ON ASYMPTOTICS OF TWO NON-UNIFORM RECURSIVE TREE MODELS

by

Ella Veronika Hiesmayr

B.S., Mathematics, University of Vienna, 2013

Submitted to the Institute for Graduate Studies in

Science and Engineering in partial fulfillment of

the requirements for the degree of

Master of Science

Graduate Program in Mathematics

Boğaziçi University

2017

ACKNOWLEDGEMENTS

First of all I want to thank my supervisor Ümit Işlak. I feel extremely lucky to have a supervisor that is passionate about his research as well as excited to transfer this enthusiasm to his students. What and how he taught me in the comparatively short time we have known each other changed my view on mathematics and probability.

I also want to thank Serdar Altok for introducing me to random graphs and sparking my interest in probability theory.

Moreover I want to thank all my professors who supported my wish to understand in the course of my mathematics studies. I especially want to thank Fatih Ecevit and Mine Çağlar for being part of my thesis commitee.

I also want to thank my fellow mathematics students, in particular Arda, Beyza, Can, Çınar, Deniz, Doğa, Eray, Erol, Gözde, Mert and Turan for the mathematical as well fun atmosphere they created during the last weeks of my writing.

I thank Anna, Carmen, Ece and Emre for their friendship and for providing me with a different perspective when I struggled during my studies.

I also want to thank Efe for his support and understanding and for doing everything he could to provide me with a good working environment.

I am thankful that Judith and Fritz, my sister and brother, always keep my interest in different areas of mathematics going and for diverting me when things did not go as expected.

Finally I want to thank my parents, Hildegard and Michael, for their support during my studies and for trusting in my decisions.

ABSTRACT

ON ASYMPTOTICS OF TWO NON-UNIFORM RECURSIVE TREE MODELS

In this thesis the properties of two kinds of non-uniform random recursive trees are studied. In the first model weights are assigned to each node, thus altering the attachment probabilities. We will call these trees weighted recursive trees. In the second model a different distribution rather than the uniform one is chosen on the symmetric group, namely a riffle shuffle distribution. These trees will be called biased recursive trees. For both of these models the number of branches, the number of leaves, the depth of nodes and some other properties are studied. The focus is on asymptotic results and the comparison with uniform random recursive trees. It will be shown that the studied properties of weighted recursive trees are close to uniform recursive trees in many cases when the number of nodes increases. In contrast biased recursive trees show a different behaviour but approach uniform recursive trees depending on the parameters of the riffle shuffle distribution.

ÖZET

İKİ FARKLI DÜZGÜN DAĞILIMA SAHİP OLMAYAN YİNELİ AĞAÇ MODELİ ÜZERİNE

Bu tezde iki çeşit düzgün dağılıma sahip olmayan yineli ağaç modelinin özellikleri incelenmektedir. İlk modelde her köşeye ağırlık vermek suretiyle bağlanma olasıklıkları değişkenlik göstermektedir. İkinci modelde ise altta yatan düzgün permütasyon dağılımı özel bir kart karma modeli ile değiştirilerek yine düzgün dağılmayan bir ağaç tipi oluşturulmaktadır. Her iki modelde de dalların sayısı, yaprakların sayısı, köşelerin derinliği gibi pek çok özellik incelenecektir. Odak noktamız asimptotik sorular ve oluşan ağaçların düzgün yineli ağaçlar ile kıyaslanması üzerinedir. Köşelerin sayısı arttıkça ağırlıklı yineli ağaçların özelliklerinin düzgün dağılmış ağaçlara benzediği gösterilecektir. Benzer şekilde, ikinci modelimizde de altta yatan parametrelere bağlı olarak ortaya çıkan dağılım düzgün dağılıma yakın olabilmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\mathbf{1}(A)$	Indicator function of the event A
$\|A\|$	Cardinality of a set A
$\mathcal{B}(\mathbb{R})$	Borel sigma algebra of $\mathbb{R}$
$\mathrm{Cov}(X, Y)$	Covariance of two random variables X and Y
$\to_d$	Convergence in distribution
$=_d$	Equality in distribution
d_{TV}	Total variation metric
d_K	Kolmogorov metric
d_W	Wasserstein metric
$\mathbb{E}[X]$	Expectation of a random variable X
$\mathcal{G}$	Standard normal distribution
H_n	Harmonic numbers of the first order: $\sum_{i=1}^{n} \frac{1}{i}$
$H_n^{(2)}$	Harmonic numbers of the second order: $\sum_{i=1}^{n} \frac{1}{i^2}$
$\mathrm{mult}(a, \vec{p})$	Multinomial distribution with parameters a and $\vec{p} = (p_1, p_2, \ldots, p_a)$
$[n]$	First n positive integers: $\{1, \ldots, n\}$
$\mathcal{O}$	Big O
$\mathbb{P}(A)$	Probability of an event A
$\mathrm{Po}(\lambda)$	Poisson distribution with expectation λ
$\mathrm{Var}(X)$	Variance of a random variable X

LIST OF ACRONYMS/ABBREVIATIONS

a-RT	a-recursive tree
a.s.	almost surely
BRT	Biased recursive tree
URT	Uniform recursive tree
URP	Uniform random permutation
WRT	Weighted recursive tree

1. INTRODUCTION

Recursive trees are rooted labeled trees where the nodes on a path from the root to any other node form an increasing sequence. Because of this property, many recursive trees can be considered to grow dynamically, by attaching every new node to one of the already present nodes. Several distributions can be applied on the set of recursive trees of size n, the most common one being the uniform distribution, i.e. every recursive tree being equally likely. Given this distribution the number of leaves, internodal distances, the number of branches, the height of the tree and various other statistics are well studied.

In this thesis we will study two non-uniform distributions on recursive trees. Our main motivation is to get a better understanding of general random recursive tree models. More precisely the goal is to grasp the behaviour of inhomogeneous recursive trees, which are recursive trees that can be constructed by attaching at every step the new node n to one node of a recursive tree of size $n-1$ according to some distribution on $[n-1]$ and independently of the structure of the tree. Our first model is a special kind of inhomogeneous tree, namely the distribution obtained when each node is assigned a fixed weight. The second model is not an inhomogeneous tree itself but provides an approximation for inhomogeneous trees as it asymptotically approaches the uniform model.

Recursive trees are interesting from a theoretical as well as a practical point of view. First of all recursive trees are connected to a wide range of other mathematical structures like permutations, branching processes and records. Because this diversity is well reflected in the history of the study uniform recursive trees, we will now give a short summary of that development. After that we will give some examples of applications of uniform recursive trees that illustrate well how profitable knowledge of more general recursive tree models is from a practical perspective. For the results that were established until the early 90's the reader is referred to the comprehensive survey [1]. Unless otherwise mentioned, the results discussed below in this section can

be found in the cited source.

Properties that were considered early on are internodal distances. In particular the expectation and variance of the distance between two nodes and as a special case of the depth of node n were solved in the seventies. Moon found a recursion for the distribution of the distance between two fixed nodes by using the fact that uniform recursive trees have a relatively simple growing rule [2]. This recursion can then be used to derive the expectation without solving the recursion for the exact distribution. In 1996 Dobrow gave the exact distribution of the distance between a fixed node and n and proves asymptotic normality for the distance between certain sequences of nodes and node n [3]. Asymptotic normality of the distance between a fixed node and n was shown by Su, Liu and Feng in 2006 by using the decomposition of the distance into a sum of random variables and applying the Lindeberg-Feller central limit theorem [4]. They moreover proved a more general result, namely that the distance between any sequence of nodes and node n is asymptotically normal.

As a special case of an internodal distance, the exact distribution of the depth of node n was established by making use of a proper recursion by Szymański in 1990 [1]. Finally, asymptotic normality of the depth of node n was proven by two different methods. Devroye used the theory of records [5] and Mahmoud calculated the limit of the moment generating function of the depth using its exact distribution which was already known [1]. The expectation and limiting distribution of the depth of node n was proved by Feng, Su and Hu by yet another method in 2005 [6]. They were able to write the depth as a sum of independent indicator random variables, which easily yields expectation, variance and a central limit theorem.

Another statistic that was studied early on is the number of nodes of a certain degree and as a special case the number of leaves. The earliest result in this direction is about the expected number of nodes of a certain degree, which was proved using a recursion by Na and Rapoport [7]. Gastwirth could establish upper bounds by first writing the random variables as a sum of Bernoulli random variables and then using Poisson approximations [8]. For the exact distribution, expectation, variance

and asymptotic distribution of the number of leaves, again several methods can be used. First of all a special case of Friedman's urn [9] can be used to derive the exact distribution [1]. Friedman's urn contains balls of two colours and grows according to a replacement rule. When applied on the leaves of uniform recursive trees, the white balls represent the internal nodes and the black balls the leaves. When a black ball is drawn, this means that the parent of the next node was a leaf and has now become an internal node. The new node definitely is a leaf. We thus put the black ball back, together with an additional white ball. Similarly, if a white ball is drawn, we put it back together with a black ball, since the internal node is still internal and the new node is a leaf. Based on the urn model and recursions derived from it, a differential equation for the moment generating function can be established, which can be solved and turns out to be the generating function of the Eulerian numbers [1]. Similarly Mahmoud and Smythe derived the expectation, variance and asymptotic normality of the number of nodes of degree 1 and 2 by a generalization of the above described urn model [10].

By the use of a recursion Najock and Heyde also derived the distribution of the number of leaves and via some well-known results on permutations they could further prove asymptotic normality of the number of leaves [11]. The emergence of the use of results on permutation statistics in leaf related problems is no coincidence, as it was later discovered that there is a bijection between uniform random permutations and uniform recursive trees, see for example [12]. Using this bijection, size-biased coupling and Stein's method, results on the convergence rate could recently be established by Zhang [13]. Altok and Işlak refined these results and studied other leaf-related properties [12] .

Moving on to another statistic, the number of nodes with a fixed number of descendants was studied by Devroye in 90's [14]. He used a connection between recursive trees and binary search trees, which were well studied then. He also uses the idea of local counters which allows many local properties to be studied by writing them as sums of Bernoulli random variables and using central limit theorems for locally dependent random variables.

The properties that were historically investigated next were two global properties, and thus combinatorial methods turned out not to suffice [1]. For these properties different probabilistic methods were developed. Concerning the expectation of the maximum degree, Szymanski proved an upper bound using the averages of the number of nodes of a certain degree developed earlier [1]. Actually the maximum degree turned out to converge almost surely to that bound, which was proved by Devroye and Lu in 1995 [15]. Later results on the distribution of the maximum degree were obtained by Goh and Schmutz [16] and Addario-Berry and Eslava [17]. Moreover Eslava investigated the height of nodes of high degree [18]. The last two papers use a connection between uniform recursive trees and Kingman's coalescence, which is a new approach yet to uniform recursive trees. Kingman's coalescence is a Markov process that starts with all singletons of $[n]$. At every step subsets or blocks can merge to form new blocks, and all blocks merge with any other at the same rate, until all elements are in the same block [19].

The other global property about which results could only be obtained in the 90's is the height of the tree, i.e. the longest path from the root to a leaf. In 1994 Pittel obtained almost sure convergence of the height of uniform recursive trees by using another way of constructing a uniform recursive tree: from a branching structure with population-dependent rate we can obtain a uniform recursive tree by attaching the i-th node to the parent of the i-th born child in the branching process [20]. By using results about the connection between the time the n-th descendant is born and the birth time of the first member of the k-th generation, almost sure convergence could be obtained. Devroye proved this result without using branching processes, this time by making use of a second moment method argument [21]. Moreover convergence results for the minimum depth of the second half of the nodes are given in that paper.

Finally another important statistic of uniform recursive trees are their branches. By using the fact that subtrees of uniform recursive trees have, conditioned on their size, the same structure as the tree itself, results on the number of leaves and the size of the subtrees were derived in the 90's by Mahmoud and Smythe [22]. In that paper a generalized urn model, similar to the one described above, and known asymptotic

results about urn statistics are used. The branching structure was further investigated in 2005 by Feng, Su and Hu, mainly by using recursive formulas for branches of a given size [6]. The number of branches is shown to be asymptotically normal while the number of branches of a fixed size converges to a Poisson random variable. Furthermore, results on the size of the largest branch were proven in the same reference.

It is clear that the literature of uniform recursive trees encompasses a wide collection of methods. Recursions play an important role in various proofs, which is not surprising given that the considered structure is recursive. But the ways used to solve these recursions, or to use these recursions without solving them explicitly, differ considerably, as was described above. Moreover recursive trees can be constructed by at least 4 different processes: by dynamically building the tree step by step, by constructing it from a permutation, by building a genealogy tree for a branching process or by using Kingsman's coalescence. This makes uniform recursive trees very interesting from a mathematical point of view, since on the one hand methods from different areas can be applied in the investigation of uniform recursive trees and on the other hand connections between different mathematical structures can be discovered through the study of uniform recursive trees. Also the bijection between uniform recursive trees and permutations immediately suggests some applications. We will now describe some of them and demonstrate the importance of more general recursive tree models.

Regarding applications, first, recursive trees can be used as a model for the spread of epidemics [6]. In that case the root stands for the first person infected, and in general node i stand for the i-th person infected. Now the second person will definitely be infected by the first one, the third person can in turn be infected by the first or the second one and so on. In a uniform recursive tree, every node is attached to any of the previous nodes with the same probability, so in the model any of the previously infected persons is equally likely to infect the next one. In [2] a different way of modelling the spread of an infection is suggested: Given a uniform recursive tree an infection starts from a node i and spreads to any node attached to i with probability p. A uniform recursive tree can thus either model the infected people or the structure the infection spreads on.

Similarly uniform recursive trees are useful in order to determine the genealogy of ancient and medieval texts. As described in [11], often the original sources of old texts are lost. By modeling the existent copies as nodes of a uniform recursive tree whose root is the original text, it is possible to reconstruct the genealogy of these texts.

Also, recursive trees are used as models for the pyramid scheme [8]. The pyramid scheme is a business model that is based on offering people a sales job where they have to pay an initial fee to participate and most of their revenue will come from recruiting new people. By letting the i-th node in a recursive tree denote the i-th person that participates, the pyramid scheme can be modeled by a uniform recursive tree, which gives estimations on the number of persons that will not even recruit enough new sellers to make up for their initial investment. More generally, a uniform recursive tree can represent a distribution network where the root is the producer, the internal nodes are suppliers and the leaves are retailers, i.e. sell the product to the consumer.

Uniform recursive trees are moreover used to model the spread of a fire in a tree [23]. This is done by first determining each edge to be either fire-proof or to be set on fire. When an edge is set on fire it burns all edges connected to it, but cannot pass any edges previously determined as fire-proof. By removing all nodes that are connected to burnt edges, only some connected components of the tree remain. In [23] these components and the number of remaining nodes are then investigated.

Having at hand different distributions than the uniform one entails much more flexibility when modeling real life problems. For many applications described above recursive trees seem to be a proper structure to represent the phenomenon in question. Whether the uniform distribution is the most appropriate one is more questionable. Using uniform recursive trees implies that all nodes are identical, or more specifically that every infected person is equally likely to infect the next one, that every book is equally likely to be copied or that every person is equally likely to recruit the next seller. This is obviously not the case in real world applications. Thus for recursive trees to be successfully used it is necessary to investigate the properties of non-uniform distributions on recursive trees.

Parallel to the development of the theory of uniform recursive trees some other recursive tree structures were already studied. One of the most common ones are probably binary recursive trees, which are described in [24] and can also be represented bijectively by permutations. The binary recursive tree is also very well studied, see for example [25]. Binary trees can moreover be generalized in a straightforward way to m-ary trees by specifying that every node can have at most m children [26].

Another possibility is to consider plane-oriented recursive trees, i.e. recursive trees where the children of each node are ordered. By choosing each such tree with equal probability and subsequently ignoring different orderings of children, this leads to a non-uniform distribution on increasing trees [1]. In this model the attachment probabilities depend on the out-degrees of the nodes [26]. Plane-oriented recursive trees were first introduced in [27] where results on node degrees were derived and compared to analogous results for the uniform model.

Yet another distribution on recursive trees recently introduced are scaled attachment random recursive trees. There the parent of each node i is chosen as $\lfloor iX_i \rfloor$, where all X_i are identically distributed on $[0, 1)$. By choosing the uniform distribution for the X_i's uniform recursive trees can be recovered. This model was introduced in [21] and subsequently some depth properties were studied.

Another natural generalization of uniform recursive trees, Hoppe trees, was recently considered in [28]. There, the root is assigned a weight θ, all other nodes get weight 1. Node i then attaches to the root with probability $\frac{\theta}{\theta+i-2}$ and to any other node with probability $\frac{1}{\theta+i-2}$. This model is associated to Hoppe's urn, which has an application in modelling the alleles of a gene with mutation rate $\theta > 0$. Concerning many properties like the number of leaves, the height and the depth of node n, Hoppe trees behave similarly to uniform recursive trees.

The first model we have chosen to study in this thesis generalizes the idea of Hoppe trees: we assign every node a weight ω_i. Node j then attaches to node $1 \le i < j$ with probability $\frac{\omega_i}{\omega_1+\cdots+\omega_{j-1}}$. For this model we first give a coupling construction from

a uniform recursive tree on n nodes. We then study the number of branches and the depth of node n and give their expectation and variance, as well as some conditions under which asymptotic normality holds. We moreover derive explicit values for the expectation and the variance for some examples of weight sequences.

For the number of leaves we had to restrict ourselves to a model where the first k nodes have weight θ and the rest weight 1. For this case we first use a martingale argument to get expressions for the expectation and the variance. We then introduce another coupling between uniform recursive trees and weighted recursive trees like the one just described with the additional restriction that $\theta \in \mathbb{N}$. Moreover we introduce a coupling between Hoppe trees and weighted recursive trees where the first k nodes have weight $\omega_i \in \mathbb{R}^+$ and the rest have weight 1. This coupling easily allows inferences about the number of leaves of these trees, based on known results about uniform recursive trees and Hoppe trees.

In our model we choose fixed weights for each node, so every step is independent of the structure of the already present tree. This property is crucial in many arguments we use, such as couplings and martingales, as well as in applications of various central limit theorems. It is also worth noting that by defining the tree model via attachment probabilities, implies that we directly use some of the approaches on uniform recursive trees described above, but cannot use others.

Introducing weights is also interesting from the point of view of applications since it allows to introduce diversity among the nodes. In the other non-uniform distributions discussed above, all nodes have the same behaviour, or in other words attract nodes according to the same rule. When a recursive process does not satisfy such conditions, weighted recursive trees can be used to model it more precisely. Moreover the properties of weighted recursive trees and how much they differ from the uniform model can be interpreted as an indicator for the stability of a process. It is reasonable to assume that it is in general more probable for some nodes to get children as others. For example some persons might be more likely to infect others, some copies of ancient texts are more probable to have been copied again and some people might be more likely to recruit

new people. Thus it is interesting to see how much fluctuation in the attachment probabilities of a phenomenon can be tolerated when modelling with uniform recursive trees. If weighted recursive trees have the same asymptotic behaviour as uniform recursive trees under some conditions, we can still model processes that satisfy these conditions by the uniform model.

The second model we consider, introduced in [12], is based on a completely different approach based on the bijection between the symmetric group and recursive trees. Instead of the uniform distribution we choose a biased riffle shuffle distribution on the symmetric group and then consider the trees obtained from these permutations. We use the aforementioned connections between properties of permutations and recursive trees in order to derive results on the number of branches, the number of nodes with at least k descendants and the depth of node n.

Among the non-uniform distributions on the symmetric group the biased riffle shuffle distribution was chosen for several reasons. First of all the possibility to construct inverse riffle shuffles from random variables makes the model tractable. Moreover riffle shuffle permutations are themselves theoretically and practically important [12]. Finally biased riffle shuffle distributions vary a lot depending on the chosen parameters. Accordingly, the properties of the corresponding trees can differ more or less from the uniform case. For example the number of leaves and the number of branches can be limited if appropriate parameters are chosen. This allows us to model more diverse recursive phenomena with more precision. The two models we have chosen hence have the advantage of reflecting the diversity of approaches there are to uniform recursive trees and also being interesting for applications.

The rest of the thesis is organized as follows: Chapter 2 is devoted to providing the necessary background on graph theory and probability theory techniques that are to be used below. In particular, we begin with a review of some basic definitions and results from graph theory, especially about trees, and then introduce the tree statistics that are to be investigated later on. Also, we include some facts about permutations, and some relevant theorems and methods from probability theory that we will use

subsequently.

In Chapter 3 we then introduce different representations of uniform recursive trees and discuss the current literature on them. This will allow us to evaluate the distance between statistics of our models and the ones from the uniform case. Here, some non-uniform recursive tree models from the literature are also discussed.

Next we begin the actual topic of this thesis: In Chapter 4 we introduce the weighted recursive tree model and give results about its number of branches, the depth of node n and its number of leaves. Furthermore, we introduce two couplings of uniform recursive trees and weighted recursive trees.

We introduce the second model we will consider in Chapter 5. Since it is based on a different distribution on the symmetric group we first define riffle shuffle permutations and based on them biased recursive trees. Subsequently we review known results about their number of leaves and then give our results about the number of branches, the number of nodes with at least k descendants and the depth of node n. In the conclusion we will give an overview of the problems we could not solve and some further generalizations we think might be interesting.

2. PRELIMINARIES

2.1. Graph Theory

2.1.1. First Definitions and Properties

A *graph* G consists of a set of *vertices*, also called *nodes* or *points*, V, denoted by $V(G)$ and a set of *edges*, also called *lines*, E, denoted by $E(G)$, such that each element of E is an unordered pair of elements of V. The *size* of G, denoted by $|G|$ is the number of vertices of G. We will denote the number of edges of G by $e(G)$. A graph of size n can have between 0 and $\binom{n}{2}$ edges. A graph of size n with $\binom{n}{2}$ edges is called a *complete n-graph*, and is denoted by K_n. A graph of size n without edges is called an *empty n-graph* and denoted by E_n. The graph $K_1 = E_1$ is called *trivial*. Below, we will only focus on a special type of graphs, trees, which will always be of finite size.

A graph is called *labeled*, if all its vertices have a name. If a graph is labeled, not only the structure of the graph matters, but also between which nodes these edges exist. Thus, for each graph with vertex set V and edge set E, there are $|V|$ labelings. The number of labeled graphs of size n is $2^{\binom{n}{2}}$, since each edge can either be present or absent in the graph [29]. From now on we will consider all graphs to be labeled.

If two vertices v and w are *joined* by an edge $e = \{v, w\}$, they are said to be *adjacent* or *neighbouring*, and v, w are called *endpoints* of e. An edge will sometimes be denoted vw for convenience. Edges are said to be *adjacent* if they have a common vertex. A graph without loops, i.e. no edges of the form $\{v, v\}$, and no multiple edges is called *simple*. We will only deal with simple graphs in this thesis.

The set of vertices adjacent to a vertex v is denoted by $A(v)$ and the *degree* of a vertex is denoted by $d(v) := |A(v)|$. Here, $|\cdot|$ is used for the cardinality of the underlying set. A vertex of degree 0 is called an *isolated vertex*. Since every edge

has two endpoints, the well-known hand shaking lemma says that $\sum_{i=1}^{n} d(v_i) = 2e(G)$, where $V(G) = \{v_1, \ldots, v_n\}$.

A *path* P is a sequence of vertices $(v_1, v_2, \ldots, v_i)$ with the property that $\{v_1 v_2, v_2 v_3, \ldots, v_{i-1} v_i\} \subseteq E$. We sometimes also write $v_1 v_2 \ldots v_i$ to mean the path through v_1 to v_i. v_1 and v_i are called *endpoints of P* and P is said to *join* v_1 and v_i. The *length* of P is denoted by $\ell := e(P)$. A path is called *simple* if no vertex occurs more than once. A graph is *connected* if between any two vertices in V there exists a path joining them, which, in particular, implies that there is no isolated vertex.

If the edges are ordered pairs of vertices, G is said to be a *directed* graph, and a given edge e is written as (v, w) or vw. In the directed case, the edges vw and wv are not the same, and the edge vw can only be used to go from v to w in a path. In a directed graph the edge vw is said to *start* at vertex v and *end* at vertex w. A directed graph $G = (V, E)$ is called an *oriented* graph if $E \cap E^{-1} = \varnothing$; i.e. between any two vertices, there can only be an edge in one direction. For a detailed account of graph theory, we refer to the texts [29–32].

2.1.2. Basics about Trees

Given a graph $G = (V, E)$, a *cycle* is a path $v_1 v_2 \ldots v_\ell v_1$, with $v_i \in V$ for $i = 1, \ldots, l$; i.e. a cycle is a path with only one endpoint. A graph without any cycles is called a *forest*, and a connected graph without any cycles is called a *tree*, and is usually denoted by $\mathcal{T}$. In a tree all nodes of degree 1 are called *leaves*. Every tree with at least one edge has at least two leaves [33]. This can be seen by considering the longest path in $\mathcal{T}$ with distinct nodes, and concluding by contradiction that its endpoints must be leaves. The following proposition summarizes some equivalent formulations of a tree structure.

Proposition 2.1 ([32,33]). *The following are equivalent:*

(i) $\mathcal{T}$ *is a tree.*

(ii) *Any two vertices in* $\mathcal{T}$ *are connected by a single path in* $\mathcal{T}$.

(iii) $\mathcal{T}$ *is minimally connected, i.e.* $\mathcal{T}$ *is connected but* $\mathcal{T} \setminus e$ *is disconnected for every* $e \in E(\mathcal{T})$.

(iv) $\mathcal{T}$ *is maximally acyclic, i.e.* $\mathcal{T}$ *contains no cycle but* $\mathcal{T} \cup vw$ *does, for any two non-adjacent vertices* $v, w \in \mathcal{T}$.

(v) $\mathcal{T}$ *is connected and has* $n - 1$ *edges.*

(vi) $\mathcal{T}$ *has no cycles and* $n - 1$ *edges.*

Cayley's theorem states that there are n^{n-2} undirected labeled trees on n vertices [29]. A *rooted* tree is a labeled tree where one node is specified as the *root*. Cayley's theorem, in particular, implies that there are n^{n-1} rooted labeled trees on n vertices. In a rooted *plane* or *planted* tree the children of each node have a left-to-right ordering [29].

An oriented tree where all edges are directed outwards from the root is called *branching tree* [29]. In a branching tree, leaves are defined slightly differently, since the root should not be considered a leaf, even if it has degree 1. Thus, for a directed graph, we define the *outdegree*, $d_+(v) := |\{e \in E(\mathcal{T}) : e = (v, w), w \in V(\mathcal{T})\}|$, i.e. the number of edges starting in v. In a branching tree, a leaf is a vertex with outdegree 0. Since the trees we will consider are all branching trees, we will from now one use the word leaf in this sense. Similarly to the outdegree, the indegree $d_-(v)$ is defined as the number of edges ending in v, i.e. $d_-(v) := |\{e \in E(\mathcal{T}) : e = (w, v), w \in V(\mathcal{T})\}|$. Hence, in a branching tree, the root has indegree 0 and all other nodes have indegree 1. By definition of the degree of a vertex, it is in general true in an oriented graph that $d(v) = d_+(v) + d_-(v)$.

In a branching tree, if $vw \in E(\mathcal{T})$, w is called a *child* of v and v is called *parent* of w. Similarly, if there is a $y \in V(\mathcal{T})$ such that vy and yw are in $E(\mathcal{T})$, w is called *grandchild* of v and v is called *grandparent* of w. Every node can only have one parent

and one grandparent, since otherwise we would get a cycle. On the other hand, a node can have several children and all the nodes that have the same parent are called *siblings*. There is a single path from the root r to each vertex v and all nodes in this path are called *ancestors* of v. Similarly if v lies on the path from r to another node w, w is called a *descendant* of v [29].

There are many subcategories of trees. Some of these subcategories can be obtained by restricting the number of children a node can have, the most common one being the restriction to two children. Unfortunately there are different definitions for binary trees. In [29] a *binary tree* is defined as a rooted plane tree in which every node can have at most 2 children. However, a binary tree is also sometimes defined as a rooted tree where each node can have at most 2 children [32,33]. We will use the term binary tree is this latter sense, i.e. we will the not consider the children to be ordered. In general, for $m \in \mathbb{N}$, an *m-ary tree* is a rooted labeled tree where every node has at most m children [26,33].

The trees we will consider are another subcategory of trees, called *increasing* trees. These are rooted trees such that on every path from the root to a node the labels of the nodes are increasing [33]. It is also possible to consider the intersection of these two subcategories, i.e. increasing binary trees or increasing m-ary trees.

2.1.3. Statistics of Interest

In this subsection, we define certain tree statistics that are studied under various branching structures. Most of these will be directly or indirectly handled in the following chapters for the non-uniform recursive tree models we discuss. Let $\mathcal{T}$ be a tree.

Recall that a vertex is called a leaf if it has no children. The *number of leaves* of a branching tree is denoted by $\mathcal{L}_{\mathcal{T}}$. For $|E(\mathcal{T})| > 1$, $1 \leq \mathcal{L}_{\mathcal{T}} \leq n-1$, which can be seen by construction or by considering that for a tree $\sum_{v \in V(E)} d_-(v) = n-1$, because the total number of edges is $n-1$. Similarly, we can also consider the *number of nodes*

with at least degree k, i.e. the number of nodes $v \in \mathcal{T}$, with $d(v) \geq k$.

The *number of branches* $\mathcal{B}_\mathcal{T}$ of a rooted tree is the number of children of the root. Clearly, we have $1 \leq \mathcal{B}_\mathcal{T} \leq n-1$. Since every branch is again a branching structure, with the child of the root as the new root, all statistics defined for trees can also be considered on branches. In particular, we will be interested in the size of the branches, i.e. in the number of nodes that are descendants of a given child of the root. Let w be a child of the root and b_w the branch rooted at w. If we define $Anc(v)$ as the set of ancestors of node v, then $|b_w| = |\{v \in V : w \in Anc(v)\}|$.

The *depth* $\mathcal{D}_v$ of node v in a branching tree is the length of the path from the root to v or equivalently the number of ancestors of v. For $v \neq r$, we have $1 \leq \mathcal{D}(v) \leq n-1$. Similarly, the *distance* from one node to another, denoted by $\mathcal{D}_{vw}$, is the length of the path from v to w. For general graphs the distance between two nodes is the length of the shortest path from v to w but since in a tree there is only one path between any two nodes, this amounts to the same.

The *height* $\mathcal{H}_\mathcal{T}$ of a rooted tree is the length of the longest path from the root to a leaf. As for the other statistics $1 \leq \mathcal{H}_\mathcal{T} \leq n-1$, again with the same examples. When comparing the height of two trees, $\mathcal{T}$ and $\mathcal{T}'$, we say that $\mathcal{T}$ is *taller* than $\mathcal{T}'$ and $\mathcal{T}'$ is *flater* or *shorter* than $\mathcal{T}$, when $\mathcal{H}_{\mathcal{T}'} < \mathcal{H}_\mathcal{T}$.

2.2. Permutations

There are several ways to define permutations, we will only give two here, see [34]. While permutations can be defined for any finite set, by giving each element of the set a label from 1 to n we can consider that all permutations are defined on $[n]$. First of all a *permutation* of size n is an linear ordering of the set $[n]$, say $\pi_1\pi_2\ldots\pi_n$, where every integer from 1 to n only appears once. In other words, a permutation is a word with letters from $[n]$, such that each letter appears exactly once. Equivalently a permutation can be considered as a bijective map $\pi : [n] \to [n]$, by defining $\pi(i) = \pi_i$.

Corresponding to these two ways of defining permutations there are two different notations for permutations. Corresponding to the view of a permutation as a bijection, a permutation π can be represented as

$$
\begin{matrix}
1 & 2 & 3 & \ldots & n-1 & n \\
\pi(1) & \pi(2) & \pi(3) & \ldots & \pi(n-1) & \pi(n).
\end{matrix}
$$

We call this the *Cauchy representation*. Corresponding to the view of a permutation as a list of the numbers from 1 to n, a permutation also has a *word* or *one-line representation*,

$$
\begin{matrix}
\pi(1) & \pi(2) & \pi(3) & \ldots & \pi(n-1) & \pi(n).
\end{matrix}
$$

Since the word notation is more compact and thus easier to include in the text we will prefer it most of the time.

The set of all permutations of size n is called *symmetric group* of size n and is denoted by S_n. The number of permutations of $[n]$ is $n!$, since $\pi(1)$ can be chosen among n elements, $\pi(2)$ among $n-1$ elements, and so on. A *uniform random permutation*, or URP of $[n]$ is a permutation chosen uniformly among all permutations of $[n]$.

There are several properties of permutations that will be important later, when we use them as representations for trees. For a permutation π of $[n]$, an *inversion* is a pair $(i, j) \in [n] \times [n]$, such that $i < j$ and $\pi(i) > \pi(j)$. For $1 \le i \le n-1$, the permutation π has a *descent* in i if $\pi(i) > \pi(i+1)$ and an *ascent* in i if $\pi(i) < \pi(i+1)$. For example the permutation $\pi = 41562837$ has 3 descents: in 1, 4 and 6, and 4 ascents in 2,3,5 and 7.

Another concept we will often use are records and anti-records. A record is an element that is greater than all previous ones and an anti-record an element that is

smaller than all previous ones. More precisely: a permutation π has a *record* in i if $\pi(i) > \{\pi(1), \ldots, \pi(i-1)\}$ and an *anti-record* in i if $\pi(i) < \{\pi(1), \ldots, \pi(i-1)\}$ [34]. Every permutation has a record and an anti-record in $\pi(1)$.

Records and anti-records are equally distributed in uniform random permutations as we can see by the following bijection: Consider the map $f : S_n \to S_n$ defined by $f : \pi \to \rho = n - \pi$, i.e. for all $i = 1, \ldots, n$, we have $\rho(i) = n - \pi(i)$. This map is bijective and if π has an anti-record in i, then ρ has a record in i by construction. Thus the distributions of the number of anti-records and records in URPs are equal.

Nevzorov summarizes many results about records of sequences of random variables in [35]. The theory of records is very rich and connected to several mathematical structures and also has many applications. The results in [35] also apply for the records in a uniform random permutation by the following standard construction of a uniform recursive permutation from random variables, as described for instance in [12]. Let $Y_1, \ldots, Y_n$ be independent uniformly distributed random variables over $(0, 1)$. Then the *rank* R_i of Y_i is equal to j if Y_i is the j-th-largest among $Y_1, \ldots, Y_n$. The sequence $(R_1, \ldots, R_n)$ is distributed as a uniform random permutation of $[n]$, see [5]. In our case we will often only consider permutations where $\pi(1) = 1$ and thus use permutations of $\{2, 3, \ldots, n\}$. This means that we mostly use $Y_2, \ldots, Y_n$ to construct the permutation we need.

By considering a permutation as a bijection it can also be represented by its cycle structure. Since we only consider permutations of finite sets, for every $i \in [n]$, there will be a unique $\ell \in [n]$, such that $\pi^\ell(i) = i$. We can thus define a *cycle* of *length* ℓ of a permutation as a sequence $(i, \pi(i), \pi^2(i), \ldots, \pi^{\ell-1}(i))$, where $\pi^\ell(i) = i$. For $k = 0, 1, 2, \ldots$, the cycles $(\pi^k(i)\pi^{k+1}(i)\ldots\pi^{k+\ell-1}(i))$ are all representations of the same cycle. Since every element of $[n]$ can only be in one cycle, the different cycles of a permutation are distinct and thus partition $[n]$. Hence we can write every permutation as a product of at most n distinct cycles, $C_1, \ldots, C_m$, i.e. $\pi = C_1 C_2 \ldots C_m$. The cycles have different representations because they can start with any member. Moreover the cycles can also be ordered in several ways. Thus the cycle representation of a

permutation is not unique.

To guarantee uniqueness it is common to start every cycle with its smallest element and order the cycles according to this first element from largest to smallest. In this way, even if we remove the parenthesis, we know that a new cycle starts every time there is an anti-record in the permutation. Thus in this standard notation, every permutation has a unique cycle representation. For example, given the permutation 439782516, we determine the first cycle by considering $\pi(1) = 4$, then $\pi(4) = 7$, $\pi(7) = 5$, $\pi(5) = 8$, $\pi(8) = 1$, which gives the cycle (14758) in standard notation. Then we take the smallest element not in the first cycle, which is in this case 2, and proceed similarly, and so on. Finally we get $\pi = (2396)(14758)$. The advantage of this notation is that one can recover the cycles even if the parenthesis are removed: every time an anti-record, i.e. a new smallest element appears, a new cycle starts.

2.3. Probability Theory

We start with a brief explanation of the method of indicators as it will be used several times throughout the thesis. The basic idea is to write a discrete random variable X as a sum of Bernoulli random variables. Often X is the number of something, as in the following example.

Example 2.2 ([36]). At a party n men throw their hat in the air, and then every man chooses one of the hats randomly. Let X be the random variable denoting the number of men who choose their own hat. In order to calculate $\mathbb{E}[X]$ we define the Bernoulli random variables $X_i := \mathbf{1}(i\text{-th man finds his own hat})$. Then we have

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1. \tag{2.1}$$

When we use this method we will often consider the distribution of the limit of such sums of indicators. The asymptotic results we will prove are results of convergence in distribution.

Definition 2.3 ([37]). Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables with cumulative distribution function $F_n(x)$. Let X be another random variable with cumulative distribution function $F(x)$. If for all x at which F is continuous

$$\lim_{n \to \infty} F_n(x) = F(x) \tag{2.2}$$

the sequence (X_n) is said to *converge in distribution* to X and we write $X_n \xrightarrow{d} X$.

When comparing random variables we will use three different probability metrics. Let μ and ν be two probability measures. Then each of the metrics we will consider have the form

$$d_{\mathcal{H}}(\mu, \nu) = \sup_{h \in \mathcal{H}} \left| \int h(x) d\mu(x) - \int h(x) d\nu(x) \right| \tag{2.3}$$

where $\mathcal{H}$ is some family of functions. By extension this also gives a distance function for random variables: if X and Y are random variables with respective laws μ and ν, then $d_{\mathcal{H}}(X, Y) = d_{\mathcal{H}}(\mu, \nu)$. Depending on the set $\mathcal{H}$, this form gives rise to different metrics.

Definition 2.4. (i) The *Kolmogorov metric* is obtained by setting $\mathcal{H} = \{\mathbf{1}(x \leq a) : a \in \mathbb{R}\}$, and is denoted by d_K.

(ii) We get the *Wasserstein metric* if we set $\mathcal{H} = \{h : \mathbb{R} \to \mathbb{R} : |h(x) - h(y)| \leq |x - y|\}$, and we denote it by d_W. This is the main metric used for approximations by continuous distributions.

(iii) The *total variation metric* is obtained by choosing $\mathcal{H} = \{\mathbf{1}(x \in A) : A \in \mathcal{B}(\mathbb{R})\}$ and is denoted by d_{TV}. This metric is commonly used for approximations by discrete distributions.

These metrics have some important properties, which can be found in [38].

Proposition 2.5 ([38]). *(i) For any two random variables X and Y,*

$$d_K(X, Y) \leq d_{TV}(X, Y). \tag{2.4}$$

(ii) *If the density function of a random variable X is bounded by a constant C, and Y is any random variable,*

$$d_K(X,Y) \leq \sqrt{2C d_W(X,Y)}. \tag{2.5}$$

(iii) *If the random variables X and Y take values in a discrete space A,*

$$d_{TV}(X,Y) = \frac{1}{2} \sum_{a \in A} |\mathbb{P}(X = a) - \mathbb{P}(Y = a)|. \tag{2.6}$$

(iv) *The Kolmogorov metric gives the maximum distance between distribution functions. Thus, if for a sequence of random variables $X_1, X_2, \ldots$ and a random variable Y, $d_W(X_n, Y) \to 0$, then $(X_n)_{n \in \mathbb{N}}$ converges to Y in distribution.*

For proofs of items $(i), (ii)$, and (iv) see [38] and for (iii) see [39]. Using Stein's method several bounds on these metrics for different kinds of random variables are then given in [38]. While we will mainly use these bounds to determine rates of convergence, Stein's method is a general tool to determine the distance between two probability measures. It was introduced by Stein in [40] as a method to bound the error in approximations of sums of random variables with the normal distribution. Later the method was generalized and applied on bounds of approximations of more general random variables as well as with other distributions. The first part of the main idea of Stein's method is to bound the distance between the random variable we want to approximate and a well-known distribution by the expectation of a functional of the random variable we want to approximate. The second part consists in methods to bound the expectation of that functional [38].

In order to analyze the asymptotic behavior of the statistics introduced in the previous section we also need some limit theorems. In subsequent theorems we use the following notation: in general $Y_1, Y_2, \ldots$ denote a sequence of random variables and we

define $\mu_i := \mathbb{E}[Y_i]$ and $\sigma_i^2 := \mathrm{Var}(Y_i)$. Also we set

$$W_n := \frac{\sum_{i=1}^n Y_i - \mu_i}{(\sum_{i=1}^n \sigma_i^2)^{\frac{1}{2}}}. \tag{2.7}$$

Note that we have $\mathbb{E}[W_n] = 0$ and $\mathrm{Var}(W_n) = 1$. The following theorems give conditions on when W_n converges to a standard normal distribution.

Theorem 2.6 (Lindeberg-Feller's central limit theorem, [41])**.** *Let $Y_1, Y_2, \ldots$ be independent random variables such that $\mu_i < \infty$ and $\sigma_i^2 < \infty$ for all $i = 1, 2, \ldots$. Define $s_n^2 = \sum_{i=1}^n \mathrm{Var}(Y_i)$ and let W_n be as above. Assume that for all $\epsilon > 0$,*

$$\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}\left[(Y_i - \mu_i)^2 \mathbf{1}(|Y_i - \mu_i| > \epsilon s_n)\right] = 0. \tag{2.8}$$

Then

$$\lim_{n \to \infty} W_n =_d \mathcal{G}. \tag{2.9}$$

Another condition for sums of independent random variables to converge, that implies the Lindeberg-Feller condition, is Liapounov's condition.

Theorem 2.7 (Liapounov's central limit theorem, [37])**.** *Let $Y_1, Y_2, \ldots$ be independent random variables and μ_i, σ_i^2 and W_n be as above. Now assume that $\mathbb{E}\left[|Y_i - \mu_i|^3\right] < \infty$ for all $i = 1, 2, \ldots$ and that*

$$\lim_{n \to \infty} \frac{\sum_{i=1}^n \mathbb{E}[|Y_i - \mu_i|^3]}{(\sum_{i=1}^n \sigma_i^2)^{\frac{3}{2}}} = 0. \tag{2.10}$$

Then

$$\lim_{n \to \infty} W_n =_d \mathcal{G}. \tag{2.11}$$

The following case of the Liapounov theorem will suffice for our purposes.

Theorem 2.8 (Liapounov's central limit theorem for sums of independent Bernoulli random variables, [37]). *Let $Y_1, Y_2, \ldots$ be independent Bernoulli random variables with parameter p_i, $i = 1, 2, \ldots$. Letting W_n be as above we have*

$$W_n = \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n p_i}{\left(\sum_{i=1}^n p_i(1 - p_i)\right)^{\frac{1}{2}}}. \tag{2.12}$$

If the infinite series $\sum_{i=1}^\infty p_i(1 - p_i)$ diverges, then

$$\lim_{n \to \infty} W_n =_d \mathcal{G}. \tag{2.13}$$

In connection to asymptotic convergence to the normal random variable, the following result is also of interest.

Theorem 2.9 ([42]). *Let W_n be a sequence of random variables. If there is a sequence of real numbers λ_n such that $\lambda_n \to \infty$ and*

$$d_{TV}(W_n, \mathrm{Po}(\lambda_n)) \xrightarrow[n \to \infty]{} 0 \tag{2.14}$$

then

$$\lim_{n \to \infty} \frac{W_n - \lambda_n}{\sqrt{\lambda_n}} =_d \mathcal{G}. \tag{2.15}$$

As a bound on the distance between sums of Bernoulli random variables and a Poisson distribution, we will use the following result.

Theorem 2.10 (Law of small numbers, [38]). *Let $Y_1, Y_2, \ldots, Y_n$ are independent indicator random variables with $\mathbb{P}(Y_i = 1) = p_i$, $Y = \sum_{i=1}^n Y_i$ and $\mu = \mathbb{E}[Y] = \sum_{i=1}^n p_i$.*

Then

$$d_{TV}(Y, \mathrm{Po}(\mu)) \leq \min\{1, \mu^{-1}\} \sum_{i=1}^{n} p_i^2. \tag{2.16}$$

As we will also consider several cases of sums of locally dependent random variables, we will now give two bounds for approximations of sums of locally dependent random variables by the normal distribution.

Definition 2.11 ([38]). Let $(Y_1, Y_2, \ldots, Y_n)$ be a collection of random variables. For each i, we call N_i the *dependency neighbourhood* of Y_i, if Y_i is independent of $\{Y_j\}_{j \notin N_i}$ and $i \in N_i$.

For random variables with such dependency neighbourhoods the following convergence theorem is true:

Theorem 2.12 ([38]). *Let $Y_1, Y_2, \ldots, Y_n$ be random variables for which $\mathbb{E}[Y_i^4] < \infty$ and $\mathbb{E}[Y_i] = 0$ holds. Let moreover N_i be the dependency neighbourhoods of $(Y_1, \ldots, Y_n)$ and define $D := \max_{1 \leq i \leq n}\{|N_i|\}$. Finally set $\sigma^2 = \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right)$ and define $W := \sum_{i=1}^{n} \frac{Y_i}{\sigma}$. Then*

$$d_W(W, \mathcal{G}) \leq \frac{D^2}{\sigma^3} \sum_{i=1}^{n} \mathbb{E}[|Y_i|^3] + \frac{\sqrt{28}D^{\frac{3}{2}}}{\sqrt{\pi}\sigma^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}[Y_i^4]}. \tag{2.17}$$

Note that this theorem can also be applied to independent random variables by setting $D = 1$, since $i \in N_i$ for all i. There is another more restricted definition of local dependence, which will mostly suffice for our cases.

Definition 2.13. For $m \in \mathbb{N}_0$, a sequence $(Y_i)_{i \in \mathbb{N}}$ of random variables is called *m-dependent* if for all $i \in \mathbb{N}$, the sets $\{Y_j, j \leq i\}$ and $\{Y_j, i + m < j\}$ are independent.

Note that a sequence of random variables is independent if it is 0-dependent in the above sense. Definition 2.13 is a special case of Definition 2.11 where the dependency

sets do not have to be sets of random variables with consecutive indices. In order to derive asymptotic results for m-dependent random variables, we will need a special case of theorem 9.4 in [43]. There, the random variables are indexed over $\mathbb{N}^d$, so we need to take $d = 1$. This theorem was also proved by above mentioned Stein's method.

Theorem 2.14. *Let* $Y_1, Y_2, \ldots$ *be a sequence of zero-mean m-dependent random variables and* $W_n := \frac{\sum_{i=1}^n Y_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{\frac{1}{2}}}$. *Then for all* $p \in (2, 3]$,

$$d_W(W_n, \mathcal{G}) \leq 75(10m + 1)^{p-1} \sum_{i=1}^n \mathbb{E}[|Y_i|^2]. \tag{2.18}$$

In the analysis of the number of leaves of a special kind of weighted recursive tree, we will moreover need the following concentration inequality for martingale difference sequences, which is Theorem 3.13 from [44].

Theorem 2.15 ([44]). *Let* $Y_1, Y_2, \ldots, Y_n$ *be a martingale difference sequence with* $a_i < Y_i < b_i$ *for each* i, *for suitable constants* a_i, b_i. *Then for any* $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right| \geq t\right) \leq 2e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}. \tag{2.19}$$

By using a coupling construction we will moreover be able to bound some differences of random variables. In order to derive results about the asymptotic behaviour of the unknown random variable from the asymptotic distribution of the known one, we will use Slutsky's Theorem.

Theorem 2.16 (Slutsky's theorem [41]). *Let* X_n *and* Y_n *be sequences of random variables such that* $X_n \to_d X$ *and* $Y_n \to_d c$ *for* $c \in \mathbb{R}$. *Then*

$$\lim_{n \to \infty} X_n + Y_n =_d X + c. \tag{2.20}$$

3. REVIEW OF UNIFORM RECURSIVE TREES

3.1. Definition of Uniform Recursive Trees

In following chapters we will consider different distributions on the space of recursive trees. *Increasing* or *recursive* trees are branching trees whose vertices are labeled by $\{1, \ldots, n\}$ such that on every path starting from the root the labels are increasing, which implies that node 1 is the root. This property allows us to picture many increasing trees as growing dynamically, such that an increasing tree of size n is obtained by joining node n to an increasing tree of size $n - 1$ according to some rule [26]. We will see later that this is not the case for biased recursive trees. Figure 3.1 shows all increasing trees on 4 vertices.
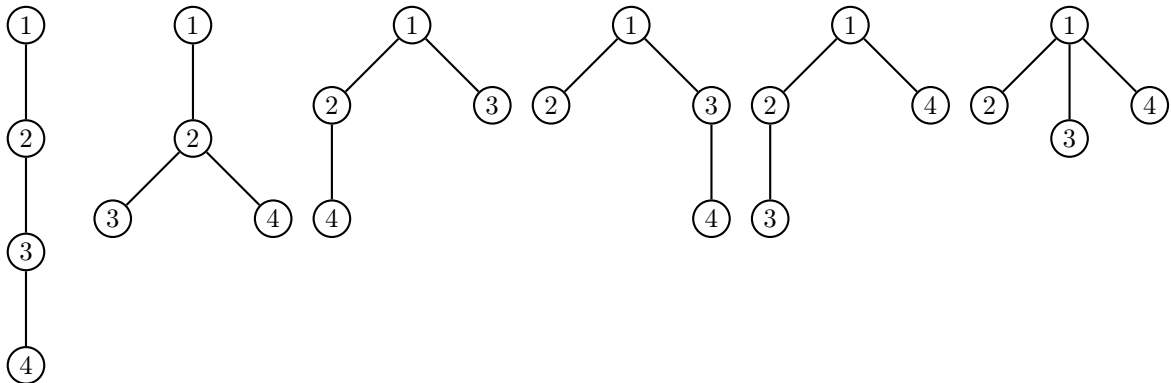
Figure 3.1. All increasing trees on 4 vertices.

A *uniform recursive tree*, abbreviated by URT, of size n is a random tree that is chosen uniformly among all increasing trees of size n. Since attaching node n to any of the nodes $\{1, \ldots, n - 1\}$ gives an increasing tree, there are $(n - 1)!$ recursive trees of size n.

3.1.1. Recursive Construction of a Uniform Recursive Tree

This is equivalent to the following recursive construction principle. A URT $\mathcal{T}_n$ of size n is obtained from a URT $\mathcal{T}_{n-1}$ of size $n - 1$ by joining node n to any of the

nodes $\{1, \ldots, n-1\}$ with equal probability. If we start from scratch, first node 1 is added as the root, and node 2 is attached to node 1. Then, node 3 is either attached to node 1 or to node 2 with equal probability $\frac{1}{2}$. In general node i attaches to any of the nodes $\{1, \ldots, i-1\}$ with probability $\frac{1}{i-1}$ [12]. It is important to note that every step is independent of the previous ones for URTs since the structure of $\mathcal{T}_{n-1}$ does not affect which node becomes the parent of n. We will call this way of constructing a URT *construction principle* and refer to the probabilities for nodes to attach to other nodes as the *attachment probabilities*.

3.1.2. Construction of a Recursive Tree from a Permutation

Besides the two alternative definitions of a URT described above, we may also make use of random permutations to generate such trees. Given a permutation π of $\{2, \ldots, n\}$, we construct a URT as follows: First, 2 is attached to 1, then 3 is connected to 1 if it is to the left of 2, otherwise to 2. In general the node i is attached to the rightmost node to the left of i that is less than i. If there is no smaller number than i to its left, i is attached to 1. Thus $i = \pi(s)$ attaches to node $j = \pi(r)$ when $r = \max\{t \in \{1, \ldots, s-1\} : \pi(t) < \pi(s)\}$, where we set $\pi(1) = 1$. We will call this way of constructing a URT the *construction from a permutation*. Figure 3.2 shows an example of the step by step construction of a recursive tree corresponding to a permutation. Similarly, given a URT $\mathcal{T}_n$, we can construct the corresponding permutation by writing 1 to the very left, 2 on its right and then step by step every node i directly on the right of the node it is attached to.

