n \rho_n \rho_{J_n}^{-1} \mathcal{D} \Sigma_{J_n}^{-1}, \quad b_n := \Sigma_n \rho_n \left(F_n^{\{1\}} + F_n^{\{2\}} \right).$$

We conclude the current subsection with a lemma that will be used in the proof of Proposition 2.1.3 and can be proved similarly to Lemma 2.1 in [61].

Lemma 2.2.3. *Let V_1, V_2, W_1, W_2 be random variables in $\mathbb{R}^d$ such that (V_1, V_2) and $(V_1 + W_1, V_2 + W_2)$ are ζ_3-compatible. Then, we have*

$$\zeta_3(V_1 + W_1, V_2 + W_2) \leq \zeta_3(V_1, V_2) + \sum_{i=1}^{2} \left(\|V_i\|_3^2 \|W_i\|_3 + \frac{\|V_i\|_3 \|W_i\|_3^2}{2} + \frac{\|W_i\|_3^3}{2} \right).$$

2.2.5 Preparatory lemmata

In this section, we collect some technical lemmata that are needed for the proof of Proposition 2.1.3 in the next section. These concern asymptotics for the coefficients that appear in decompositions (2.19) and (2.26) of Z_n and N_n, respectively. We first look at the asymptotics of the coefficients arising in (2.26).

Lemma 2.2.4. *For all* $1 \le s < \infty$, *as* $n \to \infty$,

$$\left\| A_n^{\{1\}} - \sqrt{U} \cdot \mathrm{Id}_{\mathbb{R}^{q-1}} \right\|_s \longrightarrow 0 \qquad \text{and} \qquad \left\| A_n^{\{2\}} - \sqrt{1-U} \cdot \mathcal{D} \right\|_s \longrightarrow 0.$$

Proof. We first check almost sure convergence. Both $\sqrt{I_n/n}, \sqrt{(I_n \log I_n)/(n \log n)} \to \sqrt{U}$ and $\sqrt{J_n/n}, \sqrt{(J_n \log J_n)/(n \log n)} \to \sqrt{1-U}$ a.s. as $n \to \infty$. Also, because $I_n \to \infty$ a.s. as $n \to \infty$, both $\Sigma_n, \Sigma_{I_n}^{-1} \to \mathrm{Id}_{\mathbb{R}^{q-1}}$. The claim now follows for all $1 \le s < \infty$ by an application of the dominated convergence theorem. □

Lemma 2.2.5. *Let* $k \in \{1, \dots, p\}$, *i.e.* $\sigma_k > \frac{1}{2}$. *As* $n \to \infty$,

$$\left\| \left(\frac{I_n}{n} \right)^{\lambda_k} - U^{\lambda_k} \right\|_3 = O\left(n^{-1/2}\right).$$

Proof. Let $n \ge 1$ be arbitrary and recall that conditionally on $\{U = u\}$, I_n is binomially $\mathrm{Bin}(n-1, u)$-distributed. Let $B_{n-1,u}$ be a random variable with $\mathcal{L}(B_{n-1,u}) = \mathrm{Bin}(n-1, u)$. Then

$$\begin{aligned}
\left\| \left(\frac{I_n}{n} \right)^{\lambda_k} - U^{\lambda_k} \right\|_3^3 &= \int_0^1 \mathbb{E}\left[\left| \left(\frac{B_{n-1,u}}{n} \right)^{\lambda_k} - u^{\lambda_k} \right|^3 \right] du \\
= \int_0^{1/n} \mathbb{E}\left[\left| \left(\frac{B_{n-1,u}}{n} \right)^{\lambda_k} - u^{\lambda_k} \right|^3 \right] du &+ \int_{1/n}^1 \mathbb{E}\left[\left| \left(\frac{B_{n-1,u}}{n} \right)^{\lambda_k} - u^{\lambda_k} \right|^3 \right] du \\
=: \quad I^{(1)}(n) + I^{(2)}(n).
\end{aligned}$$

On $\left[0, \frac{1}{n}\right]$, the triangle inequality and $|\alpha + \beta|^s \le 2^{s-1}(|\alpha|^s + |\beta|^s)$ imply that

$$\begin{aligned}
I^{(1)}(n) &\le \int_0^{1/n} \mathbb{E}\left[\left(\left| \left(\frac{B_{n-1,u}}{n} \right)^{\sigma_k} \right| + |u^{\sigma_k}| \right)^3 \right] du \\
&\le 4 \int_0^{1/n} \mathbb{E}\left[\left| \frac{B_{n-1,u}}{n} \right|^{3\sigma_k} \right] + \mathbb{E}\left[u^{3\sigma_k} \right] du \\
&\le 4 \left(n^{-3\sigma_k - 1} \mathbb{E}\left[B_{n-1,1/n}^3 \right]^{3\sigma_k} + \frac{1}{3\sigma_k + 1} n^{-3\sigma_k - 1} \right) = O\left(n^{-3\sigma_k - 1} \right)
\end{aligned}$$

as $n \to \infty$.
On $\left[\frac{1}{n}, 1\right]$, the triangle inequality and $|\alpha + \beta|^s \le 2^{s-1}(|\alpha|^s + |\beta|^s)$ imply that

$$\begin{aligned}
& I^{(2)}(n) \\
\le \quad & 4 \int_{1/n}^1 \mathbb{E}\left[\left| \left(\frac{B_{n-1,u}}{n} \right)^{\sigma_k} - u^{\sigma_k} \right|^3 \right] du && (2.27) \\
& + \; 4|\mu_k|^3 \int_{1/n}^1 \mathbb{E}\left[\left| \left(\frac{B_{n-1,u}}{n} \right)^{\sigma_k} \log\left(\frac{B_{n-1,u}}{nu} \right) \right|^3 \right] du \\
=: \quad & 4 \left(J^{(1)}(n) + |\mu_k|^3 \cdot J^{(2)}(n) \right). && (2.28)
\end{aligned}$$

For $\varepsilon \in (0,1)$, let $E_n := \{B_{n-1,u} > (1-\varepsilon)un\}$. We further split

$$\begin{aligned}
J^{(1)}(n) &= \int_{1/n}^{1} \mathbb{E}\left[\left|\left(\frac{B_{n-1,u}}{n}\right)^{\sigma_k} - u^{\sigma_k}\right|^3 \mathbb{1}_{E_n}\right] du + \int_{1/n}^{1} \mathbb{E}\left[\left|\left(\frac{B_{n-1,u}}{n}\right)^{\sigma_k} - u^{\sigma_k}\right|^3 \mathbb{1}_{E_n^c}\right] du \\
&=: J^{(1,1)}(n) + J^{(1,2)}(n), \\
J^{(2)}(n) &= \int_{1/n}^{1} \mathbb{E}\left[\left|\left(\frac{B_{n-1,u}}{n}\right)^{\sigma_k} \log\left(\frac{B_{n-1,u}}{nu}\right)\right|^3 \mathbb{1}_{E_n}\right] du \\
&\quad + \int_{1/n}^{1} \mathbb{E}\left[\left|\left(\frac{B_{n-1,u}}{n}\right)^{\sigma_k} \log\left(\frac{B_{n-1,u}}{nu}\right)\right|^3 \mathbb{1}_{E_n^c}\right] du \\
&=: J^{(2,1)}(n) + J^{(2,2)}(n).
\end{aligned}$$

In order to show that $J^{(1,2)}(n)$ and $J^{(2,2)}(n)$ are sufficiently small, we can use the fact that $\mathbb{P}(E_n^c)$ is small. More precisely, Chernoff's inequality implies that

$$\mathbb{P}(E_n^c) \leq \exp\left(\frac{\varepsilon^2}{2} nu\right).$$

In a second preliminary note, we remark that for all $\alpha \geq 1$, there exists $C > 0$ such that

$$\mathbb{E}\left[|B_{n,u} - nu|^\alpha\right] \leq C(nu)^{\alpha/2} \tag{2.29}$$

for $u \geq 1/n$. Using Bernstein's inequality, this follows from

$$\begin{aligned}
\mathbb{E}\left[|B_{n,u} - nu|^3\right] &= \alpha \int_0^\infty y^{\alpha-1} \mathbb{P}\left(|B_{n,u} - nu| > y\right) dy \\
&\leq 2\alpha \int_0^\infty y^{\alpha-1} \exp\left(-\frac{y^2}{2u(1-u)n + 2y/3}\right) dy \\
&\leq 2\alpha \left(\int_0^{6nu} y^\alpha \exp\left(-\frac{y^2}{6nu}\right) dy + \int_{6nu}^\infty y^{\alpha-1} \exp(-y) dy\right) \\
&\leq C_1 (nu)^{\alpha/2} + C_2 \\
&\leq C(nu)^{\alpha/2}
\end{aligned}$$

as $nu \geq 1$.

Now,

$$\begin{aligned}
J^{(1,2)}(n) &= \int_{1/n}^{1} \mathbb{E}\left[\left|\left(\frac{B_{n-1,u}}{n}\right)^{\sigma_k} - u^{\sigma_k}\right|^3 \mathbb{1}_{E_n^c}\right] du \\
&\le n^{-3\sigma_k} \int_{1/n}^{1} \mathbb{E}\left[|(B_{n-1,u})^{\sigma_k} - (nu)^{\sigma_k}|^6\right]^{1/2} \mathbb{E}\left[\mathbb{1}_{E_n^c}\right]^{1/2} du \\
&\le n^{-3\sigma_k} \int_{1/n}^{1} \mathbb{E}\left[|B_{n-1,u} - nu|^{6\sigma_k}\right]^{1/2} \mathbb{E}\left[\mathbb{1}_{E_n^c}\right]^{1/2} du \\
&\overset{(2.29)}{\le} \sqrt{C} n^{-3\sigma_k/2} \int_{1/n}^{1} u^{3\sigma_k/2} \exp\left(-\frac{\varepsilon^2}{4} nu\right) du = O\left(n^{-3\sigma_k - 1}\right).
\end{aligned}$$

For $J^{(2,2)}(n)$, let $h_{\sigma_k} : [0, \infty) \to \mathbb{R}$ be the function $h_{\sigma_k}(x) := x^{\sigma_k} |\log(x)|$ (convention: $0 \cdot \log 0 := 0$). Then $\sup_{x \in [0,1]} |h_{\sigma_k}(x)| = \frac{1}{\sigma_k e} < \frac{2}{e} < 1$. It follows that

$$\begin{aligned}
J^{(2,2)}(n) &\le \int_{1/n}^{1} u^{3\sigma_k} \mathbb{E}\left[\left|h_{\sigma_k}\left(\frac{B_{n-1,u}}{nu}\right)\right|^3 \mathbb{1}_{E_n^c}\right] du \le \int_{1/n}^{1} u^{3\sigma_k} \exp\left(-\frac{\varepsilon^2}{2} nu\right) du \\
&= O\left(n^{-3\sigma_k - 1}\right).
\end{aligned}$$

We now turn to $J^{(1,1)}(n)$ and $J^{(2,1)}(n)$. Set $\varphi_{\sigma_k} : (-1, \infty) \to \mathbb{R}, \varphi_{\sigma_k}(x) := (1 + x)^{\sigma_k}$. Then for $x \ge -\varepsilon$, $|\varphi'_{\sigma_k}(x)| \le \sigma_k (1 - \varepsilon)^{\sigma_k - 1}$.

$$\begin{aligned}
J^{(1,1)}(n) &\le \int_{\frac{1}{n}}^{1} u^{3\sigma_k} \mathbb{E}\left[\left|\varphi_{\sigma_k}\left(\frac{B_{n-1,u} - nu}{nu}\right) - 1\right|^3 \mathbb{1}_{E_n}\right] du \\
&\le \sigma_k (1 - \varepsilon)^{\sigma_k - 1} n^{-3} \int_{\frac{1}{n}}^{1} u^{3\sigma_k - 3} \mathbb{E}\left[|B_{n-1,u} - nu|^3\right] du \\
&\overset{(2.29)}{\le} C n^{-3/2} \int_{\frac{1}{n}}^{1} u^{3\sigma_k - 3/2} du = O\left(n^{-3/2}\right).
\end{aligned}$$

Similarly, as $B_{n-1,u} > (1 - \varepsilon) nu$ on E_n,

$$\begin{aligned}
J^{(2,1)}(n) &\le \int_{1/n}^{1} u^{3\sigma_k} \mathbb{E}\left[\left|h_{\sigma_k}\left(\frac{B_{n-1,u}}{nu}\right) - h_{\sigma_k}(1)\right|^3 \mathbb{1}_{E_n}\right] du \\
&\le C n^{-3/2} \int_{\frac{1}{n}}^{1} u^{3\sigma_k - 3/2} du = O\left(n^{-3/2}\right).
\end{aligned}$$

In total, we obtain that

$$\left\| \left(\frac{I_n}{n}\right)^{\lambda_k} - U^{\lambda_k} \right\|_3 = O\left(n^{-\sigma_k - 1/3}\right) + O\left(n^{-1/2}\right) = O\left(n^{-1/2}\right) \qquad \text{as } n \to \infty.$$

□

Lemma 2.2.6. *As* $n \to \infty$,

$$\|b_n\|_3 \longrightarrow 0.$$

Proof. By the triangle inequality,

$$\|b_n\|_3 \leq \|\Sigma_n\|_{\mathrm{op}} \sum_{j=1}^{2} \left\|\rho_n F_n^{\{j\}}\right\|_3.$$

We have $\mathcal{L}\left(\left(I_n, U, \Xi_k^{\{1\}}\right)\right) = \mathcal{L}\left(\left(J_n, 1-U, \Xi_k^{\{2\}}\right)\right)$, where $\Xi_k^{\{1\}}$ is independent of (I_n, U). The triangle inequality implies

$$\left\|\rho_n F_n^{\{2\}}\right\|_3 \leq \frac{4}{\sqrt{n}} \sum_{k=1}^{\frac{p-1}{2}} n^{\sigma_{2k}} \left\|\Xi_{2k}^{\{1\}}\right\|_3 \left\|\left(\frac{I_n}{n}\right)^{\lambda_{2k}} - U^{\lambda_{2k}}\right\|_3 = \frac{4}{\sqrt{n}} \sum_{k=1}^{\frac{p-1}{2}} O\left(n^{\sigma_{2k}-1/2}\right) = o(1)$$

by Lemma (2.2.5). Also, for $n \to \infty$,

$$\begin{aligned}
\left\|\rho_n F_n^{\{1\}}\right\|_3 \leq \frac{2}{\sqrt{n}} &\left(\sum_{k=1}^{\frac{p-1}{2}} \left\|G_{2k,n}(I_n) - n^{\lambda_{2k}} g_{2k}(U)\right\|_3\right. \\
&\left. + \frac{1}{\sqrt{\log(n)}} \|G_{p+1,n}(I_n)\|_3 + \sum_{k=\frac{p+3}{2}}^{q/2-1} \|G_{2k,n}(I_n)\|_3\right) \\
\leq \frac{2}{\sqrt{n}} &\left(\sum_{k=1}^{\frac{p-1}{2}} \frac{2}{\Gamma(1+\lambda_{2k})} n^{\sigma_{2k}} \left\|\left(\frac{I_n}{n}\right)^{\lambda_{2k}} - U^{\lambda_{2k}}\right\|_3\right. \\
&\left. + \frac{1}{\sqrt{\log(n)}} \|G_{p+1,n}(I_n)\|_3 + \sum_{k=\frac{p+3}{2}}^{q/2-1} \|G_{2k,n}(I_n)\|_3\right) + o(1) = o(1)
\end{aligned}$$

as before. Now, the sequence $(\|\Sigma_n\|_{\mathrm{op}})_{n\geq 0}$ is convergent and thus bounded, which implies the claim. □

Finally, we use recursion (2.26) for N_n to show that the sequence $(\|N_n\|_3)_{n\geq 0}$ is bounded.

Lemma 2.2.7. *As* $n \to \infty$,

$$\|N_n\|_3 = O(1).$$

Proof. Recall that the composition vector X_n takes only finitely many values, the random variables Ξ_k have finite absolute moments of arbitrary order, see (2.13), and $\|\Sigma_n\|_{\mathrm{op}} \to 1$. Hence, we have $\|N_n\|_3 < \infty$ for all $n \geq 0$.

Recursion (2.26) implies that

$$\|N_n\| \leq \mathcal{Y}^{\{1\}} + \mathcal{Y}^{\{2\}} + \|b_n\|,$$

where $\mathcal{Y}^{\{1\}} := \left\|A_n^{\{1\}}\right\|_{op}\left\|N_{I_n}^{\{1\}}\right\|$, $\mathcal{Y}^{\{2\}} := \left\|A_n^{\{2\}}\right\|_{op}\left\|N_{J_n}^{\{2\}}\right\|$. For all $n \geq 0$,

$$\begin{aligned}\mathbb{E}\left[\|N_n\|^3\right] \leq \mathbb{E}\left[\left(\mathcal{Y}^{\{1\}}\right)^3\right] + \mathbb{E}\left[\left(\mathcal{Y}^{\{2\}}\right)^3\right] + \mathbb{E}\left[\|b_n\|^3\right] + 3\mathbb{E}\left[\left(\mathcal{Y}^{\{1\}}\right)^2\mathcal{Y}^{\{2\}}\right] \\ + 3\mathbb{E}\left[\left(\mathcal{Y}^{\{2\}}\right)^2\mathcal{Y}^{\{1\}}\right] + 3\mathbb{E}\left[\left(\mathcal{Y}^{\{1\}}\right)^2\|b_n\|\right] + 3\mathbb{E}\left[\mathcal{Y}^{\{1\}}\|b_n\|^2\right] \\ + 3\mathbb{E}\left[\left(\mathcal{Y}^{\{2\}}\right)^2\|b_n\|\right] + 3\mathbb{E}\left[\mathcal{Y}^{\{2\}}\|b_n\|^2\right] + 6\mathbb{E}\left[\mathcal{Y}^{\{1\}}\mathcal{Y}^{\{2\}}\|b_n\|\right]. \qquad (2.30)\end{aligned}$$

Set

$$\beta_n := 1 \vee \max_{0\leq k\leq n} \mathbb{E}\left[\|N_k\|^3\right].$$

By Lemma 2.2.6, $\mathbb{E}\left[\|b_n\|^3\right] \to 0$ as $n \to \infty$. Also,

$$\mathbb{E}\left[\left(\mathcal{Y}^{\{j\}}\right)^3\right] = \mathbb{E}\left[\left\|A_n^{\{j\}}\right\|_{op}^3 \sum_{k=0}^{n-1} \mathbb{1}_{\{I_n=k\}}\mathbb{E}\left[\|N_k\|^3\right]\right] \leq \mathbb{E}\left[\left\|A_n^{\{j\}}\right\|_{op}^3\right]\beta_{n-1}$$

for $j = 1, 2$.

To bound the summand $\mathbb{E}\left[\left(\mathcal{Y}^{\{1\}}\right)^2\mathcal{Y}^{\{2\}}\right]$, note that $\left\|A_n^{\{1\}}\right\|_{op}$ and $\left\|A_n^{\{2\}}\right\|_{op}$ are uniformly bounded in n. This implies that after conditioning on I_n, there is a constant $D > 0$ such that

$$\begin{aligned}\mathbb{E}\left[\left(\mathcal{Y}^{\{1\}}\right)^2\mathcal{Y}^{\{2\}}\right] &\leq D\mathbb{E}\left[\sum_{k=0}^{n-1} \mathbb{1}_{\{I_n=k\}}\mathbb{E}\left[\|N_k\|^2\right]\mathbb{E}\left[\|N_{n-1-k}\|\right]\right] \\ &\leq D\left(\max_{0\leq k\leq n-1}\|N_k\|_2^2\right)\left(\max_{0\leq k\leq n-1}\|N_k\|_1\right).\end{aligned}$$

Now, by construction, $\mathrm{Cov}(N_n) = M_q$ for all $n \geq n_0$, so $\max_{0\leq k\leq n-1}\|N_k\|_2^2 < K$ for some $K > 0$ and hence $\mathbb{E}\left[\left(\mathcal{Y}^{\{1\}}\right)^2\mathcal{Y}^{\{2\}}\right] = O(1)$. The same applies to $\mathbb{E}\left[\left(\mathcal{Y}^{\{2\}}\right)^2\mathcal{Y}^{\{1\}}\right]$.

All other summands in (2.30) can be bounded using Hölder's inequality. Combining all these bounds leads to the estimate

$$\mathbb{E}\left[\|N_n\|^3\right] \leq \left(\mathbb{E}\left[\left\|A_n^{\{1\}}\right\|_{op}^3 + \left\|A_n^{\{2\}}\right\|_{op}^3\right] + o(1)\right)\beta_{n-1} + O(1).$$

The asymptotics in Lemma 2.2.4 further imply

$$\mathbb{E}\left[\|N_n\|^3\right] \leq \left(\mathbb{E}\left[U^{3/2} + (1-U)^{3/2}\right] + o(1)\right)\beta_{n-1} + O(1) = \left(\frac{4}{5} + o(1)\right)\beta_{n-1} + O(1).$$

Now there exist $J \in \mathbb{N}$ and a constant $0 < E < \infty$ such that for all $n \geq J$, $\mathbb{E}\left[\|N_n\|^3\right] \leq (9/10)\beta_{n-1} + E$. Induction on n gives that for all $n \geq 0$, $\mathbb{E}\left[\|N_n\|^3\right] \leq \max\{\beta_J, 10E\}$. □

2.2.6 Proof of Proposition 2.1.3

Based on the asymptotics that were derived in the last two subsections, finally a proof of Proposition 2.1.3 is given. Convergence of $(N_n)_{n\geq 0}$ is shown with respect to the Zolotarev metric, which implies convergence in distribution of $(Z_n)_{n\geq 0}$.

Proof of Proposition 2.1.3. Proposition 2.1.3 states that as $n \to \infty$, $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}$, where $\mathcal{L}(\mathcal{N}) = \mathcal{N}(0, M_q)$. In order to prove this, we show that

$$\zeta_3(N_n, \mathcal{N}) \longrightarrow 0 \quad \text{as} \quad n \to \infty,$$

as Z_n does not have the correct covariance matrix. However, this is sufficient, as the difference $Z_n - N_n$ tends to 0 in probability. Also note that here, convergence in the Zolotarev metric implies weak convergence, and that $\mathcal{N}(0, M_q)$ is the unique solution to the distributional recursion

$$\mathcal{L}(\mathcal{N}) = \mathcal{L}\left(\sqrt{U'}\mathcal{N}^{\{1\}} + \sqrt{1-U'}\mathcal{D}\mathcal{N}^{\{2\}}\right)$$

with the given mean and covariance matrix. Here, $\mathcal{N}^{\{1\}}, \mathcal{N}^{\{2\}}$ and U' are independent, $\mathcal{L}(U') =$ unif$(0,1)$ and $\mathcal{L}\left(\mathcal{N}^{\{1\}}\right) = \mathcal{L}\left(\mathcal{N}^{\{2\}}\right) = \mathcal{N}(0, M_q)$.

First, we use decomposition (2.26) for N_n to define hybrid random variables $(Q_n)_{n\geq 0}$ that link $\mathcal{L}(N_n)$ to $\mathcal{N}(0, M_q)$ as follows: Let $\mathcal{N}^{\{1\}}$ and $\mathcal{N}^{\{2\}}$ be defined on the same probability space as $(X_n)_{n\geq 0}$, independent with distribution $\mathcal{N}(0, M_q)$ and also independent of the process $(X_n)_{n\geq 0}$. We eliminate the error term in the given recursion and set

$$Q_n := A_n^{\{1\}}\left(\mathbb{1}_{\{I_n<n_0\}}N_{I_n}^{\{1\}} + \mathbb{1}_{\{I_n\geq n_0\}}\mathcal{N}^{\{1\}}\right) + A_n^{\{2\}}\left(\mathbb{1}_{\{J_n<n_0\}}N_{J_n}^{\{2\}} + \mathbb{1}_{\{J_n\geq n_0\}}\mathcal{N}^{\{2\}}\right).$$

Now, Q_n does not need to have covariance matrix M_q like $\mathcal{N}^{\{1\}}$ and $\mathcal{N}^{\{2\}}$. However, note that I_n/n converges to the uniform random variable U almost surely. Together with (2.25), (2.26) and $\|b_n\|_3 \to 0$ (Lemma 2.2.6), we obtain

$$\mathrm{Cov}(Q_n) \longrightarrow M_q, \qquad \text{as } n \to \infty.$$

Again, in order to ensure finiteness of the Zolotarev metric, the covariance matrix of Q_n has to be adjusted. Due to the convergence above, $\mathrm{Cov}(Q_n)$ has full rank for all $n \geq n_1$. Without loss of generality, we assume that $n_1 \geq n_0$. This implies that we can find a deterministic sequence of matrices $(B_n)_{n\geq 0}$ with $\mathrm{Cov}(B_n Q_n) = M_q$ for all $n \geq n_1$ and $B_n \to \mathrm{Id}_{\mathbb{R}^{q-1}}$ componentwise and in operator norm as $n \to \infty$. We write $B_n = \mathrm{Id}_{\mathbb{R}^{q-1}} + K_n$ with $(K_n)_{n\geq 0}$ tending to 0 componentwise. Hence, with $\mathcal{N}$ as before and $n \geq n_1$, each pair of N_n, $(\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n$ and $\mathcal{N}$ is ζ_3-compatible and the triangle inequality implies

$$\zeta_3(N_n, \mathcal{N}) \leq \zeta_3(N_n, (\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n) + \zeta_3((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n, \mathcal{N}). \tag{2.31}$$

The right hand side is finite for all $n \geq n_1$.

