PATTERNS IN RANDOM WORDS OF CONTEXT-FREE GRAMMARS

by

SERDAR TURKMENAFSAR

A THESIS SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Patterns in Random Words of Context-Free Grammars

submitted by Serdar Turkmenafsar in partial fulfillment of the requirements for

the degree of Master of Science

in Mathematics

Examining Committee:

Omer Angel, Professor, Department of Mathematics, UBC
Co-supervisor

Andrew Rechnitzer, Professor, Department of Mathematics, UBC
Co-supervisor

Abstract

Let T_n be a uniformly randomly selected word of size n from an irreducible unambiguous context-free grammar H. We analyze the number of times of seeing a fixed pattern within T_n by constructing a critical multitype branching process (on the context free grammar) that assigns the same probability to words of the same size (if the grammar is regular, then we construct such a Markov chain instead). We give an easy way to compute the density constants of expectations for any given pattern. We also give a lower bound for the density constant of the variance in the case that the given pattern is replaceable by another pattern. Using fringe convergence results from [5], we show that a uniformly selected fringe subtree from a uniformly selected derivation tree converges to the unconditioned branching process.

Lay Summary

Context-free grammars are combinatorial objects that often show up when analyzing statistics that are determined recursively e.g. conditioned random walks on recursive graphs. In this thesis, we pick a uniformly randomly selected word of a given size n from a context-free grammar and count the number of times a given pattern P appears inside this word. We examine the behaviour of this statistic when $n \to \infty$ by showing a connection between the uniformly selected word and a branching process.

Preface

This thesis is original, unpublished, independent work by the author, Serdar Turkmenafsar

Contents

List of Figures

Acknowledgements

I'd like to thank both of my supervisors Omer Angel and Andrew Rechnitzer for introducing me to this subject, the useful discussions we had, and the suggestions they gave me. I'd also like to thank Cindy Tan for writing and maintaining the LaTeXstyle files used to compile this thesis.

Chapter 1

Introduction to Languages

1.1 Alphabets

Let us consider a countable set Σ, called an *alphabet*. The elements of Σ are called symbols. We use Σ^* to denote the set of all finite sequences of symbols, also called *words* or *strings*.

$$\Sigma^* = \{a_1 a_2 \ldots a_n : n \in \mathbb{N}, a_i \in \Sigma \text{ for } i = 1, \ldots, n\} \tag{1.1}$$

For any two finite sequences v and ω in Σ^*, a concatenation operation can be defined. ϵ denotes the identity of this operation and is unique. We also define the length of each word in Σ^* as the length of the finite sequence, denoted $\#\cdot$. We have that $\#\epsilon = 0$.

1.2 Grammars

Definition 1.2.1. (Formal Grammar) A *formal grammar* G is a 4-tuple $(V, T, \mathcal{P}, S)$ where

1. V is a finite set of symbols, called *variables*.
2. T is a finite set of symbols, called *terminal symbols*.
3. $V \cap T = \emptyset$.
4. $\mathcal{P}$ is a set of pairs of words that are in $(V \cup T)^*$, denoted $\alpha \to \gamma$, called *productions* or *production rules*.
5. $S \in V$ is called a *start symbol*.

We sometimes use the notation $\alpha \to \beta|\gamma$ instead of saying $\alpha \to \beta$ and $\alpha \to \gamma$.

We say that α derives β, $\alpha \Rightarrow \beta$, if there exists a finite sequence of productions $\{\alpha_{i-1} \to \alpha_i\}_{i=1}^m$ with $\alpha_0 = \alpha$ and $\alpha_m = \beta$. This sequence is called a *derivation*.

$L(G) = \{\omega : \omega \in T^*, S \Rightarrow \omega\}$ is called the *language generated by a grammar*.

Chapter 2

Regular Languages

2.1 Introduction to Finite Automata

Definition 2.1.1. (DFA) A deterministic finite automata M is a 5-tuple $(V, \Sigma, \delta, v_0, T)$ where:

1. V is a finite set of states.
2. Σ is an alphabet.
3. $\delta : V \times \Sigma \to V$ is a transition function.
4. $v_0 \in V$ is the initial (start) state.
5. $T \subset V$ is the set of terminal states.

Notice that this is just a directed graph where the edges are labelled by elements of Σ, with a special starting vertex, and some terminal vertices. Also note that we cannot have double edges labelled by the same letter.

Definition 2.1.2. (Acceptance by DFAs) Let $\omega = a_1 a_2 \ldots a_n$ be a word over the alphabet Σ. We say that a DFA, M, *accepts* ω if there exists a sequence of states $\{r_0, \ldots, r_n\} \subset V$ such that the following hold:

1. $r_0 = v_0$
2. $r_{i+1} = \delta(r_i, a_{i+1})$ for $i = 0, \ldots, n-1$
3. $r_n \in F$

This is to say that given a word ω, starting from q_0 we trace a unique path following edges labelled by the letters of ω.

We label the language of the words accepted by the DFA, M, by $L(M)$. See Fig. 2.1 for an example DFA.

A DFA M is called unambiguous if each word in $L(M)$ is associated to a unique path in M. This is easy to verify, one only needs to check two conditions:

1. No edge is labeled with the empty word ϵ.
2. No two edges coming out of the same state is labeled with the same letter.

The language in Fig. 2.1 is unambiguous.

Definition 2.1.3. A language, $L \subset \Sigma*$ is *regular* if $L = L(M)$ for some DFA, M.

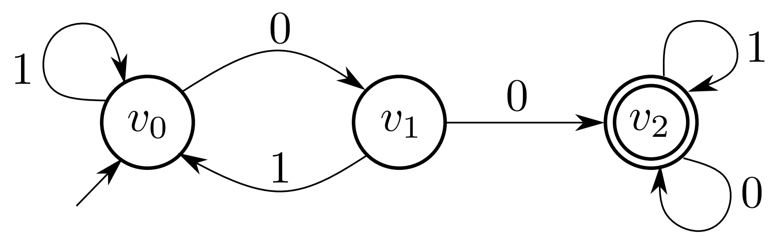


Figure 2.1: The language of all binary strings containing at least one "00".

2.2 Random Walks and Generating Functions

Since each word of a regular language is represented by a path in its DFA. We will be interested in looking at random walks on this DFA that give the same weight to each path of a given length. Therefore, we look at the markov chain called maximum entropy random walk (MERW).

Here we first state one version of the Perron-Frobenius Theorem.

Theorem 2.2.1. *(Perron-Frobenius)*

Let M be a non-negative irreducible square matrix. Then M has a unique dominant eigenvalue λ, that has exactly one positive left eigenvector φ, and one positive right eigenvector ψ. And,

$$\lim_{l\to\infty}\left(\frac{M}{\lambda}\right)^l = \psi\varphi^T \tag{2.1}$$

Note that a non-negative matrix M is irreducible if and only if its associated digraph is strongly connected.

2.3 Random Walks on DFA

Let (M_{ij}) be the adjacency matrix of an irreducible and unambiguous DFA M. We consider a markov chain where transition probabilities (P_{ij}):

$$P_{ij} = \frac{M_{ij}}{\lambda}\frac{\psi_j}{\psi_i} \text{ with } \pi_i = \varphi_i\psi_i \tag{2.2}$$

where φ and ψ are from Perron-Frobenius theorem on M, normalized so that $\varphi^T\psi = 1$. This is called the maximal entropy random walk. Notice that it has the property that for any integer l, and for any two vertices v, w, the paths of length l from v to w are equally likely since: If we suppose $\gamma = (\gamma_0, \ldots, \gamma_n)$ is a path, then

$$\mathbb{P}(\gamma) = \prod_{i=0}^{n-1} \mathbb{P}(\gamma_i \to \gamma_{i+1}) = \prod_{i=0}^{n-1} \frac{1}{\lambda} \frac{\psi_{\gamma_{i+1}}}{\psi_{\gamma_i}} = \frac{1}{\lambda^n} \frac{\psi_{\gamma_n}}{\psi_{\gamma_0}} \tag{2.3}$$

Theorem 2.3.1. *Let $M = (V, \Sigma, \delta, S, T)$ be an irreducible DFA. Let H be a pattern corresponding uniquely to a directed walk L_H in M. And let θ_n be the r.v. corresponding to the number of times H appears in a word of length n sampled uniformly. Then either for each n, θ_n is constant or:*

$$\frac{\theta_n - \mu n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2) \tag{2.4}$$

for some $\mu > 0$ and $\sigma > 0$. (The above convergence only holds on subsequences so that $\#\{$words of length $n\} > 0$)

Proof. We will first solve the special case where there is only one terminal vertex T.

Let $X = \{X_n\}$ be the markov chain on M defined by Eq. (2.2). Let $\theta_n(X)$ be the number of times the pattern H is observed in the first n steps of X (notice that $\theta_n(X)$ is measurable with respect to $\sigma(X_0, \ldots, X_n)$). We are interested in the asymptotics of $\theta_n(X)$ conditioned on $X_0 = S$ and $X_n = T$. This is because $\theta_n(X)$ conditioned on $X_0 = S$ and $X_n = T$ is the number of times the pattern H is observed in a uniformly selected word of size n (since we are assuming that the only terminal state is T).

We already know that there is a central limit theorem of $\theta_n(X)$ conditioned on $X_0 = S$ (and not conditioned on the future), i.e.

$$\frac{\theta_n(X) - \mu n}{\sqrt{n}} \to^d N(0, \sigma^2) \tag{2.5}$$

for some $\mu, \sigma > 0$ and for the initial distribution $X_0 = S$. The only exception to this is if the whole DFA M is a directed cycle, in which case the random variable $\theta_n(X)$ is a constant for each n. Suppose that M is not a directed cycle.

This gives a CLT when no future conditioning is present. Now we prove the conditioned result.

Let $g_n(X) = e^{is\frac{\theta_n(X) - \mu n}{\sqrt{n}\sigma}}$ (where $i = \sqrt{-1}$ and s an arbitrary real parameter). We know that the characteristic function $\mathbb{E}(g_n(X)|X_0 = S)$ converges to the characteristic function of a stardard normal variable. Now our goal is to show that the conditioned characteristic function $\mathbb{E}(g_n(X)|X_0 = S, X_n = T)$ also converges to this. Note that $g_n(X)$ is measurable w.r.t. $\sigma(X_0, \ldots, X_n)$. And that for simplicity, we suppose $\mathbb{P}(X_0 = S, X_n = T) > 0$, otherwise one can just work on a subsequence of n, say $k(n)$ such that $\mathbb{P}(X_0 = S, X_{k(n)} = T) > 0$ and the results would be identical.

Let $f(n)$ be a sequence such that $f_n = o(\sqrt{n}) \to \infty$. (We are using the small o notation here.) Notice that stoping the pattern counting after the first $n - f(n)$ steps in an n step random walk will not alter the gaussian law since

$$\left| \frac{\theta_n(X) - \mu n}{\sigma\sqrt{n}} - \frac{\theta_{n-f(n)}(X) - \mu(n - f(n))}{\sigma\sqrt{n}} \right| \lesssim \frac{f(n)}{\sqrt{n}} \to 0 \text{ a.s.} \tag{2.6}$$

Then, we define the modified characteristic function with $g'_n(X) = \exp\left(is\frac{\theta_{n-f(n)}(X)-c(n-f(n))}{C\sqrt{n}}\right)$ where $g'_n(X)$ is measurable w.r.t. $\sigma(X_0,\dots,X_{n-f(n)})$. From Eq. (2.6), we also see that the characteristic function $\mathbb{E}(g'_n(X)|X_0 = S)$ converges to the CF of a standard normal. We claim the following:

$$\mathbb{E}_\pi(g'_n(X)|X_0 = S, X_n = T) = [1 + o(1)]\mathbb{E}_\pi(g'_n(X)|X_0 = S) \tag{2.7}$$

which would conclude our proof since we know that $\mathbb{E}(g_n(X)|X_0 = S)$ converges to the CF of a standard normal. To show the above result, we first show the following, let $\{x_t\}$ be a sequence in the sample space of X, then,

$$\mathbb{P}_\pi(X_t = x_t \text{ for } t \le n-f(n)|X_0 = S, X_n = T) = [1+o(1)]\mathbb{P}_\pi(X_t = x_t \text{ for } t \le n - f(n)|X_0 = S) \tag{2.8}$$

where $o(1)$ does not depend on the sequence $\{x_t\}$. This holds true because,

$$\begin{aligned}
&\mathbb{P}_\pi(X_t = x_t\ \forall t \le n - f(n)|X_0 = S, X_n = T) = \frac{\mathbb{P}_\pi(\{X_t = x_t\ \forall t \le n - f(n)\} \cap \{X_n = T\})}{\mathbb{P}_\pi(X_0 = S, X_n = T)} \\
&= \frac{\mathbb{P}_\pi(X_n = T|X_t = x_t\ \forall t \le n - f(n))\mathbb{P}_\pi(X_t = x_t\ \forall t \le n - f(n))}{\mathbb{P}_\pi(X_n = T|X_0 = S)\mathbb{P}_\pi(X_0 = S)} \\
&= \frac{\mathbb{P}_\pi(X_n = T|X_{n-f(n)} = x_{n-f(n)})}{\mathbb{P}_\pi(X_n = T|X_0 = S)}\frac{\mathbb{P}_\pi(X_t = x_t\ \forall t \le n - f(n))}{\mathbb{P}_\pi(X_0 = S)} \\
&= \frac{P^{[f(n)]}_{x_{n-f(n)},T}}{P^{[n]}_{S,T}}\mathbb{P}_\pi(X_t = x_t\ \forall t \le n - f(n)|X_0 = S)
\end{aligned} \tag{2.9}$$

