Contents

Abstract

"Perhaps an Erdos quote linking probability and graph theory"

A *random recursive tree* is a finite nested family of labelled trees, $\{T_t\}_{t=1}^{n}$ built by iteratively attaching a new vertex to one of the existing vertices uniformly at random. At step t in this iteration let $a_t = |\mathrm{Aut}(T_t)|$ where $\mathrm{Aut}(T_t)$ is the group of automorphisms of tree T_t. MacArthur, ([**?**], Personal Communication), performed numerical calculations which suggest that in the limit $t \to \infty$

$$(a_t)^{\frac{1}{t}} \to \mathcal{V}$$

where $\mathcal{V}$ is an unknown constant redolent of the Fibonacci sequence.

A random Fibonacci sequence is a simple variation of the usual Fibonaci sequence in which $F(1) = F(2) = 1$ and F is recursively defined for $n > 2$:

$$F(n) = F(n-1) \pm F(n-2),$$

where the addition and negation are chosen equiprobably. Viswanath [**?**] showed that almost surely $|f(t)|^{\frac{1}{t}} \to 1.13198824$ (a number that has subsequently become known as Viswanath's constant) as $t \to \infty$ although his proof was computational in nature [1]. Macarthur made the following conjecture.

Conjecture 0.0.1. *In the limit $t \to \infty, a_t^{\frac{1}{t}} \to \mathcal{V}$ where $\mathcal{V}$ is Viswinath's constant.*

The primary goal of this project is to prove Conjecture 0.0.1.

[1] More recently Viswanath's constant was calculated to another 5 decimal places [**?**]

Chapter 1

Introduction

1.1 Attachment trees

The set of vertices and the set of edges of some tree T are denoted $V(T)$ and $E(T)$ respectively. A *random attachment tree*, $T = \{T_t\}_{t=1}^{n}$ on n vertices is a nested sequence of labelled, rooted trees:

$$T_1 \subset T_2 \subset \cdots \subset T_n.$$

Tree T_1 consists of 1 vertex and no edges. At subsequent time $t = 2, 3, \ldots, n$ a vertex, v, is chosen from $V(T_{t-1})$ according to some attachment model and a new vertex labelled t is attached to v via an edge.

Note that each T_i is indeed a tree since it has i vertices and $i - 1$ edges and each T_i is obviously connected.

If the attachment model is uniform random attachment then we say that $\{T_t\}_{t=1}^{n}$ is a *random recursive tree* on n vertices. For succinctness we will henceforth refer to a random recursive trees as a RRT. A random recursive d-ary tree is an attachment tree withe following attachment model. The probability of a new vertex is uniformly distributed amongst the vertices with degree less than d hence the proabbility of attaching to an existing vertex with degree d is 0.

For any tree T let $d(v, w)$ be the length of the (unique) shortest path between any pair of vertices $v, w \in V(T)$. Each vertex (apart from 1) of an attachment tree is adjacent to precisely one vertex with a lower label. Furthermore, we define the *father* of each vertex v to be the vertex v' adjacent to v such that $d(v', 1) < d(v, 1)$. By the attachment process the father of a vertex is unique hence well-defined.

For example, consider the attachment tree T in Figure 1.1, the distance $d(1, 10) = 2$ and the father of vertex 13 is vertex 8.

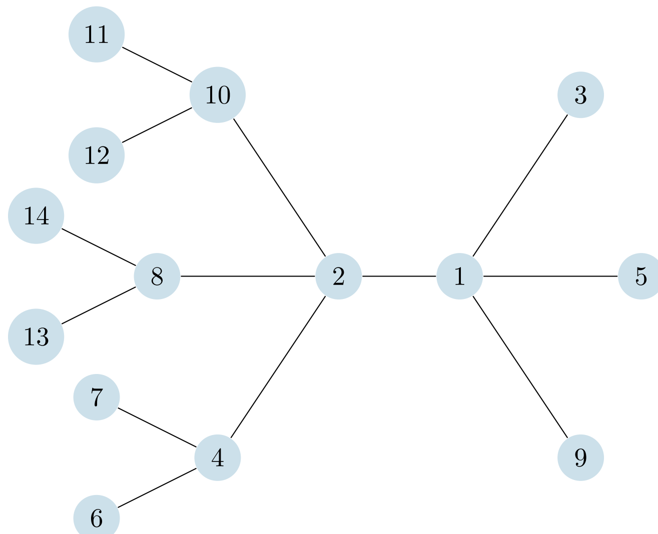

Figure 1.1: An example of an attachment tree $T = \{T_t\}_{t=1}^{14}$.

1.2 The automorphism group of labelled trees

Two labelled trees T_1 and T_2 are considered the same and called isomorphic if and only if there is a 1-1 map $\alpha : V(T_1) \to V(T_2)$ which preserves adjacency and the labelling. If $T_1 = T_2$ then α is called an automorphism. The collection of all automorphisms of tree T is denoted $\mathrm{Aut}(T)$ and constitutes a group.

We denote the symmetric group on n objects by S_n and the set of all labelled trees by $\tilde{\mathcal{T}}_\mathrm{n}$. We can act on $\tilde{\mathcal{T}}_\mathrm{n}$ with S_n by permuting vertices i.e. for some $\sigma \in S_n$ and $T \in \tilde{\mathcal{T}}_\mathrm{n}$ we define $\sigma \cdot T$ to be the tree T' such that $(\sigma(v), \sigma(w))$ is an edge of T' if and only if (v, w) is an edge of T. We will abuse notation by writing $\sigma \in \mathrm{Aut}(T)$ if $\sigma \cdot T = T$.

The orbits of this action are the unlabelled trees on n vertices. Denote the number of unlabelled trees on n vertices by t_n then by Burnside's lemma:

$$t_n = \frac{1}{|S_n|} \sum_{\sigma \in S_n} |\mathrm{fix}(\sigma)| \tag{1.2.1}$$

Fix a permutation, $\sigma \in S_n$, and consider $\mathrm{fix}(\sigma) = \{T \in \tilde{\mathcal{T}}_\mathrm{n} : \sigma \cdot T = T\}$. The permutation $\sigma \in \mathrm{Aut}(T)$ for all $T \in \mathrm{fix}(\sigma)$. To make this clear consider the indicator function:

$$I(\sigma, T) = \begin{cases} 1 & \text{if } \sigma \in \mathrm{Aut}(T) \\ 0 & \text{otherwise.} \end{cases}$$

The expected order of automorphism group of a labelled tree on n vertices

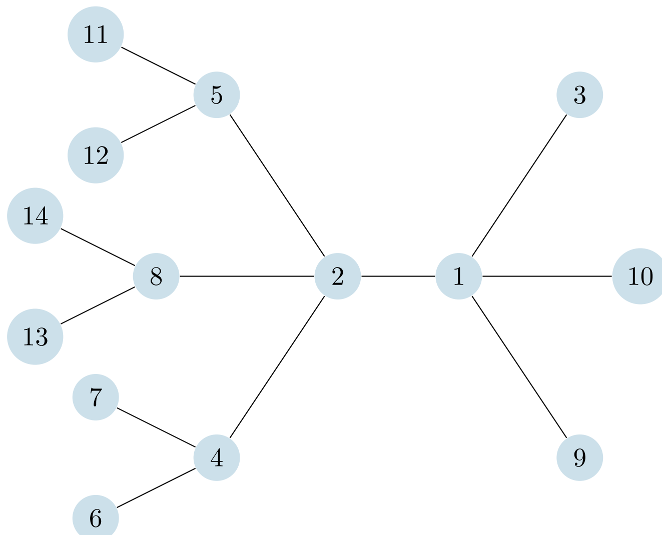

Figure 1.2: This the tree $\sigma \cdot T$ where σ is the transposition $(5, 10)$ and T is the tree in Figure 1.1. Note that σ is *not* an automorphism of T.

is:

$$\mathbb{E}(\mathrm{Aut}_n(T)) = \frac{1}{|\tilde{\mathcal{T}}_\mathrm{n}|} \sum_{T \in \tilde{\mathcal{T}}_\mathrm{n}} |\mathrm{Aut}(T)| \tag{1.2.2}$$

$$= \frac{1}{|\tilde{\mathcal{T}}_\mathrm{n}|} \sum_{\sigma \in S_n} \sum_{T \in \tilde{\mathcal{T}}_\mathrm{n}} I(\sigma, T) \tag{1.2.3}$$

$$= \frac{1}{|\tilde{\mathcal{T}}_\mathrm{n}|} \sum_{\sigma \in S_n} |\mathrm{fix}(\sigma)| \tag{1.2.4}$$

$$= \frac{t_n |S_n|}{|\tilde{\mathcal{T}}_\mathrm{n}|}. \tag{1.2.5}$$

1.3 Automorphisms of random recursive trees

It is a result of Pólya that the automorphism group of a tree belongs to the class of permutation groups which contains the symmetric groups and is closed under taking direct and wreath products [?]. Therefore there exists a direct product decomposition for any tree T:

$$\mathrm{Aut}(T) \cong A_1 \times A_2 \times \cdots \times A_p \times B_1 \times B_2 \times \cdots \times B_q \tag{1.3.1}$$

Such that $A_i \cong S_{x(i)}$ and $B_j = S_{y_1(j)} \wr S_{y_2(j)} \wr \cdots \wr S_{y_{k_j}(j)}$.

