UNIVERSIDADE FEDERAL DE PERNAMBUCO

Tese de Doutorado em Matemática Computacional

Substitution Operators

Andréa Vanessa Rocha

Orientador: Andrei Toom

Março, 2009

Rocha, Andréa Vanessa
Substitution operators / Andréa Vanessa Rocha - Recife : O Autor, 2009.
iv, 85 folhas

Tese (doutorado) – Universidade Federal de Pernambuco. CCEN. Matemática Computacional, 2009.

Inclui bibliografia.

1. Probabilidade. 2. Processos estocásticos. Título.

519.2 CDD (22. ed.) MEI2009-049

UFPE 60ANOS

UNIVERSIDADE FEDERAL DE PERNAMBUCO
Centro de Ciências Exatas e da Natureza
Doutorado em Matemática Computacional

PARECER DA BANCA EXAMINADORA DE DEFESA DE TESE DE DOUTORADO

ANDRÉA VANESSA ROCHA

"SUBSTITUTION OPERATORS"

A Banca Examinadora composta pelos Professores ANDRÉ TOOM(Presidente), KLAUS LEITE PINTO VASCONCELOS e MANOEL JOSÉ MACHADO SOARES LEMOS, todos da Universidade Federal de Pernambuco; GAUSS MOUTINHO CORDEIRO, da Universidade Federal Rural de Pernambuco; e ANATOLI IAMBARTSEV, da Universidade de São Paulo, considera a candidata:

(X) APROVADA ( ) REPROVADA

Secretaria do Programa de Doutorado em Matemática Computacional do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, aos 24 dias do mês de março de 2009.

A Toom

PROF. ANDRÉ TOOM
PRESIDENTE

Klaus Leite Pinto Vasconcellos

PROF. KLAUS LEITE PINTO VASCONCELOS
1º EXAMINADOR

Manoel Lemos

PROF. MANOEL JOSÉ MACHADO SOARES LEMOS
2º EXAMINADOR

Gauss M Cordeiro

PROF. GAUSS MOUTINHO CORDEIRO
3º EXAMINADOR

PROF. ANATOLI IAMBARTSEV
4º EXAMINADOR

Av. Prof. Luis Barros Freire, s/nº, Cidade Universitária - CEP: 50670-901 - Recife - PE
FONE/FAX: (81) 2126-8412 - E-mail: ppgmc@ufpe.br - Home Page: www.ppgmc.ufpe.br

Acknowledgments

First of all, I would like to thank the Heavenly Father for giving me the privilege of being a part of His family, for all the blessings that are given to me, as well as to His son Jesus Christ by his unconditional love, which gave me the ability to develop this work.

To my family, which is my foundation: my beloved parents Albertino (in memoriam) and Leonice; my brothers Marco Aurélio and Marcos Vinícius; and especially to my great love Alexandre for his continuous support, and his optimism throughout the development of this thesis.

To Professor Andrei Toom who has been not only an advisor but also a friend. I also thank him for all the seriousness and confidence during the development of this work.

To Klaus Vasconcellos and Sóstenes Lins for their dedication to teaching. The courses they taught me during this period, together with the courses given by prof. Toom, helped me improve my knowledge in mathematics.

To the staff of CCEN - UFPE, in particular to Valéria Bittencourt for being an excellent secretary.

To professors Anatoli Iambartsev, Gauss Cordeiro, Manoel Lemos and Klaus Vasconcellos for their participation in my committee, and all the sugestions made to improve the overall quality of this thesis.

To my doctoral colleagues: Ademakson, Alex, Calitéia, Glória, Lenira and Liliam. They made my doctoral journey much more enjoyable.

To CAPES for the financial support.

Abstract

We study a new kind of random processes in discrete time, which we call *substitution processes*. Since time is discrete, we can define a process by presenting an operator transforming one probability measure on a certain space into another. It is noteworthy that the bulk of studies of locally interacting particle processes is based on the assumption that the set of sites, also called the space, does not change in the process of interaction. There are only a few works where the sites themselves may appear and disappear in the course of the process, and this work is one of them. Consider a non-empty finite set $\mathcal{A}$ called *alphabet*, whose elements are called *letters*. Finite sequences of letters are called *words* and *length* of a word is the number of letters in it. Bi-infinite sequences of letters, that is, elements of $\mathcal{A}^{\mathbb{Z}}$, are called *configurations*. $\mathcal{M}$ denotes the set of translation-invariant probability measures on $\mathcal{A}^{\mathbb{Z}}$. We can informally define a generic substitution operator as the operator from $\mathcal{M}$ to $\mathcal{M}$, which substitutes each occurrence of a word G (G needs to satisfy some condition) in a configuration by another word H with probability ρ, where $\rho \in [0,1]$, independently from each other. Our main contribution is a definition and study of substitution operators in general. The most interesting case of the substitution operator, in our opinion, is the case when G and H have different lengths. The very definition of a substitution operator in this situation is non-trivial and needs an involved preparation. In fact, we had to develop a certain theory of approximation of measures by sequences of words to deal with this case. One difficulty with this case is that the substitution operators are not linear. However, we prove that all of them are *fine*, which in our context seems to be the next best thing after linearity. We also prove that all substitution operators are continuous and use this to make conclusions about existence of their invariant measures, which help us to study their ergodicity. These properties give us hope to relate our work with various needs of applied sciences.

Keywords: Substitution operators; Substitution Processes; Markov Processes.

Resumo

Nós estudamos um novo tipo de processo estocástico a tempo discreto, que nós chamamos de *processos de substituição*. Como o tempo é discreto, nós podemos definir este processo através de um operador que transforma uma medida de probabilidade (em um certo espaço) em outra. Vale a pena notar que a maioria dos estudos de processos de partículas interagentes se baseia na suposição de que o conjunto de sítios, também chamado de espaço, não muda ao longo da interação. Existem apenas poucos trabalhos onde os sítios podem aparecer e desaparecer durante a realização do processo, e este trabalho é um deles. Considere então um conjunto finito não-vazio $\mathcal{A}$ chamado de *alfabeto*, cujos elementos são chamados de *letras*. Sequências finitas de letras são chamadas de *palavras* e o *comprimento* de uma palavra é o número de letras que existe nela. Os elementos de $\mathcal{A}^{\mathbb{Z}}$ (sequências bi-infinitas de letras) são chamados de *configurações*. Denotamos por $\mathcal{M}$ o conjunto de medidas de probabilidade invariantes por translação. Nós podemos definir um operador de substituição genérico como um operador de $\mathcal{M}$ em $\mathcal{M}$ que substitui cada ocorrência de uma palavra G (onde G precisa satisfazer uma certa condição) em uma configuração por outra palavra H com probabilidade ρ, onde $\rho \in [0,1]$, independentemente das outras ocorrências. A nossa maior contribuição é dada pela definição e estudo dos operadores de substituição em geral. O caso mais interessante do operador de substituição, na nossa opinião, é o caso em que G e H têm comprimentos diferentes; a própria definição do operador neste caso é não-trivial. De fato, foi preciso desenvolver uma teoria de aproximação de medidas por sequências de palavras para lidar com este caso. Uma dificuldade com este caso é que os operadores de substituição não são lineares. No entanto, foi possível provar que todos eles são *finos*, que nos parece ser a melhor propriedade depois da linearidade. Nós também provamos que todos os operadores de substituição são contínuos e nós utilizamos este fato para obter conclusões a respeito da existência de medidas invariantes por estes operadores, o que nos ajuda a estudar ergodicidade do processo de substituição. Por fim, nós esperamos que estas propriedades possam ser relacionadas com certas necessidades de ciências aplicadas.

Palavras-Chave: Operadores de Substituição; Processos de Substituição; Processos Markovianos.