where $P^t_{i,j}$ is the t-step transition matrix, thus, we are done proving Eq. (2.8) since $\frac{P^{[f(n)]}_{x_{n-f(n)},T}}{P^{[n]}_{S,T}} \to 1$ also notice that this does not depend on the sequence $\{x_t\}$. This actually also proves Eq. (2.7) since $g'_n(X)$ is in $\sigma(X_0,\dots,X_{n-f(n)}) \subset \sigma(X_0,\dots,X_n)$. More explicitly,

$$\begin{aligned}
g'_n(X) &= \sum_{x_0,\dots,x_n} g'_{x_0,\dots,x_{n-f(n)}}\mathbf{1}_{\{X_t=x_t\ \forall t\le n-f(n)\}} \text{ for some } g'_{x_0,\dots,x_{n-f(n)}} \in \mathbb{C} \\
\therefore\ \mathbb{E}_\pi(g'_n(X)|X_0 = S, X_n = T) &= \sum_{x_0,\dots,x_n} g'_{x_0,\dots,x_{n-f(n)}}\mathbb{E}_\pi(\mathbf{1}_{\{X_t=x_t\ \forall t\le n-f(n)\}}|X_0 = S, X_n = T) \\
\text{(using Eq. (2.8))} &= \sum_{x_0,\dots,x_n} g'_{x_0,\dots,x_{n-f(n)}}[1 + o(1)]\mathbb{E}_\pi(\mathbf{1}_{\{X_t=x_t\ \forall t\le n-f(n)\}}|X_0 = S) \\
&= [1 + o(1)]\mathbb{E}_\pi(g'_n(X)|X_0 = S)
\end{aligned} \tag{2.10}$$

and this shows that indeed $\mathbb{E}_\pi(g'_n(X)|X_0 = S, X_n = T)$ converges to the CF of a standard normal. This concludes the proof since we already have that

$$\mathbb{E}_\pi(|g_n(X) - g'_n(X)|\big|X_0 = S, X_n = T) \to 0 \tag{2.11}$$

by DCT (using 2 as the upper bound and) using Eq. (2.6). ■

Proof of Theorem 2.3.1 for multiple terminal vertices. Now recall that the proof of Theorem 2.3.1 we supposed that the DFA had only one terminal vertex, now we fix this. Suppose that the DFA has the terminal vertices $T = \{t_1,\dots,t_m\}$. Construct m identical copies of the DFA M called M_i or DFA_i, each having only one terminal vertex t_i. And on each one we define the same markov chain

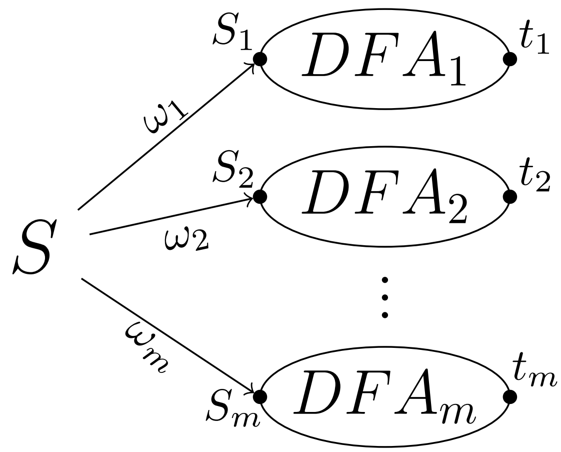


Figure 2.2: Dealing with multiple terminal vertices of Theorem 2.3.1.

from Eq. (2.2). Then join them as in Fig. 2.2 to form an NFA: where the starting vertex is S and all the edges leaving S are labelled with the "empty word" (so that these edges don't affect any word). And the transition probabilities are given by w_i in the following equation:

$$\omega_i = \frac{1/\pi(t_i)}{\sum_{j=1}^{m} \frac{1}{\pi(t_j)}} \tag{2.12}$$

Notice that in this NFA, all large words have approximately equal probability since $\lim_{n\to\infty} \mathbb{P}(X_n = t_i | X_0 = S) = \pi(t_i)$. And on each component, we can apply the one terminal vertex version of the theorem. (Again the limits should be taken over a subsequence $k(n)$ so that $\mathbb{P}(X_{k(n)} = t_i | X_0 = S) > 0$) ■

Chapter 3

Context-Free Languages

3.1 Introduction to CLFs

Definition of CFLs

Definition 3.1.1. (CFL) A grammar $H = (V, T, \mathcal{P}, S)$ (as defined in Definition 1.2.1) is context-free iff every production $\alpha \to \gamma$ is such that $\alpha \in V$. A language L is *context-free* iff it is generated by a context-free grammar.

Definition 3.1.2. A derivation of a word for a grammar is a sequence of grammar rule/production applications that transform the start symbol into the string.

A derivation of a string in a CFG can also be described by a finite rooted tree satisfying the following.

1. The root is labelled with S.
2. Each non-leaf vertex is labelled by an element of V.
3. Each leaf vertex is labelled by an element of T.
4. The direct descendants of a node N correspond to a production of the form $N \to \beta$.
5. The labels on the leaves read from left to right form the derived word.

Definition 3.1.3. We say that a context-free grammar G is *unambiguous* if for every word in G, there is a unique derivation tree. A language is unambiguous if it is generated by an unambiguous grammar.

Definition 3.1.4. We say that a context-free grammar G is *irreducible* if from any variable $i \in V$, one can obtain any other variable $j \in V$ by some consecutive application of production rules in $\mathcal{P}(G)$. I.e. there is only one equivalence class of V under the equivalence relation of mutual reachability. (This equivalence class is called the communicating class.)

Example 3.1.5 (Dyck Language, i.e. Language of Balanced Brackets)**.** We define the grammar $G = (V, T, \mathcal{P}, S)$ with $V = \{S\},\ T = \{")", "("\}$, and the following production rules:

$$S \to \epsilon,\ S \to (S)S \tag{3.1}$$

these rules can be shorthanded with the notation $S \to \epsilon \mid (S)S$.

The word $(())()$ can be derived in the following manner:

$$S \to (S)S \to ((S)S)S \to ((S)S)(S)S \implies (())() \tag{3.2}$$

in the last step marked with " $\implies$ " we shorthand the 4 steps of substituting ϵ for the S'es. The unique derivation tree associated to this derivation is given in Fig. 3.1.

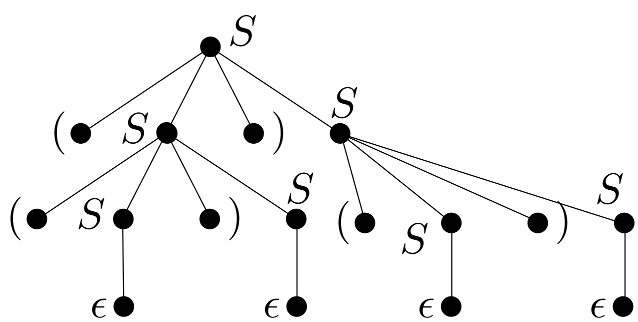


Figure 3.1: The derivation tree corresponding to the word $(())()$ in Example 3.1.5.

Example 3.1.6 (Second block of b's are of double size)**.** We define the grammar $G = (V, T, \mathcal{P}, S)$ with $V = \{S, A\},\ T = \{a, b\}$, and the following production rules:

$$\begin{aligned} S &\to bSbb \mid A \\ A &\to aA \mid \epsilon \end{aligned} \tag{3.3}$$

Here $L(G) = \{b^n a^m b^{2n} : n \geq 0, m \geq 0\}$

Before we move forward with the rest of this chapter, we introduce some notation building on the definition of grammars given in Definition 1.2.1. Let $G = (V, T, \mathcal{P}, S)$ be a formal grammar, and $\omega \in (V \cup T)^*$, then

- $\#\omega \equiv \#_T\omega :=$ number of characters of T present in ω (e.g. $\#bSbb = 3$ in Example 3.1.6).
- $\vec{\#}\omega \equiv \vec{\#}_V\omega := \langle \#_{i_1}\omega, \ldots, \#_{i_d}\omega \rangle$ where $V = \{i_1, \ldots, i_d\}$ and $\#_{i_k}\omega :=$ number of $i_k \in V$ present in ω.
  (e.g. $\vec{\#}bSbb = \langle 1, 0 \rangle$ in Example 3.1.6)
- We use the notation above also for derivation trees.
- $\mathcal{P}_i \equiv \mathcal{P}_i(G) := \{p \in \mathcal{P} : \mathrm{LHS}(p) = i\}$.
- Given $\mathbf{x}, \mathbf{k} \in \mathbb{R}^d$, $\mathbf{x}^{\mathbf{k}} := \prod_{m=1}^{d} (x_i)^{k_i}$, and $\mathbf{x} * \mathbf{k}$ denotes the entrywise product, i.e. $\mathbf{x} * \mathbf{k} = \mathrm{diag}(\mathbf{x})\mathbf{k}$
- For a set V, $|V| := \mathrm{card}(V)$.

Note that we sometimes use the notation T for trees, in cases where we think this is confusing, we will refer to the set of terminal variables by $T(G)$.

3.2 General Irreducible Context Free Grammars

Let H be an unambiguous irreducible context free grammar where V is the set of non-terminal variables ($|V| = d$), and $T(H)$ is the set of terminal letters.

Definition 3.2.1. (Redundant Variables) A variable $i \in V$ is called redundant if no finite words can be obtained by starting a derivation from i.

Note: No finite word can use a variable that belongs to a set of redundant variables. So removing these variables would not alter the language.

Maximal Entropy Branching on Words

Then, we define the following function:

Definition 3.2.2. (Characteristic Function of a CFG) Let $H = (V, T, \mathcal{P}, S)$ be a context-free grammar. For each $i \in V$, define $F_i : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$

$$F_i(z, \mathbf{x}) = \sum_{w: i \to w \in \mathcal{P}} z^{\#(w)} \mathbf{x}^{\vec{\#}(w)} \tag{3.4}$$

Now, we will want to find a pair $(\frac{1}{\lambda}, \psi) \in \mathbb{R} \times \mathbb{R}^m$ such that $\mathbf{F}(\frac{1}{\lambda}, \psi) = \psi$. And such a pair must exist as long as we can guarantee that there are no empty productions in our grammar.

Proposition 3.2.3. *Let $H = (V, T, \mathcal{P}, S)$ be an unambiguous context-free grammar. For each $i \in V$, define the following combinatorial generating function:*

$$N_i(z) = \sum_n \#\{\text{words of size } n \text{ starting from } i\}\ z^n \tag{3.5}$$

Then, $\mathbf{N}(z) = \mathbf{F}(z, N(z))$.

Proof. This very standard result is left to the reader. References can be found in [3] if needed. ∎

Proposition 3.2.4. *Let H be an unambiguous context-free grammar, then there exists a pair $(\frac{1}{\lambda}, \psi) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^m_{\geq 0}$ such that $\mathbf{F}(\frac{1}{\lambda}, \psi) = \psi$.*

Proof. Let V be the set of variables for the CFG H, and T be the set of letters (alphabet) of H. For each $i \in V$, define the following combinatorial generating function:

$$N_i(z) = \sum_n \#\{\text{words of size } n \text{ starting from } i\}\ z^n \tag{3.6}$$

Now notice that we have $\mathbf{N}(z) = \mathbf{F}(z, \mathbf{N}(z))$. Moreover, there exists m such that for all $n \geq 0$ and all $i \in V$,

$$\#\{\text{words of size } n \text{ starting from } i\} \leq m^n \tag{3.7}$$

This is obvious if we set $m = \#T$. Thus notice that, $\psi := \mathbf{N}(\frac{1}{m+1}) < \infty$. Then we are done since

$$F\left(\frac{1}{m+1}, \psi\right) = \psi \tag{3.8}$$

■

We also need it so that $\psi_i > 0$ for all $i \in V$. Thus, the following proposition follows:

Proposition 3.2.5. *Suppose that* $(\lambda > 0, \psi \geq \mathbf{0})$ *solves* $\mathbf{F}(\frac{1}{\lambda}, \psi) = \psi$*, then* $\psi_i > 0$ *for all* $i \in V$ *if the context-free grammar has no redundant variables.*

Proof. We will prove the contrapositive of this. Suppose there exists a $i \in V$ such that $\psi_i = 0$. Then define the set $I = \{i \in V : \psi_i = 0\} \neq \emptyset$. Then we have that for all $i \in I$, $F_i\left(\frac{1}{\lambda}, \psi\right) = 0$. Thus, the right hand side of each production of $i \in I$ must be including an element of I. But this implies that all elements of I are redundant variables, since a derivation tree could never terminate once encountering a variable of I. ■

Now, let H be an unambiguous irreducible context-free grammar with no empty productions and with no productions of the form $i \longrightarrow j$ that has no redundant variables. Then, we use the last proposition to define the following branching process.

$$\mathbb{P}(i \to w) = \frac{1}{\lambda^{\eta(w)}} \frac{\psi^{\vec{\#}(w)}}{\psi_i} \tag{3.9}$$