First, we show that $\zeta_3((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n, \mathcal{N}) = \mathrm{o}(1)$ via an upper bound for ζ_3 given by the minimal L_3-metric ℓ_3. The minimal L_3-metric ℓ_3 is given by

$$\ell_3(X, Y) := \ell_3(\mathcal{L}(X), \mathcal{L}(Y)) := \inf\{\|X' - Y'\|_3 : \mathcal{L}(X) = \mathcal{L}(X'), \mathcal{L}(Y) = \mathcal{L}(Y')\}, \tag{2.32}$$

for all random vectors X, Y with $\|X\|_3, \|Y\|_3 < \infty$. For a ζ_3-compatible pair (X, Y), we have the inequality, see [19, Lemma 5.7],

$$\zeta_3(X, Y) \leq \left(\|X\|_3^2 + \|Y\|_3^2\right) \ell_3(X, Y).$$

As $\sup_{n\geq 0} \|Q_n\|_3 < \infty$ by Lemma 2.2.4 and the properties of the Gaussian distribution, also $\|(\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n\|_3$ is uniformly bounded in n. So there exists a finite constant $C > 0$ with

$$\zeta_3((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n, \mathcal{N}) \leq C\ell_3((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n, \mathcal{N})$$

for all $n \geq n_1$. In order to upper bound the latter ℓ_3-distance, note that the random vectors $\mathcal{N}$ and $\sqrt{U}\mathcal{N}^{\{1\}} + \sqrt{1-U}\mathcal{D}\mathcal{N}^{\{2\}}$ are identically distributed. Thus

$$\begin{aligned}
&\zeta_3((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n, \mathcal{N}) \\
&\quad\leq C\ell_3((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n, \mathcal{N}) \\
&\quad\leq C\left\|\left((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)A_n^{\{1\}} - \sqrt{U}\mathrm{Id}_{\mathbb{R}^{q-1}}\right)\mathcal{N}^{\{1\}} + \left((\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)A_n^{\{2\}} - \sqrt{1-U}\mathcal{D}\right)\mathcal{N}^{\{2\}}\right\|_3 \\
&\quad\leq C\left(\left\|(\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)A_n^{\{1\}} - \sqrt{U}\mathrm{Id}_{\mathbb{R}^{q-1}}\right\|_3 \left\|\mathcal{N}^{\{1\}}\right\|_3\right. \\
&\qquad\quad \left. + \left\|(\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)A_n^{\{2\}} - \sqrt{1-U}\mathcal{D}\right\|_3 \left\|\mathcal{N}^{\{2\}}\right\|_3\right) \longrightarrow 0.
\end{aligned}$$

To bound the first summand in (2.31), we split N_n into two parts and consider the vector

$$\Phi_n := A_n^{\{1\}}N_{I_n}^{\{1\}} + A_n^{\{2\}}N_{J_n}^{\{2\}}, \qquad n \geq 1.$$

An application of Lemma 2.2.3 to the sums $N_n = \Phi_n + b_n$ and $(\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n = Q_n + K_nQ_n$ gives for $n \geq n_1$ that

$$\begin{aligned}
\zeta_3(N_n, (\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n) \leq \zeta_3(\Phi_n, Q_n) + \|\Phi_n\|_3^2\|b_n\|_3 + \frac{1}{2}\|\Phi_n\|_3\|b_n\|_3^2 + \frac{1}{2}\|b_n\|_3^3 \\
\left(\|K_n\|_{\mathrm{op}} + \frac{1}{2}\|K_n\|_{\mathrm{op}}^2 + \frac{1}{2}\|K_n\|_{\mathrm{op}}^3\right)\|Q_n\|_3^3.
\end{aligned}$$

By construction, $\|K_n\|_{\mathrm{op}} \to 0$ and by Lemma 2.2.6, $\|b_n\|_3 \to 0$. Also, by Lemma 2.2.7, $\sup_{n\geq 1} \|\Phi_n\|_3 < \infty$ and $\sup_{n\geq 1} \|Q_n\|_3 < \infty$. This yields

$$\zeta_3(N_n, (\mathrm{Id}_{\mathbb{R}^{q-1}} + K_n)Q_n) \leq \zeta_3(\Phi_n, Q_n) + o(1).$$

The previous estimates and (2.31) imply for $n \geq n_1$

$$\begin{aligned}
\zeta_3(N_n, \mathcal{N}) \leq &\zeta_3\left(A_n^{\{1\}}N_{I_n}^{\{1\}} + A_n^{\{2\}}N_{J_n}^{\{2\}},\right. \\
&\left. A_n^{\{1\}}\left(\mathbb{1}_{\{I_n<n_0\}}N_{I_n}^{\{1\}} + \mathbb{1}_{\{I_n\geq n_0\}}\mathcal{N}^{\{1\}}\right) + A_n^{\{2\}}\left(\mathbb{1}_{\{J_n<n_0\}}N_{J_n}^{\{2\}} + \mathbb{1}_{\{J_n\geq n_0\}}\mathcal{N}^{\{2\}}\right)\right) + o(1).
\end{aligned} \tag{2.33}$$

Let $\Delta(n) := \zeta_3(N_n, \mathcal{N})$, which is finite for $n \geq n_0$. The right hand side in (2.33) is finite for all $n \geq 1$. In the expectations defining the Zolotarev distance, we condition on the value of I_n. To this end, let $(N_0^{[1]}, \ldots, N_{n-1}^{[1]}), (N_0^{[2]}, \ldots, N_{n-1}^{[2]})$ be i.i.d. with distribution $\mathcal{L}(N_0, \ldots, N_{n-1})$.

We then make use of independence and the fact that ζ_3 is $(3,+)$-ideal and satisfies (2.24) to get

$$\begin{aligned}
&\zeta_3\left(A_n^{\{1\}}N_{I_n}^{\{1\}}+A_n^{\{2\}}N_{J_n}^{\{2\}}, A_n^{\{1\}}\left(\mathbb{1}_{\{I_n<n_0\}}N_{I_n}^{\{1\}}+\mathbb{1}_{\{I_n\geq n_0\}}\mathcal{N}^{\{1\}}\right)+A_n^{\{2\}}\left(\mathbb{1}_{\{J_n<n_0\}}N_{J_n}^{\{2\}}+\mathbb{1}_{\{J_n\geq n_0\}}\mathcal{N}^{\{2\}}\right)\right)\\
&\leq \frac{1}{n}\sum_{k=n_0}^{n-1-n_0}\zeta_3\left(\Sigma_n\rho_n\rho_k^{-1}\Sigma_k^{-1}N_k^{[1]}+\Sigma_n\rho_n\rho_{n-1-k}\mathcal{D}\Sigma_{n-1-k}^{-1}N_{n-1-k}^{[2]},\right.\\
&\qquad\qquad\left.\Sigma_n\rho_n\rho_k^{-1}\Sigma_k^{-1}\mathcal{N}^{\{1\}}+\Sigma_n\rho_n\rho_{n-1-k}^{-1}\mathcal{D}\Sigma_{n-1-k}^{-1}\mathcal{N}^{\{2\}}\right)\\
&\quad+\frac{1}{n}\sum_{k=0}^{n_0-1}\zeta_3\left(\Sigma_n\rho_n\rho_{n-1-k}\mathcal{D}\Sigma_{n-1-k}^{-1}N_{n-1-k}^{[2]},\Sigma_n\rho_n\rho_{n-1-k}^{-1}\mathcal{D}\Sigma_{n-1-k}^{-1}\mathcal{N}^{\{2\}}\right)\\
&\quad+\frac{1}{n}\sum_{k=n-n_0}^{n-1}\zeta_3\left(\Sigma_n\rho_n\rho_k^{-1}\Sigma_k^{-1}N_k^{[1]},\Sigma_n\rho_n\rho_k^{-1}\Sigma_k^{-1}\mathcal{N}^{\{1\}}\right)\\
&\leq \frac{1}{n}\sum_{k=n_0}^{n-1-n_0}\left(\|\rho_n\rho_k^{-1}\|_{op}^3\|\Sigma_n\|_{op}^3\|\Sigma_k^{-1}\|_{op}^3\zeta_3\left(N_k^{[1]},\mathcal{N}^{\{1\}}\right)\right.\\
&\qquad\qquad\left.+\|\rho_n\rho_{n-1-k}^{-1}\|_{op}^3\|\Sigma_n\|_{op}^3\|\Sigma_{n-1-k}^{-1}\|_{op}^3\zeta_3\left(N_{n-1-k}^{[2]},\mathcal{N}^{\{2\}}\right)\right)\\
&\quad+\frac{1}{n}\sum_{k=0}^{n_0-1}\|\rho_n\rho_{n-1-k}^{-1}\|_{op}^3\|\Sigma_n\|_{op}^3\|\Sigma_{n-1-k}^{-1}\|_{op}^3\zeta_3\left(N_{n-1-k}^{[2]},\mathcal{N}^{\{2\}}\right)\\
&\quad+\frac{1}{n}\sum_{k=n-n_0}^{n-1}\|\rho_n\rho_k^{-1}\|_{op}^3\|\Sigma_n\|_{op}^3\|\Sigma_k^{-1}\|_{op}^3\zeta_3\left(N_k^{[1]},\mathcal{N}^{\{1\}}\right)\\
&=\frac{2}{n}\sum_{k=n_0}^{n-1}\|\rho_n\rho_k^{-1}\|_{op}^3\|\Sigma_n\|_{op}^3\|\Sigma_k^{-1}\|_{op}^3\zeta_3\left(N_k^{[1]},\mathcal{N}^{\{1\}}\right).
\end{aligned}$$

Note that $\|\rho_n\rho_{I_n}^{-1}\|_{op}^3=\left(\frac{I_n}{n}\right)^{3/2}$ in both cases $6\mid q$ and $6\nmid q$. Hence, for $6\mid q$ and $n\geq n_1$,

$$\Delta(n)\leq 2\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{3/2}\|\Sigma_n\|_{op}^3\|\Sigma_{I_n}^{-1}\|_{op}^3\Delta(I_n)\mathbb{1}_{\{I_n\geq n_0\}}\right]+o(1).$$

A standard argument shows that $\zeta_3(N_n,\mathcal{N})\to 0$ as $n\to\infty$ (see [61], for example). □

We can now prove Theorems 2.1.1 and 2.1.2:

Proof of Theorem 2.1.1. Note that $6\nmid q$ implies that there is no $k\in\{1,\ldots,q\}$ with $\sigma_k=\frac{1}{2}$.

We obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k}\Xi_k v_k\right)\\
&= \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{p}\left[u_k(X_n - \mathbb{E}[X_n]) - n^{\lambda_k}\Xi_k\right] v_k + \sum_{k=p+1}^{q} u_k(X_n - \mathbb{E}[X_n])v_k\right)\\
&= 2Z_n^{(1)}\Re(v_2) - 2Z_n^{(2)}\Im(v_2) + 2Z_n^{(3)}\Re(v_4) - 2Z_n^{(4)}\Im(v_4)\cdots + \mathbb{1}_{\{q \text{ even}\}}Z_n^{(q-1)}v_q\\
&\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma^{(q)}\right),
\end{aligned}
$$

by Proposition 2.1.3 and the continuous mapping theorem, where $\Sigma^{(q)}$ is as in the statement of Theorem 2.1.1. Recall that M_q is a diagonal matrix with non-zero diagonal entries. As $\Re(v_2), \Im(v_2), \dots, v_q$ (assuming $2 \mid q$) are linearly independent, the rank of the covariance matrix $\Sigma^{(q)}$ is $q-1$. □

Proof of Theorem 2.1.2. Note that $6 \mid q$ implies that there are two eigenvalues with real parts $\sigma_{q/3} = \sigma_{q/3+1} = \frac{1}{2}$. Rearranging terms as in the proof of Theorem 2.1.1, we obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{n\log(n)}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k}\Xi_k v_k\right)\\
&= \frac{1}{\sqrt{\log(n)}}\sum_{k=1, k\neq q/6}^{q/2-1} 2\left(Z_n^{(2k-1)}\Re(v_{2k}) - Z_n^{(2k)}\Im(v_{2k})\right)\\
&+ 2\left(Z_n^{(q/3-1)}\Re(v_{q/3}) - Z_n^{(q/3)}\Im(v_{q/3})\right) + \frac{1}{\sqrt{\log(n)}}Z_n^{(q-1)}v_q \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma^{(q)}\right),
\end{aligned}
$$

by Proposition 2.1.3 and Slutsky's Lemma, where $\Sigma^{(q)}$ is as in Theorem 2.1.2. Again, $\Re(v_{q/3})$ and $\Im(v_{q/3})$ are linearly independent and $(M_q)_{q/3-1,q/3-1}, (M_q)_{q/3,q/3} > 0$. As all other coefficients tend to zero in probability, $\Sigma^{(q)}$ has rank 2. □

2.3 Construction of limits

The current section is devoted to a closer study of the random variables $\Xi_1, \dots \Xi_p$ that arise in the cyclic urn process. In the previous section, we claimed their existence by means of martingale arguments. Using a relation between the cyclic urn dynamics and the random binary search tree dynamics, we below extend this existence result and provide an explicit construction of the random variables $\Xi_1, \dots, \Xi_p$ from a sequence of i.i.d. uniform random variables $(U_n)_{n\geq 1}$. From this representation, the self-similarity relation (2.18) can be read off as well. A related explicit construction of the Quicksort limit can be found in [7].
However, recent developments in the limit theory of combinatorial Markov chains shed a different light on the above-mentioned constructions: It turns out that there is a deeper connection between these representations and *convergence* of the random binary search tree itself. The random binary search tree per se is a transient Markov chain; however, methods

from discrete potential theory can be applied to yield a limit object that is a random variable in the Markov chain's *Doob-Martin boundary.* This approach is developed in [22] for the general setting of so-called *trickle down* processes and more specifically for the random binary search tree in [32]. The idea behind this construction is to provide a unifying perspective for convergence of various functionals of combinatorial Markov chains. In our case, it turns out that the representation of the martingale limits $\Xi_1, \dots, \Xi_p$ with respect to $(U_n)_{n\geq 1}$ is indeed a deterministic function of the Doob-Martin limit of the random binary search tree.
The proceeding is as follows: First, a brief overview on the general construction of the Doob-Martin compactification is given. Second, following [32], the construction is applied to the random binary search tree sequence and some results about its limit are stated. Finally, $\Xi_1, \dots, \Xi_p$ are constructed and their connection to the Doob-Martin limit as well as some futher properties are established.

2.3.1 Doob-Martin compactification for discrete-time Markov chains

In order to put the following construction into perspective, we briefly explain the concept of the Doob-Martin compactification for the class of Markov chains that is treated in [22]. We therefore follow the display in [22]. For further information, see [17] or [79].
Assume that S is a countable set that is partially ordered by "$\leq$"; and furthermore, that there exists a unique minimal element $e \in S$. Let now $X = (X_n)_{n\in\mathbb{N}_0}$ be a discrete-time transient Markov chain with state space S and transition matrix P, that can reach any state from e with positive probability. More precisely, for all states $x \in S$, $\sum_{n=0}^{\infty} P^n(e,x) < \infty$ (transience) and there is $n = n_x \in \mathbb{N}_0$ such that $P^n(e,x) > 0$ (x may be reached from e). In our case, S is the set of all finite binary trees and e the tree that consists of the root node only. Indeed, in [22], even a stronger form of transience is assumed, namely that the chain may only progress to greater elements and $P(x,y) = 0$ unless $x < y$.
Now, the goal is to embed S into a compact space that provides a limit for X. The proceeding is quite similar to the proceeding in the Stone-Čech compactification: The Green kernel $G : S \times S \to \mathbb{R}_+$ of P is defined as

$$G(x,y) := \sum_{n=0}^{\infty} P^n(x,y) = \mathbb{P}_x\left(X_n = y \text{ for some } n \in \mathbb{N}_0\right) =: \mathbb{P}_x(X \text{ hits } y)$$

for $x, y \in S$. The assumptions imply that $0 < G(e,y) < \infty$ for all $y \in S$.
Using the Green kernel, we further define an appropriate set F of functions on S in which we will embed the space. A function $f : S \to \mathbb{R}_+$ is termed non-negative superharmonic (respectively, non-negative harmonic) if

$$\sum_{y\in S} P(x,y)f(y) =: Pf(x) \leq f(x)$$

for all $x \in S$ (respectively, $Pf(x) = f(x)$ for all $x \in S$). Let F be the set of all non-negative superharmonic functions on S that take the value 1 at the minimal element e. It can be shown that F is a compact convex metrisable subset of the locally convex topological vector space

$\mathbb{R}^S$. The Martin kernel with reference state e is given by

$$K(x,y) := \frac{G(x,y)}{G(e,y)} = \frac{\mathbb{P}_x(X \text{ hits } y)}{\mathbb{P}_e(X \text{ hits } y)}, \quad x,y \in S.$$

The functions $K(\cdot, y)$ are non-harmonic elements of F for all $y \in S$. More precisely, they are exactly the non-harmonic extreme points of F. Set

$$\iota : S \to F, \quad \iota(x) := K(\cdot, x).$$

The Riesz decomposition theorem implies that the map ι is injective. This allows us to identify the state space S with its image $\iota(S) \subset F$. Furthermore, $\iota(S)$ is dense in its compact closure $\overline{\iota(S)}$ in F. The compact metrisable space $\overline{\iota(S)}$, constructed from S via e and P is called the *Doob-Martin compactification* of S. Its *Doob-Martin boundary* is given by

$$\partial S := \overline{\iota(S)} \setminus \iota(S).$$

We close this section by noting that a sequence $(y_n)_{n\in\mathbb{N}}$ in S converges to a point in the Doob-Martin compactification $\overline{\iota(S)}$ if and only if the sequence of real numbers $(K(x, y_n))_{n\in\mathbb{N}}$ converges for all $x \in S$.

2.3.2 Doob-Martin boundary for the BST chain

The Doob-Martin compactification for the BST chain is also determined in [22]. We present a more direct approach from [32] that uses an algorithmic construction. The latter approach has the advantage that it leads to a description of the limit in terms of the input sequence of the algorithm. Correspondingly, the account of the results in the present subsection proceeds along the lines of [32].
Before we address ourselves to the binary search tree algorithm, we need to introduce some notation. Binary trees are formally defined by means of a label set $\mathbb{V}$, where

$$\mathbb{V}_k := \{0,1\}^k, \quad \mathbb{V} := \{0,1\}^* := \bigcup_{k=0}^{\infty} \mathbb{V}_k, \qquad \partial\mathbb{V} := \{0,1\}^\infty. \tag{2.34}$$

In the above, $\mathbb{V}_0 = \{\emptyset\}$, and for $k \in \mathbb{N}$, $\mathbb{V}_k$ is the set of words of length k in the alphabet $\{0,1\}$. Further, $\mathbb{V}$ is the set of all finite words in $\{0,1\}$ and $\partial\mathbb{V}$ the set of all infinite words in $\{0,1\}$. By interpreting each $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \mathbb{V}$ as an individual of the population $\mathbb{V}$, we may use a family-based language for relations between words. The father $\pi(\vartheta)$, the left child $\vartheta 0$ and the right child $\vartheta 1$ of ϑ are defined as

$$\pi(\vartheta) := (\vartheta_1, \dots, \vartheta_{k-1}), \quad \text{if} \quad k \geq 1, \tag{2.35}$$
$$\vartheta 0 := (\vartheta_1, \dots, \vartheta_k, 0), \tag{2.36}$$
$$\vartheta 1 := (\vartheta_1, \dots, \vartheta_k, 1), \tag{2.37}$$

respectively. More generally, we write $\vartheta \leq \kappa$ for $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \mathbb{V}$, $\kappa = (\kappa_1, \dots, \kappa_\ell) \in \mathbb{V}$, if $k \leq \ell$ and $\vartheta_j = \kappa_j$ for $j = 1, \dots, k$, that is, if ϑ is an ancestor of κ. Furthermore, for each $\vartheta, \kappa \in \mathbb{V}$, there exists a unique last common ancestor $\vartheta \wedge \kappa$. This defines a partial order (the prefix order) on $\mathbb{V}$, and we denote by $|\vartheta|$ the index k of the set $\mathbb{V}_k$ that contains ϑ, or, in other

words, the generation of $\vartheta \in \mathbb{V}$. The partial order and the concept of a last common ancestor can easily be extended to $\kappa \in \partial\mathbb{V}$ and $\vartheta, \kappa \in \mathbb{V} \cup \partial\mathbb{V}$, respectively.
A *binary tree* is a subset $t \subset \mathbb{V}$ of the label set such that

(i) $\emptyset \in t$.

(ii) For all $\vartheta \in t \setminus \{\emptyset\}$, $\pi(\vartheta) \in t$.

The countable set of all finite binary trees is denoted by $\mathbb{B}$, while $\mathbb{B}_n := \{t \in \mathbb{B} : \#t = n\}$ denotes the trees that have exactly n elements. $\mathbb{B}$ is the state space of the BST Markov chain, and its unique minimal element is $\{\emptyset\}$ in the prefix order. Binary trees can also be depicted as directed graphs, whose nodes are the elements of t ($\emptyset$ is the root node) and whose edges are the pairs $(\pi(\vartheta), \vartheta)$, for $\vartheta \in t \setminus \{\emptyset\}$.
Finally, the elements $\vartheta \in t$ are called *internal nodes* of t. On the contrary, elements of $\partial t := \{\vartheta \in \mathbb{V} : \pi(\vartheta) \in t, \vartheta \notin t\}$, the set of children of leaf nodes in t, are called *external nodes* of t.

BST chain. Binary search trees (BSTs) are fundamental data structures that gain additional importance from their connection to Quicksort, which is one of the most popular sorting algorithms. We present a recursive construction of labelled BSTs for a given deterministic sequence $(x_i)_{i\in\mathbb{N}}$ of pairwise distinct real numbers. This recursive construction is called the binary search tree (BST) algorithm, and the elements $x_1, x_2, \dots$ are the keys that are stored in the binary tree.
The BST algorithm generates a sequence of growing labelled binary trees $((t_n, \ell_n))_{n\in\mathbb{N}}$. Here, $t_n \in \mathbb{B}_n$ and ℓ_n is a function $\ell_n : t_n \to \{x_1, \dots, x_n\}$ that assigns an element of $\{x_1, \dots, x_n\}$ to each node in t_n. Alternatively, $\ell_n(\vartheta)$ is the key that is stored in the node ϑ. The binary trees are successively grown from $(x_i)_{i\in\mathbb{N}}$ as follows: At time $n = 1$, $t_1 = \{\emptyset\}$ and $\ell_1(\emptyset) = x_1$. If (t_n, ℓ_n) is already constructed, the tree t_{n+1} is obtained by adding an element of ∂t_n to t_n and by storing x_{n+1} in this element. The external node by which the tree is augmented is determined by the following procedure: At time $n + 1$, the key x_{n+1} is “inserted” at the root node of the tree t_n. Starting from there, it traverses some of the internal nodes of t_n. The way of x_{n+1} through the tree is prescribed by the values $\ell_n(\vartheta)$ of the internal nodes it encounters: If it passes by a node ϑ holding a key $\ell_n(\vartheta)$ greater than x_{n+1}, it moves to its left child, otherwise to its right child. The key stays with the first unoccupied node κ it visits, which is an external node by definition. We then set $t_{n+1} := t_n \cup \{\kappa\}$ and $\ell_{n+1}(\kappa) := x_{n+1}$, $\ell_{n+1}(\vartheta) := \ell_n(\vartheta)$ for all $\vartheta \in t_n$.
The random binary search tree is obtained by an application of this procedure to a sequence of random keys. More precisely, let $(U_i)_{i\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables with $\mathcal{L}(U_1) = \text{unif}(0, 1)$ and let $((T_n, L_n))_{n\in\mathbb{N}}$ be the sequence of labelled binary trees that are generated by this sequence. In particular, T_n is the random binary tree associated with the first n uniform keys. For a more detailed description of random binary search trees, see [54], for example.
In the following, we also make use of an alternative description of the growth dynamics of the sequence $(T_n)_{n\in\mathbb{N}}$: Recall that for each $n \geq 1$, T_n is generated by $U_1, \dots, U_n$. It is immediate from the definition of the BST algorithm, that only the rank of U_{n+1} relative to $U_1, \dots, U_n$ is important for the choice of the next node. By only looking at T_n, we do not know the exact values of $U_1, \dots U_n$. However, there are $n+1$ possible ranks for U_{n+1} among $U_1, \dots U_n$. There are also $n+1$ possible external nodes. The crucial observation is that each rank of U_{n+1}

corresponds to exactly one external node, which can be seen by showing that the allocation of possible ranks $1, \dots, n+1$ to external nodes is injective.
As a consequence, each external node is chosen with equal probability $\frac{1}{n+1}$ and then joined to the tree. Thus, $(T_n)_{n\in\mathbb{N}}$ is a Markov chain with state space $\mathbb{B}$, start at $T_1 = \{\emptyset\}$ and transition probabilities

$$\begin{cases} P(t, t \cup \{\vartheta\}) = \frac{1}{1+\#t}, & \vartheta \in \partial t, \\ 0, & \text{otherwise.} \end{cases} \tag{2.38}$$

$(T_n)_{n\in\mathbb{N}}$ is called the BST chain.
Denote by

$$0 =: U_{(n:0)} < U_{(n:1)} < \cdots < U_{(n:n)} < U_{(n:n+1)} := 1$$

the augmented order statistics among $U_1, \dots, U_n$. These partition the unit interval into $n+1$ subintervals of lengths $U_{(n:0)}, U_{(n:1)} - U_{(n:0)}, \dots, 1 - U_{(n:n)}$, and given $U_1, \dots, U_n$, the probability that U_{n+1} falls into intervall $(U_{(n:i)}, U_{(n:i+1)})$ is the interval length $U_{(n:i+1)} - U_{(n:i)}$. We now attach such an interval to each node $\vartheta \in T_n$, whose length gives the probability that a future node is a descendant of ϑ.
More precisely, denote by V_ϑ the key among $(U_i)_{i\in\mathbb{N}}$ that is inserted into node ϑ. We then construct recursively splitting intervals in the following way: Set $I_\emptyset := [0, 1]$, whose length corresponds to the probability that a future node is a descendant of the root node. If $I_\vartheta = [L_\vartheta, R_\vartheta]$ is already defined for $\vartheta \in \mathbb{V}$, we set $I_{\vartheta 0} = [L_\vartheta, V_\vartheta]$ and $I_{\vartheta 1} := [V_\vartheta, R_\vartheta]$. If ϑ is occupied by U_{n+1} for some $n \in \mathbb{N}$, I_ϑ is the same as the interval $[U_{(n:i)}, U_{(n:i+1)}]$ that encloses U_{n+1}, in the notation above. Furthermore, the length of the interval I_ϑ is denoted by $|I_\vartheta|$.