That this relation gives a bijection between URTs and URPs can be seen by the symmetry of the recursive constructions. The tree and the permutation corresponding to it can be constructed simultaneously, since we can construct a uniform random permutation by inserting the numbers successively, see [12]. Given any permutation π of $\{2, \ldots, k\}$ we can construct a permutation π' of $\{2, \ldots, k+1\}$ by inserting $k+1$ at the ultimate left or right or between any $\pi(i)$ and $\pi(i+1)$ for $i = 2, \ldots, k-1$. In total there are k spots where we can insert $k+1$. If at each step we choose the spot for the next number uniformly, this process gives a uniform random permutation π' on

$\underline{1}6387254$ $1\mathbf{6}387\underline{2}54$ $16\mathbf{3}87254$ $1637\mathbf{8}2\underline{5}\underline{4}$ $16378\mathbf{2}\underline{5}4$
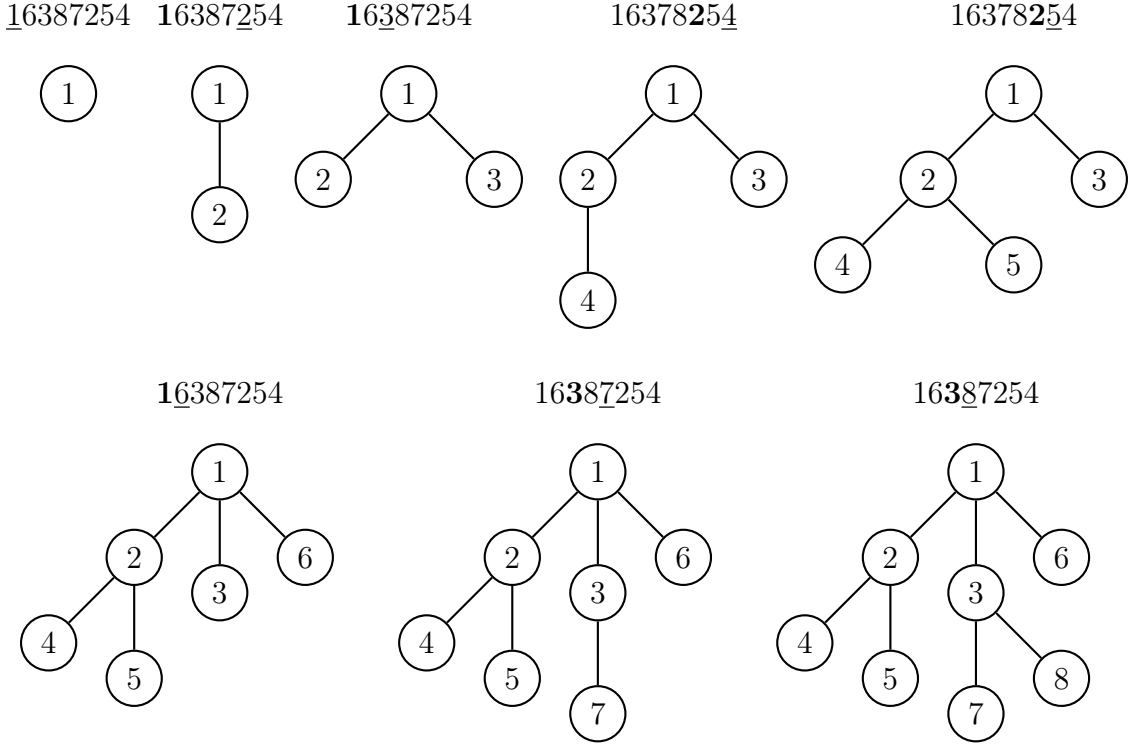
Figure 3.2. Step by step construction of the recursive tree corresponding to 16387254. The newly attached node is underlined and its parent bold.

$\{2, \ldots, k+1\}$. Note that if we insert $k+1$ between π_i and π_{i+1}, $\pi'(s) = \pi(s)$ for all $s \leq i$, $\pi(i+1) = k+1$ and $\pi'(s) = \pi(s-1)$ for all $s > i$.

3.1.3. Simultaneous Construction of a URP in Cycle Notation and a URT

By using the cycle notation for permutations, we can define another way of defining a bijection between URPs and URTs. We will show how we can simultaneously construct a URP in cycle notation and a URT [25]: We start with the node 1. At the first step we attach 2 to 1. The corresponding permutation is $\pi = (2)$. At each following step, we proceed as follows: if j is attached to the root in the URT, we add a cycle of length 1 that only contains j. If j is attached to a node $i > 1$, we add j to the cycle of i right after i. Thus there are in total $j - 1$ slots we can put j in. If we choose uniformly among all possibilities, we get a URT and a URP in cycle notation simultaneously.

Since we want the permutation to have the standard cycle notation, when we start a new cycle, we put it in the leftmost place. This is the only thing we need to do, since when adding a number to an existent cycle, it will always be bigger than the already present ones, hence the condition that every cycle starts with its smallest element is automatically satisfied. It is easy to see that we simultaneously get a unique URT and a unique URP by following this process, so this is a bijection. See Figure 3.3 for an example.
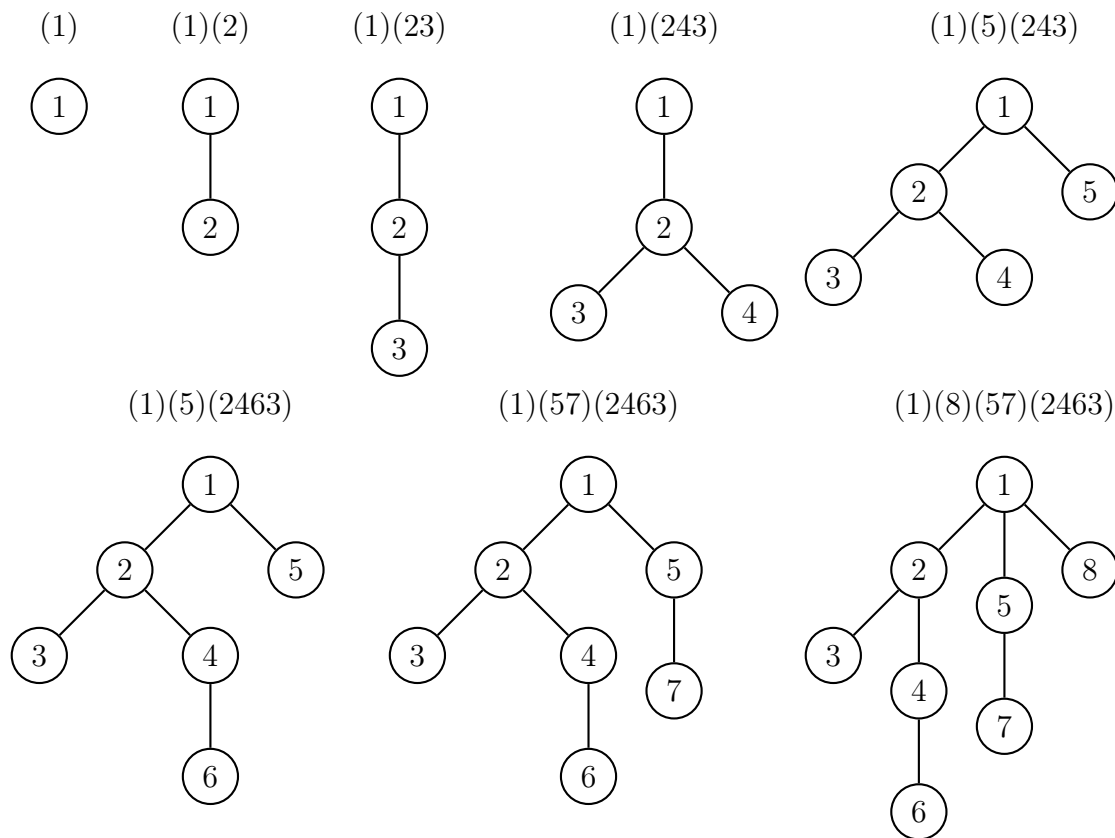
Figure 3.3. Simultaneous construction of a uniform recursive tree and a permutation in cycle notation.

When comparing the two permutations one gets for the same tree by these two bijections, one can see that the numbers are actually in the same order, we just interpret the sequence in two different ways: either as a permutation in Cauchy notation or in cycle notation. Thus we get the cycle representation of a URT from the Cauchy representation by simply starting a new cycle at every anti-record.

By simultaneously considering a uniform recursive tree and a uniform random permutation in cycle notation in this way, we can see that the construction is also equivalent to a *Chinese restaurant process*: the cycles can stand for round tables and every node corresponds to a customer. When a new customer arrives, s/he can either join an already existing table, or sit at a new one [25].

3.2. A Brief Literature Review on Uniform Recursive Trees

Various aspects of URTs are well studied, and in particular the statistics we introduced in Section 2.1.3 are deeply understood in most cases. In this section, we briefly go over some known results on URTs. For a more detailed survey on the subject see [1] or the relevant chapters in [25].

As mentioned in the introduction there are several ways of studying the leaves of a URT. By using the bijection between URTs and URPs described above, the expectation and variance of the number of leaves in a URT can easily be derived. First we observe that i is a leaf if and only if none of the nodes in $\{i+1, \ldots, n\}$ is attached to i. This is the case if and only if none of $\{i+1, \ldots, n\}$ is inserted in the spot to the right of i. Thus $i = \pi(r)$ is a leaf if and only if $\pi(r) > \pi(r+1)$, because this implies by the construction principle that $\pi(r)$ cannot be the closest smaller label to the left of any $j > i$. Moreover $\pi(n)$ definitely is a leaf. Thus we can write $\mathcal{L}_n$, the number of leaves of the URT $\mathcal{T}_n$, as

$$\mathcal{L}_n =_d \sum_{r=2}^{n-1} \mathbf{1}(\pi(r) > \pi(r+1)) + 1. \tag{3.1}$$

Since the permutation is uniformly random, $\mathbb{E}(\mathbf{1}(\pi(r) > \pi(r+1)) = 1) = 1/2$, and we can immediately conclude that $\mathbb{E}[\mathcal{L}_n] = \frac{n}{2}$. Similar considerations yield $\mathrm{Var}(\mathcal{L}_n) = \frac{n}{12}$. Concerning the limiting distribution of $\mathcal{L}_n$, we moreover have the following theorem:

Theorem 3.1 ([12]). *Let $\mathcal{T}_n$ be a URT of size n and $\mathcal{L}_n$ the number of leaves of $\mathcal{T}_n$.*
Then

$$d_K\left(\frac{\mathcal{L}_n - n/2}{\sqrt{n/12}}, \mathcal{G}\right) \leq \frac{C}{\sqrt{n}} \tag{3.2}$$

and for any $x > 0$

$$\max\left\{\mathbb{P}\left(\mathcal{L}_n - \frac{n}{2} \geq x\right), \mathbb{P}\left(\mathcal{L}_n - \frac{n}{2} \leq -x\right)\right\} \leq e^{-\frac{2x^2}{n}}. \tag{3.3}$$

As mentioned in the introduction, the expectation of the number of nodes of degree k was also established early on.

Theorem 3.2 ([7]). *Let C_n^k denote the number of nodes of degree k in a URT $\mathcal{T}_n$ of size n. Then*

$$\frac{\mathbb{E}\left[C_n^k\right]}{n} \xrightarrow{n \to \infty} \frac{1}{2^k}. \tag{3.4}$$

There are several methods to analyze the distribution of the number of branches, $\mathcal{B}_n$ in a URT. We will introduce three of them here because we will need them later. The first approach is to use the construction principle, see for example [6]. Since the number of branches is equal to the number of children of node 1, we set $X_i = \mathbf{1}(i$ is a child of $1)$, and write $\mathcal{B}_n$ as a sum of these indicator random variables:

$$\mathcal{B}_n = \sum_{i=2}^{n} X_i = 1 + \sum_{i=3}^{n} X_i. \tag{3.5}$$

Since it is equally probable for node i to attach to any of the already present nodes, $\mathbb{P}(X_i = 1) = \frac{1}{i-1}$ for $i = 2, \ldots, n$. This implies $\mathbb{E}[X_i] = \frac{1}{i-1}$ and $\text{Var}(X_i) = \frac{i-2}{(i-1)^2}$.

Thus, we get

$$\mathbb{E}[\mathcal{B}_n] = \sum_{i=2}^{n} \mathbb{E}[X_i] = \sum_{i=2}^{n} \frac{1}{i-1} = \sum_{i=1}^{n-1} \frac{1}{i} = H_{n-1} \tag{3.6}$$

and for large n we have $\mathbb{E}[B_n] = \ln(n) + \mathcal{O}(1)$. Also

$$\text{Var}(\mathcal{B}_n) = \sum_{i=2}^{n} \text{Var}(X_i) = \sum_{i=2}^{n} \frac{i-2}{(i-1)^2} = \sum_{i=1}^{n-1} \frac{i-1}{i^2} = H_{n-1} - H_{n-1}^{(2)} \tag{3.7}$$

and for large n we have $\text{Var}(\mathcal{B}_n) = \ln(n) + \mathcal{O}(1)$. Moreover these results allow us to derive a central limit theorem for $\mathcal{B}_n$ by using the Lindeberg-Feller theorem.

Theorem 3.3 ([6]). *Let $\mathcal{B}_n$ denote the number of branches of a URT $\mathcal{T}_n$ of size n. Then*

$$\frac{\mathcal{B}_n - \ln(n)}{\sqrt{\ln(n)}} \xrightarrow{d} \mathcal{G}. \tag{3.8}$$

We can moreover use the representation of $\mathcal{T}_n$ as a random permutation in order to calculate the number of branches. We can observe that $\pi(2)$ attaches to 1 and is thus the first node of a branch of T_n. If $\pi(3) > \pi(2)$ it will attach to $\pi(2)$, if $\pi(4) > \pi(2)$ it will attach to $\pi(2)$ or $\pi(3)$ and so on. As long as no $\pi(r) < \pi(2)$ appears, all nodes will be part of the branch starting with node $\pi(2)$. Let $r = \min\{r = 3, \ldots, n : \pi(r) < \pi(2)\}$ then $\pi(r)$ also attaches to 1 and is thus the start of the next branch. Similarly, as long as no smaller number comes up, all subsequent nodes will be part of this second branch. In general, $\pi(r)$ attaches to 1 if and only if $\pi(r) = \min\{\pi(2), \ldots, \pi(r)\}$ which means that $\pi(r)$ must be an anti-record. Thus, the number of branches of $\mathcal{T}_n$ is equal to the number of antirecords of π and we can write the number of branches $\mathcal{B}_n$ of $\mathcal{T}_n$ as another sum of indicator random variables:

$$\mathcal{B}_n = \sum_{r=2}^{n} \mathbf{1}(\pi(r) = \min\{\pi(1), \ldots, \pi(r)\}). \tag{3.9}$$

As expected the number of records in a URP and thus also the number of anti-records satisfy the same central limit theorem as $\mathcal{B}_n$, see [5, 45]. In [5] the equality in distribution between the number of branches and the number of records of a permutation was discovered without refering to the permutation representation of a URT. This way of analysing the number of branches in a URT has the advantage that records of sequences of random variables and thus random permutations are well studied, see for example [35]. Thus all properties true for records of uniform permutations are bijectively also true for the branches of uniform recursive trees.

The third possibility to calculate the number of branches of a URT is by using the representation of URTs as permutations in cycle representation. This interpretation allows us to equate the number of branches with the number of cycles in a random permutation which is well known, see for example [25].

We can also consider the number of branches of a given size. We have that:

Theorem 3.4 ([6]). *For $m \in \mathbb{N}$, let $\beta_{m,n}$ denote the number of branches of size m in $\mathcal{T}_n$. Then*

(i) for $m \in \mathbb{N}$, $\beta_{m,n} \xrightarrow{n \to \infty} \mathrm{Po}\left(\frac{1}{m}\right)$ and

(ii) for $n > m > \frac{n-1}{2}$, $\beta_{m,n}$ can only take the values 0 or 1 and $\mathbb{P}(\beta_{m,n} = 1) = \frac{1}{m}$.

Moreover the covariance matrix for $(\beta_{m,n})_{1 \leq m \leq n}$ is derived and it is shown that the number of branches of different sizes are asymptotically independent for any sequence of integers $(m_1, \ldots, m_\ell)$, see [6].

As a final note, these results allow some conclusions concerning the size of the largest branch.

Theorem 3.5 ([6]). *Let ν_n denote the size of the largest branch of $\mathcal{T}_n$. Then*

(i) $\mathbb{P}\left(\nu_n > \frac{n-1}{2}\right) \xrightarrow{n \to \infty} \ln(2)$ and

(ii) $\nu_n \xrightarrow[a.s.]{n \to \infty} \infty$.

Also in connection to the size of specific components the number of nodes with k descendants can be studied. In [14], Devroye uses a connection between URTs and binary trees and local counters to analyse this statistic.

Theorem 3.6 ([14]). *Let $X_{k,n}$ denote the number of leaves with exactly k descendants in a URT. Moreover we define $\alpha_k = \frac{1}{(k+2)(k+1)}$, $\gamma_k = \frac{1}{(2k+3)(2k+2)(k+1)}$ and $\sigma_k = \alpha_k(1-\alpha_k) - 2(k+1)\alpha_k^2 + 2\gamma_k$. Then*

(i) $\frac{X_{k,n}}{n} \xrightarrow{n\to\infty} \alpha_k$ in probability and

(ii) $\frac{X_{k,n}-n\alpha_k}{\sqrt{n}} \xrightarrow{n\to\infty} \mathcal{N}(0,\sigma_k^2)$ in distribution.

The depth of node n can also be studied by several different methods, of which we will introduce three. Let $\mathcal{D}_n$ denote the depth of node n in a URT $\mathcal{T}_n$ of size n. It turns out that $\mathcal{D}_n$ and $\mathcal{B}_n$ have the same distribution. Again there are several methods to show that this is the case.

The first possibility is to describe the depth of node n is to consider the path P_n from 1 to n in $\mathcal{T}_n$. Every node in P_n is an ancestor of n and the number of nodes in the path, except n, thus gives the depth of the node n:

$$\mathcal{D}_n = \sum_{i=2}^{n} \mathbf{1}(i \in P_n) = \sum_{i=2}^{n-1} \mathbf{1}(i \in P_n) + 1. \tag{3.10}$$

It can then be shown that $\mathbb{P}(i \in P_n) = \frac{1}{i}$ and that the indicator random variables $\mathbf{1}(i \in P_n)$ are mutually independent, which allows to conclude that $\mathcal{D}_n$ has the same distribution as $\mathcal{B}_n$ [6].

On the other hand, the depth of node n can also be seen as follows: By construction, the parent of node n is a node a_1 uniformly distributed among $\{1,\ldots,n-1\}$. Its grandparent is then a node uniformly distributed among $\{1,\ldots,a_1-1\}$ and so on until we reach $a_m = 1$. The depth of node n is then m. Similarly, the position of the last record in a random permutation, b_1, is uniformly distributed on $\{1,\ldots,n-1\}$. The position of the 2nd to last record, b_2, is then uniformly distributed on $\{1,\ldots,b_1-1\}$

and so on, until $b_m = 1$. m then gives the number of records in the permutation. Therefore, we get

Theorem 3.7 ([5]). $\mathcal{D}_n$ *is distributed as the number of records in a uniform random permutation of* $\{2, \ldots, n\}$.

We already saw that the number of branches is distributed as the number of records in a uniform random permutation, implying that the depth and the number of branches are equally distributed as well.

There is a third method to determine the depth of node n by recursions, which was historically the first one used, see [1]. It is based on a recursion for the expected number of nodes in the k-th generation, which allows to derive the depth of node n by the following relation. Let μ_n^k denote the number of nodes in the k-th generation, where the generation of the root is 0. Then $\mathbb{P}(\mathcal{D}_n = k) = \frac{\mathbb{E}\left[\mu_{n-1}^{k-1}\right]}{n-1}$. The expectation and variance can then directly be computed from the distribution.

Finally there is another, easier way of determining the depth of node n by the use of recursion without first determining the exact distribution described in [1]. It is a corollary of a result about internodal distances, which we will now give.

In [2] the expectation and variance of the distance between two nodes in a random recursive tree are derived via a simple recursion. Let us denote the distance between i and j by $\mathcal{D}_{i,j}$. If $i < j$, the distance between i and j is longer by 1 than the distance between the node j attaches to and i. In other words if j attaches to k where $1 \leq k \leq j-1$, then we have $\mathcal{D}_{i,j} = \mathcal{D}_{i,k} + 1$. This gives the recursion

$$\mathbb{P}(\mathcal{D}_{i,j} = d) = \frac{1}{j-1}[\mathbb{P}(\mathcal{D}_{1,i} = d-1) + \mathbb{P}(\mathcal{D}_{2,i} = d-1) \\ + \cdots + \mathbb{P}(\mathcal{D}_{i,j-1} = d-1)]. \tag{3.11}$$

This recursion can be used to prove the following theorem.

Theorem 3.8 ([2]). *Let $\mathcal{D}_{i,j}$ denote the distance between i and j where $1 \leq i < j \leq n$ in a random recursive tree. Then*

$$\mathbb{E}[\mathcal{D}_{i,j}] = H_i + H_{j-1} - 2 + \frac{1}{i} \ and$$
$$\mathrm{Var}(\mathcal{D}_{i,j}) = H_i + H_{j-1} - 3H_i^{(2)} - H_{j-1}^{(2)} + 4 - 4\frac{H_i}{i} + \frac{3}{i} - \frac{1}{i^2}. \tag{3.12}$$

In [3] the exact distribution for $\mathcal{D}_{i,n}$ is given and asymptotic results for $\mathcal{D}_{i_n,n}$ for some special cases of $(i_n)_{n \in \mathbb{N}}$ are given. The asymptotic distribution of the distance between a fixed node and n is shown to be normal in [4].

Theorem 3.9 ([4]). *Let $\mathcal{D}_{i,n}$ denote the distance between nodes i and n in a URT $\mathcal{T}_n$. Then*

$$\frac{\mathcal{D}_{i,n} - \ln(n)}{\sqrt{\ln(n)}} \xrightarrow[d]{n \to \infty} \mathcal{G}. \tag{3.13}$$

Moreover, asymptotic normality of $\mathcal{D}_{i_n,n}$ for any sequence $(i_n)_{n \in \mathbb{N}}$ is shown.

Theorem 3.10 ([4]). *Let $\mathcal{D}_{i_n,n}$ denote the distance between nodes i_n and n in a URT $\mathcal{T}_n$, where $i_n \geq n - 1$ for all n. Then*

$$\frac{\mathcal{D}_{i_n,n} - \ln(n) - \ln(i_n)}{\sqrt{\ln(n) + \ln(i_n)}} \xrightarrow[d]{n \to \infty} \mathcal{G}. \tag{3.14}$$

As already mentioned in the introduction there are two main global properties of URTs that were investigated: the maximum degree and the height. First let us consider the maximum degree. Let the degree of node i in $\mathcal{T}_n$ be denoted by $\delta_{n,i}$ and Δ_n be the maximum degree of any vertex in $\mathcal{T}_n$. Since $\mathbb{E}[\delta_{n,1}] > \ln(n)$, and $\Delta_n > \delta_{n,1}$, we get $\mathbb{E}[\Delta_n] > \ln(n)$. Moreover we have the following result.

Theorem 3.11 ([15]). *Let $\mathcal{T}_n$ be a URT and let Δ_n denote the maximum degree among the vertices of $\mathcal{T}_n$. Then,*

$$\frac{\Delta_n}{\ln(n)} \xrightarrow[a.s.]{n\to\infty} 1 \qquad and \qquad \frac{\mathbb{E}[\Delta_n]}{\ln(n)} \xrightarrow{n\to\infty} 1. \tag{3.15}$$

For results on the distribution of the maximum degree, see [16, 17].

Concerning the height of a URT the following is known.

Theorem 3.12 ([20]). *Let $\mathcal{T}_n$ be a URT and let $\mathcal{H}_n$ denote the height of $\mathcal{T}_n$ i.e. the longest path from the root to any other node. Then with probability 1*

$$\frac{\mathcal{H}_n}{\ln(n)} \xrightarrow{n\to\infty} e. \tag{3.16}$$

In [20] this result is proved by using the connection between a continuous time branching process and URTs. In [21] results about the height of a more general family of random recursive trees are obtained, implying in particular the above result, by using a second moment method.

Moreover we have a more precise estimation of the expectation of the height of a uniform recursive tree, again derived by the use of a branching process.

Theorem 3.13 ([46]). *Let $\mathcal{T}_n$ be a URT and let $\mathcal{H}_n$ denote the height of $\mathcal{T}_n$. Then*

$$\mathbb{E}[\mathcal{H}_n] = e\ln(n) - \frac{3}{2}\ln\ln(n) + \mathcal{O}(1). \tag{3.17}$$

3.3. Non-uniform Recursive Trees Considered in the Literature

There has been a tremendous effort to study uniform recursive trees as the results above imply. There also are a few papers considering different kinds of non-uniform

recursive trees. We now give some examples of such distributions.

3.3.1. Non-uniform Recursive Trees via External Nodes

In [26] recursive trees are described by the use of external nodes. By variations on the number of external nodes each node has, one can thus change the attachment probabilities and the distribution on the recursive trees. In general, recursive trees can be constructed by attaching node n according to some rule to a node of a recursive tree of size $n-1$. In order to specify that rule one can add external nodes at every place the new node can be attached to and then choose one of these according to some distribution. In the uniform model there is one external node at every node of $\mathcal{T}_{n-1}$ and one of them is chosen uniformly. This node then becomes the internal node n and a new external node is created at n and at the parent of n.

This model can be generalized by adding at each step n, after attaching node n, α external nodes at the parent of n and β external nodes at n. Among all the external nodes the place for $n+1$ is then again chosen uniformly. Using this principle the following distributions can for example be obtained. For plane-oriented trees the number of external nodes of a node i is equal to the out-degree of i +1. This model is called plane-oriented because it corresponds to the distribution obtained by considering the children of each node as ordered, then choosing one of these trees uniformly and finally identifying all trees that only differ because of the ordering of the children in order to get a distribution on recursive trees [1]. This model was introduced in [27] as one of the first non-uniform models. In that paper moreover properties related to the degrees of nodes where obtained and compared to the uniform case. For $m \in \mathbb{N}$, m-oriented trees are a generalization of this model where every node has $m-1$ times its outdegree +1 external nodes.

Binary trees are a tree model where every node can have exactly 2 children. For $m \in \mathbb{N}$, m-ary trees are a generalization of binary trees where every node can have exactly m children. These trees can also be defined via the use of external nodes. Table 3.1 shows the α and β values for these tree models [26].

Table 3.1. Distributions on recursive trees for some values of α and β.

Recursive tree model	α	β
Uniform recursive	1	1
Plane-oriented	2	1
m-oriented	m	1
Binary	0	2
m-ary	0	m

Figure 3.4 demonstrates the construction of a 2-oriented tree with the use of external nodes. The nodes without label are external nodes. At each step one of the external nodes is chosen uniformly chosen as the place of the newly added node.
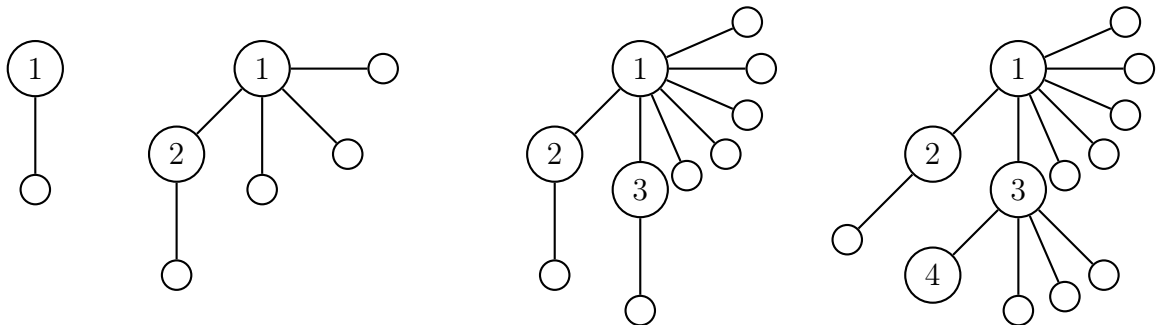
Figure 3.4. Example of the construction of a 2-oriented tree with external nodes.

Dobrow and Smythe then give a recursive formula for the distribution of $\mathcal{D}_{i,n}$, the distance between node i and node n, and the exact distribution of $\mathcal{D}_{1,n}$, the depth of node n, for general α and β. By using Poisson approximations they then prove asymptotic normality of $\mathcal{D}_{i,n}$ for all the models introduced above.

3.3.2. Introducing Choice

Instead of altering the distribution on the set of recursive trees by introducing external nodes, it is possible to introduce choice according to some criterion. The idea is to choose at each step several potential parents one of which is then chosen according

to some criterion. In [47], k possible parents are selected for each node. The node is then attached to the one among these with the smallest or biggest distance to the root or the lowest or highest degree. These are called the *smallest-depth*, *highest-depth*, *lowest-degree* and *highest-degree model* respectively. The relevant properties of these models are then shown to qualitatively differ from the uniform one.

While for both m-oriented trees and the highest degree model the choice of the parent depends on the degree, the two models differ considerably. In the construction of plane-oriented trees knowledge about the degrees of all nodes is necessary. On the other hand in the highest-degree model only knowledge about the degree of k nodes is necessary at each step. The authors of [47] refer to this as global vs. local knowledge.

In [48] a different choice criterion is introduced because the ones investigated in [47] are intractable in many cases. They propose to either choose the node with the minimal or the maximal label among the k potential parents and call this the *label model*. In addition to giving the exact distribution of the depths in these trees, they also show that these label models can be used as bounds for the depth models introduced in [47].

3.3.3. Scaled Attachment Random Recursive Trees

Another generalization of uniform random recursive trees, that goes into a very different direction are *scaled attachment random recursive trees*, abbreviated SARRT, which were introduced in [21]. To construct an SARRT with $n + 1$ nodes, we need n identically distributed independent random variables $X_1, \ldots, X_n$ on $[0, 1)$. The root gets label 0. The parent of i is then chosen as $\lfloor iX_i \rfloor$. Like the uniform recursive tree, the attachment probabilities do not depend on structural properties of the tree constructed before node i is attached. This was not the case for the recursive tree models introduced in [26], since there the attachment probability depends on the number of children of the already present nodes. The models introduced in [47] also relies on knowledge of the structure of $\mathcal{T}_{n-1}$ in order to choose the parent of n since we need to know the depth or degree of the potential parents. SARRTs moreover include URTs: when we choose

the uniform distribution for the X_i, the random variable $\lfloor iX_i \rfloor$ is uniformly distributed over $[i]$.

The explicit example given in [21] is the SARRT where $X_i = \max\{U_1, U_2, \ldots, U_k\}$ or $X_i = \min\{U_1, U_2, \ldots, U_k\}$ with $k \in \mathbb{N}$ and U_ℓ, $\ell = 1, \ldots, k$, uniformly distributed over $[0, 1)$. This is equivalent to choosing uniformly k potential parents for each node i and among these attaching i to the youngest respectively oldest parent, where labels with higher label are considered to be younger. In this way SARRTs recover the model introduced in [48]. In [21] the authors then go on to prove asymptotic results on the height of the tree, the depth of node n and the minimal depth of the nodes $\lceil \frac{n}{2} \rceil$ to n, for general X_i and depending on the common distribution of the X_i.

Apart from different distributions on recursive trees, recursive structures other than trees are also studied. One possibility is to consider random recursive forests, i.e. allowing disconnected components. A way in the other direction is to allow every new node to join several previously present nodes.

3.3.4. Random Recursive Forests

Another generalization of random recursive trees are random recursive forests, which were introduced in [49]. A *random recursive forest* is constructed in a similar way as a random recursive tree. The difference is that each node n can either attach to any of the nodes 1 to $n - 1$ or become the root of a new tree. In the general model, at each step n, a number $y_n \in \{0, \ldots, n-1\}$ is chosen according to probability $\vec{p} = (p_n(y))_{y=0,\ldots,n-1}$. If $y_n = 0$, node n is the root of a new tree. If $y_n \in \{1, \ldots, n-1\}$, y_n is the parent of n. See Figure 3.5 for an example of a random recursive forest.

For the general model the expectations and variances of the outdegrees of all nodes i and as a corollary the expected number of components are given. Moreover the expected number of leaves and of nodes of out-degree 1 are given. The rest of the results are given for uniform recursive forests, so forests where $p_n(y) = \frac{1}{n}$ for all $y = 0, 1, \ldots, n - 1$. For the uniform model asymptotic normality of the number of
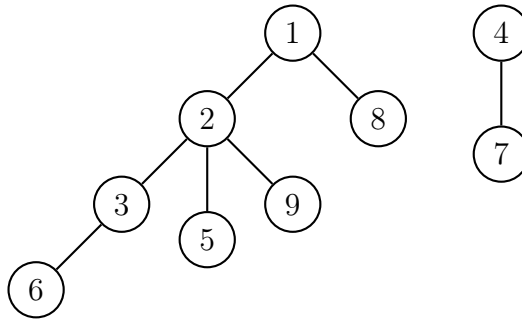
Figure 3.5. Example of a random recursive forest.

components is shown and results about the expected number of nodes of out-degree k and the maximum out-degree are derived. Moreover the distribution of the label of the root of the last component and of the label of the root of the component containing node i are given.

A uniform recursive forest of size n can easily be constructed from a uniform recursive of size $n+1$ by deleting the root. The nodes that are the start of the branches in the URT then become the roots of the components in the uniform recursive forests. Most results mentioned above thus follow more or less directly from results on URT.

3.3.5. Uniform Random Recursive Directed Acyclic Graphs

As already mentioned, we can also allow nodes to attach to several nodes. In [15] a model was introduced where initially m roots are present and every additional node attaches uniformly to r already present nodes, with r and m in $\mathbb{N}$. This yields a *uniform random recursive directed acyclic graph*, short URRD, which can be used to model circuits [21]. In [15] asymptotic results about the maximum degree of such structures are derived. Similarly as for recursive trees, various distributions different from the uniform one can be considered on recursive directed acyclic graphs. When letting the attachment probability depend on the degree of the present nodes, this in particular yields the preferential attachment model [21]. Also see [50] for results on paths in uniform random recursive acyclic graphs.

3.3.6. Hoppe Trees

Another non-uniform distribution on random recursive trees considered in the literature are Hoppe trees. These trees were introduced by Leckey and Neininger on the basis of Hoppe's urn [28]. In Hoppe's urn one starts with a black ball of weight θ. All other balls will have weight 1. We will call all balls that are not black coloured. At each step one of the balls is drawn from the urn with a probability proportional to its weight. The drawn ball is always put back. Moreover, when the black ball is drawn, a ball with a new colour is added to the urn. When a coloured ball is drawn, a new ball of the same colour is added to the urn.

On the basis of this process a random recursive tree called *Hoppe tree* can be constructed: The black ball represents the root and gets label 1. The other balls get their label from the time they were added to the urn. If we consider the black ball as the first ball added to the urn, the n-th ball added to the urn thus gets label n. The label of the ball we draw at time n is the label of the parent of node n in the recursive tree. Each colour in the urn then corresponds to one branch of the tree.

At each time n the probability to draw the black ball is $\frac{\theta}{\theta+n-2}$ and the probability for any coloured ball to be drawn is $\frac{1}{\theta+n-2}$. We thus get a random recursive tree with the following construction principle: let $c_{i,n}$ denote the event that i is the parent of n, then

$$\mathbb{P}(c_{i,n} = 1) = \begin{cases} \frac{\theta}{\theta+n-2} & \text{for } i = 1 \\ \frac{1}{\theta+n-2} & \text{for } i = 2, \ldots, n-1. \end{cases} \tag{3.18}$$

The Hoppe distribution on recursive trees is equivalent to the θ-biased distribution on permutations as described in [51]. We will call these permutations *Hoppe permutations*. θ-biased permutations can be constructed by using a variation of the Chinese restaurant process described in Section 3.1.3. At each step, i either starts a new cycle with probability $\frac{\theta}{\theta+i-2}$ or joins any other cycle after any other integer with

probability $\frac{1}{\theta+i-2}$. The number of cycles of θ-biased permutations and the distribution of the cycle sizes are known and thus imply the corresponding results for branches of Hoppe trees, see [51, 52]. In particular we have the following theorem.

Theorem 3.14 ([51, 52]). *Let $\mathcal{B}_n^\theta$ denote the number of branches of a Hoppe tree of size n. Then*

$$\mathbb{E}\left[\mathcal{B}_n^\theta\right] = \theta \sum_{i=1}^n \frac{1}{\theta + i - 1} \tag{3.19}$$

and

$$\mathrm{Var}\left(\mathcal{B}_n^\theta\right) = \theta \sum_{i=1}^n \frac{n-1}{(\theta + j - 1)^2}. \tag{3.20}$$

Moreover asymptically

$$\frac{\mathcal{B}_n^\theta - \theta \ln(n)}{\sqrt{\theta \ln(n)}} \xrightarrow[d]{n \to \infty} \mathcal{G}. \tag{3.21}$$

In [28], results about the depth of node n and the height, the number of leaves and the internal path length of a Hoppe tree of n nodes are given. We will now give the theorems from [28] that concern statistics we will also investigate.

Theorem 3.15 ([28]). *Let $\mathcal{D}_n^\theta$ denote the depth of the n-th node of a Hoppe tree and let, for $i = 1, \ldots, n - 2$, B_i be independent Bernoulli random variables with $\mathbb{P}(B_i = 1) = \frac{1}{\theta+1}$. Then*

$$\mathcal{D}_n^\theta =_d 1 + \sum_{i=1}^{n-2} B_i. \tag{3.22}$$

This easily gives

$$\mathbb{E}\left[\mathcal{D}_n^\theta\right] = \ln(n) + \mathcal{O}(1)$$

$$\mathrm{Var}\left(\mathcal{D}_n^\theta\right) = \ln(n) + \mathcal{O}(1)$$

$$\frac{\mathcal{D}_n^\theta - \mathbb{E}[\mathcal{D}_n^\theta]}{\sqrt{\mathrm{Var}(\mathcal{D}_n^\theta)}} \xrightarrow[d]{n\to\infty} \mathcal{G} \ \ and$$

$$d_{TV}(\mathcal{D}_n^\theta, \mathrm{Poisson}(\mathbb{E}\left[\mathcal{D}_n^\theta\right])) = \mathcal{O}\left(\frac{1}{\ln(n)}\right). \tag{3.23}$$

Theorem 3.16 ([28]). *Let $\mathcal{H}_n^\theta$ denote the height of a Hoppe tree with n nodes. Then*

$$\mathbb{E}\left[\mathcal{H}_n^\theta\right] = e\ln(n) - \frac{3}{2}\ln\ln n + \mathcal{O}(1) \ \ and$$

$$\mathrm{Var}\left(\mathcal{H}_n^\theta\right) = \mathcal{O}(1). \tag{3.24}$$

Theorem 3.17 ([28]). *Let $\mathcal{L}_n^\theta$ denote the number of leaves of a Hoppe tree with $n \geq 2$ nodes. Then*

$$\mathbb{E}\left[\mathcal{L}_n^\theta\right] = \frac{n}{2} + \frac{\theta-1}{2} + \mathcal{O}\left(\frac{1}{n}\right)$$

$$\mathrm{Var}\left(\mathcal{L}_n^\theta\right) = \frac{n}{12} + \frac{\theta-1}{12} + \mathcal{O}\left(\frac{1}{n}\right)$$

$$\mathbb{P}(|\mathcal{L}_n^\theta - \mathbb{E}\left[\mathcal{L}_n^\theta\right]| \geq t) \leq 2e^{-\frac{6t^2}{n+\theta+1}} \ \ for\ all\ t > 0 \ \ and$$

$$\frac{\mathcal{L}_n^\theta - \mathbb{E}\left[\mathcal{L}_n^\theta\right]}{\sqrt{\mathrm{Var}\left(\mathcal{L}_n^\theta\right)}} \xrightarrow[d]{n\to\infty} \mathcal{G}. \tag{3.25}$$

The first distribution we will consider in this thesis is a generalization of Hoppe trees so this brings us to the next chapter.

4. WEIGHTED RECURSIVE TREES

4.1. Definition

We now consider a recursive tree model obtained by varying the weights of each node. In a URT the weight of each node can be considered to be 1. We now give each node $i = 1, \ldots, n$ a weight $\omega_i \in \mathbb{R}$, thus every weighted recursive tree model can be characterized by its sequence of weights, denoted $(\omega_i)_{i \in \mathbb{N}}$. These weights affect the attachment probabilities: let $c_{i,j}$ denote the event that j is attached to i in the j-th construction step. Then for $i < j$ we have

$$\mathbb{P}(c_{i,j} = 1) = \frac{\omega_i}{\sum_{k=1}^{j-1} \omega_k}. \tag{4.1}$$

We will call these trees weighted recursive tree, or WRT. This is a generalization of Hoppe trees, which correspond to the case $\omega_1 = \theta > 0$ and $\omega_i = 1$ for $i = 2, \ldots, n$.

WRTs share with URTs the property that the construction steps are mutually independent: whether j attaches to i is not dependent on the structure of the tree at time j, but only on the weight sequence. We already mentioned this property above and saw that this is not true for several distributions considered in the literature. For example for the trees considered by Dobrow and Smythe in [26] where α, i.e. the number of external nodes created at the parent, was different from 1, the number of external nodes of node i, and thus the attachment probabilities, were dependent on the outdegree of i at time j. The second model we will consider, biased recursive trees, does not have this independence property either.

Because the construction steps in weighted recursive trees are mutually independent, the weighted recursive tree model is a kind of a more general model, which we call *inhomogeneous trees*, where j attaches with probability $p_{j,i}$ to a node i, such that $\sum_{i=1}^{j-1} p_{j,i} = 1$. In some sense, in inhomogeneous trees, the weight of each node can change at each step, while in weighted recursive trees the weight is fixed.

Because of technicalities we will not be able to tackle all problems for a general sequence of weights. Where it will be necessary we will make some assumptions on the weight sequence $(\omega_i)_{i\in\mathbb{N}}$, which will be specified at the beginning of each section or subsection. For the number of leaves, for example, we will assume that for some $k \in \mathbb{N}$, $\omega_i = 1$ for $i > k$ and $\omega_i = \theta$ for $1 \leq i \leq k$.

In the rest of this section we will proceed as follows: In Section 4.2 we will discuss a method of constructing WRTs from URTs. It can be applied to any weighted recursive tree, but is much easier in the case where only the first k nodes get a weight θ that is different from 1. We then go on by giving results about the number of branches and the depth of node n of general weighted recursive trees. In both cases we will be able to write the random variable in question as a sum of independent Bernoulli random variables. As we will see the depth of node n and the number of branches are not equidistributed, as it is the case for URTs. We then derive results on the number of leaves of a special case of weighted recursive trees, where only the first k nodes get a weight different from 1. This is done by using a martingale approach. Finally we introduce a second coupling that can only be applied on weighted recursive trees where for some $k \in \mathbb{N}$, $\omega_i = 1$ for $i > k$. The counterpart of this restriction is that it is much simpler and deterministic and thus allows for easy conclusions concerning the asymptotic behaviour of the number of leaves of weighted recursive tree of this form.

4.2. Constructing Weighted Recursive Trees from Uniform Recursive Trees

We begin our discussion of weighted recursive trees by presenting a coupling construction that can be used to generate a WRT once we are given a URT. Given a URT we will relocate some nodes with a certain probability. By relocation we mean that a node j is detached from its parent and attached to another node with all its descendants. In Figure 4.1 such a relocation is illustrated.

The coupling construction can be applied step by step, as the new edges appear, or at the end, after the whole URT is constructed. In other words, we can either at
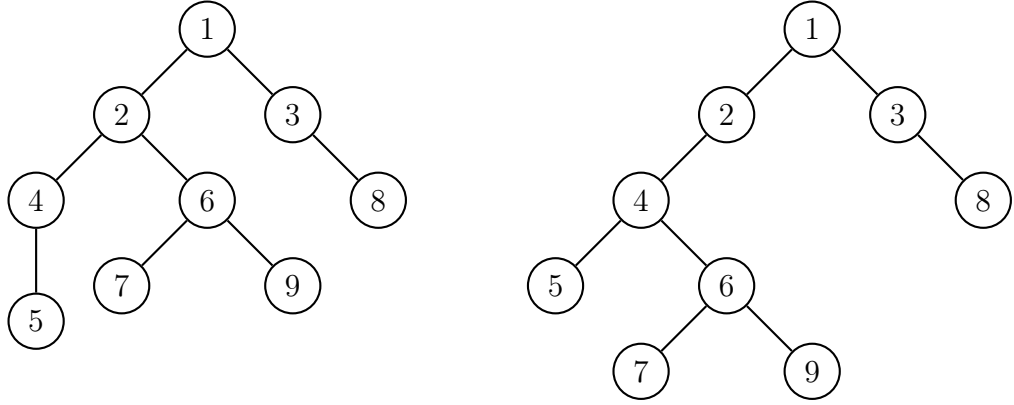
Figure 4.1. Node 6 is relocated from node 2 to node 4.

each step attach a new node uniformly to any of the present nodes and then relocate it according to the rule we will now describe, or we can construct a URT with n nodes and then apply the relocation rule for every node. This amounts to the same result since the structure of the tree at any time does not affect which node is chosen as the parent of the next node.

Let us further remark that in the latter case we can relocate the nodes in any order we like, since we always relocate the node with all its descendants. As we will see, the relocation process only depends on the weight of the old and the new parent, and is thus independent of any structural properties of the tree. Thus this coupling has the property that all adjustments we make, i.e. all relocations, are mutually independent. So we do not lose the important property of WRTs and URTs that each construction step is independent while rearranging one to get the other.

4.2.1. General Case

First construct a URT. We will rearrange it by relocating nodes in order to get a WRT with weight sequence $(\omega_i)_{i \in \mathbb{N}}$. For any two nodes $3 \leq i < j$, node j can only be either relocated to or from node i, not both. This depends on whether its weight is bigger or smaller than the average weight at time j, i.e. among the first $j - 1$ nodes. More precisely, let us denote the average weight at time j by $\overline{\omega_j} := \frac{\sum_{k=1}^{j-1} \omega_k}{j-1}$. If $\omega_i < \overline{\omega_j}$ node j can only be relocated *from* i and if $\omega_i > \overline{\omega_j}$, node j can only be relocated

to node i. This restriction is suggested by the fact that $\omega_i > \overline{\omega_j}$ is equivalent to $\frac{1}{j-1} < \frac{\omega_i}{\omega_1 + \cdots + \omega_{j-1}}$. We will show that under this condition we can construct a coupling producing a WRT with weight sequence $(\omega_i)_{i \in \mathbb{N}}$.

For $j \geq 3$, if in the URT j is attached to a node i with $\omega_i < \overline{\omega_j}$, we relocate it with probability $r_{i,j}$ to another node such that

$$
\begin{aligned}
&\mathbb{P}(j \text{ at } i \text{ in WRT}) = \mathbb{P}(j \text{ at } i \text{ in reconstructed tree}) \\
&\Leftrightarrow \mathbb{P}(j \text{ at } i \text{ in WRT}) = \mathbb{P}(j \text{ at } i \text{ in URT})(1 - \mathbb{P}(j \text{ is relocated from } i)) \\
&\Leftrightarrow \mathbb{P}(j \text{ at } i \text{ in WRT})) = \mathbb{P}(j \text{ at } i \text{ in URT})(1 - r_{i,j}) \\
&\Leftrightarrow \frac{\omega_i}{\sum_{k=1}^{j-1} \omega_k} = (1 - r_{i,j})\frac{1}{j-1} \\
&\Leftrightarrow r_{i,j} = 1 - \frac{(j-1)\omega_i}{\sum_{k=1}^{j-1} \omega_k} \\
&\Leftrightarrow r_{i,j} = \frac{\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_i}{\sum_{k=1}^{j-1} \omega_k}.
\end{aligned}
\tag{4.2}
$$

We only get a positive value for $r_{i,j}$ if $\omega_i < \frac{\sum_{k=1}^{j-1} \omega_k}{j-1}$, which is consist with our assumption that we only relocate from nodes whose weight is strictly less than the average at time j. It is important to remember that the average of the weights can change at each step.

Now we want to relocate j to a node that has a weight larger than the average in such a way that we get the right attachment probabilities. Let $p_{i,j,h} = \mathbb{P}(j \text{ relocated to } i | j \text{ relocated from } h)$. We will see that we get the right probabilities if we choose

$$
p_{i,j,h} = \frac{(j-1)\omega_i - \sum_{k=1}^{j-1} \omega_k}{\sum_{\substack{\ell=1 \\ \omega_\ell > \overline{\omega_j}}}^{j-1} \left[(j-1)\omega_\ell - \sum_{k=1}^{j-1} \omega_k\right]}.
\tag{4.3}
$$

That $p_{i,j,h}$ does not depend on h implies that given a node j is relocated it does not matter which node it is relocated from. We thus get

$\mathbb{P}(j$ at i in reconstructed tree$)$

$= \mathbb{P}(j$ at i in URT$) + \mathbb{P}(j$ relocated to $i)$

$= \mathbb{P}(j$ at i in URT$) + \sum_{\substack{h=1 \\ \omega_h < \omega_j}}^{j-1} \mathbb{P}(j$ at h in URT$)\mathbb{P}(j$ relocated $|j$ at h in URT$)$

$$\cdot \mathbb{P}(j \text{ relocated to } i | j \text{ relocated from } h)$$

$$= \frac{1}{j-1} + \sum_{\substack{h=1 \\ \omega_h < \omega_j}}^{j-1} \frac{1}{j-1} r_{h,j} p_{i,j,h}$$

$$= \frac{1}{j-1} + \sum_{\substack{h=1 \\ \omega_h < \omega_j}}^{j-1} \frac{1}{j-1} \frac{\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_h}{\sum_{k=1}^{j-1} \omega_k} \frac{(j-1)\omega_i - \sum_{k=1}^{j-1} \omega_k}{\sum_{\substack{\ell=1 \\ \omega_\ell > \omega_j}}^{j-1} \left[(j-1)\omega_\ell - \sum_{k=1}^{j-1} \omega_k\right]} \tag{4.4}$$

$$= \frac{1}{j-1} \left[1 + \frac{\sum_{\substack{h=1 \\ \omega_h < \omega_j}}^{j-1} \left[\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_h\right]}{\sum_{k=1}^{j-1} \omega_k} \frac{(j-1)\omega_i - \sum_{k=1}^{j-1} \omega_k}{\sum_{\substack{\ell=1 \\ \omega_\ell > \omega_j}}^{j-1} \left[(j-1)\omega_\ell - \sum_{k=1}^{j-1} \omega_k\right]} \right]$$

$$= \frac{1}{j-1} \left[1 + \frac{1}{\sum_{k=1}^{j-1} \omega_k} \frac{(j-1)\omega_i - \sum_{k=1}^{j-1} \omega_k}{1} \right]$$

$$= \frac{1}{j-1} \left[1 + \frac{(j-1)\omega_i}{\sum_{k=1}^{j-1} \omega_k} - \frac{\sum_{k=1}^{j-1} \omega_k}{\sum_{k=1}^{j-1} \omega_k} \right]$$

$$= \frac{\omega_i}{\sum_{k=1}^{j-1} \omega_k}$$

$$= \mathbb{P}(j \text{ at } i \text{ in WRT}).$$

In the sixth line we used that

$$
\begin{aligned}
\sum_{\substack{h=1 \\ \omega_h < \overline{\omega_j}}}^{j-1} & \left[\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_h \right] - \sum_{\substack{l=1 \\ \omega_\ell > \overline{\omega_j}}}^{j-1} \left[(j-1)\omega_\ell - \sum_{k=1}^{j-1} \omega_k \right] \\
&= \sum_{\substack{h=1 \\ \omega_h < \overline{\omega_j}}}^{j-1} \left[\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_h \right] + \sum_{\substack{\ell=1 \\ \omega_\ell > \overline{\omega_j}}}^{j-1} \left[\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_\ell \right] \\
&\quad + \sum_{\substack{h=1 \\ \omega_h = \overline{\omega_j}}}^{j-1} \left[\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_h \right] \\
&= \sum_{h=1}^{j-1} \left[\sum_{k=1}^{j-1} \omega_k - (j-1)\omega_h \right] \\
&= (j-1) \sum_{k=1}^{j-1} \omega_k - \sum_{h=1}^{j-1} (j-1)\omega_h \\
&= 0.
\end{aligned}
$$

(4.5)

Thus the probabilities for the reconstructed tree correspond to the ones in the weighted recursive tree we wanted to get. The values of $r_{i,j}$ and $p_{i,j,h}$ imply that, if $\omega_i = \overline{\omega_j}$, node j is neither relocated from nor to i.

4.2.2. Special Case: When the First k Nodes Have Weight θ

Let now $(\omega_i)_{i \in \mathbb{N}}$ be such that for some $k \in \mathbb{N}$, $\theta \in \mathbb{R}^+$, $\omega_i = \theta$ for $i \leq k$ and $\omega_i = 1$ for $i > k$. As mentioned above we can either first construct a URT and then rearrange all nodes or relocate the new node at each step. Since the first k nodes all have the same weight, for $1 \leq i < j < k + 1$ we have

$$
\mathbb{P}(j \text{ attaches to } i) = \frac{\theta}{(j-1)\theta} = \frac{1}{j-1}. \tag{4.6}
$$

Hence we don't need to change anything for the first $k + 1$ nodes.

We now need to differentiate between the case $\theta > 1$ and $\theta < 1$, because in the first case the probability that a node attaches to the first k nodes increases in the

WRT compared to the URT and in the second case this probability decreases. So in the first case we will need to relocate nodes from the nodes $k+1, \ldots, n-1$ to the nodes $1, \ldots, k$ and in the second case the other way round. The exact rules of relocation will be specified below.

(i) Let first $\theta < 1$. We have $\mathbb{P}(j$ attaches to i in the URT$) = \frac{1}{j-1}$ and for $j > k$ and $i \leq k$ we would like to have $\mathbb{P}(j$ at i in reconstructed tree$) = \frac{\theta}{k\theta + j - k - 1}$. Let $r_{i,j}$ be the probability that j is relocated given that it is attached to some $i \leq k$. Then we want

$$\mathbb{P}(j \text{ at } i \text{ in WRT}) = \mathbb{P}(j \text{ at } i \text{ in reconstructed tree})$$

$$\Leftrightarrow \frac{\theta}{k\theta + j - k - 1} = \mathbb{P}(j \text{ at } i \text{ in URT})\mathbb{P}(j \text{ is not relocated})$$

$$\Leftrightarrow \frac{\theta}{k\theta + j - k - 1} = \frac{1}{j-1}(1 - r_{i,j}) \tag{4.7}$$

$$\Leftrightarrow 1 - r_{i,j} = \frac{\theta(j-1)}{k\theta + j - k - 1}$$

$$\Leftrightarrow r_{i,j} = \frac{k\theta + j - k - 1 - \theta(j-1)}{k\theta + j - k - 1}.$$

If the node $j > k$ is relocated from a node $i \leq k$, we attach it with uniform probability to any node $k < h \leq j - 1$. We now show that this gives the right probabilities.