This process has the following bi-variate offspring generating function attached to it:

$$\begin{gathered} f_i(z, \mathbf{x}) = \sum_{w: i \to w \in \mathcal{P}} \mathbb{P}(i \to w) z^{\#(w)} \mathbf{x}^{\vec{\#}(w)} = \frac{1}{\psi_i} F_i\left(\frac{z}{\lambda}, \psi * \mathbf{x}\right) \\ \therefore \quad \psi * \mathbf{f}(z, \mathbf{x}) = \mathbf{F}(\frac{z}{\lambda}, \psi * \mathbf{x}) \end{gathered} \tag{3.10}$$

Proposition 3.2.6. *A branching process defined by Eq. (3.9) on an unambiguous CFG gives the same probability to words of the same size.*

Proof. Let t be a derivation tree for a word w. Let $\mathcal{P}(t)$ denote the multiset of productions that are used in the derivation tree t. Let T be a random tree produced by the BP (branching process) defined by Eq. (3.9). Then:

$$\mathbb{P}(T = t) = \prod_{p \in \mathcal{P}(t)} \mathbb{P}(p) \tag{3.11}$$

Let $[\mathcal{P}(t)]_n$ denote the multiset of productions in t that originate from a vertex of depth n in t, i.e. $[\mathcal{P}(t)]_n = \{p \in \mathcal{P}(t) : \text{LHS}(p) \text{ is of depth } n.\}$. Then,

$$\mathbb{P}(T = t) = \prod_{p \in \mathcal{P}(t)} \frac{1}{\lambda^{\#(\text{RHS}(p))}} \frac{\psi^{\vec{\#}(\text{RHS}(p))}}{\psi_{\text{LHS}(p)}} = \frac{1}{\lambda^{\# w}} \prod_{p \in \mathcal{P}(t)} \frac{\psi^{\vec{\#}(\text{RHS}(p))}}{\psi_{\text{LHS}(p)}} \tag{3.12}$$

where for $\omega \in (V \cup T)^*$, $\#(\omega)$:= # of terminal letters in ω, and for each $i \in V$, $y_i(\omega)$:= # of variables in ω. Notice that since derivation trees eventually end in all terminal letters, every variable that appears in the RHS of a production is also on the LHS of another production (i.e. all leaves are terminal letters not variables). Thus, $\prod_{p \in \mathcal{P}(t)} \frac{\psi^{\vec{\#}(\text{RHS}(p))}}{\psi_{\text{LHS}(p)}} = 1$,

$$\mathbb{P}(T = t) = \frac{1}{\lambda^{\# w}} \tag{3.13}$$

■

Lemma 3.2.7. *Let $H = (V, T, \mathcal{P}, S)$ be an unambiguous context-free grammar. Let $i \implies \alpha$ be a derivation (as defined in Section 1.2) in H where $i \in V$ and $\alpha \in (V \cup T)^*$. Then,*

$$N_i(z) \geq z^{\#\alpha} [\mathbf{N}(z)]^{\vec{\#}\alpha} \tag{3.14}$$

where $\mathbf{N}(z)$ is defined in Proposition 3.2.3.

Proof. Since a derivation is a sequence of consequently applied productions, it suffices to prove this result for a production (then proceed by induction). Suppose $i \to \omega$ is a production in H, then consider,

$$N_i(z) = F_i(z, \mathbf{N}(z)) = \sum_{w: i \to w \in \mathcal{P}} z^{\#(w)} \mathbf{N}(z)^{\vec{\#}(w)} \geq z^{\#(\omega)} \mathbf{N}(z)^{\vec{\#}(\omega)} \tag{3.15}$$

The first equality follows from Proposition 3.2.3. Thus, we are done. ∎

Theorem 3.2.8. *Let $H = (V, T, \mathcal{P}, S)$ be a CFG that:*

(i) has no redundant variables,

(ii) is irreducible (i.e. there is only one communicating class),

(iii) is not equivalent to a DFA (i.e. $\exists (i \to w) \in \mathcal{P}$ s.t. $\left\|\vec{\#}(w)\right\|_1 \geq 2$)

Then, $\{N_i(z)\}_{i \in V}$ (defined in Proposition 3.2.3) have a common radius of convergence. Let $0 < \frac{1}{\lambda} \leq 1$ be the common radius of convergence.

Then $\psi_i := N_i(\frac{1}{\lambda}) < \infty$ and $\psi = \mathbf{F}(\frac{1}{\lambda}, \psi)$, and the resulting BP from Eq. (3.9) is critical.

Proof. For each $i \in V$, define the following combinatorial generating function,

$$N_i(z) = \sum_n \#\{\text{words of size } n \text{ starting from } i\} z^n \tag{3.16}$$

and $\mathbf{N}(z) = \sum_{i \in V} N_i(z) \mathbf{e}_i$ where $\{\mathbf{e}_i\}_{i \in V}$ are taken as an orthonormal basis. Let $\frac{1}{\lambda}$ be the radius of convergence of $\mathbf{N}(z)$. We need to show that λ is well-defined i.e. all generating functions $\{N_i(z)\}_{i \in V}$ have the same radius of convergence.

Fix $z > 0$. We claim that $A = \{i \in V : N_i(z) < \infty\}$ is a communicating class (if it is not empty) i.e. starting from any $i \in A$, one can produce any other state in A using production rules of H (this is already given since the language is assumed to be irreducible) and one cannot produce any $j \notin A$. Thus it suffices to show that no variable $i \in A$ can include a $j \notin A$ in the RHS of any of its productions.

First note that $N_i(z) > 0$ for all $i \in V$ since the language contains no redundant variables. Let $i \in A$, then pick any $p \in \mathcal{P}_i$, suppose $j \in V$ appears in the RHS of p (i.e. $\#_j \mathrm{RHS}(p) > 0$), then, using $\mathbf{F}(z, \mathbf{N}(z)) = \mathbf{N}(z)$, we get:

$$N_i(z) = (\cdots \geq 0) + (\cdots > 0) N_j(z) \tag{3.17}$$

Thus, $N_i(z) < \infty \implies N_j(z) < \infty$ thus, $i \in A \implies j \in A$. This shows that A is a communicating class. Since the language is irreducible by assumption $A = V$ and thus $\{N_i(z)\}_{i \in V}$ have a common radius of convergence $\frac{1}{\lambda}$. Also note that the radius of convergence $\frac{1}{\lambda} \leq 1$.

We claim that $\lim_{z \uparrow \frac{1}{\lambda}} \mathbf{N}(z) = \mathbf{N}(\frac{1}{\lambda}) = \psi < \infty$. All we need to do here is to show that $N_i(\frac{1}{\lambda}) < \infty$ for all $i \in V$. Note that so far, we have not used assumption (iii), the proof of the prior statement will require (iii).

For contradiction, suppose that for some $i \in V$, $N_i(\frac{1}{\lambda}) = \infty$, then since A is a communicating class, we have that $N_i(\frac{1}{\lambda}) = \infty$ for all $i \in V$. By assumtion (iii), there exists $i \in V$ such that $(i \to w) \in \mathcal{P}$ where $\left\|\vec{\#}w\right\|_1 \geq 2$. Then, w can be decomposed as $w = w_1 i_1 w_2 i_2 w_3$ where $w_1, w_2, w_3 \in (V \cup T)^*$ and $i_1, i_2 \in V$. Moreover, because the grammar is irreducible, there must be a way to obtain i from i_1 and from i_2, thus there exists a derivation so that $i \implies \omega_1 i \omega_2 i \omega_3$ where $\omega_1, \omega_2, \omega_3 \in (V \cup T)^*$ (and derivations are defined in Section 1.2).

Let $M = \lambda^{\#\omega_1 + \#\omega_2 + \#\omega_3} \geq 1$. Pick $\mu < \frac{1}{\lambda}$ such that $\mathbf{N}(\mu) > M\mathbf{1}$ (thus $M < N_i(\mu) < \infty$) (this can be done because we assumed $\lim_{z \uparrow \frac{1}{\lambda}} \mathbf{N}(z) = \mathbf{N}(\frac{1}{\lambda}) = \infty$). Then notice that the derivation $i \implies \omega_1 i \omega_2 i \omega_3$ can be applied on itself $k \in \mathbb{N}$ times, (for example $i \implies \omega_1 i \omega_2 i \omega_3 \implies \omega_1 \omega_1 i \omega_2 i \omega_3 \omega_2 i \omega_3 \implies \dots$). This can be used with Lemma 3.2.7 to give that for all $k \in \mathbb{N}$, we have:

$$N_i(\mu) \geq \frac{1}{\mu^{k(\#\omega_1 + \#\omega_2 + \#\omega_3)}} [N_i(\mu)]^k > \frac{M^k}{\mu^{k(\#\omega_1 + \#\omega_2 + \#\omega_3)}} \to \infty \text{ as } k \to \infty \tag{3.18}$$

This gives a contradiction since $\mu < \frac{1}{\lambda} \implies N_i(\mu) < \infty$. This proves that $\psi = \mathbf{N}(\frac{1}{\lambda}) < \infty$.

Now it remains to show that the branching process genereted by Eq. (3.9) is critical. Eq. (3.10) implies that the mean matrix of this branching process is similar to the matrix $D_2\mathbf{F}(\frac{1}{\lambda}, \psi)$. Here, $\partial_1 \mathbf{F}$ and $D_2\mathbf{F}$ refer to $\frac{d}{dz}\mathbf{F}(z, \mathbf{x})$ and $\frac{d}{d\mathbf{x}}\mathbf{F}(z, \mathbf{x})$ respectively. Thus, it suffices to show that the dominant eigenvalue of $D_2\mathbf{F}(\frac{1}{\lambda}, \psi)$ is 1. (Note that for $z > 0$, Perron-Frobenius Theorem applies to $D_2\mathbf{F}(z, \mathbf{N}(z))$ and the concept of a dominant eigenvalue makes sense)

Consider the Jacobian $J(z_0) := \det[I - D_2\mathbf{F}(z_0, \mathbf{N}(z_0))]$. The multivariate version of implicit function theorem implies that if this Jacobian is non-zero, then $\mathbf{N}(z)$ can be analytically extended around z_0. However, since we know that this is impossible at the radius of convergence, thus we must have:

$$\det\left[I - D_2\mathbf{F}\left(\frac{1}{\lambda}, \psi\right)\right] = 0 \tag{3.19}$$

This shows that 1 is a eigenvalue of the branching process, a similar argument is given in the proof of Theorem VII.6 (DLW Theorem) from [3]. Now we show that 1 is the maximum eigenvalue. Suppose (for contradiction) the maximum eigenvalue is $\rho > 1$, then the branching process defined by Eq. (3.9) is supercritical. Let T be a tree generated by this branching process, then for each $i \in V$, define $\gamma_i = \mathbb{P}_i(T < \infty)$. Since the process is supercritical, irreducible, and each production has positive weight (by Proposition 3.2.5), we know that $\mathbf{0} < \gamma < \mathbf{1}$. Let $\mathbf{f}(z, \mathbf{x})$ be the offspring generating function defined in Eq. (3.10), then we have that $\mathbf{f}(1, \gamma) = \gamma$. Then Eq. (3.10) implies that:

$$\psi * \gamma = \mathbf{F}(\frac{1}{\lambda}, \psi * \gamma) \tag{3.20}$$

where $*$ is the entrywise product operation on vectors. Now we have two solutions to the equation $\mathbf{x} = \mathbf{F}(\frac{1}{\lambda}, \mathbf{x})$, one being ψ and the other being $\psi_2 := \gamma * \psi < \psi$. Then we can use Eq. (3.9) to define a new branching process using $(\frac{1}{\lambda}, \gamma * \psi)$. If this new BP is still supercritical, we can repeat this procedure again to produce a third solution to $\mathbf{x} = \mathbf{F}(\frac{1}{\lambda}, \mathbf{x})$ so that $\psi_3 < \psi_2 < \psi$. We can continue repeating this inductively until we no longer have a supercritical process, and such a time must come because $\mathbf{x} = \mathbf{F}(\frac{1}{\lambda}, \mathbf{x})$ can only have a finite number of solutions. Thus, let $\bar{\psi}$ be the solution such that the BP defined by $(\frac{1}{\lambda}, \bar{\psi})$ using Eq. (3.9) is not supercritical.