Doob-Martin compactification. One can view the Doob-Martin limit of the BST chain as an extension of the above probabilities to infinite words.
More formally, the Doob-Martin compactification of the BST chain is given in [22]: Let $[0, 1]^{\mathbb{V}}$ be the product space endowed with the topology of pointwise convergence, which turns it into a compact space. We define the size of the subtree of t rooted at ϑ as

$$\tau(t, \vartheta) := \#\{\kappa \in t : \vartheta \le \kappa\}. \tag{2.39}$$

The space $\mathbb{B}$ can be embedded into $[0, 1]^{\mathbb{V}}$ via the standardised subtree functional

$$\iota : \mathbb{B} \to [0, 1]^{\mathbb{V}}, \quad \iota(t) := \left(\vartheta \mapsto \frac{\tau(t, \vartheta)}{\#t}\right)$$

for $t \in \mathbb{B}$. Note also that $\frac{\tau(t,\vartheta)}{\#t}$ is roughly the probability that an external node of the subtree rooted at ϑ is chosen for replacement in the BST algorithm. The Doob-Martin compactification of the BST chain is given by the closure $\bar{\mathbb{B}}$ of the embedding ι of $\mathbb{B}$ into the compact space $[0, 1]^{\mathbb{V}}$.
Now by construction, the previously transient BST chain $(T_n)_{n\ge 1}$ converges almost surely in the Doob-Martin compactification $\bar{\mathbb{B}}$. Let X_∞ denote its limit, which takes values in the Doob-Martin boundary $\partial\mathbb{B}$. It can be seen that $\partial\mathbb{B}$ is homeomorphic to the set of probability measures μ on $(\partial\mathbb{V}, \mathcal{B}(\partial\mathbb{V}))$, where $\mathcal{B}(\partial\mathbb{V})$ is the σ-field generated by the projections on the

component spaces. Alternatively, $\mathcal{B}(\partial\mathbb{V})$ is generated by the sets

$$A_\vartheta := \{\kappa \in \partial\mathbb{V} : \kappa \geq \vartheta\}, \qquad \vartheta \in \mathbb{V}. \tag{2.40}$$

It is straightforward to verify that these sets form a π-system that generates $\mathcal{B}(\partial\mathbb{V})$. This in turn implies that each measure μ on $(\partial\mathbb{V}, \mathcal{B}(\partial\mathbb{V}))$ is completely determined by its values $\mu(A_\vartheta)$ for $\vartheta \in \mathbb{V}$. On these sets, convergence of a sequence $(t_n)_{n\in\mathbb{N}}$ in $\mathbb{B}$ to μ in the Doob-Martin topology requires that

$$\mu(A_\vartheta) = \lim_{n\to\infty} \frac{\tau(t_n, \vartheta)}{\# t_n} \quad \text{for all} \quad \vartheta \in \mathbb{V}.$$

Similarly, if $(\mu_n)_{n\in\mathbb{N}}$ is a sequence of elements of $\partial\mathbb{V}$ instead, Doob-Martin convergence is implied by $\mu_n(A_\vartheta) \to \mu(A_\vartheta)$ for all $\vartheta \in \mathbb{V}$.

The general theory in [22] also implies that X_∞ generates the tail σ-field

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(\{X_m : m \geq n\})$$

associated with the sequence $(X_n)_{n\in\mathbb{N}}$. A description of $\mathcal{L}(X_\infty)$ is given in [22].

Grübel [32] obtains a more direct representation of X_∞ in terms of the input sequence $(U_n)_{n\in\mathbb{N}}$ by an alternative interpretation of weak convergence in the Doob-Martin topology. More precisely, the partially ordered set $\mathbb{V}$ can be equipped with a metric $d_\mathbb{V}$, defined by

$$d_\mathbb{V}(\vartheta, \kappa) := 2^{-|\vartheta\wedge\kappa|} - \frac{1}{2}\left(2^{-|\vartheta|} + 2^{-|\kappa|}\right), \quad \vartheta, \kappa \in \mathbb{V}. \tag{2.41}$$

This metric induces the discrete topology on $\mathbb{V}$, and the completion of $\mathbb{V}$ with respect to $d_\mathbb{V}$ is given by $\bar{\mathbb{V}} := \mathbb{V} \cup \partial\mathbb{V}$. Now $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$ is a compact and separable metric space, and weak convergence in the Doob-Martin topology is equivalent to weak convergence of probability measures on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$: The generating sets of $\mathcal{B}(\partial\mathbb{V})$ find their counterparts in the sets

$$\bar{A}_\vartheta := \{\kappa \in \bar{\mathbb{V}} : \kappa \geq \vartheta\}, \qquad \vartheta \in \mathbb{V}.$$

These sets are open and closed; and $\mathcal{B}(\bar{\mathbb{V}})$ is generated by the π-system $\{\bar{A}_\vartheta : \vartheta \in \mathbb{V}\}$. A combination of the uniqueness of extension theorem and the Portemanteau lemma implies that a sequence $(\mu_n)_{n\geq 1}$ converges weakly to a probability measure μ on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$ if and only if

$$\lim_{n\to\infty} \mu_n(\bar{A}_\vartheta) = \mu(\bar{A}_\vartheta) \tag{2.42}$$

for all $\vartheta \in \mathbb{V}$.

Assume now that $(T_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{B}$ that converges to X_∞ in the Doob-Martin topology. For a subset $M \subset \mathbb{V}$, let unif(M) denote the uniform distribution on the elements of M, such that unif(M) is a probability measure on $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$. Then, as $n \to \infty$,

$$\text{unif}(T_n)(\bar{A}_\vartheta) = \frac{\tau(T_n, \vartheta)}{n} \longrightarrow X_\infty(A_\vartheta) = X_\infty(\bar{A}_\vartheta)$$

as $X_\infty(\mathbb{V}) = 0$. Conversely, if $\mathrm{unif}(T_n)(\bar{A}_\vartheta) \to \mu(\bar{A}_\vartheta)$ for all $\vartheta \in \mathbb{V}$, the sequence $(T_n)_{n\in\mathbb{N}}$ in $\mathbb{B}$ converges in the Doob-Martin topology.
If we thus identify finite subsets M of $\mathbb{V}$ with the uniform distribution $\mathrm{unif}(M)$ on M, which is an element of $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$, convergence of $(T_n)_{n\in\mathbb{N}}$ in the Doob-Martin topology is equivalent to weak convergence of the probability measures $(\mathrm{unif}(T_n))_{n\geq 1}$ on the metric space $(\bar{\mathbb{V}}, d_{\mathbb{V}})$. For more details on this, we refer the reader to [32]. We now state Theorem 1 from [32], which is the promised representation of the random measure X_∞ in terms of the input sequence.

Theorem 2.3.1 (Grübel). *Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with $\mathcal{L}(U_1) = \mathrm{unif}(0,1)$ and let $(T_n)_{n\in\mathbb{N}}$ be the sequence of random binary trees generated by the BST algorithm with input sequence $(U_n)_{n\in\mathbb{N}}$.*

(a) With probability 1, the sequence $\mathrm{unif}(T_n)$ converges weakly to a random probability measure X_∞ on $(\partial\mathbb{V}, \mathcal{B}(\partial\mathbb{V}))$ as $n \to \infty$.

(b) For each $\vartheta \in \mathbb{V}$,

$$X_\infty(A_\vartheta) = |I_\vartheta|. \tag{2.43}$$

(c) The random variables

$$\xi_\vartheta := \frac{X_\infty(A_{\vartheta 0})}{X_\infty(A_\vartheta)}, \quad \vartheta \in \mathbb{V}, \tag{2.44}$$

are independent with distribution $\mathcal{L}(\xi_\vartheta) = \mathrm{unif}(0,1)$ for all $\vartheta \in \mathbb{V}$.

Thus, the probability $X_\infty(A_\vartheta)$ of A_ϑ under X_∞ is the length of the interval I_ϑ associated with ϑ.
Finally, note that part (c) of the theorem implies the following alternative representation of $X_\infty(A_\vartheta)$ as a product of independent uniform random variables: For $\vartheta = (\vartheta_1, \ldots, \vartheta_k) \in \mathbb{V}$,

$$X_\infty(A_\vartheta) = \prod_{j=0}^{k-1} \tilde{\xi}_{(\vartheta_1,\ldots,\vartheta_j)}, \qquad \text{where} \tag{2.45}$$

$$\tilde{\xi}_{(\vartheta_1,\ldots,\vartheta_j)} := \begin{cases} \xi_{(\vartheta_1,\ldots,\vartheta_j)}, & \text{if } \vartheta_{j+1} = 0, \\ 1 - \xi_{(\vartheta_1,\ldots,\vartheta_j)}, & \text{if } \vartheta_{j+1} = 1. \end{cases}$$

Here, as usual, the empty sequence is identified with $\emptyset$.
In the next subsection, we will use a connection of the BST chain and the cyclic urn to show that the limits $\Xi_1, \ldots, \Xi_p$ can be written as functions of the tree limit X_∞. More precisely, we define functions of X_∞ that then turn out to coincide with the martingale limits almost surely. The distributional fixed point equations for $\Xi_1, \ldots, \Xi_p$ can be read off from this representation as well.

2.3.3 Embedding of the cyclic urn process

Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with $\mathcal{L}(U_1) = \mathrm{unif}(0,1)$. As in the preceding subsection, we construct the random BST chain $(T_n)_{n\in\mathbb{N}}$ from $(U_n)_{n\in\mathbb{N}}$. One can now observe that the dynamics of the random binary search tree algorithm give rise to a

cyclic urn process $(X_n)_{n\geq 0}$ within the tree. Note that as usual, we start the urn process with one ball of colour 0 and omit this choice from the notation in the following.
The cyclic urn is embedded into the evolution of the random binary search tree by labelling its external nodes by the types of the balls. Let us agree that the set of external nodes at time 0 is given by the single element $\{\emptyset\}$, which is turned into an internal node at time 1. This initial external node is labelled 0. Whenever an external node of type $j \in \{0, \dots, q-1\}$ is replaced by an internal node, its new left external node is labelled j and its new right external node is labelled $(j+1) \mod q$. This is the embedding of [47, Section 6.3], see also [12]: In the language of [47], the binary search tree on the nodes holding the keys $U_1, \dots, U_n$ in the rooted complete infinite binary tree is the 0-associated tree for a cyclic urn process at time n.
Now the external nodes of the tree sequence $(T_n)_{n\in\mathbb{N}}$ (together with the external node at time 0) form a cyclic urn process in the following manner: At time 0, both the tree and the urn consist of one external node respectively one ball of colour 0. In each following step, an external node is chosen uniformly at random in the tree, according to the binary search tree dynamics, and replaced by two external nodes, one of the same colour and one of the next higher colour. This corresponds to drawing a ball from the cyclic urn and returning it to the urn together with one ball of the following colour. Thus, if we denote by $X_n^{(1)}, \dots, X_n^{(q)}$ the numbers of external nodes having colours $0, \dots, q-1$ in the binary search tree on the nodes holding the keys $U_1, \dots, U_n$, then the vector-valued process $(X_n)_{n\geq 0}$ describes the evolution of a cyclic urn's composition vector.

We now adapt (2.17) as well as the derived decompositions to the present setting. The self-similarity of the binary tree transfers to the constructed process in the following way: Note that the complete binary tree decomposes into a left and a right subtree with root nodes 0 and 1, respectively. Let I_n denote the number of internal nodes in the left subtree of T_n. Consequently, I_n is uniformly distributed on $\{0, \dots, n-1\}$ and conditionally on $\{U_1 = u\}$, I_n has distribution $\mathrm{Bin}(n-1, u)$. The size of the right subtree is denoted by $J_n := n-1-I_n$. For the decomposition of the process, for each $j \geq 0$, we denote by $X_j^{[0],\{1\}}$ the vector of the numbers of various types of external nodes in the left subtree at the first time the left subtree reaches size j. Analogously, let $X_j^{[1],\{2\}}$ be the vector of the numbers of various types of external nodes in the right subtree at the first time the right subtree reaches size j. Again, upper indices $[0]$ and $[1]$ denote the type of the external root vertex. Apparently, for $n \geq 1$, the total number of external nodes of a certain colour is given by the sum of the numbers of external nodes of this colour in the two subtrees, so we have the identity

$$X_n = X_{I_n}^{[0],\{1\}} + X_{J_n}^{[1],\{1\}} = X_{I_n}^{[0],\{1\}} + R_{\mathrm{Cyc}} X_{J_n}^{[0],\{2\}}.$$

The sequences $\left(X_n^{[0],\{1\}}\right)_{n\geq 0}$ and $\left(X_n^{[1],\{2\}}\right)_{n\geq 0}$ denote the composition vectors of the cyclic urns given by the evolutions of the left and right subtrees of the root of the binary search tree. They are independent of I_n. We have set $\left(X_n^{[0],\{1\}}\right)_{n\geq 0} := \left(R_{\mathrm{Cyc}}^t X_n^{[1],\{1\}}\right)_{n\geq 0}$, and note that due to identity (2.16), $\left(X_n^{[0],\{1\}}\right)_{n\geq 0}$ is a cyclic urn process started with one ball of type 0 at time 0. This corresponds to shifting the labels in the right subtree by one to the left.
In a next step, for each $k \in \{1, \dots, p\}$, we define a sequence of random variables $\left(N_n^{(k)}\right)_{n\geq 0}$

that converges to Ξ_k. More precisely, we set $N_0^{(k)} := 0$ and for $n \geq 1$,

$$N_n^{(k)} := u_k(X_n - \mathbb{E}[X_n])/n^{\lambda_k},$$

such that

$$N_n^{(k)} = \left(\frac{I_n}{n}\right)^{\lambda_k} N_{I_n}^{(k),\{1\}} + \lambda_k \left(\frac{J_n}{n}\right)^{\lambda_k} N_{J_n}^{(k),\{2\}} + n^{-\lambda_k} G_{k,n}(I_n).$$

Here $N_\ell^{(k),\{1\}} := u_k \left(X_\ell^{[0],\{1\}} - \mathbb{E}\left[X_\ell^{[0],\{1\}}\right]\right)/\ell^{\lambda_k}$, $N_\ell^{(k),\{2\}} := u_k \left(X_\ell^{[0],\{2\}} - \mathbb{E}\left[X_\ell^{[0],\{2\}}\right]\right)/\ell^{\lambda_k}$ and $G_{k,n}$ is defined in (2.20). Note that $\left(N_n^{(k)}\right)_{n\geq 0}$ is almost surely convergent with limit Ξ_k, as

$$N_n^{(k)} = \frac{\Gamma(n+1+\lambda_k)}{\Gamma(n+1)n^{\lambda_k}} M_n^{(k)}.$$

The recurrence for $N_n^{(k)}$ implies that the martingale limits $\Xi_1, \dots, \Xi_p$ give rise to decompositions of the form

$$\Xi_k = U_1^{\lambda_k} \Xi_k^{\{1\}} + \lambda_k (1-U_1)^{\lambda_k} \Xi_k^{\{2\}} + g_k(U_1). \tag{2.46}$$

$U_1, \Xi_k^{\{1\}}, \Xi_k^{\{2\}}$ are independent and $\Xi_k^{\{1\}}$ and $\Xi_k^{\{2\}}$ have the same distribution as Ξ_k. This is equation (2.18), but note that the uniform random variable is U_1 here, the first random variable in the sequence $(U_n)_{n\in\mathbb{N}}$ that generates the BST chain. Further recall that

$$g_k : [0,1] \to \mathbb{C}, \qquad g_k(u) = \frac{1}{\Gamma(1+\lambda_k)} \left(u^{\lambda_k} + \lambda_k (1-u)^{\lambda_k} - 1\right).$$

We now use the definitions of the current section to rewrite the random variables $\Xi_1, \dots, \Xi_p$ via the sequence $(U_n)_{n\in\mathbb{N}}$. To this end, we define a candidate

$$\Psi_k := \sum_{s=0}^{\infty} \sum_{|\vartheta|=s} \lambda_k^{\sum_{i=1}^{s} \vartheta_i} X_\infty(A_\vartheta)^{\lambda_k} g_k(\xi_\vartheta) \tag{2.47}$$

for $1 \leq k \leq p$, which is a function of X_∞. Recall that $\Xi_1 = 0$ almost surely, so we do not have to worry about $k = 1$. We further define

$$\Psi_k^{\{1\}} := \sum_{s=0}^{\infty} \sum_{|\vartheta|=s} U_1^{-\lambda_k} \lambda_k^{\sum_{i=1}^{s} \vartheta_i} X_\infty(A_{0\vartheta})^{\lambda_k} g_k(\xi_{0\vartheta}), \tag{2.48}$$

$$\Psi_k^{\{2\}} := \sum_{s=0}^{\infty} \sum_{|\vartheta|=s} (1-U_1)^{-\lambda_k} \lambda_k^{\sum_{i=1}^{s} \vartheta_i} X_\infty(A_{1\vartheta})^{\lambda_k} g_k(\xi_{1\vartheta}). \tag{2.49}$$

Theorem 2.3.2. *(a) For all $2 \leq k \leq p$, the limits $\Psi_k, \Psi_k^{\{1\}}$ and $\Psi_k^{\{2\}}$ exist almost surely and in quadratic mean. Furthermore, $\Psi_k, \Psi_k^{\{1\}}$ and $\Psi_k^{\{2\}}$ are identically distributed.*

(b) For all $2 \leq k \leq p$, we have

$$\Psi_k = U_1^{\lambda_k}\Psi_k^{\{1\}} + \lambda_k(1-U_1)^{\lambda_k}\Psi_k^{\{2\}} + g_k(U_1) \tag{2.50}$$

almost surely and $U_1, \Psi_k^{\{1\}}, \Psi_k^{\{2\}}$ are independent.

(c) For all $2 \leq k \leq p$, as $n \to \infty$,

$$N_n^{(k)} \xrightarrow{a.s.} \Psi_k. \tag{2.51}$$

Hence, $\Xi_k = \Psi_k$ almost surely.

Proof. The proof of (a) is similar to the proof of Theorem 6 in [32]: First recall from Theorem 2.3.1 that $(\xi_\vartheta)_{\vartheta\in\mathbb{V}}$ is a family of i.i.d. random variables with $\mathcal{L}(\xi_\emptyset) = \mathrm{unif}(0,1)$. We also noted in (2.45) that for each $\vartheta \in \mathbb{V}$, the random variable $X_\infty(A_\vartheta)$ is a function of the variables ξ_κ with $\kappa < \vartheta$. Taken together, these facts imply that for all $\vartheta \in \mathbb{V}$, $X_\infty(A_\vartheta)$ and $g_k(\xi_\vartheta)$ are independent.

For $m \geq 0$, we now set $\mathcal{G}_m := \sigma(\{\xi_\vartheta : |\vartheta| \leq m\})$ as well as

$$\Psi_{k,m} := \sum_{s=0}^{m} \sum_{|\vartheta|=s} \lambda_k^{\sum_{i=1}^{s}\vartheta_i} X_\infty(A_\vartheta)^{\lambda_k} g_k(\xi_\vartheta).$$

We aim to show that $(\Psi_{k,m})_{m\geq 0}$ is an almost surely convergent martingale with limit Ψ_k. To this end, we compute the conditional expectation

$$\begin{aligned}
\mathbb{E}\left[\Psi_{k,m+1}|\mathcal{G}_m\right] &= \Psi_{k,m} + \sum_{|\vartheta|=m+1} \lambda_k^{\sum_{i=1}^{m+1}\vartheta_i} \mathbb{E}[X_\infty(A_\vartheta)^{\lambda_k} g_k(\xi_\vartheta)|\mathcal{G}_m] \\
&= \Psi_{k,m} + \sum_{|\vartheta|=m+1} \lambda_k^{\sum_{i=1}^{m+1}\vartheta_i} X_\infty(A_\vartheta)^{\lambda_k} \mathbb{E}[g_k(\xi_\vartheta)] \\
&= \Psi_{k,m},
\end{aligned}$$

where we have used the fact that $\mathbb{E}[g_k(U')] = 0$ for all $2 \leq k \leq p$ for $\mathcal{L}(U') = \mathrm{unif}(0,1)$. Similarly,

$$\begin{aligned}
\mathbb{E}\left[|\Psi_{k,m+1} - \Psi_{k,m}|^2 \Big| \mathcal{G}_m\right] &= \mathbb{E}\left[\left|\sum_{|\vartheta|=m+1} \lambda_k^{\sum_{i=1}^{m+1}\vartheta_i} X_\infty(A_\vartheta)^{\lambda_k} g_k(\xi_\vartheta)\right|^2 \Bigg| \mathcal{G}_m\right] \\
&= \sum_{|\vartheta|=m+1} |X_\infty(A_\vartheta)^{\lambda_k}|^2 \mathbb{E}\left[|g_k(\xi_\vartheta)|^2\right]
\end{aligned}$$

and thus with $C := \mathbb{E}\left[|g_k(\xi_\vartheta)|^2\right] < \infty$,

$$\mathbb{E}\left[|\Psi_{k,m+1} - \Psi_{k,m}|^2\right] = C\left(\frac{2}{1+2\sigma_k}\right)^{m+1}$$

for all $m \geq 0$. As $\sigma_k > \frac{1}{2}$, it follows that $(\Psi_{k,m}, \mathcal{G}_m)_{m\geq 0}$ is an L^2-bounded martingale and

almost surely convergent with limit Ψ_k, as we may see by considering real part and imaginary part separately, for example. In particular, Ψ_k is well-defined and has finite second moment. The claim for $\Psi_k^{\{1\}}$ and $\Psi_k^{\{2\}}$ can be shown analogously.
To show that Ψ_k and $\Psi_k^{\{1\}}$ are identically distributed, note that $\Psi_k^{\{1\}}$ may alternatively be written as

$$\Psi_k^{\{1\}} = \sum_{s=0}^{\infty} \sum_{|\vartheta|=s} \lambda_k^{\sum_{i=1}^{s} \vartheta_i} \prod_{j=0}^{s-1} (\tilde{\xi}_{(0,\vartheta_1,\cdots,\vartheta_j)})^{\lambda_k} g_k(\xi_{0\vartheta}).$$

Now, the first summand ($s = 0$ in the outer sum) is just $g_k(\xi_0)$. For $s \geq 1$, every ϑ with $|\vartheta| = s$, the corresponding inner summand is a product of three terms: the first is just $\lambda_k^{\sum_{i=1}^{s} \vartheta_i}$ as in the definition of Ψ_k. The second is a product of s independent unif$(0,1)$ random variables to the power of λ_k that are independent of the uniform $\xi_{0\vartheta}$, while the third is g_k applied to the independent $\xi_{0\vartheta}$. This representation makes it clear that $\Psi_k^{\{1\}}$ is defined for the left subtree in exactly the same way in which Ψ_k is defined for the whole tree, and they are identically distributed due to the self-similarity of the tree.
It can be shown similarly that Ψ_k and $\Psi_k^{\{2\}}$ are identically distributed.

Part (b): We decompose the series defining Ψ_k into the word $\vartheta = \emptyset$, those words that start with 0 and those that start with 1. This yields with $\xi_\emptyset = U_1$ that

$$\begin{aligned}
\Psi_k &= g_k(\xi_\emptyset) + \sum_{s=0}^{\infty} \sum_{|\vartheta|=s} \lambda_k^{0+\sum_{i=1}^{s} \vartheta_i} X_\infty(A_{0\vartheta})^{\lambda_k} g_k(\xi_{0\vartheta}) + \sum_{s=0}^{\infty} \sum_{|\vartheta|=s} \lambda_k^{1+\sum_{i=1}^{s} \vartheta_i} X_\infty(A_{1\vartheta})^{\lambda_k} g_k(\xi_{1\vartheta}) \\
&= g_k(U_1) + U_1^{\lambda_k} \sum_{s=0}^{\infty} U_1^{-\lambda_k} \sum_{|\vartheta|=s} \lambda_k^{\sum_{i=1}^{s} \vartheta_i} X_\infty(A_{0\vartheta})^{\lambda_k} g_k(\xi_{0\vartheta}) \\
&\quad + \lambda_k (1-U_1)^{\lambda_k} \sum_{s=0}^{\infty} (1-U_1)^{-\lambda_k} \sum_{|\vartheta|=s} \lambda_k^{\sum_{i=1}^{s} \vartheta_i} X_\infty(A_{1\vartheta})^{\lambda_k} g_k(\xi_{1\vartheta}) \\
&= U_1^{\lambda_k} \Psi_k^{\{1\}} + \lambda_k (1-U_1)^{\lambda_k} \Psi_k^{\{2\}} + g_k(U_1).
\end{aligned}$$

Furthermore, $U_1, \Psi_k^{\{1\}}, \Psi_k^{\{2\}}$ are independent, as they are functions of disjoint sets of the random variables $(\xi_\vartheta)_{\vartheta \in \mathbb{V}}$.

Part (c): Note that $\left(N_n^{(k)}\right)_{n \geq 0}$ converges to Ξ_k also in L^2. Using the recursion for $N_n^{(k)}$ and part (b) of the current theorem, we will show that the L^2-distance of $N_n^{(k)}$ and Ψ_k tends to zero as well, that is

$$a_n := \mathbb{E}\left[\left|N_n^{(k)} - \Psi_k\right|^2\right] \xrightarrow{n \to \infty} 0.$$

Here, $a_0 = \mathbb{E}\left[|\Psi_k|^2\right]$. We decompose

$$\begin{aligned} N_n^{(k)} - \Psi_k &= \left(\left(\frac{I_n}{n}\right)^{\lambda_k} N_{I_n}^{(k),\{1\}} - U_1^{\lambda_k}\Psi_k^{\{1\}}\right) \\ &\quad + \lambda_k \left(\left(\frac{J_n}{n}\right)^{\lambda_k} N_{J_n}^{(k),\{2\}} - (1-U_1)^{\lambda_k}\Psi_k^{\{2\}}\right) + \left(n^{-\lambda_k} G_{n,k}(I_n) - g_k(U_1)\right) \\ &=: W_1 + \lambda_k \cdot W_2 + W_3. \end{aligned}$$

Conditionally on U_1 and I_n, W_1 and W_2 are independent with expectation 0, and W_3 is constant. Because the random variables are centered, it follows that for $n \geq 1$,

$$\mathbb{E}\left[\left|N_n^{(k)} - \Psi_k\right|^2 \middle| I_n, U_1\right] = \mathbb{E}\left[|W_1|^2 \middle| I_n, U_1\right] + \mathbb{E}\left[|W_2|^2 \middle| I_n, U_1\right] + |W_3|^2$$

almost surely. Taking the expectation on both sides gives

$$a_n = \mathbb{E}\left[|W_1|^2\right] + \mathbb{E}\left[|W_2|^2\right] + \mathbb{E}\left[|W_3|^2\right],$$

$n \geq 1$. We now analyse the three terms separately.
As for $\mathbb{E}\left[|W_1|^2\right]$: We further subdivide this term into

$$\begin{aligned} \mathbb{E}\left[|W_1|^2\right] &= \mathbb{E}\left[\left|\left(\frac{I_n}{n}\right)^{\lambda_k}\left(N_{I_n}^{(k),\{1\}} - \Psi_k^{\{1\}}\right)\right|^2\right] + \mathbb{E}\left[\left|\Psi_k^{\{1\}}\right|^2 \left|\left(\frac{I_n}{n}\right)^{\lambda_k} - U_1^{\lambda_k}\right|^2\right] \\ &\quad + 2\Re\left(\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{\lambda_k} \overline{\Psi_k^{\{1\}}\left(N_{I_n}^{(k),\{1\}} - \Psi_k^{\{1\}}\right)\left(\left(\frac{I_n}{n}\right)^{\lambda_k} - U_1^{\lambda_k}\right)}\right]\right) \\ &=: W_{1,1} + W_{1,2} + W_{1,3} \leq W_{1,1} + W_{1,2} + 2\sqrt{W_{1,1}W_{1,2}}. \end{aligned}$$

Lemma 2.3.1. *For* $n \geq 1$*, we have*

$$W_{1,1} = \frac{1}{n}\sum_{j=0}^{n-1}\left(\frac{j}{n}\right)^{2\sigma_k} a_j. \tag{2.52}$$

Proof of Lemma 2.3.1. Since $\Psi_k^{\{1\}}$ is independent of I_n, conditioning on I_n gives

$$W_{1,1} = \mathbb{E}\left[\sum_{j=0}^{n-1} \mathbb{1}_{\{I_n=j\}}\left(\frac{j}{n}\right)^{2\sigma_k}\mathbb{E}\left[\left|N_j^{(k),\{1\}} - \Psi_k^{\{1\}}\right|^2\right]\right] = \frac{1}{n}\sum_{j=0}^{n-1}\left(\frac{j}{n}\right)^{\sigma_k} a_j.$$

Note that for all $j \geq 0$, $\mathbb{E}\left[\left|N_j^{(k),\{1\}} - \Psi_k^{\{1\}}\right|^2\right] = \mathbb{E}\left[\left|N_j^{(k)} - \Psi_k\right|^2\right]$. This is due to the fact that $\left(\left(N_n^{(k),\{1\}}\right)_{n\geq 0}, \Psi_k^{\{1\}}\right)$ and $\left(\left(N_n^{(k)}\right)_{n\geq 0}, \Psi_k\right)$ are identically distributed. □

Finally, it is an immediate consequence of Lemma 2.2.5 that $W_{1,2} = O\left(n^{-1}\right)$ as $n \to \infty$.