Contents

Chapter 1

Introduction

The bulk of modern studies of locally interacting particle processes is based on the assumption that the set of sites, also called the space, does not change in the process of interaction. Elements of this space, also called components, may be in different states, e.g. 0 and 1, often interpreted as absence vs. presence of a particle, and may go from one state to another, which may be interpreted as birth or death of a particle, but the sites themselves do not appear or disappear in the process of functioning. Operators and processes which do not create or eliminate sites will be called *constant-length* ones.

However, in the last dozen of years some attention started to be paid to two new similar kinds of processes with one-dimensional local interaction, where sites may appear and disappear in the course of functioning of a process. In all of them we choose a non-empty finite set $\mathcal{A}$ called *alphabet*, the elements of which are called *letters*.

One kind of these processes are called *random grammars* [2, 3, 5, 6, 7, 8]. All of them have a finite space, whose elements are finite sequences of letters, and are special cases of Markov chains with continuous time and countable sets of states.

Another kind of processes, called *variable-length processes*, [9, 10, 11, 13, 14, 15], are closer to our approach. They have infinite space $\mathbb{Z}$ and discrete time and have a continual set of states $\mathcal{A}^{\mathbb{Z}}$. Until now, their study was reduced to several special cases and it is our contribution to present a general framework, which includes all these cases.

The main purpose of this thesis is to define rigorously and study a large class of variable-length processes, called *substitution processes.* Since all the processes considered here have discrete time, they can be defined in terms of *substitution operators* acting on certain probability measures, and these operators are our main subject of study. Let us give an insight into how a substitution operator works. The set of states of our process is $\mathcal{A}^{\mathbb{Z}}$, the set of configurations. We denote by $\mathcal{M}$ the set of translation-invariant probability measures on $\mathcal{A}^{\mathbb{Z}}$. A generic substitution operator acts from $\mathcal{M}$ to $\mathcal{M}$ roughly as follows. Let us call a *word* a finite sequence of letters. *Length* of a word is the number of letters in it. There is the empty word, whose length is zero. Given two words G and H (where G must satisfy a certain condition of self-avoiding, which will be presented below) and a real number $\rho \in [0, 1]$, a *substitution operator*, informally speaking, substitutes every occurrence of the word G in any configuration by the word H with a probability ρ (and leaves this occurrence unchanged with a probability $1-\rho$). Our definition can be used only if all the occurrences of G in any configuration do not overlap, and this is why we need a special assumption about G.

It is natural to ask which studies in concrete sciences may be relevant to those cited above and our studies. Some answers to this question have been presented in the quoted papers. In particular, some connections with statistical and quantum physics have been pointed at.

Before going into detail in our approach let us mention a few more of those many areas of science where a study of such transformations under which units may appear and disappear may be relevant.

One is error correction in information transmission. When a long sequence of symbols is transmitted at a long distance, all kinds of local distortions may occur including those which change the length of the sequence and it may be useful to approximate them with certain classes of abstract transformations. See for example: *In mathematics, computer science, telecommunication, and information theory, error detection and correction has great practical importance in maintaining data (information) integrity across noisy channels and less-than-reliable storage media.* [16]

Another area is molecular biology. It is well-known that DNA and RNA are polymers, that is, chains of units called nucleotides. These polymers can produce other polymers along themselves many times. It is important for us that the relation between the units of the old and new polymers is not exactly one-to-one. The acts of violation of the 1-1 principle are called *insertions* and *deletions* and all of them together are called *indels.* It seems appropriate for us to idealize these processes as iterations of transformation operators. See for example: *The polymerase chain reaction is an extremely versatile technique for copying DNA. In brief, PCR allows a single DNA sequence to be copied (millions of times), or altered in predetermined ways.* [18]

Still another area where our theoretical considerations may be relevant is historical linguistics. It is well-known that evolution of natural languages is not a chaotic mess of individual changes, but a sequence of systematic changes. Lately the well-known linguist Andrei Zaliznyak put it this way: *... the external forms of words of a language do not change individually for every word; they change according to the functioning of so called phonetical changes: at any period of time any language is subject to certain phonetical transitions, which act on all those words, which include certain phonemes* *or certain combinations of phonemes, without exception ... Even the most astonishing transformation of a shape of a word, which may occur in the course of history is not a result of a random singular substitution of sounds, but a display of systematic phonetical changes, which occured in a certain language at a certain time.*[19]

We hope that our study may serve as a first step towards mathematical theories of real processes with variable sets of units. Also we must note that all our constructions would be much simpler if we dealt only with finite sequences. Our main technical difficulties come from our desire to deal with infinite sequences. This choice is supported by the fact that our interest in processes with infinite space is shared by most researchers. (Of course, interest in finite spaces remains valid.) In fact, mathematicians have deep reasons to study infinite systems, but we leave explanation of this fact to philosophers of science.

Chapter 2

Basic Definitions

Throughout this text $\mathcal{A}$ is a non-empty finite set called *alphabet*, which we assume to be fixed. Its elements are called *letters* and finite sequences of letters are called *words*. The number of letters in a word W is called its *length* and denoted by $|W|$. Any letter may be considered as a word of length one. There is the empty word, denoted by Λ, whose length is zero. The set of words in a given alphabet $\mathcal{A}$ is called *dictionary* and denoted by $\mathrm{dic}(\mathcal{A})$.

We assume that our alphabet contains no brackets or commas. Given any finite sequence of words $W_1, \ldots, W_k$ (perhaps separated by commas or put in brackets), we denote by $\mathrm{concat}(W_1, \ldots, W_k)$ and call their *concatenation* the word obtained by writing all these words one after another in that order in which they are listed, all brackets and commas eliminated. In particular, W^n means concatenation of n words, everyone of which is a copy of W. If $n = 0$, the word W^n is empty, $W^0 = \Lambda$.

We denote by $\mathbb{Z}$ the set of integer numbers and $\mathcal{A}^{\mathbb{Z}}$ the set of bi-infinite sequences whose terms are elements of $\mathcal{A}$. We consider probability measures on the σ-algebra $\mathbb{A}^{\mathbb{Z}}$ on the product space $\mathcal{A}^{\mathbb{Z}}$ generated by the product of discrete topologies on all copies of $\mathcal{A}$. We note that since $\mathcal{A}$ is finite, it is compact in the discrete topology, and by Tychonoff's compact theorem, $\mathcal{A}^{\mathbb{Z}}$ is also compact. As usual, shifts on $\mathbb{Z}$ generate shifts on $\mathcal{A}^{\mathbb{Z}}$ and shifts on $\mathbb{A}^{\mathbb{Z}}$. We call a measure μ on $\mathcal{A}^{\mathbb{Z}}$ *uniform* if it is invariant under all shifts. In this case for any word $W = (a_1, \ldots, a_n)$, where $a_1, \ldots, a_n$ is the sequence of letters that forms the word W, the right side of

$$\mu(W) = \mu(a_1, \ldots, a_n) = \mu(s_{i+1} = a_1, \ldots, s_{i+n} = a_n)$$

is one and the same for all $i \in \mathbb{Z}$, whence we may use the left side as a shorter denotation.

We denote by $\mathcal{M}$ the set of uniform probability measures on $\mathcal{A}^{\mathbb{Z}}$. Since $\mathcal{A}^{\mathbb{Z}}$ is compact, $\mathcal{M}$ is also compact. Any uniform measure is determined by its values on all words and it is a probability measure if its value on the empty word equals 1.

In order for values $\mu(W)$ to form a uniform probability measure, it is necessary and sufficient that: all the numbers $\mu(W)$ must be non-negative, μ on the empty word must equal one and for any letter a and any word W (including the empty one) we must have

$$\mu(W) = \sum_{a \in \mathcal{A}} \mu(W, a) = \sum_{a \in \mathcal{A}} \mu(a, W),$$

where (W, a) and (a, W) are concatenations of the word W and the letter a in the two possible orders.