Let T be sampled from the original BP generated by $(\frac{1}{\lambda}, \psi)$ and T' be independently sampled from the new BP generated by $(\frac{1}{\lambda}, \bar{\psi})$. Then we know that $\mathbb{P}(\#T < \infty) < 1$ and $\mathbb{P}(\#T' < \infty) = 1$, but this is a contradiction since if we pick any non-random derivation tree, t, of a word, then Proposition 3.2.6 gives us that:

$$\mathbb{P}(T = t) = \mathbb{P}(T' = t) = \frac{1}{\lambda^{\#t}} \tag{3.21}$$

■

Conditioning on Word Size

Consider the critical branching process on H given by Theorem 3.2.8. Sample a word/tree T by running this branching process, this word is finite almost surely because the process is critical. Now for each $i \in V$, we define the following generating functions.

$$\begin{aligned} N_i(z, \mathbf{u}) &= \sum_{n=0}^{\infty} \sum_{\mathbf{k} \in \mathbb{N}^m} \mathbb{P}_i(\#T = n, \vec{\#}T = \mathbf{k}) z^n \mathbf{u}^{\mathbf{k}} \\ N_i(z) = N_i(z, \mathbf{1}) &= \sum_{n=0}^{\infty} \mathbb{P}_i(\#T = n) z^n \end{aligned} \tag{3.22}$$

Proposition 3.2.9.

$$\begin{aligned} \mathbf{N}(z, \mathbf{u}) &= \mathbf{u} * \mathbf{f}(z, \mathbf{N}(z, \mathbf{u})) \\ \therefore \quad \psi * \mathbf{N}(z, \mathbf{u}) &= \mathbf{u} * \mathbf{F}\left(\frac{z}{\lambda}, \psi * \mathbf{N}(z, \mathbf{u})\right) \end{aligned} \tag{3.23}$$

Proof. Proving the first equation suffices since the second equation follows from the first one after applying Eq. (3.10). We use the notation $[z^n, \mathbf{u}^{\mathbf{k}}]\mathbf{N}(z, \mathbf{u})$ to denote the coefficient of the term $z^n, \mathbf{u}^{\mathbf{k}}$ in $\mathbf{N}(z, \mathbf{u})$. Fix a $i \in V$, then,

$$\begin{aligned} [z^n, \mathbf{u}^{\mathbf{k}}]\mathrm{RHS}_i &= \sum_{w : i \to w \in \mathcal{P}} \mathbb{P}(i \to w) [z^{n-\#(w)}, \mathbf{u}^{(\mathbf{k}-\mathbf{e}_i)}]\, [\mathbf{N}(z, \mathbf{u})]^{\vec{\#}(w)} \\ \text{(by induction)} &= \sum_{w : i \to w \in \mathcal{P}} \mathbb{P}(i \to w) \mathbb{P}_{\vec{\#}(w)}(\#T = n - \#(w), \vec{\#}T = \mathbf{k} - \mathbf{e}_i) \\ &= \mathbb{P}_i(\#T = n, \vec{\#}T = \mathbf{k}) = [z^n, \mathbf{u}^{\mathbf{k}}]\mathrm{LHS}_i \end{aligned} \tag{3.24}$$

where $\mathbb{P}_{\vec{\#}(w)}(\#T = n, \vec{\#}T = \mathbf{k})$ refers to the probability of $\#T = n$ and $\vec{\#}T = \mathbf{k}$ where we start a forest T with multiple roots ($\#_i(w)$ many roots of type i for each type $i \in V$). The step that needs induction for justification is that:

$$[z^n, \mathbf{u}^{\mathbf{k}}]\, [\mathbf{N}(z, \mathbf{u})]^{\mathbf{m}} = \mathbb{P}_{\mathbf{m}}(\#T = n, \vec{\#}T = \mathbf{k}) \tag{3.25}$$

■

Perturbation Analysis of the Resolvent

As we will see in Proposition 3.2.10 and similar results, one matrix function we care about a lot is $\mathcal{R}(z) := [I - D_2\mathbf{f}(z, \mathbf{N}(z))]^{-1}$, the resolvent, where I is the identity matrix and $D_2 := \frac{\partial}{\partial \mathbf{x}}$ on $\mathbf{f}(z, \mathbf{x})$. We will do a perturbation analysis of this quantity.

We will keep assuming that the language is irreducible, that is to say that $D_2\mathbf{f}(1,\mathbf{1}) = \mathbb{M}$ is an irreducible matrix. Now notice that this implies whenever $z \in (0,1]$, $\mathbb{M}(z) := D_2\mathbf{f}(z,\mathbf{N}(z))$ is also irreducible since for all $i \in V$, $N_i(z) > 0$ for all $z \in (0,1]$. And thus, $[D_2\mathbf{f}(z,\mathbf{N}(z))]_{i,j} > 0$ if and only if $\mathbb{M}_{i,j} > 0$. We will employ results from Chapter 2 of [4] to analyse $\mathcal{R}(z)$. But this requires $\mathbb{M}(z)$ to be analytic. To satisfy this condition, notice that Proposition 3.2.9 tells us that $\mathbf{N}(z)$ may be expanded as a Puiseux series around $(1-z)$ when $z \in [0,1]$, and combined with the fact that $\mathbf{N}(1) = \mathbf{1}$, this gives us that $\exists p \in \mathbb{N}$ such that $\mathbf{N}(1-z^p)$ is analytic around $z = 0$ and the power series converges when $z \in [0,1]$. This implies that $\mathbb{M}(1-z^p)$ is also analytic. In this setting, we can use Equation $II - (1.22)$ from [4], to say that:

$$R(1-z^p) = -\sum_{k=1}^{s}\left[\frac{P'_k(z)}{1-\rho'_k(z)} + \frac{\mathbb{D}'_k(z)}{(1-\rho'_k(z))^2} + \cdots + \frac{[\mathbb{D}'_k(z)]^{m_k-1}}{(1-\rho'(z))^{m_k}}\right] \tag{3.26}$$

where each $\rho'_k(z)$ for $k = 1,\ldots,s$ is the eigenvalue of the matrix $D_2\mathbf{f}(1-z^p,\mathbf{N}(1-z^p))$, in a neighbourhood of $z = 0$ with m_k being its multiplicity. And $P'_k, \mathbb{D}'_k$ are matrices that are analytic in a neighbourhood of 0. Notice that due to Perron-Frobenius theorem, we have that $\rho'_1(0) = 1$, $m_1 = 1$, $\mathbb{D}_1(z) = 0$, and $\rho'_1(z) > |\rho'_k(z)|$ for all $z \in (0,1]$ for all $k > 1$.

For $z \in [0,1)$, substituting $z := (1-z)^{\frac{1}{p}}$ gives us:

$$R(z) = -\sum_{k=1}^{s}\left[P_k(z)\frac{1}{1-\rho_k(z)} + \mathbb{D}_k(z)\frac{1}{(1-\rho_k(z))^2} + \cdots + [\mathbb{D}_k(z)]^{m_k-1}\frac{1}{(1-\rho(z))^{m_k}}\right] \tag{3.27}$$

in a neighbourhood of 1. Where each $\rho_k(z)$ is the eigenvalue of $\mathbb{M}(z)$. Moreover, all of $P_k(z), \mathbb{D}_k(z), \rho_k(z)$ admit a Puiseux series of the form $\sum_{n=n_0}^{\infty} a_n(1-z)^{n/p}$ with $n_0 \geq 0$. With $\rho_1(1) = 1, P_1(1) = \lim_n \mathbb{M}^n$, $\mathbb{D}_1(z) = 0$, and $\rho'_1(z) > |\rho'_k(z)|$ for all z in a neighbourhood of 1.

General Density Result

Let H be a context-free grammar where the BP defined on it is given by Theorem 3.2.8. We use the notations $\partial_1\mathbf{f}$ and $D_2\mathbf{f}$ to refer to the following partial derivatives of $\mathbf{f}(z,\mathbf{x})$: $\frac{\partial \mathbf{f}}{\partial z}, \frac{\partial \mathbf{f}}{\partial \mathbf{x}}$.

Proposition 3.2.10. *Let H be an irreducible CFG such that Theorem 3.2.8 applies. Then,*

$$\begin{aligned}\mathbf{N}'(z) &= [I - D_2\mathbf{f}(z,\mathbf{N}(z))]^{-1}\,\partial_1\mathbf{f}(z,\mathbf{N}(z)) \\ &= \left[\mathbb{L}(z)\frac{1}{1-\rho(z)} + \mathbb{D}(z)\right]\partial_1\mathbf{f}(z,\mathbf{N}(z))\end{aligned} \tag{3.28}$$

where $\rho(z), \mathbb{L}(z), \mathbb{D}(z)$ have Puiseux series of the form $\sum_{n=n_0}^{\infty} a_n(1-z)^{n/p}$ in a neighbourhood of 1 where $n_0 \geq 0$. $\rho(z)$ is monotone and corresponds to the largest eigenvalue of $\mathbb{M}(z) := D_2\mathbf{f}(z,\mathbf{N}(z))$. We have $\rho(1) = 1$ and $\mathbb{L}(1) = \lim_n \mathbb{M}^n$

Proof. This follows from Eq. (3.27). ∎

Proposition 3.2.11. *Let H be an irreducible CFG such that Theorem 3.2.8 applies. Then,*

$$\begin{aligned}\partial_u|_{u=1}\mathbf{N}(z,u) &= [I - D_2\mathbf{f}(z,\mathbf{N}(z))]^{-1} f_1(z,\mathbf{N}(z))\mathbf{e}_1 \\ &= \left[\mathbb{L}(z)\frac{1}{1-\rho(z)} + \mathbb{D}(z)\right] f_1(z,\mathbf{N}(z))\mathbf{e}_1\end{aligned} \tag{3.29}$$

where $\rho(z), \mathbb{L}(z), \mathbb{D}(z), \mathbf{N}(z)$ have Puiseux series of the form $\sum_{n=n_0}^{\infty} a_n(1-z)^{n/p}$ in a neighbourhood of 1 where $n_0 \geq 0$. $\rho(z)$ is monotone and corresponds to the largest eigenvalue of $\mathbb{M}(z) := D_2\mathbf{f}(z,\mathbf{N}(z))$. We have $\rho(1) = 1$ and $\mathbb{L}(1) = \lim_n \mathbb{M}^n$.

Proof. This follows from Eq. (3.27). ■

Corollary 3.2.12. *(Density Result)*
Let H be an unambiguous CFG that satisfies the assumptions of Theorem 3.2.8. Then, for any type of variable $i \in V$,

$$\mathbb{E}_S(\#i|\#W = n) \sim Cn \tag{3.30}$$

where $\#W$ denotes the size of a uniformly randomly selected word. Moreover, let $\mathbb{M} = D_2\mathbf{f}(1,\mathbf{1})$ be the critical mean matrix of the branching process defined by Theorem 3.2.8, let $\mathbb{L} = \lim_{n\to\infty}\mathbb{M}^n$, and let $\mathbf{E} = \partial_1\mathbf{f}(1,\mathbf{1})$ so that E_i is the expected number of letters i produces in one generation. Then,

$$C = \frac{\mathbb{L}_{1,i}}{\mathbf{e}_1^T\mathbb{L}\mathbf{E}} \equiv \frac{\mathbb{L}_{1,i}}{\sum_j \mathbb{L}_{1,j}E_j} \tag{3.31}$$

Proof. This result leverages the fact that if we have two combinatorial generating functions $A(z) = \sum_n a_n z^n$ and $B(z) = \sum_n b_n z^n$ with dominant singularities at $z = \rho$. Then if there exists a constant $C > 0$ so that $CA(z) - B(z)$ does not have a singularity at $z = \rho$, then $a_n \sim Cb_n$. ■

3.3 General Pattern Results

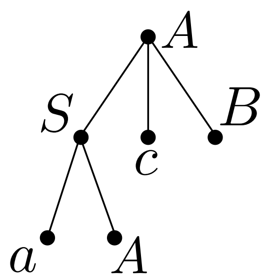
Before we look at general pattern results, we must first define what we mean by a pattern. A pattern is a finite sequence of production applications that transform a variable ($\in V$) into a word in $(V \cup T)^*$. So essentially, a production is an incomplete word derivation (where a derivation is defined in Definition 3.1.2) that can start from any variable symbol.

Definition 3.3.1. (Tree Patterns) A tree pattern T in a CFG, H, is a rooted finite connected subtree of a derivation tree T^* such that if a vertex v and one of its offspring are in T, then all offspring of v are in T. (See Fig. 3.2 for an example.)

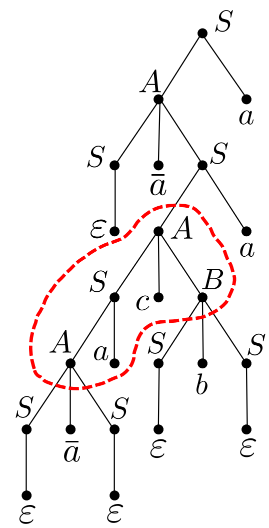
Equivalently, a tree pattern T is a rooted finite connected subtree of a derivation tree T^* that can be obtained from T^* by selecting a vertex v and deleting all of its descendants (but not v) repeatedly.

Another equivalent definition is that a tree pattern T is a tree that can be constructed from a root vertex by consecutive applications of productions in the language H. Thus, we sometimes represent the pattern as a sequence of productions and denote the this pattern as $P = \{p_v\}_{v\in V(T)}$ where each p_v represents the production v uses to obtain its offspring in T. When v has no offspring, we say $p_v = \emptyset$.

And a final equivalent representation of a tree pattern T is an incomplete derivation. (Derivations



(a) A tree pattern



(b) A derivation tree with the pattern (a) identified.

Figure 3.2: Tree pattern (from Example 5.0.2)

are defined at Definition 3.1.2)

In this section, our goal is to count tree patterns in random words.

Definition 3.3.2. (Tree Contraction) Given a tree pattern T, we contract T into a production $p(T)$ in the following manner. Let ω be the word in $[V(H)]^*$ that would result if T is interpreted as a complete derivation tree. i.e., ω is the word produced when T is interpreted as a derivation. Then $p(T) := \text{type}(\rho) \to \omega$ where ρ is the root of T. For example the pattern in Fig. 3.2a would be contracted to the production $A \to AacB$

Depth of a rooted tree T is the length of the largest path from the root. And for a rooted tree T, $[T]_n$ denotes the subtree of T that includes all edges and vertices of distance n or less. (The first n generations of T.)