Let $h > k$, then

$$\mathbb{P}(j \text{ at } h \text{ in reconstructed tree })$$

$$= \mathbb{P}(j \text{ at } h \text{ in URT})$$

$$+ \sum_{\ell=1}^{k} \mathbb{P}(j \text{ at } \ell \text{ in URT})\mathbb{P}(j \text{ relocated } |j \text{ at } l)\mathbb{P}(j \text{ relocated to } h)$$

$$= \frac{1}{j-1} + \sum_{\ell=1}^{k} \frac{1}{j-1} \frac{k\theta + j - k - 1 - (j-1)\theta}{k\theta + j - k - 1} \frac{1}{j-1-k}$$

$$= \frac{1}{j-1}\left(1 + \frac{k(k\theta + j - k - 1) - k(j-1)\theta}{(k\theta + j - k - 1)(j-1-k)}\right) \tag{4.8}$$

$$= \frac{(k\theta + j - k - 1)(j - 1 - k) + k(k\theta + j - k - 1) - k(j-1)\theta}{(j-1)(k\theta + j - k - 1)(j-1-k)}$$

$$= \frac{k\theta j - k\theta - k^2\theta + (j-k-1)^2 + k^2\theta + k(j-k-1) - kj\theta + k\theta}{(j-1)(k\theta + j - k - 1)(j-1-k)}$$

$$= \frac{(j-k-1)(j-1-k+k)}{(j-1)(k\theta + j - k - 1)(j-1-k)}$$

$$= \frac{1}{k\theta + j - k - 1}$$

$$= \mathbb{P}(j \text{ attaches to } h \text{ in WRT}).$$

(ii) Let now $\theta > 1$. In the URT $\mathbb{P}(j \text{ attaches to } i) = \frac{1}{j-1}$. But for $j > i$ and $i > k$ we want in the reconstructed tree $\mathbb{P}(j \text{ attaches to } i) = \frac{1}{k\theta+j-k-1}$. Let $r_{i,j}$ be the probability that j is relocated given that it is attached to some $i > k$. Then we want

$$\mathbb{P}(j \text{ at } i \text{ in WRT}) = \mathbb{P}(j \text{ at } i \text{ in reconstructed tree})$$

$$\Leftrightarrow \mathbb{P}(j \text{ at } i \text{ in WRT}) = \mathbb{P}(j \text{ at } i \text{ in URT})\mathbb{P}(j \text{ is not relocated})$$

$$\Leftrightarrow \frac{1}{k\theta + j - k - 1} = \frac{1}{j-1}(1 - r_{i,j}) \tag{4.9}$$

$$\Leftrightarrow 1 - r_{i,j} = \frac{(j-1)}{k\theta + j - k - 1}$$

$$\Leftrightarrow r_{i,j} = \frac{k(\theta - 1)}{k\theta + j - k - 1}.$$

If node j is relocated from a node $h > k$, we attach j uniformly to a node $1 \le i \le k$. Again we check if we get the right probabilities.

Let $i \leq k$, then

$$\mathbb{P}(j \text{ at } i \text{ in reconstructed tree })$$

$$= \mathbb{P}(j \text{ at } i \text{ in URT})$$

$$+ \sum_{\ell=k+1}^{j-1} \mathbb{P}(j \text{ at } \ell \text{ in URT})\mathbb{P}(j \text{ relocated } |j \text{ at } l)\mathbb{P}(j \text{ relocated to } i)$$

$$= \frac{1}{j-1} + \sum_{\ell=k+1}^{j-1} \frac{1}{j-1} \frac{k(\theta-1)}{k\theta+j-k-1} \frac{1}{k}$$

$$= \frac{1}{j-1} \left(1 + \frac{(j-k-1)k(\theta-1)}{(k\theta+j-k-1)k}\right) \tag{4.10}$$

$$= \frac{1}{j-1} \frac{j-1+k(\theta-1)+(j-k-1)(\theta-1)}{k\theta+j-k-1}$$

$$= \frac{1}{j-1} \frac{(j-1)\theta}{k\theta+j-k-1}$$

$$= \frac{\theta}{k\theta+j-k-1}$$

$$= \mathbb{P}(j \text{ at } i \text{ in WRT}).$$

We will now start to study some tree statistics for WRTs. All probabilities in the next sections refer to probabilities in WRTs.

4.3. The Number of Branches

As before the number of branches is denoted by $\mathcal{B}_n^\omega$ and is equal to the number of nodes attaching to 1. We have the following theorem

Theorem 4.1. *Let $\mathcal{B}_n^\omega$ denote the number of branches in a WRT with weight sequence $(\omega_i)_{i\in\mathbb{N}}$. Then*

$$\mathbb{E}[\mathcal{B}_n^\omega] = \omega_1 \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \omega_k} \tag{4.11}$$

and

$$\text{Var}\left(\mathcal{B}_n^\omega\right) = \sum_{i=1}^{n-1} \frac{\omega_1}{\sum_{k=1}^{i} \omega_k} - \sum_{i=1}^{n-1} \frac{\omega_1^2}{(\sum_{k=1}^{i} \omega_k)^2} = \omega_1 \sum_{i=2}^{n-1} \frac{\sum_{k=2}^{i} \omega_k}{(\sum_{k=1}^{i} \omega_k)^2}. \tag{4.12}$$

Proof. Since the construction steps are independent we can write $\mathcal{B}_n^\omega$ as a sum of independent Bernoulli random variables $b_i^\omega = \mathbf{1}(\text{node } i \text{ attaches to node } 1)$. Thus

$$\mathcal{B}_n^\omega = \sum_{i=2}^n b_i^\omega \tag{4.13}$$

and

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{B}_n^\omega\right] &= \sum_{i=2}^n \mathbb{E}[b_i^\omega] = \sum_{i=2}^n \mathbb{P}(i \text{ attaches to } 1) = \sum_{i=2}^n \frac{\omega_1}{\sum_{k=1}^{i-1}\omega_k} \\
&= \omega_1 \sum_{i=2}^n \frac{1}{\sum_{k=1}^{i-1}\omega_k} = \omega_1 \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i}\omega_k}.
\end{aligned}
\tag{4.14}
$$

By independence of the b_i^ω it is also easy to calculate the variance:

$$
\begin{aligned}
\mathrm{Var}\left(\mathcal{B}_n^\omega\right) &= \sum_{i=2}^n \mathbb{E}\left[(b_i^\omega)^2\right] - \sum_{i=2}^n \mathbb{E}[b_1^\omega]^2 = \sum_{i=2}^n \frac{\omega_1}{\sum_{k=1}^{i-1}\omega_k} - \sum_{i=2}^n \frac{\omega_1^2}{(\sum_{k=1}^{i-1}\omega_k)^2} \\
&= \sum_{i=2}^n \frac{\omega_1(\sum_{k=1}^{i-1}\omega_k) - \omega_1^2}{(\sum_{k=1}^{i-1}\omega_k)^2} = \omega_1 \sum_{i=3}^n \frac{\sum_{k=2}^{i-1}\omega_k}{(\sum_{k=1}^{i-1}\omega_k)^2} = \omega_1 \sum_{i=2}^{n-1} \frac{\sum_{k=2}^{i}\omega_k}{(\sum_{k=1}^{i}\omega_k)^2}.
\end{aligned}
\tag{4.15}
$$

$\square$

4.3.1. Some Examples of Weight Sequences

We now give the exact values of the expected number of branches and their variance for some examples of weight sequences. For the Hoppe tree, we write $\mathcal{B}_n^\theta$ for

the number of branches and get

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{B}_n^\theta\right] &= \theta \sum_{i=2}^{n} \frac{1}{\theta + i - 2} \\
&= \theta \sum_{i=0}^{n-2} \frac{1}{\theta + i} = \theta \sum_{i=0}^{n-2} \frac{1}{i+1} + \frac{1}{\theta+i} - \frac{1}{i+1} \\
&= \theta H_{n-1} + \theta \sum_{i=0}^{n-2} \frac{1-\theta}{(\theta+i)(i+1)} \\
&= \theta H_{n-1} + \theta(1-\theta) \sum_{i=0}^{n-2} \frac{1}{i^2 + (\theta+1)i + \theta}.
\end{aligned}
\tag{4.16}
$$

For the variance we similarly get

$$
\begin{aligned}
\operatorname{Var}\left(\mathcal{B}_n^\theta\right) &= \theta \sum_{i=3}^{n} \frac{i-2}{(\theta+i-2)^2} = \theta \sum_{i=3}^{n} \frac{1}{i-2} + \frac{i-2}{(\theta+i-2)^2} - \frac{1}{i-2} \\
&= \theta \sum_{i=1}^{n-2} \frac{1}{i} + \theta \sum_{i=1}^{n-2} \frac{i^2 - (\theta+i)^2}{(\theta+i)^2 i} \\
&= \theta H_{n-2} - \theta^3 \sum_{i=1}^{n-2} \frac{1}{i(\theta+i)^2} - 2\theta^2 \sum_{i=1}^{n-2} \frac{1}{(\theta+i)^2}.
\end{aligned}
\tag{4.17}
$$

Thus asymptotically as $n \to \infty$, we get in consistence with Theorem 3.14

$$
\mathbb{E}\left[\mathcal{B}_n^\theta\right] = \theta \ln(n) + \mathcal{O}(1) \quad \text{and} \quad \operatorname{Var}\left(\mathcal{B}_n^\theta\right) = \theta \ln(n) + \mathcal{O}(1).
\tag{4.18}
$$

For the weighted recursive tree with $\omega_i = \theta$ for $i = 1, \ldots, k$ and $\omega_i = 1$ for $i > k$, we write $T_n^{\theta^k}$ for the tree and $\mathcal{B}_n^{\theta^k}$ for the number of branches. We get

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{B}_n^{\theta^k}\right] &= \theta \sum_{i=1}^{k} \frac{1}{i\theta} + \theta \sum_{i=k+1}^{n-1} \frac{1}{k\theta + i - k} \\
&= \theta \sum_{i=1}^{n-1} \frac{1}{i} + \theta \sum_{i=k+1}^{n-1} \frac{1}{\theta(k-1)+i} - \frac{1}{i} + \sum_{i=1}^{k} \frac{1}{i} - \frac{\theta}{i} \\
&= \theta \sum_{i=1}^{n-1} \frac{1}{i} + k\theta(\theta-1) \sum_{i=k+1}^{n-1} \frac{1}{(k(\theta-1)+i)i} + (1-\theta) \sum_{i=1}^{k} \frac{1}{i}.
\end{aligned}
\tag{4.19}
$$

Also

$$
\begin{aligned}
&\mathrm{Var}\left(\mathcal{B}_n^{\theta^k}\right) \\
&= \theta \sum_{i=2}^{k} \frac{(i-1)\theta}{(i\theta)^2} + \theta \sum_{i=k+1}^{n-1} \frac{(k-1)\theta + i - k}{(k\theta + i - k)^2} \\
&= \theta \sum_{i=2}^{n-1} \frac{(i-1)}{i^2} + \theta \sum_{i=2}^{k} \frac{(i-1)\theta}{(i\theta)^2} - \theta \sum_{i=2}^{k} \frac{(i-1)}{i^2} \\
&\quad + \theta \sum_{i=k+1}^{n-1} \frac{k(\theta-1)+i-\theta}{(k(\theta-1)+i)^2} - \theta \sum_{i=k+1}^{n-1} \frac{(i-1)}{i^2} \\
&= \theta \sum_{i=2}^{n-1} \frac{(i-1)}{i^2} + (1-\theta) \sum_{i=2}^{k} \frac{(i-1)}{i^2} \\
&\quad + \theta \sum_{i=k+1}^{n-1} \left[\frac{k(\theta-1)i^2 + i^3 - \theta i^2 - k^2(\theta-1)^2 i + k^2(\theta-1)^2}{(k(\theta-1)+i)^2 i^2} \right. \\
&\qquad\qquad\quad \left. + \frac{-i^3 + i^2 - 2i^2 k(\theta-1) + 2k(\theta-1)i}{(k(\theta-1)+i)^2 i^2} \right] \\
&= \theta \sum_{i=2}^{n-1} \frac{(i-1)}{i^2} + (1-\theta) \sum_{i=2}^{k} \frac{(i-1)}{i^2} \\
&\quad + \theta \sum_{i=k+1}^{n-1} \frac{-i^2(k+1)(\theta-1) + ik(\theta-1)(k(\theta-1)+2) + k^2(\theta-1)^2}{(k(\theta-1)+i)^2 i^2} \\
&= \theta \left(H_{n-1} - H_{n-1}^{(2)} \right) + \mathcal{O}(1).
\end{aligned}
\tag{4.20}
$$

Asymptotically this gives, as for the Hoppe tree,

$$\mathbb{E}\left[\mathcal{B}_n^{\theta^k}\right] = \theta\ln(n) + \mathcal{O}(1) \text{ and } \mathrm{Var}\left(\mathcal{B}_n^{\theta^k}\right) = \theta\ln(n) + \mathcal{O}(1). \qquad (4.21)$$

Concerning more general models the following can be said. If the weights are bounded from below and above, the expectation and variance of the number of branches will still be equal to $\mathcal{O}(\ln(n))$ asymptotically.

Let for all i, $0 < m < \omega_i < M$, then

$$\omega_1 \sum_{i=1}^{n-1} \frac{1}{iM} < \omega_1 \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i}\omega_k} < \omega_1 \sum_{i=1}^{n-1} \frac{1}{im}$$

$$\frac{\omega_1}{M} H_{n-1} < \mathbb{E}\left[\mathcal{B}_n^{\omega}\right] < \frac{\omega_1}{m} H_{n-1} \qquad (4.22)$$

and

$$\omega_1 \sum_{i=2}^{n-1} \frac{(i-1)m}{(iM)^2} < \omega_1 \sum_{i=2}^{n-1} \frac{\sum_{k=2}^{i}\omega_k}{(\sum_{k=1}^{i}\omega_k)^2} < \omega_1 \sum_{i=2}^{n-1} \frac{(i-1)M}{(im)^2}$$

$$\frac{\omega_1 m}{M^2} \sum_{i=2}^{n-1} \frac{i-1}{i^2} < \mathrm{Var}\left(\mathcal{B}_n^{\omega}\right) < \frac{\omega_1 M}{m^2} \sum_{i=2}^{n-1} \frac{i-1}{i^2} \qquad (4.23)$$

$$\frac{\omega_1 m}{M^2} H_{n-1} + \mathcal{O}(1) < \mathrm{Var}\left(\mathcal{B}_n^{\omega}\right) < \frac{\omega_1 M}{m^2} H_{n-1} + \mathcal{O}(1).$$

The situation is very different when the weights are not bounded. We now give some examples of weight sequences that give a different asymptotic behaviour.

(i) Let $(\omega_i)_{i\in\mathbb{N}} = (i)_{i\in\mathbb{N}}$. In this case we get from the above formulas

$$\mathbb{E}\left[\mathcal{B}_n^{\omega}\right] = \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} k} = \sum_{i=1}^{n-1} \frac{2}{(i+1)i}$$

$$= 2\sum_{i=1}^{n-1}\left(\frac{1}{i} - \frac{1}{i+1}\right) = 2\left(1 - \frac{1}{n}\right) \xrightarrow[n\to\infty]{} 2 \qquad (4.24)$$

and

$$\begin{aligned}
\operatorname{Var}\left(\mathcal{B}_n^\omega\right) &= \sum_{i=2}^{n-1} \frac{\frac{i(i+1)}{2} - 1}{\left(\frac{i(i+1)}{2}\right)^2} \\
&= \sum_{i=1}^{n-1} \frac{\frac{i(i+1)}{2} - 1}{\left(\frac{i(i+1)}{2}\right)^2} \\
&= \sum_{i=1}^{n-1} \frac{2i(i+1)}{(i(i+1))^2} - \frac{4}{(i(i+1))^2} \\
&= \sum_{i=1}^{n-1} \frac{2}{i(i+1)} - 4\left(\frac{1}{i(i+1)}\right)^2 \\
&= \sum_{i=1}^{n-1} \frac{2}{i(i+1)} - 4\left(\frac{1}{i} - \frac{1}{i+1}\right)^2 \\
&= \sum_{i=1}^{n-1} \frac{2}{i(i+1)} - 4\left(-\frac{2}{i(i+1)} + \frac{1}{i^2} + \frac{1}{(i+1)^2}\right) \\
&= \sum_{i=1}^{n-1} \frac{10}{i(i+1)} - 4\left(\frac{1}{i^2} + \frac{1}{(i+1)^2}\right) \\
&= 10 \sum_{i=1}^{n-1}\left(\frac{1}{i} - \frac{1}{i+1}\right) - 4 \sum_{i=1}^{n-1}\left(\frac{1}{i^2} + \frac{1}{(i+1)^2}\right) \\
&= 10\left(1 - \frac{1}{n}\right) - 4 \sum_{i=1}^{n-1}\left(\frac{1}{i^2} + \frac{1}{(i+1)^2}\right) \\
&\xrightarrow{n\to\infty} 10 - 4\left(2\frac{\pi^2}{6} - 1\right) = 14 - \frac{4\pi^2}{3} < 1.
\end{aligned} \tag{4.25}$$

(ii) Let $(\omega_i)_{i\in\mathbb{N}} = \left(\frac{1}{i}\right)_{i\in\mathbb{N}}$. In this case we get

$$\mathbb{E}\left[\mathcal{B}_n^\omega\right] = \omega_1 \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^i \omega_k} = \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^i \frac{1}{k}} = \sum_{i=1}^{n-1} \frac{1}{H_i} \tag{4.26}$$

and

$$\operatorname{Var}\left(\mathcal{B}_n^\omega\right) = \sum_{i=2}^{n-1} \frac{1}{H_i} - \sum_{i=2}^{n-1} \frac{1}{H_i^2}. \tag{4.27}$$

(iii) Let $(\omega_i)_{i \in \mathbb{N}} = \left(\frac{1}{i^2}\right)_{i \in \mathbb{N}}$. In this case

$$\mathbb{E}\left[\mathcal{B}_n^\omega\right] = \omega_1 \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \omega_k} = \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \frac{1}{k^2}} \tag{4.28}$$

hence

$$\frac{6}{\pi^2}(n-1) \leq \mathbb{E}\left[\mathcal{B}_n^\omega\right] \leq n-1. \tag{4.29}$$

Also

$$\begin{aligned}
\text{Var}\left(\mathcal{B}_n^\omega\right) &= \sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \frac{1}{k^2}} - \sum_{i=1}^{n-1} \frac{1}{\left(\sum_{k=1}^{i} \frac{1}{k^2}\right)^2} \\
&= \sum_{i=2}^{n-1} \frac{1}{\sum_{k=1}^{i} \frac{1}{k^2}} \left(1 - \frac{1}{\sum_{k=1}^{i} \frac{1}{k^2}}\right).
\end{aligned} \tag{4.30}$$

Now we have on the one hand, for $0 \leq a \leq 1$, that $0 \leq a(1-a) \leq \frac{1}{4}$. On the other hand $1 - \frac{1}{\sum_{k=1}^{i} \frac{1}{k^2}}$ is increasing in i for $2 \leq i$ and $1 - \frac{1}{\sum_{k=1}^{2} \frac{1}{k^2}} = \frac{1}{5}$. Thus we get

$$\frac{6}{5\pi^2}(n-2) \leq \text{Var}\left(\mathcal{B}_n^\omega\right) \leq \frac{1}{4}(n-2). \tag{4.31}$$

4.3.2. Central Limit Theorem

As the number of branches can be written as a sum of independent random variables, we can apply Theorem 2.8, the Liapounov central limit theorem for Bernoulli random variables if $\text{Var}(\mathcal{B}_n^\omega)$ diverges.

Theorem 4.2. *If* $\text{Var}(\mathcal{B}_n^\omega)$ *diverges,* $\mathcal{B}_n^\omega$ *the number of branches of a weighted random recursive tree converges in distribution to a normal random variable:*

$$\frac{\mathcal{B}_n^\omega - \mathbb{E}[\mathcal{B}_n^\omega]}{\sqrt{\text{Var}(\mathcal{B}_n^\omega)}} \xrightarrow[d]{n \to \infty} \mathcal{G}. \tag{4.32}$$

In particular this is the case

(i) *if there is an $i > 1$ s.t. $\omega_i > 0$ and $\left(\sum_{k=1}^{i} \omega_k\right)_{i \in \mathbb{N}}$ converges*

(ii) *and if $\left(\sum_{k=1}^{i} \omega_k\right)_{i \in \mathbb{N}}$ and $\left(\sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \omega_k}\right)_{n \in \mathbb{N}}$ diverge but $\left(\sum_{i=1}^{n-1} \frac{1}{\left(\sum_{k=1}^{i} \omega_k\right)^2}\right)_{n \in \mathbb{N}}$ converges.*

Proof. In order to apply Liapounov's central limit theorem for sums of Bernoulli random variables we need to show that $\mathrm{Var}(\mathcal{B}_n^\omega)$ diverges. Now we have

$$\mathrm{Var}(\mathcal{B}_n^\omega) = \omega_1 \sum_{i=2}^{n-1} \frac{\sum_{k=2}^{i} \omega_k}{\left(\sum_{k=1}^{i} \omega_k\right)^2} = \omega_1 \sum_{i=2}^{n-1} \frac{1}{\sum_{k=1}^{i} \omega_k} \frac{\sum_{k=2}^{i} \omega_k}{\sum_{k=1}^{i} \omega_k}. \tag{4.33}$$

If $\omega_1 > 0$ and for all $k > 1$, $\omega_k = 0$ this sum is zero so we do not have convergence. That the distribution of the branches is not normal in that case is obvious because all nodes will attach to 1, so the number of branches is just $n - 1$.

Let us assume this is not the case, so there is an $i > 1$ such that $\omega_i > 0$. We now differentiate two cases. First let us assume that $\sum_{k=1}^{i-1} \omega_k$ converges to a number $a \in \mathbb{R}$. Then

$$\frac{\sum_{k=2}^{i-1} \omega_k}{\left(\sum_{k=1}^{i-1} \omega_k\right)^2} \xrightarrow[i \to \infty]{} \frac{a - \omega_1}{a^2} > 0 \tag{4.34}$$

thus $\sum_{i=2}^{n} \mathrm{Var}(b_i^\omega)$ diverges.

Let now $\sum_{k=1}^{i-1} \omega_k$ diverge. Then

$$\mathrm{Var}(\mathcal{B}_n^\omega) = \sum_{i=2}^{n} \frac{\sum_{k=2}^{i-1} \omega_k}{\left(\sum_{k=1}^{i-1} \omega_k\right)^2} = \sum_{i=2}^{n} \frac{1}{\sum_{k=1}^{i-1} \omega_k} - \frac{\omega_1}{\left(\sum_{k=1}^{i-1} \omega_k\right)^2}. \tag{4.35}$$

Thus in that case if $\sum_{i=2}^{n} \frac{1}{\sum_{k=1}^{i-1} \omega_k}$ diverges and $\sum_{i=2}^{n} \frac{1}{\left(\sum_{k=1}^{i-1} \omega_k\right)^2}$ converges, $\mathrm{Var}(\mathcal{B}_n^\omega)$ diverges. $\qquad \square$

Remark 4.3. Additionally to the statements in Theorem 4.2 we know that if the series $\sum_{i=2}^{n} \frac{1}{\sum_{k=1}^{i-1} \omega_k}$ converges, i.e. if the weights increase too fast, the variance is finite and we cannot apply Theorem 4.2. This is for example the case for $(\omega_i)_{i \in \mathbb{N}} = (i)_{i \in \mathbb{N}}$. The other case where we cannot say anything general is when $\sum_{k=1}^{i-1} \omega_k$, $\sum_{i=2}^{n} \frac{1}{\sum_{k=1}^{i-1} \omega_k}$ and $\sum_{i=2}^{n} \frac{1}{\left(\sum_{k=1}^{i-1} \omega_k\right)^2}$ all diverge. This is for example the case if $\omega_i = \frac{1}{i}$, since then $\sum_{k=1}^{i-1} \omega_k = \mathcal{O}(\ln(i))$.

Example 4.4. For the other examples above, considering the order of divergence of the sum mostly allows to decide whether the central limit theorem holds. For example, if $\sum_{k=2}^{i-1} \omega_k = \mathcal{O}(i)$, then $\frac{1}{\sum_{k=1}^{i-1} \omega_k} = \mathcal{O}(\frac{1}{i})$ and $\frac{1}{\left(\sum_{k=1}^{i-1} \omega_k\right)^2} = \mathcal{O}(\frac{1}{i^2})$ and thus $\mathrm{Var}(\mathcal{B}_n^\omega)$ diverges. This is in particular the case when $\omega_i = \theta$ for $i = 1, \ldots, k$ and $\omega_i = 1$ for $i > k$. On the other hand, if $\sum_{k=1}^{i-1} \omega_k = \mathcal{O}(i^2)$, then $\frac{\sum_{k=2}^{i-1} \omega_k}{\left(\sum_{k=1}^{i-1} \omega_k\right)^2} = \mathcal{O}(\frac{1}{i^2})$ and thus $\sum_{i=2}^{n} \mathrm{Var}(b_i^\omega)$ converges. This is for example the case for $\omega_i = i$.

We have another possibility of proving convergence to a normal random variable.

Theorem 4.5. *If*

$$\frac{\sum_{i=1}^{n-1} \left(\frac{1}{\sum_{k=1}^{i} \omega_k}\right)^2}{\sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \omega_k}} \xrightarrow{n \to \infty} 0 \tag{4.36}$$

and $\mathbb{E}\left[\mathcal{B}_n^\omega\right]$ diverges, then

$$\frac{\mathcal{B}_n^\omega - \mathbb{E}\left[\mathcal{B}_n^\omega\right]}{\sqrt{\mathbb{E}\left[\mathcal{B}_n^\omega\right]}} \xrightarrow{n \to \infty} \mathcal{G}. \tag{4.37}$$

Proof. Let $\mu_n := \mathbb{E}\left[\mathcal{B}_n^\omega\right]$. By Theorem 2.10

$$d_{TV}(\mathcal{B}_n^\omega, \mathrm{Po}(\mu_n)) \leq \min \left\{1, \left(\sum_{i=1}^{n-1} \frac{\omega_1}{\sum_{k=1}^{i} \omega_k}\right)^{-1}\right\} \sum_{i=1}^{n-1} \left(\frac{\omega_1}{\sum_{k=1}^{i} \omega_k}\right)^2$$

$$= \frac{\sum_{i=1}^{n-1} \left(\frac{\omega_1}{\sum_{k=1}^{i} \omega_k}\right)^2}{\sum_{i=1}^{n-1} \frac{\omega_1}{\sum_{k=1}^{i} \omega_k}} \tag{4.38}$$

$$= \omega_1 \frac{\sum_{i=1}^{n-1} \left(\frac{1}{\sum_{k=1}^{i} \omega_k}\right)^2}{\sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \omega_k}}.$$

Thus, if $\dfrac{\sum_{i=1}^{n-1} \left(\frac{1}{\sum_{k=1}^{i} \omega_k}\right)^2}{\sum_{i=1}^{n-1} \frac{1}{\sum_{k=1}^{i} \omega_k}} \xrightarrow{n \to \infty} 0$ and $\mathbb{E}\left[\mathcal{B}_n^\omega\right]$ diverges, we can apply Theorem 2.9 and get the result. $\square$

Remark 4.6. Theorem 4.5 implies that if for a given WRT the weight sequence satisfies $\omega_n \xrightarrow{n \to \infty} 0$, $\sum_{i=1}^{\infty} \omega_i = \infty$ and $\mathbb{E}\left[\mathcal{B}_n^\omega\right]$ diverges, then the number of branches of the corresponding WRT is asymptotically normal. This is implied by the fact that if $\omega_n \xrightarrow{n \to \infty} 0$, we have $\frac{\sum_{i=1}^{n} \omega_i^2}{\sum_{i=1}^{n} \omega_i} \xrightarrow{n \to \infty} 0$, which can be proved as follows: Let $\varepsilon > 0$ and N be such that for all $n > N$, $\omega_n < \varepsilon$. Then we have for $n > N$,

$$\begin{aligned} \frac{\sum_{i=1}^{n} \omega_i^2}{\sum_{i=1}^{n} \omega_i} &= \frac{\sum_{i=1}^{N} \omega_i^2}{\sum_{i=1}^{n} \omega_i} + \frac{\sum_{i=N+1}^{n} \omega_i^2}{\sum_{i=1}^{n} \omega_i} \\ &\leq \frac{\sum_{i=1}^{N} \omega_i}{\sum_{i=1}^{n} \omega_i^2} + \varepsilon \frac{\sum_{i=N+1}^{n} \omega_i}{\sum_{i=1}^{n} \omega_i} \leq \frac{\sum_{i=1}^{N} \omega_i^2}{\sum_{i=1}^{n} \omega_i} + \varepsilon \xrightarrow{n \to \infty} \varepsilon. \end{aligned} \tag{4.39}$$

Hence $\frac{\sum_{i=1}^{n} \omega_i^2}{\sum_{i=1}^{n} \omega_i} \xrightarrow{n \to \infty} 0$.

Remark 4.7. All of the above theorems only hold if $\mathrm{Var}\left(\mathcal{B}_n^\omega\right)$ diverges, which implies that none covers cases where $\mathbb{E}\left[\mathcal{B}_n^\omega\right]$ is finite. This follows from (4.11) and (4.12) which show that if the variance diverges, the expectation diverges to.

4.3.3. Rate of Convergence

We can use some theorems obtained by Stein's method to obtain bounds on the rate of convergence. We have

Theorem 4.8. *Let $\mathcal{B}_n^\omega$ denote the number of branches of a weighted recursive tree. Then*

$$d_W\left(\frac{\mathcal{B}_n^\omega - E[\mathcal{B}_n^\omega]}{\sqrt{\mathrm{Var}(\mathcal{B}_n^\omega)}}, \mathcal{G}\right) \leq \frac{1}{\sqrt{\mathrm{Var}(\mathcal{B}_n^\omega)}} \frac{\sqrt{28} + \sqrt{\pi}}{\sqrt{\pi}}. \tag{4.40}$$

This bound decreases to 0 if and only if $\mathrm{Var}(\mathcal{B}_n^\omega)$ diverges, which is the same criterion we had for the CLT.

Proof. In order to apply Theorem 2.12 we will consider $Y_i = b_{i+1}^\omega - \mathbb{E}[b_{i+1}^\omega]$ and hence $Y = \frac{\sum_{i=2}^n b_i^\omega - \mathbb{E}[b_i^\omega]}{\sqrt{\mathrm{Var}(\mathcal{B}_n^\omega)}} = \frac{\mathcal{B}_n^\omega - \mathbb{E}[\mathcal{B}_n^\omega]}{\sqrt{\mathrm{Var}(\mathcal{B}_n^\omega)}}$. First of all the Y_i are mutually independent thus giving $D = 1$. Now for $i = 1, \ldots, n-1$,

$$
\begin{aligned}
\mathbb{E}\left[|Y_i|^3\right] &= \mathbb{E}\left[\left|b_i^\omega - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right|^3\right] \\
&= \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^3 \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}} \\
&\quad + \left(\frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^3 \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right) \\
&= \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right) \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}} \\
&\quad \cdot \left[\left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^2 + \left(\frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^2\right] \\
&= \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right) \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}} \\
&\quad \cdot \left[1 - 2\frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)\right] \\
&< \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right) \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}} \\
&= \mathrm{Var}(b_i^\omega).
\end{aligned}
\tag{4.41}
$$

We used that for $0 < a < 1$, $0 < a(1-a) \leq \frac{1}{4}$, which implies $1 > 1 - 2a(1-a) > \frac{1}{2}$. Hence we get

$$\frac{D^2}{\sigma^3} \sum_{i=1}^{n-1} \mathbb{E}\left[|Y_i|^3\right] < \frac{1}{\sigma^3} \sum_{i=2}^n \mathrm{Var}(b_i^\omega) = \frac{1}{\sigma}. \tag{4.42}$$

Similarly

$$\mathbb{E}\left[Y_i^4\right] = \mathbb{E}\left[\left(b_i^\omega - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^4\right]$$

$$= \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^4 \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}$$

$$+ \left(\frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^4 \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)$$

$$= \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right) \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}$$

$$\cdot \left[\left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^3 + \left(\frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)^3\right] \qquad (4.43)$$

$$= \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right) \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}$$

$$\cdot \left[1 - 3\frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right)\right]$$

$$< \left(1 - \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}\right) \frac{\omega_1}{\omega_1 + \cdots + \omega_{i-1}}$$

$$= \mathrm{Var}(b_i^\omega).$$

We again used that $0 < a(1-a) < \frac{1}{4}$ for $0 < a < 1$, which implies $1 > 1 - 3a(1-a) > \frac{1}{4}$. So we get

$$\frac{\sqrt{28}D^{\frac{3}{2}}}{\sqrt{\pi}\sigma^2}\sqrt{\sum_{i=1}^{n-1}\mathbb{E}\left[Y_i^4\right]} < \frac{\sqrt{28}}{\sqrt{\pi}\sigma^2}\sqrt{\sum_{i=2}^{n}\mathrm{Var}(b_i^\omega)} = \frac{\sqrt{28}}{\sqrt{\pi}\sigma}. \qquad (4.44)$$

Hence in total we have

$$d_W\left(\frac{\mathcal{B}_n^\omega - \mathbb{E}[\mathcal{B}_n^\omega]}{\sigma}, \mathcal{G}\right) \leq \frac{1}{\sigma} + \frac{\sqrt{28}}{\sqrt{\pi}\sigma}. \qquad (4.45)$$

$\square$

Finally we will now give a result concerning the size of branches in Hoppe trees. We hope to generalize it to more general WRT models in the future.

4.3.4. Largest Branch in a Hoppe Tree

We denote by $\mathcal{B}_{n,i}$ the number of branches of size i in $\mathcal{T}_n$, a recursive tree of size n. Now we define

$$\nu_n(\mathcal{T}_n) := \max\{i \in [n-1] : \mathcal{B}_{n,i} \geq 1\} \tag{4.46}$$

to be the number of nodes in the largest branch of that tree. In [6], it was shown that

$$\lim_{n\to\infty} \mathbb{P}\left(\nu_n(\mathcal{T}_n) \geq \frac{n}{2}\right) = \ln 2 \tag{4.47}$$

when $\mathcal{T}_n$ is a URT on n vertices. The purpose of this section and the next theorem is to extend the result of [6] to Hoppe trees, and to provide more details about the asymptotic distribution, via exploiting the relation between Hoppe trees and Hoppe permutations, which were introduced in 3.3.6. Further, the result in (4.47) is now extended to an explicit expression for the limit $\lim_{n\to\infty} \mathbb{P}\left(\nu_n(\mathcal{T}_n) \geq cn\right)$ for $c \in [1/2, 1]$.

Theorem 4.9. *(i.) Let $\mathcal{T}_n^\theta$ be a Hoppe tree. Then $\frac{\nu_n(\mathcal{T}_n^\theta)}{n}$ converges weakly to a random variable ν whose cumulative distribution function is given by*

$$F_\theta(x) = e^{\gamma\theta} x^{\theta-1} \Gamma(\theta) p_\theta(1/x) \text{ for } x > 0 \tag{4.48}$$

where γ is Euler's constant and

$$p_\theta(x) = \frac{e^{-\gamma\theta} x^{\theta-1}}{\Gamma(\theta)} \left(1 + \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k!} \int \cdots \int_{\mathcal{S}_k(x)} \left(1 - \sum_{j=1}^{k} y_j\right)^{\theta-1} \frac{dy_1 \cdots dy_k}{y_1 \cdots y_k}\right) \tag{4.49}$$

with

$$\mathcal{S}_k(x) = \left\{y_1 > \frac{1}{x}, \ldots, y_k > \frac{1}{x}, \sum_{j=1}^{k} y_j < 1\right\}. \tag{4.50}$$

(ii) When $\theta = 1$, we obtain the following for the largest branch in a URT: $\frac{\nu_n(\mathcal{T}_n)}{n}$ converges weakly to a random variable ν whose cumulative distribution function is given by

$$F_1(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int \cdots \int_{\mathcal{S}_k(x)} \frac{dy_1...dy_k}{y_1...y_k} & \text{if } x \in [0,1] \\ 1 & \text{if } x > 1 \end{cases} \tag{4.51}$$

where $\mathcal{S}_k(x)$ is as before.

Also, for any c in $\left[\frac{1}{2}, 1\right]$, we have

$$\lim_{n \to \infty} \mathbb{P}(\nu_n(\mathcal{T}_n) \leq cn) = 1 - \ln(c^{-1}). \tag{4.52}$$

In particular, we have

$$\mathbb{E}[\nu] \geq \frac{n}{2}. \tag{4.53}$$

Proof. (i.) First, we translate the problem into a random permutation setting. We have

$$\nu_n(\mathcal{T}_n^\theta) =_d \max\{i \in [n-1] : C_{n-1,i}(\theta) \geq 1\} =: \alpha_n(\theta) \tag{4.54}$$

where $C_{n-1,i}(\theta)$ is the number of cycles of length i in a θ-biased Hoppe permutation. In this setting, Kingman [53] shows that $\frac{\alpha_n(\theta)}{n}$ converges in distribution to a random variable α with cumulative distribution function

$$F_\theta(x) = e^{\gamma\theta} x^{\theta-1} \Gamma(\theta) p_\theta\left(\frac{1}{x}\right) \quad \text{for } x > 0 \tag{4.55}$$

where γ is Euler's constant,

$$p_\theta(x) = \frac{e^{-\gamma\theta}}{\Gamma(\theta)} \left(1 + \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k!}\right) \int \cdots \int_{\mathcal{S}_k(x)} \left(1 - \sum_{j=1}^{k} y_i\right)^{\theta-1} \frac{dy_1 \ldots dy_k}{y_1 \ldots y_k} \qquad (4.56)$$

and

$$\mathcal{S}_k(x) = \left\{ y_1 > \frac{1}{x}, \ldots, y_k > \frac{1}{x}, \sum_{j=1}^{k} y_j < 1 \right\}. \qquad (4.57)$$

This proves the first part.

(ii) Setting $\theta = 1$ in the argument of (i) and recalling that the random permutation in this case reduces to a uniformly random permutation immediately reveals the result.

For the second claim, we first note Watterson [54] shows that the derivative of $F_1(x)$ over $[1/2, 1]$ simplifies to

$$f_1(x) = \frac{1}{x}. \qquad (4.58)$$

Hence, for any $c \in \left[\frac{1}{2}, 1\right]$

$$\lim_{n\to\infty} \mathbb{P}(\nu_n(\mathcal{T}_n) \le cn) = \mathbb{P}(\nu \le cn) = \int_c^1 \frac{1}{x} dx = \ln(1/c). \qquad (4.59)$$

We have

$$\mathbb{E}[\nu] \ge \int_{1/2}^1 x \frac{1}{x} dx = \frac{1}{2}. \qquad (4.60)$$

$\square$

Remark 4.10. The value $\lim_{n\to\infty} \frac{\nu_n(\mathcal{T}_n)}{n}$ is known to be the Golomb-Dickman constant in the literature. Its exact value is known to be $0.62432998854....$

4.4. Depth of Node n

Theorem 4.11. *Let $\mathcal{D}_n^\omega$ denote the depth of node n in a WRT $\mathcal{T}_n^\omega$ and let Z_n^ω denote the set of ancestors of n. Let moreover $A_{i,n}^\omega := \mathbf{1}(i \in Z_n^\omega)$. Then*

$$\mathcal{D}_n^\omega = 1 + \sum_{i=2}^{n-1} A_{i,n}^\omega. \tag{4.61}$$

The $A_{i,n}^\omega$ are mutually independent Bernoulli random variables with

$$\mathbb{P}(A_{i,n}^\omega = 1) = \frac{\omega_i}{\sum_{j=1}^{i} \omega_j}. \tag{4.62}$$

This directly yields the expectation and the variance:

$$\mathbb{E}\left[\mathcal{D}_n^\omega\right] = 1 + \sum_{i=2}^{n-1} \frac{\omega_i}{\sum_{j=1}^{i} \omega_j} = \sum_{i=1}^{n-1} \frac{\omega_i}{\sum_{j=1}^{i} \omega_j} \tag{4.63}$$

and

$$\mathrm{Var}\left(\mathcal{D}_n^\omega\right) = \sum_{i=2}^{n-1} \frac{\omega_i}{\sum_{j=1}^{i} \omega_j}\left(1 - \frac{\omega_i}{\sum_{j=1}^{i} \omega_j}\right) = \sum_{i=2}^{n-1} \frac{\omega_i \sum_{j=1}^{i-1} \omega_j}{(\sum_{j=1}^{i} \omega_j)^2}. \tag{4.64}$$

Proof. In a rooted tree, the depth of a node is equal to its number of ancestors, since these determine the path from the root to the node. Using that 1 definitely is an ancestor of n, in the notation of the theorem we thus get

$$\mathcal{D}_n^\omega = 1 + \sum_{i=2}^{n-1} A_{i,n}^\omega. \tag{4.65}$$

We will first find the distribution law of the $A_{i,n}^\omega$ and then show mutual independence. For the distribution law we will use the method used in [6]: we will first find the values for $n-1$ and $n-2$ and then proceed by induction.

The node $n-1$ can only be an ancestor of n if it is the parent of n, so we get

$$\mathbb{P}(n-1 \in Z_n^\omega) = \frac{\omega_{n-1}}{\sum_{i=1}^{n-1} \omega_i}. \tag{4.66}$$

Similarly $n-2$ can only be an ancestor of n if it is the parent of n or if it is the grandparent of n, in which case $n-2$ needs to be the parent of $n-1$ who needs to be the parent of n. This gives

$$\begin{aligned}
\mathbb{P}(n-1 \in Z_n^\omega) &= \underbrace{\frac{\omega_{n-2}}{\sum_{i=1}^{n-1} \omega_i}}_{n-2 \text{ is parent of } n} + \underbrace{\frac{\omega_{n-2}}{\sum_{i=1}^{n-2} \omega_i} \frac{\omega_{n-1}}{\sum_{i=1}^{n-1} \omega_i}}_{n-2 \text{ is grandparent } ofn} \\
&= \frac{\omega_{n-2} \sum_{i=1}^{n-2} \omega_i + \omega_{n-2}\omega_{n-1}}{\sum_{i=1}^{n-1} \omega_i \sum_{i=1}^{n-2} \omega_i} \\
&= \frac{\omega_{n-2} \sum_{i=1}^{n-1} \omega_i}{\sum_{i=1}^{n-1} \omega_i \sum_{i=1}^{n-2} \omega_i} \\
&= \frac{\omega_{n-2}}{\sum_{i=1}^{n-2} \omega_i}.
\end{aligned} \tag{4.67}$$

We will now show by induction that for all $j = 2, \ldots, n-1$,

$$\mathbb{P}(j \in Z_n^\omega) = \frac{\omega_j}{\sum_{i=1}^{j} \omega_j}. \tag{4.68}$$

Let the above be true for all $j \geq i+1$ and let $C_{i,j}^\omega$ denote the event that j is a child of i. Then

$$\mathbb{P}(i \in Z_n^\omega) = \sum_{j=i+1}^{n-1} \mathbb{P}(j \in Z_n^\omega, C_{i,j}^\omega) + \mathbb{P}(C_{i,n}^\omega). \tag{4.69}$$

Since $C_{i,j}^\omega$ only relates to the j-th step of the construction process and $j \in Z_n^\omega$ only depends on the $j+1$-th, $\ldots$, n-th steps, these two events are independent. We thus

get

$$
\mathbb{P}(i \in Z_n^\omega) = \sum_{j=i+1}^{n-1} \mathbb{P}(j \in Z_n^\omega)\mathbb{P}(C_{i,j}^\omega) + \mathbb{P}(C_{i,n}^\omega)
$$

$$
= \sum_{j=i+1}^{n-1} \left(\frac{\omega_j}{\sum_{k=1}^{j} \omega_k} \frac{\omega_i}{\sum_{k=1}^{j-1} \omega_k} \right) + \frac{\omega_i}{\sum_{j=1}^{n-1} \omega j}. \tag{4.70}
$$

To simplify this expression we first note that we can factor out ω_i and that the following holds:

$$
\frac{\omega_{i+1}}{\sum_{k=1}^{i+1} \omega_k \sum_{k=1}^{i} \omega_k} + \frac{\omega_{i+2}}{\sum_{k=1}^{i+2} \omega_k \sum_{k=1}^{i+1} \omega_k}
$$

$$
= \frac{\omega_{i+1} \sum_{k=1}^{i+2} \omega_k + \omega_{i+2} \sum_{k=1}^{i} \omega_k}{\sum_{k=1}^{i+2} \omega_k \sum_{k=1}^{i+1} \omega_k \sum_{k=1}^{i} \omega_k}
$$

$$
= \frac{(\omega_{i+1} + \omega_{i+2}) \sum_{k=1}^{i} \omega_k + \omega_{i+1}(\omega_{i+1} + \omega_{i+2})}{\sum_{k=1}^{i+2} \omega_k \sum_{k=1}^{i+1} \omega_k \sum_{k=1}^{i} \omega_k} \tag{4.71}
$$

$$
= \frac{(\omega_{i+1} + \omega_{i+2}) \sum_{k=1}^{i+1} \omega_k}{\sum_{k=1}^{i+2} \omega_k \sum_{k=1}^{i+1} \omega_k \sum_{k=1}^{i} \omega_k}
$$

$$
= \frac{\omega_{i+1} + \omega_{i+2}}{\sum_{k=1}^{i+2} \omega_k \sum_{k=1}^{i} \omega_k}.
$$

In general the following holds for $l \in \mathbb{N}$:

$$
\frac{\omega_{i+1} + \omega_{i+2} + \cdots + \omega_{i+l}}{\sum_{k=1}^{i} \omega_k \sum_{k=1}^{i+l} \omega_k} + \frac{\omega_{i+l+1}}{\sum_{k=1}^{i+l+1} \omega_k \sum_{k=1}^{i+l} \omega_k}
$$

$$
= \frac{(\omega_{i+1} + \cdots + \omega_{i+l}) \sum_{k=1}^{i+l+1} \omega_k + \omega_{i+l+1} \sum_{k=1}^{i} \omega_k}{\sum_{k=1}^{i+l+1} \omega_k \sum_{k=1}^{i+l} \omega_k \sum_{k=1}^{i} \omega_k}
$$

$$
= \frac{(\omega_{i+1} + \cdots + \omega_{i+l} + \omega_{i+l+1}) \sum_{k=1}^{i} \omega_k + (\omega_{i+1} + \cdots + \omega_{i+l}) \sum_{k=i+1}^{i+l+1} \omega_k}{\sum_{k=1}^{i+l+1} \omega_k \sum_{k=1}^{i+l} \omega_k \sum_{k=1}^{i} \omega_k} \tag{4.72}
$$

$$
= \frac{(\omega_{i+1} + \cdots + \omega_{i+l+1}) \sum_{k=1}^{i+l} \omega_k}{\sum_{k=1}^{i+l+1} \omega_k \sum_{k=1}^{i+l} \omega_k \sum_{k=1}^{i} \omega_k}
$$

$$
= \frac{\omega_{i+1} + \cdots + \omega_{i+l+1}}{\sum_{k=1}^{i+l+1} \omega_k \sum_{k=1}^{i} \omega_k}.
$$

By using this equality $n - i - 2$ times, we thus get

$$
\begin{aligned}
\mathbb{P}(i \in Z_n^\omega) &= \omega_i \left(\frac{\omega_{i+1} + \cdots + \omega_{n-1}}{\sum_{k=1}^i \omega_k \sum_{k=1}^{n-1} \omega_k} + \frac{1}{\sum_{k=1}^{n-1} \omega_k} \right) \\
&= \omega_i \left(\frac{\omega_{i+1} + \cdots + \omega_{n-1} + \sum_{k=1}^i \omega_k}{\sum_{k=1}^i \omega_k \sum_{k=1}^{n-1} \omega_k} \right) \qquad (4.73) \\
&= \frac{\omega_i}{\sum_{k=1}^i \omega_k}.
\end{aligned}
$$

Now we will show that the events $A_{i,n}^\omega$ are mutually independent for $j = 2, \ldots, n-1$. For this we will use the method used in [28]: for any $2 \le k \le n-2$ and $2 \le j_k < \cdots < j_2 < j_1 \le n-1$ consider the event that all j_i's and only the j_i's are ancestors of n. We will denote this event by E. Then

$$
E := (\mathbf{1}(j_i \in Z_n^\omega) = 1, \mathbf{1}(j \in Z_n^\omega) = 0, \text{ for } j \ne j_i, i = 1, \ldots, k). \qquad (4.74)
$$

By the structure of the recursive tree, to realize this event, n must be a child of j_1, j_1 a child of $j_2, \ldots, j_{k-1}$ a child of j_k and j_k a child of 1. In general for $i = 1, \ldots, k-1$, j_i must be a child of j_{i+1}. It does not matter what nodes $j \ne j_i$ attach to. Hence, by

the attachment probabilities we get:

$$\mathbb{P}(E) = \mathbb{P}(j_i \in Z_n^\omega, j \notin Z_n^\omega, \text{ for } j \neq j_i, i = 1, \ldots, k)$$

$$= \underbrace{\frac{\omega_{j_1}}{\sum_{\ell=1}^{n-1} \omega_\ell}}_{n \text{ child of } j_1} \prod_{i=1}^{k-1} \underbrace{\frac{\omega_{j_{i+1}}}{\sum_{\ell=1}^{j_i-1} \omega_\ell}}_{j_i \text{ child of } j_{i+1}} \underbrace{\frac{\omega_1}{\sum_{\ell=1}^{j_k-1} \omega_\ell}}_{j_k \text{ child of } 1}$$

$$= \omega_1 \omega_{j_1} \cdots \omega_{j_k} \prod_{i=1}^{n-1} \frac{1}{\sum_{\ell=1}^{i} \omega_\ell} \prod_{\substack{1 < j < n \\ j \neq j_i, i=1,\ldots,k}} \left(\sum_{\ell=1}^{j-1} \omega_\ell \right)$$

$$= \prod_{i=1}^{k} \frac{\omega_{j_i}}{\sum_{\ell=1}^{j_i} \omega_\ell} \prod_{\substack{1 < j < n \\ j \neq j_i, i=1,\ldots,k}} \left(\frac{\sum_{\ell=1}^{j-1} \omega_\ell}{\sum_{\ell=1}^{j} \omega_\ell} \right) \frac{\omega_1}{\sum_{\ell=1}^{1} \omega_\ell} \qquad (4.75)$$

$$= \prod_{i=1}^{k} \frac{\omega_{j_i}}{\sum_{\ell=1}^{j_i} \omega_\ell} \prod_{\substack{1 < j < n \\ j \neq j_i, i=1,\ldots,k}} \left(1 - \frac{\omega_j}{\sum_{\ell=1}^{j} \omega_\ell} \right)$$

$$= \prod_{i=1}^{k} \mathbb{P}(j_i \in Z_n^\omega) \prod_{\substack{1 < j < n \\ j \neq j_i, i=1,\ldots,k}} \mathbb{P}(j \notin Z_n^\omega).$$

This implies that the events $\mathbf{1}(i \in Z_n^\omega)$ are mutually independent. Hence the $A_{i,n}^\omega$ are mutually independent Bernoulli random variables, which immediately gives expectation and variance of $\mathcal{D}_n^\omega$ as stated in the theorem. $\qquad \square$

Remark 4.12. While the number of branches and the depth of node n are identically distributed in URTs this is not generally the case for WRTs. Both of these statistics can be written as sums of Bernoulli random variables but these are not identically distributed in general.

4.4.1. Some Examples of Weight Sequences

We will now consider the examples of weight sequences we also considered in Section 4.3 to illustrate the above remark. For the Hoppe tree, we write $\mathcal{D}_n^\theta$ for the

depth of node n and have by Theorem 3.15

$$
\mathbb{E}\left[\mathcal{D}_n^\theta\right] = 1 + \sum_{i=1}^{n-2} \frac{1}{\theta+1} = \ln(n) + \mathcal{O}(1) \text{ and}
$$

$$
\mathrm{Var}\left(\mathcal{D}_n^\theta\right) = \sum_{i=1}^{n-2} \frac{\theta+i-1}{(\theta+i)^2} = \ln(n) + \mathcal{O}(1).
$$

(4.76)

This implies that in a Hoppe tree the distribution of the depth of node n is asymptotically close to the distribution of the depth in URTs, which is not the case for the number of branches of a Hoppe tree.

For the WRT with $\omega_i = \theta$ for $i = 1, \ldots, k$ and $\omega_i = 1$ for $i > k$, we call the tree $T_n^{\theta^k}$ and write $\mathcal{D}_n^{\theta^k}$ for the depth of node n. We get

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^{\theta^k}\right] &= \sum_{i=1}^{k} \frac{\theta}{i\theta} + \sum_{i=k+1}^{n-1} \frac{1}{k\theta+i-k} \\
&= \sum_{i=1}^{k} \frac{1}{i} + \sum_{i=k+1}^{n-1} \frac{1}{i} + \sum_{i=k+1}^{n-1} \frac{1}{k\theta+i-k} - \sum_{i=k+1}^{n-1} \frac{1}{i} \\
&= \sum_{i=1}^{n-1} \frac{1}{i} + \sum_{i=k+1}^{n-1} \frac{-k\theta+k}{(k\theta+i-k)i} \\
&= \sum_{i=1}^{n-1} \frac{1}{i} + \sum_{i=k+1}^{n-1} \frac{k(1-\theta)}{(k(\theta-1)+i)i}.
\end{aligned}
$$

(4.77)

This shows that the expectation of the depth of node n in a θ^k-RT is smaller than the expectation in a URT by a constant term if $\theta > 1$ and bigger by a constant term if $\theta < 1$.