In addition to this algebraic interpretation of $\mathrm{Aut}(T_n)$ there is a pleasing geometric realisation of equation 1.3.1. In this geometric realisation each symmetric factor S_k corresponds to a hub vertex adjacent to k paths of length n, each of which terminates in a leaf. Each wreath product factor, B_i, corresponds to an extended symmetric induced subtree [**?**]. We denote an induced subtree consisting of a hub vertex adjacent to k paths of length n a (n,k)-star. For succinctness we will refer to $(1,k)$-stars simply as k-stars. These (n,k)-stars and extended symmetric branches are known as *network motifs* and have been described as the building blocks of many real-world networks [**?**].l

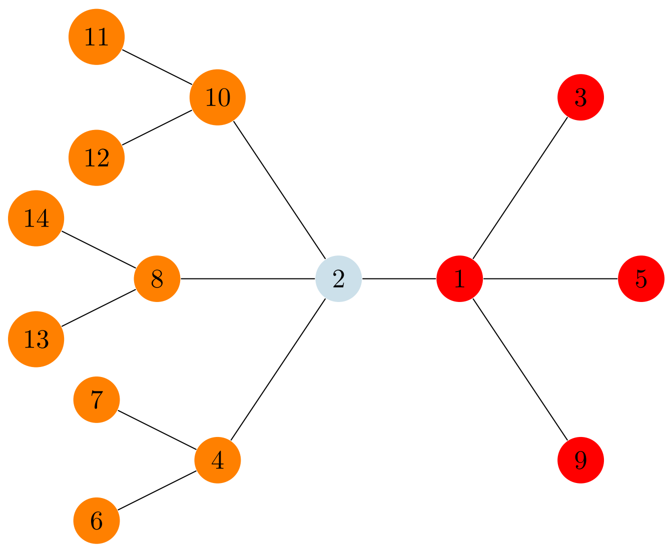

Figure 1.3: An example of a random recursive tree, T, such that $\mathrm{Aut}(T) \cong S_3 \times S_2 \wr S_3$. The red vertices indicate an induced subtree $\tilde{T}_{v_1}$ isomorphic to the bipartite graph $k_{1,3}$ which contributes S_3 to $\mathrm{Aut}(T_{14})$. The orange nodes highlight an induced subtree, $\tilde{T}_{v_2}$ of T_{14} isomorphic to an extended symmetric subtree and contributes $S_2 \wr S_3$ to $\mathrm{Aut}(T)$.

Let T be a tree. There is a natural way to split $\mathrm{Aut}(T) = A_1 \times \cdots \times A_p \times B_1 \times \cdots \times B_q$ into two subgroups. The direct product of symmetric groups form the *elementary subgroup*:

$$\mathcal{E}(T) = A_1 \times A_2 \times \cdots \times A_p$$

The direct product of wreath products of symmetric groups form the *complex subgroup*:

$$\mathcal{C}(T) = B_1 \times B_2 \times \cdots \times B_q$$

The elementary subgroup captures the contribution that (n,k)-stars make to the automorphism group and the complex subgroup captures the contribution that the extended symmetric branches make to the automorphism group. The order of the automorphism group of tree T can also be split as follows:

$$|\mathrm{Aut}(T)| = |\mathcal{E}(T)||\mathcal{C}(T)|$$

This begs the question: does the order of either the elementary or the complex subgroup dominate the other? MacArthur [**?**] made the following additional conjecture:

Conjecture 1.3.1. *Let $\{T_t\}_{t=1}^{n}$ be a RRT. In the limit as $n \to \infty$, $|\mathcal{E}(T_n)|^{\frac{1}{n}} = \mathcal{V}$, while in the limit as $n \to \infty$, $|\mathcal{C}(T_n)|^{\frac{1}{n}} = 1$.*

If conjecture 1.3.1 were true then to prove conjecture 0.0.1 it is enough to calculate the limiting behaviour of (n, k)-stars.

In Chater 2.1 we refine a bijection of Prüfer between labelled trees on n vertices and sequences called Prüfer sequences by restricting this bijection to random recursive trees. As a result there is an obvious way to give the set $\mathcal{T}_n$ additional structure by thinking it as a topological and then a measure space.

In Chapter 3 we exploit another bijection, this time between $\mathcal{T}_n$ and a certain class of integer valued functions. We define what it is to be an automorphism of these functions and use a varaiant of equation **??** to calculate the expected number of k-stars in a random recursive tree on n vertices.

In Chapter 4 we construct several further bijections between $\mathcal{T}_n$ and a class of binary trees, and a set of recursively defined permutations via a combinatorial setup known as a Chinese Restraunt Process. We show that many of the results from the theory of random permutations can be fruitfully applied to RRTs. For example, we calculate the expected number of leaves in a RRT on n vertices. We conclude Chapter 4 by suggesting a possible future path in order to calculate the expected number of (k)-stars in a RRT.

Finally, in Chapter 5 we build a statistical device called a Polya Urn scheme to output the features of a typical RRT such as the degree distribution, the number of (n, k)-stars and the number of extended symmetric branches. We then give limits for the order of limiting order of the elementary subgroup $|\mathcal{E}(T)|$. We end Chapter 5 by proving that conjecture 1.3.1 is false.

Chapter 2

A Measure on Random Recursive Trees

2.1 Prüfer sequences

Cayley was first to prove that there are n^{n-2} labelled trees on n vertices [**?**]. A later proof due to Prüfer uses a correspondence between particular sequences called *Prüfer sequences* and labelled trees [**?**].

Let the set of the first n integers be denoted $\mathbb{N}_n = \{1, 2, 3, \ldots n\}$. A Prüfer sequence $P_n = (a[1], a[2], \ldots a[n-2])$ is a sequence of length $n-2$ where each entry $a[i] \in \mathbb{N}_n$. For example $P = (2, 1, 4, 5, 4, 2, 2, 2, 5, 7, 7)$ is a Prüfer sequence.

If A and B are sets, we denote the set of elements in A and not in B by $A \backslash B$. Let $L(P_n)$ be the set of all elements $l \in P_n \backslash \mathbb{N}_n$. For example, $L(P) := \{3, 6, 8, 9, 10, 11, 12, 13\}$.

The set of Prüfer sequences of length $n-2$ is denoted $\mathcal{P}_n$. Prüfer showed that there exists the following bijection.

$$f : \mathcal{P}_n \to \tilde{\mathcal{T}}_n$$

via the following dynamic correspondence. Let $P_n = (a[1], a[2], \ldots a[n-2])$ be a Prüfer sequence.

(i) At time $t = 0$ define graph G_0 to have n vertices and no edges and define $L_0 = L(P_n)$.

(ii) At times $t = 1, 2, \ldots, n-2$ construct G_t from G_{t-1} by adding the edge $(a[t], \min(L_{t-1})$ and the set L_t is build from L_{t-1} by removing element $\min(L_{t-1})$ and, if for all i such that $a[i] = a[t]$, $i \leq t$ $a[t]$ then add $a[t]$ to L_{t-1}.

(iii) Let D be the number of distinct entries of P_n. Since D elements are added to L_0 and $n-2$ elements are removed to form L_{t-1}:

$$|L_{n-1}| = |L_0| + D - (n-2).$$

Therefore $|L_{n-1}| = 2$. Finally, G_{n-1} is formed by adding an edge between the two vertices in L_{n-1}.

Graph G_{n-1} consists of n vertices and $n-1$ edges. Since for every $x \in \mathbb{N}_n$ there exists an i such that $x \in L_i$ every vertex in G_{n-1} is incident to at least one edge so $G_{n-1} \in \tilde{\mathcal{T}}_n$.

Given a labelled tree T_n we can generate a Prüfer sequence by removing the smallest leaf of T_n at time i and assigning $a[i]$ to be the vertex adjacent to i for $i = 1, 2, \ldots, n-2$. For example $P = (1, 1, 4, 4, 2, 1, 2, 10, 10, 2, 8, 8)$ is the unique Prüfer sequence that corresponds to tree T in Figure 1.1. Map f is a bijection and as a corollary:

$$|\tilde{\mathcal{T}_n}| = |\mathcal{P}_n| = n^{n-2}$$

We could restrict f^{-1} to RRTs as $f^{-1}|_{\mathcal{T}_n} : \mathcal{T}_n \to \mathcal{P}_n$ and ask the question: what form does the image of $f^{-1}|_{\mathcal{T}_n}$ take?

Define another bijection $f^* : \P_n \to \tilde{\mathcal{T}}_{\rm n}$ via a similar correspondence to map f. We define L and G_0 as before and at times $t = 1, 2, \ldots n-2$ we construct G_t from G_{t-1} by adding the edge $(a[t], \max(L_{t-1})$. We can uniquely construct the preimage of labelled tree $T \in \tilde{\mathcal{T}}_{\rm n}$ by removing the leaf with the largest label and setting $a[1]$ to be the vertex adjacent to that leaf. Entry $a[2]$ is the vertex attached to the new greatest value leaf and so on. For example, the Prüfer sequence that corresponds to T in Figure 1.1 under f^* is:

$$P = (8, 8, 10, 10, 2, 1, 2, 4, 4, 1, 2, 1).$$

If $T \in \mathcal{T}_n$ then $a[1]$ is the father of vertex n, entry $a[2]$ is the father of $n-1$ and in general $a[i]$ is the father of vertex $n+1-i$ in the corresponding Prüfer sequence.