As for $\mathbb{E}\left[|W_2|^2\right]$: As we have done for $\mathbb{E}\left[|W_1|^2\right]$, we further subdivide in

$$\begin{aligned}\mathbb{E}\left[|W_2|^2\right] &= \mathbb{E}\left[\left|\left(\frac{J_n}{n}\right)^{\lambda_k}\left(N_{J_n}^{(k),\{2\}} - \Psi_k^{\{2\}}\right)\right|^2\right] + \mathbb{E}\left[\left|\Psi_k^{\{2\}}\right|^2\left|\left(\frac{J_n}{n}\right)^{\lambda_k} - (1-U_1)^{\lambda_k}\right|^2\right] \\ &\quad + 2\Re\left(\mathbb{E}\left[\left(\frac{J_n}{n}\right)^{\lambda_k}\overline{\Psi_k^{\{2\}}}\left(N_{J_n}^{(k),\{2\}} - \Psi_k^{\{2\}}\right)\overline{\left(\left(\frac{J_n}{n}\right)^{\lambda_k} - (1-U_1)^{\lambda_k}\right)}\right]\right) \\ &=: W_{2,1} + W_{2,2} + W_{2,3} \le W_{2,1} + W_{2,2} + 2\sqrt{W_{2,1}W_{2,2}}.\end{aligned}$$

Similar to the previous calculations, it can be shown that

$$W_{2,1} = \frac{1}{n}\sum_{j=0}^{n-1}\left(\frac{j}{n}\right)^{2\sigma_k} a_j. \tag{2.53}$$

for $n \ge 1$ and that $W_{2,2} = O\left(n^{-1}\right)$ as $n \to \infty$.

As for $\mathbb{E}\left[|W_3|^2\right]$:

Lemma 2.3.2. *As* $n \to \infty$,

$$\mathbb{E}\left[\left|n^{-\lambda_k} G_{n,k}(I_n) - g_k(U_1)\right|^2\right] = O\left(n^{-1}\right). \tag{2.54}$$

Proof of Lemma 2.3.2. A combination of Lemma 2.2.5 and Stirling's formula implies that

$$\begin{aligned}&\mathbb{E}\left[|W_3|^2\right] \\ &= \mathbb{E}\left[\left|\frac{\Gamma(I_n+1+\lambda_k)}{\Gamma(I_n+1)n^{\lambda_k}} - U_1^{\lambda_k} + \lambda_k\left(\frac{\Gamma(J_n+1+\lambda_k)}{\Gamma(J_n+1)n^{\lambda_k}} - (1-U_1)^{\lambda_k}\right) + \frac{\Gamma(n+1+\lambda_k)}{\Gamma(n+1)n^{\lambda_k}} - 1\right|^2\right] \\ &\qquad \cdot \left|\frac{1}{\Gamma(1+\lambda_k)}\right|^2 \\ &\le \left|\frac{2}{\Gamma(1+\lambda_k)}\right|^2\left(\mathbb{E}\left[\left|\frac{\Gamma(I_n+1+\lambda_k)}{\Gamma(I_n+1)n^{\lambda_k}} - U_1^{\lambda_k}\right|^2\right] + \mathbb{E}\left[\left|\frac{\Gamma(J_n+1+\lambda_k)}{\Gamma(J_n+1)n^{\lambda_k}} - (1-U_1)^{\lambda_k}\right|^2\right]\right. \\ &\quad \left. + \left|\frac{\Gamma(n+1+\lambda_k)}{\Gamma(n+1)n^{\lambda_k}} - 1\right|^2\right) = O\left(n^{-1}\right) \qquad \text{as} \quad n \to \infty.\end{aligned}$$

□

Conclusion. The preceding calculations have shown that there exists some constant $C > 0$ such that for all $n \ge 1$,

$$a_n \le \frac{2}{n}\sum_{j=0}^{n-1}\left(\frac{j}{n}\right)^{2\sigma_k} a_j + Cn^{-1} + \frac{C}{\sqrt{n}}\sqrt{\frac{2}{n}\sum_{j=0}^{n-1}\left(\frac{j}{n}\right)^{2\sigma_k} a_j}.$$

Claim: As $n \to \infty$, for all $\alpha \in (0, 2\sigma_k - 1)$,

$$a_n = O\left(n^{-\alpha}\right).$$

We prove the claim via induction on $n \geq n_0$.

$$\begin{aligned}
a_{n+1} &\leq \frac{2}{n+1}\sum_{j=0}^{n}\left(\frac{j}{n+1}\right)^{2\sigma_k} a_j + C(n+1)^{-1} + \frac{C}{\sqrt{n+1}}\sqrt{\frac{2}{n+1}\sum_{j=0}^{n}\left(\frac{j}{n+1}\right)^{2\sigma_k} a_j} \\
&\overset{\text{I.H.}}{\leq} 2D(n+1)^{-\alpha}\frac{1}{n+1}\sum_{j=0}^{n}\left(\frac{j}{n+1}\right)^{2\sigma_k-\alpha} + C(n+1)^{-1} \\
&+ \frac{C}{\sqrt{n+1}}\sqrt{2D(n+1)^{-\alpha}\frac{1}{n+1}\sum_{j=0}^{n}\left(\frac{j}{n+1}\right)^{2\sigma_k-\alpha}} \\
&\leq \frac{2D}{2\sigma_k+1-\alpha}(n+1)^{-\alpha} + C(n+1)^{-1} + \frac{C}{\sqrt{n+1}}\sqrt{\frac{2D}{2\sigma_k+1-\alpha}(n+1)^{-\alpha}} \\
&\leq D(n+1)^{-\alpha}
\end{aligned}$$

for some constant $D > 0$ and n sufficiently large. This implies

$$a_n \longrightarrow 0, \qquad n \to \infty,$$

and thus (c). □

3 General Case

This chapter serves the only purpose to prove Theorem 1.2.5 and to make the proof more transparent. It closely follows the preprint [58].

3.1 Projections and martingales

The key to the proof of Theorem 1.2.5 is an understanding of the order of magnitude and the asymptotic behaviour of certain components of X_n, namely the projection coefficients $\pi_k(Y_n)$. Most of the following facts are reformulations of known results and listed here to keep the text as self-contained as possible.

We first note that for each $n \in \mathbb{N}_0$, the conditional expectation of the next state, given the history of the urn process, takes the particular form

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \left(\mathrm{Id}_{\mathbb{C}^q} + \frac{R}{rn + |X_0|}\right) X_n. \tag{3.1}$$

This yields a vector-valued martingale

$$\left(\prod_{j=N}^{n-1} \left(\mathrm{Id}_{\mathbb{C}^q} + \frac{R}{rj + |X_0|}\right)^{-1} X_n\right)_{n \geq N}$$

for some $N \in \mathbb{N}_0$ sufficiently large such that the occurring matrix inverses exist. This observation can be found below Definition 2.1 in [67].
The particular form of $\mathbb{E}[X_{n+1}|\mathcal{F}_n]$ leads to complex-valued martingales via projections on the eigenspaces of R. This idea is implicit in the work of Smythe [74] and more explicit in the proof of Theorem 3.5 in [67] for certain projections. We adopt it for all eigenspace projections.

Lemma 3.1.1 (Projection martingales). *(i) Let $k \in \{1, \ldots, q\}$ be such that λ_k satisfies $\lambda_k + |X_0| \notin r\mathbb{Z}_- := \{0, -r, -2r, \ldots\}$. Define*

$$\gamma_n^{(k)} := \prod_{j=0}^{n-1} \left(1 + \frac{\lambda_k}{rj + |X_0|}\right),$$

which is different from zero for all $n \geq 0$. Then

$$\mathbb{E}[\pi_k(X_n)] = \gamma_n^{(k)} \pi_k(X_0) = \frac{\Gamma\left(\frac{|X_0|}{r}\right) \pi_k(X_0)}{\Gamma\left(\frac{|X_0|+\lambda_k}{r}\right)} \cdot n^{\frac{\lambda_k}{r}} + O\left(n^{\Re\left(\frac{\lambda_k}{r}\right)-1}\right),$$

as $n \to \infty$. $\left(M_n^{(k)}\right)_{n\geq 0}$, *defined by*

$$M_n^{(k)} := \left(\gamma_n^{(k)}\right)^{-1} \cdot \pi_k(Y_n),$$

is a complex-valued martingale with mean zero.

(ii) Let $k \in \{1, \dots, q\}$ *be such that* λ_k *satisfies* $\lambda_k + |X_0| \in r\mathbb{Z}_- = \{0, -r, -2r, \dots\}$. *Define*

$$\gamma_n^{(k)} := \prod_{j=-\frac{\lambda_k+|X_0|}{r}+1}^{n-1} \left(1 + \frac{\lambda_k}{rj + |X_0|}\right),$$

which is different from zero for all $n \geq -\frac{\lambda_k+|X_0|}{r} + 1$. *Then*

$$\mathbb{E}[\pi_k(X_n)] = 0,$$

for all $n \geq -\frac{\lambda_k+|X_0|}{r} + 1$. $\left(M_n^{(k)}\right)_{n\geq -\frac{\lambda_k+|X_0|}{r}+1}$, *defined by*

$$M_n^{(k)} := \left(\gamma_n^{(k)}\right)^{-1} \cdot \pi_k(Y_n),$$

is a complex-valued martingale with mean zero.

Proof. Let $k \in \{1, \dots, q\}$ and $n \geq 0$. As a direct consequence of (3.1) for all $n \geq 0$,

$$\mathbb{E}[\pi_k(X_{n+1})|\mathcal{F}_n] = \left(1 + \frac{\lambda_k}{rn + |X_0|}\right) \pi_k(X_n)$$

almost surely and $\left(M_n^{(k)}\right)_n$ is a martingale in each of the two cases. In particular,

$$\mathbb{E}[\pi_k(X_n)] = \prod_{j=0}^{n-1} \left(1 + \frac{\lambda_k}{rj + |X_0|}\right) \pi_k(X_0),$$

which is zero for $n \geq -\frac{\lambda_k+|X_0|}{r} + 1$ in the second case. In the first case, by Stirling's formula,

$$\gamma_n^{(k)} = \frac{\Gamma\left(\frac{|X_0|}{r}\right)}{\Gamma\left(\frac{|X_0|+\lambda_k}{r}\right)} \cdot \frac{\Gamma\left(n + \frac{|X_0|}{r} + \frac{\lambda_k}{r}\right)}{\Gamma\left(n + \frac{|X_0|}{r}\right)} = \frac{\Gamma\left(\frac{|X_0|}{r}\right)}{\Gamma\left(\frac{|X_0|+\lambda_k}{r}\right)} \cdot n^{\frac{\lambda_k}{r}} + O\left(n^{\Re\left(\frac{\lambda_k}{r}\right)-1}\right)$$

as $n \to \infty$. This implies the claim. □

The "effect" $\gamma_n^{(k)}$ on the various projections resulting from the transition (3.1) is also studied in [43], see in particular Lemmata 5.4 and 5.5.

The martingales of the preceding proposition can be divided into two classes: convergent and non-convergent martingales. The corresponding eigenvalues are sometimes referred to as "big" and "small", respectively. The remainder of this section will be devoted to properties of the convergent martingales and their limits. In the context of m-ary search trees, Lemma 3 in [13] is precisely the following result.

Lemma 3.1.2 (Martingale limits)**.** *For each* $k \in \{1, \dots, q\}$ *such that* $\Re(\lambda_k) > r/2$*, there exists a complex-valued mean zero random variable* Ξ_k *such that*

$$M_n^{(k)} \to \frac{\Gamma(|X_0|/r + \lambda_k/r)}{\Gamma(|X_0|/r)} \quad \Xi_k$$

almost surely and in L^2 *as* $n \to \infty$*.*

Remark 9. The random variables Ξ_k in Theorem 1.2.5 and Lemma 3.1.2 are identical.

Proof. We apply the L^2-martingale convergence theorem and show boundedness of second moments.

$$\mathbb{E}\left[|\pi_k(X_{n+1})|^2 \middle| \mathcal{F}_n\right] = \left(1 + \frac{2\Re(\lambda_k)}{rn + |X_0|}\right) |\pi_k(X_n)|^2 + \sum_{j=1}^{q} \frac{X_n^{(j)}}{rn + |X_0|} |\pi_k(\Delta_j)|^2.$$

Set $C_k := \sum_{j=1}^{q} |\pi_k(\Delta_j)|^2 = \sum_{j=1}^{q} |\lambda_k|^2 \left|u_k^{(j)}\right|^2$, where $\Delta_1, \dots, \Delta_q$ are the possible increments that are defined in the introduction of the thesis. With this,

$$\mathbb{E}\left[|\pi_k(X_{n+1})|^2 \middle| \mathcal{F}_n\right] \le \left(1 + \frac{2\Re(\lambda_k)}{rn + |X_0|}\right) |\pi_k(X_n)|^2 + C_k$$

and thus

$$\begin{aligned} &\mathbb{E}\left[|\pi_k(X_n)|^2\right] \\ &\le \prod_{j=0}^{n-1} \left(1 + \frac{2\Re(\lambda_k)}{rj + |X_0|}\right) \mathbb{E}[|\pi_k(X_0)|^2] + C_k \prod_{j=1}^{n-1} \left(1 + \frac{2\Re(\lambda_k)}{rj + |X_0|}\right) \sum_{m=0}^{n-1} \prod_{j=1}^{m} \left(1 + \frac{2\Re(\lambda_k)}{rj + |X_0|}\right)^{-1} \\ &= \prod_{j=0}^{n-1} \left(1 + \frac{2\Re(\lambda_k)}{rj + |X_0|}\right) \left(\mathbb{E}[|\pi_k(X_0)|^2] + C_k \left(1 + \frac{2\Re(\lambda_k)}{|X_0|}\right)^{-1} \sum_{m=0}^{n-1} \prod_{j=1}^{m} \left(1 + \frac{2\Re(\lambda_k)}{rj + |X_0|}\right)^{-1}\right) \\ &= O\left(n^{2\Re(\lambda_k)/r}\right) \end{aligned}$$

as $n \to \infty$, because $\Re(\lambda_k) > r/2$. By Jensen's inequality, $|\mathbb{E}[\pi_k(X_n)]|^2 = O\left(n^{2\Re(\lambda_k)/r}\right)$ as $n \to \infty$. Thus

$$\mathbb{E}\left[\left|M_n^{(k)}\right|^2\right] = \left|\gamma_n^{(k)}\right|^{-2} \left(\mathbb{E}\left[|\pi_k(X_n)|^2\right] - |\mathbb{E}\left[\pi_k(X_n)\right]|^2\right) = O(1)$$

as $n \to \infty$. By the L^2-martingale convergence theorem, $M_n^{(k)}$ converges almost surely and in L^2 to a complex-valued random variable. □

Remark 10. Janson [43] offers an explanation of the difference between small and large urns that is related to the calculation in the proof of Lemma 3.1.2: It is immediate from this proof that for all eigenvalues λ_k,

$$\mathbb{E}\left[|\pi_k(X_n)|^2\right] = \prod_{j=0}^{n-1}\left(1+\frac{2\Re(\lambda_k)}{rj+|X_0|}\right)|\pi_k(X_0)|^2 + \sum_{m=0}^{n-1}\sum_{j=1}^{q}\mathbb{E}\left[\frac{X^{(j)}_{n-1-m}}{r(n-1-m)+|X_0|}\right]|\pi_k(\Delta_j)|^2\prod_{\ell=n-m}^{n-1}\left(1+\frac{2\Re(\lambda_k)}{r\ell+|X_0|}\right).$$

If we interpret the summand corresponding to index $n-m-1$ in the sum on the right hand side as the influence of the m^{th} draw on the state at time n, this term is (very roughly) of the order $(n/m)^{2\Re(\lambda_k)/r}$.

Depending on convergence or non-convergence of the series $\sum_{m=1}^{\infty} m^{-2\Re(\lambda_k)/r}$, we apply a different scaling. For large eigenvalues λ_k, $\sum_{m=1}^{\infty} m^{-2\Re(\lambda_k)/r}$ is convergent and we consider

$$\begin{aligned}
&n^{-2\Re(\lambda_k)/r}\left(\mathbb{E}\left[|\pi_k(X_n)|^2\right] - |\mathbb{E}\left[\pi_k(X_n)\right]|^2\right)\\
= \quad & n^{-2\Re(\lambda_k)/r}\sum_{m=0}^{n-1}\sum_{j=1}^{q}\mathbb{E}\left[\frac{X^{(j)}_{n-1-m}}{r(n-1-m)+|X_0|}\right]|\pi_k(\Delta_j)|^2\prod_{\ell=n-m}^{n-1}\left(1+\frac{2\Re(\lambda_k)}{r\ell+|X_0|}\right)+O\left(n^{-1}\right).
\end{aligned}$$

In the above expression, $n^{2\Re(\lambda_k)/r}$ cancels. For a large eigenvalue λ_k, the terms in the sum $\sum_{m=1}^{n-1} m^{-2\Re(\lambda_k)/r}$ decrease fast. In this case, the sum above is dominated by the overproportional influence of the first few draws. Almost sure limits which depend on the initial state X_0 arise as a consequence of this strong long-term dependency.

If λ_k is strictly small, $\sum_{m=1}^{\infty} m^{-2\Re(\lambda_k)/r}$ is divergent and we consider

$$\begin{aligned}
&n^{-1}\left(\mathbb{E}\left[|\pi_k(X_n)|^2\right] - |\mathbb{E}[\pi_k(X_n)]|^2\right)\\
= \quad & n^{-1}\sum_{m=0}^{n-1}\sum_{j=1}^{q}\mathbb{E}\left[\frac{X^{(j)}_{n-1-m}}{r(n-1-m)+|X_0|}\right]|\pi_k(\Delta_j)|^2\prod_{\ell=n-m}^{n-1}\left(1+\frac{2\Re(\lambda_k)}{r\ell+|X_0|}\right)+O\left(n^{-1}\right)
\end{aligned}$$

instead. This scaling leads to convergence of the right hand side, again. However, in contrast to the rescaled sum for large eigenvalues, each single summand tends to zero as $n\to\infty$. More precisely, for $\varepsilon>0$ sufficiently small and n large, the total value of the right hand side does not change by more than $\delta>0$, if we only consider contributions to the sum of draws after time εn. This explains why the influences of the first draws and the initial state are negligible in the long run for small projections.

Finally, if $\Re(\lambda_k)=1/2$, $\sum_{m=1}^{\infty} m^{-2\Re(\lambda_k)/r}$ is still divergent. In this case, we consider

$$\begin{aligned}
&(n\log(n))^{-1}\left(\mathbb{E}\left[|\pi_k(X_n)|^2\right] - |\mathbb{E}[\pi_k(X_n)]|^2\right)\\
= \quad & (n\log(n))^{-1}\sum_{m=0}^{n-1}\sum_{j=1}^{q}\mathbb{E}\left[\frac{X^{(j)}_{n-1-m}}{r(n-1-m)+|X_0|}\right]|\pi_k(\Delta_j)|^2\prod_{\ell=n-m}^{n-1}\left(1+\frac{2\Re(\lambda_k)}{r\ell+|X_0|}\right)+O\left(n^{-1}\right).
\end{aligned}$$

Again, the right hand side is convergent under this scaling, and each single summand is negligible. However, the long-term dependency is more pronounced. For example, we cannot make the contribution from the first εn summands arbitrarily small by choosing $\varepsilon>0$ sufficiently small. In this case, we can only neglect a sublinear number n^{ε} of draws for some small $\varepsilon \in (0, 1)$ to obtain all but a small fraction of the sum at time n.

Remark 11. Recall the asymptotic proportions $D^{(1)}, \dots, D^{(a+1)}\Gamma_{a+1}, \dots, D^{(a+1)}\Gamma_{a+c}$ of the $a + c$ dominant classes in Theorem 1.2.3 and below. For $k \in \{1, \dots, a\}$, we have the relation

$$\frac{\Xi_k}{r} = D^{(k)} - \frac{\pi_k(X_0)}{|X_0|}. \tag{3.2}$$

Analogously, for $k \in \{a + 1, \dots, a + c\}$, there is the relation

$$\frac{\Xi_k}{r} = D^{(a+1)}\Gamma_k - \frac{\pi_k(X_0)}{|X_0|}.$$

The proportion vector V of Theorem 1.2.3 and below may thus alternatively be written as a sum over the martingale limits of Lemma 3.1.2:

$$V = \sum_{k=1}^{a+c} \left(\frac{\Xi_k}{r} + \frac{\pi_k(X_0)}{|X_0|} \right) \cdot \nu_k. \tag{3.3}$$

Further note that all (real) random variables $D^{(1)}, \dots, D^{(a+1)}\Gamma_{a+1}, \dots, D^{(a+1)}\Gamma_{a+c}$ are strictly positive almost surely under (A1)-(A5): This is clear for $D^{(1)}, \dots, D^{(a+1)}$ as these are the components of a Dirichlet vector. To see this for $D^{(a+1)}\Gamma_{a+1}, \dots, D^{(a+1)}\Gamma_{a+c}$, we compare the original urn process with a slightly modified process that only has (isolated) dominant classes. More precisely, each time a ball of colour i in a category 3 class is drawn, instead of following the original rules, give all its children *outside the class of* i colour i, too. In the modified urn, the classes remain unchanged, but all category 3 classes become dominant. For example,

$$R = \begin{pmatrix} 5 &&&&&&& \\ & 5 &&&&&& \\ && 1 & 1 &&&& \\ && 2 & -2 &&&& \\ &&& 1 & 4 &&& \\ && 1 & 2 && 2 & 1 & \\ &&& 2 & 1 & 3 & 4 & \\ && 1 & 1 &&&& 5 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 5 &&&&&&& \\ & 5 &&&&&& \\ && 3 & 1 &&&& \\ && 2 & 4 &&&& \\ &&&& 5 &&& \\ &&&&& 2 & 1 & \\ &&&&& 3 & 4 & \\ &&&&&&& 5 \end{pmatrix}.$$

Recall that the random variable $D^{(a+1)}\Gamma_{a+j}$ gives the asymptotic proportion of balls in class $\mathcal{C}_{a+j}$ in the original urn. One can now couple the two processes by drawing balls according to the original process, looking at the counterpart of the drawn ball in the derived process, and then follow the appropriate rules. Now, for the derived process, by standard results on the Pólya urn, the proportion of balls in class $\mathcal{C}_{a+j}$ tends to a positive limit almost surely. As there are at least as many balls of this class in the original urn as in the derived urn, the class also has a positive proportion in the original urn.

Corollary 3.1.1 (Random limits). *Under conditions (A1)-(A5), $\Xi_1, \dots, \Xi_{a+c}$ are almost surely non-degenerate unless r is simple. In this case, $\Xi_1 = 0$ almost surely. Furthermore, $\Xi_{a+c+1}, \dots, \Xi_p$ are almost surely non-degenerate.*

Proof. Note that for the limits $\Xi_1, \ldots, \Xi_a$ that correspond to the proportions of the isolated category 1 classes, the claim immediately follows from Theorem 1.2.3 and identity (3.2): For $k \in \{1, \ldots, a\}$,

$$\mathrm{Var}(\Xi_k) = r^2 \mathrm{Var}\left(D^{(k)}\right),$$

where Dirichlet vector D has parameters $\frac{|(X_0)_{\mathcal{C}_1}|}{r}, \ldots, \frac{|(X_0)_{\mathcal{C}_a}|}{r}, \frac{|X_0|}{r} - \sum_{j=1}^{a} \frac{|(X_0)_{\mathcal{C}_k}|}{r}$.

More generally and without reference to Theorem 1.2.3, we can use orthogonality of martingale increments to see that for $k \in \{1, \ldots, p\}$,

$$\begin{aligned}\mathbb{E}\left[|\Xi_k|^2\right] &= \mathbb{E}\left[\left|\Xi_k - M_0^{(k)}\right|^2\right] = \sum_{j=0}^{\infty} \mathbb{E}\left[\left|M_{j+1}^{(k)} - M_j^{(k)}\right|^2\right] \\ &= \sum_{j=0}^{\infty} \left|\gamma_{j+1}^{(k)}\right|^{-2} \mathbb{E}\left[\left|\pi_k(X_{j+1} - X_j) - \frac{\lambda_k}{rj + |X_0|}\pi_k(X_j)\right|^2\right].\end{aligned}$$

The requirement of zero variance completely determines the evolution of projection k in each draw: The expression on the right hand side is only equal to zero if for all $j \geq 0$,

$$\pi_k(X_{j+1} - X_j) = \frac{\lambda_k}{rj + |X_0|}\pi_k(X_j) \tag{3.4}$$

almost surely (note that $\lambda_k \neq 0$). This in particular means that the value of $\pi_k(X_{j+1} - X_j)$ is independent of the colour of the $(j+1)^{\text{th}}$ ball drawn from the urn. We will see that this is not possible under our assumptions.