Similarly

$$\text{Var}\left(\mathcal{D}_n^{\theta k}\right) = \sum_{i=2}^{k} \frac{\theta^2(i-1)}{(i\theta)^2} + \sum_{i=k+1}^{n-1} \frac{k\theta+i-1-k}{(k\theta+i-k)^2}$$

$$= \sum_{i=2}^{k} \frac{i-1}{i^2} + \sum_{i=k+1}^{n-1} \frac{i-1}{i^2} + \sum_{i=k+1}^{n-1} \frac{k\theta+i-1-k}{(k\theta+i-k)^2} - \sum_{i=k+1}^{n-1} \frac{i-1}{i^2}$$

$$= \sum_{i=2}^{n-1} \frac{i-1}{i^2} + \sum_{i=k+1}^{n-1} \frac{(k(\theta-1)+i-1)i^2 - ((k(\theta-1)+i)^2(i-1)}{(k(\theta-1)+i)^2 i^2}$$

$$= \sum_{i=2}^{n-1} \frac{i-1}{i^2} + \sum_{i=k+1}^{n-1} \frac{k(\theta-1)(i^2+i+k(\theta-1))}{(i(k(\theta-1)+i))^2}.$$

(4.78)

Asymptotically this gives, as for URTs,

$$\mathbb{E}\left[\mathcal{D}_n^{\theta k}\right] = \ln(n) + \mathcal{O}(1) \text{ and } \text{Var}\left(\mathcal{D}_n^{\theta k}\right) = \ln(n) + \mathcal{O}(1).$$

(4.79)

More generally, as for the number of branches, if the weights are bounded from below and above, the expectation and variance of the depth of node n will still be equal to $\mathcal{O}(\ln(n))$ asymptotically. Let $0 < m < \omega_i < M$ for all i, then

$$\sum_{i=1}^{n-1} \frac{m}{iM} < \sum_{i=1}^{n-1} \frac{\omega_i}{\sum_{j=1}^{i}\omega_j} < \sum_{i=1}^{n-1} \frac{M}{im}$$

$$\Leftrightarrow \frac{m}{M}H_{n-1} < \mathbb{E}\left[\mathcal{D}_n^\omega\right] < \frac{M}{m}H_{n-1}.$$

(4.80)

And similarly

$$\sum_{i=2}^{n-1} \frac{(i-1)m^2}{(iM)^2} < \sum_{i=2}^{n-1} \frac{\omega_i \sum_{j=1}^{i-1}\omega_j}{(\sum_{j=1}^{i}\omega_j)^2} < \sum_{i=2}^{n-1} \frac{(i-1)M^2}{(im)^2}$$

$$\Leftrightarrow \frac{m^2}{M^2}\sum_{i=2}^{n-1}\frac{i-1}{i^2} < \text{Var}\left(\mathcal{D}_n^\omega\right) < \frac{M^2}{m^2}\sum_{i=2}^{n-1}\frac{i-1}{i^2}$$

$$\Leftrightarrow \frac{m^2}{M^2}H_{n-1} + \mathcal{O}(1) < \text{Var}\left(\mathcal{D}_n^\omega\right) < \frac{M^2}{m^2}H_{n-1} + \mathcal{O}(1).$$

(4.81)

Thus in the examples above the qualitative behaviour of the depth of node n is similar to the uniform case, since it is of order $\ln(n)$. For the branches the situation was very different when the weights were not bounded. We will now consider these same examples and some others in order to compare the behaviours and get an idea of the range of values we can get for the expectation and variance of the depth of node n in WRTs.

(i) Let $(\omega_i)_{i\in\mathbb{N}} = (i)_{i\in\mathbb{N}}$. In this case we get from Theorem 4.11

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^\omega\right] &= \sum_{i=1}^{n-1} \frac{i}{\sum_{j=1}^{i} j} = \sum_{i=1}^{n-1} \frac{2i}{(i+1)i} \\
&= 2\sum_{i=1}^{n-1} \frac{1}{i+1} = 2\ln(n) + \mathcal{O}(1)
\end{aligned}
\tag{4.82}
$$

and

$$
\begin{aligned}
\mathrm{Var}\left(\mathcal{D}_n^\omega\right) &= \sum_{i=2}^{n-1} \frac{i\sum_{j=1}^{i-1} j}{\left(\sum_{j=1}^{i} j\right)^2} = \sum_{i=2}^{n-1} \frac{2i^2(i-1)}{(i(i+1))^2} = 2\sum_{i=2}^{n-1} \frac{(i-1)}{(i+1)^2} \\
&= 2\sum_{i=2}^{n-1} \frac{1}{i+1} - 2\sum_{i=2}^{n-1} \frac{2}{(i+1)^2} = 2\ln(n) + \mathcal{O}(1).
\end{aligned}
\tag{4.83}
$$

(ii) We can get a similar result for a more general case. Let $(\omega_i)_{i\in\mathbb{N}} = (i^k)_{i\in\mathbb{N}}$ for $k \in \mathbb{N}$. In this case

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^\omega\right] &= \sum_{i=1}^{n-1} \frac{i^k}{\sum_{j=1}^{i} j^k} = \sum_{i=1}^{n-1} \frac{1}{\frac{1}{i^k}\frac{i^{k+1}}{k+1} + \mathcal{O}(i^k)} \\
&= \sum_{i=1}^{n-1} \frac{1}{\frac{i}{k+1} + \mathcal{O}(1)} = (k+1)\ln(n) + \mathcal{O}(1)
\end{aligned}
\tag{4.84}
$$

and

$$\begin{aligned}
\mathrm{Var}\left(\mathcal{D}_n^\omega\right) &= \sum_{i=2}^{n-1} \frac{i^k \left(\frac{i^{k+1}}{k+1} + \mathcal{O}(i^k)\right)}{\left(\frac{i^{k+1}}{k+1} + \mathcal{O}(i^k)\right)^2} \\
&= \sum_{i=2}^{n-1} (\mathcal{O}(1) + \mathcal{O}\left(\frac{1}{i}\right) \frac{i^k}{\frac{i^{k+1}}{k+1} + \mathcal{O}(i^k)} \\
&= \sum_{i=2}^{n-1} \left(\mathcal{O}(1) + \mathcal{O}\left(\frac{1}{i}\right)\right) \frac{1}{\frac{i}{k+1} + \mathcal{O}(1)} \\
&= (k+1)\ln(n) + \mathcal{O}(1).
\end{aligned}$$ (4.85)

(iii) Let $(\omega_i)_{i\in\mathbb{N}} = \left(\frac{1}{i}\right)_{i\in\mathbb{N}}$. In this case we get

$$\mathbb{E}\left[\mathcal{D}_n^\omega\right] = \sum_{i=1}^{n-1} \frac{\frac{1}{i}}{\sum_{j=1}^i \frac{1}{j}} = \sum_{i=1}^{n-1} \frac{1}{i H_i} = \sum_{i=1}^{n-1} \frac{1}{i(\ln(i) + \mathcal{O}(1))} \xrightarrow{n\to\infty} \infty$$ (4.86)

and

$$\mathrm{Var}\left(\mathcal{D}_n^\omega\right) = \sum_{i=2}^{n-1} \frac{\frac{1}{i} H_{i-1}}{i H_i^2} = \sum_{i=2}^{n-1} \frac{1}{i H_i} - \sum_{i=2}^{n-1} \frac{1}{i^2 H_i^2} \xrightarrow{n\to\infty} \infty.$$ (4.87)

(iv) Let $(\omega_i)_{i\in\mathbb{N}} = \left(\frac{1}{i^2}\right)_{i\in\mathbb{N}}$. Then

$$\mathbb{E}\left[\mathcal{D}_n^\omega\right] = \sum_{i=1}^{n-1} \frac{\frac{1}{i^2}}{\sum_{j=1}^i \frac{1}{j^2}}$$ (4.88)

and since for all $i \in \mathbb{N}$ we have $1 < \sum_{j=1}^i \frac{1}{j^2} < \frac{\pi^2}{6}$ we get

$$\begin{aligned}
& \frac{6}{\pi^2} \sum_{i=1}^{n-1} \frac{1}{i^2} \le \mathbb{E}\left[\mathcal{D}_n^\omega\right] \le \sum_{i=1}^{n-1} \frac{1}{i^2} \\
\Rightarrow\ & \frac{6}{\pi^2} \frac{\pi^2}{6} \le \mathbb{E}\left[\mathcal{D}_n^\omega\right] \le \frac{\pi^2}{6} \\
\Rightarrow\ & 1 \le \mathbb{E}\left[\mathcal{D}_n^\omega\right] \le \frac{\pi^2}{6}.
\end{aligned}$$ (4.89)

And since we have

$$\text{Var}\left(\mathcal{D}_n^\omega\right) = \sum_{i=1}^{n-1} \frac{\frac{1}{i^2} \sum_{j=1}^{i-1} \frac{1}{j^2}}{\left(\sum_{j=1}^{i} \frac{1}{j^2}\right)^2} \tag{4.90}$$

and for all $i \geq 2$, it holds that $\frac{4}{5} \leq \frac{\sum_{j=1}^{i-1} \frac{1}{j^2}}{\left(\sum_{j=1}^{i} \frac{1}{j^2}\right)^2} \leq 1$, we get

$$\frac{4}{5} \sum_{i=2}^{n-1} \frac{1}{i^2} \leq \text{Var}\left(\mathcal{D}_n^\omega\right) \leq \sum_{i=2}^{n-1} \frac{1}{i^2}$$

$$\Rightarrow \frac{4}{5}\frac{1}{4} \leq \text{Var}\left(\mathcal{D}_n^\omega\right) \leq \frac{\pi^2}{6} - 1 \tag{4.91}$$

$$\Rightarrow 0,2 \leq \text{Var}\left(\mathcal{D}_n^\omega\right) \leq 0,65.$$

We will also get a finite expectation and variance for $\left(\frac{1}{i^k}\right)_{i\in\mathbb{N}}$ where $k \in \mathbb{N}$ by similar computations.

(v) Let $(\omega_i)_{i\in\mathbb{N}} = (\ln(i))_{i\in\mathbb{N}}$. Then

$$\mathbb{E}\left[\mathcal{D}_n^\omega\right] = \sum_{i=1}^{n-1} \frac{\ln(i)}{\sum_{j=1}^{i} \ln(j)} = \sum_{i=1}^{n-1} \frac{\ln(i)}{\ln(i!)}$$

$$> \sum_{i=1}^{n-1} \frac{\ln(i)}{\ln(i^i)} = \sum_{i=1}^{n-1} \frac{1}{i} = \ln(n) + \mathcal{O}(1). \tag{4.92}$$

Also

$$\text{Var}\left(\mathcal{D}_n^\omega\right) = \sum_{i=1}^{n-1} \frac{\ln(i)}{\sum_{j=1}^{i} \ln(j)} - \sum_{i=1}^{n-1} \frac{(\ln(i))^2}{\left(\sum_{j=1}^{i} \ln(j)\right)^2}$$

$$= \sum_{i=1}^{n-1} \frac{\ln(i)}{\ln(i!)} - \sum_{i=1}^{n-1} \frac{(\ln(i))^2}{(\ln(i!))^2}$$

$$> \sum_{i=1}^{n-1} \frac{\ln(i)}{\ln(i^i)} - \sum_{i=1}^{n-1} \frac{(\ln(i))^2}{\left(\ln(i^{\frac{i}{2}})\right)^2} \tag{4.93}$$

$$= \sum_{i=1}^{n-1} \frac{1}{i} - \sum_{i=1}^{n-1} \frac{4}{i^2} = \ln(n) + \mathcal{O}(1).$$

(vi) Let $(\omega_i)_{i\in\mathbb{N}} = (a^{i-1})_{i\in\mathbb{N}}$. First let us consider the case $a < 1$. Then

$$\mathbb{E}\left[\mathcal{D}_n^\omega\right] = \sum_{i=1}^{n-1} \frac{a^{i-1}}{\sum_{j=1}^i a^{j-1}} < \sum_{i=1}^{n-1} a^{i-1} < \frac{1}{1-a} \tag{4.94}$$

and similarly

$$\mathrm{Var}\left(\mathcal{D}_n^\omega\right) = \sum_{i=1}^{n-1} \frac{a^{i-1}}{\sum_{j=1}^i a^{j-1}} - \sum_{i=1}^{n-1} \frac{(a^{i-1})^2}{\left(\sum_{j=1}^i a^{j-1}\right)^2} < \frac{1}{1-a}. \tag{4.95}$$

Thus we have a finite expectation and variance. On the other hand, if $a > 1$,

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^\omega\right] &= \sum_{i=1}^{n-1} \frac{a^{i-1}}{\sum_{j=1}^i a^{j-1}} = \sum_{i=0}^{n-2} \frac{(a-1)a^i}{a^{i+1}-1} = \sum_{i=0}^{n-2} \frac{a-1}{a-\frac{1}{a^i}} \\
&> \sum_{i=0}^{n-2} \frac{a-1}{a} = (n-2)\frac{a-1}{a}
\end{aligned} \tag{4.96}$$

and

$$\begin{aligned}
\mathrm{Var}\left(\mathcal{D}_n^\omega\right) &= \sum_{i=1}^{n-1} \frac{a^{i-1}}{\sum_{j=1}^i a^{j-1}}\left(1 - \frac{a^{i-1}}{\sum_{j=1}^i a^{j-1}}\right) = \sum_{i=0}^{n-2} \frac{(a-1)a^i}{a^{i+1}-1}\frac{\sum_{j=0}^{i-1} a^j}{\sum_{j=0}^i a^j} \\
&= \sum_{i=0}^{n-2} \frac{(a-1)a^i}{a^{i+1}-1}\frac{a^i-1}{a^{i+1}-1} > (a-1)\sum_{i=0}^{n-2} \frac{a^{2i}-a^i}{(a^{i+1})^2} > (a-1)\sum_{i=0}^{n-2} \frac{1-\frac{1}{a^i}}{a^2} \\
&> (a-1)\sum_{i=0}^{n-2} \frac{1}{a^2} - \frac{1}{a^{i+2}} = \frac{a-1}{a^2}n + \mathcal{O}(1).
\end{aligned} \tag{4.97}$$

4.4.2. Central Limit Theorem

As the depth of node n can be written as a sum of independent Bernoulli random variables, we can apply Theorem 2.8, as we did for the number of branches.

Theorem 4.13. *Let $\mathcal{D}_n^\omega$ denote the depth of node n in a weighted random recursive tree $\mathcal{T}_n^\omega$ with n nodes. If $\mathrm{Var}(\mathcal{D}_n^\omega) = \sum_{i=2}^{n-1} \frac{\omega_i}{\sum_{j=1}^i \omega_j} - \left(\frac{\omega_i}{\sum_{j=1}^i \omega_j}\right)^2$ diverges, then*

$$\frac{\mathcal{D}_n^\omega - \mathbb{E}\left[\mathcal{D}_n^\omega\right]}{\sqrt{\mathrm{Var}(\mathcal{D}_n^\omega)}} \xrightarrow[d]{n\to\infty} \mathcal{G}. \tag{4.98}$$

Proof. The result follows directly from Theorem 2.8, i.e. Liapounov's central limit theorem for sums of independent Bernoulli random variables. □

Remark 4.14. We could not find a general condition for the weight sequence that implies divergence of the variance. We can though say the following:

(i) When $\sum_{i=1}^{n} \omega_i$ converges, the variance does not diverge. This can be seen by bounding $\frac{\omega_i}{\sum_{j=1}^{i} \omega_j}$ from above by $\frac{\omega_i}{\omega_1}$.

(ii) If the variance diverges the expectation diverges too.

4.4.3. Rate of Convergence

We can derive a rate of convergence for the depth of node n that is similar to the rate of convergence for the number of branches.

Theorem 4.15. *Let $\mathcal{D}_n^{\omega}$ denote the number of branches of a weighted recursive tree. Then*

$$d_W\left(\frac{\mathcal{D}_n^{\omega} - E[\mathcal{D}_n^{\omega}]}{\sqrt{\operatorname{Var}(\mathcal{D}_n^{\omega})}}, \mathcal{G}\right) \leq \frac{1}{\sqrt{\operatorname{Var}(\mathcal{D}_n^{\omega})}} \frac{\sqrt{28} + \sqrt{\pi}}{\sqrt{\pi}}. \tag{4.99}$$

This bound decreases to 0 if and only if $\operatorname{Var}(\mathcal{D}_n^{\omega})$ diverges, which is the same criterion as we had for the CLT.

Proof. The proof follows steps similar to the proof of Theorem 4.8. □

This concludes the results about the depth of node n, we will continue with another statistic: the number of leaves.

4.5. Number of Leaves

We will compute the number of leaves for a specific kind of weighted recursive tree only. In this section we assume that the weight sequence $(\omega_i)_{i \in \mathbb{N}}$ is such that for

some $\theta > 0$,

$$
\omega_i = \begin{cases} \theta \text{ for } 1 \le i \le k \\ 1 \text{ otherwise.} \end{cases} \tag{4.100}
$$

We will call such WRTs θ^k-RTs and denote them by $\mathcal{T}_n^{\theta^k}$. The techniques used here should also apply for more general tree models, but with some accompanying cumbersome notation.

Theorem 4.16. *Let $\mathcal{L}_n^{\theta^k}$ denote the number of leaves of a WRT $\mathcal{T}_n^{\theta^k}$. Then*

(i) *for some constant C' such that $|C'| < |k(\theta - 1)| + \frac{k(\theta+1)}{2}$,*

$$
\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right] = \frac{n}{2} + C' + \mathcal{O}\left(\frac{1}{n}\right) \tag{4.101}
$$

(ii)

$$
\mathrm{Var}\left(\mathcal{L}_n^{\theta^k}\right) = \frac{n}{12} + \mathcal{O}(1) \text{ and} \tag{4.102}
$$

(iii) *for all $t \ge 0$,*

$$
\mathbb{P}\left(\left|\mathcal{L}_n^{\theta^k} - \mathbb{E}\left[L_n^{\theta^k}\right]\right| \ge t\right) \le 2e^{-6\frac{t^2}{k(\theta-1)+n+2}} e^{\frac{6t(k\theta(k-1))}{(k(\theta-1)+n-1)(k(\theta-1)+n+2)}}. \tag{4.103}
$$

Proof. The proof will make use of a martingale argument. First, note that for $n \ge 2$ we have

$$
\mathcal{L}_n^{\theta^k} = \mathcal{L}_{n-1}^{\theta^k} + Y_n^{\theta^k} \tag{4.104}
$$

where

$$
Y_n^{\theta^k} = \begin{cases} 1 \text{ if the parent of } n \text{ was not a leaf at time } n-1 \\ 0 \text{ otherwise} \end{cases} \tag{4.105}
$$

and $\mathcal{L}_1^{\theta^k} = 0$, implying in particular $\mathcal{L}_2^{\theta^k} = 1$.

We will now construct the martingale that we will use throughout the proof. We want to find an expression for $\mathbb{E}\left[\mathcal{L}_n^{\theta^k} | \mathcal{L}_{n-1}^{\theta^k}\right]$, so that we can use a martingale to study the properties of $\mathcal{L}_n^{\theta^k}$. First of all let $N_n^{\theta^k}$ denote the number of leaves among the first k nodes at time n. We have for $n \geq k+1$,

$$
\begin{aligned}
\mathbb{E}\left[Y_n^{\theta^k} \Big| \mathcal{L}_{n-1}^{\theta^k}\right] &= \sum_{j=0}^{k} \mathbb{E}\left[Y_n^{\theta^k} \Big| \mathcal{L}_{n-1}^{\theta^k}, N_{n-1}^{\theta^k} = j\right] \mathbb{P}\left(N_{n-1}^{\theta^k} = j\right) \\
&= \sum_{j=0}^{k} \mathbb{P}\left(Y_n^{\theta^k} = 1 \Big| \mathcal{L}_{n-1}^{\theta^k}, N_{n-1}^{\theta^k} = j\right) \mathbb{P}\left(N_{n-1}^{\theta^k} = j\right) \\
&= \sum_{j=0}^{k} \left(1 - \mathbb{P}\left(Y_n^{\theta^k} = 0 \Big| \mathcal{L}_{n-1}^{\theta^k}, N_{n-1}^{\theta^k} = j\right)\right) \mathbb{P}\left(N_{n-1}^{\theta^k} = j\right) \\
&= \sum_{j=0}^{k} \left(1 - \frac{\mathcal{L}_{n-1}^{\theta^k} - j}{k\theta + n - 1 - k} - \frac{\theta j}{k\theta + n - 1 - k}\right) \mathbb{P}\left(N_{n-1}^{\theta^k} = j\right) \quad (4.106) \\
&= \sum_{j=0}^{k} \mathbb{P}\left(N_{n-1}^{\theta^k} = j\right) - \sum_{j=0}^{k} \frac{\mathcal{L}_{n-1}^{\theta^k}}{k\theta + n - 1 - k} \mathbb{P}\left(N_{n-1}^{\theta^k} = j\right) \\
&\quad + \sum_{j=0}^{k} \frac{(1-\theta)j}{k\theta + n - 1 - k} \mathbb{P}\left(N_{n-1}^{\theta^k} = j\right) \\
&= 1 - \frac{\mathcal{L}_{n-1}^{\theta^k}}{k\theta + n - 1 - k} + \frac{(1-\theta)\mathbb{E}\left[N_{n-1}^{\theta^k}\right]}{k\theta + n - 1 - k}.
\end{aligned}
$$

We thus get

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{L}_n^{\theta^k} \Big| \mathcal{L}_{n-1}^{\theta^k}\right] &= \mathbb{E}\left[\mathcal{L}_{n-1}^{\theta^k} + Y_n^{\theta^k} \Big| \mathcal{L}_{n-1}^{\theta^k}\right] \\
&= \mathcal{L}_{n-1}^{\theta^k} + \mathbb{E}\left[Y_n^{\theta^k} \Big| \mathcal{L}_{n-1}^{\theta^k}\right] \\
&= \mathcal{L}_{n-1}^{\theta^k} + 1 - \frac{\mathcal{L}_{n-1}^{\theta^k}}{k\theta + n - 1 - k} + \frac{(1-\theta)\mathbb{E}\left[N_{n-1}^{\theta^k}\right]}{k\theta + n - 1 - k} \qquad (4.107) \\
&= \frac{k(\theta - 1) + n - 2}{k(\theta - 1) + n - 1} \mathcal{L}_{n-1}^{\theta^k} + \frac{(1-\theta)\left(\mathbb{E}\left[N_{n-1}^{\theta^k}\right] - k\right) + n - 1}{k(\theta - 1) + n - 1}.
\end{aligned}
$$

We now set $X_n^{\theta^k} := a(n)\mathcal{L}_n^{\theta^k} + b(n)$ and try to find $a(n)$ and $b(n)$ such that $\mathbb{E}\left[X_n^{\theta^k}\middle|\mathcal{L}_{n-1}^{\theta^k}\right] = X_{n-1}^{\theta^k}$. Now we have that

$$
\begin{aligned}
&\mathbb{E}\left[X_n^{\theta^k}\middle|\mathcal{L}_{n-1}^{\theta^k}\right] = X_{n-1}^{\theta^k}\\
&\Leftrightarrow a(n)\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\middle|\mathcal{L}_{n-1}^{\theta^k}\right] + b(n) = a(n-1)\mathcal{L}_{n-1}^{\theta^k} + b(n-1)\\
&\Leftrightarrow a(n)\left(\frac{k(\theta-1)+n-2}{k(\theta-1)+n-1}\mathcal{L}_{n-1}^{\theta^k} + \frac{(1-\theta)\left(\mathbb{E}\left[N_{n-1}^{\theta^k}\right]-k\right)+n-1}{k(\theta-1)+n-1}\right) + b(n)\\
&\qquad = a(n-1)\mathcal{L}_{n-1}^{\theta^k} + b(n-1)\\
&\Leftrightarrow \left(a(n)\frac{k(\theta-1)+n-2}{k(\theta-1)+n-1} - a(n-1)\right)\mathcal{L}_{n-1}^{\theta^k}\\
&\qquad + a(n)\frac{(1-\theta)\left(\mathbb{E}\left[N_{n-1}^{\theta^k}\right]-k\right)+n-1}{k(\theta-1)+n-1} + b(n) - b(n-1) = 0.
\end{aligned}
\tag{4.108}
$$

To simplify this equation we first set the factor of $\mathcal{L}_n^{\theta^k}$ equal to 0, which gives

$$
a(n)\frac{k(\theta-1)+n-2}{k(\theta-1)+n-1} - a(n-1) = 0 \Leftrightarrow \frac{k(\theta-1)+n-2}{k(\theta-1)+n-1} = \frac{a(n-1)}{a(n)}.
\tag{4.109}
$$

Since we do not have boundary conditions we choose $a(n) = k(\theta-1)+n-1$.

Now we similarly need

$$
\begin{aligned}
&a(n)\frac{(1-\theta)\left(\mathbb{E}\left[N_{n-1}^{\theta^k}\right]-k\right)+n-1}{k(\theta-1)+n-1} + b(n) - b(n-1) = 0\\
&\Leftrightarrow b(n) - b(n-1) = (\theta-1)\left(\mathbb{E}\left[N_{n-1}^{\theta^k}\right]-k\right) - (n-1).
\end{aligned}
\tag{4.110}
$$

This is true for

$$
b(n) = (\theta-1)\left(\sum_{i=k+1}^{n}\mathbb{E}\left[N_{i-1}^{\theta^k}\right]-k\right) - \sum_{i=k+1}^{n} i - 1.
\tag{4.111}
$$

We thus define for $n \geq k + 1$,

$$X_n^{\theta^k} := (k(\theta - 1) + n - 1)\mathcal{L}_n^{\theta^k} + (\theta - 1)\left(\sum_{i=k+1}^{n} \mathbb{E}\left[N_{i-1}^{\theta^k}\right] - k\right) - \sum_{i=k+1}^{n} i - 1 \quad (4.112)$$

and thus have $\mathbb{E}\left[X_n^{\theta^k} | \mathcal{L}_{n-1}^{\theta^k}\right] = X_{n-1}^{\theta^k}$.

We can now use this martingale to get the expectation. Since $X_n^{\theta^k}$ is a martingale we get for all $n \geq k + 1$,

$$\mathbb{E}\left[X_n^{\theta^k}\right] = \mathbb{E}\left[X_{k+1}^{\theta^k}\right]. \quad (4.113)$$

Also we have

$$
\begin{aligned}
\mathbb{E}\left[X_{k+1}^{\theta^k}\right] &= (k(\theta - 1) + k + 1 - 1)\mathbb{E}\left[\mathcal{L}_{k+1}^{\theta^k}\right] + (1 - \theta)\left(\mathbb{E}\left[N_k^{\theta^k}\right] - k\right) + k \\
&= k\theta\frac{k+1}{2} + (1 - \theta)\left(\frac{k}{2} - k\right) + k \\
&= k\theta\frac{k+1}{2} + \frac{k}{2}(1 + \theta) \\
&= \frac{k}{2}(k\theta + 1 + 2\theta)
\end{aligned}
\quad (4.114)
$$

where the second equation comes from the fact that up to time $k + 1$ the tree has the same attachment probabilities as a uniform recursive tree.

This gives us an expression for $\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right]$:

$$
\begin{aligned}
\mathbb{E}\left[X_n^{\theta^k}\right] &= \frac{k}{2}(k\theta + 1 + 2\theta) \\
&\Leftrightarrow a(n)\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right] + b(n) = \frac{k}{2}(k\theta + 1 + 2\theta) \\
&\Leftrightarrow \mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right] = \frac{1}{a(n)}\left(\frac{k}{2}(k\theta + 1 + 2\theta) - b(n)\right).
\end{aligned}
\quad (4.115)
$$

So we have

$$
\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right]
$$

$$
= \frac{1}{k(\theta - 1) + n - 1}
$$

$$
\cdot \left(\frac{k}{2}(k\theta + 1 + 2\theta) - (\theta - 1) \left(\sum_{i=k+1}^{n} \mathbb{E}\left[N_{i-1}^{\theta^k}\right] - k \right) + \sum_{i=k+1}^{n} i - 1 \right)
$$

$$
= \frac{k(k\theta + 1 + 2\theta)}{2(k(\theta - 1) + n - 1)} + \frac{1 - \theta}{k(\theta - 1) + n - 1} \left(\sum_{i=k+1}^{n} \mathbb{E}\left[N_{i-1}^{\theta^k}\right] - k \right)
$$

$$
+ \frac{1}{k(\theta - 1) + n - 1} \sum_{i=k+1}^{n} i - 1 \tag{4.116}
$$

$$
= \frac{1}{k(\theta - 1) + n - 1} \frac{(n - k)(n - k - 1)}{2} + C + \mathcal{O}\left(\frac{1}{n}\right)
$$

$$
= \frac{n}{2} + \frac{(n - k)(n - k - 1) - n(k(\theta - 1) + n - 1)}{2(k(\theta - 1) + n - 1)} + C + \mathcal{O}\left(\frac{1}{n}\right)
$$

$$
= \frac{n}{2} + \frac{n(-\theta - 1)k + k + k^2}{2(k(\theta - 1) + n - 1)} + C + \mathcal{O}\left(\frac{1}{n}\right)
$$

$$
= \frac{n}{2} + C' + \mathcal{O}\left(\frac{1}{n}\right)
$$

where $|C| < |k(\theta - 1)|$ and $|C'| < |k(\theta - 1)| + \frac{k(\theta + 1)}{2}$. In the fifth line we used that for all $i = k + 1, \ldots, n$, we have $|\mathbb{E}[N_{i-1}^{\theta^k}] - k| < k$ since at any time there can be between 0 and $k - 1$ leaves among the first k nodes.

We will now derive the concentration equality. Because it is easier to manipulate zero-martingales we now set

$$
Z_n^{\theta^k} := X_n^{\theta^k} - \mathbb{E}\left[X_{k+1}^{\theta^k}\right]
$$

$$
= (k(\theta - 1) + n - 1)\mathcal{L}_n^{\theta^k} + (\theta - 1) \left(\sum_{i=k+1}^{n} \mathbb{E}\left[N_{i-1}^{\theta^k}\right] - k \right)
$$

$$
- \sum_{i=k+1}^{n} i - 1 - \frac{k}{2}(k(\theta + 2) + 1). \tag{4.117}
$$

Then in particular

$$Z_{k+1}^{\theta^k} = X_{k+1}^{\theta^k} - \mathbb{E}\left[X_{k+1}^{\theta^k}\right] = k\theta\left(\mathcal{L}_{k+1}^{\theta^k} - \mathbb{E}\left[\mathcal{L}_{k+1}^{\theta^k}\right]\right). \tag{4.118}$$

We have

$$\begin{aligned}
Z_i^{\theta^k} &- Z_{i-1}^{\theta^k} \\
&= X_i^{\theta^k} - X_{i-1}^{\theta^k} \\
&= (k(\theta - 1) + i - 1)\mathcal{L}_i^{\theta^k} - (1 - \theta)\sum_{j=k}^{i-1}\left(\mathbb{E}\left[N_j^{\theta^k}\right] - k\right) - \sum_{j=k}^{i-1} j \\
&\quad - \left[(k(\theta - 1) + i - 2)\mathcal{L}_{i-1}^{\theta^k} - (1 - \theta)\sum_{j=k}^{i-2}\left(\mathbb{E}\left[N_j^{\theta^k}\right] - k\right) - \sum_{j=k}^{i-2} j\right] \\
&= (k(\theta - 1) + i - 1)\mathcal{L}_i^{\theta^k} - (k(\theta - 1) + i - 2)\mathcal{L}_{i-1}^{\theta^k} \\
&\quad - (1 - \theta)\mathbb{E}\left[N_{i-1}^{\theta^k}\right] + (1 - \theta)k - i + 1.
\end{aligned} \tag{4.119}$$

Now we use that

$$\mathcal{L}_i^{\theta^k} = \mathcal{L}_{i-1}^{\theta^k} + Y_i^{\theta^k} \tag{4.120}$$

and

$$\mathbb{E}\left[Y_i^{\theta^k}\right] = 1 - \frac{E\left[\mathcal{L}_{i-1}^{\theta^k}\right] - \mathbb{E}\left[N_{i-1}^{\theta^k}\right]}{k(\theta - 1) + i - 1} - \frac{\theta E\left[N_{i-1}^{\theta^k}\right]}{k(\theta - 1) + i - 1} \tag{4.121}$$

to get

$$
\begin{aligned}
& Z_i^{\theta^k} - Z_{i-1}^{\theta^k} \\
& = Z_i^{\theta^k} - Z_{i-1}^{\theta^k} - (k(\theta - 1) + i - 1) \\
& \qquad \cdot \left[E\left[Y_i^{\theta^k}\right] - \left(1 - \frac{E\left[\mathcal{L}_{i-1}^{\theta^k}\right] - \mathbb{E}\left[N_{i-1}^{\theta^k}\right]}{k(\theta-1)+i-1} - \frac{\theta E\left[N_{i-1}^{\theta^k}\right]}{k(\theta-1)+i-1} \right) \right] \\
& = (k(\theta-1)+i-1)\left(\mathcal{L}_{i-1}^{\theta^k} + Y_i^{\theta^k}\right) - (k(\theta-1)+i-2)\mathcal{L}_{i-1}^{\theta^k} \\
& \quad - (1-\theta)\mathbb{E}\left[N_{i-1}^{\theta^k}\right] + (1-\theta)k - i + 1 - (k(\theta-1)+i-1)E\left[Y_i^{\theta^k}\right] \\
& \quad + k(\theta-1)+i-1 - \mathbb{E}\left[\mathcal{L}_{i-1}^{\theta^k}\right] + \mathbb{E}\left[N_{i-1}^{\theta^k}\right] - \theta\mathbb{E}\left[N_{i-1}^{\theta^k}\right] \\
& = \mathcal{L}_{i-1}^{\theta^k} - \mathbb{E}\left[\mathcal{L}_{i-1}^{\theta^k}\right] + [k(\theta-1)+i-1]\left[Y_i^{\theta^k} - E\left[Y_i^{\theta^k}\right]\right].
\end{aligned}
\tag{4.122}
$$

Since $Z_i^{\theta^k}$ is only defined for $k+1 \le i \le n$, the factor $k(\theta-1)+i-1$, which actually is the sum of all weights at time $i-1$, is positive for all such i. Setting $R_i^{\theta^k} := \mathcal{L}_{i-1}^{\theta^k} - \mathbb{E}\left[\mathcal{L}_{i-1}^{\theta^k}\right] - [k(\theta-1)+i-1]E\left[Y_i^{\theta^k}\right]$, we get

$$
R_i^{\theta^k} \le Z_i^{\theta^k} - Z_{i-1}^{\theta^k} = R_i^{\theta^k} + [k(\theta-1)+i-1]Y_i^{\theta^k} \le R_i^{\theta^k} + k(\theta-1)+i-1. \tag{4.123}
$$

This allows us to derive a concentration inequality via Theorem 2.15. In our case, the martingale starts with $k+1$ and we thus get

$$
\mathbb{P}\left(\left| \sum_{i=k+1}^{n-1} Z_{i+1}^{\theta^k} - Z_i^{\theta^k} \right| \ge t \right) \le 2e^{-\frac{2t^2}{\sum_{i=k+1}^{n}(k(\theta-1)+i-1)^2}} \tag{4.124}
$$

which is equivalent to

$$
\mathbb{P}\left(\left| Z_n^{\theta^k} - Z_{k+1}^{\theta^k} \right| \ge t \right) \le 2e^{-\frac{2t^2}{\sum_{i=k+1}^{n}(k(\theta-1)+i-1)^2}}. \tag{4.125}
$$

We can now bound $\sum_{i=k+1}^{n}(k(\theta-1)+i-1)^2$ by an integral bound. Since $(k(\theta-1)+i-1)^2$ is an increasing function, we have

$$
\begin{aligned}
\sum_{i=k+1}^{n}(k(\theta-1)+i-1)^2 &\leq \int_{k+1}^{n}(k(\theta-1)+x-1)^2+(k(\theta-1)+n-1)^2 \\
&= \frac{(k(\theta-1)+x-1)^3}{3}\bigg|_{x=k+1}^{n}+(k(\theta-1)+n-1)^2 \\
&= \frac{(k(\theta-1)+n-1)^3}{3}-\frac{(k\theta)^3}{3}+(k(\theta-1)+n-1)^2 \\
&\leq \frac{(k(\theta-1)+n-1)^3}{3}+(k(\theta-1)+n-1)^2.
\end{aligned}
\tag{4.126}
$$

Thus,

$$
e^{-\frac{2t^2}{\sum_{i=k+1}^{n}(k(\theta-1)+i-1)^2}} \leq e^{-\frac{2t^2}{\frac{(k(\theta-1)+n-1)^3}{3}+(k(\theta-1)+n-1)^2}} = e^{-\frac{6t^2}{(k(\theta-1)+n-1)^3+3(k(\theta-1)+n-1)^2}}. \tag{4.127}
$$

In order to derive a concentration equality for $\mathcal{L}_n^{\theta^k}$ we notice first of all that

$$
\begin{aligned}
Z_n^{\theta^k} &= (k(\theta-1)+n-1)\mathcal{L}_n^{\theta^k}+(\theta-1)\left(\sum_{i=k+1}^{n}\mathbb{E}\left[N_{i-1}^{\theta^k}\right]-k\right) \\
&\quad -\sum_{i=k+1}^{n}i-1-\frac{k}{2}-\mathbb{E}\left[X_{k+1}^{\theta^k}\right] \\
&= (k(\theta-1)+n-1)\left(\mathcal{L}_n^{\theta^k}-\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right]\right).
\end{aligned}
\tag{4.128}
$$

Moreover

$$
\left|Z_{k+1}^{\theta^k}\right|=\left|\theta k\left(\mathcal{L}_{k+1}^{\theta^k}-\mathbb{E}\left[\mathcal{L}_{k+1}^{\theta^k}\right]\right)\right|\leq\theta k\left(k-\frac{k+1}{2}\right)=\frac{\theta k(k-1)}{2} \tag{4.129}
$$

where we use that the expected number of leaves is that of a URT until time $k+1$.

Thus for all $t \geq 0$,

$$
\begin{aligned}
\mathbb{P}&\left(\left|\mathcal{L}_n^{\theta^k} - \mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right]\right| \geq t\right) \\
&= \mathbb{P}\left(k(\theta-1)+n-1\right)\left|\mathcal{L}_n^{\theta^k} - \mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right]\right| \geq (k(\theta-1)+n-1)t\right) \\
&= \mathbb{P}\left(\left|Z_n^{\theta^k}\right| \geq (k(\theta-1)+n-1)t\right) \\
&= \mathbb{P}\left(\left|Z_n^{\theta^k}\right| - \left|Z_{k+1}^{\theta^k}\right| \geq (k(\theta-1)+n-1)t - \left|Z_{k+1}^{\theta^k}\right|\right) \\
&\leq \mathbb{P}\left(\left|Z_n^{\theta^k}\right| - \left|Z_{k+1}^{\theta^k}\right| \geq (k(\theta-1)+n-1)t - \frac{k\theta(k-1)}{2}\right) \\
&\leq \mathbb{P}\left(\left|\left|Z_n^{\theta^k}\right| - \left|Z_{k+1}^{\theta^k}\right|\right| \geq (k(\theta-1)+n-1)t - \frac{k\theta(k-1)}{2}\right) \\
&\leq \mathbb{P}\left(\left|Z_n^{\theta^k} - Z_{k+1}^{\theta^k}\right| \geq (k(\theta-1)+n-1)t - \frac{k\theta(k-1)}{2}\right) \\
&\leq 2e^{-\frac{6\left((k(\theta-1)+n-1)t - \frac{k\theta(k-1)}{2}\right)^2}{(k(\theta-1)+n-1)^3 + 3(k(\theta-1)+n-1)^2}} \\
&= 2e^{-6\frac{(k(\theta-1)+n-1)^2 t^2 + \left(\frac{k\theta(k-1)}{2}\right)^2 - 2(k(\theta-1)+n-1)t\frac{k\theta(k-1)}{2}}{(k(\theta-1)+n-1)^3 + 3(k(\theta-1)+n-1)^2}} \\
&= 2e^{-6\frac{t^2}{(k(\theta-1)+n-1)+3}} e^{-3\frac{(k\theta(k-1))^2 - 4(k(\theta-1)+n-1)t(k\theta(k-1))}{2(k(\theta-1)+n-1)^3 + 6(k(\theta-1)+n-1)^2}} \\
&\leq 2e^{-6\frac{t^2}{k(\theta-1)+n+2}} e^{\frac{12t(k\theta(k-1))}{2(k(\theta-1)+n-1)^2 + 6(k(\theta-1)+n-1)}} \\
&\leq 2e^{-6\frac{t^2}{k(\theta-1)+n+2}} e^{\frac{6t(k\theta(k-1))}{(k(\theta-1)+n-1)(k(\theta-1)+n+2)}}.
\end{aligned}
\tag{4.130}
$$

We will now calculate the variance, $\mathrm{Var}\left(\mathcal{L}_n^{\theta^k}\right)$. We will use that

$$
Z_n^{\theta^k} = \frac{k(\theta-1)+n-1}{k(\theta-1)+n-2} Z_{n-1}^{\theta^k} + [k(\theta-1)+n-1]\left(Y_n^{\theta^k} - \mathbb{E}\left[Y_n^{\theta^k}\right]\right)
\tag{4.131}
$$

which we obtain by observing the following. First of all

$$
Z_n^{\theta^k} = Z_{n-1}^{\theta^k} + \mathcal{L}_{n-1}^{\theta^k} - \mathbb{E}\left[\mathcal{L}_{n-1}^{\theta^k}\right] + [k(\theta-1)+n-1]\left(Y_n^{\theta^k} - \mathbb{E}\left[Y_n^{\theta^k}\right]\right).
\tag{4.132}
$$

Also

$$\mathcal{L}_{n-1}^{\theta^k} = \frac{1}{k(\theta-1)+n-2}\left(Z_{n-1}^{\theta^k} + E\left[X_{k+1}^{\theta^k}\right]\right.$$
$$\left. - (\theta-1)\left(\sum_{i=k}^{n-2}\mathbb{E}\left[N_i^{\theta^k}\right] - k\right) + \sum_{i=k}^{n-2}i\right) \tag{4.133}$$

and similarly

$$\mathbb{E}\left[\mathcal{L}_{n-1}^{\theta^k}\right] = \frac{1}{k(\theta-1)+n-2}\left(\mathbb{E}\left[Z_{n-1}^{\theta^k}\right] + E\left[X_{k+1}^{\theta^k}\right]\right.$$
$$\left. - (\theta-1)\left(\sum_{i=k}^{n-2}\mathbb{E}\left[N_i^{\theta^k}\right] - k\right) + \sum_{i=k}^{n-2}i\right). \tag{4.134}$$

Moreover we use that for all n

$$\mathbb{E}\left[Z_n^{\theta^k}\right] = 0. \tag{4.135}$$

Using (4.131) we get,

$$\mathbb{E}\left[Z_n^{\theta^k\,2}\right] = \left(\frac{k(\theta-1)+n-1}{k(\theta-1)+n-2}\right)^2\mathbb{E}\left[Z_{n-1}^{\theta^k\,2}\right]$$
$$+ 2\frac{(k(\theta-1)+n-1)^2}{k(\theta-1)+n-2}\mathbb{E}\left[Z_{n-1}^{\theta^k}\left(Y_n^{\theta^k} - \mathbb{E}\left[Y_n^{\theta^k}\right]\right)\right] \tag{4.136}$$
$$+ (k(\theta-1)+n-1)^2\mathbb{E}\left[\left(Y_n^{\theta^k} - \mathbb{E}\left[Y_n^{\theta^k}\right]\right)^2\right].$$

Using again $\mathbb{E}\left[Z_n^{\theta^k}\right] = 0$ for all n and the fact that $Z_{n-1}^{\theta^k}$ is $\mathcal{L}_1^{\theta^k}, \ldots, \mathcal{L}_{n-1}^{\theta^k}$ measurable, we get

$$
\begin{aligned}
\mathbb{E}&\left[Z_{n-1}^{\theta^k}\left(Y_n^{\theta^k} - \mathbb{E}\left[Y_n^{\theta^k}\right]\right)\right] \\
&= \mathbb{E}\left[Z_{n-1}^{\theta^k} Y_n^{\theta^k}\right] - \mathbb{E}\left[Z_{n-1}^{\theta^k} \mathbb{E}\left[Y_n^{\theta^k}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Z_{n-1}^{\theta^k} Y_n^{\theta^k}\,\Big|\,\mathcal{L}_1^{\theta^k}, \ldots, \mathcal{L}_{n-1}^{\theta^k}\right]\right] \\
&= \mathbb{E}\left[Z_{n-1}^{\theta^k} \mathbb{E}\left[Y_n^{\theta^k}\,\Big|\,\mathcal{L}_1^{\theta^k}, \ldots, \mathcal{L}_{n-1}^{\theta^k}\right]\right] \\
&= \mathbb{E}\left[Z_{n-1}^{\theta^k}\left(1 - \frac{\mathcal{L}_{n-1}^{\theta^k}}{k\theta + n - 1 - k} + \frac{(1-\theta)\mathbb{E}\left[N_{n-1}^{\theta^k}\right]}{k\theta + n - 1 - k}\right)\right] \\
&= \mathbb{E}\left[Z_{n-1}^{\theta^k}\right] - \frac{\mathbb{E}\left[Z_{n-1}^{\theta^k}\mathcal{L}_{n-1}^{\theta^k}\right]}{k\theta + n - 1 - k} + \mathbb{E}\left[Z_{n-1}^{\theta^k}\right]\frac{(1-\theta)\mathbb{E}\left[N_{n-1}^{\theta^k}\right]}{k\theta + n - 1 - k} \\
&= -\mathbb{E}\left[\frac{Z_{n-1}^{\theta^k}}{k\theta + n - 1 - k}\right. \\
&\qquad \left. \cdot\left(\frac{Z_{n-1}^{\theta^k} + \mathbb{E}\left[X_{k+1}^{\theta^k}\right] + (\theta-1)\sum_{i=k}^{n-2}\left(\mathbb{E}\left[N_i^{\theta^k}\right] - k\right) - \sum_{i=k}^{n-2} i}{k(\theta-1) + n - 2}\right)\right] \\
&= -\frac{\mathbb{E}\left[Z_{n-1}^{\theta^k\,2}\right]}{(k(\theta-1) + n - 1)(k(\theta-1) + n - 2)}.
\end{aligned}
\tag{4.137}
$$

Moreover

$$
\begin{aligned}
\mathbb{E}\left[Y_n^{\theta^k}\right] &= 1 - \frac{\mathbb{E}\left[\mathcal{L}_{n-1}^{\theta^k}\right]}{k(\theta-1) + n - 1} + \frac{(1-\theta)\mathbb{E}\left[N_{n-1}^{\theta^k}\right]}{k(\theta-1) + n - 1} \\
&= 1 - \frac{\mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(1) + \frac{n}{2}}{k(\theta-1) + n - 1} + \frac{(1-\theta)\mathbb{E}\left[N_{n-1}^{\theta^k}\right]}{k(\theta-1) + n - 1} \\
&= \frac{1}{2} + \mathcal{O}\left(\frac{1}{n}\right)
\end{aligned}
\tag{4.138}
$$

and

$$\begin{aligned}
\operatorname{Var}\left(Y_n^{\theta^k}\right) &= \mathbb{E}\left[Y_n^{\theta^k}\right] - \mathbb{E}\left[Y_n^{\theta^k}\right]^2 \\
&= \frac{1}{2} + \mathcal{O}\left(\frac{1}{n}\right) - \left(\frac{1}{2} + \mathcal{O}\left(\frac{1}{n}\right)\right)^2 \\
&= \frac{1}{4} + \mathcal{O}\left(\frac{1}{n}\right).
\end{aligned} \tag{4.139}$$

We thus get the recursion

$$\begin{aligned}
\mathbb{E}\left[Z_n^{\theta^k\,2}\right] &= \left(\frac{k(\theta-1)+n-1}{k(\theta-1)+n-2}\right)^2 \mathbb{E}\left[Z_{n-1}^{\theta^k\,2}\right] \\
&\quad - 2\frac{(k(\theta-1)+n-1)^2}{k(\theta-1)+n-2}\frac{\mathbb{E}\left[Z_{n-1}^{\theta^k\,2}\right]}{(k(\theta-1)+n-1)(k(\theta-1)+n-2)} \\
&\quad + (k(\theta-1)+n-1)^2\left(\frac{1}{4}+\mathcal{O}\left(\frac{1}{n}\right)\right) \\
&= \frac{(k(\theta-1)+n-1)^2 - 2(k(\theta-1)+n-1)}{(k(\theta-1)+n-2)^2}\mathbb{E}\left[Z_{n-1}^{\theta^k\,2}\right] \\
&\quad + (k(\theta-1)+n-1)^2\left(\frac{1}{4}+\mathcal{O}\left(\frac{1}{n}\right)\right) \\
&= \frac{(k(\theta-1)+n-3)(k(\theta-1)+n-1)}{(k(\theta-1)+n-2)^2}\mathbb{E}\left[Z_{n-1}^{\theta^k\,2}\right] \\
&\quad + (k(\theta-1)+n-1)^2\left(\frac{1}{4}+\mathcal{O}\left(\frac{1}{n}\right)\right).
\end{aligned} \tag{4.140}$$

This is equivalent to

$$\begin{aligned}
\frac{k(\theta-1)+n-2}{k(\theta-1)+n-1}&\mathbb{E}\left[Z_n^{\theta^k\,2}\right] \\
&= \frac{k(\theta-1)+n-3}{k(\theta-1)+n-2}\mathbb{E}\left[Z_{n-1}^{\theta^k\,2}\right] \\
&\quad + (k(\theta-1)+n-2)\,(k(\theta-1)+n-1)\left(\frac{1}{4}+\mathcal{O}\left(\frac{1}{n}\right)\right).
\end{aligned} \tag{4.141}$$

Now set

$$W_n^{\theta^k} := \frac{k(\theta-1)+n-2}{k(\theta-1)+n-1}\mathbb{E}\left[Z_n^{\theta^k\,2}\right] \tag{4.142}$$

then we have

$$W_n^{\theta^k} = W_{n-1}^{\theta^k} + (k(\theta - 1) + n - 2)(k(\theta - 1) + n - 1)\left(\frac{1}{4} + \mathcal{O}\left(\frac{1}{n}\right)\right). \qquad (4.143)$$

This is satisfied for $n \geq 2$ by

$$W_n^{\theta^k} = \sum_{i=2}^{n}(k(\theta - 1) + i - 2)(k(\theta - 1) + i - 1)\left(\frac{1}{4} + \mathcal{O}\left(\frac{1}{n}\right)\right). \qquad (4.144)$$

Hence

$$\mathbb{E}\left[Z_n^{\theta^k 2}\right] = \frac{k(\theta - 1) + n - 1}{k(\theta - 1) + n - 2}$$
$$\cdot \sum_{i=2}^{n}(k(\theta - 1) + i - 2)(k(\theta - 1) + i - 1)\left(\frac{1}{4} + \mathcal{O}\left(\frac{1}{n}\right)\right). \qquad (4.145)$$

And thus

$$
\begin{aligned}
&\operatorname{Var}\left(\mathcal{L}_n^{\theta^k}\right) \\
&= \operatorname{Var}\left(\frac{1}{k(\theta-1)+n-1}\left[Z_n^{\theta^k} - (\theta-1)\sum_{i=k}\left(\mathbb{E}\left[N_i^{\theta^k}\right]-k\right)\right.\right. \\
&\hspace{6cm}\left.\left. + \sum_{i=k}^{n-1} i + \mathbb{E}\left[X_{k+1}^{\theta^k}\right]\right]\right) \\
&= \frac{1}{(k(\theta-1)+n-1)^2}\left(\mathbb{E}\left[Z_n^{\theta^k\,2}\right] - \mathbb{E}\left[Z_n^{\theta^k}\right]^2\right) \\
&= \frac{1}{(k(\theta-1)+n-1)^2}\frac{k(\theta-1)+n-1}{k(\theta-1)+n-2} \\
&\qquad \sum_{i=2}^{n}(k(\theta-1)+i-2)\,(k(\theta-1)+i-1)\left(\frac{1}{4}+\mathcal{O}\left(\frac{1}{n}\right)\right) \\
&= \frac{1}{(k(\theta-1)+n-1)(k(\theta-1)+n-2)} \\
&\qquad \sum_{i=2}^{n}(k(\theta-1)+i-2)\,(k(\theta-1)+i-1)\left(\frac{1}{4}+\mathcal{O}\left(\frac{1}{n}\right)\right) \\
&= \frac{1}{(k(\theta-1)+n-1)(k(\theta-1)+n-2)}\frac{1}{4}\sum_{i=2}^{n}i^2 + \mathcal{O}(1) \\
&= \frac{1}{(k(\theta-1)+n-1)(k(\theta-1)+n-2)}\frac{1}{4}\left(\frac{(2n+1)(n+1)n}{6}-1\right) + \mathcal{O}(1) \\
&= \frac{n}{12} + \mathcal{O}(1).
\end{aligned}
\tag{4.146}
$$

$\square$

Remark 4.17. We can also get some results by writing $\mathcal{L}_n^{\omega}$ as the sum of $\mathbf{1}(\ell_i^{\omega})$ where ℓ_i^{ω} denotes the event that i is a leaf in a WRT. It follows from the construction principle that

$$
\mathbb{P}\left(\ell_i^{\omega}\right) = \prod_{j=i+1}^{n}\left(1 - \frac{\omega_i}{\omega_1+\cdots+\omega_{j-1}}\right).
\tag{4.147}
$$

After some manipulation this expression becomes

$$
\mathbb{P}\left(l_i^{\omega}\right) = \frac{\omega_1+\cdots+\omega_{i-1}}{\omega_1+\cdots+\omega_{n-1}}\prod_{j=i+1}^{n-1}\left(1 + \frac{\omega_j-\omega_i}{\omega_1+\cdots+\omega_{j-1}}\right).
\tag{4.148}
$$

Moreover we get the exact expression for the expectation of the number of leaves of a θ^k-tree by writing $\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right] = \sum_{i=2}^{n} \mathbb{E}\left[\mathbf{1}\left(\ell_i^{\theta^k}\right)\right]$ and using the above expression. After some computations we get by this method

$$\mathbb{E}\left[\mathcal{L}_n^{\theta^k}\right] = \frac{n}{2} + \frac{k(\theta-1)}{2} + \frac{k\theta(1-k\theta)}{2(k(\theta-1)+n-1)} + \frac{k-1}{2} \prod_{i=1}^{n-1-k} \frac{\theta(k-1)+i}{\theta k + i}. \quad (4.149)$$

Remark 4.18. It is possible to derive a CLT by using this martingale, but we will now introduce another coupling construction, that allows much easier inferences concerning the distribution of the number of leaves.