Given two words $W = (a_1, \ldots, a_m)$ and $V = (b_1, \ldots, b_n)$, where $|W| \leq |V|$, we call the integer numbers in the interval $[0,\ n - m]$ *positions* of W in V. We say that W *enters* V at a position k if

$$\forall\ i \in \mathbb{Z} : 1 \leq i \leq m \Rightarrow a_i = b_{i+k}.$$

We call a word W *self-overlapping* if there is a word V such that $|V| < 2 \cdot |W|$ and W enters V at two different positions. A word is called *self-avoiding* if it is not self-overlapping. In particular, the empty word, every word consisting of one letter and every word consisting of two different letters are self-avoiding.

It is known that self-avoiding words are not very rare: in fact for any alphabet with at least two letters the number of self-avoiding words of length n divided by the number of all words of length n tends to a positive limit when $n \to \infty$ and this limit tends to one when the number of letters in the alphabet tends to infinity. [1]

We denote by $\#(W \text{ in } V)$ the number of positions at which W enters V. If W is the empty word, it enters any word V at $|V|+1$ positions. If $|W| \le |V|$, we call the *frequency* of a word W in a word V and denote by $\text{freq}(W \text{ in } V)$ the number of positions at which W enters V divided by the total number of positions of W in V, that is, the fraction

$$\text{freq}(W \text{ in } V) = \frac{\#(W \text{ in } V)}{|V|-|W|+1}. \tag{2.1}$$

Notice that the frequency of the empty word in any word is 1. If $|W| > |V|$, the set of positions of W in V is empty and the frequency of W in V is zero by definition.

We call a *pseudo-measure* any map $\mu : \text{dic}(\mathcal{A}) \to \mathbb{R}$. In particular, any measure $\mu \in \mathcal{M}$ is a pseudo-measure.

Definition 2.0.1. For every word $V \in \text{dic}(\mathcal{A})$ we define the corresponding pseudo-measure, denoted by V' and defined by the rule $V'(W) = \text{freq}(W \text{ in } V)$ for all words W.

Since the set $\text{dic}(\mathcal{A})$ is countable, we can enumerate it in some order

$$\text{dic}(\mathcal{A}) = \{W_1,\ W_2,\ W_3, \ldots\}. \tag{2.2}$$

Also we choose a sequence of positive numbers $w_1,\ w_2,\ w_3, \ldots$, whose sum equals 1. Thus, given a pseudo-measure ν, the norm of ν defined by the sequence is as

$$\|\nu\| = \sum_{i=1}^{\infty} w_i \cdot |\nu(W_i)|. \tag{2.3}$$

We denote by $\mathcal{M}_{norm}$ the set of those pseudo-measures whose norm is finite. Thus $\mathcal{M}_{norm}$ is a normed linear space which contains $\mathcal{M}$. Having a norm, we define a metric dist in the space by the usual rule $\operatorname{dist}(\mu,\ \nu) = \|\mu - \nu\|$.

Definition 2.0.2. We say that a sequence of pseudo-measures $\nu_1,\ \nu_2,\ \nu_3, \ldots$ converges to a pseudo-measure $\mu \in \mathcal{M}$ if

$$\|\nu_n - \mu\| \to 0 \text{ as } n \to \infty.$$

Our main definition is the following:

Definition 2.0.3. We say that a sequence (V_n) of words $V_1,\ V_2,\ V_3, \ldots \in \operatorname{dic}(\mathcal{A})$ converges to a measure $\mu \in \mathcal{M}$ if any of the following equivalent conditions holds:

a) for every word $W \in \operatorname{dic}(\mathcal{A})$ the frequency of W in V_n tends to $\mu(W)$ as $n \to \infty$.

b) the sequence (V_n') converges to μ.

Remark 2.0.4. Notice that since the frequencies of all W in a given V are zeros for all W longer than V, the convergence in Definition 2.0.3 is possible only if the length of V_n tends to ∞ as $n \to \infty$.

Lemma 2.0.5. *The equivalence in Definition 2.0.3 holds. That is, let $V_1,\ V_2,\ V_3, \ldots$ be a sequence of words, and $V_1',\ V_2',\ V_3', \ldots$ be the sequence of their pseudo-measures. Then (V_n) converges to μ if and only if (V_n') converges to μ.*

Proof: *First we prove in one direction.* Suppose that (V_n) converges to μ and let $\varepsilon > 0$ be any real number. Then choose J so large that

$$\sum_{i=J+1}^{\infty} w_i < \frac{\varepsilon}{4}.$$

Further, let us consider the enumeration of the words that defines the norm, that is, consider an enumeration of $\mathrm{dic}(\mathcal{A})$, for instance $\mathrm{dic}(\mathrm{A}) = \{W_1,\ W_2,\ W_3, \ldots\}$ such that

$$\|\mu\| = \sum_{i=1}^{\infty} w_i \cdot |\mu(W_i)|.$$

Let $i \in \{1, \ldots, J\}$. Then we choose n_i so large that

$$|\mathrm{freq}(W_i \ \text{ in } \ V_{n_i}) - \mu(W_i)| < \frac{\varepsilon}{2^{i+1} \cdot w_i}$$

and let $N = \max\{n_1, \ldots, n_J\}$. Then, since $|V'_N(W)| \leq 1$ and $|\mu(W)| \leq 1$ for any word W, we have

$$\begin{aligned}
\|V'_N - \mu\| &= \sum_{i=1}^{\infty} w_i \cdot |V'_N(W_i) - \mu(W_i)| \\
&= \sum_{i=1}^{J} w_i \cdot |V'_N(W_i) - \mu(W_i)| + \sum_{i=J+1}^{\infty} w_i \cdot |V'_N(W_i) - \mu(W_i)| \\
&\leq \sum_{i=1}^{J} w_i \cdot |V'_N(W_i) - \mu(W_i)| + 2 \cdot \sum_{i=J+1}^{\infty} w_i \\
&\leq \sum_{i=1}^{J} \frac{\varepsilon}{2^{i+1}} + \frac{\varepsilon}{2} \leq \varepsilon,
\end{aligned}$$

which concludes the proof.

Proof in the other direction. Let $\|V'_n - \mu\| \to 0$. Then, for any $W \in \mathrm{dic}(\mathcal{A})$ there exists $j \in \{1,\ 2,\ 3, \ldots\}$ such that $W = W_j$. Therefore

$$|\text{freq}(W \ \text{ in } \ V_n) - \mu(W)| = |\text{freq}(W_j \text{ in } V_n) - \mu(W_j)| \leq \frac{1}{w_j} \|V'_n - \mu\|,$$

which tends to zero as $n \to \infty$.

Lemma 2.0.5 is proved in both directions.

The following definition is more technical:

Definition 2.0.6. Given a real number $\varepsilon > 0$ and a natural number r, a word V is said to $(\varepsilon,\ r)$-*approximate* a measure $\mu \in \mathcal{M}$ if for every word $W \in \text{dic}(\mathcal{A})$,

$$|W| \leq r \Rightarrow |\text{freq}(W \ \text{ in } \ V) - \mu(W)| \leq \varepsilon.$$

Lemma 2.0.7. *A sequence (V_n) of words converges to a measure μ if and only if for any positive $\varepsilon > 0$ and any natural r there is n_0 such that for every $n \geq n_0$ the word V_n $(\varepsilon,\ r)$-approximates μ.*

Proof in one direction: Suppose that (V_n) converges to μ. We want to prove that

$$\left.\begin{array}{c} \forall\ \varepsilon > 0,\ \forall\ r \in N\ \exists\ n_0,\ \forall\ n \geq n_0,\ \forall\ W \in \text{dic}(\mathcal{A}): \\ |W| \leq r \Rightarrow |\text{freq}(W \ \text{ in } \ V_n) - \mu(W)| \leq \varepsilon. \end{array}\right\} \tag{2.4}$$

Let us choose W such that $0 < |W| \leq r$. Since (V_n) converges to μ,

$$\lim_{n \to \infty} \text{freq}(W \ \text{ in } \ V_n) = \mu(W),$$

that is

$$\forall\ \varepsilon' > 0\ \ \exists\ n_W\ \ \forall\ n \geq n_W : |\text{freq}(W \ \text{ in } \ V_n) - \mu(W)| \leq \varepsilon'.$$

Taking $\varepsilon' = \varepsilon$ and n_0 equal to the maximum of n_W over all these non-empty W, whose length does not exceed r, we obtain (2.4).