First assume that there is an initial configuration X_0 that is compatible with (A1)-(A5) and has $\pi_k(X_0) = 0$. Under this initial configuration, $\pi_k(X_j) = 0$ for all $j \geq 0$ almost surely because of (3.4), and with probability one,

$$0 = \pi_k(X_{j+1} - X_j) = \lambda_k u_k^{(N_{j+1})}$$

for all $j \geq 0$. Here, N_{j+1} denotes the colour obtained in the $(j+1)^{\text{th}}$ draw. For each colour $f \in \{1, \ldots, q\}$, there is $n \in \mathbb{N}_0$ with $\mathbb{P}\left(X_n^{(f)} > 0\right) > 0$, due to assumption (A5). The last equation yields that $u_k^{(f)} = 0$. So $u_k = 0$, which is a contradiction.

The last paragraph shows that there is no admissible choice of initial configuration such that $\pi_k(X_0) = 0$. For $\pi_k(X_0) \neq 0$, $\pi_k(X_{j+1} - X_j) \neq 0$ for all j almost surely because of (3.4). If there is more than one dominant colour or if λ_k belongs to a category 3 class, it follows immediately from our choice of left eigenvectors that $\pi_k(X_{j+1} - X_j) \neq 0$ for all j almost surely is not possible: According to the mechanism of the urn, once there is a ball of a dominant class in the urn, there are balls of its class in the urn at all future times. So almost surely, there is a time N at which there are balls of all dominant classes in the urn. From this point on, there is a positive probability of drawing balls that lead to no change in the class under consideration. Now assume that there is only one dominant class and that λ_k belongs to this class. As the proportions of balls in this class converge to a positive limit with probability one, there is a time N from which on there are balls of each colour of the dominant class in the urn. This implies that $u_k^{(i)} = u_k^{(j)}$ for all colours i, j in this class. So $\lambda_k = r$, and $u_k = (1, \ldots, 1)^t$ is the only projection that induces a deterministic limit. □

Remark 12. In Theorem 3.6 in [43] the following related result is shown for strictly small, irreducible, balanced and tenable urns: If $u \in \mathbb{R}^q$, then $u^t X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_u^2)$ and $\sigma_u^2 = 0$ if and only if for every $n \geq 0$, $\mathrm{Var}(u^t X_n) = 0$, i.e. if $u^t X_n$ is deterministic.

More precise results on distributional properties of the random variables $\Xi_1, \dots, \Xi_p$ can be found in [51, 57], for example. We conclude the current section with the following lemma.

Lemma 3.1.3 (Speed of convergence)**.** *Let* $k \in \{1, \dots, p\}$. *Then*

$$\left\| \frac{\Gamma(|X_0|/r + \lambda_k/r)}{\Gamma(|X_0|/r)} \Xi_k - M_n^{(k)} \right\|_2 = O\left(n^{1/2 - \Re(\lambda_k)/r}\right), \qquad n \to \infty. \tag{3.5}$$

Proof. We use the decomposition

$$\begin{aligned}
\left\| \frac{\Gamma(|X_0|/r + \lambda_k/r)}{\Gamma(|X_0|/r)} \Xi_k - M_n^{(k)} \right\|_2^2 &= \sum_{j=n}^{\infty} \mathbb{E}\left[\left| M_{j+1}^{(k)} - M_j^{(k)} \right|^2 \right] \\
&= \sum_{j=n}^{\infty} \left| \gamma_{j+1}^{(k)} \right|^{-2} \mathbb{E}\left[\left| \pi_k(X_{j+1} - X_j) - \frac{\lambda_k}{rj + |X_0|} \pi_k(X_j) \right|^2 \right] \\
&= \sum_{j=n}^{\infty} \left| \gamma_{j+1}^{(k)} \right|^{-2} \left(\mathbb{E}\left[|\pi_k(X_{j+1} - X_j)|^2 \right] - \left| \frac{\lambda_k}{rj + |X_0|} \right|^2 \mathbb{E}\left[|\pi_k(X_j)|^2 \right] \right) \\
&\leq \sum_{j=n}^{\infty} \left| \gamma_{j+1}^{(k)} \right|^{-2} \mathbb{E}\left[|\pi_k(X_{j+1} - X_j)|^2 \right] \\
&\leq C n^{1 - 2\Re(\lambda_k)/r}
\end{aligned}$$

as $|\pi_k(X_{j+1} - X_j)|^2$ can only take q values, independently of j. □

We finally prove Proposition 1.1.4.

Proof of Proposition 1.1.4. Recall that we consider a generating matrix of the form

$$R = \begin{pmatrix} a_{1,1} & 0 \\ a_{2,1} & r \end{pmatrix}$$

with $a_{1,1} > 0$ and $a_{1,1} + a_{2,1} = r$. If we set

$$M_n := \prod_{j=0}^{n-1} \left(1 + \frac{a_{1,1}}{rj + |X_0|} \right)^{-1} \left(X_n^{(1)} - \mathbb{E}\left[X_n^{(1)} \right] \right),$$

$(M_n)_{n\geq 0}$ is a martingale. A similar calculation as in Lemma 3.1.2 leads to

$$\mathbb{E}\left[\left(X_n^{(1)}\right)^2\right] = \frac{\Gamma\left(\frac{|X_0|}{r}\right)\Gamma\left(n+\frac{2a_{1,1}+|X_0|}{r}\right)}{\Gamma\left(n+\frac{|X_0|}{r}\right)}X_0^{(1)}$$
$$\cdot\left(\frac{X_0^{(1)}}{\Gamma\left(\frac{|X_0|+2a_{1,1}}{r}\right)}+\frac{a_{1,1}^2}{\Gamma\left(\frac{|X_0|+a_{1,1}}{r}\right)}\sum_{m=0}^{n-1}\frac{\Gamma\left(m+\frac{|X_0|+a_{1,1}}{r}\right)}{(rm+2a_{1,1}+|X_0|)\,\Gamma\left(m+\frac{|X_0|+2a_{1,1}}{r}\right)}\right).$$

This implies that $(M_n)_{n\geq 0}$ is almost surely convergent with limit $\frac{\Gamma\left(\frac{|X_0|+a_{1,1}}{r}\right)}{\Gamma\left(\frac{|X_0|}{r}\right)}\Xi$. Note that it does not matter whether $a_{1,1}/r < 1/2$ or not in this case, as the size of $X_n^{(1)}$ is sublinear. An application of Corollary 3.5 from [34] further implies the convergence

$$\sqrt{\frac{a_{1,1}\Gamma\left(\frac{|X_0|+a_{1,1}}{r}\right)}{X_0^{(1)}\Gamma\left(\frac{|X_0|}{r}\right)}}\,n^{-\frac{a_{1,1}}{2r}}\left(X_n^{(1)}-\mathbb{E}\left[X_n^{(1)}\right]-n^{\frac{a_{1,1}}{r}}\Xi\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,1)$$

as $n\to\infty$. This in turn implies that as $n\to\infty$,

$$\frac{1}{\sqrt{n^{\frac{a_{1,1}}{r}}}}\left(X_n-\mathbb{E}[X_n]-n^{\frac{a_{1,1}}{r}}\Xi v_2\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\frac{a_{1,1}X_0^{(1)}\Gamma\left(\frac{|X_0|}{r}\right)}{\Gamma\left(\frac{|X_0|+a_{1,1}}{r}\right)}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right). \tag{3.6}$$

□

3.2 Proof of Theorem 1.2.5

For case 1 of Theorem 1.2.5, we set

$$Z_n := \frac{1}{\sqrt{n}}\begin{pmatrix}
\Re\left(\pi_1(Y_n)-\gamma_n^{(1)}\frac{\Gamma(|X_0|/r+\lambda_1/r)}{\Gamma(|X_0|/r)}\Xi_1\right)\\
\Im\left(\pi_1(Y_n)-\gamma_n^{(1)}\frac{\Gamma(|X_0|/r+\lambda_1/r)}{\Gamma(|X_0|/r)}\Xi_1\right)\\
\Re\left(\pi_2(Y_n)-\gamma_n^{(2)}\frac{\Gamma(|X_0|/r+\lambda_2/r)}{\Gamma(|X_0|/r)}\Xi_2\right)\\
\Im\left(\pi_2(Y_n)-\gamma_n^{(2)}\frac{\Gamma(|X_0|/r+\lambda_2/r)}{\Gamma(|X_0|/r)}\Xi_2\right)\\
\vdots\\
\Re\left(\pi_p(Y_n)-\gamma_n^{(p)}\frac{\Gamma(|X_0|/r+\lambda_p/r)}{\Gamma(|X_0|/r)}\Xi_p\right)\\
\Im\left(\pi_p(Y_n)-\gamma_n^{(p)}\frac{\Gamma(|X_0|/r+\lambda_p/r)}{\Gamma(|X_0|/r)}\Xi_p\right)\\
\Re(\pi_{p+1}(Y_n))\\
\Im(\pi_{p+1}(Y_n))\\
\vdots\\
\Re(\pi_q(Y_n))\\
\Im(\pi_q(Y_n))
\end{pmatrix}.$$

If we are in case 2 on the contrary, we define

$$Z_n := \frac{1}{\sqrt{n\log(n)}} \begin{pmatrix} \Re(\pi_1(Y_n) - \gamma_n^{(1)} \frac{\Gamma(|X_0|/r+\lambda_1/r)}{\Gamma(|X_0|/r)} \Xi_1) \\ \Im(\pi_1(Y_n) - \gamma_n^{(1)} \frac{\Gamma(|X_0|/r+\lambda_1/r)}{\Gamma(|X_0|/r)} \Xi_1) \\ \Re(\pi_2(Y_n) - \gamma_n^{(2)} \frac{\Gamma(|X_0|/r+\lambda_2/r)}{\Gamma(|X_0|/r)} \Xi_2) \\ \Im(\pi_2(Y_n) - \gamma_n^{(2)} \frac{\Gamma(|X_0|/r+\lambda_2/r)}{\Gamma(|X_0|/r)} \Xi_2) \\ \vdots \\ \Re(\pi_p(Y_n) - \gamma_n^{(p)} \frac{\Gamma(|X_0|/r+\lambda_p/r)}{\Gamma(|X_0|/r)} \Xi_p) \\ \Im(\pi_p(Y_n) - \gamma_n^{(p)} \frac{\Gamma(|X_0|/r+\lambda_p/r)}{\Gamma(|X_0|/r)} \Xi_p) \\ \Re(\pi_{p+1}(Y_n)) \\ \Im(\pi_{p+1}(Y_n)) \\ \vdots \\ \Re(\pi_q(Y_n)) \\ \Im(\pi_q(Y_n)) \end{pmatrix}.$$

In each case, the vectors Z_n serve the purpose to study the joint fluctuations of all projections. Note that some components of Z_n may be equal or 0. For example, Z_n in chapter 2 is a trimmed version of Z_n of the current chapter (that is, we deleted all zero or equal components). The aim of the next section is to show convergence in distribution of Z_n to a multivariate mixed Gaussian distribution. More precisely, similar to Proposition 2.1.3, we prove the following theorem.

Theorem 3.2.1. *As* $n \to \infty$,

$$Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_V),$$

where $\mathcal{N}(0, \Sigma_V)$ *is a mixed Gaussian random vector with mixture components* $V^{(1)}, \ldots, V^{(q)}$ *and covariance matrix* Σ_V *defined in (3.7) to (3.9) and (3.10) to (3.11).*

Before we give the formulas for the entries of the covariance matrix Σ_V, consider again a standard Pólya urn scheme with $q \in \mathbb{N}$ colours and replacement matrix

$$R_{\text{Pólya}} := \begin{pmatrix} r & 0 & 0 & \cdots & 0 & 0 \\ 0 & r & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & & & & r & 0 \\ 0 & & \cdots & & 0 & r \end{pmatrix}$$

for $r \in \mathbb{N}$. It is well-known (see [23] or Theorem 1.2.3, for example) that

$$\frac{X_n}{rn + |X_0|} \xrightarrow{\text{a.s.}} D, \qquad n \to \infty,$$

where D is a $(q \times 1)$-Dirichlet distributed random vector with parameters $\left(X_0^{(1)}/r, \ldots, X_0^{(q)}/r\right)$. In this case, Theorem 1.2.5 takes the form

$$\sqrt{\frac{1}{r^2 n}}(X_n - rD) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \begin{pmatrix} D^{(1)}(1-D^{(1)}) & -D^{(1)}D^{(2)} & \dots & -D^{(1)}D^{(q)} \\ -D^{(1)}D^{(2)} & D^{(2)}(1-D^{(2)}) & \dots & -D^{(2)}D^{(q)} \\ \vdots & & \ddots & \vdots \\ -D^{(1)}D^{(q-1)} & -D^{(2)}D^{(q-1)} & & -D^{(q-1)}D^{(q)} \\ -D^{(1)}D^{(q)} & -D^{(2)}D^{(q)} & \dots & D^{(q)}(1-D^{(q)}) \end{pmatrix}\right)$$

as $n \to \infty$.

The matrix Σ_V: First case of Theorem 1.2.5. First of all, asymptotically, the components of Z_n decompose into three stochastically independent classes. These are

- components associated with eigenvalues λ_k such that $\lambda_k = r$,
- components associated with eigenvalues λ_k such that $\frac{r}{2} < \Re(\lambda_k) < r$ and
- components associated with eigenvalues λ_k such that $\Re(\lambda_k) < \frac{r}{2}$.

Consequently, $(\Sigma_V)_{k,\ell} = 0$ if k and ℓ belong to different classes, and we can give the definitions of the components of Σ_V for each of the three classes separately.
In order to describe the entries of Σ_V, recall that we earlier denoted the geometric multiplicity of the eigenvalue r by $a + c$. The first class consists of components $Z_n^{(1)}, \dots, Z_n^{(2(a+c))}$ (even though all even components are zero, as the projections are real) and for $k, \ell \in \{1, \dots, a + c\}$,

$$(\Sigma_V)_{2k-1,2\ell-1} := \begin{cases} r^2\left(\frac{\Xi_k}{r} + \frac{\pi_k(X_0)}{|X_0|}\right)\left(1 - \left(\frac{\Xi_k}{r} + \frac{\pi_k(X_0)}{|X_0|}\right)\right), & k = \ell, \\ -r^2\left(\frac{\Xi_k}{r} + \frac{\pi_k(X_0)}{|X_0|}\right)\left(\frac{\Xi_\ell}{r} + \frac{\pi_\ell(X_0)}{|X_0|}\right), & k \neq \ell. \end{cases} \tag{3.7}$$

Note that this is exactly the asymptotic variance (respectively, covariance) of the different types in a Pólya urn in the example above. This observation agrees with our interpretation of the classes $\mathcal{C}_1, \dots, \mathcal{C}_{a+c}$ as supercolours. The additional factor r^2 is due to a different scaling in the theorems.
If $k, \ell \in \{a + c + 1, \dots, p\}$, the asymptotic covariances between real parts, imaginary parts and real and imaginary parts of components of this class are given by

$$(\Sigma_V)_{2k-1,2\ell-1} := \sum_{m=1}^{q} V^{(m)} \Re\left(\frac{\left(\frac{\bar\lambda_k+\bar\lambda_\ell}{r} - 1\right)\lambda_k\lambda_\ell u_k^{(m)} u_\ell^{(m)}}{2\left|\frac{\lambda_k+\lambda_\ell}{r} - 1\right|^2} + \frac{\left(\frac{\bar\lambda_k+\lambda_\ell}{r} - 1\right)\lambda_k\bar\lambda_\ell \bar u_k^{(m)} u_\ell^{(m)}}{2\left|\frac{\lambda_k+\bar\lambda_\ell}{r} - 1\right|^2}\right),$$

$$(\Sigma_V)_{2k,2\ell} := \sum_{m=1}^{q} V^{(m)} \Re\left(-\frac{\left(\frac{\bar\lambda_k+\bar\lambda_\ell}{r} - 1\right)\lambda_k\lambda_\ell u_k^{(m)} u_\ell^{(m)}}{2\left|1 - \frac{\lambda_k+\lambda_\ell}{r}\right|^2} + \frac{\left(\frac{\bar\lambda_k+\lambda_\ell}{r} - 1\right)\lambda_k\bar\lambda_\ell u_k^{(m)} \bar u_\ell^{(m)}}{2\left|1 - \frac{\lambda_k+\bar\lambda_\ell}{r}\right|^2}\right) \text{ and}$$

$$(\Sigma_V)_{2k-1,2\ell} := \sum_{m=1}^{q} V^{(m)} \Im\left(\frac{\left(\frac{\bar\lambda_k+\bar\lambda_\ell}{r} - 1\right)\lambda_k\lambda_\ell u_k^{(m)} u_\ell^{(m)}}{2\left|1 - \frac{\lambda_k+\lambda_\ell}{r}\right|^2} + \frac{\left(\frac{\bar\lambda_k+\lambda_\ell}{r} - 1\right)\lambda_k\bar\lambda_\ell u_k^{(m)} \bar u_\ell^{(m)}}{2\left|1 - \frac{\lambda_k+\bar\lambda_\ell}{r}\right|^2}\right), \tag{3.8}$$

respectively.
Similarly, if $k, \ell \in \{p+1, \dots, q\}$, the asymptotic covariances between real parts, imaginary parts and real and imaginary parts of components of this class are given by

$$(\Sigma_V)_{2k-1,2\ell-1} := \sum_{m=1}^{q} V^{(m)} \Re \left(\frac{\left(1 - \frac{\bar{\lambda}_k + \bar{\lambda}_\ell}{r}\right) \lambda_k \lambda_\ell u_k^{(m)} u_\ell^{(m)}}{2 \left| \frac{\lambda_k + \lambda_\ell}{r} - 1 \right|^2} + \frac{\left(1 - \frac{\bar{\lambda}_k + \lambda_\ell}{r}\right) \lambda_k \bar{\lambda}_\ell \bar{u}_k^{(m)} u_\ell^{(m)}}{2 \left| \frac{\lambda_k + \bar{\lambda}_\ell}{r} - 1 \right|^2} \right),$$

$$(\Sigma_V)_{2k,2\ell} := \sum_{m=1}^{q} V^{(m)} \Re \left(-\frac{\left(1 - \frac{\bar{\lambda}_k + \bar{\lambda}_\ell}{r}\right) \lambda_k \lambda_\ell u_k^{(m)} u_\ell^{(m)}}{2 \left| 1 - \frac{\lambda_k + \lambda_\ell}{r} \right|^2} + \frac{\left(1 - \frac{\bar{\lambda}_k + \lambda_\ell}{r}\right) \lambda_k \bar{\lambda}_\ell u_k^{(m)} \bar{u}_\ell^{(m)}}{2 \left| 1 - \frac{\lambda_k + \bar{\lambda}_\ell}{r} \right|^2} \right) \quad \text{and}$$

$$(\Sigma_V)_{2k-1,2\ell} := \sum_{m=1}^{q} V^{(m)} \Im \left(\frac{\left(1 - \frac{\bar{\lambda}_k + \bar{\lambda}_\ell}{r}\right) \lambda_k \lambda_\ell u_k^{(m)} u_\ell^{(m)}}{2 \left| 1 - \frac{\lambda_k + \lambda_\ell}{r} \right|^2} + \frac{\left(1 - \frac{\bar{\lambda}_k + \lambda_\ell}{r}\right) \lambda_k \bar{\lambda}_\ell u_k^{(m)} \bar{u}_\ell^{(m)}}{2 \left| 1 - \frac{\lambda_k + \bar{\lambda}_\ell}{r} \right|^2} \right), \tag{3.9}$$

respectively. Note the close analogy of formulas (3.8) and (3.9).

The matrix Σ_V: Second case of Theorem 1.2.5. In this case, where there is at least one dominant class that gives rise to an eigenvalue λ_ℓ with $\Re(\lambda_\ell)/r = 1/2$, the matrix Σ_V has a lot more zero entries than in the previous case. This is due to the different scaling. The only components of Z_n that do not necessarily have zero variance in the limit are

- components associated with dominant eigenvalues λ_k such that $\Re(\lambda_k) = \frac{r}{2}$.

For these k, if $\Im(\lambda_k) \neq 0$,

$$(\Sigma_\nu)_{2k-1,2k-1} = (\Sigma_\nu)_{2k,2k} = \frac{|\lambda_k|^2}{2} \sum_{m=1}^{q} V^{(m)} \left| u_k^{(m)} \right|^2. \tag{3.10}$$

If $\Im(\lambda_k) = 0$, $(\Sigma_\nu)_{2k,2k} = 0$ and

$$(\Sigma_\nu)_{2k-1,2k-1} = \lambda_k^2 \sum_{m=1}^{q} V^{(m)} \left(u_k^{(m)} \right)^2. \tag{3.11}$$

Remark 13. The variances and covariances take a simpler form if one considers the complex quantities instead. Note that the covariance structure in particular implies that

$$Z_n^{(2k-1)} + i Z_n^{(2k)} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \sum_{m=1}^{q} V^{(m)} \frac{|\lambda_k|^2 \left| u_k^{(m)} \right|^2}{\left| \frac{2\Re(\lambda_k)}{r} - 1 \right|} \right), \qquad n \to \infty \tag{3.12}$$

and

$$\mathbb{E}\left[\left(Z_n^{(2k-1)} + i Z_n^{(2k)} \right) \left(Z_n^{(2\ell-1)} - i Z_n^{(2\ell)} \right) \right] \xrightarrow{n \to \infty} \sum_{m=1}^{q} V^{(m)} \frac{\left(\frac{\bar{\lambda}_k + \lambda_\ell}{r} - 1 \right) \lambda_k \bar{\lambda}_\ell u_k^{(m)} \bar{u}_\ell^{(m)}}{\left| \frac{\bar{\lambda}_k + \lambda_\ell}{r} - 1 \right|^2} \tag{3.13}$$

for large eigenvalues λ_k, λ_ℓ and analogously for small eigenvalues.

Summary of covariance structure: First case.

1. The "supercolour" components $Z_n^{(1)}, \dots, Z_n^{(2(a+c))}$ are asymptotically independent of all other components.
2. $Z_n^{(i)}$ and $Z_n^{(j)}$ are asymptotically independent for $i \in \{1, \dots, 2p\}$ and $j \in \{2p+1, \dots, 2q\}$.
3. If the eigenvalue λ_k belongs to a class of category 3, $Z_n^{(2k-1)}$ and $Z_n^{(2k)}$ tend to zero in probability. That is, the fluctuations of projections corresponding to eigenvalues λ_k of category 3 classes vanish in the $\sqrt{n}$ scaling. This might be due to the fact that there are too little draws from these classes compared to the other classes. So Theorem 3.2.1 says nothing about the fluctuations within these classes (or, at least, nothing particularly interesting), as the draws from the dominant colours dominate in the limit and there is too little fluctuation among the remaining colours.

Summary of covariance structure: Second case.

1. The fluctuations of components $Z_n^{(2k-1)}, Z_n^{(2k)}$ with $\Re(\lambda_k) \neq \frac{r}{2}$ are at most of order $\sqrt{n}$, as in the first case. However, they tend to zero in probability due to the $\sqrt{n \log(n)}$ scaling.
2. For each dominant λ_k with $\Re(\lambda_k) = \frac{r}{2}$, $Z_n^{(2k-1)}$ and $Z_n^{(2k)}$ are asymptotically independent. Furthermore, for different dominant k, ℓ with $\Re(\lambda_k) = \Re(\lambda_\ell) = \frac{r}{2}$, $Z_n^{(2k-1)}, Z_n^{(2k)}, Z_n^{(2\ell-1)}$ and $Z_n^{(2\ell)}$ are asymptotically independent.