4.6. A Coupling View of Some Special Kinds of WRTs

We will now introduce three coupling constructions for special kinds of WRTs. They all have in common that they can only be applied to WRTs for which there is a $k \in \mathbb{N}$ such that the weights are constant for all nodes i with $i > k$. This is a quite restrictive assumption, but we will see that the first and the third coupling allow very easy conclusions concerning several statistics.

4.6.1. Construction of a θ^k-RT from a URT in the Case $\theta \in \mathbb{N}^+$

In the case where $\theta \in \mathbb{N}$, we can use the following reconstruction to get a θ^k-RT from a URT. To emphasize the assumption, we write $\theta = m$ from now on. First construct a URT on $mk + n - k$ nodes. We write $\mathcal{T}_{mk+n-k}$ for this URT. Since we want the weight of the first k nodes to be m we then join several nodes into one in the following way. To avoid confusion let us denote the nodes in the URT by i and the nodes in the reconstructed tree by i^T. Then we have the following:

- The nodes $1, \ldots, m$ will become the node 1^T,
- $m+1, \ldots, 2m$ will become the node 2^T $\ldots$
- $(k-1)m+1, \ldots, km$ become k^T.

The new node i^T gets all the children of $(i-1)m+1, \ldots, im$. But since we joined several nodes into one and the nodes $(j-1)m+1, \ldots, jm$ might have different parents, for $1 < j \le k$, we set the parent of j^T as the parent of $(j-1)m+1$, i.e. of the node with the smallest label among those that become j^T. If in the URT the parent of $(j-1)m+1$ is any of the nodes $(i-1)m+1, \ldots, im$, the parent of j^T is i^T.

For $j > k$, we set $j^T = j + k(m-1)$, so all nodes after k only correspond to a single node, we just need to "translate" the names of the nodes to take into account that we used mk nodes instead of k for the first k nodes in the reconstructed tree. If the parent of node $j + k(m-1)$ for $j > k$ is among the first km nodes of the URT, we check into which range this node falls and the parent of j^T is chosen as above. In other words if, for $1 \le i \le k$, the parent of node $j+k(m-1)$ is one of $(i-1)m+1, \ldots, im$, the parent of node j^T is node i^T. If the parent of node $j+k(m-1)$ is equal to $h+k(m-1)$ with $h > k$, the parent of node j^T is node h^T. The tree we thus obtain is called $\mathcal{T}_n^{m^k}$. We can easily verify that the probabilities are right.

- For $i < j \le k$,

$$\mathbb{P}(j^T \text{ attaches to } i^T)$$
$$= \mathbb{P}((j-1)m+1 \text{ attaches to any of the nodes } (i-1)m+1, \ldots, im)$$
$$= \frac{m}{(j-1)m} \tag{4.150}$$
$$= \frac{1}{j-1}.$$

- For $i \le k < j$,

$$\mathbb{P}(j^T \text{ attaches to } i^T)$$
$$= \mathbb{P}(j+k(m-1) \text{ attaches to any of the nodes } (i-1)m+1, \ldots, im)$$
$$= \frac{m}{j+k(m-1)-1} \tag{4.151}$$
$$= \frac{m}{j-1-k+km}.$$

- For $k < i < j$,

$$\mathbb{P}(j^T \text{ attaches to } i^T)$$
$$= \mathbb{P}(j + k(m-1) \text{ attaches to } i + k(m-1))$$
$$= \frac{1}{j + k(m-1) - 1}$$
$$= \frac{1}{j - 1 - k + km}. \tag{4.152}$$

Let now $\mathcal{L}_{mk+n-k}$ denote the number of leaves of $\mathcal{T}_{mk+n-k}$ and $\mathcal{L}_n^{m^k}$ denote the number of leaves of $\mathcal{T}_n^{m^k}$. Then $\mathcal{L}_{mk+n-k}$ can be used to bound $\mathcal{L}_n^{m^k}$. First of all if a node $i > km$ is a leaf in $\mathcal{T}_{mk+n-k}$, the corresponding node in $\mathcal{T}_n^{m^k}$, which is $i - k(m-1)^T$, is also a leaf. The reconstruction process thus only affects the children of the nodes i^T with $1 \leq i \leq k$, so we can have at most k additional leaves.

For $2 \leq i \leq k$, we note that $2m + 1, 3m + 1, \ldots, (k-1)m + 1$, determine the parent of $3^T, 4^T, \ldots k^T$ respectively. Similarly the nodes $km+1, km+2, \ldots, km+n-k$ determine the parent of $(k+1)^T, (k+2)^T, \ldots n^T$ respectively. Hence if any of the above are children of one of $(i-1)m+1, (i-1)m+2, \ldots, im$, the node i^T will not be a leaf. Thus for $i \leq k$, the node i^T is a leaf if and only if there is no node with the label $2m+1, 3m+1, \ldots, (k-1)m+1$ and no node with label $km+1, km+2, \ldots, km+n-k$ that attaches to any of the nodes $(i-1)m+1, (i-1)m+2, \ldots, im$. This can theoretically be used to calculate the exact probability that i^T is leaf for $1 \leq i \leq k$.

But without calculating these exact probabilities we can conclude the following. For each $2 \leq i \leq k$ we can at most "loose" $m-1$ leaves since if all $(i-1)m+1, \ldots, im$ are leaves in $\mathcal{T}_{mk+n-k}$, i^T will be a leaf in $\mathcal{T}_n^{m^k}$. Hence we can conclude that

$$\mathcal{L}_{mk+n-k} + k > \mathcal{L}_n^{m^k} > \mathcal{L}_{mk+n-k} - k(m-1). \tag{4.153}$$

Let moreover $\mathcal{L}_n$ denote the number of leaves in a URT on n nodes, then

$$
\begin{aligned}
& \mathbb{E}[\mathcal{L}_{mk+n-k}] + k > \mathbb{E}\left[\mathcal{L}_n^{m^k}\right] > \mathbb{E}[\mathcal{L}_{mk+n-k}] - k(m-1) \\
& \Leftrightarrow \frac{n+k(m-1)}{2} + k > \mathbb{E}\left[\mathcal{L}_n^{m^k}\right] > \frac{n+k(m-1)}{2} - k(m-1) \\
& \Leftrightarrow \frac{n+k(m+1)}{2} > \mathbb{E}\left[\mathcal{L}_n^{m^k}\right] > \frac{n-k(m-1)}{2} \\
& \Leftrightarrow \mathbb{E}[\mathcal{L}_n] + \frac{k(m+1)}{2} > \mathbb{E}\left[\mathcal{L}_n^{m^k}\right] > \mathbb{E}[\mathcal{L}_n] - \frac{k(m-1)}{2} \\
& \Rightarrow \left| \mathbb{E}[\mathcal{L}_n^{m^k}] - \mathbb{E}[\mathcal{L}_n] \right| \leq \frac{k(m+1)}{2}.
\end{aligned}
\tag{4.154}
$$

4.6.2. Construction of a θ^k-RT from a URT in the Case $\frac{1}{\theta} \in \mathbb{N}^+$

In this case the construction turns out to be a little bit more complicated. When using the idea introduced above, in order to reduce the weight of the first k nodes given a URT on an appropriate number of nodes we will now specify, we need to increase the weight of the remaining $n - k$ nodes. This means that the reconstruction affects a much bigger number of nodes. This in turn implies that the coupling is not as useful as the previous one but we nevertheless introduce it, to show the same principle can be applied.

We now assume that the first k nodes have weight θ with $\theta = \frac{1}{m}$ for some $m \in \mathbb{N}$ and that the remaining ones weight 1. Such a recursive tree on n nodes has a total weight of $k\frac{1}{m} + (n-k)$. Since the actual value of the weights does not matter, but only the weights relative to each other, it is equivalent to construct a tree with the first k nodes having weight 1 and the remaining ones m. This would result in a total weight of $k + (n-k)m$. So we first construct a URT with $nm - k(m-1)$ nodes.

We leave the first k nodes as they are, so $i^T = i$ for $i \leq k$. Then, for $i \geq k+1$ we need to join several nodes together. In particular:

- Nodes $k+1, \ldots, k+m$ are joined together to form node $(k+1)^T$,
- nodes $k+m+1, \ldots, k+2m$ are joined together to form node $(k+2)^T$ ...

- nodes $k + (i - k - 1)m + 1, \ldots, k + (i - k)m$ form node i^T $\ldots$
- nodes $k + (n - k - 1)m + 1, \ldots, k + (n - k)m$ form node n^T.

Again, as before, the parent of i^T in the reconstructed tree is the parent of $k + (i - k - 1)m + 1$ in the URT and all children of the nodes we join together become children of the new node. Except for the first k nodes, we need to remember that only if a node has a child with label $k + \ell m + 1$ for some $0 \leq \ell \leq n - k - 1$, the corresponding node will also have a child in the reconstructed tree.

We now check that this rearrangement gives us the right attachment probabilities.

- For $i < j \leq k+1$ both the attachment probabilities in the URT and the weighted recursive tree correspond to a URT, so there is nothing to check.
- For $i \leq k < j$,

$$
\begin{aligned}
&\mathbb{P}\left(j^T \text{ attaches to } i^T\right) \\
&= \mathbb{P}(k + (j - k - 1)m + 1 \text{ attaches to } i) \\
&= \frac{1}{k + (j - k - 1)m} \\
&= \frac{\frac{1}{m}}{\frac{k}{m} + j - k - 1}.
\end{aligned}
\tag{4.155}
$$

- For $k < i < j$,

$$
\begin{aligned}
&\mathbb{P}(j^T \text{ attaches to } i^T) \\
&= \mathbb{P}(k + (j - k - 1)m + 1 \text{ attaches to} \\
&\qquad \text{one of } k + (i - k - 1)m + 1, \ldots, k + (i - k)m) \\
&= \frac{m}{k + (j - k - 1)m} \\
&= \frac{1}{\frac{k}{m} + j - k - 1}.
\end{aligned}
\tag{4.156}
$$

While this reconstruction gives the tree we wanted, i.e. $\mathcal{T}_n^{\frac{1}{m}^k}$, it does not allow easy conclusions concerning the number of leaves since the coupling affects the children of

all but a finite number of nodes.

4.6.3. Constructing a Special Kind of WRT from a Hoppe Tree

Finally we will introduce a third coupling construction that can be applied for a special case of WRTs that includes the two models presented above. We will now rearrange a Hoppe tree instead of a URT in order to get a WRT. Let $(\omega_i)_{i \in \mathbb{N}}$ be such that there is a $k \in \mathbb{N}$ such that $\omega_i = 1$ for $i > k$. We can then construct the corresponding WRT from a Hoppe tree using an inverse process than above. Instead of joining nodes into one we will now split the root into k nodes.

First construct a Hoppe tree with $n - k + 1$ nodes and with θ, the weight of the root, equal to $\sum_{i=1}^{k} \omega_i$. Then construct a weighted recursive tree of size k corresponding to $(\omega_i)_{i \in \mathbb{N}}$. Now we replace the root of the Hoppe tree by this weighted recursive tree of size k. For $i \geq 2$, node i in the Hoppe tree becomes node $i + k - 1^T$ in the reconstructed tree. Then for all $i \geq 2$, if i is a child of 1 in the Hoppe tree, $i + k - 1^T$ becomes a child of one of the nodes $1^T, \ldots, k^T$ in the reconstructed tree proportional to their weights. This means that if i is a child of 1 in the Hoppe tree, $i + k - 1^T$ will become a child of node j^T in the reconstructed tree with probability $\frac{\omega_j}{\sum_{\ell=1}^{k} \omega_\ell}$. Let us check that this gives the right probabilities.

- For $1 \leq i \leq k < j$,

$$
\begin{aligned}
\mathbb{P}(j^T &\text{ is child of } i^T \text{ in reconstructed tree}) \\
&= \mathbb{P}(j - k + 1 \text{ is child of 1 in Hoppe tree}) \frac{\omega_i}{\sum_{\ell=1}^{k} \omega_\ell} \\
&= \frac{\sum_{\ell=1}^{k} \omega_\ell}{j - k + 1 - 2 + \sum_{\ell=1}^{k} \omega_\ell} \frac{\omega_i}{\sum_{\ell=1}^{k} \omega_\ell} \\
&= \frac{\omega_i}{j - 1 - k + \sum_{\ell=1}^{k} \omega_\ell}.
\end{aligned}
\tag{4.157}
$$

- For $k < i < j$,

$$\mathbb{P}(j^T \text{ is child of } i^T \text{ in reconstructed tree})$$
$$= \mathbb{P}(j - k + 1 \text{ is child of } i - k + 1 \text{ in Hoppe tree})$$
$$= \frac{1}{j - k + 1 - 2 + \sum_{\ell=1}^{k} \omega_\ell}$$
$$= \frac{1}{j - 1 + \sum_{\ell=1}^{k} \omega_\ell - k}. \qquad (4.158)$$

4.6.4. Using the Coupling for WRT Statistics Analysis

The coupling constructions given in previous subsections are useful in understanding various WRT statistics with the aid of well known results on URTs and Hoppe trees. We already gave an example in subsection 4.6.1. We demonstrate this briefly here for the number of leaves and the height of a WRT $\mathcal{T}_n^\omega$ with $(\omega_i)_{i \in \mathbb{N}}$ such that $\omega_i = 1$ for $i > k$.

First focusing on the number of leaves, the reconstruction process implies that all the leaves of the Hoppe tree are still leaves in the reconstructed tree, since we do not change any relation among the nodes $2, \ldots, n - k + 1$ of the Hoppe tree or respectively $k + 1^T, \ldots n^T$ of the reconstructed tree. There can be at most $k - 1$ additional leaves among the first k nodes. Thus,

$$\mathcal{L}_{n-k+1}^\theta \leq \mathcal{L}_n^\omega \leq \mathcal{L}_{n-k+1}^\theta + k - 1$$
$$\Rightarrow \frac{n - k + 1}{2} + \frac{\sum_{i=1}^k \omega_i - 1}{2} + \mathcal{O}\left(\frac{1}{n}\right) \leq \mathbb{E}\left[\mathcal{L}_n^\omega\right]$$
$$\leq \frac{n - k + 1}{2} + \frac{\sum_{i=1}^k \omega_i - 1}{2} + k - 1 + \mathcal{O}\left(\frac{1}{n}\right) \qquad (4.159)$$
$$\Rightarrow \frac{n}{2} + \frac{\sum_{i=1}^k \omega_i - k}{2} + \mathcal{O}\left(\frac{1}{n}\right) \leq \mathbb{E}\left[\mathcal{L}_n^\omega\right] \leq \frac{n}{2} + \frac{\sum_{i=1}^k \omega_i + k}{2} - 1 + \mathcal{O}\left(\frac{1}{n}\right)$$
$$\Rightarrow \mathbb{E}\left[\mathcal{L}_n^\omega\right] = \frac{n}{2} + \mathcal{O}(1).$$

In a similar way, one can make conclusions about the variance, concentration and asymptotic distribution of the number of leaves in a WRT with $\omega_i = 1$ for $i > k$ for

some $k \in \mathbb{N}$.

Theorem 4.19. *Let $\mathcal{L}_n^\omega$ denote the number of leaves of a WRT of size n with $(\omega_i)_{i \in \mathbb{R}}$ such that there is a $k \in \mathbb{N}$ such that for all $i > k$ we have $\omega_i = 1$. Then $\mathcal{L}_n^\omega$ is asymptotically normal.*

Proof. In order to derive a central limit theorem for $\mathcal{L}_n^\omega$, we write

$$\frac{\mathcal{L}_n^\omega - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]}{\sqrt{\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta)}} = \frac{\mathcal{L}_{n-k+1}^\omega - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]}{\sqrt{\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta)}} - \frac{\mathcal{L}_{n-k+1}^\omega - \mathcal{L}_n^\theta}{\sqrt{\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta)}}. \tag{4.160}$$

Now we have by Theorem 3.17 that

$$\frac{\mathcal{L}_{n-k+1}^\omega - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]}{\sqrt{\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta)}} \xrightarrow[d]{n \to \infty} \mathcal{G} \tag{4.161}$$

and by (4.159) that

$$\left| \frac{\mathcal{L}_{n-k+1}^\omega - \mathcal{L}_n^\theta}{\sqrt{\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta)}} \right| \leq \frac{k}{\sqrt{\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta)}} \xrightarrow{a.s.} 0 \tag{4.162}$$

since $\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta) \xrightarrow{n \to \infty} \infty$. Now we can apply Slutsky's theorem, which we stated as Theorem 2.16 and conclude that

$$\frac{\mathcal{L}_n^\omega - \mathbb{E}[\mathcal{L}_{n-k+1}^\theta]}{\sqrt{\mathrm{Var}(\mathcal{L}_{n-k+1}^\theta)}} \xrightarrow[d]{n \to \infty} \mathcal{G}. \tag{4.163}$$

$\square$

As a second example, we discuss the height. Let $\mathcal{H}_n^\omega$ denote the height of a WRT with $(\omega_i)_{i \in \mathbb{N}}$ such that $\omega_i = 1$ for $i > k$. Let moreover $\mathcal{D}_{1,i}^\omega$ and $\mathcal{D}_{1,i}^\theta$ denote the distance between the root and node i in the reconstructed tree and the original tree respectively. Then for any node i^T, the path from the root to i^T corresponds to the path from the

root to the corresponding node in the original tree, $i - k + 1$, except that we might have an additional path among the first k nodes instead of the first edge. Thus $\mathcal{D}_{1,i}^\omega$ is at least as big as $\mathcal{D}_{i-k+1}^\theta$.

Also $\mathcal{D}_{1,i}^\omega$ is at most $k - 1$ bigger than the distance between the root and the corresponding node in the original tree: For $i \leq k$, $\mathcal{D}_{1,i}^\omega$ is at most $k - 1$ anyway. For $i > k$, node $i - k + 1$ in the original tree becomes node i^T in the reconstructed tree. Let $j - k + 1$ be the first node on the path from 1 to $i - k + 1$ in the original tree. Then in the reconstructed tree j^T will be attached to some h, where $1 \leq h \leq k$. Thus for all $k + 1 \leq i \leq n$, there is some $h \leq k$ such that,

$$D_{1,i}^\omega = \mathcal{D}_{1,h}^\omega + 1 + \mathcal{D}_{j-k+1,i-k+1}^\theta = \mathcal{D}_{1,h}^\omega + \mathcal{D}_{1,i-k+1}^\theta. \tag{4.164}$$

Also $\mathcal{D}_{1,h}^\omega \leq k - 1$, so we have

$$\mathcal{D}_{1,i-k+1}^\theta \leq D_{1,i}^\omega \leq \mathcal{D}_{1,i-k+1}^\theta + k - 1 \tag{4.165}$$

which implies that

$$\max_{i=1,\ldots,n-k+1}\{\mathcal{D}_{1,i}\} \leq \max_{i=1,\ldots,n}\{\mathcal{D}_{1,i}\} \leq \max_{i=1,\ldots,n-k+1}\{\mathcal{D}_{1,i}\} + k - 1. \tag{4.166}$$

Thus from $\mathcal{H}_n = \max_{i=1,\ldots,n}\{\mathcal{D}_{1,i}\}$, we can derive that

$$\mathcal{H}_{n-k+1}^\theta \leq \mathcal{H}_n^\omega \leq \mathcal{H}_{n-k+1}^\theta + k - 1. \tag{4.167}$$

According to Theorem 3.16 we have $\mathbb{E}[\mathcal{H}_n^\theta] = e\ln(n) - \frac{3}{2}\ln\ln(n) + \mathcal{O}(1)$. Now we have

$$
\begin{aligned}
e\ln(n - &k + 1) - \frac{3}{2}\ln\ln(n - k + 1) \\
&= e\left(\ln(n) + \ln\left(1 - \frac{k-1}{n}\right)\right) \\
&\quad - \frac{3}{2}\left(\ln\ln(n) + \ln\left(1 + \frac{\ln\left(1 - \frac{k-1}{n}\right)}{\ln(n)}\right)\right) \\
&= e\ln(n) - \frac{3}{2}\ln\ln(n) + \mathcal{O}(1).
\end{aligned}
\tag{4.168}
$$

Thus we get for $k > n$,

$$
e\ln(n) - \frac{3}{2}\ln\ln(n) + \mathcal{O}(1) \leq \mathbb{E}\left[\mathcal{H}_n^\omega\right] \leq e\ln(n) - \frac{3}{2}\ln\ln(n) + \mathcal{O}(1) + k - 1 \tag{4.169}
$$

which implies

$$
\mathbb{E}\left[\mathcal{H}_n^\omega\right] = e\ln(n) - \frac{3}{2}\ln\ln(n) + \mathcal{O}(1). \tag{4.170}
$$

The CLT for $\mathcal{H}_n^\omega$ can be derived similarly as above by using Slutsky's Theorem.

5. BIASED RECURSIVE TREES

In Chapter 3 we introduced the representation of recursive trees as permutations and showed that there is a bijection between uniform random permutations and uniform random recursive trees. In [12] Altok and Işlak raise the question how the recursive trees change when a different distribution on S_n is chosen. In particular, the properties of the random recursive trees that are obtained from a riffle shuffle distribution on S_n are studied.

In this chapter after introducing riffle-shuffle permutations, we will be analyzing the associated p-biased recursive trees. Our study of this tree model will begin by reviewing some results from [12] on the number of leaves. Then we will present our further investigations on p-biased recursive trees. In particular, we will be working on the number of branches, the number of nodes with at least k descendants and the depth of node n.

5.1. Definitions and Basics

Riffle shuffle permutations are based on a common method to shuffle cards: a deck of cards is first cut into two piles of approximately equal size and these piles are then riffled together, so that the cards of the two piles interleave. A mathematical model for riffle shuffles is given in [55].

Definition 5.1 ([55])**.** In order to *riffle shuffle* a deck of n cards, the deck is first cut into two piles according to a binomial distribution, so that the probability that the first pile has k cards is $\frac{\binom{n}{k}}{2^n}$. Intuitively this means that the deck is approximately cut in half. The two piles are then riffled together by dropping the cards face down one by one proportionally to the size of the remaining piles. More precisely, let A_1 and A_2 denote the sizes of the remaining piles. Then the probability that the next card comes from pile i is equal to $\frac{A_i}{A_1+A_2}$, $i = 1, 2$.

This model is the simplest version of a riffle shuffle and was later generalized into biased riffle shuffles, where the deck is cut into a piles and the sizes of the piles are chosen according to a multinomial distribution, see [56].

Definition 5.2 ([56]). A *p-biased riffle shuffle permutation* is obtained by first cutting a deck of n cards into a piles by determining the pile sizes according to $\text{mult}(a; \vec{p})$ with $\vec{p} = (p_1, p_2, \ldots, p_a)$. This means that pile sizes $b_1, \ldots, b_a$ are chosen with probability $\binom{n}{b_1, \ldots, b_a} \prod_{i=1}^{a} p_i^{b_i}$. Given the piles, they are riffled such that any of the $\binom{n}{b_1, \ldots, b_a}$ ways of interleaving is equally likely. The order of the cards in each pile is not changed in this process. We call the resulting distribution on S_n the *p-biased riffle shuffle distribution*. When $\vec{p}$ is the uniform distribution over $[a]$, i.e. $\vec{p} = (\frac{1}{a}, \ldots, \frac{1}{a})$, the resulting permutation is called an *a-shuffle*.

The first definition of a riffle shuffle we gave above corresponds to a 2-shuffle. The shuffling of the cards after determining the sizes of the piles was described in different ways in these first two definitions. It is standard in the literature that these ways are equivalent. The following are a few more equivalent descriptions of p-biased riffle shuffles.

Theorem 5.3 ([55, 56]). *The following are equivalent.*

(i) *γ is a p-biased riffle shuffle.*

(ii) *A deck of n cards is cut according to $\text{mult}(a, \vec{p})$ and then the cards are dropped face down one by one proportional to the size of the remaining piles. More precisely, let A_i denote the remaining size of each pile, then the probability that the next card comes from pile i is equal to $\frac{A_i}{A_1 + \cdots + A_a}$, $i = 1, \ldots a$.*

(iii) *A deck of n cards is cut according to $\text{mult}(a, \vec{p})$. The piles are then sequentially riffled together: first riffle pile 1 and 2 according to the rule in Definition 5.1. Then riffle this combined pile with pile 3 according to that rule. Continue until all piles are riffled together.*

(iv) *Partition the interval $[0, 1]$ into a subintervals of length $\frac{1}{a}$. Then drop n points into this interval according to the following rule: choose subinterval i with probability p_i and then drop the point uniformly in this interval. Subsequently label the points*

from 1 to n according to their order from smallest to largest and apply the function $f(x) = ax$ (mod 1) to rearrange their order.

(v) γ is the inverse of a permutation constructed in the following way: First assign a digit from 1 to a to each card according to $\vec{p}$. Then reorder the cards by first taking all cards with digit 1, then all cards with digit 2, and so on, without changing the order of cards having the same digit.

Given a p-biased riffle shuffle permutation γ the corresponding recursive tree is constructed in the usual way: take 1 as the root, attach 2 to 1, then attach every node i to the node with the label that is the rightmost number to the left of i that is smaller than i in γ. If there is no such element, attach i to 1.

Definition 5.4 ([12]). A tree constructed from a permutation over $\{2, \ldots, n\}$ having the p-biased riffle shuffle distribution is called a *p-biased recursive tree*, short BRT, and denoted by $\mathcal{T}_n^p$. When $\vec{p}$ is the uniform distribution over $[a]$ we call the corresponding tree an *a-recursive tree*, short *a*-RT, and denote it by $\mathcal{T}_n^a$.

In Figure 5.1 all 2-RTs on 4 vertices are depicted. We can see that not all permutations of n numbers can be obtained as 2-shuffles, for example the permutation (432) is not a 2-shuffle and the recursive tree on 4 vertices where all nodes are children of 1 is thus not a 2-RT. This can be seen by considering that either 2 and 3 or 3 and 4 must be in the same pile and thus not all number can be in reversed order. Moreover we can see that the probability of obtaining the permutation (234) is very high compared to the others. The reason for this is that no matter how the deck is cut, just putting the piles on top of each other in the original order is one way of interleaving them.

What makes this model very different from the weighted model and the other generalizations mentioned in the literature is that we do not have a way of picturing its dynamic growth. We mentioned in Chapter 3 that for many kinds of increasing trees we can obtain a tree of size n by attaching n according to some rule to a node of a tree of size $n - 1$. BRTs are neither defined in that way nor can such a rule be easily derived from the definition. In [12], given the pile sizes, a way of constructing a

1234	1243	1324	1342	1423
$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
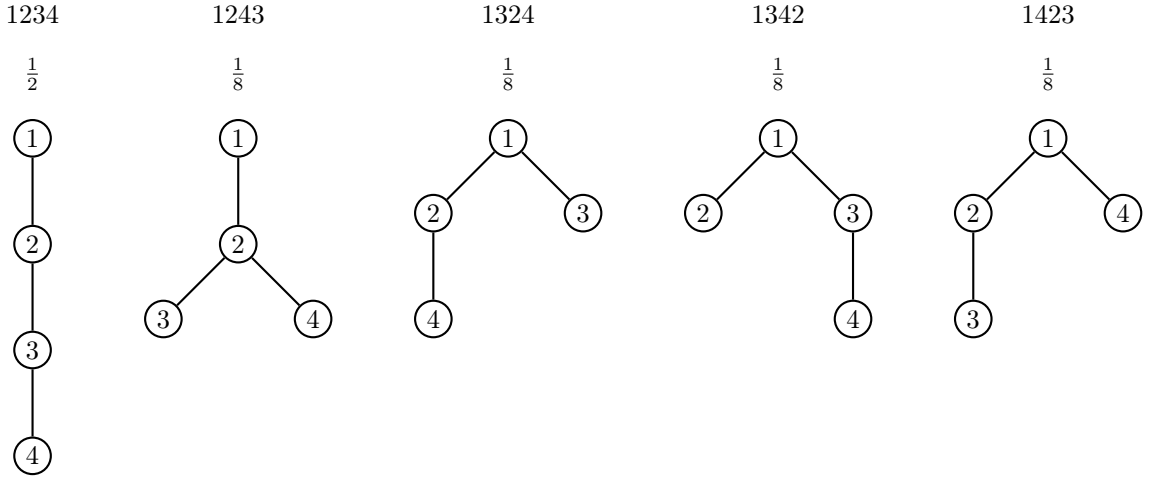
Figure 5.1. All 2-recursive trees on 4 vertices with corresponding permutation and probability.

BRT step by step is given but it cannot be considered as a dynamic growth because we cannot construct a BRT on n vertices given a BRT on $n-1$ vertices using this method. We thus need to base our results about BRTs on the permutation representation. The description of p-biased riffle shuffles via their inverse will be crucial in many arguments.

5.2. Number of Leaves

We now review some results from [12] on the distribution of the number of leaves in p-biased recursive trees.

Theorem 5.5. *Let $\mathcal{T}_n^p$ be a BRT and $\mathcal{L}_n^p$ the number of leaves of $\mathcal{T}_n^p$. Then*

$$\mathbb{E}[\mathcal{L}_n^p] = (n-2)\frac{1 - \sum_{s=1}^a p_s^2}{2} + 1 \tag{5.1}$$

and

$$\begin{aligned}
\mathrm{Var}(\mathcal{L}_n^p) =& (n-2)\left(\frac{1 - \sum_{s=1}^a p_s^2}{2} - \left(\frac{1 - \sum_{s=1}^a p_s^2}{2}\right)^2\right) \\
&+ 2(n-3)\left(\sum_{1 \le s_1 < s_2 < s_3 \le a} p_{s_1} p_{s_2} p_{s_3} - \left(\frac{1 - \sum_{s=1}^a p_s^2}{2}\right)^2\right).
\end{aligned} \tag{5.2}$$

When $\vec{p}$ is the uniform distribution over $[a]$, these expressions simplify to

$$\mathbb{E}[\mathcal{L}_n^a] = 1 + \frac{(n-2)(a-1)}{2a} \ \text{ and } \text{Var}(\mathcal{L}_n^a) = (n-2)\frac{a^2-1}{4a^2} - (n-3)\frac{a^2-1}{6a^2}. \quad (5.3)$$

In particular for fixed n asymptotically we get the same values as for URTs

$$\lim_{a\to\infty} \mathbb{E}[\mathcal{L}_n^a] = \frac{n}{2} \ \text{ and } \lim_{a\to\infty} \text{Var}(\mathcal{L}_n^a) = \frac{n}{12}. \quad (5.4)$$

Proof. As already mentioned in the review of uniform recursive trees, i is a leaf if and only if it is greater than its right neighbour in the permutation representation. So given a tree constructed from a permutation γ with $\gamma(j) = i$, i is a leaf if and only if $\gamma(j) > \gamma(j+1)$. Moreover the last element of the permutation is always a leaf. This allows us to write the number of leaves of a p-biased recursive tree as a sum of indicator random variables. So we have

$$\mathcal{L}_n^p =_d \sum_{i=2}^{n-1} \mathbf{1}(\gamma(i) > \gamma(i+1)) + 1. \quad (5.5)$$

In the description of the inverse of a riffle shuffle each card is assigned a digit from 1 to a with probability $p_1, \ldots, p_a$ and the cards are then rearranged according to the digits. This implies that when $i < j$, i will come before j in the inverse permutation if and only if card i gets a digit that is smaller or equal to the digit of j. Since $\gamma(i) > \gamma(i+1)$ if and only if $i+1$ comes before i in the inverse permutation, $\gamma(i) > \gamma(i+1)$ is equivalent to i getting a bigger digit than $i+1$, i.e. to $X_i > X_{i+1}$.

Let X_i, for $2 \le i \le n$ be independent random variables with distribution $\vec{p}$, over $\{1, \ldots, a\}$, then $\mathbf{1}(\gamma(i) > \gamma(i+1)) = \mathbf{1}(X_i > X_{i+1})$ and we can thus write

$$\mathcal{L}_n^p =_d \sum_{i=2}^{n-1} \mathbf{1}(X_i > X_{i+1}) + 1. \quad (5.6)$$

Since all X_i are identically distributed we have $\mathbb{P}(X_i > X_{i+1}) = \mathbb{P}(X_{i+1} > X_i)$. Moreover $\mathbb{P}(X_i > X_{i+1}) + \mathbb{P}(X_i = X_{i+1}) + \mathbb{P}(X_i < X_{i+1}) = 1$ and $\mathbb{P}(X_i = X_{i+1}) = \sum_{s=1}^{a} p_s^2$, thus

$$\mathbb{P}(X_i > X_{i+1}) = \frac{1 - \sum_{s=1}^{a} p_s^2}{2}. \tag{5.7}$$

So we get

$$\mathbb{E}[\mathcal{L}_n^p] = \sum_{i=2}^{n-1} \mathbb{E}[\mathbf{1}(X_i > X_{i+1})] + 1 = (n-2)\frac{1 - \sum_{s=1}^{a} p_s^2}{2} + 1. \tag{5.8}$$

Let now $\ell_i = \mathbf{1}(X_i > X_{i+1})$, then $\mathcal{L}_n^p = \sum_{i=2}^{n-1} \ell_i + 1$ and by writing $\mathrm{Var}(\mathcal{L}_n^p) = \sum_{i=2}^{n-1} \mathrm{Var}(\ell_i) + 2\sum_{2 \leq i < j \leq n-1} \mathrm{Cov}(\ell_i, \ell_j)$ Altok and Işlak get

$$\begin{aligned}
\mathrm{Var}(\mathcal{L}_n^p) =& (n-2)\left(\frac{1 - \sum_{s=1}^{a} p_s^2}{2} - \left(\frac{1 - \sum_{s=1}^{a} p_s^2}{2}\right)^2\right) \\
&+ 2(n-3)\left(\sum_{1 \leq s_1 < s_2 < s_3 \leq a} p_{s_1} p_{s_2} p_{s_3} - \left(\frac{1 - \sum_{s=1}^{a} p_s^2}{2}\right)^2\right).
\end{aligned} \tag{5.9}$$

The results for the uniform case can be derived by substituting $\frac{1}{a}$ for all p_s in these expressions. □

As in the uniform case, the distribution of the number of leaves tends to a normal distribution when $\vec{p}$ is non-degenerate, i.e. when there is no $i = 1, \ldots, a$ such that $p_i = 1$.

Theorem 5.6 ([12]). *Let $\mathcal{L}_n^p$ denote the number of leaves in a p-biased recursive tree and assume that $\vec{p}$ is non degenerate. Then for any $n \in \mathbb{N}$ there is a $C \in \mathbb{R}$ such that*

$$d_K \left(\frac{\mathcal{L}_n^p - \mathbb{E}[\mathcal{L}_n^p]}{\sqrt{\mathrm{Var}(\mathcal{L}_n^p)}}, \mathcal{G} \right) \leq \frac{C}{\sqrt{n}}. \tag{5.10}$$

Moreover a result about the distance between the number of leaves of URTs and of p-biased recursive trees is given.

Theorem 5.7 ([12]). *Let us denote the number of leaves in a URT, p-BRT and a-RT of size n as $\mathcal{L}_n$, $\mathcal{L}_n^p$ and $\mathcal{L}_n^a$ respectively. Then we have the following:*

(i) *For $n \geq 3$,*

$$d_{TV}(\mathcal{L}_n, \mathcal{L}_n^p) \leq \binom{n-1}{2} \sum_{s=1}^{a} p_s^2. \tag{5.11}$$

(ii) *For $a \geq n \geq 3$ and $\vec{p}$ the uniform distribution this bound can be improved and we get*

$$d_{TV}(\mathcal{L}_n, \mathcal{L}_n^a) \leq 1 - \frac{a!}{(a-n)!\, a^n}. \tag{5.12}$$

(iii) *These two bounds imply that $\mathcal{L}_n^a$ converges in distribution to $\mathcal{L}_n$ as $a \to \infty$ and that $\mathcal{L}_n^p$ converges in distribution to $\mathcal{L}_n$ as $a \to \infty$ if $\vec{p}$ is non-degenerate.*

(iv) *For a given a, among all distributions on $[a]$, the uniform distribution maximizes the expected number of leaves of $\mathcal{T}_n^p$. Thus if $\vec{p}$ is any distribution on $[a]$, then $\mathbb{E}[\mathcal{L}_n^p] \leq \mathbb{E}[\mathcal{L}_n^a]$.*

The proofs of Theorems 5.6 and 5.7 can be found in [12].

We now move on to giving new results about BRTs, namely, on the number of branches, the number of nodes with at least k descendants and the depth of node n.

5.3. Number of Branches

We will present two different methods to calculate the number of branches in a BRT. While we use the first one to get explicit expressions we still mention the second approach because it might be helpful for other problems.

5.3.1. Anti-records

As we have seen in Chapter 3, in URTs there are several possibilities to obtain results on the number of branches. For WRTs we used the attachment probabilities, which allowed us to write the number of branches as a sum of independent random variables. For BRTs we will use the observation that the number of branches of a recursive tree is equal to the number of anti-records in its permutation representation. First of all let us observe that in a riffle shuffle permutation, for $\gamma(i)$ to be an anti-record, it must be the first card of one of the a piles. There can thus be at most a anti-records. Moreover, as we consider riffle shuffles of $\{2, \ldots, n\}$, there definitely is an anti-record at $\gamma(2)$. Using this approach we will prove the following theorem.

Theorem 5.8. *Let $\mathcal{B}_n^p$ denote the number of branches in a p-biased recursive tree $\mathcal{T}_n^p$. Then*

$$\mathbb{E}[\mathcal{B}_n^p] = \sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-1} \right) + p_a. \tag{5.13}$$

Moreover we have the following asymptotic result

$$\mathbb{E}[\mathcal{B}_n^p] \xrightarrow{n \to \infty} \sum_{s=1}^{a} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell}. \tag{5.14}$$

When $\vec{p}$ is the uniform distribution over $[a]$, letting $\mathcal{B}_n^a$ denote the number of branches in an a-RT $\mathcal{T}_n^a$,

$$\mathbb{E}[\mathcal{B}_n^a] = \sum_{s=1}^{a-1} \frac{1}{s} \left(1 - \left(1 - \frac{s}{a} \right)^{n-1} \right) + \frac{1}{a} \tag{5.15}$$

and asymptotically

$$\mathbb{E}[\mathcal{B}_n^a] \xrightarrow{n \to \infty} H_a. \tag{5.16}$$

Moreover, when a tends to infinity, the expectation tends to the expectation for URTs, namely

$$\mathbb{E}[\mathcal{B}_n^a] \xrightarrow{a \to \infty} H_{n-1}. \tag{5.17}$$

Proof. We will use the inverse formulation for biased riffle shuffles in order to study the number of anti-records. When $X_i < \min\{X_2, \ldots, X_{i-1}\}$, we get at i the first card from a pile with cards that are smaller than all the previous ones, hence there is an anti-record at i. Thus for $3 \le i < n$, and γ a p-biased riffle shuffle, $\gamma(i)$ is an anti-record if and only if X_i is strictly less than $X_2, \ldots X_{i-1}$ where X_j, $j = 2, \ldots, n$, are independent random variables such that $X_j = s$, $s \in [a]$, with probability p_s. By independence of the X_i's we get for $3 \le i < n$,

$$
\begin{aligned}
&\mathbb{P}(X_i < \min\{X_2, \ldots, X_{i-1}\}) \\
&= \sum_{s=1}^{a} \mathbb{P}(X_i < \min\{X_2, \ldots, X_{i-1}\} | X_i = s) \mathbb{P}(X_i = s) \\
&= \sum_{s=1}^{a} \mathbb{P}(s < \min\{X_2, \ldots, X_{i-1}\}) \mathbb{P}(X_i = s) \\
&= \sum_{s=1}^{a-1} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} p_s.
\end{aligned}
\tag{5.18}
$$

We only sum from 1 to $a - 1$ because for $i \ge 3$, $\min\{X_2, \ldots, X_{i-1}\} \le a$ and thus $\mathbb{P}(a < \min\{X_2, \ldots, X_{i-1}\}) = 0$.

Let $A_i = \mathbf{1}(\gamma(i)$ is an anti-record), then $\mathcal{B}_n^p = \sum_{i=2}^{n} A_i$ and $\gamma(2)$ definitely is an anti-record. We need to assume $p_1 \ne 0$ from now on in order to avoid division by 0. This is not an important restriction though, since given $\vec{p}$, we can define $\vec{q}$ as $\vec{p}$ without the p_i's that are zero. Then the distributions obtained from $\vec{p}$ and $\vec{q}$ are the same.

Now,

$$
\begin{aligned}
\mathbb{E}[\mathcal{B}_n^p] &= 1 + \sum_{i=3}^{n} \mathbb{E}[A_i] \\
&= 1 + \sum_{i=3}^{n} \mathbb{P}(X_i < \min\{X_2, \ldots, X_{i-1}\}) \\
&= 1 + \sum_{i=3}^{n} \sum_{s=1}^{a-1} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} p_s \\
&= 1 + \sum_{s=1}^{a-1} p_s \sum_{i=1}^{n-2} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \\
&= 1 + \sum_{s=1}^{a-1} p_s \left(\frac{1 - (\sum_{\ell=s+1}^{a} p_\ell)^{n-1}}{1 - \sum_{\ell=s+1}^{a} p_\ell} - 1 \right) \\
&= \sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-1} \right) + p_a.
\end{aligned}
\tag{5.19}
$$

Since $p_1 > 0$ implies $\left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-1} \xrightarrow{n \to \infty} 0$ for all s,

$$
\mathbb{E}[\mathcal{B}_n^p] \xrightarrow{n \to \infty} \sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} + p_a = \sum_{s=1}^{a} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell}.
\tag{5.20}
$$

In particular, for $\vec{p}$ the uniform distribution over $[a]$ we have $p_i = \frac{1}{a}$, so we get

$$
\begin{aligned}
\mathbb{E}[\mathcal{B}_n^a] &= \sum_{s=1}^{a-1} \frac{\frac{1}{a}}{\sum_{\ell=1}^{s} \frac{1}{a}} \left(1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a} \right)^{n-1} \right) + \frac{1}{a} \\
&= \sum_{s=1}^{a-1} \frac{1}{s} \left(1 - \left(\frac{a-s}{a} \right)^{n-1} \right) + \frac{1}{a} \\
&= \sum_{s=1}^{a-1} \frac{1}{s} \left(1 - \left(1 - \frac{s}{a} \right)^{n-1} \right) + \frac{1}{a}
\end{aligned}
\tag{5.21}
$$

and

$$
\mathbb{E}[\mathcal{B}_n^a] \xrightarrow{n \to \infty} \sum_{s=1}^{a} \frac{1}{s} = H_a.
\tag{5.22}
$$

Moreover, for a fixed n, if $a \to \infty$, for fixed n the expectation tends to the expectation of the number of branches for URTs. To show this we will manipulate the second line of 5.21 and then use the following identity:

$$\sum_{s=1}^{a} s^k = \frac{a^{k+1}}{k+1} + \mathcal{O}(a^k) \text{ for all } k \in \mathbb{N}, k \geq 0. \tag{5.23}$$

We have

$$
\begin{aligned}
\mathbb{E}[\mathcal{B}_n^a] &= \sum_{s=1}^{a-1} \frac{1}{s} \left(1 - \left(\frac{a-s}{a} \right)^{n-1} \right) + \frac{1}{a} \\
&= \sum_{s=1}^{a-1} \frac{1}{a} \frac{1 - \left(\frac{a-s}{a} \right)^{n-1}}{1 - \frac{a-s}{a}} + \frac{1}{a} \\
&= \sum_{s=1}^{a-1} \frac{1}{a} \sum_{\ell=0}^{n-2} \left(\frac{a-s}{a} \right)^{\ell} + \frac{1}{a} \\
&= \sum_{\ell=0}^{n-2} \frac{1}{a^{\ell+1}} \sum_{s=1}^{a-1} s^{\ell} + \frac{1}{a} \\
&= \sum_{\ell=0}^{n-2} \frac{1}{a^{\ell+1}} \left[\frac{a^{\ell+1}}{\ell+1} + \mathcal{O}(a^{\ell}) \right] + \frac{1}{a} \\
&= \sum_{\ell=0}^{n-2} \frac{1}{\ell+1} + \mathcal{O}\left(\frac{1}{a} \right) \\
&\xrightarrow{a \to \infty} H_{n-1}.
\end{aligned}
\tag{5.24}
$$

$\square$

Theorem 5.9. *Let $\mathcal{B}_n^p$ denote the number of branches of a p-BRT. Then for $\vec{p}$ such that $p_1 > 0$,*

$$\text{Var}(\mathcal{B}_n^p) =$$

$$\sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-1} \right) - \sum_{s=1}^{a-1} p_s$$

$$- \sum_{s=1}^{a-1} p_s^2 \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{2(n-2)}}{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2}$$

$$- 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right) \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{n-2}}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}$$

$$+ 2 \sum_{s=2}^{a-1} \frac{p_s \sum_{\ell=s+1}^{a} p_\ell}{\sum_{\ell=1}^{s} p_\ell} \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right) \qquad (5.25)$$

$$- 2 \sum_{s=2}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{s-1} \frac{p_r \sum_{q=r+1}^{a} p_q}{\sum_{q=1}^{r} p_q} \frac{1}{\sum_{q=r+1}^{s} p_q}$$

$$\cdot \left(\left(\sum_{q=r+1}^{a} p_q \right)^{n-3} - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right)$$

$$- 2 \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^2 \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{n-3}}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}$$

$$+ 2 \sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{n-1} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right)$$

and

$$\mathrm{Var}(\mathcal{B}_n^p) \xrightarrow{n\to\infty}$$

$$\sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} - \sum_{s=1}^{a-1} p_s$$

$$- \sum_{s=1}^{a-1} p_s^2 \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \frac{1}{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2}$$

$$- 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right) \frac{1}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}$$

$$+ 2 \sum_{s=2}^{a-1} \frac{p_s \sum_{\ell=s+1}^{a} p_\ell}{\sum_{\ell=1}^{s} p_\ell} \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q}$$

$$- 2 \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^2 \frac{1}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}. \tag{5.26}$$

Proof. See Appendix A. $\qquad\qquad\square$

Theorem 5.10. *Let $\mathcal{B}_n^a$ denote the number of branches in an a-RT. Then*

$$\mathrm{Var}(\mathcal{B}_n^a) = \sum_{s=1}^{a-1} \frac{1}{s} \left(1 - \left(\frac{a-s}{a} \right)^{n-1} \right) - \frac{a-1}{a}$$

$$- \sum_{s=1}^{a-1} \frac{1}{a^2} \left(\frac{a-s}{a} \right)^2 \frac{1 - \left(\frac{a-s}{a} \right)^{2(n-2)}}{1 - \left(\frac{a-s}{a} \right)^2}$$

$$- 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} \frac{1}{a^2} \frac{a-s}{a} \frac{a-r}{a} \frac{1 - \left(\frac{a-s}{a} \frac{a-r}{a} \right)^{n-2}}{1 - \frac{a-s}{a} \frac{a-r}{a}}$$

$$+ 2 \sum_{s=2}^{a-1} \frac{a-s}{as} \sum_{r=1}^{s-1} \frac{1}{r} \left(1 - \left(\frac{a-s}{a} \right)^{n-3} \right) \tag{5.27}$$

$$- 2 \sum_{s=2}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{s-1} \frac{a-r}{r} \frac{1}{s-r} \left(\left(\frac{a-r}{a} \right)^{n-3} - \left(\frac{a-s}{a} \right)^{n-3} \right)$$

$$- 2 \sum_{s=1}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{a-1} \frac{1}{r} \left(\frac{a-r}{a} \right)^2 \frac{1 - \left(\frac{a-s}{a} \frac{a-r}{a} \right)^{n-3}}{1 - \frac{a-s}{a} \frac{a-r}{a}}$$

$$+ 2 \sum_{s=1}^{a-1} \frac{a-s}{sa} \sum_{r=1}^{a-1} \frac{1}{r} \left(\frac{a-r}{a} \right)^{n-1} \left(1 - \left(\frac{a-s}{a} \right)^{n-3} \right)$$

and

$$\text{Var}(\mathcal{B}_n^a) \xrightarrow{n\to\infty} H_a - \frac{a-1}{a} + \frac{2}{a}\sum_{s=2}^{a-1}\frac{a-s}{s}\sum_{r=1}^{s-1}\frac{1}{r}$$
$$-\frac{1}{a^2}\sum_{s=1}^{a-1}\frac{s^2}{a^2-s^2}$$
$$-\frac{2}{a^2}\sum_{s=1}^{a-2}\sum_{r=1}^{a-s+1}\frac{sr}{a^2-sr} \tag{5.28}$$
$$-\frac{2}{a^2}\sum_{s=1}^{a-1}\sum_{r=1}^{a-1}\frac{1}{r}\frac{s(a-r)^2}{a^2-s(a-r)}.$$

Moreover, for fixed n, when a increases the variance approaches the variance from the uniform case:

$$\text{Var}(\mathcal{B}_n^a) \xrightarrow{a\to\infty} H_{n-1} - H_{n-1}^{(2)}. \tag{5.29}$$

Proof. See Appendix A. $\square$

Remark 5.11. There are various results in the literature on central limit theorems for the number of records in random words, for example [57,58]. We hope to adapt these for obtaining asymptotic results on the number of branches in a subsequent work.

5.3.2. Sequential Shuffling

We now present another possibility to calculate the expectation and variance of the number of branches of a BRT. As we described in Section 5.1, in a riffle shuffle, after choosing the pile sizes, we might first shuffle pile 1 and 2 such that if there are A_1 cards remaining in pile 1 and A_2 cards in pile 2, the probability that the next card is from pile i is $\frac{A_i}{A_1+A_2}$ for $i = 1, 2$. More generally, for $i = 2, \ldots, n$, after shuffling the first $i-1$ piles, we shuffle the obtained shuffled pile with the i-th pile such that if A_i are the cards remaining in the i-th pile and B_{i-1} the cards remaining in the already shuffled pile, the probability that the next card comes from the i-th pile is $\frac{A_i}{B_{i-1}+A_i}$.

Only the first card of each pile can be the start of a branch, since all other cards will definitely have a smaller card of the same pile to their left. Hence we only need to consider where these are. Let us call the first cards of each pile $F_1, F_2, \ldots, F_a$ and the piles $S_1, \ldots, S_a$. If pile i is empty F_i refers to no card. We know that all cards of $S_1, \ldots, S_i$ are smaller than F_{i+1}. Hence, if S_{i+1} is non-empty, F_{i+1} attaches to 1 if and only if it comes before any card from piles with lower index. Since each time we riffle a new pile this only depends on the first step, the remaining pile sizes actually are the initial pile sizes in our case.

Set $I_i = \mathbf{1}(F_i$ attaches to 1), then we get

$$\mathbb{P}(I_i = 1) = \begin{cases} 0, & \text{if } b_1, \ldots, b_i = 0 \\ \frac{b_i}{b_1 + \cdots + b_i}, & \text{otherwise.} \end{cases} \tag{5.30}$$

Or alternatively

$$\mathbb{E}[I_i] = \mathbb{P}(I_i = 1) = \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}}. \tag{5.31}$$

Now $\mathcal{B}_n^p = \sum_{i=1}^a I_i$, and we can write, conditioned on the pile sizes,

$$\mathbb{E}[\mathcal{B}_n^p | b_1, \ldots, b_a] = \sum_{i=1}^a \mathbb{E}[I_i | b_1, \ldots, b_a] = \sum_{i=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}}. \tag{5.32}$$

Since the pile sizes are chosen according to the multinomial distribution this gives

$$\mathbb{E}[\mathcal{B}_n^p] = \sum_{\substack{b_1, \ldots, b_a \\ b_1 + \cdots + b_a = n}} \binom{n}{b_1, \ldots, b_a} \prod_{i=1}^a p_i^{b_i} \sum_{j=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}} \tag{5.33}$$

where we used $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$ with $X = \mathcal{B}_n^p$ and $Y = (b_1, \ldots, b_a)$.

We can give this a nicer form by considering for all $j = 1, \ldots, a$ the expression $\sum_{\substack{b_1,\ldots,b_a \\ b_1+\cdots+b_a=n}} \binom{n}{b_1,\ldots,b_a} \prod_{i=1}^{a} p_i^{b_i} \frac{b_j}{\max\{1,b_1+\cdots+b_j\}}$ separately.