In the other direction the proof is straightforward.

Lemma 2.0.7 is proved.

We define a *random word* X on an alphabet $\mathcal{A}$ as a probability distribution on $\mathrm{dic}(\mathcal{A})$ which is concentrated on a finite subset of $\mathrm{dic}(\mathcal{A})$. A random word is determined by its non-negative components X_V for all V whose sum is 1 and the set $\{V : X_V > 0\}$ is finite, where X_V denotes the probability that $X = V$. We denote by Ω the set of random words on the alphabet $\mathcal{A}$.

Definition 2.0.8. We define the mean length of any random word X as

$$E|X| = \sum_{V \in \mathrm{dic}(\mathcal{A})} X_V \cdot |V|.$$

Definition 2.0.9. We define the expected number of entrances of a word W in a random word X as

$$E\#(W \ \text{ in } \ X) = \sum_{V \in \mathrm{dic}(\mathcal{A})} X_V \cdot \#(W \ \text{ in } \ V).$$

Definition 2.0.10. We define the mean frequency of a word W in a random word X as

$$\mathrm{freq}(W \ \text{ in } \ X) = \frac{E\#(W \ \text{ in } \ X)}{E|X| - |W| + 1}. \tag{2.5}$$

For any random word X we define the corresponding pseudo-measure X' by the rule $X'(W) = \mathrm{freq}(W \ \text{ in } \ X)$ for all W.

Definition 2.0.11. We say that a sequence (X_n) of random words $X_1,\ X_2,\ X_3, \ldots \in \Omega$ converges to a measure $\mu \in \mathcal{M}$ if any of the following equivalent conditions holds:

a) for every word $W \in \mathrm{dic}(\mathcal{A})$ the mean frequency of W in X_n tends to $\mu(W)$ as $n \to \infty$

b) (X_n') tends to μ as $n \to \infty$.

Definition 2.0.12. Given a positive number $\varepsilon > 0$ and a natural number r, we say that a random word X $(\varepsilon,\ r)$*-approximates* a measure $\mu \in \mathcal{M}$ if for every non-empty word $W \in \text{dic}(\mathcal{A})$,

$$|W| \leq r \Rightarrow |\text{freq}(W \text{ in } X) - \mu(W)| \leq \varepsilon.$$

Definition 2.0.13. Let ν be any pseudo-measure. Then let $\nu|_{|W|\leq r}$ be the pseudo-measure obtained from ν by truncation to the words whose length is not greater than r. Thus, for every word W

Aqui ele diz que a notação é confusa e sugere $\nu \mathrm{I}_{|W|\leq r}(W)$.

$$\nu|_{|W|\leq r}(W) = \begin{cases} \nu(W), & \text{if } |W| \leq r, \\ 0, & \text{if } |W| > r. \end{cases}$$

Then we say that a pseudo-measure ν $(\varepsilon,\ r)$ *-approximates* a measure $\mu \in \mathcal{M}$ if

$$\|\nu|_{|W|\leq r} - \mu|_{|W|\leq r}\| < \varepsilon.$$

Lemma 2.0.14. *If a random word X $(\varepsilon,\ r)$-approximates a measure $\mu \in \mathcal{M}$ in the sense of Definition 2.0.12, then the pseudo-measure X' associated with X also $(\varepsilon,\ r)$-approximates the same μ in the sense of Definition 2.0.13. Conversely, if a pseudo-measure X' associated with the random word X $(\varepsilon,\ r)$-approximates a measure $\mu \in \mathcal{M}$, then X $(\varepsilon',\ r)$ -approximates the same μ, where*

$$\varepsilon' = \frac{\varepsilon}{w} \quad \textit{and} \quad w = \min\{w_i; |W_i| \leq r\}.$$

Notice that w defined as that minimum is well-defined since the set of words such that $|W_i| \leq r$ is finite.

Proof in one direction: If X $(\varepsilon,\ r)$-*approximates* a measure $\mu \in \mathcal{M}$, then

$$\begin{aligned} \|X'|_{|W|\le r} - \mu|_{|W|\le r}\| &= \sum_{i=1}^{\infty} w_i |X'|_{|W|\le r}(W_i) - \mu|_{|W|\le r}(W_i)| \\ &= \sum_{i:|W_i|\le r} w_i |X'|_{|W|\le r}(W_i) - \mu|_{|W|\le r}(W_i)| \qquad (2.6) \\ &\le \varepsilon \sum_{i:|W_i|\le r} w_i \\ &\le \varepsilon \cdot \sum_{i=1}^{\infty} w_i = \varepsilon, \qquad (2.7) \end{aligned}$$

which concludes this part of the proof.

In the other direction. Let X' $(\varepsilon,\ r)$-*approximate* μ. Then by (2.6) and (2.7) we have

$$\sum_{i:|W_i|\le r} w_i |X'|_{|W|\le r}(W_i) - \mu|_{|W|\le r}(W_i)| \le \varepsilon,$$

which implies trivially that for any word W, such that $|W| \le r$, there exists $j \in \{i; |W_i| \le r\}$ such that $W = W_j$ and

$$w_j |X'|_{|W|\le r}(W_j) - \mu|_{|W|\le r}(W_j)| \le \varepsilon.$$

Thus for any W, such that $|W| \le r$,

$$|X'|_{|W|\le r}(W) - \mu|_{|W|\le r}(W)| \le \frac{\varepsilon}{w},$$

where $w = \min\{w_i\}$, with i such that $|W_i| \le r$.

Lemma 2.0.14 is proved.

Theorem 2.0.15. *For any $\mu \in \mathcal{M}$, any $\varepsilon > 0$ and any natural r there is a random word which $(\varepsilon,\ r)$-approximates μ.*

Proof: If $\mathcal{A}$ contains only one letter, the theorem is trivial. So let $\#(\mathcal{A}) \geq 2$. As usual, for any real x we denote by $[x]$ the greatest integer number which is not greater than x and by $]x[$ the least integer number which is not less than x. Consider the following parameters

$$s = \left]\frac{4r}{\varepsilon}\right[\quad \text{and} \quad N = \#(\mathcal{A})^s.$$

Hence $r/s \leq \varepsilon/4$. There are N words in $\mathrm{dic}(\mathcal{A})$, whose lengths equal s and we denote them by $U_1, \ldots, U_N$. Further, let us consider the random word X such that

$$\forall i = 1, \ldots, N, P(X = U_i) = \mu(U_i).$$

Since $\sum_i \mu(U_i) = 1$, our random word is well defined. Furthermore, for any position k of W in U_i, that is, for any $k \in [1, s - |W| + 1]$, we define

$$\#(W \ \text{ in } \ U_i)_k = \begin{cases} 1 & \text{if } W \text{ enters } U_i \text{ at the position } k, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\sum_{k=1}^{s-|W|+1} \#(W \ \text{ in } \ U_i)_k = \#(W \ \text{ in } \ U_i) \tag{2.8}$$

and

$$\sum_{i=1}^{N} (\#(W \ \text{ in } \ U_i)_k \cdot \mu(U_i)) = \mu(W),$$

for all k fixed. Summing this over k yields

$$\sum_{k=1}^{s-|W|+1} \sum_{i=1}^{N} (\#(W \ \text{ in } \ U_i)_k \cdot \mu(U_i)) = (s - |W| + 1) \cdot \mu(W),$$

whence

$$\sum_{i=1}^{N} \left(\mu(U_i) \cdot \sum_{k=1}^{s-|W|+1} \#(W \ \text{ in } \ U_i)_k \right) = (s - |W| + 1) \cdot \mu(W). \tag{2.9}$$