Finally, note the following structural difference between the components of Z_n: The first $2p$ components of Z_n describe the fluctuation of sums of the form $(\sum_{k=n}^{\infty} I_k)_n$ and hence the way in which the convergent martingales converge to their limits. The following components describe the fluctuation of sums of the form $(\sum_{k=0}^{n} I_k)_n$. One way to deal with these different types of fluctuation is to use classical theorems for triangular arrays. In the proof of Theorem 3.2.1, we will employ Corollary 3.1 from [34]:

Proposition 3.2.2. *Let $\{S_{n,j}, \mathcal{F}_{n,j}, 1 \leq j \leq k_n, n \geq 1\}$ be a zero-mean, square-integrable martingale array with increments $I_{n,j}$ and let η^2 be an a.s. finite random variable. Suppose that for all $\varepsilon > 0$,*

$$\sum_{j=1}^{k_n} \mathbb{E}[I_{n,j}^2 \mathbb{1}_{\{|I_{n,j}| > \varepsilon\}} | \mathcal{F}_{n,j-1}] \xrightarrow{\mathbb{P}} 0, \tag{3.14}$$

and

$$\sum_{j=1}^{k_n} \mathbb{E}[I_{n,j}^2 | \mathcal{F}_{n,j-1}] \xrightarrow{\mathbb{P}} \eta^2, \tag{3.15}$$

and $\mathcal{F}_{n,j} \subseteq \mathcal{F}_{n+1,j}$ for $1 \leq j \leq k_n$, $n \geq 1$. Then

$$S_{n,k_n} = \sum_{j=1}^{k_n} I_{n,j} \xrightarrow{\mathcal{L}} Z,$$

where the random variable Z *has characteristic function*

$$\phi(t) = \mathbb{E}\left[\exp\left(-\frac{1}{2}\eta^2 t^2\right)\right].$$

Proof of Theorem 3.2.1. We will prove Theorem 3.2.1 via the Cramér-Wold device. Consequently, we study weak convergence of all linear combinations of the components of Z_n. Let $\alpha_1, \ldots, \alpha_{2q} \in \mathbb{R}$. Our aim is to apply Proposition 3.2.2 to an appropriate martingale array. To this end, in simple words, we decompose the sum $\alpha_1 Z_n^{(1)} + \ldots + \alpha_{2q} Z_n^{(2q)}$ into a sum of weighted martingale differences. We set $\mathcal{F}_{n,i} := \sigma(X_0, \ldots, X_i)$. This definition satisfies the condition on the filtration in Proposition 3.2.2.
First consider the case where for all dominant eigenvalues, $\Re(\lambda_k) \neq r/2$.
We rewrite the given linear combination $\alpha_1 Z_n^{(1)} + \ldots + \alpha_{2q} Z_n^{(2q)}$ as a sum of martingale differences. We simultaneously consider real and imaginary part of each eigenspace coefficient. For $1 \leq k \leq p$, write

$$\begin{aligned}
&\alpha_{2k-1} Z_n^{(2k-1)} + \alpha_{2k} Z_n^{(2k)} \\
&= \frac{1}{\sqrt{n}} \sum_{j=n}^{\infty} \left(\alpha_{2k-1} \Re\left(\gamma_n^{(k)}\left(M_j^{(k)} - M_{j+1}^{(k)}\right)\right) + \alpha_{2k} \Im\left(\gamma_n^{(k)}\left(M_j^{(k)} - M_{j+1}^{(k)}\right)\right)\right) \\
&= \frac{1}{\sqrt{n}} \left(\alpha_{2k-1} \Re\left(\gamma_n^{(k)}\right) + \alpha_{2k} \Im\left(\gamma_n^{(k)}\right)\right) \sum_{j=n}^{\infty} \Re\left(M_j^{(k)} - M_{j+1}^{(k)}\right) \\
&+ \frac{1}{\sqrt{n}} \left(\alpha_{2k} \Re\left(\gamma_n^{(k)}\right) - \alpha_{2k-1} \Im\left(\gamma_n^{(k)}\right)\right) \sum_{j=n}^{\infty} \Im\left(M_j^{(k)} - M_{j+1}^{(k)}\right) \\
&=: \beta_{2k-1}(n) \sum_{j=n}^{\infty} \Re\left(M_j^{(k)} - M_{j+1}^{(k)}\right) + \beta_{2k}(n) \sum_{j=n}^{\infty} \Im\left(M_j^{(k)} - M_{j+1}^{(k)}\right).
\end{aligned}$$

Set $g := \max\left\{-\frac{\lambda_k + |X_0|}{r} + 1 : 1 \leq k \leq q, \lambda_k + |X_0| \in r\mathbb{Z}_-\right\}$. Then for $p+1 \leq k \leq q$,

$$\begin{aligned}
&\alpha_{2k-1} Z_n^{(2k-1)} + \alpha_{2k} Z_n^{(2k)} \\
&= \frac{1}{\sqrt{n}} \left(\alpha_{2k-1} \Re\left(\gamma_n^{(k)}\right) + \alpha_{2k} \Im\left(\gamma_n^{(k)}\right)\right) \sum_{j=g}^{n-1} \Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right) \\
&+ \frac{1}{\sqrt{n}} \left(\alpha_{2k} \Re\left(\gamma_n^{(k)}\right) - \alpha_{2k-1} \Im\left(\gamma_n^{(k)}\right)\right) \sum_{j=g}^{n-1} \Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right) \\
&+ \frac{1}{\sqrt{n}} \left(\alpha_{2k-1} \Re\left(\gamma_n^{(k)}\right) + \alpha_{2k} \Im\left(\gamma_n^{(k)}\right)\right) \Re\left(M_g^{(k)}\right) \\
&+ \frac{1}{\sqrt{n}} \left(\alpha_{2k} \Re\left(\gamma_n^{(k)}\right) - \alpha_{2k-1} \Im\left(\gamma_n^{(k)}\right)\right) \Im\left(M_g^{(k)}\right) \\
&=: \beta_{2k-1}(n) \sum_{j=g}^{n-1} \Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right) + \beta_{2k}(n) \sum_{j=g}^{n-1} \Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right) + r_n(k).
\end{aligned}$$

So setting $r_n := \sum_{k=p+1}^{q} r_n(k)$,

$$\sum_{k=1}^{2q} \alpha_k Z_n^{(k)} = \sum_{k=1}^{p} \left(\beta_{2k-1}(n) \sum_{j=n}^{\infty} \Re\left(M_j^{(k)} - M_{j+1}^{(k)}\right) + \beta_{2k}(n) \sum_{j=n}^{\infty} \Im\left(M_j^{(k)} - M_{j+1}^{(k)}\right) \right)$$
$$+ \sum_{k=p+1}^{q} \left(\beta_{2k-1}(n) \sum_{j=g}^{n-1} \Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right) + \beta_{2k}(n) \sum_{j=g}^{n-1} \Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right) \right) + r_n.$$

For n fixed, this is an *infinite* martingale difference array plus error term r_n which tends to 0 almost surely as $n \to \infty$. We now cut off the tail of the series to work with a finite martingale difference array. More precisely, we choose a sequence $(k_n)_{n \geq 0} \uparrow \infty$ appropriately and write

$$\sum_{k=1}^{2q} \alpha_k Z_n^{(k)} = \sum_{k=1}^{p} \left(\beta_{2k-1}(n) \sum_{j=n}^{k_n} \Re\left(M_j^{(k)} - M_{j+1}^{(k)}\right) + \beta_{2k}(n) \sum_{j=n}^{k_n} \Im\left(M_j^{(k)} - M_{j+1}^{(k)}\right) \right)$$
$$+ \sum_{k=p+1}^{q} \left(\beta_{2k-1}(n) \sum_{j=g}^{n-1} \Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right) + \beta_{2k}(n) \sum_{j=g}^{n-1} \Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right) \right) + \varepsilon_n$$

such that $\varepsilon_n \to 0$ in L^2. The following lemma shows that $(k_n)_{n \geq 0} = (n^2)_{n \geq 0}$ is sufficient.

Lemma 3.2.1. *Let*

$$\varepsilon_n := \sum_{k=1}^{p} \left(\beta_{2k-1}(n) \sum_{j=n^2+1}^{\infty} \Re\left(M_j^{(k)} - M_{j+1}^{(k)}\right) + \beta_{2k}(n) \sum_{j=n^2+1}^{\infty} \Im\left(M_j^{(k)} - M_{j+1}^{(k)}\right) \right) + r_n.$$

Then

$$\varepsilon_n \xrightarrow{L^2} 0, \qquad n \to \infty.$$

Proof. It is easy to see that r_n tends to zero in L^2 as $\Re(\lambda_k) < r/2$ for all summands in this term. The remaining part follows immediately from Lemma 3.1.3. □

For the proof of Theorem 3.2.1, it is thus sufficient to show weak convergence of the martingale difference array

$$\sum_{k=1}^{p} \left(\beta_{2k-1}(n) \sum_{j=n}^{n^2} \Re\left(M_j^{(k)} - M_{j+1}^{(k)}\right) + \beta_{2k}(n) \sum_{j=n}^{n^2} \Im\left(M_j^{(k)} - M_{j+1}^{(k)}\right) \right)$$
$$+ \sum_{k=p+1}^{q} \left(\beta_{2k-1}(n) \sum_{j=g}^{n-1} \Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right) + \beta_{2k}(n) \sum_{j=g}^{n-1} \Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right) \right)$$

which perfectly fits into the setting of Proposition 3.2.2. We now check the conditions. Depending on the summation index j, there are two types of increments $I_{n,j}$. We use the shorthand

$$I_{n,j} := \begin{cases} \sum_{k=p+1}^{q} \left(\beta_{2k-1}(n)\Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right) + \beta_{2k}(n)\Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right)\right), & j < n, \\ \sum_{k=1}^{p} \left(\beta_{2k-1}(n)\Re\left(M_j^{(k)} - M_{j+1}^{(k)}\right) + \beta_{2k}(n)\Im\left(M_j^{(k)} - M_{j+1}^{(k)}\right)\right), & j \geq n. \end{cases}$$

The absolute value of these increments is deterministically bounded: For $j < n$,

$$\begin{aligned} |I_{n,j}| &\leq \sum_{k=p+1}^{q} |\beta_{2k-1}(n)| \left|\Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right)\right| + |\beta_{2k}(n)| \left|\Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right)\right| \\ &\leq C \sum_{k=p+1}^{q} n^{\Re(\lambda_k)/r-1/2} \left(\left|\Re\left(M_{j+1}^{(k)} - M_j^{(k)}\right)\right| + \left|\Im\left(M_{j+1}^{(k)} - M_j^{(k)}\right)\right|\right) \\ &\leq \sqrt{2}C \sum_{k=p+1}^{q} n^{\Re(\lambda_k)/r-1/2} \left|M_{j+1}^{(k)} - M_j^{(k)}\right| \\ &\leq D \quad n^{-1/2} \sum_{k=p+1}^{q} \left(\frac{n}{j}\right)^{\Re(\lambda_k)/r} = O\left(n^{\max\{\Re(\lambda_{p+1})/r, 0\}-1/2}\right) \end{aligned}$$

for all $j < n$. Here, $C, D > 0$ are constants and the uniform bound tends to 0 as $n \to \infty$. Analogously, for $n \leq j \leq n^2$,

$$|I_{n,j}| \leq C \quad n^{-1/2} \sum_{k=1}^{p} \left(\frac{n}{j}\right)^{\Re(\lambda_k)/r} = O\left(n^{-1/2}\right)$$

for all $n \leq j \leq n^2$ as $n \to \infty$. Again, $C > 0$ is a constant and the uniform bound tends to 0 as $n \to \infty$.
Fix $\varepsilon > 0$. Then by the above, for N big enough, for all $j = 1, \ldots, N^2$ we have $|I_{N,j}| < \varepsilon$, so in particular

$$\sum_{j=g}^{n^2} \mathbb{E}[I_{n,j}^2 \mathbb{1}_{\{|I_{n,j}| > \varepsilon\}} | \mathcal{F}_{n,j-1}] = 0$$

for all $n \geq N$. This implies condition (3.14).
We now turn to condition (3.15), which is computationally the hardest.
We rewrite the increments as

$$I_{n,j}^2 = \left(\xi_{n,j}^t (X_{j+1} - X_j) - \eta_{n,j}^t \frac{X_j}{rj + |X_0|}\right)^2$$

where

$$\xi_{n,j} := \sum_{k\in K} \frac{1}{\sqrt{n}} \left[\left(\alpha_{2k-1} \Re\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) + \alpha_{2k} \Im\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) \right) \Re(u_k) + \left(\alpha_{2k} \Re\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) - \alpha_{2k-1} \Im\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) \right) \Im(u_k) \right]$$

and

$$\eta_{n,j} := \sum_{k\in K} \frac{1}{\sqrt{n}} \left[\left(\alpha_{2k-1} \Re\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) + \alpha_{2k} \Im\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) \right) \Re(\lambda_k u_k) + \left(\alpha_{2k} \Re\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) - \alpha_{2k-1} \Im\left(\frac{\gamma_n^{(k)}}{\gamma_{j+1}^{(k)}}\right) \right) \Im(\lambda_k u_k) \right]$$

and the set K is either equal to $\{1, \dots, p\}$ or $\{p+1, \dots, q\}$, depending on j. With this,

$$\sum_{j=g}^{n^2} \mathbb{E}[I_{n,j}^2 | \mathcal{F}_{n,j}] = \sum_{j=g}^{n^2} \mathbb{E}\left[\left(\xi_{n,j}^t (X_{j+1} - X_j) - \eta_{n,j}^t \frac{X_j}{rj + |X_0|} \right)^2 | \mathcal{F}_j \right]$$
$$= \sum_{j=g}^{n^2} \sum_{m=1}^{q} \frac{X_j^{(m)}}{rj + |X_0|} \left(\xi_{n,j}^t \Delta_m - \eta_{n,j}^t \frac{X_j}{rj + |X_0|} \right)^2. \tag{3.16}$$

We show that (3.16) converges almost surely by looking at the different terms separately: Recall that each of the $\xi_{n,j}$ and $\eta_{n,j}$ is a sum over different eigenspace components.
In the first part of the sum where $j \le n-1$, there are only λ_k with $\Re(\lambda_k)/r < 1/2$, so the product of component k with component ℓ in the square $\left(\xi_{n,j}^t \Delta_m - \eta_{n,j}^t \frac{X_j}{rj+|X_0|}\right)^2$ is asymptotically equivalent to

$$\frac{1}{n} \sum_{j=g}^{n-1} \sum_{m=1}^{q} \frac{X_j^{(m)}}{rj + |X_0|} \left(\frac{n}{j}\right)^{\frac{\Re(\lambda_k + \lambda_\ell)}{r}}$$
$$\cdot \left(\left(\alpha_{2k-1} \Re\left(\lambda_k \left(u_k^{(m)} - \frac{u_k^t X_j}{rj + |X_0|} \right) \right) + \alpha_{2k} \Im\left(\lambda_k \left(u_k^{(m)} - \frac{u_k^t X_j}{rj + |X_0|} \right) \right) \right) \cos\left(\frac{\Im(\lambda_k)}{r} \log\left(\frac{n}{j}\right) \right) \right.$$
$$\left. + \left(\alpha_{2k} \Re\left(\lambda_k \left(u_k^{(m)} - \frac{u_k^t X_j}{rj + |X_0|} \right) \right) - \alpha_{2k-1} \Im\left(\lambda_k \left(u_k^{(m)} - \frac{u_k^t X_j}{rj + |X_0|} \right) \right) \right) \sin\left(\frac{\Im(\lambda_k)}{r} \log\left(\frac{n}{j}\right) \right) \right)$$
$$\cdot \left(\left(\alpha_{2\ell-1} \Re\left(\lambda_\ell \left(u_\ell^{(m)} - \frac{u_\ell^t X_j}{rj + |X_0|} \right) \right) + \alpha_{2\ell} \Im\left(\lambda_\ell \left(u_\ell^{(m)} - \frac{u_\ell^t X_j}{rj + |X_0|} \right) \right) \right) \cos\left(\frac{\Im(\lambda_\ell)}{r} \log\left(\frac{n}{j}\right) \right) \right.$$
$$\left. + \left(\alpha_{2\ell} \Re\left(\lambda_\ell \left(u_\ell^{(m)} - \frac{u_\ell^t X_j}{rj + |X_0|} \right) \right) - \alpha_{2\ell-1} \Im\left(\lambda_\ell \left(u_\ell^{(m)} - \frac{u_\ell^t X_j}{rj + |X_0|} \right) \right) \right) \sin\left(\frac{\Im(\lambda_\ell)}{r} \log\left(\frac{n}{j}\right) \right) \right).$$

Recall that the vector $\frac{X_j}{rj+|X_0|}$ converges to V almost surely. It follows that for $\lambda_k \neq r$, $u_k^t \frac{X_j}{rj+|X_0|} \to 0$ almost surely. So the sum above is almost surely asymptotically equivalent to

$$
\begin{aligned}
&\frac{1}{2n}\sum_{j=g}^{n-1}\sum_{m=1}^{q} V^{(m)}\left(\frac{n}{j}\right)^{\frac{\Re(\lambda_k+\lambda_\ell)}{r}}\\
&\cdot\Big(\Big((\alpha_{2k-1}\alpha_{2\ell-1}-\alpha_{2k}\alpha_{2\ell})\Re\left(\lambda_k\lambda_\ell u_k^{(m)}u_\ell^{(m)}\right)+(\alpha_{2k-1}\alpha_{2\ell}+\alpha_{2k}\alpha_{2\ell-1})\Im\left(\lambda_k\lambda_\ell u_k^{(m)}u_\ell^{(m)}\right)\Big)\\
&\cdot\cos\left(\frac{\Im(\lambda_k+\lambda_\ell)}{r}\log\left(\frac{n}{j}\right)\right)+\Big((\alpha_{2k-1}\alpha_{2\ell-1}+\alpha_{2k}\alpha_{2\ell})\Re\left(\lambda_k\bar{\lambda}_\ell u_k^{(m)}\bar{u}_\ell^{(m)}\right)\\
&+(\alpha_{2k}\alpha_{2\ell-1}-\alpha_{2k-1}\alpha_{2\ell})\Im\left(\lambda_k\bar{\lambda}_\ell u_k^{(m)}\bar{u}_\ell^{(m)}\right)\Big)\cos\left(\frac{\Im(\lambda_k-\lambda_\ell)}{r}\log\left(\frac{n}{j}\right)\right)\\
&+\Big((\alpha_{2k-1}\alpha_{2\ell}+\alpha_{2k}\alpha_{2\ell-1})\Re\left(\lambda_k\lambda_\ell u_k^{(m)}u_\ell^{(m)}\right)+(\alpha_{2k}\alpha_{2\ell}-\alpha_{2k-1}\alpha_{2\ell-1})\Im\left(\lambda_k\lambda_\ell u_k^{(m)}u_\ell^{(m)}\right)\Big)\\
&\cdot\sin\left(\frac{\Im(\lambda_k+\lambda_\ell)}{r}\log\left(\frac{n}{j}\right)\right)+\Big((\alpha_{2k}\alpha_{2\ell-1}-\alpha_{2k-1}\alpha_{2\ell})\Re\left(\lambda_k\bar{\lambda}_\ell u_k^{(m)}\bar{u}_\ell^{(m)}\right)\\
&-(\alpha_{2k-1}\alpha_{2\ell-1}+\alpha_{2k}\alpha_{2\ell})\Im\left(\lambda_k\bar{\lambda}_\ell u_k^{(m)}\bar{u}_\ell^{(m)}\right)\Big)\sin\left(\frac{\Im(\lambda_k-\lambda_\ell)}{r}\log\left(\frac{n}{j}\right)\right)\Big)\\
&\to\alpha_{2k-1}\alpha_{2\ell-1}(\Sigma_V)_{2k-1,2\ell-1}+\alpha_{2k}\alpha_{2\ell}(\Sigma_V)_{2k,2\ell}+\alpha_{2k-1}\alpha_{2\ell}(\Sigma_V)_{2k-1,2\ell}+\alpha_{2k}\alpha_{2\ell-1}(\Sigma_V)_{2k,2\ell-1}
\end{aligned}
$$

as $n\to\infty$.
The same calculation shows convergence for summands k,ℓ with $1/2<\Re(\lambda_k)/r,\Re(\lambda_\ell)/r<1$. Furthermore, there are no summands with $\Re(\lambda_k)/r\le 1/2$ and $\Re(\lambda_\ell)/r>1/2$. Let now $1/2<\Re(\lambda_k)/r<1$ and $\lambda_\ell=r$. Then the component k with component ℓ product converges to

$$
\begin{aligned}
&r^2\sum_{m=1}^{q}V^{(m)}\left(\frac{\Xi_\ell}{r}+\frac{\pi_\ell(X_0)}{|X_0|}-u_\ell^{(m)}\right)\left(\alpha_{2k-1}\alpha_{2\ell-1}\cdot\Re\left(u_k^{(m)}\right)+\alpha_{2k}\alpha_{2\ell-1}\Im\left(u_k^{(m)}\right)\right)\\
=\;&0\\
=\;&\alpha_{2k-1}\alpha_{2\ell-1}(\Sigma_V)_{2k-1,2\ell-1}\alpha_{2k-1}\alpha_{2\ell}(\Sigma_V)_{2k-1,2\ell}.
\end{aligned}
$$

Finally, if $\lambda_k=\lambda_\ell=r$, the product tends to

$$
\begin{aligned}
&\alpha_{2k-1}\alpha_{2\ell-1}r^2\sum_{m=1}^{q}V^{(m)}\left(u_k^{(m)}-\left(\frac{\Xi_k}{r}+\frac{\pi_k(X_0)}{|X_0|}\right)\right)\left(u_\ell^{(m)}-\left(\frac{\Xi_\ell}{r}+\frac{\pi_\ell(X_0)}{|X_0|}\right)\right)\\
=&\begin{cases}-\alpha_{2k-1}\alpha_{2\ell-1}r^2\left(\frac{\Xi_k}{r}+\frac{\pi_k(X_0)}{|X_0|}\right)\left(\frac{\Xi_\ell}{r}+\frac{\pi_\ell(X_0)}{|X_0|}\right), & k\neq\ell\\ \alpha_{2k-1}^2r^2\left(\frac{\Xi_k}{r}+\frac{\pi_k(X_0)}{|X_0|}\right)\left(1-\left(\frac{\Xi_k}{r}+\frac{\pi_k(X_0)}{|X_0|}\right)\right), & k=\ell\end{cases}\\
=&\alpha_{2k-1}\alpha_{2\ell-1}(\Sigma_V)_{2k-1,2\ell-1}.
\end{aligned}
$$

In total, this implies that

$$
\sum_{j=g}^{n^2}\mathbb{E}[I_{n,j}^2|\mathcal{F}_{n,j}]\xrightarrow{a.s.}\sum_{i,j=1}^{2q}\alpha_i\alpha_j(\Sigma_V)_{i,j}=(\alpha_1,\dots,\alpha_{2q})\Sigma_V(\alpha_1,\dots,\alpha_{2q})^t.
$$

By Proposition 3.2.2, $\alpha_1 Z_n^{(1)}+\ldots+\alpha_{2q}Z_n^{(2q)}$ converges weakly to a random variable with characteristic function $\psi(t) = \mathbb{E}[\exp(-1/2\left((\alpha_1,\ldots,\alpha_{2q})\Sigma_V(\alpha_1,\ldots,\alpha_{2q})^t\right)^2 t^2)]$. Now the Cramér-Wold device implies weak convergence of $(Z_n)_{n\geq 0}$ to a mixed multivariate normal distribution with covariance matrix Σ_V given V: If the almost sure limit V is deterministic, then the asymptotic distribution of Z_n is the normal distribution, as can be seen from the characteristic function. If the almost sure limit V is not deterministic, then, conditionally on V, Z_n converges to some normal limit law, so the asymptotic distribution of Z_n is a mixed normal distribution.

We finally consider the case where there is some k such that $\Re(\lambda_k)/r = 1/2$. By very similar calculations, there is also weak convergence of $\alpha_1 Z_n^{(1)} + \ldots + \alpha_{2q} Z_n^{(2q)}$ to a random variable with characteristic function $\mathbb{E}[\exp(-1/2\left((\alpha_1,\ldots,\alpha_{2q})\Sigma_V(\alpha_1,\ldots,\alpha_{2q})^t\right)^2 t^2)]$, where Σ_V is defined in equations (3.10) and (3.11). Due to the scaling, the matrix Σ_V has a lot more zero entries in this case. Again, the Cramér-Wold device implies Theorem 3.2.1 in this case. □

Proof of Theorem 1.2.5. It remains to show that under our assumptions, in both cases of Theorem 1.2.5, the covariance matrix A_V given the proportions of the supercolours V is of the stated form and almost surely has positive entries in the specified positions.
Because all complex conjugates of eigenvectors in our basis are also eigenvectors in the basis, we have the almost sure asymptotic equivalence

$$\frac{1}{\sqrt{n\ell_n}}\left(Y_n - \sum_{k=1}^{p} n^{\frac{\lambda_k}{r}}\Xi_k v_k\right) \sim \sum_{k=1}^{q}\left(Z_n^{(2k-1)}\Re(v_k) - Z_n^{(2k)}\Im(v_k)\right) = MZ_n,$$

where $\ell_n = 1$ in case 1 of the theorem and $\ell_n = \log(n)$ in case 2. This shows that $A_V = M\Sigma_V M^t$ in both cases.
For a fixed colour j, the conditional variance of colour j is the conditional variance of

$$\sum_{k=1}^{q}\left(Z_n^{(2k-1)}\Re\left(v_k^{(j)}\right) - Z_n^{(2k)}\Im\left(v_k^{(j)}\right)\right). \tag{3.17}$$

Now these are exactly linear combinations of the components of Z_n as considered in the last proof.

First case: Non-dominant colours. Let j be a non-dominant colour. By our choice of right eigenvectors, for all dominant colours k, we have $v_k^{(j)} = 0$. So (3.17) reduces to a sum over eigenvectors from category 3 classes. But all variances and covariances of category 3 projections $Z_n^{(2k-1)}, Z_n^{(2k)}$ are zero in the limit and thus, $(A_V)_{j,j} = 0$.