Corollary 2.1.1. *Prüfer sequence $P = (a[1], a[2], \ldots, a[n-2])$ corresponds to a RRT $T \in T_n$ if and only if $a[i] \in \mathbb{N}_{n-i}$*

Corollary 2.1.2.

$$|T_n| = (n-1)!$$

Note that $a[i]$ corresponds to the father of node $n-i$ in T_n and as therefore L is the set of leaves of T_n.

2.2 The space of random recursive trees

By Section 2.1 every RRT can be associated with a unique nested set of Prüfer sequences:

$$P_{n-1} \supset P_{n-2} \supset \cdots \supset P_3$$

where each $P_3 = (a[1])$ with $a[1] \in \mathbb{N}_2$ and subsequent P_i are built from P_{i-1} by the addition of $a[i-2] \in \mathbb{N}_i$. In fact $a[i]$ is the father of vertex $i+2$ for $i = 1, 2, \ldots, n-2$. The set, $\mathcal{T}_n$, of all RRTs on n vertices can be considered as the set $\mathcal{P}_n$ of all nested of Prüfer sequences of length n. We can make $\mathcal{P}_n$ a topological product space,

$$\mathcal{P}_n = \prod_{i=2}^{n-1} \mathbb{N}_i,$$

where each $\mathbb{N}_i$ is equipt with the discrete topology. We can further define a measure on $\mathcal{P}_n$ by appealing to the following Theorem of Tao [**?**]:

Theorem 1. *Let A be an arbitrary set. For each $\alpha \in A$ let $(\Omega_\alpha, \mathcal{F}_\alpha, P_\alpha)$ be a probability space such that Ω_α is a locally compact, σ-compact metric space with Borel σ-algebra $\mathcal{F}_\alpha$ then there exists a unique probability measure*

$$P_A = \prod_{\alpha \in A} P_\alpha \textit{ on } \left(\prod_{\alpha \in A} \Omega_\alpha, \prod_{\alpha \in A} \mathcal{F}_\alpha, \right)$$

Such that $P_A \left(\prod_{\alpha \in A} E_\alpha \right) = \prod_{\alpha \in A} P_\alpha(E_\alpha)$.

Furthermore, whenever $E_\alpha \in \mathcal{F}_\alpha$ one has $E_\alpha = \Omega_\alpha$ for all but finitely many α.

Corollary 2.2.1. *There exists a unique measure μ on $\mathcal{P}_n$.*

Proof. Let $A = \mathbb{N}_n$, then for each $\alpha \in \mathbb{N}_n$ let $X_\alpha = \mathbb{N}_\alpha$ with the discrete metric (and the discrete σ-algebra) and the uniform probability measure μ_α in the statement of Theorem 1. $\square$

For each $i \in \mathbb{N}_\alpha$ we have coordinate projection maps $\pi_i : \mathcal{P}_n \to \mathbb{N}_i$ defined by $\pi_1(P_n) := 1$ and $\pi_i(P_n) := a[i]$ for $i = 2, 3 \ldots, n-1$. Therefore a RRT can be interpreted as a sequence of random variables taking values in $\mathbb{N}_\alpha$ for $\alpha = 1, 2, \ldots n-1$. The measure $P_{\mathbb{N}_n}$ is a uniform probability measure on $\mathcal{T}_n$.

Chapter 3

Random Recursive Trees as Functions

Let $\mathcal{F}_n$ be the set of functions $f : \mathbb{N}_n \longrightarrow \mathbb{N}_n$ such that $f(1) = 1$ and $f(i) < i$ for $i = 2, 3, \ldots n$.

Lemma 3.0.2. *There is a bijection, $\alpha : \mathcal{T}_n \to \mathcal{F}_n$.*

Proof. Let $T = \{T_t\}_{t=1}^{n}$ be a RRT on n vertices. Construct $\alpha(T)$ by assigning $f(1) = 1$ and $f(i)$ to be the father of vertex i for $i = 2, 3, \ldots n$. On the other hand consider a function $f \in \mathcal{F}_n$; we construct the preimage under α of f as follows. Let T_1 be the tree on 1 vertex and 0 edges and build tree T_i from T_{i-1} by attaching vertex i to vertex $f(i)$ via an edge. This preimage is clearly unique, therefore α is a bijection. □

Corollary 3.0.3. $|\mathcal{T}_n| = (n-1)!$

Proof. Since $|\mathcal{T}_n| = |\mathcal{F}_n|$ it is enough to enumerate $\mathcal{F}_n$. Given any function $f \in \mathcal{F}_n$ by definition $f(i) \in \mathbb{N}_{i-1}$ for $i = 2, 3, \ldots, n$. We can think of this as 1 choice for $f(2)$, 2 choices for $f(3)$ and in general $i - 1$ possible choices for $f(i)$.

$$|\mathcal{F}_n| = (n-1)!$$

□

3.1 Expected automorphism group of a random recursive trees

In Section 1.2 we noted that the expected order of automorphism group of a labelled tree on could be calculated via the action of S_n on $\tilde{\mathcal{T}}_{\mathrm{n}}$ and applying Burnside's Lemma. This begs the question: is it possible to build a similar argument for the space of RRTs?

There does not exist an analogous group action of S_n on $\mathcal{T}_n$. For example, let T be the RRT shown in Figure 1.1. If we forget that T is a RRT and think of T as a labelled tree we can act T with S_n. The action of the transposition $(1,5)$ on T is shown in Figure 3.1 and is *not* a RRT since vertex 5 is attached to 3 vertices with smaller labels.

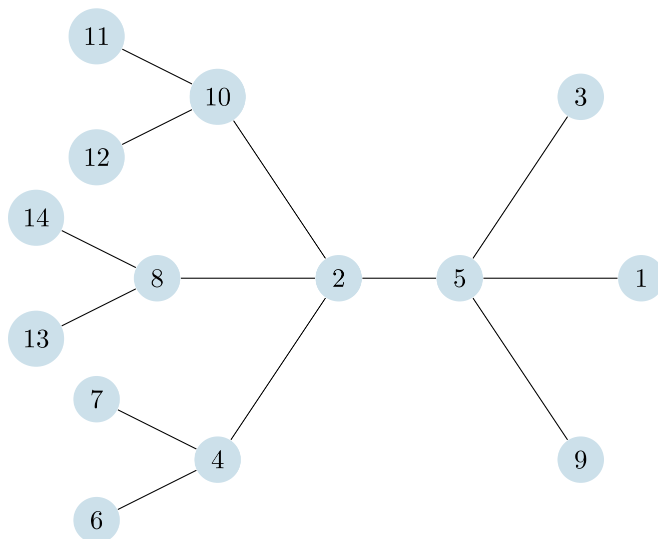

Figure 3.1: Tree $(1,5)\cdot T$

Even though S_n does not act on $\mathcal{T}_n$ we can define $\sigma \cdot T$ to be the tree such that $(\sigma(v),\sigma(w)) \in E(\sigma\cdot T)$ if and only if $(v,w)\in E(T)$. If $\sigma\cdot T\in\mathcal{T}_n$ then there exists a function $f'\in\mathcal{F}_n$ corresponding to $\sigma\cdot T$. We describe the precise form of f' in Lemma **??**.

Lemma 3.1.1. **??** *Let $T \in T_n$ correspond to $f\in\mathcal{F}_n$ and $\sigma\cdot T\in \mathrm{Aut}(T)$ correspond to function $f'\in\mathcal{F}_n$, then f_n has the following form:*

$$f' = \begin{pmatrix} 1 & \sigma(2) & \sigma(3) & \sigma(4) & \dots & \sigma(n) \\ 1 & \sigma(f(2)) & \sigma(f(3)) & \sigma(f(4)) & \dots & \sigma(f(n)) \end{pmatrix}$$

Proof. Let $T' = \sigma\cdot T$, there exists some function g corresponding to T' such that:

$$g = \begin{pmatrix} 1 & \sigma(2) & \sigma(3) & \sigma(4) & \dots & \sigma(n) \\ 1 & g(\sigma(2)) & g(\sigma(3)) & g(\sigma(4)) & \dots & g(\sigma(n)) \end{pmatrix}$$

Where $g(\sigma(i))$ is the father of $\sigma(i)$ but it is clear that the father of $\sigma(i)$ is $\sigma(f(i))$ hence $g(i)=\sigma(f(i))$ for $i=2,3,\dots,n$. $\square$

Corollary 3.1.2. *Let $T\in\mathcal{T}_n$ and $\sigma\in S_n$. Then $\sigma\cdot T\in\mathcal{T}_n$ if and only if $\sigma(f(i))<\sigma(i)$.*

Let $T \in T_n$. $\sigma \in S_n$ and recall the definition of the indicator function from Section 1.2:

$$I(\sigma, T) = \begin{cases} 1 & \text{if } \sigma \in \mathrm{Aut}(T) \\ 0 & \text{otherwise.} \end{cases}$$

The expected size of the automorphism group of a RRT is:

$$\frac{1}{|\mathcal{T}_n|} \sum_{T \in \mathcal{T}_n} |\mathrm{Aut}(T)| = \frac{1}{|\mathcal{T}_n|} \sum_{T \in \mathcal{T}_n} \sum_{\sigma \in S_n} I(\sigma, T) \tag{3.1.1}$$

$$= \frac{1}{|\mathcal{T}_n|} \sum_{\sigma \in S_n} \sum_{T \in T_n} I(\sigma, T) \tag{3.1.2}$$

$$\tag{3.1.3}$$

For each permutation $\sigma \in S_n$ we define:

$$J(\sigma) = \frac{1}{|\mathcal{T}_n|} I(\sigma, T)$$

In this section we will calculate $J(\sigma)$ when $\sigma = (p_1, p_2, \ldots, p_m)$ is a cycle of length $1 \leq m < n$ with the additional condition $p_1 < p_2 < \cdots < p_m$.