Definition 3.3.3. Finally, given a tree pattern T, we will define S_n, the set of all tree patterns similar to T but differ from T in n-th generation, defined by:

$$S_n(T) = \left\{ T' \text{ is a tree pattern in } H : \begin{array}{lc} (1) & T' \text{ has depth } n, \text{ i.e. } T' = [T']_n \\ (2) & [T']_{n-1} = [T]_{n-1}. \\ (3) & \text{if } v \in [T']_{n-1} = [T]_{n-1}, \text{ then, } \#_T v > 0 \iff \#_{T'} v > 0 \\ (4) & T' \neq [T]_n \end{array} \right\} \tag{3.32}$$

for $n \geq 1$, where $\#_T v$ denotes the number of offspring $v \in V(T)$ has in the tree T. Notice that this is essentially the set of tree patterns in H of depth n that agree with T on the first $n-1$ generations and differ from T in at least one vertex of depth n (with one more caveat about which vertices can have offspring).

To make analysing patterns easier, we will reduce the case of counting patterns into counting a single edge. Before we make any changes on the language H, apply Theorem 3.2.8 to define the critical branching process. Then, we will transform the branching process in the following manner:

Lemma 3.3.4. *(Memory Lemma) Suppose that H is an unambiguous grammar that satisfies the assumptions of Theorem 3.2.8 and has the critical branching process defined on it so that $\mathbf{F}(\frac{1}{\lambda}, \psi) = \psi$. Let T be a tree pattern in this grammar starting from $j \in V$ and ending at $w \in (V(H))^*$.*

Then there exists another unambiguous grammar H' such that:

(i) it is equivalent to H (i.e. $L(H) = L(H')$),

(ii) $V(H') = V(H)$, and $\mathcal{P}(H) \setminus \{p_1\} \subset \mathcal{P}(H')$,

(iii) satisfies the assumptions of Theorem 3.2.8, and has a critical branching process, $\mathbb{P}'$, defined on it so that $\mathbf{F}'(\frac{1}{\lambda}, \psi) = \psi$,

(iv) for all $\omega \in L(H)$, and all $i \in V$, $\mathbb{P}'_i(\omega) = \mathbb{P}_i(\omega)$,

(v) if T is a non-self-containing pattern, then the number of occurrences of the pattern P in H corresponds to the number of occurrences of the production rule $p(T)$ in H'

Moreover, under the critical BP of H', one has that:

$$\mathbb{P}'\left(p(T)\right) = \prod_{v \in T}^{N} \mathbb{P}(p_T(v)) \tag{3.33}$$

Proof. Let N be the depth of T. If $N = 1$, take $H' = H$, this gives us all of $(i) - (v)$. Thus suppose $N \geq 2$.

In this case, consider the collection of sets $\{S_n(T)\}_{n=1}^{N}$ from Definition 3.3.3 (note that $S_k = \emptyset$ for $k \geq N + 1$). Then we convert all of these trees into productions by contracting them using Definition 3.3.2 so that $p(S_n) = \{p(T') : T' \in S_n\}$. Then we can define the new language H' in the following manner, $V(H') := V(H)$,

$$\mathcal{P}(H') = (\mathcal{P}(H) \setminus \{p_1\}) \cup \left(\bigcup_{n=1}^{N} p(S_n)\right) \cup \{p(T)\} \tag{3.34}$$

where p_1 is the production used by the root of T to branch in T.

Now we will prove the desired properties of this language. We first prove that the two languages are equivalent. We do this by picking derivation trees from both languages and converting them to derivation trees in the other one. Let's prove the easier direction first. Pick a derivation tree Y in H'. If Y contains no productions from $\left(\bigcup_{n=1}^{N} p(S_n)\right) \cup \{p(T)\}$, then it is already a derivation tree in H. Thus, suppose Y has productions from $\left(\bigcup_{n=1}^{N} p(S_n)\right) \cup \{p(T)\}$. Let p be such a production, then $p = p(T')$ for some tree pattern T' in H. Then one sees that simply replacing the production p with the tree T' in Y will turn it into a tree in H of the same word (and this relation is injective).

For the other direction, pick a derivation tree Y in H. If Y only contains productions in $\mathcal{P}(H) \setminus \{p_1\}$, then it is already a derivation tree in H'. Then suppose Y contains $p_1 \in \mathcal{P}(H)$. Let p be a production in Y with minimal depth such that type$(p) = p_1$. Here the idea is that p is the start of a pattern that is equal to T (in which case it can be replaced by $p(T)$) or it differs from T in a generation $n \leq N$ (in which case it can be replaced by $p(T')$ where $T' \in S_n$). Let Y_p be the subtree of Y started from p (this is a pattern in H). We know that $[Y_p]_1 = [T]_1$ since they start with the same production. In this context, we say that $[Y_p]_n \equiv [T]_n$ if the following hold true:

1. $[Y_p]_{n-1} \equiv [T]_{n-1}$.
2. $\forall v \in [T]_{n-1} \equiv [Y_p]_{n-1}, \#_T v > 0 \implies \rho_T(v) = \rho_{Y_p}(v)$

where $\rho_T(v)$ represents the production that v uses to branch in T, and $\rho_{Y_p}(v)$ is the production used to branch out v in Y_p (allowed to be $= \emptyset$).

Let $K = \inf\{1 \leq n \leq N : [T]_n \not\equiv [Y_p]_n\} \wedge (N+1)$ (where $\inf \emptyset = \infty$). Then $K \in [1, N+1] \cap \mathbb{N}$, now notice that if $K = N+1$, then $[Y_p]_K$ can be replaced by $p(T)$ in Y by contraction. Otherwise if $K \leq N$, then $[Y_p]_K$ can be replaced by some production in $p(S_K)$ since all patterns that differ from T in the K-th generation are in S_K. Thus, we were able to replace $[Y_p]_K$ (and therefore p) with a production from $\mathcal{P}(H')$. Repeat this from top down until Y has no p_1 productions left (the reason we go from top down is so that at each iteration, the subtree we consider is a pattern in H), thus, becoming a tree in H'. This proves equivalence of languages. And we can see that this preserves unambiguity since if we have two different derivation trees for the same word in H', we easily can convert it to two different derivation trees in H using the algorithm from the easy direction.

Properties (iii)-(v) are easy to see once one notices that the following probability assignment for each production in $\mathcal{P}(H')$ works:

$$\mathbb{P}_{H'}(p) = \begin{cases} \mathbb{P}_H(p) & \text{if } p \in \mathcal{P}(H) \setminus \{p_1\} \subset \mathcal{P}(H') \\ \prod_{v \in V(T')} \mathbb{P}_H(\rho_{T'}(v)) & \text{if } p = p(T') \in \left(\bigcup_{n=1}^N p(S_n)\right) \cup \{p(T)\} \end{cases} \tag{3.35}$$

■

In the previous section, we looked the number of times a certain node appears, conditioned on the size of the word. To analyze general patterns, we will first count the number of times a certain production/edge is used in the BP, then we will generalize this to all patterns.

Suppose that $\rho = (j \to w_0) \in \mathcal{P}$ is the production we want to count. Then, for all $i \in V$, define the following generating function:

$$\begin{aligned} N_i(z, q) &= \sum_{n=0}^{\infty} \sum_{\mathbf{k} \in \mathbb{N}^m} \mathbb{P}_i(\#T = n, \#_\rho T = k) z^n q^k \\ N_i(z) = N_i(z, 1) &= \sum_{n=0}^{\infty} \mathbb{P}_i(\#T = n) z^n \end{aligned} \tag{3.36}$$

where $\#_\rho T$ denotes the number of times the production ρ appears in the tree T.

Consider the function $\mathbf{F}(z, \mathbf{x}, q)$ defined below where $\mathbf{F}(z, \mathbf{x}, 1) = \mathbf{F}(z, \mathbf{x})$ and this new function is the same as $\mathbf{F}(z, \mathbf{x})$ except F_j is modified so that the multinomial term corresponding to the edge

$p = (j \to w_0)$ is multiplied by a q:

$$F_i(z, \mathbf{x}, q) = \sum_{p=(i\to w)\in\mathcal{P}_i} z^{\#(w)} \mathbf{x}^{\vec{\#}(w)} q^{\mathbf{1}\{p=\rho\}} \tag{3.37}$$

We also modify $\mathbf{f}(z, \mathbf{x})$ similarly to obtain $\mathbf{f}(z, \mathbf{x}, q)$. We will also use the following notation to refer to the corresponding partial derivatives of $\mathbf{f}(z, \mathbf{x}, q)$, where $\partial_1 \mathbf{f} = \frac{\partial \mathbf{f}}{\partial z}$, $D_2\mathbf{f} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}$, and $\partial_3 \mathbf{f} = \frac{\partial \mathbf{f}}{\partial q}$

Proposition 3.3.5.

$$\mathbf{N}(z, q) = \mathbf{f}(z, \mathbf{N}(z, q), q) \tag{3.38}$$

Proof. We use the notation $[z^n, q^k]\mathbf{N}(z)$ to denote the coefficient of the term z^n, q^k in $\mathbf{N}(z, q)$. We will now show that the RHS is equal to LHS. Fix $i \in V$,

$$\begin{aligned}
\mathrm{RHS}_i &= \sum_{p=(i\to\omega)\in\mathcal{P}_i} \mathbb{P}(p) z^{\#(\omega)} (\mathbf{N}(z, q))^{\vec{\#}(\omega)} q^{\mathbf{1}\{p=\rho\}} \\
[z^n, q^k]\mathrm{RHS}_i &= \sum_{p=(i\to\omega)\in\mathcal{P}_i} \mathbb{P}(p)[z^{n-\#(\omega)}, q^{k-\mathbf{1}\{p=\rho\}}][\mathbf{N}(z, q)]^{\vec{\#}(\omega)} \\
&= \sum_{p=(i\to\omega)\in\mathcal{P}_i} \mathbb{P}(i \to \omega)\mathbb{P}_{\vec{\#}(\omega)}(\#T = n - \#(\omega), \#_\rho T = k - \mathbf{1}\{p = \rho\}) \\
&= \mathbb{P}_i(\#T = n, \#_\rho T = k)
\end{aligned} \tag{3.39}$$

where $\mathbb{P}_{\vec{\#}(w)}(\#T = n, \#_\rho T = k)$ refers to the probability of $\#T = n$ and $\#_\rho T = k$ where we start a forest T with multiple roots ($\#_i(w)$ many roots of type i for each type $i \in V$). The step that needs induction for justification is that:

$$[z^n, q^k][\mathbf{N}(z, \mathbf{u})]^{\mathbf{m}} = \mathbb{P}_{\mathbf{m}}(\#T = n, \#_\rho T = k) \tag{3.40}$$

■

Now in order to find $\mathbb{E}(\#j \to w_0 \text{ occurs}|\#T = n)$, we follow the same method from previous section. Since we already know Proposition 3.2.10, we can directly jump to finding $\partial_q|_{q=1}\mathbf{N}$.

Proposition 3.3.6. *Let H be a grammar that satisfies the assumptions of Theorem 3.2.8, then we have that*

$$\partial_q|_{q=1}\mathbf{N} = [I - D_2\mathbf{f}(z, \mathbf{N}(z), 1)]^{-1}\mathbb{P}(j \to w_0) z^{\#(w_0)} \mathbf{N}(z)^{\vec{\#}(w_0)} \mathbf{e}_j \tag{3.41}$$

Proof. This simply follows from differentiating ■

Corollary 3.3.7. *Let H be a grammar that satisfies the assumptions of Theorem 3.2.8, then we have that*

$$\mathbb{E}_S(\#(j \to w_0)|\#W = n) \sim Cn \tag{3.42}$$

where $\#W$ denotes the size of a uniformly randomly selected word. Moreover, let $\mathbb{M} = D_2\mathbf{f}(1, \mathbf{1})$ be the critical mean matrix of the branching process defined by Theorem 3.2.8, let $\mathbb{L} = \lim_{n\to\infty} \mathbb{M}^n$, and let $\mathbf{E} = \partial_1\mathbf{f}(1, \mathbf{1})$ so that E_i is the expected number of letters i produces in one generation. Then,

$$C = \frac{\mathbb{P}(j \to w_0)\mathbb{L}_{S,j}}{\mathbf{e}_S^T \mathbb{L}\mathbf{E}} \equiv \mathbb{P}(j \to w_0) \frac{\mathbb{L}_{S,j}}{\sum_i \mathbb{L}_{S,i} E_i} \tag{3.43}$$

Proof. This corollary follows from Proposition 3.3.6 and Proposition 3.2.10. ■

In the next corollary, we will denote the different variables of a language by S_i with S_1 denoting S

Corollary 3.3.8. *(Easy way to Find Densities) Let H be a grammar that satisfies the assumptions of Theorem 3.2.8, let $m = |V|$, and let $\mathbf{C}$ be defined as below,*

$$\mathbf{C} = \lim_{n\to\infty} \frac{1}{n} \begin{bmatrix} \mathbb{E}(\#_1 T \mid \#T = n) \\ \mathbb{E}(\#_2 T \mid \#T = n) \\ \vdots \\ \mathbb{E}(\#_1 T \mid \#T = n) \end{bmatrix} \tag{3.44}$$

Then $\mathbf{C}$ is the solution to the following linear system of equations:

$$\begin{gathered} \mathbf{C}^T = \mathbf{C}^T \mathbb{M} \\ \mathbf{C} \cdot \mathbf{E} = 1 \end{gathered} \tag{3.45}$$

where $\mathbb{M} = D_2\mathbf{f}(1, \mathbf{1})$ is the mean matrix, and $\mathbf{E} = \partial_1 \mathbf{f}(1, \mathbf{1})$ is the mean letters vector.