Second case: Dominant colours. Suppose that j is a dominant colour in class $\mathcal{C}_m$. Again, by our choice of right eigenvectors, the sum (3.17) reduces to a sum over eigenvectors associated with $\mathcal{C}_m$, as $v_k^{(j)} \neq 0$ only if v_k is an eigenvector associated with class $\mathcal{C}_m$. If $|\mathcal{C}_m| = 1$, then (3.16) is equal to

$$(A_V)_{j,j} = r^2\left(\frac{\Xi_m}{r} + \frac{\pi_k(X_0)}{|X_0|}\right)\left(1 - \left(\frac{\Xi_m}{r} + \frac{\pi_k(X_0)}{|X_0|}\right)\right),$$

which is positive almost surely by Remark 11, because there is at least one other dominant class by our assumptions.
Let $|\mathcal{C}_m| > 1$. In this case, we have to look at the different components of the variance of

colour j coming from the projections associated with class $\mathcal{C}_m$ in more detail. Recall that the sum (3.16) is an approximation of the conditional variance of (3.17), if the coefficients are chosen appropriately. Now, each of the n^2+1-g summands in (3.16) is non-negative and thus any sum over less terms yields a lower bound for the whole sum. Only considering part of the sum has the advantage that the variances and covariances of the fluctuations in the sum grow at a different speed in the beginning (respectively end) of the urn. For example, for n large and $\varepsilon \in (0,1)$, we can either sum from g to εn or from $\varepsilon^{-1}n$ to n^2 to get a lower bound. In the first case, there are only summands with $\Re(\lambda_k), \Re(\lambda_\ell) \le r/2$. A calculation as in the proof of Theorem 3.2.1 shows that the contribution coming from the fluctuations in projections π_k, π_ℓ to the sum (3.16) cut off at εn is at most of order $\varepsilon^{1-\Re(\lambda_k+\lambda_\ell)/r}$ (again, with coefficients chosen appropriately). In the second case, there are only summands with $\Re(\lambda_k), \Re(\lambda_\ell) > r/2$. The contribution coming from the fluctuations in projections π_k, π_ℓ to the sum (3.16) without the first $\varepsilon^{-1}n$ summands with coefficients chosen appropriately is at most of order $\varepsilon^{\Re(\lambda_k+\lambda_\ell)/r-1}$. In particular, the contribution to the variance of colour j from projections with real part close to $r/2$ (and nonzero coefficients) is the greatest.
We now choose k^* such that among all possible eigenvalues $\lambda_k \neq 0$ associated to $\mathcal{C}_m$, the distance $|\Re(\lambda_{k^*})/r - 1/2|$ is minimal and $\left|v_{k^*}^{(j)}\right| > 0$. Note that this is possible as the Perron-Frobenius eigenvalue r associated with $\mathcal{C}_m$ satisfies these conditions, for example.
In case 1 of Theorem 1.2.5, there are only the cases $\Re(\lambda_{k^*})/r < 1/2$ and $\Re(\lambda_{k^*})/r > 1/2$. Assume $\Re(\lambda_{k^*})/r < 1/2$. If $\lambda_{k^*} \in \mathbb{R}$ is real and a simple eigenvalue, there is only one dominant term of order $\varepsilon^{1-2\Re(\lambda_{k^*})/r}$, which has coefficient

$$\frac{|\lambda_{k^*}|^2 \left|v_{k^*}^{(j)}\right|^2}{1-2\lambda_{k^*}/r} \sum_{m=1}^{q} V^{(m)} \left|u_{k^*}^{(m)}\right|^2.$$

This is positive almost surely. On the contrary, if $\lambda_{k^*} \in \mathbb{C}\backslash\mathbb{R}$ is a simple eigenvalue, we choose the eigenvalue λ_{k^*} from the complex conjugated pair with equal real parts that has $\Im(\lambda_{k^*}) > 0$. Furthermore, we choose ε small enough and such that $2\Im(\lambda_{k^*})/r \log(\varepsilon)$ is a negative multiple of 2π and then cut off at εn (ε is chosen such that the sine-terms that arise from the cutting off of the sum vanish, and the cosine terms are equal to one). With this choice, the dominant term is of order $\varepsilon^{1-2\Re(\lambda_{k^*})/r}$. It has coefficient

$$\begin{aligned}&\frac{|\lambda_{k^*}|^2 \left|v_{k^*}^{(j)}\right|^2}{2\Re(1-2\lambda_{k^*}/r)} \sum_{m=1}^{q} V^{(m)} \left|u_{k^*}^{(m)}\right|^2 \\ &\quad + \frac{1}{2|1-2\lambda_{k^*}/r|^2} \sum_{m=1}^{q} V^{(m)} \Re\left((1-2\bar{\lambda}_{k^*}/r)\lambda_{k^*}^2 \left(u_{k^*}^{(m)}\right)^2 \left(\bar{v}_{k^*}^{(j)}\right)^2\right),\end{aligned}$$

which is positive (recall $\left|v_{k^*}^{(j)}\right| > 0$ and $\Re(\lambda_{k^*}), \Im(\lambda_{k^*}) > 0$).
It remains to consider the case where λ_{k^*} is a multiple eigenvalue. This case is more involved as we have to ensure that the various terms of equal size do not cancel. We do this by the following trick: The conditional variance of colour j is independent of the choice of bases $\{u_1, \dots, u_q\}$ and $\{v_1, \dots, v_q\}$. In particular, we may choose these bases in a certain way, for each realisation of V:
Assume that $\lambda \in \{\lambda_1, \dots, \lambda_q\}$ is a multiple eigenvalue of submatrix $T_{i,i}$ (or $Q_{i,i}$) with $\Im(\lambda) \ge 0$,

such that its eigenspace has dimension $s \geq 2$. We wish to choose the corresponding left eigenvectors in a particular way. Let $V \in \mathbb{R}^q$ be the random vector defined in Theorem 1.2.3 and the paragraph below. Its entries are strictly positive on $T_{i,i}$ (or $Q_{i,i}$) with probability one and zero on all classes of category 3. We may thus use V restricted to the components in $T_{i,i}$ (or $Q_{i,i}$) to define a scalar product on $\mathbb{R}^\ell$, where ℓ is the number of types in $T_{i,i}$ (or $Q_{i,i}$). Suppose that $\tilde{u}_k, \tilde{u}_{k+1}, \dots, \tilde{u}_{k+s-1}$ is a choice of basis of the eigenspace under consideration, whose components are zero on all other dominant classes. Let $A := \{\mathfrak{R}(\tilde{u}_{k+j}) : 0 \leq j \leq s-1, \mathfrak{R}(\tilde{u}_{k+j}) \neq 0\} \cup \{\mathfrak{I}(\tilde{u}_{k+j}) : 0 \leq j \leq s-1, \mathfrak{I}(\tilde{u}_{k+j}) \neq 0\}$. The elements of A are linearly independent. By the standard theory, we may now choose left eigenvectors $u_k, \dots, u_{k+s-1}$ for λ such that for all $x \neq y$,

$$\langle \mathfrak{R}(u_x), \mathfrak{R}(u_y) \rangle_V := \sum_{m=1}^{q} V^{(m)} \mathfrak{R}\left(u_x^{(m)}\right) \mathfrak{R}\left(u_y^{(m)}\right) = 0,$$

$$\langle \mathfrak{I}(u_x), \mathfrak{I}(u_y) \rangle_V := \sum_{m=1}^{q} V^{(m)} \mathfrak{I}\left(u_x^{(m)}\right) \mathfrak{I}\left(u_y^{(m)}\right) = 0,$$

$$\langle \mathfrak{R}(u_x), \mathfrak{I}(u_y) \rangle_V := \sum_{m=1}^{q} V^{(m)} \mathfrak{R}\left(u_x^{(m)}\right) \mathfrak{I}\left(u_y^{(m)}\right) = 0.$$

(Furthermore, because u_x has non-zero components on exactly one dominant class (and maybe on some non-dominant components), multiple eigenvalues from *different* dominant classes are automatically "orthogonal" with respect to $\langle \cdot, \cdot \rangle_V$.) We additionally assume the other properties of the dual bases from section 1.2.

Now, with this choice, the covariances of the associated components of Z are zero. Hence each projection associated with this eigenvalue yields a lower bound on the variance.

If $\mathfrak{R}(\lambda_k)/r > 1/2$, we choose ε small enough, start the sum at $\varepsilon^{-1}n$ and proceed analogously. The dominant term now is of order $\varepsilon^{2\mathfrak{R}(\lambda_k)/r-1}$ and has non-zero coefficient.

If there are a big eigenvalue λ_k and a small eigenvalue λ_ℓ such that $|\mathfrak{R}(\lambda_k)/r - 1/2| = |\mathfrak{R}(\lambda_\ell)/r - 1/2|$, we can choose either.

In case 2 of Theorem 1.2.5, there are eigenvalues λ_k associated with class $\mathcal{C}_m$ such that $\mathfrak{R}(\lambda_k)/r = 1/2$. This case is even simpler as everything that is not killed by the scaling is independent. The asymptotic variance of colour j is then the weighted sum over all variances of real and imaginary parts of the $r/2$-projections in the class, which are not identically zero, because all different summands are asymptotically uncorrelated.

□

4 Further Applications

Below, we consider three examples which illustrate the statement of Theorem 1.2.5 for both small and large urns. Within each of the models, a phase transition takes place. Note however, that the result for small urns is covered in [74] and [40].

4.1 m-ary search tree

The evolution of the vector of the number of nodes containing $0, 1, \dots, m-2$ keys in an $m-$ary search tree under the uniform permutation model can be encoded by the following urn model: We have generating matrix

$$R_m = \begin{pmatrix} -1 & 0 & & & & m \\ 2 & -2 & & & & \\ & 3 & -3 & & & \\ & & & \ddots & & \\ & & & & m-1 & -(m-1) \end{pmatrix}$$

and $X_0 = (1, 0, \dots, 0)^t$. It is well known, cf. [13], that

$$\frac{X_n}{n+1} \xrightarrow{a.s.} \frac{1}{H_m - 1} \left(\frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{m} \right)^t$$

as $n \to \infty$, where H_m denotes the m^{th} Harmonic number. This limit is deterministic and all its components are strictly positive, so we expect convergence in distribution to a non-mixed Gaussian distribution. The eigenvalues of R_m are given by the solutions z of the equation

$$m! = \prod_{k=1}^{m-1} (z + k).$$

If $m \leq 26$, there are no eigenvalues with real part greater than $1/2$. In this case, Theorem 1.2.5 confirms the well-known result that $(X_n - \mathbb{E}[X_n])/\sqrt{n} \to \mathcal{N}$ in distribution: Mahmoud and Pittel [56] show that when $m \leq 15$, the limiting distribution is normal. The result was later extended to include $m \leq 26$ by Lew and Mahmoud [53].
For $m > 26$, there is at least one eigenvalue with real part greater than $1/2$ and it is known that for all such m, there is no eigenvalue whose real part is equal to $1/2$. Chern and Hwang [15] prove that when $m \geq 27$, the space requirement centered by its mean and scaled by its standard deviation does not have a limiting distribution. Here, Theorem 1.2.5 can be

applied to give the result

$$\frac{1}{\sqrt{n}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k}\Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_V).$$

4.2 B-urns

A B-tree with integer-valued parameter $m \geq 2$ is a search tree, where keys are stored in internal nodes of the tree and its leaves (or gaps) represent future insertion possibilities. See [11] for an exposition of B-trees, where the authors present two algorithms to construct B-trees: the prudent algorithm and the optimistic algorithm. It turns out that the urn process obtained from considering the optimistic algorithm has a generating matrix that is more convenient. When constructed from the so-called optimistic algorithm (see [11]), there are $m-1$ different types of gaps. As explored in [11], the evolution of the gap process $(X_n)_{n\geq 0}$ under the random permutation model on the keys can be modelled as a Pólya urn with generating matrix

$$R_B = \begin{pmatrix} -m & & & & 2m \\ m+1 & -(m+1) & & & \\ & m+2 & & & \\ & & \ddots & & \\ & & & 2m-1 & -(2m-1) \end{pmatrix}.$$

The eigenvalues of R_B are given by the solutions of the equation

$$\frac{2m!}{m!} = \prod_{k=m}^{2m-1} (z+k).$$

Its complex roots are all simple, and 1 is the Perron-Frobenius eigenvalue. Furthermore, two distinct eigenvalues that have the same real part are conjugated. The left and right eigenvectors of R_B are also explicitly calculated in [11].

If $m \leq 59$, there are no eigenvalues with real part greater than $1/2$. In this case, Theorem 1.2.5 confirms that

$$\frac{X_n - \mathbb{E}[X_n]}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_V).$$

For $m \geq 60$, Theorem 1.2.5 states that

$$\frac{1}{\sqrt{n}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k}\Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_V).$$

In real life implementations, m is taken around 100 or more, and thus, the fluctuating term is of significance in applications. See [11] for pictures of this behaviour.

4.3 Congruence classes of depths in random BSTs

Janson [39] studies congruence classes of depths in three different random trees, among them the random binary search tree (BST). For a definition of the random BST, see section 2.3.2. Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with $\mathcal{L}(U_1) = \text{unif}(0,1)$ and $(T_n)_{n\in\mathbb{N}}$ be the sequence of BSTs constructed from the uniform keys. Furthermore, for fixed $q \in \mathbb{N}$, $q \geq 2$, and each $n \geq 1$, let $\tilde{X}_n$ be the vector of the numbers of internal nodes at depth $0, \dots, q-1 \mod q$ of T_n. With eigenvalues and eigenvectors as in chapter 2, (2.2) and (2.3), set

$$\Sigma_q = \sum_{j=2}^{q} \frac{1}{|3 - 4\Re(\lambda_j)|} v_j v_j^*.$$

Janson [39], Theorem 2.7, proves the following results about the asymptotic behaviour of $(\tilde{X}_n)_{n\in\mathbb{N}}$:

Theorem 4.3.1 (Janson). *(i) If* $2 \leq q \leq 8$, *then*

$$n^{-1/2}\left(\tilde{X}_n - \frac{n}{q}1\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_q) \qquad \textit{as } n \to \infty.$$

Furthermore, the moments of $\left(n^{-1/2}\left(\tilde{X}_n - \frac{n}{q}1\right)\right)_{n\geq 1}$ *converge to the moments of* $\mathcal{N}(0, \Sigma_q)$.

(ii) If $q \geq 9$, *let* $\alpha := 2\cos(2\pi/q) - 1 > 1/2$ *and* $\beta := 2\sin(2\pi/q)$. *Then*

$$n^{-1/2}\left(\tilde{X}_n - \frac{n}{q}1\right) - \Re(n^{i\beta}\tilde{W}) \xrightarrow{a.s.} 0 \qquad \textit{as } n \to \infty,$$

for some complex random vector $\tilde{W} = W v_2$, *where* W *is a complex-valued random variable and* v_2 *is defined in (2.3).*

In particular, the variance of each component $\tilde{X}_n^{(j)}$ is of order n for $q \leq 8$, but larger for $q \geq 9$. The above asymptotics already strongly resemble the asymptotics of cyclic urns from chapter 2. Indeed, Janson uses an urn model to derive the quoted results, which we will now present.

First, it turns out to be more convenient to work with external rather than with internal nodes, see section 2.3.2. Note that every internal vertex has two children, while every external vertex has none. Furthermore, let $\left(X_n^{(j)}\right)_{j=0}^{\infty}$ and $\left(V_n^{(j)}\right)_{j=0}^{\infty}$ denote the profiles of the internal and external nodes in T_n, respectively. That is, $X_n^{(j)}$ ($V_n^{(j)}$) denotes the number of internal (external) nodes at depth j in T_n. The internal and external nodes are related in the following way: As each internal node has exactly two children on the next level,

$$2X_n^{(j-1)} = X_n^{(j)} + V_n^{(j)}, \qquad j \geq 1.$$

Furthermore, $X_n^{(0)} + V_n^{(0)} = 1$. For congruence classes modulo q, this means that for every j,

$$2\tilde{X}_n^{(j-1)} - \tilde{X}_n^{(j)} = \tilde{V}_n^{(j)} - \delta_{j,0}. \tag{4.1}$$

Here, for $q \geq 2$ fixed, we used $\tilde{V}_n := \left(\tilde{V}_n^{(1)}, \dots, \tilde{V}_n^{(q)}\right)^t$ to denote the vector of external nodes in the random BST at depths $0, 1, \dots, q-1 \mod q$ after the insertion of n keys.
Now, the evolution of the congruence classes of depths of the external nodes can be viewed as the evolution of a generalised Pólya urn scheme. At each step, an external node is picked uniformly at random for replacement and turned into an internal node with two new external nodes at the next level attached. The growth of the external nodes can be thus be modelled by an urn scheme with generating matrix

$$R_{\mathrm{BST}} := \begin{pmatrix} -1 & 0 & 0 & \cdot & \cdot & 0 & 2 \\ 2 & -1 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 2 & -1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & 2 & -1 \end{pmatrix} \in \mathbb{R}^{q \times q} \tag{4.2}$$

and $\tilde{V}_0 = e_1$. Note that R_{BST} is also circulant like the cyclic urn matrix, and thus has the same left and right eigenvectors as (2.3) and eigenvalues

$$\lambda_1' := 1, \lambda_2' := 2\omega - 1, \lambda_3' := 2\omega^{-1} - 1, \dots, \lambda_q' := 2\omega^{\lceil \frac{q}{2} \rceil} - 1.$$

Hence, the condition $\Re(\lambda_2) < 1/2$ becomes $\cos(2\pi/q) < 3/4$, which holds for $q \leq 8$, while for $q \geq 9$ we have $\cos(2\pi/q) > 3/4$ and thus $\Re(\lambda_2) > 1/2$. Also note that for all $q \geq 2$ and $1 \leq k \leq q$, $\cos(2\pi k/q) \neq 3/4$, see [77]. Consequently, for all $q \geq 2$, there are no eigenvalues with real part equal to $1/2$.
As we have seen, Janson [39] obtains normal convergence of $\left(\tilde{V}_n\right)_{n \geq 0}$ when $q \leq 8$, and the covariance matrix of the limit is given by A_q. Here,

$$A_q = \sum_{j=2}^{q} \frac{|\lambda_j'|^2}{|1 - 2\Re(\lambda_j')|} v_j v_j^*.$$

For $q \geq 9$, there are oscillations. However, Theorem 1.2.5, case (i), implies that also

$$\frac{1}{\sqrt{n}} \left(\tilde{V}_n - \mathbb{E}[\tilde{V}_n] - \sum_{k=1}^{p} n^{\lambda_k} \Xi_k v_k \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_q).$$

Again, in this example, the covariance matrix is comparatively simple to compute.
Using relation (4.1), one arrives at

$$\frac{1}{\sqrt{n}} \left(\tilde{X}_n - \mathbb{E}[\tilde{X}_n] - \sum_{k=1}^{p} n^{\lambda_k} \Xi_k v_k \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_q).$$

4.4 Open questions

The central limit theorem analogues of the current thesis contribute to a more thorough understanding of the asymptotic properties of generalised Pólya-Eggenberger urn processes. Some directions for future research along the direction taken in the present work are given below.

To begin with, recall that the Gaussian limit in Theorem 1.2.5 is degenerate if the urn is reducible and the largest eigenvalue is simple. The two-colour case (Proposition 1.1.4) suggests that weak convergence to a (mixed) normal distribution still holds, if a different scaling is applied. From this, the natural question arises as to the correct scaling in this case.

Secondly, it might be of interest to determine the distribution of the random variables $\Gamma_{a+1}, \dots, \Gamma_{a+c}$ introduced below Theorem 1.2.3, if there is a simple characterisation. Their isolated counterparts $D^{(1)}, \dots, D^{(a+1)}$ constitute a Dirichlet distributed random vector, but as category 2 classes of colours experience influencing from outside, the situation is more involved in this case.

Thirdly, one could try to relax conditions (A1) and (A2). For example, it seems very plausible that condition (A1) can be removed by considering generalised eigenspaces rather than eigenspaces. The task of removing the balance condition (A2) is probably much harder, but nevertheless very interesting.

Furthermore, a functional version of Theorem 1.2.5 would be of great interest. For two-colour urns and small urns, results of this type are known and can be found in [29, 40].

Finally, as mentioned in Chapter 2, various other combinatorial structures related to algorithms exhibit periodic behaviour that resembles the oscillating nature of some urn models. Now the question is, if similar central limit theorems can be derived for these structures. The answer to this question is not immediate, as urn models possess some additional properties that cannot be drawn on in other situations. This is work in progress.

Bibliography

[1] D. Aldous, B. Flannery, and J. L. Palacios. Two applications of urn processes: the fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains. *Probability in the engineering and informational sciences*, 2(3):293–307, 1988.

[2] K. B. Athreya and S. Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *The Annals of Mathematical Statistics*, 39(6):1801–1817, 1968.

[3] K. B. Athreya and P. E. Ney. *Branching processes*. Springer Science & Business Media, 2012.

[4] A. Bagchi and A. K. Pal. Asymptotic normality in the generalized Pólya–Eggenberger urn model, with an application to computer data structures. *SIAM Journal on Algebraic Discrete Methods*, 6(3):394–405, 1985.

[5] R. Bayer. Binary B-trees for virtual memory. In *Proceedings of the 1971 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control*, pages 219–235. ACM, 1971.

[6] R. Bayer and E. McCreight. Organization and maintenance of large ordered indices. In *Proceedings of the 1970 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control*, pages 107–141. ACM, 1970.

[7] P. Bindjeme and J. A. Fill. Exact L^2-distance from the limit for QuickSort key comparisons. *arXiv:1201.6445*, 2012.

[8] A. Bose, A. Dasgupta, and K. Maulik. Multicolor urn models with reducible replacement matrices. *Bernoulli*, 15(1):279–295, 2009.

[9] D. Branson. An urn model and the coalescent in neutral infinite-alleles genetic processes. *Lecture Notes - Monograph Series*, pages 174–192, 1991.

[10] F. T. Bruss and C. A. O'Cinneide. On the maximum and its uniqueness for geometric random samples. *Journal of Applied Probability*, 27(03):598–610, 1990.

[11] B. Chauvin, D. Gardy, N. Pouyanne, and D.-H. Ton-That. B-urns. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 13:605–634, 2016.

[12] B. Chauvin, C. Mailler, and N. Pouyanne. Smoothing equations for large Pólya urns. *Journal of Theoretical Probability*, 28(3):923–957, 2015.

[13] B. Chauvin and N. Pouyanne. m-ary search trees when $\mathfrak{m} \geq 27$: A strong asymptotics for the space requirements. *Random Structures & Algorithms*, 24(2):133–154, 2004.

[14] H.-H. Chern, M. Fuchs, and H.-K. Hwang. Phase changes in random point quadtrees. *ACM Transactions on Algorithms (TALG)*, 3(2):40–91, 2007.

[15] H.-H. Chern and H.-K. Hwang. Phase changes in random m-ary search trees and generalized Quicksort. *Random Structures & Algorithms*, 19(3-4):316–358, 2001.

[16] H.-H. Chern and H.-K. Hwang. Transitional behaviors of the average cost of Quicksort with median-of-(2t+ 1). *Algorithmica*, 29(1-2):44–69, 2001.

[17] J. L. Doob. *Discrete potential theory and boundaries.* United States Air Force, Office of Scientific Research, 1958.

[18] Z. Drezner and N. Farnum. A generalized binomial distribution. *Communications in Statistics - Theory and Methods*, 22(11):3051–3063, 1993.

[19] M. Drmota, S. Janson, and R. Neininger. A functional limit theorem for the profile of search trees. *The Annals of Applied Probability*, 18(1):288–333, 2008.

[20] F. Eggenberger and G. Pólya. Über die Statistik verketteter Vorgänge. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 3(4):279–289, 1923.

[21] P. Ehrenfest and T. Ehrenfest. Über zwei bekannte Einwände gegen Boltzmanns H-Theorem. *Physikalische Zeitschrift*, 8:311–314, 1907.

[22] S. N. Evans, R. Grübel, and A. Wakolbinger. Trickle-down processes and their boundaries. *Electron. J. Probab*, 17(1):1–58, 2012.

[23] W. Feller. *An introduction to probability theory and its applications*, volume 1-2. John Wiley & Sons, 2008.

[24] P. Flajolet, P. Dumas, and V. Puyhaubert. Some exactly solvable models of urn process theory. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, pages 59–118. Discrete Mathematics and Theoretical Computer Science, 2006.

[25] D. A. Freedman. Bernard Friedman's urn. *The Annals of Mathematical Statistics*, pages 956–970, 1965.

[26] B. Friedman. A simple urn model. *Communications on Pure and Applied Mathematics*, 2(1):59–70, 1949.

[27] C. Goldschmidt and B. Haas. A line-breaking construction of the stable trees. *Electronic Journal of Probability*, 20(1):1–24, 2015.

[28] R. Gouet. A martingale approach to strong convergence in a generalized Pólya-Eggenberger urn model. *Statistics & Probability Letters*, 8(3):225–228, 1989.

[29] R. Gouet. Martingale functional central limit theorems for a generalized Pólya urn. *The Annals of Probability*, pages 1624–1639, 1993.

[30] R. Gouet. Strong convergence of proportions in a multicolor Pólya urn. *Journal of Applied Probability*, 34(02):426–435, 1997.

[31] R. Grübel. Persisting randomness in randomly growing discrete structures: graphs and search trees. *arXiv:1407.0808*, 2014.

[32] R. Grübel. Search trees: metric aspects and strong limit theorems. *The Annals of Applied Probability*, 24(3):1269–1297, 2014.

[33] R. Grübel and Z. Kabluchko. A functional central limit theorem for branching random walks, almost sure weak convergence and applications to random trees. *The Annals of Applied Probability*, 26(6):3659–3698, 2016.

[34] P. Hall and C. C. Heyde. *Martingale limit theory and its application.* Academic press, 2014.

[35] C. C. Heyde. Asymptotics and criticality for a correlated Bernoulli process. *Australian & New Zealand Journal of Statistics*, 46(1):53–57, 2004.

[36] C. Holmgren and S. Janson. Asymptotic distribution of two-protected nodes in ternary search trees. *arXiv:1403.5573*, 2014.

[37] C. Holmgren, S. Janson, and M. Šileikis. Multivariate normal limit laws for the numbers of fringe subtrees in $\mathfrak{m}$-ary search trees and preferential attachment trees. *arXiv:1603.08125*, 2016.

[38] F. M. Hoppe. The sampling theory of neutral alleles and an urn model in population genetics. *Journal of Mathematical Biology*, 25(2):123–159, 1987.

[39] S. Janson. Limit theorems for certain branching random walks on compact groups and homogeneous spaces. *The Annals of Probability*, pages 909–930, 1983.

[40] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications*, 110(2):177–245, 2004.

[41] S. Janson. Congruence properties of depths in some random trees. *arXiv:math/0509471*, 2005.

[42] S. Janson. Limit theorems for triangular urn schemes. *Probability Theory and Related Fields*, 134(3):417–452, 2006.

[43] S. Janson. Mean and variance of balanced Pólya urns. *arXiv:1602.06203*, 2016.

[44] S. Janson and R. Neininger. The size of random fragmentation trees. *Probability Theory and Related Fields*, 142(3-4):399–442, 2008.

[45] S. Janson and N. Pouyanne. Moment convergence of balanced Pólya processes. *arXiv:1606.07022*, 2016.

[46] N. L. Johnson and S. Kotz. *Urn models and their application; an approach to modern discrete probability theory.* New York (USA) Wiley, 1977.

[47] M. Knape and R. Neininger. Pólya urns via the contraction method. *Combinatorics, Probability and Computing*, 23(6):1148–1186, 2014.

[48] S. Kotz, H. Mahmoud, and P. Robert. On generalized Pólya urn models. *Statistics & Probability Letters*, 49(2):163–173, 2000.

[49] M. Kuba and H. Sulzbach. On martingale tail sums in affine two-color urn models with multiple drawings. *Journal of Applied Probability*, 54(1):96–117, 2017.

[50] S. Laruelle and G. Pagès. Nonlinear randomized urn models: a stochastic approximation viewpoint. *arXiv:1311.7367*, 2013.

[51] K. Leckey. On densities for solutions to stochastic fixed point equations. *arXiv:1604.05787*, 2016.

[52] B. Levin and H. Robbins. Urn models for regression analysis, with applications to employment discrimination studies. *Law and Contemporary Problems*, 46(4):247–267, 1983.

[53] W. Lew and H. M. Mahmoud. The joint distribution of elastic buckets in multiway search trees. *SIAM Journal on Computing*, 23(5):1050–1074, 1994.

[54] H. M. Mahmoud. *Evolution of random search trees.* New York, Wiley, 1992.

[55] H. M. Mahmoud. *Pólya urn models.* CRC press, 2008.

[56] H. M. Mahmoud and B. Pittel. Analysis of the space of search trees under the random insertion algorithm. *Journal of Algorithms*, 10(1):52–75, 1989.

[57] C. Mailler. Describing the asymptotic behaviour of multicolour Pólya urns via smoothing systems analysis. *arXiv:1407.2879*, 2014.

[58] N. S. Müller. Central limit theorem analogues for multicolour urn models. *arXiv preprint 1604.02964*, 2016.

[59] N. S. Müller and R. Neininger. The CLT analogue for cyclic urns. In *2016 Proceedings of the Thirteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 121–127. SIAM, 2016.

[60] R. Neininger. *Stochastische Analyse von Algorithmen, Fixpunktgleichungen und ideale Metriken.* Habilitationsschrift, University of Frankfurt, 2004.

[61] R. Neininger. Refined Quicksort asymptotics. *Random Structures & Algorithms*, 46(2):346–361, 2015.

[62] R. Neininger and L. Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability*, 14(1):378–418, 2004.