Remark 1. *The geometric decomposition of the group of automorphisms of a tree discussed in section 1.3 means that an automorphism $\sigma \in \mathrm{Aut}(T)$ corresponds to either an (n, k)-star or an extended symmetric branch of T. In particular, $\sigma = (p_1, p_2, \ldots, p_m) \in \mathrm{Aut}(T)$ if and only if $p_1, \ldots, p_m$ are the leaves of a $(1, m')$-star in T for some $m' > m$. Equivalently $\sigma \in \mathrm{Aut}(T)$ if and only if $f(p_1) = f(p_2) = \cdots = f(p_m)$ and $1 = \deg(p_1) = \deg(p_2) = \cdots = \deg(p_m)$, where $f \in F_n$ is the unique function associated with T.*

Lemma 3.1.3. *Consider the permutation $\sigma = (p_1, p_2) \in S_{n-1}$ such that without loss of generaity $p_1 < p_2$; then*

$$J(\sigma) = \frac{(p_1 - 1)}{(n-1)(n-2)}$$

Proof. By Lemma 2, a permutation $\sigma \in \mathrm{Aut}(T)$ if and only if $f(p_1) = f(p_2)$ and $\deg(p_1) = \deg(p_2) = 1$. Since every RRT $T \in T_n$ corresponds to a unique function $f \in F_n$ we can restate remark 2 as follows: permutation $\sigma \in Aut(T)$ if and only if $f(p_1) = f(p_2)$ and p_1, p_2 are *not* contained in the image of f.

Following the proof of Corollary 3.1.2, we can think of $f(i) \in \mathbb{N}_{i-1}$ for $i = 2, 3, \ldots, n$ as $i-1$ choices for each i. The additional restriction that $f(p_1) = f(p_2)$ means that there is exactly one choice to make for $f(p_2)$. The restriction that p_1,p_2 are not contained in the image of f means that there is 1 less choice for $f(i)$ where $i > p_1$ (apart from the case $i = p_2$)) and 2 less choices for $f(i)$ when $i > p_2$. The proportion of functions f such that that $\sigma \in \mathrm{Aut}(T)$ can be calculated with the aid of the following table:

$$\begin{pmatrix} \ldots & p_1 & p_1 + 1 & \ldots & p_2 - 1 & p_2 & p_2 + 1 & \ldots & n \\ \ldots & 1 & \frac{p_1 - 1}{p_1} & \ldots & \frac{p_2 - 3}{p_2 - 2} & \frac{1}{p_2 - 1} & \frac{p_2 - 2}{p_2} & \ldots & \frac{n-3}{n-1} \end{pmatrix} \tag{3.1.4}$$

The second row of table 3.1 describes the proportion of functions f that satisfy the aforementioned conditions for each $i \in 2, 3, \ldots, n$. We can therefore read off the table:

$$J(\sigma) = \frac{(n-3)!(p_1-1)}{(p_1-2)(n-1)!} \tag{3.1.5}$$

$$= \frac{(p_1-1)}{(n-1)(n-2)} \tag{3.1.6}$$

□

The method of proof in Lemma 3.1.3 can be generalised as follows:

Lemma 3.1.4. *Let $\sigma = (p_1, p_2, \ldots, p_m) \in S_n$ such that $p_1 < p_2 < \cdots < p_m$, then:*

$$J(\sigma) = \frac{(p_1-1)(n-m-1)!}{(n-1)!}$$

Proof. Since each RRT, $T \in \mathcal{T}_n$ corresponds to a function $f \in \mathcal{F}_n$ we need only consider automorphisms of $\mathcal{F}_n$. By Remark 2, the permutation $\sigma \in \mathrm{Aut}(T)$ if and only if $f(p_1) = f(p_2) = \ldots f(p_m)$ and $p_1, p_2, \ldots, p_m$ are not contained in the image of the corresponding function f. Following the proof of Lemma 3.1.3, consider the following table:

$$\begin{pmatrix} \ldots & p_1 & p_1+1 & \ldots & p_2-1 & p_2 & p_2+1 & \ldots & p_m-1 & p_m & p_m+1 & \ldots & n \\ \ldots & 1 & \frac{p_1-1}{p_1} & \ldots & \frac{p_2-3}{p_2-2} & \frac{1}{p_2-1} & \frac{p_2-2}{p_2} & \ldots & \frac{p_m-m-1}{p_m-2} & \frac{1}{p_m-1} & \frac{p_m-m}{p_m} & \ldots & \frac{n-m-1}{n-1} \end{pmatrix}$$

Following the proof of Lemma 3.1.3 we can read off the table to ascertain:

$$J(\sigma) = \frac{(n-m-1)!(p_1-1)!}{(p_1-2)!(n-1)!} \tag{3.1.7}$$

$$= \frac{(n-m-1)!(p_1-1)}{(n-1)!} \tag{3.1.8}$$

□

Let $H_n(m)$ be the subset of S_n consisting of all permutations $\sigma = (p_1, p_2, \ldots, p_m)$ such that $p_1 < p_2 < \cdots < p_m$ and define the number of automorphisms coming from all such permutations as:

$$X_n(m) = \sum_{\sigma \in H_n(m)} J(\sigma)$$

Lemma 3.1.5. $X_n(2) = \frac{n}{3!}$

Proof. We can usefully lay out all possible permutations $\sigma \in H_n(2)$ as follows:

$$\begin{array}{c|c|c|c|c} (2,3) & (2,4) & (2,5) & \ldots & (2,n) \\ & (3,4) & (3,5) & \ldots & (3,n) \\ & & (4,5) & \ldots & (4,n) \\ & & & & \vdots \\ & & & & (n-1,n) \end{array}$$

Summing over rows and then columns and applying Lemma 3.1.3 we see that:

$$X_n(2) = \frac{1}{(n-1)(n-2)} \sum_{i=1}^{n-2} i \sum_{j=1}^{n-i-1} 1 = \frac{n}{6} \tag{3.1.9}$$

$\square$

Lemma 3.1.6. $X_n(2) = \frac{n}{4!}$

Proof. We can usefully lay out all possible permutations $\sigma \in H_n(3)$ in $n-3$ "triangles" as follows:

$$\begin{array}{c|c|c|c|c}
(2,3,4) & (2,3,5) & (2,3,6) & \dots & (2,3,n) \\
 & (2,4,5) & (2,4,6) & \dots & (2,4,n) \\
 & & (2,5,6) & \dots & (2,5,n) \\
 & & & & \vdots \\
 & & & & (2,n-1,n) \\
 & & & & \\
 & (3,4,5) & (3,4,6) & \dots & (3,4,n) \\
 & & (3,5,6) & \dots & (3,5,n) \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

Note that all permutations in a triangle take the same value for p_1. Hence summing over each triangle and a simple application of Lemma 3.1.4 gives us:

$$X_n(2) = \frac{1}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n-3} i \sum_{j=1}^{n-i-2} j = \frac{n}{4!} \tag{3.1.10}$$

$\square$

We define the *generalised triangular numbers* as follows: let $T_i^{(2)} = 1$ for $i = 1, 2, 3, \dots$, and define $T_i^{(m)} = \sum_{j=1}^{i} T_j^2$. This means that $T^{(3)}$ are the natural numbers, $T^{(4)}$ are the triangular numbers, $T^{(5)}$ are the tetrahedral numbers and so on. We can reformulate equations 3.1.9 and 3.1.10 as follows:

$$X_n(2) = \frac{1}{(n-1)(n-2)} \sum_{i=1}^{n-2} i \sum_{j=1}^{n-i-1} T_j^{(2)} \tag{3.1.11}$$

$$X_n(3) = \frac{1}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n-3} i \sum_{j=1}^{n-i-2} T_j^{(3)} \tag{3.1.12}$$

Following the proofs of Lemma 3.1.5 and 3.1.6 we can think of all possible permutations $\sigma \in H_n(m)$ in $n-m$, "$m-1$-simplicies" such that each $m-1$-simplex has the same value of p_1. We use Lemma 3.1.4 to get:

$$X_n(m) = \frac{(n-m-1)!}{(n-1)!} \sum_{i=1}^{n-m} i \sum_{j=1}^{n-i-m+1} T^{(m)}, \tag{3.1.13}$$

3.1.1 Pascal's triangle: an aside

$n = 0$:					1				
$n = 1$:				1		1			
$n = 2$:			1		2		1		
$n = 3$:		1		3		3		1	
$n = 4$:	1		4		6		4		1

Figure 3.2: Pascal's Triangle

for $m = 1, 2, \ldots n-1$.