Proof. This result follows from Corollary 3.2.12 and Corollary 3.3.7. $\mathbf{C}$ solving $\mathbf{C} \cdot \mathbf{E} = 1$ is trivial since a randomly selected node from a large tree has to produce, on average, 1 letter. And similarly, the first equation, $\mathbf{C}^T = \mathbf{C}^T \mathbb{M}$ follows from the fact that a randomly selected node from a large tree must on average produce C_i many variables of each type i. I will prove the second statement rigorously.

$$\begin{aligned} \mathbb{E}(\#_i T | \#T = n) &= \sum_{p\in\mathcal{P}(H)} \mathbb{E}(\#_p T | \#T = n) \cdot \#_i(\mathrm{RHS}(p)) \\ \therefore \quad C_i &= \sum_{p\in\mathcal{P}(H)} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}(\#_p T | \#T = n) \cdot \#_i(\mathrm{RHS}(p)) \\ \text{(by Corollary 3.3.7 and Corollary 3.2.12)} &= \sum_{p\in\mathcal{P}(H)} C_{\mathrm{LHS}(p)} \mathbb{P}(p) \#_i(\mathrm{RHS}(p)) \\ &= \sum_j C_j \sum_{p\in\mathcal{P}_j(H)} \mathbb{P}(p) \#_i(\mathrm{RHS}(p)) \\ &= \sum_j C_j \, \partial_i f_j(1, \mathbf{1}) \\ \therefore \quad \mathbf{C} &= \mathbb{M}^T \mathbf{C} \end{aligned} \tag{3.46}$$

where $\mathcal{P}_j(H) = \{p \in \mathcal{P}(H) : \mathrm{LHS}(p) = j\}$. ■

Corollary 3.3.9. *(Density of Patterns) Let H be an unambiguous grammar that satisfies the assumptions of Theorem 3.2.8, and let T be a non-self containing tree pattern and ω a word in $L(H)$, then*

1. *On a uniformly randomly chosen word of size n, there is a positive density of T and ω as $n \to \infty$*
2. *Let j denote the variable corresponding to the root of T. If all nodes of type j are either a root node or a leaf node in T, then the expectation density constant C of T is given by the following:*

$$C = C_j \prod_{p \in \mathcal{P}(T)} \mathbb{P}(p) = C_j \frac{1}{\lambda^{\#(RHS(T))}} \frac{\psi^{\vec{\#}(RHS(T))}}{\psi_j} \tag{3.47}$$

where $\mathcal{P}(T)$ is the multiset of all productions used in T and C_i is defined in Corollary 3.3.8.

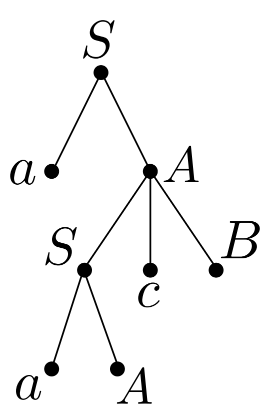
Self-containing Patterns

Since the memory lemma only works to count non-self containing patterns. However, it is worth noting that the memory lemma undercounts self containing patterns, so indeed, a lower bound on the density is achieved using memory lemma. We also prove a density result for all patterns (with exact density constants) in Corollary 4.2.2.

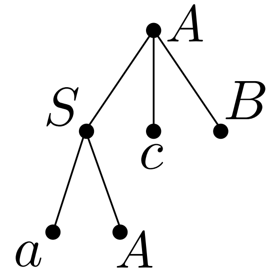
Here, it is worth noting what we mean by a self-containing pattern and noting that they can be decomposed into non self-containing patterns when applying the memory lemma. However, we will not explicitly do this since a density result will be given using different methods anyway.

Definition 3.3.10. (Self-containing Pattern) Let T be a tree pattern, for $v \in V(T)$, let T_v be the rooted sub-tree of T induced by v and its the descendants (thus, rooted at v). And let n be the depth of T_v.

Then, T is said to be self-containing if there exists a non-leaf and non-root vertex $v \in V(T)$, which has the same type as the root of T, such that $[T]_n$ can be extended (via applications of productions from $\mathcal{P}(H)$) to become T_v. (Note that this extension may not be needed and $[T]_n$ may be equal to T_v such as in Fig. 3.3a.)



(a) A self-containing tree pattern



(b) A non-self-containing tree pattern.

Figure 3.3: Self containment of patterns.

Variance Result

We will follow a similar strategy to the density result, thus, we will first prove a variance result for productions, and then extend this to patterns using the Memory Lemma. In [3], the method used to analyze occurances of patterns is to check that the bivariate generating function of the pattern satisifies the assumptions of Theorem 3.3.11, the theorem is also proved in that book. We want a density of variance so that condition (iii) in Theorem 3.3.11 can be verified. Theorem 3.3.11 tells us that the statistic we care about follows a gaussian law if we manage to verify three conditions on it's bivariate generating function. Condition (i) is trivial (especially if one is counting productions). In real examples, its hardest to check condition (iii) which ensures that the variance density is non-zero.

Theorem 3.3.11 (Algebraic Singularity Schema)**.** *flajolet2009analytic Let $F(z, u)$ be a bivariate analytic function at $(0, 0)$ with non-negative coefficients. Assume that:*

1. *(Analytic Perturbation) $\exists$ functions A, B, C analytic in $\mathcal{D} = \{|z| \leq r\} \times \{|u - 1| < \varepsilon\}$, such that the following representation holds in a subset of $\mathcal{D}$ for some $\alpha \notin \mathbb{Z}_{\leq 0}$*
$$F(z, u) = A(z, u) + B(z, u)C(z, u)^{-\alpha} \tag{3.48}$$
furthermore, assume that, in $|z| \leq r$, there exists a unique root ρ of $C(z, 1) = 0$, that this root is simple, and that $B(\rho, 1) \neq 0$.
2. *(Non-degeneracy) $\partial_z C(\rho, 1) \cdot \partial_u C(\rho, 1) \neq 0$.*
3. *(Variability) $\upsilon\left(\frac{\rho(1)}{\rho(u)}\right) \neq 0$ where $\upsilon(f) = f''(1) + f'(1) - [f'(1)]^2$*

Then the r.v. with the p.g.f. $p_n(u) = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}$ converges in distribution to a gaussian. The mean and standard deviation are asymp. linear in n.

Definition 3.3.12. (Interchangeable productions) Two productions p and p' are said to be interchangeable if $\mathrm{LHS}(p) = \mathrm{LHS}(p')$, $\eta(p) = \eta(p')$, and $\mathbf{y}(p) = \mathbf{y}(p')$ (where $\eta(p)$ = # of terminal letters in p and $y_i(p)$ = # of variables of type i appearing in p).

Definition 3.3.13. (Interchange operation) Given a derivation tree t, and two interchangeable productions p and p'. We can produce a new derivation tree t' by selecting a production of type p and replacing it by p' and reordering the sub-trees as required since the letters and variables of p and p' need not be ordered in the same manner.

Proposition 3.3.14. *Let H be an unambiguous grammar that satisfies the assumptions of Theorem 3.2.8. And let P and P' be two interchangeable productions in H. Then, $\exists K > 0$ such that*
$$\liminf_{n\to\infty} \frac{1}{n} Var(\#P \mid \#W = n) \geq K \tag{3.49}$$
and,
$$\liminf_{n\to\infty} \frac{1}{n} Var(\#P - \#P' \mid \#W = n) \geq 4K \tag{3.50}$$
Moreover, $K = \frac{qq'}{q+q'} C_{LHS(P)}$ where $q = q' = \mathbb{P}(P) = \mathbb{P}(P')$ and $C_{LHS(P)}$ is defined in Corollary 3.3.8.

Proof. Before starting the proof, fix a method of interchanging P with P' in a derivation tree, as there may be multiple ways to do this interchange for some patterns. For example, suppose you want to replace $S \to SASA$ with $S \to AASS$, you can choose which S and A should come first.

Let $T = \{t : t \text{ is a derivation tree in } H\}$ be the set of all derivation trees (starting from S) of H. We say that two trees $t, t' \in T$ are related, $t \sim t'$, if t can be converted to t' by interchanging P and P' with each other. Notice that this is an equivalence relation, let $\{T_\alpha\}_{\alpha\in\mathcal{A}}$ be the collection of equivalence classes of this relation.

Fix n, let $[T_\alpha]_n = \{t \in T_\alpha : \text{word length of } t \ = n\}$. Define a random variable X_n using the following distribution on $\mathcal{A}$,

$$\mathbb{P}(X_n = \alpha) = \frac{\#[T_\alpha]_n}{\#[T]_n} \tag{3.51}$$

(Note that X_n has finite support.) Given $\alpha \in \mathcal{A}$, define a random variable Y_α as the number of times P appears in a uniformly randomly selected member of T_α. First notice that $[T_\alpha]_n$ is either the empty set or T_α itself, since all trees in the same equivalence class must have the same word size and even the same number of total occurrences of P and P', thus, we can let N_α be the number of total occurrences of P and P' in $t \in T_\alpha$. Then, notice that Y_{X_n} is the random variable corresponding to the number of times P appears in a uniformly randomly selected member of $[T]_n$. We have that $Y_\alpha \sim \text{Bin}(N_\alpha, \kappa)$, where $\kappa = \frac{q}{q+q'}$, $q = \mathbb{P}(P)$, and $q' = \mathbb{P}(P')$.

$$\begin{aligned}
\frac{1}{n}\text{Var}(\#P \mid \#W = n) = \frac{1}{n}\text{Var}(Y_{X_n}) &= \frac{1}{n}\mathbb{E}(\text{Var}(Y_{X_n}|X_n)) + \frac{1}{n}\text{Var}(\mathbb{E}(Y_{X_n}|X_n)) \\
&\geq \frac{1}{n}\mathbb{E}(\text{Var}(Y_{X_n}|X_n)) = \mathbb{E}(N_{X_n}\kappa(1-\kappa)) = \kappa(1-\kappa)\frac{1}{n}\mathbb{E}(N_{X_n}) \\
\therefore \ \liminf_n \frac{1}{n}\text{Var}(\#P \mid \#W = n) &\geq \kappa(1-\kappa)C_{\text{LHS}(P)}(q+q') > 0
\end{aligned} \tag{3.52}$$

where $C_{\text{LHS}(P)}$ is defined in Corollary 3.3.8. Last step holds because N_{X_n} = total number of P and P' occurring in a uniformly selected member of $[T]_n$, and we already have a density result for this in Corollary 3.3.7. This proves Eq. (3.49), to prove Eq. (3.50), we define Y'_α similarly to Y_α for P' instead of P. Notice that:

$$\mathbb{E}(\text{Var}(Y_{X_n} - Y'_{X_n}|X_n)) = \mathbb{E}(\text{Var}(2Y_{X_n} - N_{X_n}|X_n)) = \mathbb{E}(\text{Var}(2Y_{X_n}|X_n)) \tag{3.53}$$

since $Y_{X_n} + Y'_{X_n} = N_{X_n}$ which has 0 variance when conditioned on X_n. ■

We prove the variance result for patterns now.

Definition 3.3.15. (Interchangeable Patterns) Let P and P' be two tree patterns. We say that these patterns are interchangeable if their tree contractions, $p(P)$ and $p(P')$ (defined in Definition 3.3.2), are interchangeable productions as defined in Definition 3.3.12.

Definition 3.3.16. (Pattern Containment) Let T and T' be two distinct tree patterns. We say that T contains T', if there exists a non-leaf vertex $v \in T$ such that T_v can be extended (by further applications of productions) to contain an occurrence T'. (Recall that T_v is the induced subgraph of v and its descendants in T.)

Remark. The definition of self-containment is the same as above, except, v is not allowed to be the root of the pattern T.

Lemma 3.3.17. *(Double Memory Lemma) Suppose that H is an unambiguous grammar that satisfies the assumptions of Theorem 3.2.8 and has the critical branching process defined on it so* *that* $\mathbf{F}(\frac{1}{\lambda}, \psi) = \psi$. *Let* T_1 *and* T_2 *be two non-self-containing tree patterns such that neither contains the other.*

Then there exists another unambiguous grammar H' *such that:*

(i) it is equivalent to H *(i.e.* $L(H) = L(H')$*),*

(ii) $V(H') = V(H)$*, and* $\mathcal{P}(H) \setminus \{p_1\} \subset \mathcal{P}(H')$*,*

(iii) satisfies the assumptions of Theorem 3.2.8, and has a critical branching process, $\mathbb{P}'$*, defined on it so that* $\mathbf{F}'(\frac{1}{\lambda}, \psi) = \psi$*,*

(iv) for all $\omega \in L(H)$*, and all* $i \in V$*,* $\mathbb{P}'_i(\omega) = \mathbb{P}_i(\omega)$*,*

(v) the number of occurrences of the patterns T_1, T_2 *in* H *corresponds to the number of occurrences of the production rules* $p(T_1)$ *and* $p(T_2)$ *in* H' *respectively.*

Moreover, under the critical BP of H'*, one has that:*

$$\mathbb{P}'\left(p(T_{1,2})\right) = \prod_{p \in \mathcal{P}(T_{1,2})} \mathbb{P}(p) \tag{3.54}$$

where $\mathcal{P}(T_{1,2})$ *denotes the multiset of productions used in* T *(so with repeats).*

Proof. We will prove this lemma by the application of Memory lemma twice. Apply the Memory Lemma (Lemma 3.3.4) to T_1. This creates the new grammar H' with $V(H') = V(H)$ and

$$\mathcal{P}(H') = (\mathcal{P}(H) \setminus \{p_1\}) \cup \left(\bigcup_{n=1}^{N} p(S_n[T]) \right) \cup \{p(T)\} \tag{3.55}$$

where $\{S_n(T)\}$ are defined in Definition 3.3.3. Now we must show that there is a tree pattern T_2' in H' that is equivalent to T_2 in H. In Memory Lemma, we have shown that each complete derivation tree in H has an equivalent derivation tree in H'. In general, this is not true for tree patterns. It is not the case that every tree pattern in H has an equivalent tree pattern in H' (for example just take the first production of T_1 to be your pattern in H). (Remark that every tree pattern in H' has an equivalent tree pattern in H.) Nevertheless, we will show that T_2 in H has an equivalent pattern in H' with the assumption that T_1 and T_2 do not contain each other.