Replacing (2.8) by (2.9) gives

$$\sum_{i=1}^{N} (\#(W \ \text{ in } \ U_i) \cdot \mu(U_i)) = (s - |W| + 1) \cdot \mu(W). \tag{2.10}$$

Let us estimate the mean frequency of W in X. First we get the lower estimation. In the present case the numerator of that fraction is

$$E\#(W \ \text{ in } \ X) = \sum_{i=1}^{N} \#(W \ \text{ in } \ U_i)\mu(U_i)$$

and an estimation for the denominator of the mean frequency is

$$E|X| - |W| + 1 \leq E|X| = s.$$

Hence

$$\begin{aligned} \text{freq}(W \ \text{ in } \ X) &\geq \frac{1}{s}\sum_{i=1}^{N} \#(W \ \text{ in } \ U_i)\mu(U_i) \\ &= \frac{s - |W| + 1}{s}\mu(W) \\ &\geq \left(1 - \frac{\varepsilon}{4}\right)\mu(W) \\ &\geq \mu(W) - \varepsilon. \end{aligned}$$

Now we want to estimate the mean frequency from above. Notice that the denominator of the mean frequency of W in X is not less than

$$E|X| - |W| + 1 \geq s - r.$$

Further,

$$\frac{s - |W| + 1}{s - r} \leq \frac{s}{s - r} \leq \frac{1}{1 - \varepsilon/4} \leq 1 + \varepsilon/2.$$

Thus we have

$$\begin{aligned} \text{freq}(W \ \text{ in } \ X) &\leq \frac{1}{s-r}\sum_{i=1}^{N}\#(W \ \text{ in } \ U_i)\mu(U_i) \\ &= \frac{s-|W|+1}{s-r}\mu(W) \\ &\leq \left(1+\frac{\varepsilon}{2}\right)\mu(W) \\ &\leq \mu(W)+\varepsilon. \end{aligned}$$

Theorem 2.0.15 is proved.

Corollary 2.0.16. *For any $\mu \in \mathcal{M}$ there is a sequence of random words which converges to it.*

Proof: Due to the previous theorem 2.0.15, for every natural n we can take a random word X_n which $(1/n,\ n)$-approximates μ. Notice that if a random word $(\varepsilon,\ r)$-approximates a measure, then it (ε', r')-approximates the same measure for any $\varepsilon' \geq \varepsilon$ and $r' \leq r$. Therefore the sequence (X_n) converges to μ.

Corollary 2.0.16 is proved

Theorem 2.0.17. *For any $\mu \in \mathcal{M}$, any $\varepsilon > 0$ and any natural r there is a word which $(\varepsilon,\ r)$-approximates μ.*

Proof: Like in the proof of Theorem 2.0.15, if $\mathcal{A}$ contains only one letter, the theorem is trivial. So let $\#(\mathcal{A}) \geq 2$. Let us introduce parameters

$$s = \left]\frac{4r}{\varepsilon}\right[, \quad N = \#(\mathcal{A})^s \quad \text{and} \quad Q = \left]\frac{4N}{\varepsilon}\right[,$$

whence

$$\frac{r}{s} \leq \frac{\varepsilon}{4} \quad \text{and} \quad \frac{N}{Q} \leq \frac{\varepsilon}{4}.$$

Like in Theorem 2.0.15, there are N words in $\mathrm{dic}(\mathcal{A})$, whose length equals s, and we denote them by $U_1,\ U_2, \ldots, U_N$. Further, for every $i = 1, \ldots, N$ we denote

$$p_i = [Q \cdot \mu(U_i)]\,,$$

where Q was defined in the beginning of the proof. Hence

$$Q \cdot \mu(U_i) - 1 < p_i \leq Q \cdot \mu(U_i). \tag{2.11}$$

For every i from 1 to N we take p_i copies of U_i and define V as their concatenations in any order, for instance

$$V = \mathrm{concat}(U_1^{p_1}, \ldots, U_N^{p_N}).$$

Let us check that the word V has the desired property, namely $(\varepsilon,\ r)$-approximates the measure μ. Since $\mu(U_1) + \cdots + \mu(U_N) = 1$, summing (2.11) over $i = 1, \ldots, N$ gives

$$Q - N < \sum_{i=1}^{N} p_i \leq Q. \tag{2.12}$$

Let us estimate the frequency of W in V. First we get the lower estimation. In the present case the numerator of that fraction is

$$\#(W \text{ in } V) \geq \sum_{i=1}^{N} \#(W \text{ in } U_i) \cdot p_i$$

and the denominator is

$$|V| - |W| + 1 \leq |V| = s \cdot \sum_{i=1}^{N} p_i \leq s \cdot Q.$$

Hence, using equation (2.10), we obtain

$$\mathrm{freq}(W \text{ in } V) \geq \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} (\#(W \text{ in } U_i) \cdot p_i) \geq$$

$$\frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} (\#(W \ \text{ in } \ U_i) \cdot (Q \cdot \mu(U_i) - 1)) =$$

$$\frac{1}{s} \cdot \sum_{i=1}^{N} (\#(W \ \text{ in } \ U_i) \cdot \mu(U_i)) - \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} \#(W \ \text{ in } \ U_i) \geq$$

$$\frac{s - |W| + 1}{s} \cdot \mu(W) - \frac{s \cdot N}{s \cdot Q} \geq \left(1 - \frac{r}{s}\right) \cdot \mu(W) - \frac{N}{Q} \geq$$

$$\left(1 - \frac{\varepsilon}{4}\right) \cdot \mu(W) - \frac{\varepsilon}{4} \geq \mu(W) - \varepsilon. \tag{2.13}$$

Now let us estimate the frequency of W in V from above. The numerator of the fraction (2.1) is not greater than

$$\sum_{i=1}^{N} (\#(W \ \text{ in } \ U_i) + |W|) \cdot p_i \leq \sum_{i=1}^{N} (\#(W \ \text{ in } \ U_i) + |W|) \cdot Q \cdot \mu(U_i)$$

and the denominator is not less than

$$|V| - |W| + 1 \geq s \cdot \sum_{i=1}^{N} p_i - r \geq s \cdot (Q - N) - r.$$

Since $\#(\mathcal{A}) \geq 2$, $r \geq 1$ and $\varepsilon \leq 1$,

$$\frac{r}{Q \cdot N} \leq \frac{r \cdot \varepsilon}{4 \cdot N^2} \leq \frac{s \cdot \varepsilon^2}{16 \cdot N^2} \leq \frac{s \cdot \varepsilon^2}{16 \cdot 4^s} = \left(\frac{s}{4^s}\right) \cdot \frac{\varepsilon^2}{16} \leq \frac{\varepsilon^2}{16}.$$

Therefore

$$\frac{s \cdot Q}{s \cdot (Q - N) - r} = \frac{1}{1 - \frac{N}{Q} - \frac{r}{Q \cdot N}} \leq \frac{1}{1 - \frac{\varepsilon}{4} - \frac{\varepsilon^2}{16}} \leq 1 + \frac{\varepsilon}{2}.$$

Thus, using (2.11), (2.12) and (2.13),

$$\text{freq}(W \ \text{ in } \ V) \leq \left(1 + \frac{\varepsilon}{2}\right) \cdot \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} ((\#(W \ \text{ in } \ U_i) + |W|) \cdot Q \cdot \mu(U_i)) \leq$$

$$\left(1+\frac{\varepsilon}{2}\right)\cdot\frac{1}{s}\cdot\left(\sum_{i=1}^{N}(\#(W\ \text{ in }\ U_i)\cdot\mu(U_i))+r\cdot\sum_{i=1}^{N}\mu(U_i)\right)\leq$$

$$\left(1+\frac{\varepsilon}{2}\right)\cdot\frac{1}{s}\cdot(s\cdot\mu(W)+r)\leq\left(1+\frac{\varepsilon}{2}\right)\cdot\left(\mu(W)+\frac{r}{s}\right)\leq$$

$$\left(1+\frac{\varepsilon}{2}\right)\cdot\left(\mu(W)+\frac{\varepsilon}{4}\right)\leq\mu(W)+\varepsilon.$$

Theorem 2.0.17 is proved.