In order to find a general form for $X_n(m)$ we will split it up into constituent parts $\frac{(n-m-1)!}{(n-1)!}$ (which we can calculate) and

$$X'_n(m) = \sum_{i=1}^{n-m} i \sum_{j=1}^{n-i-m+1} T^{(m)} \tag{3.1.14}$$

To calculate $X'_n(m)$ we must make some preliminary definitions regarding binomial coefficients and, although the suject is extremely elementary, it is appropriate to discuss Pascal's triangle because it draws together many themes of this chapter. Pascal's triangle, is an array of binomial coefficients such that the rows are enumerated from $n = 0$ at the top and the columns are enumerated from left to right beginning with $k = 0$. Row n has $n+1$ columns and the entry in the n^{th} row and k^{th} column is:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

The first 5 rows of Pascal's triangle are shown in figure 3.2. Note that *nchosek* is also the number of ways (disregarding order) that k objects can be chosen from among n objects. There is a deep connection between generalised triangular numbers and binomial coefficients []. For example, reading diagonally downwards in either direction of Pascl's triangle are the sets $T^{(2)}, T^{(3)}, T^{(4)} \ldots$ and note that:

$$\sum_{i=1}^{n} T_i^{(m)} = \binom{n+(m-2)}{(m-1)} \tag{3.1.15}$$

Finally, for the readers amusement, recall that summing the upward diagonals gives the Fibonacci sequence.

3.2 Expected automorphism group of a random recursive tree continued

By Equation 3.1.15 we can rewrite equation 3.1.14 as:

$$X'_n(m) = \sum_{i=1}^{n-m} i\binom{n-1-i}{m-1}$$

In order to prove Lemma 3.2.1 we will also need the following identity:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \tag{3.2.1}$$

Lemma 3.2.1. $X'_n(m) = \binom{n}{m+1}$ *for* $m = 2, 3, \ldots, n-1$

Proof. We will use induction on n. For the base case consider $n = 3$ and $m = 2$.

$$X'_3(2) = \sum_{i=1}^{3-2} i\binom{3-i-1}{1} = \binom{1}{1} = 1 = \binom{3}{3}.$$

Assume $X'_n(m) = \binom{n}{m+1}$ for $m = 2, 3, \ldots, n-1$ and consider $X_{n+1}(m)$. In the case that $m = n$,

$$X'_{n+1}(n) = \sum_{i=1}^{1} i\binom{m-i}{m-1} = 1 = \binom{m+1}{m+1}$$

On the other hand if $m = 1, 2, 3, \ldots, n-1$ we use the inductive hypothesis and identity 3.2.1 as follows:

$$\begin{align}
X'_{n+1}(m) &= \sum_{i=1}^{n+m} i\binom{n-i}{m-1} + (n+1-m) \tag{3.2.2}\\
&= \sum_{i=1}^{n-m} i\binom{n-i-1}{m-1} + \sum_{i=1}^{n-m} i\binom{n-i-1}{m-2} + (n+1-m) \tag{3.2.3}\\
&= X'_n(m) + (n+1-m) + \sum_{i=1}^{n+1-m} i\binom{n-i-1}{m-2} - (n+1-m) \tag{3.2.4}\\
&= X'_n(m) + X'_n(m-1) \tag{3.2.5}\\
&= \binom{n}{m+1} + \binom{n}{m} \tag{3.2.6}\\
&= \binom{n+1}{m+1} \tag{3.2.7}
\end{align}$$

□

Corollary 3.2.2. $X_n(m) = \frac{n}{(m+1)!}$

Proof.

$$X_n(m) = \frac{(n-m-1)!}{(n-1)!} \sum_{i=1}^{n-m} i \sum_{j=1}^{n-i-m+1} T^{(m)} \tag{3.2.8}$$

$$= \frac{(n-m-1)!}{(n-1)!} X'_n(m) \tag{3.2.9}$$

$$= \frac{(n-m-1)!}{(n-1)!} \binom{n}{m+1} \tag{3.2.10}$$

□

3.2.1 The expected number of k-stars

Recall that $A_n(k)$ be the expected number of k-stars in a RRT on n vertices. In this section we will calculate $A_n(m)$ in terms of $X_n(m)$. By Remark 2 given a permutation $\sigma = (p_1, p_2, \dots p_m) \in H_n(m)$ and a tree $T \in \mathcal{T}_n$ the permutation $\sigma \in \mathrm{Aut}(T)$ if and only if vertices $p_1, p_2, \dots, p_m$ are leaves *and* $f(p_1) = f(p_2) = \cdots = f(p_m)$. Equivalently, $\sigma \in \mathrm{Aut}(T)$ if and only if T has an induced subtree isomorphic to an m'-star for some $m' > m$ with $p_1, p_2, \dots, p_m$ leaves. Since T cannot contain an m' star for $m' \geq n$ we must have:

$$A_n(n-1) = X_n(n-1)$$

Note that if T has an induced subtree isomorphic to an m-star then this corresponds to $\binom{m}{k}$ ordered permutations $\sigma \in \mathrm{Aut}(T)$. In conclusion we have the following:

$$A_n(m) = X_n(m) - \binom{m+1}{m} A_n(m+1) - \binom{m+2}{m} A_n(m+2) - \cdots - \binom{n-1}{m} A_n(n-1) \tag{3.2.11}$$

Lemma 3.2.3. $A_n(n-k) = \sum_{i=0}^{k-1} \binom{n-k+i}{n-k} X_n(n-k+i)$

Proof. Use induction on k. We have already seen the case $k = 1$ so assume the following for $k = 1, 2, 3, \dots K$:

$$A_n(n-k) = \sum_{i=0}^{k-1} \binom{n-k+i}{n-k} X_n(n-k+i). \tag{3.2.12}$$

□

Chapter 4

Random Recursive Trees and Permutations

4.1 Permutations

A permutation, σ, of size n is a bijective mapping of the set $\mathbb{N}_n$ and can be represented by an array:

$$\sigma = \left(\begin{smallmatrix} 1 & 2 & \dots & n \\ \sigma_1 & \sigma_2 & \dots & \sigma_n \end{smallmatrix} \right)$$

or in *one-line notation* as $\sigma = \sigma_1 \sigma_2 \dots \sigma_n$. We denote the group of permutations of $\mathbb{N}_n$ (with composition of maps as the operation) by S_n. This group is called the symmetric group of n elements. We can also write a permutation $\sigma \in S_n$ as disjoint cycles that partition $\mathbb{N}_n$. Note that the order among the cycles is irrelevant so $(1, 2, 3)(4, 5) = (4, 5)(1, 2, 3)$ and there are k ways of writing a cycle of length k i.e.

$$(a_1, a_2, \dots, a_k) = (a_2, a_3, \dots, a_k, a_1) = \dots = (a_k, a_1, \dots, a_{k-1})$$

By writing the smallest element of each cycle last, then arranging the cycles in increasing order of their last elements the representation of any permutation in disjoint cycle form is unique. This way of writing permutations is called *canonical cycle notation*.

For example, consider the permutation

$$\sigma = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{smallmatrix} \right)$$

We write permutation $\sigma = 2314$ in one-line notation and $\sigma = (2, 3, 1)(4)$ in canonical cycle notation.

4.2 Random permutations and Chinese restaurants

Consider a sequence of permutations, $\{\sigma_m\}_{m=1}^{n}$ such that:

(i) Each permutation σ_m is a uniformly distributed permutation of $\mathbb{N}_m$ for $m = 1, 2, \ldots, n$.

(ii) If σ_m is written as a product of cycles then σ_{m-1} is derived by the deletion of element m from its cycle. For example, if $\sigma_7 = (1)(542)(673)$ then $\sigma_6 = (1)(542)(63)$.

We call these sequences *consistent random permutations*. The set of all consistent permutations of length n is denoted Σ_n.

A Chinese restraint process is a dynamic combinatorial process redolent of a RRT which partitions $\mathbb{N}_n$. Consider an evening in which n (labelled) customers patronize a restaurant according to the following probabilistic process:

(i) Customer 1 sits at some table.

(ii) Subsequently with probability $\frac{1}{m}$ customer m starts a new table and with probability $\frac{1}{m}$ customer m sits to the left of each of the $m-1$ existing customers.

If the total number of people who enter the restaurant in an evening is n the tables partition $\mathbb{N}_n$. Further, a Chinese restaurant process can also be thought of as a consistent permutation where each table corresponds to a cycle and if i is directly to the left of j in a cycle then $\sigma(i) = j$ [**?**].

Remark 2. *Consider some Chinese restaurant process that corresponds to a consistent permutation $\sigma = \{\sigma_t\}_{t=1}^n$. Since each σ_t is written in disjoint cycle notation we can insist that each σ_t is written in canonical cycle notation. Now imagine that at time t customer t sits directly to the left of customer $f(t)$; by the construction $t > f(t)$. At subsequent times, any customer that sits between t and $f(t)$ has label $t' > t$. This means that if $f(t)$ is the first customer t meets if he walks to his right down the table such that $t > f(t)$.*

Let $T = \{T_t\}_{t=1}^n$ be a RRT and for convenience we define $V(T) = \{0, 1, 2, \ldots, n-1$; tree T can be associated with a Chinese restaurant process via the following process:

(i) Vertex 0 *does not* appear in the process.

(ii) At time $t = 1$ vertex 1 sits at a table and at subsequent times t, vertex t sits to the left of its father $f(t)$.