Let $\rho \in \mathcal{P}(H)$ be the root production of T_1. We convert T_2 in the following manner, if T_2 contains no ρ, then T_2 is already a pattern in H' and we are done. So suppose T_2 contains ρ productions. Let p be a production in T_2 with minimal depth. Let N be the depth of T_2, Y_p be the subtree of T_2 generated by vertices of p and their descendants in T_2. This is a pattern in H. We know that $[Y_p]_1 = [T_1]_1$ since they start with a ρ production. We say that $[Y_p]_n \equiv [T_1]_n$ if the following two hold true:

1. $[Y_p]_{n-1} \equiv [T_1]_{n-1}$.
2. $\forall v \in [T]_{n-1} \equiv [Y_p]_{n-1}, \#_{T_1} v > 0 \implies \rho_{T_1}(v) = \rho_{Y_p}(v)$

where $\rho_{T_1}(v)$ represents the production that v uses to branch in T_1, and $\rho_{Y_p}(v)$ is the production used to branch out v in Y_p (allowed to be $= \emptyset$). And $\#_{T_1} v$ is the number of offspring of the vertex v in T_1.

Let $K = \inf\{1 \leq n \leq N : [T]_n \not\equiv [Y_p]_n\}$ (where $\inf \emptyset = \infty$). Notice that $K = \infty$ is impossible because T_2 does not contain T_1. Then if $1 \leq K \leq N$, then $[Y_p]_K$ can be replaced by some production in $p(S_K(T_1))$ since all patterns that differ from T_1 in the K-th generation are in S_K.

Then we repeat this process (from top down) until no ρ productions remain in T_2. The reason we do this top down (i.e. pick p with minimal depth) is so that Y_p is a tree pattern in H. This gives a new tree pattern T_2' in H' that is non self-containing (since T_2 was originally non self-containing). And the number of times this tree pattern occurs in a derivation tree in H' corresponds to the number of times T_2 occurs in the corresponding derivation tree in H.

Also recall that T_1 is reduced to a production, $p(T_1)$ in H'. We can apply the Memory Lemma to H' and the pattern T_2' to complete the proof. Note that the second application cannot delete the production $p(T_1)$ since T_2' cannot start by $p(T_1)$ because T_2 does not contain T_1. As in the original memory lemma, assignment of probabilities to each production is trivial. ∎

Corollary 3.3.18. *Let H be an unambiguous CFG satisfying the assumptions of Theorem 3.2.8.*

And let T and T' be two interchangeable non self-containing tree patterns such that neither of them contains the other one. Then, $\exists K > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} Var(\#T \mid \#W = n) \geq K \tag{3.56}$$

and,

$$\liminf_{n\to\infty} \frac{1}{n} Var(\#T - \#T' \mid \#W = n) \geq 4K \tag{3.57}$$

Moreover if both the patterns satisfy the assumption (b) of Corollary 3.3.9, then $K = \frac{qq'}{q+q'} C_{root(T)}$ where q and q' are respectively the product of all the probabilities of productions present in T or T'.

Proof. This follows directly from the Memory Lemma (Lemma 3.3.4) and Proposition 3.3.14. ∎

One might also want to count multiple patterns together, one example of this is given in Example 5.0.3. The density result for multiple patterns can easily be deduced from things we already did, so we will focus on variance.

Proposition 3.3.19. *Let H be an unambiguous CFG satisfying the assumptions of Theorem 3.2.8.*

Let P_1, P_1', P_2, P_2' be four distinct productions such that P_1 and P_2 are interchangeable with P_1' and P_2' respectively. Then, $\exists K > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} Var(\#P_1 + \#P_2 \mid \#W = n) \geq K \tag{3.58}$$

and,

$$\liminf_{n\to\infty} \frac{1}{n} Var([\#P_1 + \#P_2] - [\#P_1' + \#P_2'] \mid \#W = n) \geq 4K \tag{3.59}$$

Moreover, $K = \frac{pp'}{p+p'} C_{LHS(P_1)} + \frac{qq'}{q+q'} C_{LHS(P_2)}$ where $p = p' = \mathbb{P}(P_1) = \mathbb{P}(P_1')$ and $q = q' = \mathbb{P}(P_2) = \mathbb{P}(P_2')$ and $C_{LHS(P)}$ is defined in Corollary 3.3.8.

Proof. This proof will be similar to the proof of Proposition 3.3.14 and can be extended to more than 2 productions. Define T and $[T]_n$ as in the proof of Proposition 3.3.14 and fix a method of interchanging productions as in the mentioned proof.

Then we say that two trees $t, t' \in T$ are related, $t\ t'$, if t can be converted to t' by consequent interchanging of P_1 with P_1' and/or P_2 with P_2' or vice versa. This is an equivalence relation. Let

$\{T_\alpha\}_{\alpha\in\mathcal{A}}$ be the collection of equivalence classes.

The rest of the proof is very similar to that of Proposition 3.3.14. ■

Corollary 3.3.20. *Let H be an unambiguous CFG satisfying the assumptions of Theorem 3.2.8.*

Let P_1, P_1', P_2, P_2' be four distinct non self-containing tree patterns such that no pair of two patterns containing each other. Suppose P_1 and P_2 are interchangeable with P_1' and P_2' respectively. Then, $\exists K > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} Var(\#P_1 + \#P_2 |\, \#W = n) \geq K \tag{3.60}$$

and,

$$\liminf_{n\to\infty} \frac{1}{n} Var([\#P_1 + \#P_2] - [\#P_1' + \#P_2'] |\, \#W = n) \geq 4K \tag{3.61}$$

Moreover, $K = \frac{pp'}{p+p'} C_{LHS(P_1)} + \frac{qq'}{q+q'} C_{LHS(P_2)}$ where $p = p' = \mathbb{P}(P_1) = \mathbb{P}(P_1')$ and $q = q' = \mathbb{P}(P_2) = \mathbb{P}(P_2')$ and $C_{LHS(P)}$ is defined in Corollary 3.3.8.

Chapter 4

Fringe Convergence

The goal in this chapter is to prove a fringe convergence result using Theorem 3.2.8 since the the random word from the context free grammar can be reduced to a multitype branching process. This will allow us to answer the question of what a uniformly randomly picked fringe subtree looks like in a uniformly selected derivation tree of size n.

We will use this to generalize the density result of Corollary 3.3.9 to all patterns and exactly obtain the density constant for all patterns (not just non self-containing ones).

4.1 Fringe Convergence of Multitype Galton-Watson Process

The concept of fringe convergence was introduced by Aldous in [1], and theory developed mainly for single type trees. In [5], Stufler proves similar results for multitype trees. In this section, we introduce relevant definitions and results.

Definition 4.1.1 (V-type Plane Tree)**.** Let V be a countable non-empty set whose elements are called *types*. A *V-type plane tree* is a plane tree together with a map that assigns to each vertex a type from V. (In [5], author only considers trees whose vertices have finite number of offspring. i.e. *locally finite* trees.)

Definition 4.1.2 (Marked Trees)**.** Let T denote a locally finite V-type tree and let $v \in T$ be one of its vertices. We say that the tree is *marked* at v. The pair (T, v) is called a *marked tree*. The directed path from v to the root of T is called the *spine*.

Definition 4.1.3 (Fringe Subtree)**.** Let T denote a locally finite V-type tree and let $v \in T$ be one of its vertices. The tree $f(T, v)$ formed by v and all of its descendants is called the *fringe subtree* of T at v.

In this Chapter, we will be making use of a lemma shown in [5], although other results from this paper could be useful in other contexts.

Lemma 4.1.4. *Consider a conditioned multitype branching process:*

$$T_n := (T \mid \mathcal{E}_n) \tag{4.1}$$

for some family of events $\mathcal{E}_n$ satisfying $\mathbb{P}(\mathcal{E}_n) > 0$ for all n. Suppose that there is a deterministic sequence $s_n \to \infty$ satisfying for all $\varepsilon > 0$,

$$\exp(-\varepsilon s_n) = o(\mathbb{P}(\mathcal{E}_n)) \text{ and } \mathbb{P}(\#_i T_n \geq s_n) \to 1 \tag{4.2}$$

Then for any finite V-type tree t with root type i,

$$\frac{\#_t T_n}{\#_i T_n} \xrightarrow{p} \mathbb{P}(T(i) = t) \tag{4.3}$$

where $T(i)$ is a multitype branching process started at type i. Let $f(T_n, v_n^i)$ denote the fringe subtree in T_n encountered at the uniformly selected type i vertex v_n^i of T_n. Then,

$$\mathcal{L}(f(T_n, v_n^i) \mid T_n) \xrightarrow{p} \mathcal{L}(T(i)) \tag{4.4}$$

4.2 Fringe Convergence of Context-Free Grammars

Now we apply Lemma 4.1.4 to the setup we have in Theorem 3.2.8.

Proposition 4.2.1. *Let H be an unambiguous CFG satisfying the assumptions of Theorem 3.2.8. Let T_n be the derivation tree of a uniformly randomly selected word of size n. Fix a variable $i \in V$, let v_n^i denote a uniformly selected vertex of type i from T_n. Then,*

$$\mathcal{L}(f(T_n, v_n^i) \mid T_n) \xrightarrow{p} \mathcal{L}(T(i)), \tag{4.5}$$

Proof. Let T be a tree sampled according to the branching process given in Theorem 3.2.8. Let $\mathcal{E}_n := \{\#T = n\}$ (here "#" refers to the number of terminal letters). It suffices to check the assumptions of Lemma 4.1.4.

Recall that $\mathbb{P}(\#T = n)$ are the coefficients of the generating function $\mathbf{N}(z)$ which solves $N(z) = \mathbf{f}(z, \mathbf{N}(z))$ (shown in Proposition 3.2.9), Theorem VII.5 from [3] gives us that

$$\mathbb{P}(\#T = n) \sim Cn^{-\alpha} \tag{4.6}$$

where in fact $\alpha > 0$ (all of this is being done over some subsequence of n so that $\mathbb{P}(\mathcal{E}_n)$ are all positive and are asymptotically described by their algebraic singularity). Let s_n be a deterministic sequence so that $s_n = o(n)$ and $\log(n) = o(s_n)$. Then. for any $\varepsilon > 0$

$$\begin{gathered} \frac{\exp(-\varepsilon s_n)}{Cn^{-\alpha}} = \exp(-\varepsilon s_n + \alpha \log(n)) = \exp(-\varepsilon s_n[1 - o(1)]) = o(1) \\ \therefore \quad \exp(-\varepsilon s_n) = o(\mathbb{P}(\mathcal{E}_n)) \end{gathered} \tag{4.7}$$

Now we prove the second assumption,

$$\mathbb{P}(\#_i T_n \geq s_n) = \mathbb{P}\left(\frac{\#_i T}{n} \geq \frac{s_n}{n} \mid \#T = n\right) \geq \mathbb{P}\left(\frac{\#_i T}{n} \geq \frac{s_n}{n} \mid \#T \geq n\right) \to 1 \tag{4.8}$$

And the final claim can be argued either combinatorially or probabilistically since $\frac{s_n}{n} \to 0$. ∎

From this, it is easy to obtain a density result for all patterns with an exact constant (unlike Corollary 3.3.9, this gives us the exact constant for all patterns. not just non self containing ones).

Corollary 4.2.2. *Let P be a tree pattern rooted by variable i in an unambiguous CFG satisfying the assumptions of Theorem 3.2.8. Let T_n be a uniformly randomly selected derivation tree of a word of size n. Then, there exists a sequence $k(n)$ so that*

$$\begin{aligned} \frac{\#_P T_{k(n)}}{\#_i T_{k(n)}} &\xrightarrow{a.s.} \mathbb{P}(T(i) = P) \\ \therefore \quad \mathbb{E}\left(\frac{\#_P T_{k(n)}}{k(n)}\right) &\to C_i\, \mathbb{P}(T(i) = P) \end{aligned} \tag{4.9}$$

where C_i is defined as in Corollary 3.3.8 and $T(i)$ is the unconditioned branching process starting from a vertex of type i.