Corollary 2.0.18. *For any $\mu\in\mathcal{M}$ there is a sequence of words which converges to it.*

Proof: Due to the previous theorem 2.0.17, for every natural n we can find a word V_n which $(1/n,\ n)$-approximates μ. Note that if a word $(\varepsilon,\ r)$-approximates a measure, then it $(\varepsilon',\ r')$-approximates the same measure for any $\varepsilon'\geq\varepsilon$ and $r'\leq r$. Therefore, the sequence (V_n) converges to μ. *Corollary 2.0.18 is proved*

Definition 2.0.19. We say that a sequence of words (V_n) is *Cauchy* if for any word W and any $\varepsilon>0$ there is k such that

$$\forall\, m,n>k:|\operatorname{freq}(W\ \text{ in }\ V_m)-\operatorname{freq}(W\ \text{ in }\ V_n)|<\varepsilon,$$

note that n and m do not depend on the word W.

Theorem 2.0.20. *A sequence of words is Cauchy if and only if it converges to some measure.*

Proof: *In one direction.* Let us suppose that the sequence (V_n) is Cauchy. Then $\operatorname{freq}(W\ \text{ in }\ V_n)$ is a Cauchy sequence of real numbers for any W. Therefore the following limit exists:

$$\lim_{n\to\infty} \text{freq}(W \text{ in } V_n).$$

So let us define the following map having $\text{dic}(\mathcal{A})$ as its domain:

$$\mu(W) = \lim_{n\to\infty} \text{freq}(W \text{ in } V_n).$$

We want to prove that $\mu(\cdot)$ is indeed a measure. In other words, we want to prove that $\forall$ word $W : 0 \leq \mu(W) \leq 1$ and also that $\sum_a \mu(W,a) = \sum_a \mu(a,W) = \mu(W)$.

It is easy to see that $0 \leq \text{freq}(W \text{ in } V_n) \leq 1$. Therefore

$$0 \leq \lim_{n\to\infty} \text{freq}(W \text{ in } V_n) \leq 1.$$

So we have

$$0 \leq \mu(W) \leq 1.$$

We still have to show that $\sum_a \mu(W,a) = \sum_a \mu(a,W) = \mu(W)$. To do this, let us first notice that $|(W,a)| = |(a,W)| = |W|+1$. First let us deal with the case when a is on the right side, that is, we show that $\sum_a \mu(W,a) = \mu(W)$. If $a \neq b$, then (W,a) must enter V_n in a different position than that of (W,b). Moreover, if W enters V_n in a position which is not the last one, that is, if W does not enter V_n at the position $|V_n| - |W|$, then there must exist a letter, say a, at the right side of W, such that (W,a) still enters V_n at the same position.

Now let us make two observations. First, the number of entrances of W in V_n is always greater or equal than the sum over all the letters a of the numbers of entrances of (W,a) in V_n, that is, $\sum_a \#((W,a) \text{ in } V_n) \leq \#(W \text{ in } V_n)$, because if (W,a) enters V_n, then W also enters it at the same position. Second, if W enters V_n except for the last position, (W,a) enters V_n for some a, as explained before. From these observations:

$$0 \le \#(W \text{ in } V_n) - \sum_a \#((W,a) \text{ in } V_n) \le 1.$$

Now, dividing by $(|V_n| - |W| + 1)$ gives

$$0 \le \text{freq}(W \text{ in } V_n) - \sum_a \frac{\#((W,a) \text{ in } V_n)}{|V_n| - |W| + 1} \le \frac{1}{|V_n| - |W| + 1},$$

which simplifies to

$$0 \le \text{freq}(W \text{ in } V_n) - \sum_a \text{freq}((W,a) \text{ in } V_n)\frac{|V_n| - |W|}{|V_n| - |W| + 1} \le \frac{1}{|V_n| - |W| + 1},$$

Now, $1/(|V_n| - |W| + 1) \to 0$ as $n \to \infty$, because $|V_n| \to \infty$ as $n \to \infty$. For a similar reason,

$$\frac{(|V_n| - |W|)}{(|V_n| - |W| + 1)} \to 1 \text{ as } n \to \infty.$$

Therefore,

$$0 \le \lim_{n\to\infty} \text{freq}(W \text{ in } V_n) - \sum_a \lim_{n\to\infty} \text{freq}((W,a) \text{ in } V_n) \le 0,$$

that is

$$\mu(W) = \sum_a \mu(W,a).$$

The argument for concatenation from the left side is analogous. Thus, the map $\mu(\cdot)$ is indeed a measure.

In the other direction. From the definition of convergence of words to a measure we have that for each W:

$$\lim_{n\to\infty} \text{freq}(W \text{ in } V_n) = \mu(W).$$

So, let N be such that for every $m,\ n > N$, we have

$$|\,\text{freq}(W \text{ in } V_m) - \mu(W)| < \frac{\varepsilon}{2} \text{ and } |\,\text{freq}(W \text{ in } V_n) - \mu(W)| < \frac{\varepsilon}{2}.$$

Now,

$$|\,\text{freq}(W \;\; \text{in} \;\; V_m) - \text{freq}(W \;\; \text{in} \;\; V_n)|$$

$$\leq |\,\text{freq}(W \;\; \text{in} \;\; V_m) - \mu(W)| + |\,\text{freq}(W \;\; \text{in} \;\; V_n) - \mu(W)| < \varepsilon.$$

Therefore, if (V_n) tends to a measure, then it is Cauchy.

Theorem 2.0.20 is proved in both directions.

Definition 2.0.21. We say that a sequence of random words (X_n) is *Cauchy* if for any word W and any $\varepsilon > 0$ there is k such that

$$\forall\; m,\; n > k : |\,\text{freq}(W \;\; \text{in} \;\; X_m) - \text{freq}(W \;\; \text{in} \;\; X_n)| < \varepsilon.$$

Theorem 2.0.22. *A sequence of random words is Cauchy if and only if it converges to some measure.*

Proof in one direction. Let us choose a word W and suppose that the sequence (X_n) is Cauchy, thus $\text{freq}(W \;\; \text{in} \;\; X_n)$ is a Cauchy sequence of real numbers. Therefore the limit $\lim_{n\to\infty} \text{freq}(W \;\; \text{in} \;\; X_n)$ exists. So let us define the following map having the set of all words as its domain:

$$\mu(W) = \lim_{n\to\infty} \text{freq}(W \;\; \text{in} \;\; X_n).$$

We want to prove that μ is indeed a measure. In other words, we want to prove that $\forall$ 'word $W : 0 \leq \mu(W) \leq 1$ and also that $\sum_a \mu(W, a) = \sum_a \mu(a, W) = \mu(W)$.

It is easy to see that $0 \leq \text{freq}(W \;\; \text{in} \;\; X_n) \leq 1$. Then

$$0 \leq \lim_{n\to\infty} \text{freq}(W \;\; \text{in} \;\; X_n) \leq 1.$$

Therefore

$$0 \le \mu(W) \le 1.$$

We still have to show that $\sum_a \mu(W,a) = \sum_a \mu(a,W) = \mu(W)$. To do so, note initially that $|(W,a)| = |(a,W)| = |W| + 1$.