Remark 3.

Lemma 4.2.1. *There is a bijection between $\mathcal{T}_n$ and Σ_{n-1}.*

Proof. We have seen how to associate a consistent random permutation with a RRT. Note that given any consistent permutation we can construct a unique RRT via Remark 2. □

4.3 Random permutation facts

In this section we state some prominent theorems from the theory of random permutations and translate these into results about RRTs.

Let $c_1, c_2, \dots, c_n$ be integers ≥ 0 such that

$$c_1 + 2c_2 + \cdots + nc_n = n$$

A permutation $\sigma \in S_n$ is said to have type $\mathcal{C} = [c_1, c_2, \dots, c_n]$ if the decomposition of σ into disjoint cycles contains exactly c_i cycles of length i for $i = 1, 2, 3, \dots n$.

Theorem 4.3.1. *The number of permutations of type $\mathcal{C}$ is*

$$|\mathcal{C}| = \frac{n!}{\prod_{i=1}^{n} i^{c_i}(c_i!)}$$

Proof. The number of permutations of type $\mathcal{C}$ is equal to the size of the conjugacy class of any permutation $\sigma \in S_n$ of type $\mathcal{C}$. The result follows from [**?**]. $\square$

Theorem 4.3.2. *The number of Permutations of $\mathbb{N}_n$ whose decomposition has k cycles equals the unsigned Stirling number of the first kind $s(n,k)$.*

Proof. See [[**?**], Chapter 6]. $\square$

An ascent (also called a rise) of permutation $\sigma = \sigma_1\sigma_2\dots\sigma_n$ is a pair of consecutive elements (σ_i, σ_{i+1}) such that $\sigma_i < \sigma_{i+1}$ (with $1 \leq i < n$) [**?**]. For example if $\sigma = 23417586$ then the pairs $(2,3), (3,4), (1,7)$ and $(5,8)$ are all ascents of σ.

Lemma 4.3.3. *The mean number of ascents in a random permutation of size n is $\frac{1}{2}(n-1)$.*

Proof. See [**?**]. $\square$

For a fixed parameter l, an ascending run of length l is a sequence of consecutive elements $\sigma_i\sigma_{i+1}\dots\sigma_{i+l}$ such that

$$\sigma_i < \sigma_{i+1} < \cdots < \sigma_{i+l}.$$

Lemma 4.3.4. *The mean number of ascending runs of length $l-1$ in a random permutation of size n is*

$$a(n,l) = \frac{n-l+1}{l!}$$

for $2 \leq l \leq n-1$.

Proof. See [**?**] $\square$

Let $a(n, l, \sigma)$ be the number of ascending runs of length $l - 1$ in a specific permutation $\sigma \in S_n$. Then by $a(n, l)$ we mean:

$$a(n, l) = \frac{\sum_{\sigma \in S_n} a(n, l, \sigma)}{n!}$$

Note that an ascending run of length $(l - 1)$ also contains 2 (overlapping) ascending runs of length $l - 2$, 3 ascending runs of length $l - 3$ and in general i ascending runs of length $l - i$ where $2 \leq i \leq i - 2$. For example consider the permutation $\sigma = 25134768$ which contains the ascending run (1347) of length 3. Then σ also contains ascending runs (134) and (347) of length 2 and ascending runs $(13), (34)$ and (47) of length 1. We can define an *exact* ascending run of length $l - 1$ is an ascending run $\sigma_i \sigma_{i+1} \ldots \sigma_{i+l}$ such that:

$$\sigma_i < \sigma_{i+1} < \cdots < \sigma_{i+l}$$

with the additional condition that (if they exist) $\sigma_{i-1} > \sigma_i$ and $\sigma_{i+l} > \sigma_{i+l+1}$. We write $A(n, l)$ for the mean number of ascending runs of length exactly $(l-1)$ and let $A(n, l, \sigma)$ be the number of ascending runs of length exactly $l-1$ in some specific permutation $\sigma \in S_n$. Then by $A(n, l)$ we mean:

$$A(n, l) = \frac{\sum_{\sigma \in S_n} A(n, l, \sigma)}{n!}$$

Since the longest ascending run in a permutation of length n is $n - 2$ we have $A(n - 1, n, \sigma) = a(n - 1, n, \sigma)$ for all $\sigma \in S_n$. By the argument outlined above $A(n - 2, n, \sigma) = a(n - 2, n, \sigma) - 2A(n - 1, n, \sigma)$ for all $\sigma \in S_n$.

Lemma 4.3.5. *For all $\sigma \in S_n$ and $2 \leq l \leq n-3$ we have the following relation:*

$$A(n, l, \sigma) = a(n, l, \sigma) - 2a(n, l + 1, \sigma) + a(n, l + 2, \sigma)$$

Proof. We will prove Lemma 4.3.5 by induction on l. For the base case note that

$$\begin{aligned} A(n - 3, n, \sigma) &= a(n - 3, n, \sigma) - 2A(n - 2, n, \sigma) - 3A(n - 1, n, \sigma) && (4.3.1) \\ &= a(n - 3, n, \sigma) - 2(a(n - 2, n, \sigma) - 2a(n - 1, n, \sigma)) - 3a(n - 1, n, \sigma) && (4.3.2) \\ &= a(n - 3, n, \sigma) - 2a(n - 2, n, \sigma) + a(n - 1, n, \sigma) && (4.3.3) \end{aligned}$$

Now assume that

$$A(n, l, \sigma) = a(n, l, \sigma) - 2a(n, l + 1, \sigma) + a(n, l + 2, \sigma)$$

for all $k \leq l \leq n - 3$ and consider

$$\begin{aligned} A(n, k - 1, \sigma) &= a(n, k - 1, \sigma) - \sum_{l=k}^{n} (l - (k - 2))A(n, l, \sigma) && (4.3.4) \\ &= a(n, k - 1, \sigma) - 2a(n, k, \sigma) + a(n, k + 1, \sigma) && (4.3.5) \end{aligned}$$

□

Therefore $A(n,l) = a(n,l) - 2a(n,l+1) + a(n,l+2)$

Corollary 4.3.6. *The mean number of ascending runs of length* exactly $l-1$ *in a random permutation of size* n *is*

$$A(n,l) = \frac{n-l+1}{l!} - 2\frac{n-l}{l+1!} + \frac{n-l-1}{l!}$$

4.4 Applications to random recursive trees

We have seen that any RRT $\{T_t\}_{t=1}^{n}$ contains a Chinese Restaurant Process. It is clear from this construction that the number of occupied tables in the Chinese Restaurant Process at time t is the degree of vertex $v_0 \in V(T_{t+1})$. We also discussed a further bijection between Chinese Restaurant Processes with n guests and consistent random permutations $\{\sigma_t\}_{t=1}^{n}$. The number of tables in the restaurant corresponds to the number of cycles in σ_n by this bijection. If σ consists of m cycles of lengths $l_1, l_2, \ldots, l_m$ then the associated Chinese Restaurant Process consists of m occupied tables with $l_1, l_2, \ldots, l_m$ guests each. Furthermore, by Theorem 4.3.1 the number of RRTs with adjacent to c_i branches with orders $l_1, l_2, \ldots, l_m$ is

$$\frac{n!}{\prod_{i=1}^{n} i^{c_i}(c_i!)}$$

Similarly, by Theorem 4.3.2, the number of Permutations of RRTs of order $n+1$ with $\deg(v_0) = k$ equals the unsigned Stirling number of the first kind $s(n,k)$.

4.5 Theory of permutations

In Section 4.4 we successfully translated Theorem 4.3.1 and Theorem 4.3.2 into results about RRTs. In order to do the same with Lemma 4.3.3 and Lemma 4.3.4 we require a brief detour into the theory of permutations.

Let $\mathcal{P}$ be a permutation of $\mathbb{N}_n$ written in canonical cycle notation then we define the map $f : S_n \to S_n$ by letting $f(\mathcal{P})$ be the permutation written in one line notation by omitting parenthesis. For example $f[(2)(5134)(76)] = 2513476$

Lemma 4.5.1. *The map* $f : S_n \to S_n$ *is a bijection.*

Proof. See [**?**]. ∎

Definition 4.5.2. We say that i is a weak excedance of permutation $\sigma = \sigma_1\sigma_2\ldots\sigma_n$ if $\sigma_i \geq i$ [**?**].

For example if $\sigma = 253416$ then $1, 2, 3, 4$ and 6 are weak excedances.

Lemma 4.5.3. *The bijection* $f : S_n \to S_n$ *described in Lemma 4.5.1 maps permutations with* k *weak excedances to permutations with* $k-1$ *ascents.*

Proof. See [**?**] ∎

4.6 Further applications

Lemma 4.6.1. *Let σ be the permutation corresponding to some RRT T_n. The number of weak excedances in σ is equal to the number of leaves in T_n*

Proof. If $\sigma_i = i$ then by the discussion above the corresponding vertex v_i is contained in a branch of order 1 this is the case if and only if v_i is a leaf.