For $j = 1$ this gives

$$\sum_{\substack{b_1,\ldots,b_a \\ b_1+\cdots+b_a=n}} \binom{n}{b_1,\ldots,b_a} \prod_{i=1}^{a} p_i^{b_i} \frac{b_1}{\max\{1,b_1\}} = 1 - \mathbb{P}(b_1 = 0) = 1 - \left(\sum_{i=2}^{a} p_i\right)^n. \quad (5.34)$$

For $j = 2$, we get

$$\sum_{\substack{b_1,\ldots,b_a \\ b_1+\cdots+b_a=n}} \binom{n}{b_1,\ldots,b_a} \prod_{i=1}^{a} p_i^{b_i} \frac{b_2}{\max\{1,b_1+b_2\}}$$

$$= \sum_{k=0}^{n} \sum_{\substack{b_3,\ldots,b_a \\ b_3+\cdots+b_a=k}} \sum_{b_2=0}^{n-k} \binom{n}{b_1,\ldots,b_a} \prod_{i=1}^{a} p_i^{b_i} \frac{b_2}{\max\{1,n-k\}}$$

$$= \sum_{k=0}^{n-1} \sum_{\substack{b_3,\ldots,b_a \\ b_3+\cdots+b_a=k}} \frac{n!}{b_3!\ldots b_a!} \prod_{i=3}^{a} p_i^{b_i} \sum_{b_2=1}^{n-k} \frac{b_2 p_1^{n-k-b_2} p_2^{b_2}}{\max\{1,n-k\}(n-k-b_2)!\, b_2!}$$

$$= \sum_{k=0}^{n-1} \sum_{\substack{b_3,\ldots,b_a \\ b_3+\cdots+b_a=k}} \frac{n\cdots(n-k+1)}{b_3!\ldots b_a!} \prod_{i=3}^{a} p_i^{b_i} \sum_{b_2=1}^{n-k} \frac{(n-k)!\, b_2}{(n-k-b_2)!\, b_2!} \frac{p_1^{n-k-b_2} p_2^{b_2}}{\max\{1,n-k\}}.$$

$$(5.35)$$

Now we have for $1 \le k < n$,

$$\sum_{b_2=1}^{n-k} \frac{(n-k)!\, b_2}{(n-k-b_2)!\, b_2!} \frac{p_1^{n-k-b_2} p_2^{b_2}}{\max\{1,n-k\}}$$

$$= \sum_{b_2=1}^{n-k} \frac{(n-k)!}{(n-k-b_2)!\, b_2!} \frac{b_2}{n-k} p_1^{n-k-b_2} p_2^{b_2}$$

$$= \sum_{b_2=1}^{n-k} \frac{(n-k-1)\cdots(n-k-b_2+1)}{(b_2-1)!} p_1^{n-k-b_2} p_2^{b_2} \quad (5.36)$$

$$= p_2 \sum_{b_2=0}^{n-k-1} \frac{(n-k-1)\cdots(n-k-b_2)}{b_2!} p_1^{n-k-1-b_2} p_2^{b_2}$$

$$= p_2(p_1+p_2)^{n-k-1}.$$

By inserting this into 5.35 we thus get

$$p_2 \sum_{k=0}^{n-1} \sum_{\substack{b_3,\ldots,b_a \\ b_3+\cdots+b_a=k}} \frac{n\cdots(n-k+1)}{b_3!\cdots b_a!} \prod_{i=3}^{a} p_i^{b_i}(p_1+p_2)^{n-k-1}$$

$$= \frac{p_2}{p_1+p_2} \sum_{k=0}^{n-1} \sum_{\substack{b_3,\ldots,b_a \\ b_3+\cdots+b_a=k}} \frac{n\cdots(n-k+1)}{b_3!\cdots b_a!} \prod_{i=3}^{a} p_i^{b_i}(p_1+p_2)^{n-k}$$

$$= \frac{p_2}{p_1+p_2} \left[\sum_{\substack{B,b_3,\ldots,b_a \\ B+b_3+\cdots+b_a=n}} \frac{n!}{B!\,b_3!\cdots b_a!} \prod_{i=3}^{a} p_i^{b_i}(p_1+p_2)^{B} - \mathbb{P}(b_1=b_2=0) \right] \qquad (5.37)$$

$$= \frac{p_2}{p_1+p_2} \left[(p_1+p_2+\cdots+p_a)^n - (1-(p_1+p_2))^n \right]$$

$$= \frac{p_2}{p_1+p_2} \left(1 - \left(\sum_{i=3}^{n} p_i \right)^n \right).$$

The cases $j = 3, \ldots, a$ can be calculated similarly, thus we get in total

$$\mathbb{E}[\mathcal{B}_n^p] = \sum_{i=1}^{a} \frac{p_i}{p_1+\cdots+p_i} \left(1 - \left(\sum_{j=i+1}^{a} p_j \right)^n \right) \qquad (5.38)$$

which is the same expression as in Theorem 5.13.

For the variance we will use that $\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y])$. We have

$$\mathrm{Var}(\mathcal{B}_n^p|b_1,\ldots,b_a)$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{a} I_i \right)^2 \Big| b_1,\ldots,b_a \right] - \mathbb{E}\left[\sum_{i=1}^{a} I_i \Big| b_1,\ldots,b_a \right]^2$$

$$= \sum_{i=1}^{a} \mathbb{E}[I_i|b_1,\ldots,b_a] + 2\sum_{i=1}^{a}\sum_{j=i}^{a} \mathbb{E}[I_i I_j|b_1,\ldots,b_a] \qquad (5.39)$$

$$\quad - \sum_{i=1}^{a} \mathbb{E}[I_i|b_1,\ldots,b_a]^2 - 2\sum_{i=1}^{a}\sum_{j=i}^{a} \mathbb{E}[I_i|b_1,\ldots,b_a]\mathbb{E}[I_j|b_1,\ldots,b_a]$$

$$= \sum_{i=1}^{a} \mathbb{E}[I_i|b_1,\ldots,b_a] - \mathbb{E}[I_i|b_1,\ldots,b_a]^2.$$

Here we were able to get rid of the mixed terms because I_i and I_j are independent for $i < j$, given the pile sizes $b_1, \ldots, b_a$: F_j attaches to 1 if and only if it is the first card in the shuffle of $S_1, \ldots, S_{j-1}, S_j$, where $S_1, \ldots, S_{j-1}$ are already mixed together. This only depends on the sum of the first j pile sizes, i.e. $S_1 + \cdots + S_{j-1}$, and S_j, not on the order of the previously shuffled cards, particularly not on the position of F_i. Hence

$$\operatorname{Var}(\mathcal{B}_n^p | b_1, \ldots, b_a) = \sum_{i=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}} - \frac{b_i^2}{(\max\{1, b_1 + \cdots + b_i\})^2}. \quad (5.40)$$

Also

$$\operatorname{Var}(\mathbb{E}[\mathcal{B}_n^p | b_1, \ldots, b_a]) = \mathbb{E}[\mathbb{E}[\mathcal{B}_n^p | b_1, \ldots, b_a]^2] - \mathbb{E}[\mathbb{E}[\mathcal{B}_n^p | b_1, \ldots, b_a]]^2 \quad (5.41)$$

and

$$\mathbb{E}[\mathbb{E}[\mathcal{B}_n^p | b_1, \ldots, b_a]^2] = \sum_{b_1, \ldots, b_a} \binom{n}{b_1, \ldots, b_i} \prod_{i=1}^a p_i^{b_i} \left(\sum_{i=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}} \right)^2 \quad (5.42)$$

and

$$\mathbb{E}[\mathbb{E}[\mathcal{B}_n^p | b_1, \ldots, b_a]]^2 = \left(\sum_{b_1, \ldots, b_i} \binom{n}{b_1, \ldots, b_a} \prod_{i=1}^a p_i^{b_i} \sum_{i=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}} \right)^2. \quad (5.43)$$

So we get in total

$$\begin{aligned}
\operatorname{Var}(\mathcal{B}_n^p) = \sum_{b_1, \ldots, b_a} &\binom{n}{b_1, \ldots, b_a} \prod_{i=1}^a p_i^{b_i} \\
&\cdot \left[\left(\sum_{i=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}} - \frac{b_i^2}{(\max\{1, b_1 + \cdots + b_i\})^2} \right) \right. \\
&+ \left(\sum_{i=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}} \right)^2 \\
&\left. - \binom{n}{b_1, \ldots, b_a} \prod_{i=1}^a p_i^{b_i} \left(\sum_{i=1}^a \frac{b_i}{\max\{1, b_1 + \cdots + b_i\}} \right)^2 \right].
\end{aligned} \quad (5.44)$$

Remark 5.12. We argued at the beginning that the number of branches is limited by the number of piles we split the deck into, i.e. a. In fact that argument also works for the degree of any node: if two nodes are children of i, the smaller one must come later in the permutation, thus the nodes attached to i form a decreasing sequence. But in a p-biased riffle shuffle permutation any decreasing sequence must consist of cards from different piles since cards from the same pile remain in the same order. Since there are a piles, this implies that the degree of any node is limited by a. Since a-ary trees also have this property, it would be interesting to see whether a-ary recursive trees also share other properties with trees constructed from a-shuffles.

5.4. Number of Nodes with at least k Descendants

In a BRT the number of nodes with at least k descendants can also be calculated using the construction of an inverse riffle shuffle. In a recursive tree $\mathcal{T}$, node i has at least k descendants if in the permutation representation of the tree at least the k entries following i are bigger than i. This is the case if, given $\gamma(j) = i$, we have $\gamma(j+1), \ldots, \gamma(j+k) > i$. This is equivalent to $X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}$ in the inverse riffle shuffle construction. We thus get a node with a least k descendants for every X_j satisfying the above, more precisely, if $X_j \leq \{X_{j+1} \ldots X_{j+k}\}$, node $i = \gamma(j)$ is a node with at least k descendants.

Example 5.13. Let our deck consist of 8 cards. Assume that we cut it into 3 piles. We first construct the inverse riffle shuffle by assigning digits from 1 to 3 to every card as can be seen in Figure 5.2.

X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	2	3	1	3	1	1

Figure 5.2. Example of the construction of an inverse riffle shuffle permutation.

This gives the inverse permutation $\gamma^{-1} = 2578346$ and thus $\gamma = 2673845$ with the corresponding recursive tree in Figure 5.3. As can be seen the tree has two nodes with at least two descendants, 2 and 3. This corresponds to what we can derive from
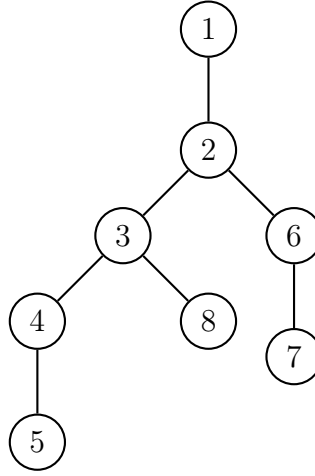
Figure 5.3. The biased recursive tree corresponding to $\gamma = 2673845$.

the permutation: $\pi(5) = 3$ and we have $X_5 \leq X_6, X_7$. Also $\pi(2) = 2$ and we have $X_2 \leq X_3, X_4$.

We will now use this observation to calculate the expectation and variance of the number of nodes with at least k descendants, and then will prove a central limit theorem by using the fact that the dependence is local.

Theorem 5.14. *Let $k \in \mathbb{N}$ and let $Y^p_{\geq k,n}$ denote the number of nodes with at least k descendants of $\mathcal{T}^p_n$, a p-biased recursive tree of size n. Then*

$$\mathbb{E}\left[Y^p_{\geq k,n}\right] = (n - k - 1) \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k + 1. \tag{5.45}$$

Moreover if $\vec{p}$ is the uniform distribution over $[a]$,

$$\mathbb{E}\left[Y^a_{\geq k,n}\right] = (n - k - 1) \frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k + 1 \tag{5.46}$$

and in particular, as $a \to \infty$, we get

$$\mathbb{E}\left[Y^a_{\geq k,n}\right] \xrightarrow{a \to \infty} \frac{n}{k + 1}. \tag{5.47}$$

Proof. Let $C_i^k = \mathbf{1}(X_i \leq \{X_{i+1}, \ldots, X_{i+k}\})$. Then

$$Y_{\geq k,n}^p =_d \sum_{i=1}^{n-k} C_i^k = \sum_{i=2}^{n-k} C_i^k + 1. \qquad (5.48)$$

Also for $2 \leq i \leq n-k$,

$$
\begin{aligned}
\mathbb{E}\left[C_i^k\right] &= \mathbb{P}(X_i \leq \{X_{i+1}, \ldots, X_{i+k}\}) \\
&= \sum_{s=1}^{a} \mathbb{P}(X_i \leq \{X_{i+1}, \ldots, X_{i+k}\}|X_i = s)\mathbb{P}(X_i = s) \\
&= \sum_{s=1}^{a} \mathbb{P}(s \leq \{X_{i+1}, \ldots, X_{i+k}\})\mathbb{P}(X_i = s) \\
&= \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k.
\end{aligned} \qquad (5.49)
$$

Thus

$$
\begin{aligned}
\mathbb{E}\left[Y_{\geq k,n}^p\right] &= \sum_{i=1}^{n-k} \mathbb{E}\left[C_i^k\right] \\
&= \sum_{i=2}^{n-k} \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k + 1 \\
&= (n-k-1) \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k + 1.
\end{aligned} \qquad (5.50)
$$

In particular, if $p_s = \frac{1}{a}$ for all s, we get

$$
\begin{aligned}
\mathbb{E}\left[Y_{\geq k,n}^a\right] &= (n-k-1) \sum_{s=1}^{a} \frac{1}{a} \left(\sum_{r=s}^{a} \frac{1}{a}\right)^k + 1 \\
&= (n-k-1)\frac{1}{a^{k+1}} \sum_{s=1}^{a} \left(\sum_{r=s}^{a} 1\right)^k + 1 \\
&= (n-k-1)\frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k + 1.
\end{aligned} \qquad (5.51)
$$

This gives asymptotically

$$
\begin{aligned}
\lim_{a \to \infty} \mathbb{E}\left[Y^a_{\geq k,n}\right] &= \lim_{a \to \infty} (n-k-1)\frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k + 1 \\
&= \lim_{a \to \infty} (n-k-1)\frac{1}{a^{k+1}}\left[\frac{a^{k+1}}{k+1} + \mathcal{O}(a^k)\right] + 1 \quad (5.52)\\
&= \frac{n}{k+1}.
\end{aligned}
$$

$\square$

Remark 5.15. We know from Theorem 3.6 that in a URT, for $A_{k,n}$, the number of nodes with exactly k descendants, $\frac{A_{k,n}}{n} \xrightarrow{n \to \infty} \frac{1}{(k+1)(k+2)}$ in probability holds. This is consistent with the expectation of the number of node with at least k descendants we found for $a \to \infty$.

Remark 5.16. For the case $k=1$ and $\vec{p}$ the uniform distribution over a this gives the expected number of internal nodes

$$
\mathbb{E}\left[Y^a_{\geq 1,n}\right] = (n-2)\frac{1}{a^2}\sum_{s=1}^{a} s + 1 = (n-2)\frac{1}{a^2}\frac{a(a+1)}{2} + 1 = \frac{n-2}{2}\frac{a+1}{a} + 1. \quad (5.53)
$$

This also follows from the expected number of leaves in an a-RT.

Theorem 5.17. *Let $k \in \mathbb{N}$ and let $Y^p_{\geq k,n}$ denote the number of nodes with at least k descendants in a p-BRT. Then*

$$
\begin{aligned}
\mathrm{Var}\left(Y^p_{\geq k,n}\right) &= \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k \left[(n-k-1) + p_1\left(2nk - 3k(k+1)\right)\right] \\
&\quad + 2\sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^k \\
&\quad \cdot \left[n-k-1-(n-2k-1)\left(\sum_{u=s}^{a} p_u\right)^k - \frac{1-\left(\sum_{u=s}^{a} p_u\right)^k}{1-\sum_{u=s}^{a} p_u}\right] \\
&\quad - \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k\right)^2 \left[n(2k+1) - (3k+1)(k+1)\right]. \quad (5.54)
\end{aligned}
$$

We moreover have

$$\lim_{n\to\infty} \frac{\mathrm{Var}(Y^p_{\geq k,n})}{n} = \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k \left(2kp_1 + 1 - (2k+1)\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k\right)$$
$$+ 2\sum_{s=2}^{a} \frac{p_s}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^k \left[1 - \left(\sum_{u=s}^{a} p_u\right)^k\right]. \tag{5.55}$$

Proof. See Appendix A. □

Corollary 5.18. *If we choose the uniform distribution over* $[a]$ *we get*

$$\mathrm{Var}\left(Y^a_{\geq k,n}\right) = \frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k \left[(n-k-1) + \frac{1}{a}(2nk - 3k(k+1))\right]$$
$$+ 2\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{r=1}^{a-s+1} r^k$$
$$\cdot \left[n - k - 1 - (n - 2k - 1)\left(\frac{a-s+1}{a}\right)^k - \frac{1 - \left(\frac{a-s+1}{a}\right)^k}{1 - \frac{a-s+1}{a}}\right]$$
$$- \left(\frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k\right)^2 [n(2k+1) - (3k+1)(k+1)] \tag{5.56}$$

and for fixed a we moreover have

$$\lim_{n\to\infty} \frac{\mathrm{Var}\left(Y^a_{\geq k,n}\right)}{n} = \frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k \left[1 + \frac{2k}{a} - \frac{2k+1}{a^{k+1}} \sum_{s=1}^{a} s^k\right]$$
$$+ \frac{2}{a^{k+1}} \sum_{s=1}^{a-1} \frac{1}{s} \left[1 - \left(\frac{a-s}{a}\right)^k\right] \sum_{r=1}^{a-s+1} r^k. \tag{5.57}$$

For fixed n we moreover get asymptotically

$$\mathrm{Var}\left(Y^a_{\geq k,n}\right) \xrightarrow{a\to\infty} \frac{n-k-1}{k+1} - 2\frac{(n-k-1)}{k+1} H_{k+1} + 2\frac{n-2k-1}{k+1} H_{2k+1}$$
$$+ \frac{2}{k+1} \sum_{\ell=0}^{k-1} H_{k+\ell+1} - \frac{n(2k+1) - (3k+1)(k+1)}{(k+1)^2}. \tag{5.58}$$

If we then take the limit as $n \to \infty$ we get

$$\lim_{n\to\infty} \lim_{a\to\infty} \frac{\mathrm{Var}\left(Y^a_{\geq k,n}\right)}{n} = \frac{1}{k+1} + \frac{2H_{2k+1} - 2H_{k+1}}{k+1} - \frac{2k+1}{(k+1)^2}. \tag{5.59}$$

Proof. See Appendix A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

Theorem 5.19. *Let $k \in \mathbb{N}$ and let $Y^p_{\geq k,n}$ denote the number of nodes with at least k descendants in a p-BRT. Then*

$$d_W\left(Y^p_{\geq k,n}, \mathcal{G}\right) \leq \frac{2k+1}{\sqrt{\mathrm{Var}\left(Y^p_{\geq k,n}\right)}} \left((2k+1) + \frac{\sqrt{28}(2k+1)^{\frac{1}{2}}}{\sqrt{\pi}}\right). \tag{5.60}$$

Proof. We will again use Theorem 2.12. For this we set $Y_i := C_i^k - \mathbb{E}[C_i^k]$ for $i = 1, \ldots, n-k$. The dependency neighbourhoods are $N_i = \{C_j^k : i - k \leq j \leq i + k\}$ for $i \leq n - 2k$ and $N_i = \{C_j^k : i - k \leq j \leq n - k\}$ for $i = n - 2k + 1, \ldots, n - k$. Hence we have $D = \max_{1 \leq i \leq n}\{|N_i|\} = 2k + 1$. Let $\sigma^2 = \mathrm{Var}\left(\sum_{i=1}^n C_i^k\right)$ and define

$$W := \sum_{i=1}^n \frac{C_i^k - \mathbb{E}[C_i^k]}{\sqrt{\mathrm{Var}(\sum_{i=1}^n C_i^k)}}. \tag{5.61}$$

We now need to estimate $\sum_{i=1}^n \mathbb{E}[|Y_i|^3]$ and $\sum_{i=1}^n \mathbb{E}[Y_i^4]$. Since the Y_i's can take values $1 - p_i$ or p_i, where $p_i = \mathbb{E}[C_i^k]$, we have as in the proof of Theorem 4.8 that

$$\begin{aligned} \mathbb{E}[|Y_i|^3] &= |1 - p_i|^3 p_i + |-p_i|^3(1 - p_i) \\ &= p_i(1 - p_i)((1 - p_i)^2 + p_i^2) \\ &= p_i(1 - p_i)(1 - 2p_i(1 - p_i)) \\ &\leq \mathrm{Var}[C_i^k] \end{aligned} \tag{5.62}$$

since for $0 < a < 1$, we have $0 < a(1 - a) < \frac{1}{4}$ and thus $1 > 1 - 2p_i(1 - p_i) > \frac{1}{2}$.

Similarly

$$
\begin{aligned}
\mathbb{E}[|Y_i|^3] &= (1-p_i)^4 p_i + p_i^4(1-p_i) \\
&= p_i(1-p_i)((1-p_i)^3 + p_i^3) \\
&= p_i(1-p_i)(1 - 3(pi(1-p_i))) \\
&\leq \mathrm{Var}(C_i^k)
\end{aligned}
\tag{5.63}
$$

since for $0 < a < 1$, we have $1 > 1 - 3p_i(1-p_i) > \frac{1}{4}$.

Thus we get

$$
\begin{aligned}
d_W(W,\mathcal{G}) &\leq \frac{(2k+1)^2}{\sigma^3} \sum_{i=1}^{n-k} \mathrm{Var}(C_i^k) + \frac{\sqrt{28}(2k+1)^{\frac{3}{2}}}{\sqrt{\pi}\sigma^2} \sqrt{\sum_{i=1}^{n} \mathrm{Var}(C_i^k)} \\
&= \frac{2k+1}{\sigma}\left((2k+1) + \frac{\sqrt{28}(2k+1)^{\frac{1}{2}}}{\sqrt{\pi}}\right).
\end{aligned}
\tag{5.64}
$$

$\square$

Remark 5.20. As we know by Theorem 5.17 that the variance is of order n, the bound in Theorem 5.19 decreases with order $\frac{1}{\sqrt{n}}$.

Remark 5.21. Since $Y_{\leq k,n}^p$, the number of nodes with at most k descendants is equal to $n - Y_{\geq k+1,n}^p$ the expectation, variance and CLT of $Y_{\leq k,n}^p$ follows directly from the above results.

Moreover the expectation and variance of the number of nodes with exactly k descendants can be calculated from the results of the previous section by some additional calculations.

Corollary 5.22. *Let $X_{k,n}^p$ denote the number of nodes with exactly k descendants in a p-BRT. Then*

$$
\begin{aligned}
\mathbb{E}[X_{k,n}^p] = (n-k-1) \sum_{s=1}^{a} p_s \left(\left(\sum_{r=s}^{a} p_r \right)^k - \left(\sum_{r=s}^{a} p_r \right)^{k+1} \right) \\
+ \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^{k+1}.
\end{aligned}
\tag{5.65}
$$

Moreover if $X_{k,n}^a$ is the number of nodes with exactly k descendants in a a-RT, then

$$
\mathbb{E}[X_{k,n}^a] = (n-k-1) \left(\frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k - \frac{1}{a^{k+2}} \sum_{s=1}^{a} s^{k+1} \right) + \frac{1}{a^{k+2}} \sum_{s=1}^{a} s^{k+1}
\tag{5.66}
$$

and asymptotically

$$
\mathbb{E}[X_{k,n}^a] \xrightarrow{a\to\infty} \frac{n}{(k+1)(k+2)}.
\tag{5.67}
$$

Proof. We have $X_{k,n}^p = Y_{\geq k,n}^p - Y_{\geq k+1,n}^p$. Thus we get from Theorem 5.14,

$$
\begin{aligned}
\mathbb{E}\left[X_{k,n}^p\right] &= \mathbb{E}\left[Y_{\geq k,n}^p\right] - \mathbb{E}\left[Y_{\geq k+1,n}^p\right] \\
&= (n-k-1) \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k + 1 - (n-k-2) \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^{k+1} - 1 \\
&= (n-k-1) \sum_{s=1}^{a} p_s \left(\left(\sum_{r=s}^{a} p_r \right)^k - \left(\sum_{r=s}^{a} p_r \right)^{k+1} \right) + \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^{k+1}.
\end{aligned}
\tag{5.68}
$$

Similarly we get from Theorem 5.14 for $\vec{p}$ uniform,

$$
\mathbb{E}[X_{k,n}^a] = (n-k-1) \left(\frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k - \frac{1}{a^{k+2}} \sum_{s=1}^{a} s^{k+1} \right) + \frac{1}{a^{k+2}} \sum_{s=1}^{a} s^{k+1}.
\tag{5.69}
$$

This implies in particular that

$$
\begin{aligned}
\mathbb{E}[X_{k,n}^a] & \\
&= (n - k - 1)\left(\frac{1}{a^{k+1}}\left[\frac{a^{k+1}}{k+1} + \mathcal{O}(a^k)\right] - \frac{1}{a^{k+2}}\left[\frac{a^{k+2}}{k+2} + \mathcal{O}(a^{k+1})\right]\right) \\
&\quad + \frac{1}{a^{k+2}}\left[\frac{a^{k+2}}{k+2} + \mathcal{O}(a^{k+1})\right] \\
&\xrightarrow{a \to \infty} (n - k - 1)\left(\frac{1}{k+1} - \frac{1}{k+2}\right) + \frac{1}{k+2} \\
&= \frac{n}{(k+1)(k+2)} - \frac{k+1}{(k+1)(k+2)} + \frac{1}{k+2} \\
&= \frac{n}{(k+1)(k+2)}.
\end{aligned}
\tag{5.70}
$$

$\square$

Remark 5.23. For the variance of the number of nodes with exactly k descendants we have that

$$
\begin{aligned}
\mathrm{Var}\left(X_{k,n}^p\right) &= \mathrm{Var}\left(Y_{\geq k,n}^p - Y_{\geq k+1,n}^p\right) \\
&= \mathrm{Var}\left(Y_{\geq k,n}^p\right) + \mathrm{Var}\left(Y_{\geq k+1,n}^p\right) - 2\mathrm{Cov}\left(Y_{\geq k,n}^p, Y_{\geq k+1,n}^p\right).
\end{aligned}
\tag{5.71}
$$

The covariance can be calculated by the same method as before, so we can obtain an expression for $\mathrm{Var}\left(X_{k,n}^p\right)$. Since the calculations as well as the expressions one gets in the end are quite long, we did not include them here. The asymptotic value of $\mathrm{Var}\left(X_{k,n}^a\right)$ as $a \to \infty$ can also be calculated using the same methods as in the proof of Theorem 5.10. As expected the value one gets in the end for $\lim_{n \to \infty} \lim_{a \to \infty} \frac{\mathrm{Var}\left(X_{k,n}^a\right)}{n}$ corresponds to the value of the same expression for the number of nodes with exactly k descendants of URTs, which is stated in Theorem 3.6.

5.5. Depth of Node n

First we observe that given a permutation representation for a recursive tree we can determine $\mathcal{D}_n^p$, the depth of node n by counting the steps we go down from that position when we go to the left until 1. It is important to note that we must take any step down if we can, and we don't go up again once we went down. This implies that

we are not looking for the longest increasing subsequence from 1 to n. Rather we are looking for all anti-records when we start at n and then go to the left until we reach 1. The following examples will clarify this difference:

Example 5.24. If $\gamma = 13456728$ the depth of node 8 is 2, because it is attached to 2 which is attached to 1. This corresponds to the number of steps down we take from 8 since we first go down to 2 and then down to 1. The longest increasing subsequence is 1345678, which is much longer.

For $\rho = 12574863$ the depth of node 8 is 3 by the construction principle. We can also see this because from 8 we go down to 4 then there is 7 and 5, which is higher so we ignore it. Then we go down to 2, and finally to 1. So we take 3 steps down. We don't count the step down from 7 to 5 since we already went down to 4 before that, so we are not interested in anything concerning nodes bigger than 4 anymore.

Thus given $\mathcal{P}_n = \gamma^{-1}(n)$, the position of n in the permutation, we get the depth of node n by calculating for all $i < \mathcal{P}_n$, the probability that $\gamma(i) < \min_{i<j<\mathcal{P}_n}\{\gamma(j)\}$. We will now use these observations to prove the following theorem:

Theorem 5.25. *Let $\mathcal{D}_n^p$ denote the depth of node n in a p-BRT. Then*

$$
\mathbb{E}[\mathcal{D}_n^p] =
$$

$$
\sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} \left[\left(\sum_{r=1}^{s'} p_r \right)^{n-1} - \left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} \right]
$$

$$
- \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \frac{\left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} - \left(\sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r \right)^{n-1}}{\sum_{r=1}^{s'-1} p_r - \sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r}
$$

$$
+ p_1 \sum_{s=2}^{a} \frac{1}{p_s}
$$

$$
\cdot \left[(n-2) \left(\sum_{r=1}^{s} p_r \right)^{n} - (n-1) \left(\sum_{r=1}^{s} p_r \right)^{n-1} \sum_{r=1}^{s-1} p_r + \sum_{r=1}^{s} p_r \left(\sum_{r=1}^{s-1} p_r \right)^{n-1} \right]
$$

$$
+ \sum_{s=2}^{a} \left(\left(\sum_{r=1}^{s} p_r \right)^{n-1} - \left(\sum_{r=1}^{s-1} p_r \right)^{n-1} \right)
$$

$$
+ \left[\sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r \right)^{n-2}}{\sum_{r=1}^{s-1} p_r} + (n-2)p_1 + 1 \right] p_1^{n-1}.
$$

(5.72)

Moreover asymptotically we have

$$
\lim_{n \to \infty} \frac{\mathbb{E}[\mathcal{D}_n^p]}{n} = p_1.
$$

(5.73)

As we will use the position of node n in order to derive its depth, we will need the following lemma.

Lemma 5.26. *Let $\mathcal{P}_n^p$ denote the position of n in a p-biased random permutation. Then for $2 \leq k < n$,*

$$
\mathbb{P}\left(\mathcal{P}_n^p = k \right) = \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r \right)^{k-2} \left(\sum_{r=1}^{s-1} p_r \right)^{n-k}
$$

(5.74)

and

$$\mathbb{P}\left(\mathcal{P}_n^p = n\right) = \sum_{s=1}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{n-2}. \tag{5.75}$$

Proof. As before we will use the construction of an inverse biased riffle shuffle permutation as introduced in Theorem 5.3. We know that n will be the last card in the last non-empty pile. Also the digits we assign each index when constructing an inverse riffle shuffle permutation define the position of the cards in the pile with that digit. In other words, the indices that get digit s, are the positions of the cards in the s-th pile. This means that n gets the position of the last index that gets the highest digit. We get

$$\mathcal{P}_n^p = \max\left\{2 \leq i \leq n : X_i \geq \{X_2, \ldots, X_n\}\right\} \tag{5.76}$$

so

$$\mathbb{P}(\mathcal{P}_n^p = k) = \mathbb{P}\left(X_k \geq \{X_2, \ldots, X_{k-1}\}, X_k > \{X_{k+1}, \ldots, X_n\}\right). \tag{5.77}$$

By conditioning on X_k we get independent events and can thus calculate this probability for $2 \leq k < n$. We get

$$
\begin{aligned}
\mathbb{P}&(\mathcal{P}_n^p = k) \\
&= \mathbb{P}\left(X_k \geq \{X_2, \ldots, X_{k-1}\}, X_k > \{X_{k+1}, \ldots, X_n\}\right) \\
&= \sum_{s=1}^{a} \mathbb{P}\left(X_k \geq \{X_2, \ldots, X_{k-1}\}, X_k > \{X_{k+1}, \ldots, X_n\} | X_k = s\right) \mathbb{P}(X_k = s) \\
&= \sum_{s=1}^{a} \mathbb{P}\left(s \geq \{X_2, \ldots, X_{k-1}\}, s > \{X_{k+1}, \ldots, X_n\}\right) \mathbb{P}(X_k = s) \\
&= \sum_{s=1}^{a} \mathbb{P}\left(s \geq \{X_2, \ldots, X_{k-1}\}\right) \mathbb{P}\left(s > \{X_{k+1}, \ldots, X_n\}\right) \mathbb{P}(X_k = s) \\
&= \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{k-2} \left(\sum_{r=1}^{s-1} p_r\right)^{n-k}.
\end{aligned}
\tag{5.78}
$$

We only sum from $s = 2$ since if $X_k = 1$ it cannot be strictly greater than $X_{k+1}, \ldots, X_n$. If $k = n$ we have the additional possibility that all X_i are equal to 1 and thus get

$$\mathbb{P}(\mathcal{P}_n^p = n) = \sum_{s=1}^{a} p_s \left(\sum_{r=1}^{s} p_r \right)^{n-2}. \tag{5.79}$$

$\square$

We can now start with the proof of Theorem 5.25.

Proof of Theorem 5.25. First of all we will find an expression for $\mathbb{E}[\mathcal{D}_n^p | \mathcal{P}_n^p = k]$. As we said above, given $\mathcal{P}_n^p$ we need to calculate the number of anti-records when we go from the position of n to the left until we reach 1. Hence, given $\mathcal{P}_n^p$, we define for $1 \leq i < \mathcal{P}_n^p$,

$$E_i := \mathbf{1} \left(\gamma(i) = \min_{i \leq k \leq \mathcal{P}_n^p} \{ \gamma(k) \} \right). \tag{5.80}$$

Then, given $\mathcal{P}_n^p$, we get $\mathcal{D}_n^p = \sum_{i=1}^{\mathcal{P}_n^p - 1} E_i$, and so

$$\mathbb{E}[\mathcal{D}_n^p | \mathcal{P}_n^p] = \sum_{i=1}^{\mathcal{P}_n^p - 1} \mathbb{E}[E_i | \mathcal{P}_n^p]. \tag{5.81}$$

We can simplify this sum by first observing that

$$\mathbb{E}(E_{\mathcal{P}_n^p - 1}) = \mathbb{P}(\gamma(\mathcal{P}_n^p - 1) < \gamma(\mathcal{P}_n^p)) = \mathbb{P}(\gamma(\mathcal{P}_n^p - 1) < n) = 1 \tag{5.82}$$

and in general for all i, we have $\gamma(i) < \gamma(\mathcal{P}_n^p) = n$, so we can rewrite E_i as

$$E_i = \mathbf{1} \left(\gamma(i) = \min_{i \leq k < \mathcal{P}_n^p} \{ \gamma(k) \} \right). \tag{5.83}$$

Moreover $\mathbb{P}(E_1) = \mathbb{P}(\gamma(1) = \min_{1 \leq k \leq \mathcal{P}_n^p} \{ \gamma(k) \}) = 1$ since $\gamma(1) = 1$. Now we will again use the inverse riffle shuffle construction to calculate the rest of these probabilities. We know that $\gamma(i) < \gamma(j)$ for all $i < j$ if and only if $X_i \leq X_j$, since this means that in

the spot j will come a higher card from the same pile or from a pile corresponding to a higher digit, thus with higher labeled cards.

Let $2 \leq i \leq \mathcal{P}_n^p$, then

$$\mathbb{E}[E_i|\mathcal{P}_n^p]$$
$$= \mathbb{P}\left(\gamma(i) = \min_{i \leq k < \mathcal{P}_n^p}\{\gamma(k)\}|\gamma(\mathcal{P}_n^p) = n\right)$$
$$= \mathbb{P}\left(X_i = \min_{i \leq k < \mathcal{P}_n^p}\{X_k\}|X_{\mathcal{P}_n^p} \geq \{X_2,\ldots,X_{\mathcal{P}_n^p-1}\}, X_{\mathcal{P}_n^p} > \{X_{\mathcal{P}_n^p+1},\ldots,X_n\}\right)$$
$$= \mathbb{P}\left(X_i = \min_{i \leq k < \mathcal{P}_n^p}\{X_k\}\right)$$
$$= \sum_{s=1}^{a} \mathbb{P}\left(X_i = \min_{i \leq k < \mathcal{P}_n^p}\{X_k\}|X_i = s\right)\mathbb{P}(X_i = s) \tag{5.84}$$
$$= \sum_{s=1}^{a} \mathbb{P}\left(s \leq \{X_{i+1},\ldots,X_{\mathcal{P}_n^p-1}\}\right)\mathbb{P}(X_i = s)$$
$$= \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^{\mathcal{P}_n^p-1-(i+1)+1}$$
$$= \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^{\mathcal{P}_n^p-i-1}.$$

We could get rid of the conditional in the third line because the X_k are mutually independent and thus the fact that $X_{\mathcal{P}_n^p} \geq \{X_2,\ldots,X_{\mathcal{P}_n^p-1}\}$ does not tell us anything about the ordering of $\{X_2,\ldots,X_{\mathcal{P}_n^p-1}\}$. Moreover $X_{\mathcal{P}_n^p} > \{X_{\mathcal{P}_n^p+1},\ldots,X_n\}$ concerns different X_is, so is also independent of E_i. We again assume $p_i > 0$ for all i in order to

avoid division by 0. In total we thus get, for $2 < k \leq n$,

$$
\begin{aligned}
\mathbb{E}[\mathcal{D}_n^p | \mathcal{P}_n^p = k] &= \sum_{i=1}^{k-1} \mathbb{E}[E_i | \mathcal{P}_n^p = k] \\
&= \sum_{i=2}^{k-1} \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^{k-i-1} + 1 \\
&= \sum_{s=1}^{a} p_s \sum_{i=2}^{k-1} \left(\sum_{r=s}^{a} p_r \right)^{k-i-1} + 1 \\
&= \sum_{s=1}^{a} p_s \sum_{i=0}^{k-3} \left(\sum_{r=s}^{a} p_r \right)^{i} + 1 \\
&= \sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r \right)^{k-2}}{1 - \sum_{r=s}^{a} p_r} + (k-2)p_1 + 1 \\
&= \sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r \right)^{k-2}}{\sum_{r=1}^{s-1} p_r} + (k-2)p_1 + 1
\end{aligned}
\tag{5.85}
$$

and

$$
\mathbb{E}[\mathcal{D}_n^p | \mathcal{P}_n^p = 2] = 1.
\tag{5.86}
$$

To get $\mathbb{E}[\mathcal{D}_n^p]$ we now need to use the tower rule, this gives

$$\mathbb{E}[\mathcal{D}_n^p]$$

$$= \mathbb{E}\left[\mathbb{E}[\mathcal{D}_n^p|\mathcal{P}_n^p = k]\right]$$

$$= \sum_{k=2}^{n} \mathbb{E}[\mathcal{D}_n^p|\mathcal{P}_n^p = k]\mathbb{P}(\mathcal{P}_n^p = k)$$

$$= \sum_{k=3}^{n-1} \mathbb{E}[\mathcal{D}_n^p|\mathcal{P}_n^p = k]\mathbb{P}(\mathcal{P}_n^p = k)$$

$$+ \mathbb{E}[\mathcal{D}_n^p|\mathcal{P}_n^p = n]\mathbb{P}(\mathcal{P}_n^p = n)$$

$$+ \mathbb{E}[\mathcal{D}_n^p|\mathcal{P}_n^p = 2]\mathbb{P}(\mathcal{P}_n^p = 2)$$

$$= \sum_{k=3}^{n-1} \left[\sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r\right)^{k-2}}{\sum_{r=1}^{s-1} p_r} + (k-2)p_1 + 1 \right]$$

$$\cdot \left[\sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{k-2} \left(\sum_{r=1}^{s-1} p_r\right)^{n-k} \right]$$

$$+ \left[\sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r\right)^{n-2}}{\sum_{r=1}^{s-1} p_r} + (n-2)p_1 + 1 \right] \left[\sum_{s=1}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{n-2} \right]$$

$$+ \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r\right)^{n-2}$$

$$= \sum_{k=3}^{n} \left[\sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r\right)^{k-2}}{\sum_{r=1}^{s-1} p_r} \right] \left[\sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{k-2} \left(\sum_{r=1}^{s-1} p_r\right)^{n-k} \right]$$

$$+ \sum_{k=3}^{n} \left[(k-2)p_1 + 1\right] \left[\sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{k-2} \left(\sum_{r=1}^{s-1} p_r\right)^{n-k} \right]$$

$$+ \left[\sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r\right)^{n-2}}{\sum_{r=1}^{s-1} p_r} + (n-2)p_1 + 1 \right] p_1^{n-1}$$

$$+ \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r\right)^{n-2} .$$

$$(5.87)$$

In order to simplify this expression we first look at the first line

$$
\sum_{k=3}^{n} \left[\sum_{s=2}^{a} p_s \frac{1 - (\sum_{r=s}^{a} p_r)^{k-2}}{\sum_{r=1}^{s-1} p_r} \right] \left[\sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r \right)^{k-2} \left(\sum_{r=1}^{s-1} p_r \right)^{n-k} \right]
$$

$$
= \sum_{k=3}^{n} \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left(1 - \left(\sum_{r=s}^{a} p_r \right)^{k-2} \right) \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-k} \tag{5.88}
$$

$$
= \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left[\sum_{k=3}^{n} \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-k} \right.
$$

$$
\left. - \sum_{k=3}^{n} \left(\sum_{r=s}^{a} p_r \right)^{k-2} \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-k} \right].
$$

Now we have

$$
\sum_{k=3}^{n} \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-k}
$$

$$
= \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2} \sum_{k=3}^{n} \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{-k+2}
$$

$$
= \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2} \left(\frac{1 - \left(\left(\sum_{r=1}^{s'} p_r \right) \left(\sum_{r=1}^{s'-1} p_r \right)^{-1} \right)^{n-1}}{1 - \left(\sum_{r=1}^{s'} p_r \right) \left(\sum_{r=1}^{s'-1} p_r \right)^{-1}} - 1 \right) \tag{5.89}
$$

$$
= \frac{\left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} - \left(\sum_{r=1}^{s'} p_r \right)^{n-1}}{\sum_{r=1}^{s'-1} p_r - \sum_{r=1}^{s'} p_r} - \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2}.
$$

Inserting this into the first part of the last line of (5.88) we get

$$
\sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \sum_{k=3}^{n} \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-k}
$$

$$
= \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left[\frac{\left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} - \left(\sum_{r=1}^{s'} p_r \right)^{n-1}}{\sum_{r=1}^{s'-1} p_r - \sum_{r=1}^{s'} p_r} - \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2} \right]
$$

$$
= \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \frac{\left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} - \left(\sum_{r=1}^{s'} p_r \right)^{n-1}}{-p_{s'}}
$$

$$
\quad - \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2}
$$

$$
= \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} \left[\left(\sum_{r=1}^{s'} p_r \right)^{n-1} - \left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} \right]
$$

$$
\quad - \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2}.
$$

(5.90)

Similarly we have that

$$
\sum_{k=3}^{n} \left(\sum_{r=s}^{a} p_r \right)^{k-2} \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-k}
$$

$$
= \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2} \sum_{k=3}^{n} \left(\sum_{r=s}^{a} p_r \right)^{k-2} \left(\sum_{r=1}^{s'} p_r \right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r \right)^{-(k+2)}
$$

$$
= \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2} \left(\frac{1 - \left(\sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r \left(\sum_{r=1}^{s'-1} p_r \right)^{-1} \right)^{n-1}}{1 - \sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r \left(\sum_{r=1}^{s'-1} p_r \right)^{-1}} - 1 \right)
$$

$$
= \left(\frac{\left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} - \left(\sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r \right)^{n-1}}{\sum_{r=1}^{s'-1} p_r - \sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r} - \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2} \right).
$$

(5.91)

Inserting this expression into the second part of the last line of (5.88) we get

$$
\sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \sum_{k=3}^{n} \left(\sum_{r=s}^{a} p_r\right)^{k-2} \left(\sum_{r=1}^{s'} p_r\right)^{k-2} \left(\sum_{r=1}^{s'-1} p_r\right)^{n-k}
$$

$$
= \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \frac{\left(\sum_{r=1}^{s'-1} p_r\right)^{n-1} - \left(\sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r\right)^{n-1}}{\sum_{r=1}^{s'-1} p_r - \sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r} \tag{5.92}
$$

$$
- \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left(\sum_{r=1}^{s'-1} p_r\right)^{n-2}.
$$

Now we look at the second line. First we separate it into two terms,

$$
\sum_{k=3}^{n} [(k-2)p_1 + 1] \left[\sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{k-2} \left(\sum_{r=1}^{s-1} p_r\right)^{n-k}\right]
$$

$$
= \sum_{k=3}^{n} (k-2)p_1 \left[\sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{k-2} \left(\sum_{r=1}^{s-1} p_r\right)^{n-k}\right] \tag{5.93}
$$

$$
+ \sum_{k=3}^{n} \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r\right)^{k-2} \left(\sum_{r=1}^{s-1} p_r\right)^{n-k}
$$

$$
:= A_1 + A_2.
$$

Now we easily get

$$
\begin{aligned}
A_2 &= \sum_{k=3}^{n} \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r \right)^{k-2} \left(\sum_{r=1}^{s-1} p_r \right)^{n-k} \\
&= \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r \right)^{n-2} \sum_{k=3}^{n} \left(\sum_{r=1}^{s} p_r \right)^{k-2} \left(\sum_{r=1}^{s-1} p_r \right)^{-k+2} \\
&= \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r \right)^{n-2} \left(\frac{1 - \left(\sum_{r=1}^{s} p_r \right)^{n-1} \left(\sum_{r=1}^{s-1} p_r \right)^{-(n-1)}}{1 - \sum_{r=1}^{s} p_r \left(\sum_{r=1}^{s-1} p_r \right)^{-1}} - 1 \right) \\
&= \sum_{s=2}^{a} p_s \left(\frac{\left(\sum_{r=1}^{s-1} p_r \right)^{n-1} - \left(\sum_{r=1}^{s} p_r \right)^{n-1}}{\sum_{r=1}^{s-1} p_r - \sum_{r=1}^{s} p_r} - \left(\sum_{r=1}^{s-1} p_r \right)^{n-2} \right) \\
&= \sum_{s=2}^{a} p_s \left(\frac{\left(\sum_{r=1}^{s} p_r \right)^{n-1} - \left(\sum_{r=1}^{s-1} p_r \right)^{n-1}}{p_s} - \left(\sum_{r=1}^{s-1} p_r \right)^{n-2} \right) \\
&= \sum_{s=2}^{a} \left(\left(\sum_{r=1}^{s} p_r \right)^{n-1} - \left(\sum_{r=1}^{s-1} p_r \right)^{n-1} \right) - \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r \right)^{n-2}.
\end{aligned}
\tag{5.94}
$$

For the first term we need a little bit more work. First of all

$$
\begin{aligned}
A_1 &= \sum_{k=3}^{n} (k-2) p_1 \left[\sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s} p_r \right)^{k-2} \left(\sum_{r=1}^{s-1} p_r \right)^{n-k} \right] \\
&= p_1 \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r \right)^{n-2} \sum_{k=3}^{n} (k-2) \left(\sum_{r=1}^{s} p_r \right)^{k-2} \left(\sum_{r=1}^{s-1} p_r \right)^{-k+2}.
\end{aligned}
\tag{5.95}
$$

We now set $x := \left(\sum_{r=1}^{s} p_r \left(\sum_{r=1}^{s-1} p_r\right)^{-1}\right)$ and only look at the sum indexed with k. Then we have a sum of the form

$$
\begin{aligned}
\sum_{k=3}^{n}(k-2)x^{k-2} &= x\sum_{k=1}^{n-2}kx^{k-1} \\
&= x\frac{d}{dx}\sum_{k=0}^{n-2}x^k \\
&= x\frac{d}{dx}\frac{1-x^{n-1}}{1-x} \\
&= x\frac{-(n-1)x^{n-2}(1-x)-(1-x^{n-1})(-1)}{(1-x)^2} \\
&= x\frac{(n-1)x^{n-1}-(n-1)x^{n-2}+1-x^{n-1}}{(1-x)^2} \\
&= x\frac{(n-2)x^{n-1}-(n-1)x^{n-2}+1}{(1-x)^2}.
\end{aligned}
$$

(5.96)

We now substitute $\left(\sum_{r=1}^{s} p_r \left(\sum_{r=1}^{s-1} p_r\right)^{-1}\right)$ for x and insert the resulting expression into A_1. This gives

$$
\begin{aligned}
A_1 &= p_1\sum_{s=2}^{a}p_s\left(\sum_{r=1}^{s-1}p_r\right)^{n-2}\left(\sum_{r=1}^{s}p_r\left(\sum_{r=1}^{s-1}p_r\right)^{-1}\right) \\
&\quad \cdot\left[\frac{(n-2)\left(\sum_{r=1}^{s}p_r\left(\sum_{r=1}^{s-1}p_r\right)^{-1}\right)^{n-1}}{\left(1-\left(\sum_{r=1}^{s}p_r\left(\sum_{r=1}^{s-1}p_r\right)^{-1}\right)\right)^2}\right. \\
&\quad\left.-\frac{(n-1)\left(\sum_{r=1}^{s}p_r\left(\sum_{r=1}^{s-1}p_r\right)^{-1}\right)^{n-2}-1}{\left(1-\left(\sum_{r=1}^{s}p_r\left(\sum_{r=1}^{s-1}p_r\right)^{-1}\right)\right)^2}\right] \\
&= p_1\sum_{s=2}^{a}p_s\cdot\left[\frac{(n-2)\left(\sum_{r=1}^{s}p_r\right)^n}{\left(\sum_{r=1}^{s-1}p_r-\sum_{r=1}^{s}p_r\right)^2}\right. \\
&\quad\left.-\frac{(n-1)\left(\sum_{r=1}^{s}p_r\right)^{n-1}\sum_{r=1}^{s-1}p_r-\sum_{r=1}^{s}p_r\left(\sum_{r=1}^{s-1}p_r\right)^{n-1}}{\left(\sum_{r=1}^{s-1}p_r-\sum_{r=1}^{s}p_r\right)^2}\right] \\
&= p_1\sum_{s=2}^{a}\frac{1}{p_s}\left[(n-2)\left(\sum_{r=1}^{s}p_r\right)^n\right. \\
&\quad\left.-(n-1)\left(\sum_{r=1}^{s}p_r\right)^{n-1}\sum_{r=1}^{s-1}p_r+\sum_{r=1}^{s}p_r\left(\sum_{r=1}^{s-1}p_r\right)^{n-1}\right].
\end{aligned}
$$

(5.97)

In total now we have as an expression for the expectation

$$\mathbb{E}[\mathcal{D}_n^p] =$$

$$\sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} \left[\left(\sum_{r=1}^{s'} p_r \right)^{n-1} - \left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} \right]$$

$$- \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2}$$

$$- \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \frac{\left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} - \left(\sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r \right)^{n-1}}{\sum_{r=1}^{s'-1} p_r - \sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r}$$

$$+ \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \left(\sum_{r=1}^{s'-1} p_r \right)^{n-2}$$

$$+ p_1 \sum_{s=2}^{a} \frac{1}{p_s} \left[(n-2) \left(\sum_{r=1}^{s} p_r \right)^{n} \right.$$

$$\left. - (n-1) \left(\sum_{r=1}^{s} p_r \right)^{n-1} \sum_{r=1}^{s-1} p_r + \sum_{r=1}^{s} p_r \left(\sum_{r=1}^{s-1} p_r \right)^{n-1} \right]$$

$$+ \sum_{s=2}^{a} \left(\left(\sum_{r=1}^{s} p_r \right)^{n-1} - \left(\sum_{r=1}^{s-1} p_r \right)^{n-1} \right) - \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r \right)^{n-2}$$

$$+ \left[\sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r \right)^{n-2}}{\sum_{r=1}^{s-1} p_r} + (n-2)p_1 + 1 \right] p_1^{n-1}$$

$$+ \sum_{s=2}^{a} p_s \left(\sum_{r=1}^{s-1} p_r \right)^{n-2} .$$

(5.98)

After some simplifications we get

$$\mathbb{E}[\mathcal{D}_n^p]$$

$$= \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} \left[\left(\sum_{r=1}^{s'} p_r \right)^{n-1} - \left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} \right]$$

$$- \sum_{s=2}^{a} \frac{p_s}{\sum_{r=1}^{s-1} p_r} \sum_{s'=2}^{a} p_{s'} \frac{\left(\sum_{r=1}^{s'-1} p_r \right)^{n-1} - \left(\sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r \right)^{n-1}}{\sum_{r=1}^{s'-1} p_r - \sum_{r=s}^{a} p_r \sum_{r=1}^{s'} p_r}$$

$$+ p_1 \sum_{s=2}^{a} \frac{1}{p_s} \left[(n-2) \left(\sum_{r=1}^{s} p_r \right)^{n} \right. \tag{5.99}$$

$$\left. - (n-1) \left(\sum_{r=1}^{s} p_r \right)^{n-1} \sum_{r=1}^{s-1} p_r + \sum_{r=1}^{s} p_r \left(\sum_{r=1}^{s-1} p_r \right)^{n-1} \right]$$

$$+ \sum_{s=2}^{a} \left(\left(\sum_{r=1}^{s} p_r \right)^{n-1} - \left(\sum_{r=1}^{s-1} p_r \right)^{n-1} \right)$$

$$+ \left[\sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{r=s}^{a} p_r \right)^{n-2}}{\sum_{r=1}^{s-1} p_r} + (n-2)p_1 + 1 \right] p_1^{n-1}.$$

Asymptotically this gives, for $\vec{p}$ such that $p_i \neq 0$ for all $i \in [a]$ and $a > 2$, for $n \to \infty$,

$$\lim_{n \to \infty} \frac{\mathbb{E}[\mathcal{D}_n^p]}{n} = \lim_{n \to \infty} p_1 \frac{(n-2) - (n-1) \sum_{r=1}^{a-1} p_r}{n p_a} = p_1 \tag{5.100}$$

since

$$p_1 \frac{(n-2) - (n-1) \sum_{r=1}^{a-1} p_r}{p_a} = p_1 \frac{(n-1) \left(1 - \sum_{r=1}^{a-1} p_r \right)}{p_a} - \frac{p_1}{p_a} = (n-1)p_1 - \frac{p_1}{p_a}. \tag{5.101}$$

$\square$

As a corollary we get for $\vec{p}$ uniformly distributed over a the following results.