Let us take first the case when a is on the right side, that is, we show that $\sum_a \mu(W,a) = \mu(W)$. Let V be any word on $\mathrm{dic}(\mathcal{A})$. If $a \neq b$, then (W,a) must enter V in a different position than that of (W,b). Moreover, if W enters V in a position which is not the last one, that is, if W does not enter V at the position $|V| - |W|$, there must exist a letter, say a, at the right side of W, such that (W,a) still enters V_n at the same position. Now we can make two remarks. First, the number of entrances of W in V is always greater or equal than the sum over all the letters a of the numbers of entrances of (W,a) in V, for if (W,a) enters V, then W also enters. Second, if W enters V except for the last position, (W,a) enters V for some a, as explained before. Therefore

$$0 \le \#(W \;\text{ in }\; V) - \sum_a \#((W,a) \;\text{ in }\; V) \le 1$$

for any word V. Then, multiplying the above expression by $(X_n)_V$, where $(X_n)_V$ stands for the probability of $X_n = V$, we get

$$0 \le \{X_n\}_V \cdot \#(W \;\text{ in }\; V) - (X_n)_V \cdot \sum_a \#((W,a) \;\text{ in }\; V) \le (X_n)_V.$$

Thus, summing over all words V, and noting that the set $\{V : X_V > 0\}$ is finite, yields

$$0 \le E\#(W \;\text{ in }\; X_n) - \sum_a E\#((W,a) \;\text{ in }\; X_n) \le 1.$$

Now, dividing by $(E|X_n| - |W| + 1)$ gives

$$0 \leq \text{freq}(W \ \text{ in } \ X_n) - \sum_a \frac{E\#((W,a) \ \text{ in } \ X_n)}{E|X_n| - |W| + 1} \leq \frac{1}{E|X_n| - |W| + 1},$$

which yields

$$0 \leq \text{freq}(W \ \text{ in } \ X_n) - \sum_a \text{freq}((W,a) \ \text{ in } \ X_n) \frac{E|X_n| - |W|}{E|X_n| - |W| + 1} \leq \frac{1}{E|X_n| - |W| + 1},$$

since $1/(E|X_n| - |W| + 1) \to 0$ as $n \to \infty$, because $E|X_n| \to \infty$ as $n \to \infty$, and by the same reason,

$$\frac{E|X_n| - |W|}{E|X_n| - |W| + 1},$$

tends to 1 as $n \to \infty$. Therefore,

$$0 \leq \lim_{n\to\infty} \text{freq}(W \ \text{ in } \ X_n) - \sum_a \lim_{n\to\infty} \text{freq}((W,a) \ \text{ in } \ X_n) \leq 0$$

that is

$$\mu(W) = \sum_a \mu(W,a).$$

The argument for a on the left side is analogous. Thus, the map $\mu(\cdot)$ is indeed a measure.

In the other direction. From the definition of convergence of words to a measure we have that for each W

$$\lim_{n\to\infty} \text{freq}(W \ \text{ in } \ X_n) = \mu(W).$$

So, let N be such that for every $m,\ n > N$, we have

$$|\,\text{freq}(W \ \text{ in } \ X_m) - \mu(W)| < \frac{\varepsilon}{2} \ \text{ and } \ |\,\text{freq}(W \ \text{ in } \ X_n) - \mu(W)| < \frac{\varepsilon}{2}.$$

Now,

$$|\operatorname{freq}(W \text{ in } X_m) - \operatorname{freq}(W \text{ in } X_n)| \leq$$

$$\leq |\operatorname{freq}(W \text{ in } X_m) - \mu(W)| + |\operatorname{freq}(W \text{ in } X_n) - \mu(W)| < \varepsilon.$$

Thus, if (X_n) tends to a measure, then it is Cauchy.

Theorem 2.0.22 is proved in both directions.

Definition 2.0.23. We denote $\mathcal{M}_{word} = \{V' : V \in \operatorname{dic}(\mathcal{A})\}$. We turn $\mathcal{M}_{word}$ into a metric space defining $\operatorname{dist}(\mu, \nu) = \|\mu - \nu\|$, where the norm $\|\cdot\|$ is the norm of the space $\mathcal{M}_{norm}$.

Definition 2.0.24. Given any metric space S, we say that a sequence of non-empty sets $S_1, S_2, S_3, \dots \subset S$ converges or tends to a point $p \in S$ if for any $\varepsilon > 0$ there is k such that

$$\forall\, n > k \;\; \forall\, q \in S_n : \operatorname{dist}(q, p) < \varepsilon.$$

Definition 2.0.25. We say that a sequence of sets $S_1, \dots, S_n, \dots \subset \mathcal{M}_{word}$ is Cauchy if

$$\forall\, \varepsilon > 0\; \exists\, N\; \forall\, n, m > N : \;\; \forall\, p \in S_n\; \forall\, q \in S_m : \operatorname{dist}(q, p) < \varepsilon.$$

Theorem 2.0.26. *A sequence of sets $S_1, S_2, S_3, \dots, \subset \mathcal{M}_{word}$ is Cauchy if and only if these sets converge to some measure (in the sense of Definition 2.0.24).*

Proof in one direction. Suppose that $S_1, \dots, S_n, \dots$ is a Cauchy sequence of sets, that is

$$\forall\, \varepsilon > 0\; \exists\, N\; \forall\, n, m > N : \;\; \forall\, p \in S_n\; \forall\, q \in S_m : \|p - q\| < \varepsilon.$$

Construct a sequence p_n by selecting for each n an element p_n in S_n. Then for every $\varepsilon > 0$ there exists an N such that for $n, m > N$ we have $\|p_n - p_m\| < \varepsilon$, since $p_n \in S_n$

and $p_m \in S_m$, and we are assuming the sequence of sets S_n to be Cauchy. Therefore, the sequence of pseudo-measures (p_n) is Cauchy. But p_n are also in $\mathcal{M}_{word}$. Thus there is some sequence of words (V_n) such that $p_n = V'_n$, and therefore the sequence (V'_n) is Cauchy. Thus, for any word W there exists j such that $W = W_j$ and therefore

$$|V'_n(W) - V'_m(W)| = |V'_n(W_j) - V'_m(W_j)| \leq \frac{1}{w_j}\|V'_n - V'_m\| \leq \frac{\varepsilon}{w_j},$$

which implies that the sequence (V_n) is Cauchy. By Theorem 2.0.20, this sequence converges to some measure μ, and by the equivalence in Definition 2.0.3, the sequence of pseudo-measures V'_n converges to one and the same measure μ. We now have to prove that the sequence of sets also converges to μ: $\forall\, \varepsilon > 0\, \exists\, k\, \forall\, n > k\, \forall p \in S_n :\ \|\mu - p\| < \varepsilon$. But this follows trivially from the fact that, since the sequence (S_n) is Cauchy, we have

$$\forall \varepsilon > 0\ \exists\ N\ \forall\ n, m > N\, \forall\ q \in S_m. : \|p_n - q\| < \varepsilon.$$

This implies that if n is so large that $\|p_n - \mu\| < \varepsilon/2$ and that $\|p_n - q\| < \varepsilon/2$ for any $q \in S_n$, then

$$\|q - \mu\| \leq \|q - p_n\| + \|p_n - \mu\| \leq \varepsilon.$$

This concludes the proof in one direction.

In the other direction. Suppose that the sequence of sets $S_1, \ldots, S_n, \ldots$ converges to some measure μ. Let $\varepsilon > 0$ and let n be so large that for all $p \in S_n$ we have $\|p - \mu\| < \varepsilon$. Then, if $m > n$, we have for any $p \in S_n$ and $q \in S_m$:

$$\|p - p\| \leq \|p - \mu\| + \|p - \mu\| \leq 2\varepsilon.$$

Theorem 2.0.26 is proved in both directions.

Chapter 3

Substitution Operators

A generic substitution operator is determined by two words G and H, where G is self-avoiding, and a real number $\rho \in [0,1]$. We denote this operator by $(G \xrightarrow{\rho} H)$. The informal idea of this operator is that it substitutes every entrance of the word G in a long word by the word H with a probability ρ or leaves it unchanged with a probability $1-\rho$ independently of states and fate of all the other components. Following some of the articles in this area, we write operators on the right side of objects (words, measures) on which they act.

3.1 General Definition of a Substitution Operator acting on Words

Our goal in this Section is to define a general substitution operator, which is denoted by $(G \xrightarrow{\rho} H)$, acting on words. If G and H are empty, the operator $(G \xrightarrow{\rho} H)$ changes nothing. Now let G or H be non-empty.