Assume that vertex $v_i \in V(T_{n+1})$ is a leaf. Then at time i for some $i < n+1$ guest i arrived at the Restaurant and sat to the left of its parent $\mathcal{P}(i)$. By construction the seat to the left customer i is either not taken, or taken by a younger sibling, $\mathcal{S}(i)$, of i or by a younger brother, $\mathcal{U}(i)$ of $\mathcal{P}(i)$. Let $\mathcal{M}(i)$ be the customer with the least label sitting at the same table as customer i. If the seat to the left of i is empty then $\sigma_{\mathcal{M}(i)} = i$. If the seat is taken by a younger sibling $\mathcal{S}(i)$ then $\sigma_{\mathcal{S}(i)} = i$. If the seat is taken by younger brother, $\mathcal{U}(i)$ of $\mathcal{P}(i)$ then $\sigma_{\mathcal{U}(i)} = i$.

Conversely let $i \in [n]$ such that $i < \sigma_i$ and assume that σ_i is not a leaf in T_{n+1}. By construction i is the eldest child of σ_i so $i > \sigma_i$ which is a contradiction. □

Corollary 4.6.2. *The mean number of leaves on a RRT T_{n+1} is $\frac{n+1}{2}$.*

Proof. Mean number of leaves,

$$\frac{n-1}{2} + \frac{\sum_{\sigma \in S_n} 1}{|S_n|}$$

□

Chapter 5

A Pólya Urn

A generalised Pólya urn, $\mathcal{U}$, contains a finite number of balls of finitely many possible types $1, 2, \ldots, q$. The content of the urn at time t is described by a vector $X_t = (X_{t,1}, X_{t,2}, \ldots X_{t,q})$ where each $X_{t,i}$ is the number of balls of type i in the urn at time t. We associate an *activity* a_i and a transition vector $\xi_i = (\xi_{i,1}, \xi_{i,2}, \ldots, \xi_{i,q})$ to each type. A generalised Pólya urn is an evolving process with initial content X_0 and at subsequent times a ball is drawn from the urn. We also assume that $\mathbb{E}|X_0|^2$. At time t a ball of type i is drawn with probability

$$P_{i,t} = \frac{a_i X_{t-1,i}}{\sum_{j=1}^{q} X_{t-1,j}}$$

The drawn ball is then returned to the urn with ξ_{ij} balls of type j for $j = 1, 2, \ldots, q$. In general we also require the following further conditions:

$$\xi_{i,j} \geq 0 \text{ if } j \neq i \tag{5.0.1}$$

$$\xi_{i,i} \geq -1 \tag{5.0.2}$$

Assume that a ball of type i is chosen from an urn, Equation 5.0.1 ensures that no balls of a type *other* than i is removed from the urn and Equation 5.0.2 means that at most one ball of type i is removed from the urn. Together these conditions prevent the removal of a ball that does not exist from a generalised Pólya urn.

Remark 4. *If $a_i = 1$ for every type i then at each time a ball is drawn from the urn it has a uniformly random type.*

5.1 Properties of Pólya urns

The crux of this section is Theorem 2 which is a limit theorem for Pólya urns. In order to state this theorem we require some further concepts from urn theory.

To every Pólya urn we can associate a matrix A such that the j^{th} column of A is defined to be $a_j \mathbb{E}(\xi_j)$ where $\mathbb{E}$ denotes expectation.

We will write $i \succ j$ if it is possible to find a ball of type j in an urn beginning with a single ball of type i. We say that a type i is *dominating* if $i \succ j$ for every type $j = 1, 2, \ldots q$. Note that $\succ$ is a transitive and reflexive relation so it partitions the set of types in to equivalence classes $C_1, C_2, \ldots, C_r$ where types $i, j \in C_k$ if $i \succ j$ and $j \succ i$. We say that some class C_k is dominating if some $i \in C_k$ is dominating.

We say that a generalised Pólya urn becomes *essentially extinct* if at some time t there does not exist a ball of dominating type.

Theorem 2. *Let A be the matrix associated to some non-essentially extinct Pólya urn process such that:*

(A1) Equations 5.0.1 and 5.0.2 are satisfied.

(A2) $\mathbb{E}(\xi_{ij}) < \infty$ for all $i, j = 1, 2, \ldots, q$.

(A3) There exists a largest real eigenvalue λ_1 of A which is positive.

(A4) λ_1 is simple.

(A5) There exists a dominating type, i, and $X_{0,i} > 0$.

(A6) λ_1 belongs to dominating type.

Let v_1 be the right eigenvector associated with λ_1 then:

$$n^{-1} X_n \to \lambda_1 v_1 \text{ almost surely as } n \to \infty$$

Janson uses Theorem 2 to prove that if $X_{n,i}$ is the number of vertices with outdegree $i (\geq 0)$ in a RRT on n vertices then:

Theorem 3. *In the limit as $n \to \infty$, $\frac{X_{n,i}}{n} \to 2^{-i-1}$ almost surely.*

Theorem 3 is an example of the strong law for large numbers. Loosely speaking, given any infinite RRT, we can think of $\frac{X_{n,i}}{n}$ as a trajectory which gets ever closer to the expected outdegree of 2^{-i-1}.

5.2 A specific Pólya urn

??

In this section we will describe a generalized Pólya urn , $\mathcal{U}_m$ that describes the distribution of random recursive d-ary tree motifs. Subsequently we will hit $\mathcal{U}_m$ with Theorem 2 to prove a limiting theorem for network motifs.

Let $\{T_t\}_{t=1}^n$ be a random recursive d-ary tree process. At time t any vertex $v \in V(T_n)$ is incident to l leaves where $0 \leq l \leq d$. We call the induced subtree of T_t with a hub node adjacent to l leaves such that each leaf has higher label than the hub an l-star (an l-star is an example of a network motif). By taking l to be maximal partition $V(T_t)$ can be partitioned into $d + 1$ sets of vertices contained in induced l-stars for $l = 1, 2, 3, d$ and the set of vertices not contained in any l-star.

Balls in $\mathcal{U}_m$ may take one of $d+1$ types that correspond to the aforementioned partition. In particular types $2, 3, \ldots d+1$ correspond to vertices in $(d-1)$-stars and type 1 corresponds to the remaining vertices (for convenience we will refer to these vertices as 0-stars). Since an l-star contains $l+1$ vertices we set activities $a_i = i$ for $i = 1, 2, \ldots d$ and $a_{d+1} = d$ for the reasons we give below.

At time $n+1$ vertex v_{n+1} is attached to i is attached via an edge to a vertex $v \in V(T_n)$. In order that the distribution of network motifs is the same as the distribution of types the probability that v is contained in an induces $i+1$-star must be the same as the probability that a ball of type i is drawn from $\mathcal{U}_m$ at time $n+1$. Furthermore, we claim that the appropriate transition vectors are as follows:

$$
\begin{aligned}
\xi_1 &= (-1, 1, 0, \ldots, 0) \\
\xi_2 &= (\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0) \\
\xi_3 &= (-0, \frac{4}{3}, -1, \frac{1}{3}, 0, \ldots, 0) \\
\xi_i &= (0, \frac{i-1}{i}, 0, \ldots, 0, \frac{i-1}{i}, -1, \frac{1}{I}, O, \ldots, 0) \\
\xi_{d+1} &= (0, 1, 0, \ldots, 0, 1, -1)
\end{aligned}
$$

of claim. For $i = 2$ assume that T_n and X_n are as usual and that one draws a ball of type 2 from U_n. Then equivalently at time $n+1$ one attaches vertex v_{n+1} via an edge to an induced 1-star of T_n for clarity let that 1-star have hub h and leaf l. With probability $\frac{1}{2}$ vertex v_{n+1} is attached to h, in which case a 2-star is created. Similarly with probability $\frac{1}{2}$ vertex v_{n+1} is attached to l in which case a 1-star and a 0-star are created. See Diagram for further details. Vectors ξ_i are built in an analogous way for $i = 3, 4, \ldots, d$. If a ball of type $d+1$ is drawn from $\mathcal{U}_m$ we could equivalently imagine that a vertex is connected via an edge to a d-star. Since each T_n is a d-ary tree $v_n + 1$ *must* be attached to a leaf of that d-star (this is why we set $a_d + 1 = d$. $\square$

5.3 Further background

In order to prove Theorem 4 we require further background from linear algebra such as the Gershgorin Circle Theorem (Theorem 5.3.1) and a short introduction to Perron-Frobenius theory.

5.3.1 The Gershgorin circle Theorem

Theorem 5.3.1. *Gershgorin circle theorem Let $A \in M_n(\mathbb{C}$, and define*

$$R_i = \sum_{i=1, i \neq j}^{n} |a_{ij}|$$

Then each eigenvalue of A is in at least one of the disks

$$D_i = \{z : |z - a_{ii}| \leq R_i$$

Somewhat surprisingly the Gershgorin circle theorem gives us a bound on the values of the eigenvalues; informally it tells us that the eigenvalues cannot be too far from the diagonal elements of A.

5.3.2 Non-negative matrices

The reference for this section is [**?**]. A square matrix $A = \{a_{ij}\}$ is said to be *non-negative* if every element $a_{ij} \geq 0$ and we write $A \geq 0$. We say that a non-negative matrix A is *irreducible* if for every a_{ij} there exists some $n \in \mathbb{N}$ such that $(a_{ij})^n > 0$.

A square matrix $A = \{a_{ij}\}$ is said to be Metzler if for all $i \neq j$, $a_{ij} \geq 0$. A Metzler matrix $A \in M_n(\mathbb{C})$ is related to a non-negative matrix T by:

$$T = \mu I + A$$

for some large enough $\mu \in \mathbb{R}$.