Proof. Since Proposition 4.2.1 shows that this setup satisfies Lemma 4.1.4, it can be shown that,

$$\frac{\#_P T_n}{\#_i T_n} \xrightarrow{p} \mathbb{P}(T(i) = P) \tag{4.10}$$

(recall that we are always working over a subsequence where $\mathbb{P}(\#T = n) > 0$). Then, there exists a subsequence $k(n)$ such that $\mathbb{P}(\#T = k(n)) > 0$ and

$$\frac{\#_P T_{k(n)}}{\#_i T_{k(n)}} \xrightarrow{a.s.} \mathbb{P}(T(i) = P) \tag{4.11}$$

Then we consider:

$$\mathbb{E}\left(\frac{\#_P T_{k(n)}}{k(n)}\right) = C_i \mathbb{E}\left(\frac{\#_P T_{k(n)}}{\#_i T_{k(n)}}\right) + \mathbb{E}\left(\left[\frac{\#_P T_{k(n)}}{\#_i T_{k(n)}} - \mathbb{P}(T(i) = P)\right]\frac{\#_i T_n}{n}\right) \tag{4.12}$$

We note that $\frac{\#_P T_{k(n)}}{\#_i T_{k(n)}} \leq 1$ a.s. since the root of P is of type i. Moreover, there exists a constant $M > 0$ so that $\frac{\#_i T_n}{n} \leq M$ a.s. This can be shown combinatorically. Then since we have that $\left|\frac{\#_P T_{k(n)}}{\#_i T_{k(n)}} - \mathbb{P}(T(i) = P)\right| \to 0$ a.s. and $\left|\frac{\#_P T_{k(n)}}{\#_i T_{k(n)}} - \mathbb{P}(T(i) = P)\right| \leq 2$ a.s. DCT gives us that:

$$\lim_{n\to\infty} \frac{1}{k(n)} \mathbb{E}\#_P T_{k(n)} = C_i \mathbb{P}(T(i) = P) \tag{4.13}$$

■

Note that here $\mathbb{P}(T(i) = P)$ is given by

$$\mathbb{P}(T(i) = P) = \prod_{p\in\mathcal{P}(P)} \mathbb{P}(p) = \frac{1}{\lambda^{\#(\mathrm{RHS}(P))}} \frac{\psi^{\vec{\#}(\mathrm{RHS}(P))}}{\psi_{\mathrm{LHS}(P)}} \tag{4.14}$$

where $\mathcal{P}(P)$ is the multiset of all the productions present in P.

Chapter 5

Examples

In this section, examples of the above methods are presented.

Example 5.0.1. (A Non-Example) This example is to demonstrate that the irreducibly assumption in Corollary 3.2.12 is necessary. Consider the Dyck language/path and uniformly randomly sample a word of size n, then consider the statistic of the number of times the path hits 0. We know that there should be no density result for this statistic.

The grammar for this example is given by:

$$\begin{aligned} S &\to \varepsilon \mid (S')M \\ M &\to \varepsilon \mid S \\ S' &\to \varepsilon \mid (S')S' \end{aligned} \tag{5.1}$$

This grammar can clearly be converted to 2-normal form. And it will still not be irreducible.

Example 5.0.2. (3 Strand Braid Group - Counting Patterns)
This example is a version of the example from Section 5 of [2]. There, they focus on the combinatoric generating functions of such examples. Consider the following infinite tree of triangles in Fig. 5.1.

In this example, we are interested in the paths of size n that start and end at e. More specifically, suppose that you pick a uniformly randomly selected path of size n that starts and ends at e. Then, we are interested in the statistic of how many times you go around a triangle in counter-clockwise direction.

We will use a CFG to represent such paths. Each word of size n in the following language uniquely corresponds to a path of size n starting and ending at the root e of Fig. 5.1.

$$\begin{aligned} S &\to \varepsilon \mid aA \mid \bar{c}B \\ A &\to S\bar{a}S \mid SbB \\ B &\to ScS \mid S\bar{b}A \end{aligned} \tag{5.2}$$

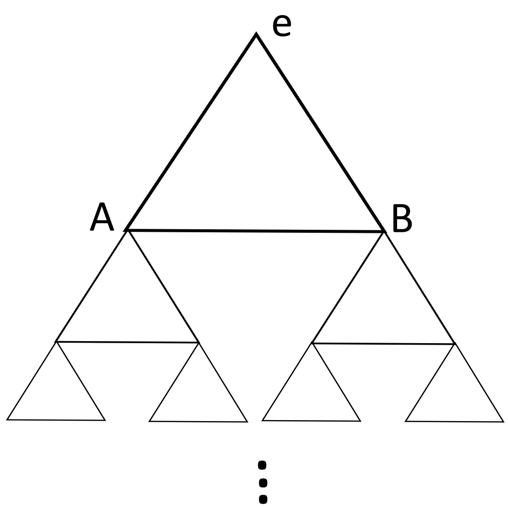


Figure 5.1: The infinite tree of triangles in Example 5.0.2

The characteristic function for this CFG is:

$$\mathbf{F}\left(z, \begin{bmatrix} s \\ a \\ b \end{bmatrix}\right) = \begin{bmatrix} 1 + za + zb \\ zs^2 + zsb \\ zb^2 + zsa \end{bmatrix} \tag{5.3}$$

Now we find the minimal λ, and ψ that solves $F(\frac{1}{\lambda}, \psi) = \psi \geq \mathbf{0}$.

$$\psi = \begin{bmatrix} 3 - \sqrt{2} \\ \frac{3}{\sqrt{2}} - 1 \\ \frac{3}{\sqrt{2}} - 1 \end{bmatrix}, \lambda = 1 + 2\sqrt{2} \tag{5.4}$$

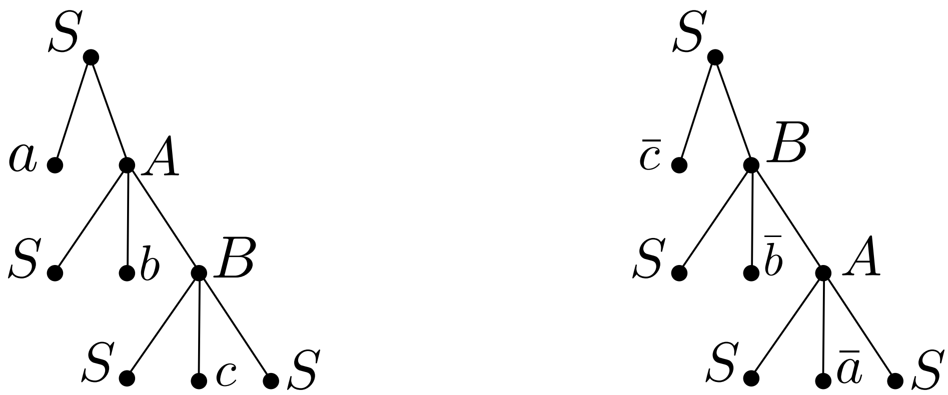
We first compute the density of S in the derivation tree of a randomly selected word of size n. For this, we will compute the mean matrix $\mathbb{M}$ and the mean letter vector:

$$\mathbb{M} = \begin{bmatrix} 0 & \frac{2-3\sqrt{2}}{2-10\sqrt{2}} & \frac{2-3\sqrt{2}}{2-10\sqrt{2}} \\ 3 - \sqrt{2} & 0 & \frac{3-\sqrt{2}}{1+2\sqrt{2}} \\ 3 - \sqrt{2} & \frac{3-\sqrt{2}}{1+2\sqrt{2}} & 0 \end{bmatrix}, \ \mathbf{E} = \begin{bmatrix} \frac{2-3\sqrt{2}}{1-5\sqrt{2}} \\ 1 \\ 1 \end{bmatrix} \tag{5.5}$$

Since the pattern we are interested (shown in Fig. 5.2a) starts with the node S, we use Corollary 3.3.8 to find the density of S. Solving for $\mathbf{C}$ in Corollary 3.3.8 gives,

$$\mathbf{C} = \begin{bmatrix} 1 \\ \frac{8+9\sqrt{2}}{12+38\sqrt{2}} \\ \frac{8+9\sqrt{2}}{12+38\sqrt{2}} \end{bmatrix} \tag{5.6}$$

This gives that the density of S is 1 i.e. $\lim_n \frac{1}{n}\mathbb{E}(\#S \mid \#W = n) = 1$. Now, we can use Corollary 3.3.9 to find the density constant of this pattern. Then the density constant C is just the



(a) Pattern corresponding to going around a triangle in counter-clockwise direction starting from S.

(b) The clockwise triangle pattern that can replace the pattern in Fig. 5.2a

Figure 5.2: Two patterns we're interested in

product of the productions used in this pattern,

$$C = \frac{1}{\lambda}\frac{\psi_A}{\psi_S} \cdot \frac{1}{\lambda}\frac{\psi_S\psi_B}{\psi_A} \cdot \frac{1}{\lambda}\frac{\psi_S^2}{\psi_B} = \frac{\psi_S^2}{\lambda^3} = \frac{(3-\sqrt{2})^2}{(1+2\sqrt{2})^3} \approx 0.0448 \tag{5.7}$$

Now to find a lower bound on the variance density, we use Corollary 3.3.18. This gives us the following lower bound:

$$\liminf_{n\to\infty} \frac{1}{n}\mathrm{Var}(\#T \mid \#W = n) \geq \frac{1}{2}\frac{(3-\sqrt{2})^2}{(1+2\sqrt{2})^3} \approx 0.0224 \tag{5.8}$$

Now we will verify these results with classical methods (those outlined in [3]) of analysing these patterns. By solving the combinatorial generating functions in Maple, we verified $\lim_n \frac{1}{n}\mathbb{E}(\#S \mid \#W = n) = 1$.

It's not obvious how one should solve for the generating function of this specific tree pattern using classical methods. So I will not verify the density constant for this example. However, a very similar statistic that can be verified using classical methods is the net number of times the walk goes around a triangle in counter-clockwise direction i.e. the statistic $\#T - \#T'$, where T and T' respectively refer to Fig. 5.2a and Fig. 5.2b. This statistic is looked at in [2]. Clearly the mean in this case is 0. However the variance lower bound in Corollary 3.3.18 can be verified using classical methods. According to our methods,

$$\liminf_{n\to\infty} \frac{1}{n}\mathrm{Var}(\#T - \#T' \mid \#W = n) \geq 4\frac{1}{2}\frac{(3-\sqrt{2})^2}{(1+2\sqrt{2})^3} \approx 0.0896 \tag{5.9}$$

And classical methods find the variance to be 0.108.

Example 5.0.3. (Tree of k, l-gons) This example is a generalization of the language given in Example 5.0.2. Here, instead of having an infinite tree of n-gons, we have k-gons and l-gons alternating in each generation as depicted in Fig. 5.3. This is another example studied in Example 5.0.2. In this case, we want to prove density result for the number of times a random walk starting from the root conditioned on ending back at the root in n steps goes around a

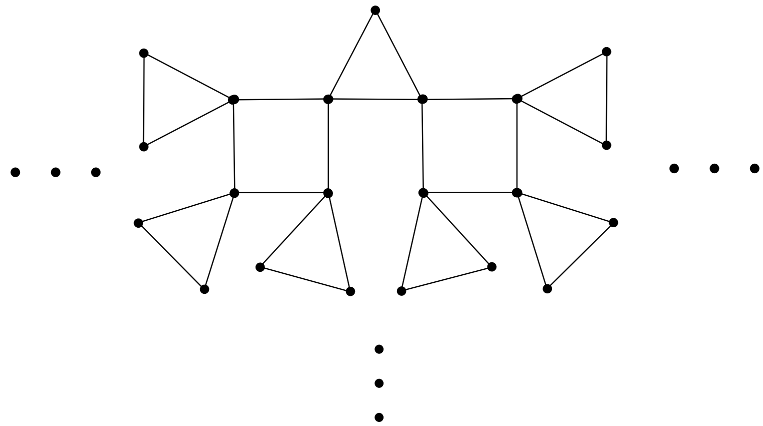


Figure 5.3: The infinite tree of $3, 4$-gons.

triangle or square loop in clockwise direction.

If we can represent this walk as an irreducible unambiguous CFG so that the loops correspond the replaceable patterns, then the density result for the expectation and variance are imminent. The following is the general language for this example where the root is a vertex of the k-gon. We use variables $S_1, \ldots, S_k$ to correspond to the vertices of the k-gon and use $S'_1, \ldots, S'_l$ to correspond to the vertices of the l-gon. We are to start each work from S_1, i.e. $S = S_1$. We will also use $z_1, \bar{z}_1 \ldots, z_k, \bar{z}_k$ and $z'_1, \bar{z}'_1, \ldots, z'_l, \bar{z}'_l$ as our terminal letters.

$$\begin{aligned} S_1 &\longrightarrow \varepsilon \mid z_1 S_2 \mid \bar{z}_k S_k \\ S_i &\longrightarrow S'_1 z_i S_{(i+1) \bmod k} \mid S'_1 \bar{z}_i S_{(i-1) \bmod k} \text{ for } 2 \le i \le k \\ S'_1 &\longrightarrow \varepsilon \mid z'_1 S'_2 \mid \bar{z}'_l S'_l \\ S'_i &\longrightarrow S_1 z'_i S'_{(i+1) \bmod l} \mid S_1 \bar{z}'_i S_{(i-1) \bmod l} \text{ for } 2 \le i \le l \end{aligned} \tag{5.10}$$

Now it's obvious that circling around triangles or squares both give us replaceable patterns (not with each other). Thus there is positive density of their expected number of occurrences and the variance. We can also extend this to the net number of times (clockwise - counter clockwise) we circle. In which case the mean will be 0 but the density result for variance stands using Corollary 3.3.20

Bibliography

[1] David Aldous. Asymptotic fringe distributions for general families of random trees. *The Annals of Applied Probability*, pages 228–266, 1991.

[2] Murray Elder and Andrew Rechnitzer. On groups with D-finite cogrowth series. In preparation, 2018.

[3] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. cambridge University press, 2009.

[4] Tosio Kato. *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media, 2013.

[5] Benedikt Stufler. Rerooting multi-type branching trees: the infinite spine case. *Journal of Theoretical Probability*, pages 1–32, 2021.