[63] R. Neininger and L. Rüschendorf. On the contraction method with degenerate limit equation. *The Annals of Probability*, 32(3B):2838–2856, 2004.

[64] R. Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, 2007.

[65] G. Pólya. Sur quelques points de la théorie des probabilités. In *Annales de l'institut Henri Poincaré*, volume 1, pages 117–161, 1930.

[66] N. Pouyanne. Classification of large Pólya-Eggenberger urns with regard to their asymptotics. *Discrete Mathematics and Theoretical Computer Science*, pages 177–245, 2005.

[67] N. Pouyanne. An algebraic approach to Pólya processes. In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 293–323, 2008.

[68] S. T. Rachev and L. Rüschendorf. Probability metrics and recursive algorithms. *Advances in Applied Probability*, 27(03):770–799, 1995.

[69] A. Rarivoarimanana. *Unbalanced Urn Models and Applications.* PhD thesis, 2014.

[70] R. Remmert. *Funktionentheorie 2.* Springer-Verlag Berlin Heidelberg New York, 1995.

[71] U. Rösler. A limit theorem for "Quicksort". *Informatique théorique et applications*, 25(1):85–100, 1991.

[72] U. Rösler. A fixed point theorem for distributions. *Stochastic Processes and their Applications*, 42(2):195–214, 1992.

[73] E. Seneta. *Non-negative matrices and Markov chains.* Springer Science & Business Media, 2006.

[74] R. T. Smythe. Central limit theorems for urn models. *Stochastic Processes and their Applications*, 65(1):115–137, 1996.

[75] H. Sulzbach. On martingale tail sums for the path length in random trees. *Random Structures & Algorithms*, 50(3):493–508, 2017.

[76] R. van der Hofstad. Random graphs and complex networks. *Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf*, 2009.

[77] J. Varona. Rational values of the arccosine function. *Open Mathematics*, 4(2):319–322, 2006.

[78] L. J. Wei and J. M. Lachin. Properties of the urn randomization in clinical trials. *Controlled clinical trials*, 9(4):345–364, 1988.

[79] W. Woess. *Denumerable Markov Chains.* EMS textbooks in mathematics, 2009.

Appendix

The complex Gamma function

The asymptotic growth of the complex Gamma function is used several times without further comment in the course of this thesis. Here, we collect some important properties, taken from [70].

For big values of $|z|$, one often wishes to approximate $\Gamma(z)$ via simpler functions, at least in the slotted plane $\mathbb{C}^- := \mathbb{C} \setminus (-\infty, 0]$. Here, the restriction to $\mathbb{C}^-$ has to be made, as the Gamma function has poles in $\mathbb{Z}_-$. More precisely, for each $\delta \in (0, \pi]$, we set $W_\delta := \{z = |z|e^{i\phi} \in \mathbb{C} \setminus \{0\} : |\phi| \leq \pi - \delta\}$. An approximation of $\Gamma(z)$ as well as the size of the "error term" $\mu(z)$ are given in the following theorem.

Theorem 4.4.1 (Stirling's formula)**.**

$$\begin{aligned}
\Gamma(z) &= \sqrt{2\pi} z^{z-\frac{1}{2}} e^{-z} e^{\mu(z)}, && z \in \mathbb{C}^-,\\
|\mu(z)| &\leq \frac{1}{8\cos^2 \frac{1}{2}\phi} \cdot \frac{1}{|z|}, && z = |z|e^{i\phi} \in \mathbb{C}^-,\\
|\mu(z)| &\leq \frac{1}{8\sin^2 \frac{1}{2}\delta} \cdot \frac{1}{|z|}, && z \in W_\delta,\ 0 < \delta \leq \pi.
\end{aligned}$$

The strength of the theorem is in the bounds for $\mu(z)$. However, it is often sufficient to know that in every W_δ, $\mu(z)$ tends to zero uniformly as $1/z$, for $|z| \to \infty$. From Stirling's formula, we have the asymptotic expansion

$$\Gamma(z+1) \sim \sqrt{2\pi z}\left(\frac{z}{e}\right)^z.$$

Here, $\sim$ means that the quotient of the left and the right hand side tends to one uniformly in every space W_δ as $z \to \infty$. A consequence that we have used throughout this thesis is, that

$$\Gamma(z+a) \sim z^a \Gamma(z), \qquad \text{for} \quad a \in \mathbb{C} \setminus \mathbb{Z}_- \ \text{ fixed.}$$

One can further show that

$$|\mu(z)| \leq \frac{1}{12}\frac{1}{\Re(z)}, \quad \Re(z) > 0, \quad \text{and} \quad |\mu(iy)| \leq \frac{1}{6}\frac{1}{|y|}, \quad y \in \mathbb{R}.$$

Moreover, in the language of this thesis, set

$$\gamma_{\ell,n}(z) := \prod_{\ell \leq k < n}\left(1 + \frac{z}{rn + |X_0|}\right) = \frac{\Gamma(n + |X_0|/r + z/r)}{\Gamma(n + |X_0|/r)} \frac{\Gamma(\ell + |X_0|/r)}{\Gamma(\ell + |X_0|/r + z/r)}.$$

Then Janson [43], Lemma 5.2, gives the following result.

Lemma 4.4.1. *Assume that* $r = 1$.

(i) For every fixed ℓ, *as* $n \to \infty$,

$$\gamma_{\ell,n}(z) = n^z \frac{\Gamma(\ell + |X_0|)}{\Gamma(\ell + |X_0| + z)}(1 + o(1))$$

uniformly for z *in any fixed compact set in the complex plane.*

(ii) As $\ell, n \to \infty$ *with* $\ell \leq n$,

$$\gamma_{\ell,n}(z) = \left(\frac{n}{\ell}\right)^z (1 + o(1))$$

uniformly for z *in any fixed compact set in the complex plane.*

Zusammenfassung

Gegenstand der vorliegenden Arbeit ist die Herleitung eines verallgemeinerten Typs von zentralem Grenzwertsatz für Urnenprozesse. Bei Urnen handelt es sich um einfache Modelle für zufällige Wachstumsprozesse, die über die Interaktion von Elementen verschiedener Typen gesteuert sind. Ein naheliegendes Studienobjekt, das auch im Mittelpunkt dieser Arbeit steht, ist die asymptotische Verteilung der Elemente auf die Typen. Diese hängt von der konkreten Form der Interaktion ab und vor allem davon, inwieweit sich selbstverstärkende oder nivellierende Effekte zwischen den Typen auf lange Sicht hin durchsetzen. Für manche Modelle ist die gegenseitige Einflussnahme gering und führt zu einem asymptotischen Verhalten, das sich seiner Natur nach nur geringfügig von dem einer Summe unabhängiger und identisch verteilter Zufallsvariablen unterscheidet. In anderen Modellen hingegen verstärken sich anfängliche Ungleichheiten hin zu Tendenzen signifikanter Größe (auf der Skala eines zentralen Grenzwertsatzes). Genauer gesagt, stabilisieren sich bestimmte Typenverhältnisse bei einem zufälligen Wert oder es entstehen logarithmisch periodische Schwankungen zufälliger Amplitude und Phase zwischen Typen. Der Schwerpunkt der vorliegenden Arbeit liegt auf den letztgenannten Modellen. Für Urnen dieser Art wird ein zentraler Grenzwertsatz hergeleitet, dessen Form sich durch die vorgenommene Normierung von der klassischen Situation unterscheidet. Die Abweichung liegt darin begründet, dass im Fall fast sicherer Tendenzen eine zufällige Zentrierung, sowie möglicherweise auch eine zufällige Skalierung, notwendig für die Konvergenz sind. In der Arbeit werden die Resultate, die dieses Vorgehen liefert, auch als *Analoga* zentraler Grenzwertsätze bezeichnet.

Urnenmodelle. Anschaulich gesprochen, handelt es sich bei einem Urnenmodell um ein gedachtes Behältnis unendlicher Kapazität – die Urne –, in der sich Kugeln verschiedener Typen befinden. Im Verlauf einer zeitlichen Entwicklung werden zufällig einzelne Kugeln aus der Urne gezogen und, in Abhängigkeit vom gezogenen Typ, andere entnommen oder hinzugefügt, sodass sich die Zusammensetzung der Urne stets ändert.
Die Untersuchung sogenannter verallgemeinerter Pólya-Eggenbergerscher Urnenmodelle nahm ihren Ausgangspunkt in der 1923 veröffentlichten Arbeit [20], die ein Zweitypen-Modell zur Chancenvermehrung durch Erfolg der Typen im Lichte verschiedener möglicher Anwendungen vorstellte. Seit der Arbeit von Pólya und Eggenberger wurde eine Vielzahl an Verallgemeinerungen und Varianten des ursprünglichen Modells untersucht, etwa Urnen mit mehr als zwei Typen, Urnen, bei denen mehrere Bälle auf einmal gezogen werden, oder Urnen, bei denen zusätzlich zu den zufälligen Zügen noch weitere Stufen von Randomisierung auftreten.
Das der vorliegenden Arbeit zugrunde liegende Modell ist das folgende: Gegeben seien eine natürliche Zahl $q \geq 2$, ein Spaltenvektor

$$X_0 = \left(X_0^{(1)}, \ldots, X_0^{(q)}\right)^t \quad \in \mathbb{N}_0^q$$

und eine $q \times q$-Matrix $R \in \mathbb{Z}^{q \times q}$ mit ganzzahligen Einträgen. Wir interpretieren q als die Anzahl verschiedener Typen, X_0 als die anfängliche Zusammensetzung der Urne und R als Schema für die Regeln, nach denen Kugeln zurückgelegt werden.
Aus den obigen Daten ergibt sich dann wie folgt der Urnenprozess $(X_n)_{n \in \mathbb{N}_0}$, wobei

$$X_n = \left(X_n^{(1)}, \ldots, X_n^{(q)}\right)^t$$

den Vektor der Kugelzahlen der verschiedenen Typen zur Zeit $n \in \mathbb{N}_0$ bezeichnet: Zur Zeit 0 befinden sich $X_0^{(i)}$ Kugeln des Typs i in der Urne, für $i = 1, \dots, q$. Unmittelbar vor Zeit $n+1$ wird uniform eine Kugel aus der Urne gezogen. Hat sie Typ i, so wird sie zurückgelegt und es werden für jedes $j \in \{1, \dots, q\}$ entweder $R_{j,i}$ Kugeln des Typs j hinzugefügt ($R_{j,i} \geq 0$) oder $R_{j,i}$ Kugeln des Typs j entfernt ($R_{j,i} < 0$). Bei dem hier untersuchten Pólya-Eggenbergerschen Urnenmodell zu den Daten X_0 und R handelt es sich also um einen Markov-Prozess $(X_n)_{n \in \mathbb{N}_0}$ in diskreter Zeit, dessen mögliche Inkremente die Spalten von R sind. Die für R und X_0 getroffenen Annahmen in der vorliegenden Arbeit sind:

(A1) R hat konstante Spaltensumme $r \geq 1$.

(A2) Für alle $i \neq j$ ist $R_{i,j} \geq 0$ und falls $R_{i,i} < 0$, dann werden $X_0^{(i)}$ und $R_{i,j}$ für alle $1 \leq j \leq q$ von $|R_{i,i}|$ geteilt.

(A3) R ist diagonalisierbar über $\mathbb{C}$.

(A4) Keine zwei Spalten von R stimmen überein.

(A5) Die anfängliche Komposition der Urne ist so gewählt, dass es zu jedem Typ $j \in \{1, \dots, q\}$ ein $n \in \mathbb{N}_0$ mit $\mathbb{P}\left(X_n^{(j)} > 0\right) > 0$ gibt.

Eigenschaft (A1) garantiert, dass die Gesamtanzahl an Kugeln in jedem Zeitschritt, unabhängig vom gezogenen Typ, um denselben positiven Betrag r zunimmt. Eigenschaft (A2) hingegen stellt sicher, dass der Prozess zu jeder Zeit wohldefiniert ist und es nicht dazu kommt, dass die Entfernung von Kugeln aus der Urne gefordert wird, die gar nicht darin enthalten sind. Matrizen mit Eigenschaft (A2) werden auch Metzler-Leontief-Matrizen genannt. (A3) erleichtert die Rechnungen. Um das Vorkommen gänzlich redundanter Typen auszuschließen, nehmen wir zudem in (A4) an, dass keine zwei Zeilen von R übereinstimmen, und in (A5), dass die Anfangszusammensetzung der Urne nicht von vornherein das Auftreten bestimmter Typen verhindert.

Ein letzter Begriff zur Unterscheidung verschiedener Urnenmodelle ist der der *Irreduzibilität*: Eine Urne wird als irreduzibel bezeichnet, wenn für jeden Typ $i \in \{1, \dots, q\}$ gilt, dass bei Start des Prozesses mit einer einzigen Kugel des Typs i auch jeder andere Typ $j \in \{1, \dots, q\}$ mit positiver Wahrscheinlichkeit mindestens einmal im Verlauf des Prozesses in der Urne erscheint. Bei den irreduziblen Modellen handelt es sich um eine wichtige und vieluntersuchte Klasse von Urnen. In der vorliegenden Arbeit ist die Annahme der Irreduzibilität *nicht* getroffen. Allerdings zerfällt jede reduzible Urne in verschiedene irreduzible Klassen, weswegen wir den Begriff an dieser Stelle erläutern. Weiter nennen wir eine solche Klasse und ihre Elemente *dominant*, falls eine Anfangskonfiguration der Urne aus Typen ausschließlich dieser Klasse dazu führt, dass im Prozess niemals Typen anderer Klassen erscheinen. Anschaulich gesprochen kann man sagen, dass diese Klassen keine „Masse“ an andere Klassen abgeben. Für eine formale Definition verweisen wir hier auf den Haupttext, Kapitel 1.2.

Mithilfe dieser Definitionen und Annahmen lässt sich nun das asymptotische Verhalten der Vektoren $(X_n)_{n \geq 0}$ für $n \to \infty$ untersuchen und damit die Konvergenz von Anteilen, die Existenz von Oszillationen oder gar zentraler Grenzwertsätze nachweisen. Dabei ist seit den 1960er Jahren aus [2] bekannt, dass die Asymptotik des Prozesses unmittelbar mit dem Spektrum der Matrix R zusammenhängt. Die Annahmen (A1) und (A2) implizieren, dass die Spaltensumme r von R stets der Eigenwert mit dem größten Realteil ist. Dabei ist nicht ausgeschlossen, dass es sich bei r um einen mehrfachen Eigenwert handelt. Alle anderen Eigenwerte haben einen strikt kleineren Realteil. Nun ist aus [30] ebenfalls bekannt, dass die Anteile fast sicher konvergieren, es aber von der geometrischen Vielfachheit des Eigenwerts r abhängt, ob der Grenzwert deterministisch oder zufällig ist. Allgemeiner haben wir die Konvergenz der Anteile

$$\frac{X_n}{rn + |X_0|} \xrightarrow{\text{f.s.}} V, \qquad n \to \infty,$$

vgl. Satz 1.2.3 und die folgenden Bemerkungen. V ist hier ein zufälliger Vektor mit Werten in $\mathbb{R}^q$.
Ganz ähnlich weiß man im Falle sogenannter irreduzibler Urnen, dass die Existenz eines zentralen Grenzwertsatzes davon abhängt, ob es Eigenwerte gibt, deren Realteile den Schwellenwert $r/2$ übersteigen. Urnen mit solchen „großen" Eigenwerten werden *groß* genannt. Andere Urnen werden dementsprechend oft als *klein* bezeichnet. Im Falle irreduzibler strikt *kleiner* Urnen (es gibt keine Eigenwerte, deren Realteil genau $r/2$ beträgt) ist wohlbekannt [40, 74], dass es einen Vektor $v_1 \in \mathbb{R}^q$ gibt, sodass

$$\frac{X_n - nv_1}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma), \qquad n \to \infty,$$

für eine geeignete Kovarianzmatrix Σ. Genauso ist im Falle der Existenz von Eigenwerten mit Realteilen der Größe $r/2$, aber keiner größeren, bei irreduziblen Matrizen bekannt, dass

$$\frac{X_n - nv_1}{\sqrt{n \log(n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma), \qquad n \to \infty,$$

wobei Σ typischerweise niedrigeren Rang als im vorigen Ergebnis aufweist.
Für große irreduzible Urnenmodelle oder reduzible Modelle mit mehrfachem Eigenwert r hingegen besitzt der Prozess eine sich selbst verstärkende Dynamik innerhalb bestimmter Farbkombinationen. Diese führt über die Zufälligkeit der ersten Züge zu nicht-deterministischen Tendenzen von größerer Ordnung als $\sqrt{n}$, welche einer „klassischen" Normierung im Rahmen eines zentralen Grenzwertsatzes entgegenstehen. Außer im Spezialfall der beispielhaften Pólya Urne liegen bislang keine Aussagen zur Fluktuation um die fast sicheren Terme vor. Diese Lücke wird in der vorliegenden Arbeit durch die Herleitung eines zentralen Grenzwertsatzes für reduzible oder große irreduzible Urnenmodelle angegangen. Dabei ist in beiden Fällen eine zufällige Zentrierung vonnöten, womit sich für reduzible Prozesse Konvergenz gegen eine gemischte Gaußsche Verteilung ergibt. Zur Formulierung des Hauptresultats seien mit $\lambda_1, \dots, \lambda_q$ die Eigenwerte von R bezeichnet, geordnet nach absteigendem Realteil und bei gleichem Realteil nach absteigendem Imaginärteil. Unter den genannten Annahmen und nach geeigneter Wahl rechter Eigenvektoren $v_1, \dots, v_q$ wird in der Arbeit das folgende Ergebnis hergeleitet, siehe auch Satz 1.2.5 im Haupttext:

1. Gibt es in keiner dominanten Klasse einen Eigenwert, dessen Realteil genau $r/2$ beträgt, so existieren $p \geq 1$ und komplex-wertige Zufallsvariablen $\Xi_1, \dots, \Xi_p$ mit Erwartung 0, sodass

$$\frac{1}{\sqrt{n}} \left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\frac{\lambda_k}{r}} \Xi_k v_k \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_V)$$

für $n \to \infty$, wobei $\mathcal{N}$ eine nichtdegenerierte, multivariate Gaußsche Mischverteilung mit Mischkomponenten $V^{(1)}, \ldots, V^{(q)}$ und Kovarianzmatrix A_V bezeichnet. Zudem ist $(A_V)_{i,i}$ genau dann positiv, wenn i ein dominanter Typ ist.

2. Gibt es eine dominante Klasse, zu der ein Eigenwert λ_k mit $\Re(\lambda_k) = r/2$ gehört, so existieren $p \geq 1$ und komplex-wertige Zufallsvariablen $\Xi_1, \ldots, \Xi_p$ mit Erwartung 0, sodass

$$\frac{1}{\sqrt{n \log(n)}} \left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\frac{\lambda_k}{r}} \Xi_k v_k \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_V)$$

für $n \to \infty$, wobei $\mathcal{N}$ eine nichtdegenerierte, multivariate Gaußsche Mischverteilung mit Mischkomponenten $V^{(1)}, \ldots, V^{(q)}$ und Kovarianzmatrix A_V bezeichnet. Zudem ist $(A_V)_{i,i}$ genau dann positiv, wenn i ein dominanter Typ ist, die zu einer der dominanten Klassen gehört, in denen es einen Eigenwert mit Realteil $r/2$ gibt.

Die genaue Gestalt der Kovarianzmatrix A_V ist in Kapitel 3 angegeben. Im Falle einer reduziblen Matrix mit einfachem Eigenwert r ist die Matrix A_V allerdings degeneriert und Satz 1.2.5 liefert ein triviales Ergebnis. Hier liegt es nahe, dass schwächer als $\sqrt{n}$ skaliert werden muss, da es nur einen dominanten Typ gibt und alle anderen Typen zu selten gezogen werden, um signifikante Beiträge zur Fluktuation zu liefern. Ein Ergebnis für diesen speziellen Fall herzuleiten, wurde in dieser Arbeit nicht versucht, mit Ausnahme der leichteren Zweitypen-Situation in Proposition 1.1.4.

Der Beweis von Satz 1.2.5 beruht auf einer spektralen Zerlegung des Urnenprozesses und Martingalmethoden. Ein weiteres Ergebnis der vorliegenden Arbeit ist die Ausarbeitung eines ergänzenden Zugangs zur Problemstellung, der Kontraktionsmethode. Dies geschieht in Kapitel 2 anhand des Beispiels der zyklischen Urne. Wir gehen nun kurz auf dieses Beispiel ein, das aus gemeinsamer Arbeit mit Herrn Prof. Dr. Neininger hervorgegangen ist.
Bei der zyklischen Urne handelt es sich um ein verallgemeinertes Pólya-Eggenbergersches Urnenmodell mit irreduzibler, zirkulanter Matrix

$$R := \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 1 \\ 1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & 1 & 0 \end{pmatrix} \in \mathbb{R}^{q \times q}.$$

Der Einfachheit halber wird in diesem Beispiel durchgehend angenommen, dass wir mit genau einer Kugel des ersten Typs starten. Mit wachsender Anzahl an Typen unterliegt die zyklische Urne einem Phasenübergang, anhand dessen sich die oben beschriebenen verschiedenen Phänomene gut erläutern lassen. Zunächst ist wohlbekannt, dass für bis zu fünf Typen ein klassischer zentraler Grenzwertsatz

$$\frac{X_n - n v_1}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma), \qquad n \to \infty,$$

mit explizit berechneter Kovarianzmatrix Σ vom Rang $q-1$ existiert, $q \in \{2, \dots, 5\}$. Für sechs Typen gilt immerhin noch

$$\frac{X_n - n v_1}{\sqrt{n \log(n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma), \qquad n \to \infty,$$

wobei Σ nur Rang zwei hat. Ab sieben Typen hingegen treten plötzlich fast sichere Oszillationen auf, die das Geschehen in zweiter Ordnung dominieren. Sei dazu $\omega := \exp(2\pi i/q) =: \sigma_2 + i\mu_2$ die q-te Einheitswurzel, zerlegt in Real- und Imaginärteil. Dann existieren für jedes $q \geq 7$ eine komplex-wertige Zufallsvariable Ξ_2, die von q und der Anfangskomposition der Urne abhängt, und determinstische Vektoren $v_1, v_2 \in \mathbb{R}^q$, sodass für $n \to \infty$

$$\frac{X_n - n v_1}{n^{\sigma_2}} - 2\Re\left(n^{i\mu_2}\left(\Xi_2 + \frac{1}{\Gamma(1+\omega)}\right) v_2\right) \xrightarrow{\text{f.s.}} 0.$$

Insbesondere existieren für jedes $q \geq 7$ unendlich viele Teilfolgen $(n_m)_{m \geq 1}$, sodass

$$\frac{X_{n_m} - n_m v_1}{\sqrt{\mathrm{Var}(X_{n_m})}}$$

mit $m \to \infty$ in Verteilung gegen verschiedene Grenzwerte konvergiert.
Es stellt sich heraus, dass bei der Formulierung eines zentralen Grenzwertsatzes für die zyklische Urne zwischen zwei Fällen unterschieden werden muss, von denen die Zusammensetzung des Spektrums von R abhängt. Dies sind die Fälle $6 \mid q$ und $6 \nmid q$. Damit erhalten wir folgende konkreteren Resultate:

1. Ist $q \geq 2$ und $6 \nmid q$, so setzen wir $p := 2\lfloor (q-1)/6 \rfloor$. In diesem Fall existieren komplex-wertige Zufallsvariablen $\Xi_1, \dots, \Xi_p$ mit Erwartung 0, sodass für $n \to \infty$

$$\frac{1}{\sqrt{n}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k} \Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma^{(q)}\right).$$

   Die Kovarianzmatrix $\Sigma^{(q)}$ hat Rang $q-1$ und ist gegeben durch

$$\Sigma^{(q)} := \sum_{k=2}^{q} \frac{1}{|2\sigma_k - 1|} v_k v_k^*,$$

   wobei die Eigenwerte $\lambda_1, \dots, \lambda_q$ und Eigenvektoren $v_1, \dots, v_q$ wie in Kapitel 2 definiert sind.

2. Für $6 \mid q$ erfordert die Normierung einen zusätzlichen $\sqrt{\log n}$ Faktor und der Rang der Kovarianzmatrix wird auf 2 reduziert: Wir setzen wieder $p := 2\lfloor (q-1)/6 \rfloor$. Es existieren komplex-wertige Zufallsvariablen $\Xi_1, \dots, \Xi_p$ mit Erwartung 0, sodass für $n \to \infty$

$$\frac{1}{\sqrt{n \log n}}\left(X_n - \mathbb{E}[X_n] - \sum_{k=1}^{p} n^{\lambda_k} \Xi_k v_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma^{(q)}\right).$$

Die Kovarianzmatrix ist gegeben durch

$$\Sigma^{(q)} := \nu_{q/3}\nu^*_{q/3} + \nu_{q/3+1}\nu^*_{q/3+1}.$$

Als ein weiteres Ergebnis von Abschnitt 2 ergibt sich eine explizite Konstruktion der zentrierenden Zufallsvariablen $\Xi_1, \ldots, \Xi_p$ aus dem oben genannten Resultat über eine Folge unabhängiger, identisch uniform auf $(0, 1)$ verteilter Zufallsvariablen $(U_i)_{i\geq 1}$. Dazu wird eine Einbettung der zyklischen Urne in den zufälligen Binärsuchbaum mit Eingabe $(U_i)_{i\geq 1}$ vorgenommen. Betrachtet man den Grenzwert des zufälligen Binärsuchbaums in seiner Doob-Martin-Kompaktifizierung, so erhalten wir zudem, dass die konstruierten Zufallsvariablen $\Xi_1, \ldots, \Xi_p$ deterministische Funktionen des Doob-Martin-Grenzwertes sind.

Ausblick. Interessante Fragestellungen, die sich aus der vorliegenden Arbeit ergeben, sind beispielsweise die folgenden: Welche Skalierung ist im Fall reduzibler Urnen mit einfachem dominantem Eigenwert r anzuwenden, um einen zentralen Grenzwertsatz herzuleiten? Lässt sich das Hauptresultat 1.2.5 zu einem funktionalen Grenzwertsatz erweitern? In der Arbeit wurde ebenfalls erwähnt, dass sich ähnliche periodische Phänomene wie im Beispiel der zyklischen Urne auch in anderen diskreten, algorithmisch motivierten Familien wiederfinden, in denen keine Martingalstrukturen bekannt sind. Die Herleitung analoger Resultate für einige dieser Strukturen befindet sich derzeit in Arbeit.