Corollary 5.27. *Let $\mathcal{D}_n^a$ denote the depth of node n in an a-RT. Then*

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^a\right] = &H_{a-1} + 1 + \frac{n-2}{a^n} - \frac{1}{a^{2n-3}} \sum_{s=1}^{a-1} \frac{(a-s)^{n-2}}{s} \\
&- \frac{1}{a^{n-2}} \sum_{s=1}^{a-1} \sum_{s'=1}^{a-1} \frac{1}{s} \frac{s'^{n-1} - ((a-s)(s'+1))^{n-1}}{ss' + s - a} \\
&+ \frac{1}{a^n} \sum_{s=1}^{a-1} (n-2)(s+1)^n - (n-1)(s+1)^{n-1}s + (s+1)s^{n-1}.
\end{aligned}
\tag{5.102}
$$

Moreover asymptotically

$$
\frac{\mathbb{E}\left[\mathcal{D}_n^a\right]}{n} \xrightarrow{n \to \infty} \frac{1}{a}
\tag{5.103}
$$

and as a approaches infinity, we get the same expectation as for URTs:

$$
\mathbb{E}\left[\mathcal{D}_n^a\right] \xrightarrow{a \to \infty} H_{n-1}.
\tag{5.104}
$$

Proof. See Appendix A. $\qquad\square$

6. CONCLUSION

We end the thesis with some concluding remarks which contain some discussions regarding the results above, and some questions which we are planning to pursue in subsequent work.

The main motivation for this thesis was to understand the 'most general' random tree model, which we call an inhomogeneous random tree, where each node has a specific probability of attaching to an already existing node. Namely, once we have i nodes present, node $i + 1$ attaches to node j, $1 \leq j \leq i$ with probability $p_{i+1,j}$ where $\sum_{k=1}^{i} p_{i+1,j} = 1$, and where the attachment is independent of the previous evolution of the tree. Both of the tree models discussed in this thesis provide approximations for inhomogeneous random trees, and our next step will be to see if our results can be translated into this more general framework by employing certain limiting arguments. We were not able to follow up asymptotic equivalence of various random tree models described in this thesis. In particular, results towards asymptotic equivalence of weighted recursive trees for varying weight sequences, and for certain classes of statistics should be of interest.

Among the problems we could not solve are the number of leaves of a general weighted recursive tree. The martingale and coupling argument we used only work when finitely many nodes are assigned a weight different from 1. There might be the possibility to use a generalization of Friedman's urn with a variable matrix. This would allow us to add at each step balls according to the weight of the newly attached node. Another way might be to use the exact probabilities we introduced, at least for specific kinds of weight sequences. This might give as an idea of what a result on asymptotic behaviour could look like.

Concerning weight sequences it would in general be of interest to find precise conditions on when central limit theorems apply and when the distribution is asymptotically close to the one of the uniform case. Especially concerning the expectation

of the depth we could not establish general conditions for a central limit theorem to hold, although many of the examples we checked exhibited a similar asymptotic behaviour. Additionally there are many statistics of interest we did not study or only for special cases, like the height, the maximum degree or the number of nodes with a certain number of descendants. We believe that the coupling constructions given in Chapter 4 will help us to analyze various other statistics of weighted recursive trees by making use of corresponding results for uniform recursive trees. This coupling can be used to obtain concentration inequalities for underlying statistics besides its use in understanding asymptotic distributions. We hope to follow this in an upcoming work.

Concerning biased recursive trees an open problem is the asymptotic distribution of the number of branches. Because of the global dependence of the anti-records, we could not apply any of the central limit theorems mentioned in the preliminaries. A different approach to the number of branches of biased recursive trees would be to analyze the cycle structure and more specifically the number of cycles of riffle shuffle permutations.

Further results on the depth are also necessary for biased recursive trees. While we could compute the expectation of the depth our method might be too complicated for computing the variance. Since the expectation only depends on p_1 when n is large, it would be interesting to see if and how the other parameters affect the variance.

We saw that in a biased recursive tree constructed from a riffle shuffle permutation based on the cutting of the deck into a piles, the maximum degree is a. Thus an interesting question would be to compare a-ary recursive tree with biased recursive trees and especially a-recursive trees. It is not clear if this common restriction on the degree of the nodes is the only property these trees have in common or if they are more similar that one might expect at first sight. In order to get insights concerning this relation a dynamic construction of biased recursive trees would be very useful. If such a construction exists it would probably be very different from the construction principles we know. Of course such a dynamic growth rule would be useful for many other questions as well.

Apart from the problems we could not solve for the models we investigated, there are moreover other non-uniform models that would be worth investigation in order to understand the behaviour of inhomogeneous recursive trees. We might gain further insight by considering weighted tree models where the weights change with time, or are also randomly distributed. Another possibility would be to use the bijection between binary recursive trees and permutations, see for example [51]. Extending uniform binary trees via other random permutation distributions in order to understand all sorts of random binary trees also seems a promising direction to follow.

Throughout the study of biased recursive trees, the use of biased riffle shuffles instead of uniformly random permutations stemmed from the fact that certain statistics could be expressed in terms of independent random variables. There is another random permutation framework allowing such use of independence, namely unfair permutations. See [59, 60] for the definition and analysis of various statistics of unfair permutations. We are not sure one would gain more insight for inhomogeneous random trees by replacing riffle shuffles with unfair permutations, and we believe that this should be checked in subsequent work. Since in that model the rank of i is determined by the maximum of i identically distributed independent uniform random variables, it probably has similarities to models where each node can choose among k potential parents.

REFERENCES

1. Smythe, R. T. and H. M. Mahmoud, "A Survey of Recursive Trees", *Theory of Probability and Mathematical Statistics*, Vol. 51, pp. 1–27, 1995.

2. Moon, J. W., "The distance between nodes in recursive trees", T. P. McDonough and V. C. Mavron (Editors), *London Mathematical Society Lecture Note Series 13: Combinatorics. Proceedings of the British Combinatorial Conference 1973*, pp. 125–132, Cambridge University Press, 1974.

3. Dobrow, R. P., "On the distribution of distances in recursive trees", *Journal of Applied Probability*, Vol. 33, No. 3, pp. 749–757, September 1996.

4. Su, C., J. Liu and Q. Feng, "A note on the distance in random recursive trees", *Statistics and Probability Letters*, Vol. 76, pp. 1748–1755, 2006.

5. Devroye, L., "Applications of the Theory of Records in the Study of Random Trees", *Acta Informatica*, Vol. 26, pp. 123–130, 1988.

6. Feng, Q., C. Su and Z. Hu, "Branching structure of uniform recursive trees", *Science in Chine Series A: Mathematics*, Vol. 48, No. 6, pp. 769–784, 2005.

7. Na, H. S. and A. Rapoport, "Distribution of Nodes of a Tree by Degree", *Mathematical Biosciences*, Vol. 6, pp. 313–329, 1970.

8. Gastwirth, J. L., "A Probability Model of a Pyramid Scheme", *The American Statistician*, Vol. 31, No. 2, pp. 79–82, MAY 1977.

9. Friedman, B., "A simple urn model", *Communications on Pure and Applied Mathematics*, Vol. 2, No. 1, pp. 59–70, March 1949.

10. Mahmoud, H. M. and R. T. Smythe, "Asymptotic Joint Normality of Outdegrees

of Nodes in Random Recursive Trees", *Random Structures & Algorithms*, Vol. 3, No. 3, pp. 255–266, 1992.

11. Najock, D. and C. C. Heyde, "On the Number of Terminal Vertices in Certain Random Trees with an Application to Stemma Construction in Philology", *Journal of Applied Probability*, Vol. 19, No. 3, pp. 675–680, September 1982.

12. Altok, S. and Ü. Işlak, "On leaf related statistics in recursive tree models", *Statistics and Probability Letters*, Vol. 121, pp. 61–69, February 2017.

13. Zhang, Y., "On the number of leaves in a random recursive tree", *Brazilian Journal of Probability and Statistics*, Vol. 29, No. 4, pp. 897–908, 2015.

14. Devroye, L., "Limit Laws for Local Counters in Random Binary Search Trees", *Random Structures & Algorithms*, Vol. 2, No. 3, pp. 303–315, 1991.

15. Devroye, L. and J. Lu, "The Strong Convergence of Maximal Degrees in Uniform Random Recursive Trees and Dags", *Random Structures & Algorithms*, Vol. 7, No. 1, pp. 1–14, 1995.

16. Goh, W. and E. Schmutz, "Limit distribution for the maximum degree of a random recursive tree", *Journal of Computational and Applied Mathematics*, Vol. 142, pp. 61–82, 2002.

17. Addario-Berry, L. and L. Eslava, "High Degrees in Recursive Trees", *Electronic Notes in Discrete Mathematics*, Vol. 50, pp. 451–456, 2015.

18. Eslava, L., "Depth of Vertices with High Degrees in Random Recursive Trees", *eprint arXiv:1611.07466*, November 2016.

19. Berestycki, N., "Recent progress in coalescent theory", *Ensaios Matemáticos*, Vol. 16, pp. 1–193, 2009.

20. Pittel, B., "Note on the Heights of Random Recursive Trees and Random m-ary

Search Trees", *Random Structures & Algorithms*, Vol. 5, No. 2, pp. 337–347, 1994.

21. Devroye, L., O. Fawzi and N. Fraiman, "Depth Properties of Scaled Attachment Random Recursive Trees", *Random Structures & Algorithms*, Vol. 41, No. 1, pp. 66–98, 2011.

22. Mahmoud, H. and R. T. Smythe, "On the distribution of leaves in rooted subtrees of recursive trees", *The Annals of Applied Probability*, Vol. 1, No. 3, pp. 406–418, 1991.

23. Marzouk, C., "Fires on large recursive trees", *Stochastic Processes and their Application*, Vol. 265-289, 2016.

24. Bergeron, F., P. Flajolet and B. Salvy, "Varieties of Increasing Trees", *17th Colloquium on Trees in Algebra and Programming*, pp. 24–48, Rennes, France, 1992.

25. Drmota, M., *Random Trees. An Interplay between Combinatorics and Probability*, Springer, Vienna, 2009.

26. Dobrow, R. P. and R. T. Smythe, "Poisson approximations for functionals of random trees", *Random Structures & Algorithms*, Vol. 9, No. 12, pp. 79 – 92, August 1996.

27. Szymanśki, J., "On a nonuniform random recursive tree", M. Karoński and Z. Palka (Editors), *Random Graphs '85 (Annals of Discrete Mathematics 33)*, pp. 297–306, Elsevier, Amsterdam, 1987.

28. Leckey, K. and R. Neininger, "Asymptotic analysis of Hoppe trees", *eprint arxiv:1202.2439*, 2012.

29. West, D. B., *Introduction to Graph Theory*, Pearson Education, Delhi, 2nd edn., 2001.

30. Bollobás, B., *Modern Graph Theory*, Springer, New York, 1998.

31. Golumbic, M. C., *Algorithmic Graph Theory and Perfect Graphs*, Elsevier, Amsterdam, 2nd edn., 2004.

32. Diestel, R., *Graph Theory*, Springer, Heidelberg, New York, 3rd (electronic) edn., 2005.

33. Işlak, Ü., *Combinatorial Analysis*, Boğaziçi University, İstanbul, 2017, `https://sites.google.com/site/umitislak/math-361-summer-2017`, accessed at August 2017.

34. Stanley, R. P., *Enumerative Combinatorics*, Vol. 1, Cambridge University Press., 2nd (online version) edn., 2011.

35. Nevzorov, V. B., *Records: Mathematical Theory*, American Mathematical Society, Providence, Rhode Island, 2001.

36. Işlak, Ü., *Basic Theory of Probability and Statistics*, Boğaziçi University, 2017, `https://sites.google.com/site/umitislak/math-345-summer-2017`, accessed at August 2017.

37. DeGroot, M. H. and M. J. Schervish, *Probability and Statistics*, Addison-Wesley, Boston, 4th edn., 2012.

38. Ross, N., "Fundamentals of Stein's Method", *Probability Surveys*, Vol. 8, pp. 210–293, 2011.

39. Levin, D. A., Y. Peres and E. L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, Providence, Rhode Island, 2nd (online) edn., 2008.

40. Stein, C., "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables", *Proceedings of the Sixth Berkeley Symposion on Mathematical Statistics and Probabilityory*, Vol. II: Probability Theory, pp. 583–602, University of California Press, 1972.

41. Ash, R. B., *Probability and Measure Theory*, Academic Press, San Diego, California, 2nd edn., 2000.

42. Barbour, A. D., L. Holst and S. Janson, *Poisson Approximation*, Clarendon Press, Oxford, 1992.

43. Chen, L. H., L. Goldstein and Q.-M. Shao, *Normal Approximation by Stein's Method*, Springer, Berlin Heidelberg, 2010.

44. McDiarmid, C., "Concentration", M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed (Editors), *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248, Springer, Berlin Heidelberg, 1998.

45. Glick, N., "Breaking Records and Breaking Boards", *The American Mathematical Monthly*, Vol. 85, No. 1, pp. 2–26, January 1978.

46. Addario-Berry, L. and K. Ford, "Poisson-Dirichlet Branching Random Walks", *The Annals of Applied Probability*, Vol. 23, No. 1, pp. 283–307, 2013.

47. R. M. D'Souza, P. L. Krapivsky and C. Moore, "The power of choice in growing trees", *The European Physical Journal B*, Vol. 59, pp. 535–543, 2007.

48. Mahmoud, H. M., "The Power of Choice in the Construction of Recursive Trees", *Methodology and Computing in Applied Probability*, Vol. 12, pp. 763–773, 2010.

49. Balińska, K. T., L. V. Quintas and J. Szymanśki, "Random recursive forests", *Random Structures & Algorithms*, Vol. 5, No. 1, pp. 3–12, 1994.

50. Devroye, L. and S. Janson, "Long and short paths in uniform random recursive dags", *Arkiv för Matematik*, Vol. 49, pp. 61–77, 2011.

51. Arratia, R., A. Barbour and S. Tavaré, *Logarithmic Combinatorial Structures: a Probabilistic Approach*, European Mathematical Society, Zürich, 2003.

52. Leckey, K., *Asymptotische Eigenschaften von Hoppe-Bäumen*, Master's Thesis, J. W. Goethe Universität, 2011/2012.

53. Kingman, J., "The population structure associated with the Ewens sampling formula", *Theoretical Population Biology*, Vol. 11, No. 2, pp. 274 – 283, 1977.

54. Watterson, G., "The stationary distribution of the infinitely-many neutral alleles diffusion model", *Journal of Applied Probability*, Vol. 13, No. 4, pp. 639–651, 1976.

55. Bayer, D. and P. Diaconis, "Trailing the dovetail shuffle to its lair", *The Annals of Applied Probability*, Vol. 2, No. 2, pp. 294–313, 1992.

56. Fulman, J., "The combinatorics of biased riffle shuffles", *Combinatorica*, Vol. 18, No. 2, pp. 173–184, 1998.

57. Gouet, R., F. J. López and G. Sanz, "Central Limit Theorems for the Number of Records in Discrete Models", *Advances in Applied Probability*, Vol. 37, No. 3, pp. 781–800, September 2005.

58. Bai, Z.-D., H.-K. Hwang and W.-Q. Liang, "Normal Approximations of the Number of Records in Geometrically Distributed Random Variables", *Random Structures & Algorithms*, Vol. 13, No. 3-4, pp. 319–334, 1998.

59. Prodinger, H., C. Schneider and S. Wagner, "Unfair permutations", *European Journal of Combinatorics*, Vol. 32, pp. 1282–1298, 2011.

60. Arslan, İ., Ü. Işlak and C. Pehlivan, "On Unfair Permutations", *arxiv preprint arXiv:1611.07275*, 2016.

61. Wilf, H. S., *generatingfunctionology*, Academic Press, internet edn., 1994, `https://www.math.upenn.edu/ wilf/gfology2.pdf`, accessed at August 2017.

APPENDIX A: PROOFS FROM CHAPTER 5

Since many of the calculations in the proofs of Chapter 5 are long, we give them in this appendix.

A.1. Proof of Theorem 5.9

Proof. Let $\mathcal{T}_n^p$ be a BRT of size n and γ be its permutation representation. Let moreover $\mathcal{B}_n^p$ denote the number of branches of $\mathcal{T}_n^p$. For the calculation of the variance difficulty arises from the dependence of the events $A_i = \mathbf{1}(\gamma(i) \text{ is an anti-record})$. We will again assume $p_1 \neq 0$.

We can write the variance as follows:

$$
\begin{aligned}
\mathrm{Var}(\mathcal{B}_n^p) = \mathrm{Var}\left(1 + \sum_{i=3}^{n} A_i\right) = \mathrm{Var}\left(\sum_{i=3}^{n} A_i\right) \\
= \sum_{i=3}^{n} \mathrm{Var}(A_i) + 2 \sum_{3 \leq i < j \leq n} \mathrm{Cov}(A_i, A_j).
\end{aligned}
\tag{A.1}
$$

Moreover we have

$$
\sum_{i=3}^{n} \mathrm{Var}(A_i) = \sum_{i=3}^{n} \mathbb{E}[A_i^2] - \sum_{i=3}^{n} \mathbb{E}[A_i]^2 = \mathbb{E}[\mathcal{B}_n^p] - 1 - \sum_{i=3}^{n} \mathbb{E}[A_i]^2
\tag{A.2}
$$

and

$$
\sum_{3 \leq i < j \leq n} \mathrm{Cov}(A_i, A_j) = \sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[A_i A_j] - \sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[A_i]\mathbb{E}[A_j].
\tag{A.3}
$$

Since the A_i are Bernoulli random variables, we have

$$
\begin{aligned}
\sum_{i=3}^{n} & \mathbb{E}[A_i]^2 \\
&= \sum_{i=3}^{n} \left(\sum_{s=1}^{a-1} p_s \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} \right)^2 \\
&= \sum_{i=3}^{n} \sum_{s=1}^{a-1} p_s^2 \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{2(i-2)} \\
&\quad + 2 \sum_{i=3}^{n} \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{i-2} \\
&= \sum_{s=1}^{a-1} p_s^2 \sum_{i=3}^{n} \left(\left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \right)^{i-2} \\
&\quad + 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \sum_{i=3}^{n} \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{i-2} \\
&= \sum_{s=1}^{a-1} p_s^2 \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \sum_{i=0}^{n-3} \left(\left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \right)^{i} \\
&\quad + 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right) \sum_{i=0}^{n-3} \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{i} \\
&= \sum_{s=1}^{a-1} p_s^2 \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{2(n-2)}}{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2} \\
&\quad + 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right) \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{n-2}}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}.
\end{aligned}
\tag{A.4}
$$

Now let $3 \leq i < j \leq n$, then

$$
\begin{aligned}
\mathbb{E}[A_i A_j] &= \mathbb{P}\left(X_i < \min\{X_2, \ldots, X_{i-1}\}, X_j < \min\{X_2, \ldots, X_{j-1}\}\right) \\
&= \mathbb{P}\left(X_j < \min\{X_2, \ldots, X_{j-1}\} | X_i < \min\{X_2, \ldots, X_{i-1}\}\right) \\
&\quad \cdot \mathbb{P}\left(X_i < \min\{X_2, \ldots, X_{i-1}\}\right) \\
&= \sum_{s=1}^{a-1} \mathbb{P}\left(X_j < \min_{k=2,\ldots,j-1}\{X_k\} | X_i < \min_{k=2,\ldots,i-1}\{X_k\}, X_i = s\right) \\
&\quad \cdot \mathbb{P}\left(X_i < \min_{k=2,\ldots,i-1}\{X_k\} | X_i = s\right) \mathbb{P}\left(X_i = s\right).
\end{aligned}
\tag{A.5}
$$

If $X_i = 1$, $X_j < X_i$ is not possible, so the above expression is only positive for $2 \leq s \leq a - 1$. For these s we have:

$$
\begin{aligned}
&\mathbb{P}\left(X_j < \min_{k=2,\ldots,j-1}\{X_k\} | X_i < \min_{k=2,\ldots,i-1}\{X_k\}, X_i = s\right) \\
&= \sum_{r=1}^{s-1} \underbrace{\mathbb{P}\left(X_j < \min_{k=2,\ldots,j-1}\{X_k\} | X_j = r, X_i = s, X_i < \min_{k=2,\ldots,i-1}\{X_k\}\right)}_{=\mathbb{P}(r < X_{i+1}, \ldots, X_{j-1})} \mathbb{P}(X_j = r) \\
&= \sum_{r=1}^{s-1} \left(\sum_{q=r+1}^{a} p_q\right)^{j-i-1} p_r.
\end{aligned}
\tag{A.6}
$$

So in total we have for $3 \leq i < j \leq n$,

$$
\mathbb{E}[A_i A_j] = \sum_{s=2}^{a-1} p_s \left(\sum_{\ell=s+1}^{a} p_\ell\right)^{i-2} \sum_{r=1}^{s-1} p_r \left(\sum_{q=r+1}^{a} p_q\right)^{j-i-1}.
\tag{A.7}
$$

Thus when summing over i and j we get

$$
\sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[A_i A_j]
$$

$$
= \sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \sum_{s=2}^{a-1} p_s \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} \sum_{r=1}^{s-1} p_r \left(\sum_{q=r+1}^{a} p_q \right)^{j-i-1}
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} p_r \sum_{i=3}^{n-1} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} \sum_{j=i+1}^{n} \left(\sum_{q=r+1}^{a} p_q \right)^{j-i-1}
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} p_r \sum_{i=3}^{n-1} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} \sum_{j=0}^{n-i-1} \left(\sum_{q=r+1}^{a} p_q \right)^{j}
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} p_r \sum_{i=1}^{n-3} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \sum_{j=0}^{n-i-3} \left(\sum_{q=r+1}^{a} p_q \right)^{j}
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} p_r \sum_{i=1}^{n-3} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \frac{1 - \left(\sum_{q=r+1}^{a} p_q \right)^{n-i-2}}{1 - \sum_{q=r+1}^{a} p_q}
$$

$$
\quad (A.8)
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \sum_{i=1}^{n-3} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i}
$$

$$
\quad - \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \sum_{i=1}^{n-3} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \left(\sum_{q=r+1}^{a} p_q \right)^{n-i-2}
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \frac{\sum_{\ell=s+1}^{a} p_\ell - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-2}}{1 - \sum_{\ell=s+1}^{a} p_\ell}
$$

$$
\quad - \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \left(\sum_{q=r+1}^{a} p_q \right)^{n-2} \sum_{i=1}^{n-3} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \left(\sum_{q=r+1}^{a} p_q \right)^{-i}
$$

$$
= \sum_{s=2}^{a-1} \frac{p_s \sum_{\ell=s+1}^{a} p_\ell}{\sum_{\ell=1}^{s} p_\ell} \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right)
$$

$$
\quad - \sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \left(\sum_{q=r+1}^{a} p_q \right)^{n-2} \left(\frac{\frac{\sum_{\ell=s+1}^{a} p_\ell}{\sum_{q=r+1}^{a} p_q} - \left(\frac{\sum_{\ell=s+1}^{a} p_\ell}{\sum_{q=r+1}^{a} p_q} \right)^{n-2}}{1 - \frac{\sum_{\ell=s+1}^{a} p_\ell}{\sum_{q=r+1}^{a} p_q}} \right).
$$

The last term can be simplified a little bit more, such that we get

$$
\sum_{s=2}^{a-1} p_s \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \left(\sum_{q=r+1}^{a} p_q\right)^{n-2} \left(\frac{\frac{\sum_{\ell=s+1}^{a} p_\ell}{\sum_{q=r+1}^{a} p_q} - \left(\frac{\sum_{\ell=s+1}^{a} p_\ell}{\sum_{q=r+1}^{a} p_q}\right)^{n-2}}{1 - \frac{\sum_{\ell=s+1}^{a} p_\ell}{\sum_{q=r+1}^{a} p_q}}\right)
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{s-1} \frac{p_r \sum_{q=r+1}^{a} p_q}{\sum_{q=1}^{r} p_q} \left(\frac{\left(\sum_{q=r+1}^{a} p_q\right)^{n-3} - \left(\sum_{\ell=s+1}^{a} p_\ell\right)^{n-3}}{\sum_{q=r+1}^{a} p_q - \sum_{\ell=s+1}^{a} p_\ell}\right) \tag{A.9}
$$

$$
= \sum_{s=2}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{s-1} \frac{p_r \sum_{q=r+1}^{a} p_q}{\sum_{q=1}^{r} p_q \sum_{q=r+1}^{s} p_q} \left(\left(\sum_{q=r+1}^{a} p_q\right)^{n-3} - \left(\sum_{\ell=s+1}^{a} p_\ell\right)^{n-3}\right).
$$

So

$$
\sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[A_i A_j] = \sum_{s=2}^{a-1} \frac{p_s \sum_{\ell=s+1}^{a} p_\ell}{\sum_{\ell=1}^{s} p_\ell} \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell\right)^{n-3}\right)
$$

$$
- \sum_{s=2}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{s-1} \frac{p_r \sum_{q=r+1}^{a} p_q}{\sum_{q=1}^{r} p_q} \frac{1}{\sum_{q=r+1}^{s} p_q} \tag{A.10}
$$

$$
\cdot \left(\left(\sum_{q=r+1}^{a} p_q\right)^{n-3} - \left(\sum_{\ell=s+1}^{a} p_\ell\right)^{n-3}\right).
$$

Moreover we have

$$\sum_{3 \le i < j \le n} \mathbb{E}[A_i]\mathbb{E}[A_j]$$

$$= \sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \sum_{s=1}^{a-1} p_s \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} \sum_{r=1}^{a-1} p_r \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{j-2}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{i=3}^{n-1} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-2} \sum_{r=1}^{a-1} p_r \sum_{j=i+1}^{n} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{j-2}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{i=3}^{n-1} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i-3} \sum_{r=1}^{a-1} p_r \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{i-1} \sum_{j=0}^{n-i-1} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{j}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} p_r \sum_{i=0}^{n-4} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{i+2} \sum_{j=0}^{n-i-4} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{j}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} p_r \sum_{i=0}^{n-4} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{i+2} \frac{1 - \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{n-i-3}}{1 - \sum_{\ell=r+1}^{a} p_\ell}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \sum_{i=0}^{n-4} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{i+2}$$

$$\quad - \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \sum_{i=0}^{n-4} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{i+2} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{n-i-3} \quad \text{(A.11)}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{2} \sum_{i=0}^{n-4} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{i}$$

$$\quad - \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{n-1} \sum_{i=0}^{n-4} \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{i}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{2} \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{n-3}}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}$$

$$\quad - \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{n-1} \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3}}{1 - \sum_{\ell=s+1}^{a} p_\ell}$$

$$= \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{2} \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{n-3}}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}$$

$$\quad - \sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{n-1} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right).$$

This gives in total

$$\mathrm{Var}(\mathcal{B}_n^p) =$$

$$\sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-1} \right) - \sum_{s=1}^{a-1} p_s$$

$$- \sum_{s=1}^{a-1} p_s^2 \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{2(n-2)}}{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2}$$

$$- 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right) \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{n-2}}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}$$

$$+ 2 \sum_{s=2}^{a-1} \frac{p_s \sum_{\ell=s+1}^{a} p_\ell}{\sum_{\ell=1}^{s} p_\ell} \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right) \tag{A.12}$$

$$- 2 \sum_{s=2}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{s-1} \frac{p_r \sum_{q=r+1}^{a} p_q}{\sum_{q=1}^{r} p_q} \frac{1}{\sum_{q=r+1}^{s} p_q}$$

$$\cdot \left(\left(\sum_{q=r+1}^{a} p_q \right)^{n-3} - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right)$$

$$- 2 \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^2 \frac{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right)^{n-3}}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}$$

$$+ 2 \sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^{n-1} \left(1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^{n-3} \right).$$

Hence we get asymptotically, for all $\vec{p}$ with $p_1 > 0$,

$$
\begin{aligned}
\operatorname{Var}(\mathcal{B}_n^p) \xrightarrow{n \to \infty} \\
\sum_{s=1}^{a-1} \frac{p_s}{\sum_{\ell=1}^{s} p_\ell} - \sum_{s=1}^{a-1} p_s \\
- \sum_{s=1}^{a-1} p_s^2 \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2 \frac{1}{1 - \left(\sum_{\ell=s+1}^{a} p_\ell \right)^2} \\
- 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} p_s p_r \left(\sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell \right) \frac{1}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell} \\
+ 2 \sum_{s=2}^{a-1} \frac{p_s \sum_{\ell=s+1}^{a} p_\ell}{\sum_{\ell=1}^{s} p_\ell} \sum_{r=1}^{s-1} \frac{p_r}{\sum_{q=1}^{r} p_q} \\
- 2 \sum_{s=1}^{a-1} p_s \sum_{\ell=s+1}^{a} p_\ell \sum_{r=1}^{a-1} \frac{p_r}{\sum_{\ell=1}^{r} p_\ell} \left(\sum_{\ell=r+1}^{a} p_\ell \right)^2 \frac{1}{1 - \sum_{\ell=s+1}^{a} p_\ell \sum_{\ell=r+1}^{a} p_\ell}.
\end{aligned}
\tag{A.13}
$$

$\square$

A.2. Proof of Theorem 5.10

Proof. If $\vec{p}$ is the uniform distribution over $[a]$, (5.25) gives

$$\mathrm{Var}(\mathcal{B}_n^a) =$$

$$\sum_{s=1}^{a-1} \frac{\frac{1}{a}}{\sum_{\ell=1}^{s} \frac{1}{a}} \left(1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a}\right)^{n-1}\right) - \sum_{s=1}^{a-1} \frac{1}{a}$$

$$- \sum_{s=1}^{a-1} \frac{1}{a^2} \left(\sum_{\ell=s+1}^{a} \frac{1}{a}\right)^2 \frac{1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a}\right)^{2(n-2)}}{1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a}\right)^2}$$

$$- 2\sum_{s=2}^{a-1} \sum_{r=1}^{s-1} \frac{1}{a^2} \left(\sum_{\ell=s+1}^{a} \frac{1}{a} \sum_{\ell=r+1}^{a} \frac{1}{a}\right) \frac{1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a} \sum_{\ell=r+1}^{a} \frac{1}{a}\right)^{n-2}}{1 - \sum_{\ell=s+1}^{a} \frac{1}{a} \sum_{\ell=r+1}^{a} \frac{1}{a}}$$

$$+ 2\sum_{s=2}^{a-1} \frac{\frac{1}{a}\sum_{\ell=s+1}^{a} \frac{1}{a}}{\sum_{\ell=1}^{s} \frac{1}{a}} \sum_{r=1}^{s-1} \frac{\frac{1}{a}}{\sum_{q=1}^{r} \frac{1}{a}} \left(1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a}\right)^{n-3}\right) \qquad \text{(A.14)}$$

$$- 2\sum_{s=2}^{a-1} \frac{1}{a^2} \sum_{\ell=s+1}^{a} \sum_{r=1}^{s-1} \frac{\frac{1}{a}\sum_{q=r+1}^{a} \frac{1}{a}}{\sum_{q=1}^{r} \frac{1}{a}} \frac{1}{\sum_{q=r+1}^{s} \frac{1}{a}}$$

$$\cdot \left(\left(\sum_{q=r+1}^{a} \frac{1}{a}\right)^{n-3} - \left(\sum_{\ell=s+1}^{a} \frac{1}{a}\right)^{n-3}\right)$$

$$- 2\sum_{s=1}^{a-1} \frac{1}{a^2} \sum_{\ell=s+1}^{a} \sum_{r=1}^{a-1} \frac{\frac{1}{a}}{\sum_{\ell=1}^{r} \frac{1}{a}} \left(\sum_{\ell=r+1}^{a} \frac{1}{a}\right)^2 \frac{1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a} \sum_{\ell=r+1}^{a} \frac{1}{a}\right)^{n-3}}{1 - \sum_{\ell=s+1}^{a} \frac{1}{a} \sum_{\ell=r+1}^{a} \frac{1}{a}}$$

$$+ 2\sum_{s=1}^{a-1} \frac{\frac{1}{a}}{\sum_{\ell=1}^{s} \frac{1}{a}} \sum_{\ell=s+1}^{a} \frac{1}{a} \sum_{r=1}^{a-1} \frac{\frac{1}{a}}{\sum_{\ell=1}^{r} \frac{1}{a}} \left(\sum_{\ell=r+1}^{a} \frac{1}{a}\right)^{n-1} \left(1 - \left(\sum_{\ell=s+1}^{a} \frac{1}{a}\right)^{n-3}\right).$$

This gives after some simplifications

$$
\begin{aligned}
\mathrm{Var}(\mathcal{B}_n^a) = \quad & \sum_{s=1}^{a-1} \frac{1}{s}\left(1 - \left(\frac{a-s}{a}\right)^{n-1}\right) - \frac{a-1}{a} \\
& - \sum_{s=1}^{a-1} \frac{1}{a^2}\left(\frac{a-s}{a}\right)^2 \frac{1 - \left(\frac{a-s}{a}\right)^{2(n-2)}}{1 - \left(\frac{a-s}{a}\right)^2} \\
& - 2\sum_{s=2}^{a-1}\sum_{r=1}^{s-1} \frac{1}{a^2}\frac{a-s}{a}\frac{a-r}{a}\frac{1 - \left(\frac{a-s}{a}\frac{a-r}{a}\right)^{n-2}}{1 - \frac{a-s}{a}\frac{a-r}{a}} \\
& + 2\sum_{s=2}^{a-1} \frac{a-s}{as}\sum_{r=1}^{s-1}\frac{1}{r}\left(1 - \left(\frac{a-s}{a}\right)^{n-3}\right) \\
& - 2\sum_{s=2}^{a-1} \frac{a-s}{a^2}\sum_{r=1}^{s-1}\frac{a-r}{r}\frac{1}{s-r}\left(\left(\frac{a-r}{a}\right)^{n-3} - \left(\frac{a-s}{a}\right)^{n-3}\right) \\
& - 2\sum_{s=1}^{a-1} \frac{a-s}{a^2}\sum_{r=1}^{a-1}\frac{1}{r}\left(\frac{a-r}{a}\right)^2\frac{1 - \left(\frac{a-s}{a}\frac{a-r}{a}\right)^{n-3}}{1 - \frac{a-s}{a}\frac{a-r}{a}} \\
& + 2\sum_{s=1}^{a-1} \frac{a-s}{sa}\sum_{r=1}^{a-1}\frac{1}{r}\left(\frac{a-r}{a}\right)^{n-1}\left(1 - \left(\frac{a-s}{a}\right)^{n-3}\right).
\end{aligned}
\tag{A.15}
$$

We can easily derive the asymptotic term for $n \to \infty$:

$$
\begin{aligned}
\operatorname{Var}(\mathcal{B}_n^a) \xrightarrow{n \to \infty} \ & \sum_{s=1}^{a} \frac{1}{s} - \frac{a-1}{a} - \sum_{s=1}^{a-1} \frac{1}{a^2} \left(\frac{a-s}{a} \right)^2 \frac{1}{1 - \left(\frac{a-s}{a} \right)^2} \\
& - 2 \sum_{s=2}^{a-1} \sum_{r=1}^{s-1} \frac{1}{a^2} \frac{a-s}{a} \frac{a-r}{a} \frac{1}{1 - \frac{a-s}{a} \frac{a-r}{a}} \\
& + 2 \sum_{s=2}^{a-1} \frac{a-s}{as} \sum_{r=1}^{s-1} \frac{1}{r} \\
& - 2 \sum_{s=1}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{a-1} \frac{1}{r} \left(\frac{a-r}{a} \right)^2 \frac{1}{1 - \frac{a-s}{a} \frac{a-r}{a}} \\
= \ & H_a - \frac{a-1}{a} + \frac{2}{a} \sum_{s=2}^{a-1} \frac{a-s}{s} \sum_{r=1}^{s-1} \frac{1}{r} \\
& - \frac{1}{a^2} \sum_{s=1}^{a-1} \frac{s^2}{a^2 - s^2} \\
& - \frac{2}{a^2} \sum_{s=1}^{a-2} \sum_{r=1}^{a-s+1} \frac{sr}{a^2 - sr} \\
& - \frac{2}{a^2} \sum_{s=1}^{a-1} \sum_{r=1}^{a-1} \frac{1}{r} \frac{s(a-r)^2}{a^2 - s(a-r)}.
\end{aligned}
\tag{A.16}
$$

We now look at the terms in Equation A.15 separately to get an asymptotic result for $a \to \infty$. First of all we know that the first line $+1$ is equal to the expectation of the number of branches, for which we already have an asymptotic result, namely that it is asymptotically equal to H_{n-1}. Again the following equation will repeatedly be used in our calculations:

$$
\sum_{s=1}^{a} s^k = \frac{a^{k+1}}{k+1} + \mathcal{O}(a^k) \text{ for all } k \in \mathbb{N}, k \geq 0.
\tag{A.17}
$$

For the second term we will now show that it converges to 0 as a goes to infinity:

$$
\begin{aligned}
\sum_{s=1}^{a-1} \frac{1}{a^2} & \left(\frac{a-s}{a}\right)^2 \frac{1-\left(\frac{a-s}{a}\right)^{2(n-2)}}{1-\left(\frac{a-s}{a}\right)^2} \\
&= \sum_{s=1}^{a-1} \frac{1}{a^2} \left(\frac{s}{a}\right)^2 \frac{1-\left(\frac{s}{a}\right)^{2(n-2)}}{1-\left(\frac{s}{a}\right)^2} \\
&= \sum_{s=1}^{a-1} \frac{1}{a^2} \left(\frac{s}{a}\right)^2 \sum_{\ell=0}^{n-3} \left(\frac{s}{a}\right)^{2\ell} \\
&= \sum_{\ell=0}^{n-3} \frac{1}{a^{2\ell+4}} \sum_{s=1}^{a-1} s^{2\ell+2} \\
&= \sum_{\ell=0}^{n-3} \frac{1}{a^{2\ell+4}} \mathcal{O}(a^{2\ell+3}) \\
&\xrightarrow{a\to\infty} 0.
\end{aligned}
$$
(A.18)

For the next term we will show that it converges to the second harmonic number plus a constant:

$$
\begin{aligned}
\sum_{s=2}^{a-1}\sum_{r=1}^{s-1} & \frac{1}{a^2}\frac{a-s}{a}\frac{a-r}{a}\frac{1-\left(\frac{a-s}{a}\frac{a-r}{a}\right)^{n-2}}{1-\frac{a-s}{a}\frac{a-r}{a}} \\
&= \sum_{s=1}^{a-2}\sum_{r=s+1}^{a-1}\frac{1}{a^2}\frac{s}{a}\frac{r}{a}\frac{1-\left(\frac{s}{a}\frac{r}{a}\right)^{n-2}}{1-\frac{s}{a}\frac{r}{a}} \\
&= \sum_{s=1}^{a-2}\sum_{r=s+1}^{a-1}\frac{1}{a^2}\frac{s}{a}\frac{r}{a}\sum_{\ell=0}^{n-3}\left(\frac{s}{a}\frac{r}{a}\right)^{\ell} \\
&= \sum_{\ell=0}^{n-3}\frac{1}{a^{2\ell+4}}\sum_{s=1}^{a-2}s^{\ell+1}\sum_{r=s+1}^{a-1}r^{\ell+1} \\
&= \sum_{\ell=0}^{n-3}\frac{1}{a^{2\ell+4}}\sum_{s=1}^{a-2}s^{\ell+1}\left[\frac{a^{\ell+2}}{\ell+2}+\mathcal{O}(a^{\ell+1})-\frac{s^{\ell+2}}{\ell+2}+\mathcal{O}(s^{\ell+1})\right] \\
&= \sum_{\ell=0}^{n-3}\frac{1}{a^{2\ell+4}}\left[a^{\ell+2}\sum_{s=1}^{a-2}\frac{s^{\ell+1}}{\ell+2}-\sum_{s=1}^{a-2}\frac{s^{2\ell+3}}{\ell+2}+\mathcal{O}(a^{\ell+1})\sum_{s=1}^{a-2}s^{\ell+1}+\sum_{s=1}^{a-2}\mathcal{O}(s^{2\ell+2})\right] \\
&= \sum_{\ell=0}^{n-3}\frac{1}{a^{2\ell+4}}\left[\frac{a^{2\ell+4}}{(\ell+2)^2}-\frac{a^{2\ell+4}}{(2\ell+4)(\ell+2)}+\mathcal{O}(a^{2\ell+3})\right] \\
&= \sum_{\ell=0}^{n-3}\left[\frac{1}{(\ell+2)^2}-\frac{1}{2(\ell+2)^2}\right]+\mathcal{O}\left(\frac{1}{a}\right) \\
&= \frac{1}{2}\left[\sum_{\ell=1}^{n-1}\frac{1}{\ell^2}-1\right]+\mathcal{O}\left(\frac{1}{a}\right) \\
&\xrightarrow{a\to\infty}\frac{1}{2}H_{n-1}^{(2)}-\frac{1}{2}.
\end{aligned}
$$

(A.19)

We will now show that the 4th and the 6th term, and the 5th and the 7th term of Equation A.15 are asymptotically equal. We will repeatedly need the following lemma in these calculations.

Lemma A.1. *Let $k,n\in\mathbb{N}$. Then*

$$
\sum_{k=1}^{n-1}(-1)^{k+1}\binom{n-1}{k}\frac{1}{k}=\sum_{k=1}^{n-1}\frac{1}{k}.
$$

(A.20)

Proof. We will prove this identity by the use of generating functions. First we calculate the generating function for $\sum_{k=1}^{n-1}(-1)^{k+1}\binom{n-1}{k}\frac{1}{k}$:

$$
\begin{aligned}
\sum_{n=1}^{\infty}\sum_{k=1}^{n-1}(-1)^{k+1}\binom{n-1}{k}\frac{1}{k}x^{n-1} &= \sum_{n=0}^{\infty}\sum_{k=1}^{n}(-1)^{k+1}\binom{n}{k}\frac{1}{k}x^{n} \\
&= \sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}\sum_{n=k}^{\infty}\binom{n}{k}x^{n} \\
&= \sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}\sum_{n=0}^{\infty}\binom{n+k}{k}x^{n+k} \\
&= \sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}x^{k}\sum_{n=0}^{\infty}\binom{n+k}{k}x^{n} \\
&= \sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}x^{k}\frac{1}{(1-x)^{k+1}} \\
&= \frac{1}{1-x}\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k}\left(\frac{x}{1-x}\right)^{k} \\
&= \frac{1}{1-x}\ln\left(1+\frac{x}{1-x}\right) \\
&= \frac{1}{1-x}\ln\left(\frac{1}{1-x}\right) \\
&= \frac{-\ln(1-x)}{1-x}.
\end{aligned}
\tag{A.21}
$$

Where we used the following two equalities:

$$
\sum_{n=0}^{\infty}\binom{n+k}{k}x^{n} = \frac{1}{(1-x)^{k+1}}
\tag{A.22}
$$

from [61] and the Taylor series for the logarithm:

$$
\ln(1+x) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{x^{k}}{k} \text{ for } |x|<1.
\tag{A.23}
$$

The generating function for $\sum_{k=1}^{n-1} \frac{1}{k}$ is easier:

$$
\begin{aligned}
\sum_{n=1}^{\infty} \sum_{k=1}^{n-1} \frac{1}{k} x^{n-1} &= \sum_{n=0}^{\infty} \sum_{k=1}^{n} \frac{1}{k} x^{n} \\
&= \sum_{k=1}^{\infty} \frac{1}{k} \sum_{n=k}^{\infty} x^{n} \\
&= \sum_{k=1}^{\infty} \frac{x^{k}}{k} \sum_{n=0}^{\infty} x^{n} \\
&= \sum_{k=1}^{\infty} \frac{x^{k}}{k} \frac{1}{1-x} \\
&= -\frac{\ln(1-x)}{1-x}.
\end{aligned}
\tag{A.24}
$$

$\square$

We will now first consider the 4th and the 6th expression and show that they are asymptotically equal. For the 4th expression we get

$$
\begin{aligned}
\sum_{s=2}^{a-1} \frac{a-s}{as} \sum_{r=1}^{s-1} \frac{1}{r} & \left(1 - \left(\frac{a-s}{a}\right)^{n-3}\right) \\
&= \sum_{s=2}^{a-1} \frac{a-s}{as} \frac{s}{a} \frac{1 - \left(\frac{a-s}{a}\right)^{n-3}}{1 - \frac{a-s}{a}} \sum_{r=1}^{s-1} \frac{1}{r} \\
&= \sum_{s=2}^{a-1} \frac{a-s}{a^{2}} \frac{1 - \left(\frac{a-s}{a}\right)^{n-3}}{1 - \frac{a-s}{a}} \sum_{r=1}^{s-1} \frac{1}{r} \\
&= \sum_{r=1}^{a-2} \frac{1}{r} \sum_{s=r+1}^{a-1} \frac{a-s}{a^{2}} \frac{1 - \left(\frac{a-s}{a}\right)^{n-3}}{1 - \frac{a-s}{a}} \\
&= \sum_{r=1}^{a-2} \frac{1}{r} \sum_{s=1}^{a-r-1} \frac{s}{a^{2}} \frac{1 - \left(\frac{s}{a}\right)^{n-3}}{1 - \frac{s}{a}} \\
&= \sum_{r=1}^{a-2} \frac{1}{r} \sum_{s=1}^{a-r-1} \frac{s}{a^{2}} \sum_{\ell=0}^{n-4} \left(\frac{s}{a}\right)^{\ell} \\
&= \sum_{r=1}^{a-2} \frac{1}{r} \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \sum_{s=1}^{a-r-1} s^{\ell+1} \\
&= \sum_{r=1}^{a-2} \frac{1}{r} \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \left[\frac{(a-r)^{\ell+2}}{l+2} + \mathcal{O}((a-r)^{\ell+1})\right].
\end{aligned}
\tag{A.25}
$$

Now we have

$$\sum_{r=1}^{a-2} \frac{1}{r} \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \mathcal{O}((a-r)^{\ell+1}) = \mathcal{O}\left(\frac{ln(a)}{a}\right). \tag{A.26}$$

We continue with the rest of the expression:

$$
\begin{aligned}
&\sum_{r=1}^{a-2} \frac{1}{r} \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \frac{(a-r)^{\ell+2}}{l+2} \\
&= \sum_{r=1}^{a-2} \frac{1}{r} \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \frac{1}{l+2} \sum_{h=0}^{l+2} \binom{l+2}{h} a^{l+2-h}(-r)^h \\
&= \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \frac{1}{l+2} \sum_{h=0}^{l+2} \binom{l+2}{h} a^{l+2-h}(-1)^h \sum_{r=1}^{a-2} r^{h-1} \\
&= \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \frac{1}{l+2} \left[\sum_{h=1}^{l+2} \binom{l+2}{h} a^{l+2-h}(-1)^h \left(\frac{a^h}{h} + \mathcal{O}(a^{h-1})\right) + a^{\ell+2} H_{a-2}\right] \\
&= \sum_{\ell=0}^{n-4} \frac{1}{l+2} \sum_{h=1}^{l+2} \binom{l+2}{h}(-1)^h \frac{1}{h} + \sum_{\ell=0}^{n-4} \frac{1}{l+2} H_{a-2} + \mathcal{O}\left(\frac{1}{a}\right) \\
&= -\sum_{\ell=0}^{n-4} \frac{H_{\ell+2}}{l+2} + \sum_{\ell=0}^{n-4} \frac{1}{l+2} H_{a-2} + \mathcal{O}\left(\frac{1}{a}\right)
\end{aligned}
\tag{A.27}
$$

where we used Lemma A.1 in the last line.