Definition 5.3.2. A Metzler matrix A is is said to be irreducible if the related non-negative matrix T is irreducible.

The theory of non-negative matrices has been widely studied and Perron-Frobenius theory yields the following theorem.

Theorem 5.3.3. *Let A be an irreducible Metzler square matrix.*

- *(i) Matrix A has an eigenvalue λ_1 such that $\lambda_1 \in \mathbb{R}$.*
- *(ii) λ_1 is associated with strictly positive left and right eigenvectors.*
- *(iii) $\lambda_1 > Re(\lambda)$ for any other other eigenvalue λ of A.*
- *(iv) λ_1 is simple.*

Theorem 5.3.4. *[**?**] Let $A = \{a_i j\}$ be an $n \times n$ matrix. Associate to A a directed graph G_A on n vertices with a directed edge from i to j whenever $a_{ij} > 0$. Then A is irreducible if and only if G_A is strongly connected.*

Theorem 4. *If $A_{d+1} = \{a_i j\}$ be the $(d+1) \times (d+1)$ matrix associated with the generalized Pólya urn U_m then A_{d+1} satisfies (A1)-(A6).*

Proof. (A1) True by construction.

(A2) true since d is finite $\xi_{ij} \leq d+1 < \infty$.

(A3) Clearly 1 is an eigenvalue of A_{d+1} with left eigenvector $u_1 = (1, 1, \ldots, 1)$. We claim that 1 is the largest real eigenvector of A_q and to prove this claim we appeal to the Gershgorin Circle Theorem. Note that for each column in A_q, the sum of the diagonal elements:

$$C_1 = 2C_2 \qquad = 2C_i = \sum_{j \neq i} |a_{ij}| = i + 1 \text{ if } 3 \leq i \leq d+1$$

Let $D_i = D(a_{ii}, C_i)$ be the disk in $\mathbb{C}$ centred at a_{ii} with radius C_i. Note that $D_1 = D(-1, 2)$, $D_2 = D(-1, 2)$ and $D_i = (-i, i+1)$ for $i = 3, 4, 5, \ldots, d+1$. Let $\mathcal{D} = \bigcup_i D_i(a_{ii}, C_i)$ (see Figure **??** for an image of $\mathcal{D}$). The largest real number in D is clearly 1, therefore by the Gergorin Circle Theorem the largest real eigenvalue λ_1 of A_{d+1} is 1.

(A4) Since no off diagonal element of A_{d+1} is negative A_{d+1} is a Metzler matrix. Therefore A_{d+1} can be associated with a non-negative matrix T by writing:

$$T_\mu = \mu I + A_{d+1}$$

For some choice of $\mu \in \mathbb{R}$. We make the choice $\mu = d+1$ so that $T_{d=1}$ is a non-negative matrix. By Theorem 5.3.2 A_{d+1} is irreducible if T_{d+1} is irreducible.

Note that all the sub diagonal and superdiagonal entries of $T_d + 1$ are positive so the graph G_A built in the way described in Theorem 5.3.4 is strongly connected so A_{d+1} is irreducible. By Theorem 5.3.3 1 is a simple eigenvalue.

(A5) True since A_{d+1} is irreducible and $X_0 \neq 0$.

(A6) True since A_{d+1} is irreducible and $X_0 \neq 0$.

□

5.4 Complex automorphisms are non-trivial

Let $\{T_t\}_{t=1}^n$ be a random recursive tree and X_n be the number of trees with leaves isomorphic to a $(2,2)$-star. If we assume that almost surely $lim_{n\to\infty} \frac{X_n}{n} \to \epsilon_{2,2}$ for some $\epsilon_{2,2} > 0$. This means that except for a set of measure 0 exceptions for all $\delta > 0$ there exists $N_\delta \in \mathbb{N}$ such that for all $n > N_\delta$

$$|\frac{X_n}{n} - \epsilon| < \delta$$

Therefore we can conclude that:

$$|X_n - n\epsilon| < n\delta \tag{5.4.1}$$
$$n\epsilon - n\delta < X_n < n\delta + n\epsilon \tag{5.4.2}$$
$$n(\epsilon - \delta) < X_n < n(\epsilon + \delta) \tag{5.4.3}$$

The part of the (complex) automorphism group coming from $(2,2)$-stars can therefore be estimated as follows. For $n > N_\delta$,

$$8^{n(\epsilon-\delta)} < |Aut_{2,2}(t_N)| = 8^{X_n} < 8^{n(\epsilon+\delta)} \tag{5.4.4}$$
$$1 < 8^{(\epsilon-\delta)} < |Aut_{2,2}(t_N)|^{\frac{1}{n}} = 8^{X_n^{\frac{1}{n}}} < 8^{(\epsilon+\delta)} \tag{5.4.5}$$

Since we can choose our $\delta << \epsilon$ and there exists a Polya urn model which shows that for all m and n there exists an $\epsilon_{n,m}$ such that $lim_{n\to\infty}\frac{X_n}{n} \to \epsilon_{n,m}$. This disproves Ben's conjecture that almost surely, in the limit as $n \to \infty$:

$$lim_{n\to\infty}|Aut_{\text{Complex}}(T_n)|^{\frac{1}{n}} \to 1$$

5.5 Convergence of the Automorphism group

In this section we will prove that $\lim_{n\to\infty} Aut(T_n)$ converges.

Let $\{T_t\}_{t=1}^n$ be a random recursive tree and X_{ni} be the number of vertices of degree i in T_n.

Theorem 5.5.1. *In the limit as $n \to \infty$, almost surely $\frac{X_{n,i}}{n} \to 2^{-i}$.*

Proof. See Janson **??** □

In other words, except for a measure 0 set of exceptions for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n > N$

$$|\frac{X_{ni}}{n} - 2^{-i}| < \epsilon.$$

We can play the same game as section **??** and write:

$$|X_{ni} - n2^{-i}| < n\epsilon \tag{5.5.1}$$
$$n2^{-i} - n\epsilon < X_{ni} < n\epsilon + n2^{-i} \tag{5.5.2}$$
$$n(2^{-i} - \epsilon) < X_{ni} < n(\epsilon + 2^{-i}) \tag{5.5.3}$$
$$\tag{5.5.4}$$

For each i case we can choose ϵ to be as small as possible so let each $\epsilon_i = 2^{-i}$

Recall that there exists a geometric decomposition of $\text{Aut}(T_n)$ into (p,k)-stars corresponding to a direct product decomposition of $\text{Aut}\, T_n$ into subgroups isomorphic to symmetric groups or wreath products of symmetric groups. The stars corresponding to a wreath product $G_1 = S_{m_1} \wr S_{m_2} \wr \cdots \wr S_{m_x}$ contribute

$|S_{m_1} \wr S_{m_2} \wr \cdots \wr S_{m_x}| = (\ldots(m_1!^{m_2} m_2!)^{m_3} \ldots m_{x-1}!)^{m_x} m_x!$ to the automorphism group $\mathrm{Aut}(T_n)$. The star corresponding to G_1 is isomorphic to the graph depicted in figure **??**. Notice that $|G_1| = \prod_{v \in V} \deg(v)!$. Therefore given some instance, T_n, of random recursive tree $\{T_n\}_{n=1}^{\infty}$, $\mathrm{Aut}(T_n)$ is bounded above by $\prod_{v \in V(T_n)} deg(v)!$, hence Equation 5.5.1 gives us the following bound in the limit as $n \to \infty$ almost surely:

$$\mathrm{Aut}(T_n) < \prod_{i=2}^{\infty} (i!)^{X_{ni}} \tag{5.5.5}$$

$$< \prod_{i=2}^{\infty} (i!)^{n(\epsilon_i + 2^{-i})} \qquad < \prod_{i=2}^{\infty} (i!)^{n(2^{-i}+2^{-i})} < \prod_{i=2}^{\infty} (i!)^{n2^{-i+1}} \tag{5.5.6}$$

This implies that $\mathrm{Aut}(T_n^{\frac{1}{n}} < \prod_{i=2}^{\infty} (i!)^{2^{-i+1}} := X$. It remains to check that the convergence of X for which we will need the following theorem.

Theorem 5.5.2. *If $b_n \neq 0$ for all n then $\prod_{n=0}^{\infty} b_n$ converges if and only if $\sum_{n=0}^{\infty} Log(b_n)$ converges.*

Proof. See the proof of Theorem 3.8.1 in [**?**]. ☐

Therefore it suffices to prove that $\sum_{n=0}^{\infty} \frac{Log(i!)}{2^{i-1}}$ converges. By Stirling's approximation and the comparison test $\sum_{n=0}^{\infty} \frac{Log(i!)}{2^{i-1}}$ converges.

Remark 5. *Assume that in the limit as $n \to \infty$ $\frac{X_n}{n} \to X$ almost surely. This means that apart from a measure zero set of sample of RRTs $\forall \epsilon > 0 \exists N_\epsilon$ such that $\forall n > N_\epsilon |\frac{X_n}{n} - X| < \epsilon$.*

This does not *mean that apart from a measure zero set of RRTs that* $\lim_{n \to \infty} X_n = Xn$. *On the other hand we can say that given any $\delta > 0$ there exists n' such that $|X_n - Xn| < \delta$ (just take $n' = N_{\delta/N_\delta}$).*