Now we look at the 6th term of Equation A.15

$$\sum_{s=1}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{a-1} \frac{1}{r} \left(\frac{a-r}{a}\right)^2 \frac{1 - \left(\frac{a-s}{a}\frac{a-r}{a}\right)^{n-3}}{1 - \frac{a-s}{a}\frac{a-r}{a}}$$

$$= \sum_{s=1}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{a-1} \frac{1}{r} \left(\frac{a-r}{a}\right)^2 \sum_{\ell=0}^{n-4} \left(\frac{a-s}{a}\frac{a-r}{a}\right)^{\ell}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{2\ell+4}} \sum_{s=1}^{a-1} s^{\ell+1} \sum_{r=1}^{a-1} \frac{1}{r}(a-r)^{\ell+2}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{2\ell+4}} \left[\frac{a^{\ell+2}}{\ell+2} + \mathcal{O}(a^{\ell+1})\right] \sum_{r=1}^{a-1} \frac{1}{r} \sum_{h=0}^{l+2} \binom{\ell+2}{h} a^{\ell+2-h} r^h (-1)^h$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{2\ell+4}} \left[\frac{a^{\ell+2}}{\ell+2} + \mathcal{O}(a^{\ell+1})\right] \sum_{h=0}^{l+2} \binom{\ell+2}{h} (-1)^h a^{\ell+2-h} \sum_{r=1}^{a-1} r^{h-1}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{2\ell+4}} \left[\frac{a^{\ell+2}}{\ell+2} + \mathcal{O}(a^{\ell+1})\right] \tag{A.28}$$

$$\cdot \left\{\sum_{h=1}^{l+2} \binom{\ell+2}{h} (-1)^h a^{\ell+2-h} \left[\frac{a^h}{h} + \mathcal{O}(a^{h-1})\right] + a^{\ell+2} H_{a-1}\right\}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{2\ell+4}} \left[\frac{a^{\ell+2}}{\ell+2} + \mathcal{O}(a^{\ell+1})\right]$$

$$\cdot \left\{a^{l+2} \sum_{h=1}^{l+2} \binom{\ell+2}{h} (-1)^h \frac{1}{h} + \mathcal{O}\left(\frac{1}{a}\right) + a^{\ell+2} H_{a-1}\right\}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{\ell+2} \sum_{h=1}^{l+2} \binom{\ell+2}{h} (-1)^h \frac{1}{h} + \mathcal{O}\left(\frac{1}{a}\right) + \sum_{\ell=0}^{n-4} \frac{1}{l+2} H_{a-1}$$

$$= -\sum_{\ell=0}^{n-4} \frac{H_{\ell+2}}{\ell+2} + \mathcal{O}\left(\frac{1}{a}\right) + \sum_{\ell=0}^{n-4} \frac{1}{l+2} H_{a-1}.$$

Hence the sum of the 4th and the 6th term is asymptotically negligible:

$$2\sum_{\ell=0}^{n-4} \frac{1}{l+2} H_{a-2} - 2\sum_{\ell=0}^{n-4} \frac{1}{l+2} H_{a-1} + \mathcal{O}\left(\frac{\ln(a)}{a}\right)$$

$$= -2\sum_{\ell=0}^{n-4} \frac{1}{l+2} \frac{1}{a-1} + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \tag{A.29}$$

$$\xrightarrow{a\to\infty} 0.$$

Finally we show that the sum of the 5th and the 7th term is asymptotically negligible:

$$
\sum_{s=2}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{s-1} \frac{a-r}{r} \frac{1}{s-r} \left(\left(\frac{a-r}{a}\right)^{n-3} - \left(\frac{a-s}{a}\right)^{n-3} \right)
$$

$$
= \sum_{s=2}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{s-1} \frac{a-r}{r} \frac{1}{a-r} \left(\frac{a-r}{a}\right)^{n-3} \frac{1 - \left(\frac{a-s}{a-r}\right)^{n-3}}{1 - \frac{a-s}{a-r}}
$$

$$
= \sum_{s=2}^{a-1} \frac{a-s}{a^2} \sum_{r=1}^{s-1} \frac{1}{r} \left(\frac{a-r}{a}\right)^{n-3} \sum_{\ell=0}^{n-4} \left(\frac{a-s}{a-r}\right)^{\ell}
$$

$$
= \frac{1}{a^{n-1}} \sum_{\ell=0}^{n-4} \sum_{s=2}^{a-1} (a-s)^{\ell+1} \sum_{r=1}^{s-1} \frac{(a-r)^{n-3-l}}{r} \tag{A.30}
$$

$$
= \frac{1}{a^{n-1}} \sum_{\ell=0}^{n-4} \sum_{r=1}^{a-2} \frac{(a-r)^{n-3-l}}{r} \sum_{s=r+1}^{a-1} (a-s)^{\ell+1}
$$

$$
= \frac{1}{a^{n-1}} \sum_{\ell=0}^{n-4} \sum_{r=1}^{a-2} \frac{(a-r)^{n-3-l}}{r} \sum_{s=1}^{a-r-1} s^{\ell+1}
$$

$$
= \frac{1}{a^{n-1}} \sum_{\ell=0}^{n-4} \sum_{r=1}^{a-2} \frac{(a-r)^{n-3-l}}{r} \left[\frac{(a-r)^{\ell+2}}{\ell+2} + \mathcal{O}((a-r)^{\ell+1}) \right].
$$

Now we have

$$
\frac{1}{a^{n-1}} \sum_{\ell=0}^{n-4} \sum_{r=1}^{a-2} \frac{(a-r)^{n-3-l}}{r} \mathcal{O}((a-r)^{\ell+1}) = \mathcal{O}\left(\frac{\ln(a)}{a}\right). \tag{A.31}
$$

Continuing with the rest we get

$$\frac{1}{a^{n-1}}\sum_{\ell=0}^{n-4}\sum_{r=1}^{a-2}\frac{(a-r)^{n-3-l}}{r}\frac{(a-r)^{\ell+2}}{\ell+2}$$

$$=\frac{1}{a^{n-1}}\sum_{\ell=0}^{n-4}\frac{1}{\ell+2}\sum_{r=1}^{a-2}\frac{(a-r)^{n-1}}{r}$$

$$=\frac{1}{a^{n-1}}\sum_{\ell=0}^{n-4}\frac{1}{\ell+2}\sum_{r=1}^{a-2}\sum_{h=0}^{n-1}\binom{n-1}{h}a^{n-1-h}(-1)^h r^{h-1}$$

$$=\frac{1}{a^{n-1}}\sum_{\ell=0}^{n-4}\frac{1}{\ell+2}\sum_{h=0}^{n-1}\binom{n-1}{h}a^{n-1-h}(-1)^h\sum_{r=1}^{a-2}r^{h-1}$$

$$=\frac{1}{a^{n-1}}\sum_{\ell=0}^{n-4}\frac{1}{\ell+2}\left\{\sum_{h=1}^{n-1}\binom{n-1}{h}a^{n-1-h}(-1)^h\left[\frac{a^h}{h}+\mathcal{O}(a^{h-1})\right]\right.$$

$$\left.+a^{n-1}\sum_{r=1}^{a-2}r^{-1}\right\}$$ (A.32)

$$=\sum_{\ell=0}^{n-4}\frac{1}{\ell+2}\left\{\sum_{h=1}^{n-1}\binom{n-1}{h}(-1)^h\left[\frac{1}{h}+\mathcal{O}\left(\frac{1}{a}\right)\right]+H_{a-2}\right\}$$

$$=\sum_{\ell=0}^{n-4}\frac{1}{\ell+2}\left\{\sum_{h=1}^{n-1}-\frac{1}{h}+H_{a-2}\right\}+\mathcal{O}\left(\frac{1}{a}\right).$$

We again used Lemma A.1 in the last line. Finally we consider the last term of Equation A.15:

$$\sum_{s=1}^{a-1}\frac{a-s}{sa}\sum_{r=1}^{a-1}\frac{1}{r}\left(\frac{a-r}{a}\right)^{n-1}\left(1-\left(\frac{a-s}{a}\right)^{n-3}\right)$$

$$=\sum_{s=1}^{a-1}\frac{a-s}{sa}\frac{s}{a}\frac{1-\left(\frac{a-s}{a}\right)^{n-3}}{1-\frac{a-s}{a}}\sum_{r=1}^{a-1}\frac{1}{r}\left(\frac{a-r}{a}\right)^{n-1}.$$ (A.33)

We now consider the term dependent on s and the term dependent on r separately. This gives

$$\sum_{s=1}^{a-1} \frac{a-s}{sa} \frac{s}{a} \frac{1-\left(\frac{a-s}{a}\right)^{n-3}}{1-\frac{a-s}{a}} = \sum_{s=1}^{a-1} \frac{a-s}{a^2} \sum_{\ell=0}^{n-4} \left(\frac{a-s}{a}\right)^{\ell}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \sum_{s=1}^{a-1} (a-s)^{\ell+1}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \sum_{s=1}^{a-1} s^{\ell+1} \qquad (A.34)$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{a^{\ell+2}} \left[\frac{a^{\ell+2}}{\ell+2} + \mathcal{O}(a^{\ell+1}) \right]$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{\ell+2} + \mathcal{O}\left(\frac{1}{a}\right)$$

and

$$\sum_{r=1}^{a-1} \frac{1}{r} \left(\frac{a-r}{a}\right)^{n-1}$$

$$= \sum_{r=1}^{a-1} \frac{1}{a^{n-1}} \frac{1}{r} \sum_{h=0}^{n-1} \binom{n-1}{h} a^{n-1-h} (-1)^h r^h$$

$$= \frac{1}{a^{n-1}} \sum_{h=0}^{n-1} \binom{n-1}{h} a^{n-1-h} (-1)^h \sum_{r=1}^{a-1} r^{h-1}$$

$$= \frac{1}{a^{n-1}} \sum_{h=1}^{n-1} \binom{n-1}{h} a^{n-1-h} (-1)^h \left[\frac{a^h}{h} + \mathcal{O}(a^{h-1}) \right] + \frac{1}{a^{n-1}} a^{n-1} H_{a-1} \qquad (A.35)$$

$$= \sum_{h=1}^{n-1} \binom{n-1}{h} (-1)^h \frac{1}{h} + \mathcal{O}\left(\frac{1}{a}\right) + H_{a-1}$$

$$= \sum_{h=1}^{n-1} -\frac{1}{h} + \mathcal{O}\left(\frac{1}{a}\right) + H_{a-1}.$$

Thus in total we have

$$\sum_{s=1}^{a-1} \frac{a-s}{sa} \sum_{r=1}^{a-1} \frac{1}{r} \left(\frac{a-r}{a}\right)^{n-1} \left(1 - \left(\frac{a-s}{a}\right)^{n-3}\right)$$

$$= \left[\sum_{\ell=0}^{n-4} \frac{1}{\ell+2} + \mathcal{O}\left(\frac{1}{a}\right)\right] \left[\sum_{h=1}^{n-1} -\frac{1}{h} + \mathcal{O}\left(\frac{1}{a}\right) + H_{a-1}\right] \qquad \text{(A.36)}$$

$$= \sum_{\ell=0}^{n-4} \frac{1}{\ell+2} \left[\sum_{h=1}^{n-1} -\frac{1}{h} + H_{a-1}\right] + \mathcal{O}\left(\frac{\ln(a)}{a}\right).$$

Implying that the sum of the fifth and the seventh term of Equation A.15 is asymptotically negligible:

$$-2\sum_{\ell=0}^{n-4} \frac{1}{\ell+2} \left\{\sum_{h=1}^{n-1} -\frac{1}{h} + H_{a-2}\right\}$$

$$+ 2\sum_{\ell=0}^{n-4} \frac{1}{\ell+2} \left[\sum_{h=1}^{n-1} -\frac{1}{h} + H_{a-1}\right] + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \xrightarrow{a \to \infty} 0. \qquad \text{(A.37)}$$

Thus finally we get

$$\text{Var}(\mathcal{B}_n^a) \xrightarrow{a \to \infty} H_{n-1} - 1 - 2\frac{1}{2}\left(H_{n-1}^{(2)} - 1\right) = H_{n-1} - H_{n-1}^{(2)}. \qquad \text{(A.38)}$$

So the variance of the number of branches in an a-biased tree converges in the case of $\vec{p}$ uniform to the variance of the number of branches of uniform recursive trees. $\qquad \square$

A.3. Proof of Theorem 5.17

Proof. We will now calculate the variance of $Y_{\geq k,n}^p$, the number of nodes with at least k descendants in a BRT. We will use the following expression for the variance:

$$
\begin{aligned}
\operatorname{Var}\left(Y_{\geq k,n}^p\right) =& \operatorname{Var}\left(\sum_{i=2}^{n-k} C_i^k\right) \\
=& \mathbb{E}\left[\left(\sum_{i=2}^{n-k} C_i^k\right)^2\right] - \mathbb{E}\left[\sum_{i=2}^{n-k} C_i^k\right]^2 \\
=& \sum_{i=2}^{n-k} \mathbb{E}\left[C_i^{k^2}\right] - \sum_{i=2}^{n-k} \mathbb{E}\left[C_i^k\right]^2 \\
& + 2\sum_{i=2}^{n-k-1}\sum_{j=i+1}^{n-k} \mathbb{E}\left[C_i^k C_j^k\right] - 2\sum_{i=2}^{n-k-1}\sum_{j=i+1}^{n-k} \mathbb{E}\left[C_i^k\right]\mathbb{E}\left[C_j^k\right].
\end{aligned}
\tag{A.39}
$$

First of all we have

$$
\sum_{i=2}^{n-k} \mathbb{E}\left[C_i^{k^2}\right] = \sum_{i=2}^{n-k} \mathbb{E}\left[C_i^k\right] = (n-k-1)\sum_{s=1}^{a} p_s\left(\sum_{r=s}^{a} p_r\right)^k
\tag{A.40}
$$

and

$$
\sum_{i=2}^{n-k} \mathbb{E}\left[C_i^k\right]^2 = (n-k-1)\left(\sum_{s=1}^{a} p_s\left(\sum_{r=s}^{a} p_r\right)^k\right)^2.
\tag{A.41}
$$

Since the events C_i^k and C_j^k are not mutually independent when $i < j \leq i+k$, we also need to express $\mathbb{E}[C_i^k C_j^k]$ for $i < j \leq i+k$.

Let $2 \leq i < j \leq i+k$, then

$$\mathbb{E}[C_i^k C_j^k]$$

$$= \mathbb{P}\left(X_i \leq \{X_{i+1}, \ldots, X_{i+k}\}, X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}\right)$$

$$= \mathbb{P}\left(X_i \leq \{X_{i+1}, \ldots, X_{j-1}\}, X_i \leq X_j, X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}\right)$$

$$= \sum_{s=1}^{a} \mathbb{P}\left(X_i \leq \{X_{i+1}, \ldots, X_{j-1}\}, X_i \leq X_j, X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}|X_i = s\right)$$

$$\cdot \mathbb{P}(X_i = s)$$

$$= \sum_{s=1}^{a} \mathbb{P}\left(X_i \leq \{X_{i+1}, \ldots, X_{j-1}\}|X_i \leq X_j, X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}, X_i = s\right)$$

$$\cdot \mathbb{P}\left(X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}|X_i \leq X_j, X_i = s\right)$$

$$\cdot \mathbb{P}\left(X_i \leq X_j|X_i = s\right) \mathbb{P}(X_i = s)$$

$$= \sum_{s=1}^{a} \mathbb{P}(X_i = s)\mathbb{P}\left(s \leq \{X_{i+1}, \ldots, X_{j-1}\}\right) \tag{A.42}$$

$$\cdot \mathbb{P}\left(s \leq X_j\right) \mathbb{P}\left(X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}|s \leq X_j\right)$$

$$= \sum_{s=1}^{a} \mathbb{P}(X_i = s)\mathbb{P}\left(s \leq \{X_{i+1}, \ldots, X_{j-1}\}\right)$$

$$\cdot \sum_{r=s}^{a} \mathbb{P}\left(X_j = r\right) \mathbb{P}\left(X_j \leq \{X_{j+1}, \ldots, X_{j+k}\}|X_j = r\right)$$

$$= \sum_{s=1}^{a} \mathbb{P}(X_i = s)\mathbb{P}\left(s \leq \{X_{i+1}, \ldots, X_{j-1}\}\right)$$

$$\cdot \sum_{r=s}^{a} \mathbb{P}\left(X_j = r\right) \mathbb{P}\left(r \leq \{X_{j+1}, \ldots, X_{j+k}\}\right)$$

$$= \sum_{s=1}^{a} p_s \left(\sum_{u=s}^{a} p_u\right)^{j-i-1} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}.$$

Now we have

$$\sum_{i=2}^{n-k-1}\sum_{j=i+1}^{n-k}\mathbb{E}[C_i^k C_j^k] - \sum_{i=2}^{n-k-1}\sum_{j=i+1}^{n-k}\mathbb{E}\left[C_i^k\right]\mathbb{E}[C_j^k]$$

$$= \sum_{i=2}^{n-2k}\sum_{j=i+1}^{i+k}\mathbb{E}[C_i^k C_j^k] - \sum_{i=2}^{n-2k}\sum_{j=i+1}^{i+k}\mathbb{E}\left[C_i^k\right]\mathbb{E}[C_j^k] \tag{A.43}$$

$$+ \sum_{i=n-2k+1}^{n-k-1}\sum_{j=i+1}^{n-k}\mathbb{E}[C_i^k C_j^k] - \sum_{i=n-2k+1}^{n-k-1}\sum_{j=i+1}^{n-k}\mathbb{E}\left[C_i^k\right]\mathbb{E}[C_j^k].$$

Where we eliminated all terms where $i + k < j$ and thus $\mathbb{E}[C_i^k C_j^k] - \mathbb{E}\left[C_i^k\right]\mathbb{E}[C_j^k] = 0$. Moreover we separated the two sums because if $i + k > n - k$, only the C_j^k until $n - k$, not until $i + k$ are relevant. We now consider all 4 expressions separately. As in the expression for the branches, we assume $p_1 > 0$ in order to avoid division by 0. We get

$$\sum_{i=2}^{n-2k}\sum_{j=i+1}^{i+k}\mathbb{E}[C_i^k C_j^k]$$

$$= \sum_{i=2}^{n-2k}\sum_{j=i+1}^{i+k}\sum_{s=1}^{a}p_s\left(\sum_{u=s}^{a}p_u\right)^{j-i-1}\sum_{r=s}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k$$

$$= \sum_{i=2}^{n-2k}\sum_{j=0}^{k-1}\sum_{s=1}^{a}p_s\left(\sum_{u=s}^{a}p_u\right)^{j}\sum_{r=s}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k$$

$$= \sum_{i=2}^{n-2k}p_1\sum_{j=0}^{k-1}\left(\sum_{u=1}^{a}p_u\right)^{j}\sum_{r=1}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k$$

$$+ \sum_{i=2}^{n-2k}\sum_{s=2}^{a}p_s\sum_{j=0}^{k-1}\left(\sum_{u=s}^{a}p_u\right)^{j}\sum_{r=s}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k \tag{A.44}$$

$$= \sum_{i=2}^{n-2k}p_1 k\sum_{r=1}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k$$

$$+ \sum_{i=2}^{n-2k}\sum_{s=2}^{a}p_s\frac{1-\left(\sum_{u=s}^{a}p_u\right)^k}{1-\sum_{u=s}^{a}p_u}\sum_{r=s}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k$$

$$= (n-2k-1)p_1 k\sum_{r=1}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k$$

$$+ (n-2k-1)\sum_{s=2}^{a}p_s\frac{1-\left(\sum_{u=s}^{a}p_u\right)^k}{\sum_{u=1}^{s-1}p_u}\sum_{r=s}^{a}p_r\left(\sum_{t=r}^{a}p_t\right)^k.$$

Also

$$
\begin{aligned}
\sum_{i=2}^{n-2k} \sum_{j=i+1}^{i+k} \mathbb{E}\left[C_i^k\right] \mathbb{E}[C_j^k] &= \sum_{i=2}^{n-2k} \sum_{j=i+1}^{i+k} \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 \\
&= \sum_{i=2}^{n-2k} k \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 \\
&= (n-2k-1)k \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 .
\end{aligned}
\tag{A.45}
$$

For the second part of the sum we get

$$\sum_{i=n-2k+1}^{n-k-1} \sum_{j=i+1}^{n-k} \mathbb{E}[C_i^k C_j^k]$$

$$= \sum_{i=n-2k+1}^{n-k-1} \sum_{j=i+1}^{n-k} \sum_{s=1}^{a} p_s \left(\sum_{u=s}^{a} p_u\right)^{j-i-1} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$= \sum_{i=n-2k+1}^{n-k-1} \sum_{j=0}^{n-k-i-1} \sum_{s=1}^{a} p_s \left(\sum_{u=s}^{a} p_u\right)^{j} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$= \sum_{i=n-2k+1}^{n-k-1} \sum_{j=0}^{n-k-i-1} p_1 \left(\sum_{u=1}^{a} p_u\right)^{j} \sum_{r=1}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$+ \sum_{i=n-2k+1}^{n-k-1} \sum_{s=2}^{a} p_s \frac{1 - \left(\sum_{u=s}^{a} p_u\right)^{n-k-i}}{1 - \sum_{u=s}^{a} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$= \sum_{i=n-2k+1}^{n-k-1} (n-k-i) p_1 \sum_{r=1}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k} \tag{A.46}$$

$$+ \sum_{i=n-2k+1}^{n-k-1} \sum_{s=2}^{a} p_s \frac{1}{1 - \sum_{u=s}^{a} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$- \sum_{i=n-2k+1}^{n-k-1} \sum_{s=2}^{a} p_s \frac{\left(\sum_{u=s}^{a} p_u\right)^{n-k-i}}{1 - \sum_{u=s}^{a} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$= \sum_{i=1}^{k-1} i p_1 \sum_{r=1}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$+ (k-1) \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}$$

$$- \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{i=n-2k+1}^{n-k-1} \left(\sum_{u=s}^{a} p_u\right)^{n-k-i} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^{k}.$$

Continuing with the previous equation we get

$$
\sum_{i=n-2k+1}^{n-k-1} \sum_{j=i+1}^{n-k} \mathbb{E}[C_i^k C_j^k]
$$

$$
= \frac{k(k-1)}{2} p_1 \sum_{r=1}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k
$$

$$
+ (k-1) \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k
$$

$$
- \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{i=1}^{k-1} \left(\sum_{u=s}^{a} p_u \right)^i \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k
$$

$$
= \frac{k(k-1)}{2} p_1 \sum_{r=1}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k
$$

$$
+ (k-1) \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k
$$

$$
- \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \left(\frac{1 - (\sum_{u=s}^{a} p_u)^k}{1 - \sum_{u=s}^{a} p_u} - 1 \right) \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k .
$$

(A.47)

Moreover we have

$$
\sum_{i=n-2k+1}^{n-k-1} \sum_{j=i+1}^{n-k} \mathbb{E}\left[C_i^k \right] \mathbb{E}[C_j^k]
$$

$$
= \sum_{i=n-2k+1}^{n-k-1} \sum_{j=i+1}^{n-k} \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2
$$

$$
= \sum_{i=n-2k+1}^{n-k-1} (n-k-i) \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2
$$

(A.48)

$$
= \sum_{i=1}^{k-1} i \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2
$$

$$
= \frac{k(k-1)}{2} \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 .
$$

After adding up these terms, we get

$$
\begin{aligned}
\mathrm{Var}(Y_n^k) &= \sum_{i=2}^{n-k} \mathbb{E}[C_i^{k2}] + 2\sum_{i=2}^{n-2k}\sum_{j=i+1}^{i+k} \mathbb{E}[C_i^k C_j^k] \\
&\quad + 2\sum_{i=n-2k+1}^{n-k-1}\sum_{j=i+1}^{n-k} \mathbb{E}[C_i^k C_j^k] - \sum_{i=2}^{n-k} \mathbb{E}\left[C_i^k\right]^2 \\
&\quad - 2\sum_{i=2}^{n-2k}\sum_{j=i+1}^{i+k} \mathbb{E}\left[C_i^k\right]\mathbb{E}[C_j^k] - 2\sum_{i=n-2k+1}^{n-k-1}\sum_{j=i+1}^{n-k} \mathbb{E}\left[C_i^k\right]\mathbb{E}[C_j^k] \\
&= (n-k-1)\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k \\
&\quad + 2(n-2k-1)p_1 k \sum_{r=1}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^k \\
&\quad + 2(n-2k-1)\sum_{s=2}^{a} p_s \frac{1-\left(\sum_{u=s}^{a} p_u\right)^k}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^k \\
&\quad + 2\frac{k(k-1)}{2}p_1 \sum_{r=1}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^k \\
&\quad + 2(k-1)\sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^k \\
&\quad - 2\sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u}\left(\frac{1-\left(\sum_{u=s}^{a} p_u\right)^k}{1-\sum_{u=s}^{a} p_u}-1\right) \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t\right)^k \\
&\quad - (n-k-1)\left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k\right)^2 \\
&\quad - 2(n-2k-1)k\left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k\right)^2 \\
&\quad - 2\frac{k(k-1)}{2}\left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r\right)^k\right)^2.
\end{aligned}
\tag{A.49}
$$

This becomes after grouping some terms

$$
\begin{aligned}
\mathrm{Var}(Y^p_{\geq k,n}) = {} & \\
& \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \left[(n-k-1) + 2(n-2k-1)p_1 k + k(k-1)p_1 \right] \\
& + 2 \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k \\
& \cdot \left[(n-2k-1)\left(1 - \left(\sum_{u=s}^{a} p_u \right)^k \right) + (k-1) - \left(\frac{1 - \left(\sum_{u=s}^{a} p_u\right)^k}{1 - \sum_{u=s}^{a} p_u} - 1 \right) \right] \\
& - \left[k(k-1) + (n-k-1) + 2(n-2k-1)k \right] \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 .
\end{aligned}
\tag{A.50}
$$

And after some simplifications we finally get

$$
\begin{aligned}
\mathrm{Var}(Y^p_{\geq k,n}) = {} & \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \left[(n-k-1) + p_1 \left(2nk - 3k(k+1) \right) \right] \\
& + 2 \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k \\
& \cdot \left[n-k-1 - (n-2k-1)\left(\sum_{u=s}^{a} p_u \right)^k - \frac{1 - \left(\sum_{u=s}^{a} p_u\right)^k}{1 - \sum_{u=s}^{a} p_u} \right] \\
& - \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 \left[n(2k+1) - (3k+1)(k+1) \right] .
\end{aligned}
\tag{A.51}
$$

We will now look at the asymptotic for a fixed k as $n \to \infty$:

$$
\lim_{n \to \infty} \frac{\mathrm{Var}(Y_{\geq k,n}^p)}{n}
$$

$$
= \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \left[\frac{n-k-1}{n} + p_1 \frac{2nk - 3k(k+1)}{n} \right]
$$

$$
+ 2 \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k
$$

$$
\cdot \left[\frac{n-k-1}{n} - \frac{n-2k-1}{n} \left(\sum_{u=s}^{a} p_u \right)^k - \frac{1}{n} \frac{1 - \left(\sum_{u=s}^{a} p_u \right)^k}{1 - \sum_{u=s}^{a} p_u} \right]
$$

$$
- \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 \left[\frac{n(2k+1)}{n} - \frac{(3k+1)(k+1)}{n} \right]
$$

$$
= \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k [2kp_1 + 1] \tag{A.52}
$$

$$
+ 2 \sum_{s=2}^{a} p_s \frac{1}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k \left[1 - \left(\sum_{u=s}^{a} p_u \right)^k \right]
$$

$$
- \left(\sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)^2 [2k+1]
$$

$$
= \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \left(2kp_1 + 1 - (2k+1) \sum_{s=1}^{a} p_s \left(\sum_{r=s}^{a} p_r \right)^k \right)
$$

$$
+ 2 \sum_{s=2}^{a} \frac{p_s}{\sum_{u=1}^{s-1} p_u} \sum_{r=s}^{a} p_r \left(\sum_{t=r}^{a} p_t \right)^k \left[1 - \left(\sum_{u=s}^{a} p_u \right)^k \right].
$$

$\square$

A.4. Proof of Corollary 5.18

Proof. If we choose the uniform distribution over $[a]$, Theorem 5.17 gives

$$
\begin{aligned}
\mathrm{Var}\left(Y_{\geq k,n}^{a}\right) = {}& \\
& \sum_{s=1}^{a} \frac{1}{a} \left(\sum_{r=s}^{a} \frac{1}{a}\right)^{k} \left[(n-k-1) + \frac{1}{a}\left(2nk - 3k(k+1)\right)\right] \\
& + 2 \sum_{s=2}^{a} \frac{1}{a} \frac{1}{\sum_{u=1}^{s-1} \frac{1}{a}} \sum_{r=s}^{a} \frac{1}{a} \left(\sum_{t=r}^{a} \frac{1}{a}\right)^{k} \\
& \cdot \left[n-k-1 - (n-2k-1)\left(\sum_{u=s}^{a} \frac{1}{a}\right)^{k} - \frac{1-\left(\sum_{u=s}^{a} \frac{1}{a}\right)^{k}}{1-\sum_{u=s}^{a} \frac{1}{a}}\right] \\
& - \left(\sum_{s=1}^{a} \frac{1}{a} \left(\sum_{r=s}^{a} \frac{1}{a}\right)^{k}\right)^{2} \left[n(2k+1) - (3k+1)(k+1)\right].
\end{aligned}
\tag{A.53}
$$

Using

$$
\sum_{s=1}^{a} \left(\sum_{r=s}^{a} \frac{1}{a}\right)^{k} = \sum_{s=1}^{a} \left(\frac{a-s+1}{a}\right)^{k} = \frac{1}{a^{k}} \sum_{s=1}^{a} s^{k}
\tag{A.54}
$$

and

$$
\begin{aligned}
& \sum_{s=2}^{a} \frac{1}{a} \frac{1}{\sum_{u=1}^{s-1} \frac{1}{a}} \sum_{r=s}^{a} \frac{1}{a} \left(\sum_{t=r}^{a} \frac{1}{a}\right)^{k} \\
& = \sum_{s=2}^{a} \frac{1}{s-1} \sum_{r=s}^{a} \frac{1}{a} \left(\frac{a-r+1}{a}\right)^{k} \\
& = \frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{r=1}^{a-s+1} r^{k}
\end{aligned}
\tag{A.55}
$$

we can simplify this expression and get

$$
\begin{aligned}
\mathrm{Var}\left(Y^a_{\geq k,n}\right) = & \frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k \left[(n-k-1)+\frac{1}{a}\left(2nk-3k(k+1)\right)\right] \\
& + 2\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{r=1}^{a-s+1} r^k \\
& \cdot \left[n-k-1-(n-2k-1)\left(\frac{a-s+1}{a}\right)^k - \frac{1-\left(\frac{a-s+1}{a}\right)^k}{1-\frac{a-s+1}{a}}\right] \\
& - \left(\frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k\right)^2 \left[n(2k+1)-(3k+1)(k+1)\right].
\end{aligned}
\tag{A.56}
$$

For $n \to \infty$ this directly gives, as it also follows from Theorem 5.17,

$$
\begin{aligned}
\lim_{n\to\infty} \frac{\mathrm{Var}\left(Y^a_{\geq k,n}\right)}{n} = & \frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k \left[1+\frac{2k}{a}-\frac{2k+1}{a^{k+1}} \sum_{s=1}^{a} s^k\right] \\
& + \frac{2}{a^{k+1}} \sum_{s=1}^{a-1} \frac{1}{s} \left[1-\left(\frac{a-s}{a}\right)^k\right] \sum_{r=1}^{a-s+1} r^k.
\end{aligned}
\tag{A.57}
$$

In order to calculate the asymptotic value as a tends to infinity, we will first of all use the following equation, that is easy to see

$$
\sum_{s=1}^{a} s^k = \frac{a^{k+1}}{k+1} + \mathcal{O}(a^k) \text{ for all } k \in \mathbb{N}, k \geq 0.
\tag{A.58}
$$

This directly gives results for the first and third term:

$$
\begin{aligned}
& \frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k \left[(n-k-1)+\frac{1}{a}\left(2nk-3k(k+1)\right)\right] \\
= & \frac{1}{a^{k+1}} \left[\frac{a^{k+1}}{k+1} + \mathcal{O}(a^k)\right] \left[(n-k-1)+\frac{1}{a}\left(2nk-3k(k+1)\right)\right] \\
= & \frac{n-k-1}{k+1} + \mathcal{O}\left(\frac{1}{a}\right)
\end{aligned}
\tag{A.59}
$$

and

$$\left(\frac{1}{a^{k+1}} \sum_{s=1}^{a} s^k\right)^2 [n(2k+1) - (3k+1)(k+1)]$$

$$= \left(\frac{1}{a^{k+1}} \left[\frac{a^{k+1}}{k+1} + \mathcal{O}(a^k)\right]\right)^2 [n(2k+1) - (3k+1)(k+1)] \qquad \text{(A.60)}$$

$$= \frac{n(2k+1) - (3k+1)(k+1)}{(k+1)^2} + \mathcal{O}\left(\frac{1}{a}\right).$$

For the middle term we need a little bit more work. First of all

$$\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{r=1}^{a-s+1} r^k$$

$$= \frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \left[\frac{(a-(s-1))^{k+1}}{k+1} + \mathcal{O}((a-(s-1))^k)\right] \qquad \text{(A.61)}$$

$$= \frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \left[\frac{(a-(s-1))^{k+1}}{k+1} + \mathcal{O}((a-(s-1))^k)\right].$$

We moreover have

$$\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \mathcal{O}((a-(s-1))^k) = \mathcal{O}\left(\frac{\ln(a)}{a}\right). \qquad \text{(A.62)}$$

For the rest we develop the term and get

$$\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \frac{(a-(s-1))^{k+1}}{k+1}$$

$$= \frac{1}{k+1} \frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{\ell=0}^{k+1} \binom{k+1}{\ell} a^{k+1-\ell}(-1)^\ell(s-1)^\ell$$

$$= \frac{1}{k+1} \frac{1}{a^{k+1}} \sum_{\ell=0}^{k+1} \binom{k+1}{\ell} a^{k+1-\ell}(-1)^\ell \sum_{s=2}^{a}(s-1)^{\ell-1}$$

$$= \frac{1}{k+1} \frac{1}{a^{k+1}} \left\{\sum_{\ell=1}^{k+1} \binom{k+1}{\ell} a^{k+1-\ell}(-1)^\ell \left[\frac{a^\ell}{\ell} + \mathcal{O}(a^{\ell-1})\right] + a^{k+1} H_{a-1}\right\} \qquad \text{(A.63)}$$

$$= \frac{1}{k+1} \left\{\sum_{\ell=1}^{k+1} \binom{k+1}{\ell}(-1)^\ell \frac{1}{\ell} + H_{a-1}\right\}$$

$$= \frac{1}{k+1} \{H_{a-1} - H_{k+1}\} + \mathcal{O}\left(\frac{1}{a}\right).$$

Hence we get

$$(n-k-1)\frac{1}{a^{k+1}}\sum_{s=2}^{a}\frac{1}{s-1}\sum_{r=1}^{a-s+1}r^k = \frac{(n-k-1)}{k+1}\{H_{a-1}-H_{k+1}\}+\mathcal{O}\left(\frac{\ln(a)}{a}\right). \quad \text{(A.64)}$$

Next we consider

$$\begin{aligned}
&\frac{1}{a^{k+1}}\sum_{s=2}^{a}\frac{1}{s-1}\sum_{r=1}^{a-s+1}r^k\left(\frac{a-s+1}{a}\right)^k \\
&= \frac{1}{a^{2k+1}}\sum_{s=2}^{a}\frac{1}{s-1}(a-s+1)^k\left[\frac{(a-(s-1))^{k+1}}{k+1}+\mathcal{O}((a-(s-1))^k)\right].
\end{aligned} \quad \text{(A.65)}$$

Similarly to before we have

$$\frac{1}{a^{2k+1}}\sum_{s=2}^{a}\frac{1}{s-1}(a-s+1)^k\,\mathcal{O}((a-(s-1))^k) = \mathcal{O}\left(\frac{\ln(a)}{a}\right). \quad \text{(A.66)}$$

Continuing with the rest of the term we get

$$\begin{aligned}
&\frac{1}{a^{2k+1}}\sum_{s=2}^{a}\frac{1}{s-1}(a-s+1)^k\frac{(a-(s-1))^{k+1}}{k+1} \\
&= \frac{1}{k+1}\frac{1}{a^{2k+1}}\sum_{s=2}^{a}\frac{1}{s-1}(a-(s-1))^{2k+1} \\
&= \frac{1}{k+1}\frac{1}{a^{2k+1}}\sum_{s=2}^{a}\frac{1}{s-1}\sum_{\ell=0}^{2k+1}\binom{2k+1}{\ell}a^{2k+1-\ell}(-1)^\ell(s-1)^\ell \\
&= \frac{1}{k+1}\frac{1}{a^{2k+1}}\sum_{\ell=0}^{2k+1}\binom{2k+1}{\ell}a^{2k+1-\ell}(-1)^\ell\sum_{s=2}^{a}(s-1)^{\ell-1} \\
&= \frac{1}{k+1}\frac{1}{a^{2k+1}}\left\{\sum_{\ell=1}^{2k+1}\binom{2k+1}{\ell}a^{2k+1-\ell}(-1)^\ell\left[\frac{a^\ell}{\ell}+\mathcal{O}(a^{\ell-1})\right]+a^{2k+1}H_{a-1}\right\} \\
&= \frac{1}{k+1}\sum_{\ell=1}^{2k+1}\binom{2k+1}{\ell}(-1)^\ell\frac{1}{\ell}+\frac{H_{a-1}}{k+1}+\mathcal{O}\left(\frac{1}{a}\right) \\
&= -\frac{H_{2k+1}}{k+1}+\frac{H_{a-1}}{k+1}+\mathcal{O}\left(\frac{1}{a}\right).
\end{aligned} \quad \text{(A.67)}$$

Thus

$$
-(n - 2k - 1)\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{r=1}^{a-s+1} r^k \left(\frac{a-s+1}{a}\right)^k
$$
$$
= \frac{n-2k-1}{k+1} H_{2k+1} - \frac{n-2k-1}{k+1} H_{a-1} + \mathcal{O}\left(\frac{\ln(a)}{a}\right). \tag{A.68}
$$

Finally we consider

$$
\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{r=1}^{a-s+1} r^k \frac{1 - \left(\frac{a-s+1}{a}\right)^k}{1 - \frac{a-s+1}{a}}
$$
$$
= \frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{\ell=0}^{k-1} \left(\frac{a-(s-1)}{a}\right)^\ell \tag{A.69}
$$
$$
\cdot \left[\frac{(a-s+1)^{k+1}}{k+1} + \mathcal{O}((a-(s-1))^k)\right].
$$

Now we have for all $\ell = 0, \ldots, k-1$,

$$
\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \left(\frac{a-(s-1)}{a}\right)^\ell \mathcal{O}((a-(s-1))^k) = \mathcal{O}\left(\frac{\ln(a)}{a}\right). \tag{A.70}
$$

Continuing with the rest we get

$$\frac{1}{a^{k+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{\ell=0}^{k-1} \left(\frac{a-(s-1)}{a}\right)^{\ell} \frac{(a-s+1)^{k+1}}{k+1}$$

$$= \frac{1}{k+1} \sum_{\ell=0}^{k-1} \frac{1}{a^{k+\ell+1}} \sum_{s=2}^{a} \frac{1}{s-1} (a-(s-1))^{k+\ell+1}$$

$$= \frac{1}{k+1} \sum_{\ell=0}^{k-1} \frac{1}{a^{k+\ell+1}} \sum_{s=2}^{a} \frac{1}{s-1} \sum_{h=0}^{k+\ell+1} \binom{k+\ell+1}{h} a^{k+\ell+1-h}(-1)^h (s-1)^h$$

$$= \frac{1}{k+1} \sum_{\ell=0}^{k-1} \frac{1}{a^{k+\ell+1}} \sum_{h=0}^{k+\ell+1} \binom{k+\ell+1}{h} a^{k+\ell+1-h}(-1)^h \sum_{s=2}^{a}(s-1)^{h-1}$$

$$= \frac{1}{k+1} \sum_{\ell=0}^{k-1} \frac{1}{a^{k+\ell+1}}$$

$$\cdot \left\{ \sum_{h=1}^{k+\ell+1} \binom{k+\ell+1}{h} a^{k+\ell+1-h}(-1)^h \left[\frac{a^h}{h} + \mathcal{O}(a^{h-1})\right] + a^{k+l+1} H_{a-1} \right\}$$

$$= \frac{1}{k+1} \sum_{\ell=0}^{k-1} \sum_{h=1}^{k+\ell+1} \binom{k+\ell+1}{h} (-1)^h \frac{1}{h} + \frac{k}{k+1} H_{a-1} + \mathcal{O}\left(\frac{1}{a}\right)$$

$$= -\frac{1}{k+1} \sum_{\ell=0}^{k-1} H_{k+\ell+1} + \frac{k}{k+1} H_{a-1} + \mathcal{O}\left(\frac{1}{a}\right).$$

(A.71)

This gives us the asymptotic value for $\mathrm{Var}(Y_{\leq k,n}^a)$:

$$\mathrm{Var}\left(Y_{\geq k,n}^a\right)$$

$$= \frac{n-k-1}{k+1} + 2\frac{(n-k-1)}{k+1}\left\{H_{a-1} - H_{k+1}\right\} + 2\frac{n-2k-1}{k+1}H_{2k+1}$$

$$- 2\frac{n-2k-1}{k+1}H_{a-1} + 2\frac{1}{k+1}\sum_{\ell=0}^{k-1} H_{k+\ell+1} - 2\frac{k}{k+1}H_{a-1}$$

$$- \frac{n(2k+1) - (3k+1)(k+1)}{2(k+1)} + \mathcal{O}\left(\frac{\ln(a)}{a}\right)$$

$$\xrightarrow{a\to\infty} \frac{n-k-1}{k+1} - 2\frac{(n-k-1)}{k+1}H_{k+1} + 2\frac{n-2k-1}{k+1}H_{2k+1}$$

$$+ \frac{2}{k+1}\sum_{\ell=0}^{k-1} H_{k+\ell+1} - \frac{n(2k+1) - (3k+1)(k+1)}{(k+1)^2}.$$

(A.72)

And in particular we get

$$\lim_{n\to\infty}\lim_{a\to\infty}\frac{\mathrm{Var}\left(Y_{\geq k,n}^a\right)}{n}=\frac{1}{k+1}+\frac{2H_{2k+1}-2H_{k+1}}{k+1}-\frac{2k+1}{(k+1)^2}. \tag{A.73}$$

□

A.5. Proof of Corollary 5.27

Proof. We will first derive the exact value for the expected value of the depth of n in an a-RT, which we denote by $\mathcal{D}_n^a$. First we exchange all p_i for $\frac{1}{a}$ in Equation 5.72. This gives

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^a\right]=&\\
&\sum_{s=2}^{a}\frac{\frac{1}{a}}{\sum_{r=1}^{s-1}\frac{1}{a}}\sum_{s'=2}^{a}\left[\left(\sum_{r=1}^{s'}\frac{1}{a}\right)^{n-1}-\left(\sum_{r=1}^{s'-1}\frac{1}{a}\right)^{n-1}\right]\\
&-\sum_{s=2}^{a}\frac{\frac{1}{a}}{\sum_{r=1}^{s-1}\frac{1}{a}}\sum_{s'=2}^{a}\frac{1}{a}\frac{\left(\sum_{r=1}^{s'-1}\frac{1}{a}\right)^{n-1}-\left(\sum_{r=s}^{a}\frac{1}{a}\sum_{r=1}^{s'}\frac{1}{a}\right)^{n-1}}{\sum_{r=1}^{s'-1}\frac{1}{a}-\sum_{r=s}^{a}\frac{1}{a}\sum_{r=1}^{s'}\frac{1}{a}}\\
&+\frac{1}{a}\sum_{s=2}^{a}\frac{1}{\frac{1}{a}}\left[(n-2)\left(\sum_{r=1}^{s}\frac{1}{a}\right)^{n}\right.\\
&\qquad\left.-(n-1)\left(\sum_{r=1}^{s}\frac{1}{a}\right)^{n-1}\sum_{r=1}^{s-1}\frac{1}{a}+\sum_{r=1}^{s}\frac{1}{a}\left(\sum_{r=1}^{s-1}\frac{1}{a}\right)^{n-1}\right]\\
&+\sum_{s=2}^{a}\left(\left(\sum_{r=1}^{s}\frac{1}{a}\right)^{n-1}-\left(\sum_{r=1}^{s-1}\frac{1}{a}\right)^{n-1}\right)\\
&+\left[\sum_{s=2}^{a}\frac{1}{a}\frac{1-\left(\sum_{r=s}^{a}\frac{1}{a}\right)^{n-2}}{\sum_{r=1}^{s-1}\frac{1}{a}}+(n-2)\frac{1}{a}+1\right]\frac{1}{a}^{n-1}.
\end{aligned}
\tag{A.74}
$$

This expression can be simplified to get

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^a\right] = {} & \sum_{s=2}^{a} \frac{1}{s-1} \sum_{s'=2}^{a} \left[\left(\frac{s'}{a}\right)^{n-1} - \left(\frac{s'-1}{a}\right)^{n-1} \right] \\
& - \sum_{s=2}^{a} \frac{1}{s-1} \sum_{s'=2}^{a} \frac{1}{a} \frac{\left(\frac{s'-1}{a}\right)^{n-1} - \left(\frac{a-s+1}{a}\frac{s'}{a}\right)^{n-1}}{\frac{s'-1}{a} - \frac{a-s+1}{a}\frac{s'}{a}} \\
& + \sum_{s=2}^{a} (n-2)\left(\frac{s}{a}\right)^{n} - (n-1)\left(\frac{s}{a}\right)^{n-1}\frac{s-1}{a} + \frac{s}{a}\left(\frac{s-1}{a}\right)^{n-1} \\
& + \sum_{s=2}^{a} \left(\left(\frac{s}{a}\right)^{n-1} - \left(\frac{s-1}{a}\right)^{n-1} \right) \\
& + \left[\sum_{s=2}^{a} \frac{1}{a} \frac{1 - \left(\frac{a-s+1}{a}\right)^{n-2}}{\frac{s-1}{a}} + (n-2)\frac{1}{a} + 1 \right] \frac{1}{a}^{n-1}.
\end{aligned}
\tag{A.75}
$$

After some more simplifications, we have

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^a\right] = {} & \frac{1}{a^{n-1}} \sum_{s=1}^{a-1} \sum_{s'=1}^{a-1} \frac{1}{s} \left[(s'+1)^{n-1} - s'^{n-1} \right] \\
& - \frac{1}{a^{n-2}} \sum_{s=1}^{a-1} \sum_{s'=1}^{a-1} \frac{1}{s} \frac{s'^{n-1} - ((a-s)(s'+1))^{n-1}}{ss' + s - a} \\
& + \frac{1}{a^{n}} \sum_{s=1}^{a-1} (n-2)(s+1)^{n} - (n-1)(s+1)^{n-1}s + (s+1)s^{n-1} \\
& + \frac{1}{a^{n-1}} \sum_{s=1}^{a-1} (s+1)^{n-1} - s^{n-1} \\
& + \frac{1}{a^{n-1}} \left[\frac{1}{a^{n-2}} \sum_{s=1}^{a-1} \frac{a^{n-2} - (a-s)^{n-2}}{s} + \frac{n-2}{a} + 1 \right].
\end{aligned}
\tag{A.76}
$$

This in turn gives

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^a\right] = {}& \frac{1}{a^{n-1}} \sum_{s=1}^{a-1} \frac{1}{s} \left(a^{n-1}-1\right) \\
& - \frac{1}{a^{n-2}} \sum_{s=1}^{a-1}\sum_{s'=1}^{a-1} \frac{1}{s} \frac{s'^{n-1}-((a-s)(s'+1))^{n-1}}{ss'+s-a} \\
& + \frac{1}{a^n} \sum_{s=1}^{a-1}(n-2)\,(s+1)^n - (n-1)\,(s+1)^{n-1}\,s + (s+1)s^{n-1} \\
& + \frac{1}{a^{n-1}} \left(a^{n-1}-1\right) \\
& + \frac{1}{a^{n-1}} \left[\frac{1}{a^{n-2}} \sum_{s=1}^{a-1} \frac{a^{n-2}-(a-s)^{n-2}}{s} + \frac{n-2}{a} + 1 \right].
\end{aligned}
\tag{A.77}
$$

And after some more simplifications

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^a\right] = {}& H_{a-1} - \frac{H_{a-1}}{a^{n-1}} \\
& - \frac{1}{a^{n-2}} \sum_{s=1}^{a-1}\sum_{s'=1}^{a-1} \frac{1}{s} \frac{s'^{n-1}-((a-s)(s'+1))^{n-1}}{ss'+s-a} \\
& + \frac{1}{a^n} \sum_{s=1}^{a-1}(n-2)\,(s+1)^n - (n-1)\,(s+1)^{n-1}\,s + (s+1)s^{n-1} \\
& + 1 - \frac{1}{a^{n-1}} \\
& + \frac{H_{a-1}}{a^{n-1}} + \frac{1}{a^{n-1}} + \frac{n-2}{a^n} + \frac{1}{a^{2n-3}} \sum_{s=2}^{a} \frac{-(a-s+1)^{n-2}}{s-1}.
\end{aligned}
\tag{A.78}
$$

Finally we get

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}_n^a\right] = {}& H_{a-1} + 1 + \frac{n-2}{a^n} - \frac{1}{a^{2n-3}} \sum_{s=1}^{a-1} \frac{(a-s)^{n-2}}{s} \\
& - \frac{1}{a^{n-2}} \sum_{s=1}^{a-1}\sum_{s'=1}^{a-1} \frac{1}{s} \frac{s'^{n-1}-((a-s)(s'+1))^{n-1}}{ss'+s-a} \\
& + \frac{1}{a^n} \sum_{s=1}^{a-1}(n-2)\,(s+1)^n - (n-1)\,(s+1)^{n-1}\,s + (s+1)s^{n-1}.
\end{aligned}
\tag{A.79}
$$

For $n \to \infty$ Equation 5.73 gives easily,

$$\frac{\mathbb{E}\left[\mathcal{D}_n^a\right]}{n} \xrightarrow{n \to \infty} \frac{1}{a}. \tag{A.80}$$

This can also be derived from the previous expression.

For the asymptotics when $a \to \infty$ we need more work. Since in the uniform case the expectation of the depth of n is equal to H_{n-1} we expect the same here. We will first look at the term in the second line of Equation A.79. In order to get its asymptotic value we will use the form it had before simplifying, as in Equation A.75. Using this we have

$$\begin{aligned}
&\frac{1}{a^{n-2}} \sum_{s=1}^{a-1} \sum_{s'=1}^{a-1} \frac{1}{s} \frac{s'^{n-1} - ((a-s)(s'+1))^{n-1}}{ss' + s - a} \\
&= \sum_{s=2}^{a} \frac{1}{s-1} \sum_{s'=2}^{a} \frac{1}{a} \frac{\left(\frac{s'-1}{a}\right)^{n-1} - \left(\frac{a-s+1}{a}\frac{s'}{a}\right)^{n-1}}{\frac{s'-1}{a} - \frac{a-s+1}{a}\frac{s'}{a}} \\
&= \sum_{s=1}^{a-1} \frac{1}{s} \sum_{s'=1}^{a-1} \frac{1}{a} \frac{\left(\frac{s'}{a}\right)^{n-1} - \left(\frac{a-s}{a}\frac{s'+1}{a}\right)^{n-1}}{\frac{s'}{a} - \frac{a-s}{a}\frac{s'+1}{a}} \\
&= \frac{1}{a} \sum_{s=1}^{a-1} \frac{1}{s} \sum_{s'=1}^{a-1} \left(\frac{s'}{a}\right)^{n-2} \frac{1 - \left(\left(\frac{a-s}{a}\frac{s'+1}{a}\right)\left(\frac{s'}{a}\right)^{-1}\right)^{n-1}}{1 - \frac{a-s}{a}\frac{s'+1}{a}\left(\frac{s'}{a}\right)^{-1}} \\
&= \frac{1}{a} \sum_{s=1}^{a-1} \frac{1}{s} \sum_{s'=1}^{a-1} \left(\frac{s'}{a}\right)^{n-2} \sum_{\ell=0}^{n-2} \left(\left(\frac{a-s}{a}\frac{s'+1}{a}\right)\left(\frac{s'}{a}\right)^{-1}\right)^{\ell} \\
&= \sum_{\ell=0}^{n-2} \frac{1}{a^{n-1+\ell}} \sum_{s=1}^{a-1} \frac{1}{s}(a-s)^{\ell} \sum_{s'=1}^{a-1} s'^{n-2-\ell}(s'+1)^{\ell}.
\end{aligned} \tag{A.81}$$

We know look at the term containing only s and the term containing only s' separately. First of all for $\ell = 0, \ldots, n - 2$,

$$
\begin{aligned}
\sum_{s=1}^{a-1} \frac{1}{s}(a - s)^\ell &= \sum_{s=1}^{a-1} \frac{1}{s} \sum_{k=0}^{\ell} (-1)^k \binom{\ell}{k} a^{\ell-k} s^k \\
&= \sum_{k=0}^{\ell} (-1)^k \binom{\ell}{k} a^{\ell-k} \sum_{s=1}^{a-1} s^{k-1} \\
&= \sum_{k=1}^{\ell} (-1)^k \binom{\ell}{k} a^{\ell-k} \left[\frac{a^k}{k} + \mathcal{O}(a^{k-1}) \right] + a^\ell H_{a-1} \\
&= a^\ell \sum_{k=1}^{\ell} (-1)^k \binom{\ell}{k} \frac{1}{k} + \mathcal{O}(a^{l-1}) + a^\ell H_{a-1}.
\end{aligned}
\tag{A.82}
$$

Now for the term depending on s' only we get for $\ell = 0, \ldots, n - 2$,

$$
\begin{aligned}
\sum_{s'=1}^{a-1} s'^{n-2-\ell}(s' + 1)^\ell &= \sum_{s'=1}^{a-1} s'^{n-2-\ell} \sum_{h=0}^{\ell} \ell \binom{\ell}{h} s'^h \\
&= \sum_{s'=1}^{a-1} s'^{n-2-\ell} \sum_{h=0}^{\ell} \binom{\ell}{h} s'^h \\
&= \sum_{h=0}^{\ell} \binom{\ell}{h} \sum_{s'=1}^{a-1} s'^{n+h-\ell-2} \\
&= \sum_{h=0}^{\ell} \binom{\ell}{h} \left[\frac{a^{n+h-\ell-1}}{n+h-\ell-1} + \mathcal{O}(a^{n+h-\ell-2}) \right].
\end{aligned}
\tag{A.83}
$$

By inserting these expressions into Equation A.81 we get

$$
\begin{aligned}
\sum_{\ell=0}^{n-2} & \frac{1}{a^{n-1+\ell}} \sum_{s=1}^{a-1} \frac{1}{s}(a-s)^\ell \sum_{s'=1}^{a-1} s'^{n-2-\ell}(s'+1)^\ell \\
&= \sum_{\ell=0}^{n-2} \frac{1}{a^{n-1+\ell}} \left[a^\ell \sum_{k=1}^{\ell}(-1)^k \binom{\ell}{k}\frac{1}{k} + \mathcal{O}(a^{l-1}) + a^\ell H_{a-1} \right] \\
&\quad \cdot \sum_{h=0}^{\ell} \binom{\ell}{h} \left[\frac{a^{n+h-\ell-1}}{n+h-\ell-1} + \mathcal{O}(a^{n+h-\ell-2}) \right] \\
&= \sum_{\ell=0}^{n-2} H_{a-1} \sum_{h=0}^{\ell} \binom{\ell}{h} \left[\frac{1}{a^{\ell-h}}\frac{1}{n+h-\ell-1} + \mathcal{O}\left(\frac{1}{a^{\ell-h+1}}\right) \right] \\
&\quad + \sum_{\ell=1}^{n-2} \left[\sum_{k=1}^{\ell}(-1)^k \binom{\ell}{k}\frac{1}{k} + \mathcal{O}\left(\frac{1}{a}\right) \right] \\
&\quad \cdot \sum_{h=0}^{\ell} \binom{\ell}{h} \left[\frac{1}{a^{\ell-h}}\frac{1}{n+h-\ell-1} + \mathcal{O}\left(\frac{1}{a^{\ell-h+1}}\right) \right] \\
&= \frac{1}{n-1} \sum_{\ell=0}^{n-2} H_{a-1} + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \\
&\quad + \frac{1}{n-1} \sum_{\ell=1}^{n-2} \sum_{k=1}^{\ell}(-1)^k \binom{\ell}{k}\frac{1}{k} + \mathcal{O}\left(\frac{1}{a}\right) \\
&= H_{a-1} - \frac{1}{n-1} \sum_{\ell=1}^{n-2} H_\ell + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \\
&= H_{a-1} - \frac{1}{n-1} \sum_{\ell=1}^{n-2} (n-1-\ell)\frac{1}{\ell} + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \\
&= H_{a-1} - \sum_{\ell=1}^{n-2} \left(\frac{1}{\ell} - \frac{1}{n-1}\right) + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \\
&= H_{a-1} - H_{n-2} + \frac{n-2}{n-1} + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \\
&= H_{a-1} - H_{n-1} + 1 + \mathcal{O}\left(\frac{\ln(a)}{a}\right)
\end{aligned}
\tag{A.84}
$$

where we again used Lemma A.1.

Now we look at the last term of Equation A.79:

$$
\frac{1}{a^n} \sum_{s=1}^{a-1} (n-2)(s+1)^n - (n-1)(s+1)^{n-1}s + (s+1)s^{n-1}
$$

$$
= \frac{1}{a^n} \sum_{s=1}^{a-1} (n-2) \sum_{\ell=0}^{n} \binom{n}{\ell} s^\ell - (n-1) \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} s^{\ell+1} + s^n + s^{n-1}
$$

$$
= \frac{1}{a^n} \sum_{s=1}^{a-1} (n-2) \sum_{\ell=0}^{n-2} \binom{n}{\ell} s^\ell - (n-1) \sum_{\ell=0}^{n-3} \binom{n-1}{\ell} s^{\ell+1} \tag{A.85}
$$

$$
= \frac{1}{a^n} \left[(n-2) \sum_{\ell=0}^{n-2} \binom{n}{\ell} \sum_{s=1}^{a-1} s^\ell - (n-1) \sum_{\ell=0}^{n-3} \binom{n-1}{\ell} \sum_{s=1}^{a-1} s^{\ell+1} \right]
$$

$$
= \mathcal{O}\left(\frac{1}{a}\right).
$$

Moreover we have

$$
\frac{n-2}{a^n} - \frac{1}{a^{2n-3}} \sum_{s=1}^{a-1} \frac{(a-s)^{n-2}}{s} = \mathcal{O}\left(\frac{1}{a}\right). \tag{A.86}
$$

Thus in total we get

$$
\mathbb{E}\left[\mathcal{D}_n^a\right] = H_{a-1} + 1 - H_{a-1} + H_{n-1} - 1 + \mathcal{O}\left(\frac{\ln(a)}{a}\right) \tag{A.87}
$$

$$
\xrightarrow{a \to \infty} H_{n-1}.
$$

$\square$