Let us choose a word V and define how $(G \xrightarrow{\rho} H)$ acts on it. Let us denote $N = \#(G \text{ in } V)$, i.e. N is the number of entrances of G in V. Since G is self-avoiding, these entrances do not overlap. Further, let $i_1, \ldots, i_N \in \{0,1\}$ and let us denote by $V(i_1, \ldots, i_N)$ the word obtained from V after replacing each entrance of G by the word H in those positions i_j where $i_j = 1$, the others left unchanged. We then define the substitution operator $(G \xrightarrow{\rho} H)$ as follows: the random word obtained from the word V is concentrated on the words $V(i_1, \ldots, i_N)$, where $i_1, \ldots, i_N \in \{0,1\}$ with probabilities

$$\mathbb{P}\left(V(G \xrightarrow{\rho} H) = V(i_1, \ldots, i_N)\right) = \rho^k (1-\rho)^{N-k}, \text{ where } k = \sum_j i_j.$$

Now let us extend this definition to random words.

Let us define the result of application of $(G \xrightarrow{\rho} H)$ to a random word X. Let X equal the words $V_1, \ldots, V_n$ with positive probabilities X_{V_j}. We define $X(G \xrightarrow{\rho} H)$ as the random word which equals the words $V_j(i_1, \ldots, i_N)$ with probabilities

$$X_{V_j} \cdot \rho^{\sum_j i_j} (1-\rho)^{N - \sum_j i_j}.$$

Lemma 3.1.1. *For any non-empty word V and $\rho \in [0,1]$ we can express the mean length of the random word $V(G \xrightarrow{\rho} H)$ in the following simple way*

$$E|V(G \xrightarrow{\rho} H)| = |V| + \rho \cdot (|H| - |G|) \cdot \#(G \text{ in } V).$$

Proof: Let us evaluate the mean length (Def. 2.0.8) of the random word $V(G \xrightarrow{\rho} H)$. We begin by noting that

$$|V(G \xrightarrow{\rho} H)| = |V| + (|H| - |G|) \cdot k,$$

where $k \sim \text{Bin}(N, \rho)$, where $N = \#(G \text{ in } V)$. Therefore,

$$E|V(G \xrightarrow{\rho} H)| = |V| + (|H| - |G|) \cdot E(k),$$

which is equal to

$$E|V(G \xrightarrow{\rho} H)| = |V| + \rho \cdot (|H| - |G) \cdot \#(G \text{ in } V).$$

Lemma 3.1.1 is proved.

Lemma 3.1.2. *For any random word X and $\rho \in [0,1]$ we can express the mean length of the random word $X(G \stackrel{\rho}{\to} H)$ in the following simple way*

$$E|X(G \stackrel{\rho}{\to} H)| = E|X| + \rho \cdot (|H| - |G|) \cdot E\#(G \text{ in } X).$$

Proof: Suppose that X may assume the words $V_1, \ldots, V_n$, then, we first note that

$$E|X(G \stackrel{\rho}{\to} H)| = \sum_{j=1}^{n} E|V_j(G \stackrel{\rho}{\to} H)|X_{V_j} \cdot \rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j}.$$

Now, we use Lemma 3.1.1 to obtain

$$\begin{aligned} E|X(G \stackrel{\rho}{\to} H)| &= \sum_{j=1}^{n}[|V_j| + \rho \cdot (|H| - |G|)\#(G \text{ in } V_j)]X_{V_j} \cdot \rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j} \\ &= E|X| + \rho \cdot (|H| - |G|)E\#(G \text{ in } X). \end{aligned}$$

Lemma 3.1.2 is proved.

3.2 Operators Acting on Random Words

Remember our usual notation (see also the Index of Symbols): $\mathcal{A}$ is an alphabet, Ω is the set of random words on $\mathcal{A}$, $\mathcal{M}$ is the set of uniform probability measures on $\mathrm{dic}(\mathcal{A})$, V' is the pseudo-measure corresponding to the word V, and $\mathcal{M}_{word} = \{V'; V \in \mathrm{dic}(\mathcal{A})\}$.

3.2.1 Consistency

Definition 3.2.1. We say that a map $P : \Omega \to \Omega$ is *consistent* if any of the following equivalent conditions holds:

a) for any $\mu \in \mathcal{M}$ and any sequence of random words $(X_n) \to \mu$ the limit

$$\lim_{n\to\infty}(X_n P)$$

exists and is one and the same for all sequences $X_n \to \mu$.

b) for any $\mu \in \mathcal{M}$ and any sequence of sets

$$S_1, S_2, \ldots, S_n, \ldots \subset \mathcal{M}_{word}$$

such that $S_n \to \mu$, the limit

$$\lim_{n\to\infty} S_n P$$

exists and is one and the same for all sequences $(S_n) \to \mu$, where $S_n P = \{VP : V \in S_n\}$.

Lemma 3.2.2. *The conditions (a) and (b) in Definition 3.2.1 are equivalent.*

Proof: We begin proving that if the limit $X_n P$ exists for all sequences $(X_n) \to \mu$, then the limit of $S_n P$ exists for all sequences $(S_n) \to \mu$.

From Definition 2.0.24 it is clear that if $S_n \to \mu$, then $V_n \to \mu$ for any sequence (V_n) such that each $V_n \in S_n$. Therefore, by assumption, we have that for each sequence (V_n), where V_n belongs to S_n, the limit $\lim_{n\to\infty} V_n P$ exists and is one and the same. Therefore the limit of sequence of sets $\lim_{n\to\infty} S_n P$ also exists and is one and the same for each sequence $(S_n) \to \mu$.

In the other direction: if for each sequence $(S_n) \to \mu$, the limit of $S_n P$ exists and is one and the same, it is enough to choose the sequence of sets $S_n = \{X_n\}$, that is the set containing only one element, which easily implies the desired result. *Lemma 3.2.2 is proved.*

Definition 3.2.3. Given any consistent map $P : \Omega \to \Omega$ (see Def. 3.2.1), and any $\mu \in \mathcal{M}$, we define μP, that is, the result of application of P to μ, as the measure (which is unique according to Definition 3.2.1), to which $(X_n P)$ converges for all $(X_n) \to \mu$.

Lemma 3.2.4. *Let* $P_1,\ P_2 : \Omega \to \Omega$ *be consistent operators. Then their superposition is also consistent.*

Proof: Consider any sequence of random words (X_n) converging to μ. Then, the sequence of random words $Q_n = X_n P_1$ tends to μP_1 (following Definition 3.2.3), hence $Q_n P_2$ tends to $\mu P_1 P_2$.

Lemma 3.2.4 is proved.

3.2.2 Extension

Definition 3.2.5. For any $\mu \in \mathcal{M}$ and any $P : \Omega \to \Omega$, we define *extension* of μ under P as the limit

$$\mathrm{Ext}(\mu|P) = \lim_{n\to\infty} \frac{E|X_n P|}{E|X_n|}$$

for any sequence of random words $(X_n) \to \mu$ if this limit exists and is one and the same for all sequences (X_n) which tend to μ.

Informally speaking, extension of a measure μ under operator P is that coefficient by which P multiplies the length of a long word approximating μ.

Lemma 3.2.6. *Suppose that* $P_1,\ P_2 : \Omega \to \Omega$ *have extensions for all measures and* P_1 *is consistent. Then their superposition* $P_1 P_2$ *also has extension for all measures and*

$$\forall\ \mu : \mathrm{Ext}(\mu|P_1 P_2) = \mathrm{Ext}(\mu|P_1) \times \mathrm{Ext}(\mu P_1|P_2).$$

Proof: Since we are assuming that P_1 is consistent, we have by Definition 3.2.3 that $X_n P_1 \to \mu P_1$ as $n \to \infty$ for any sequence (X_n) of random words converging to μ. Thus, since we are assuming that P_2 has extension for all measures $\mu \in \mathcal{M}_{norm}$, Definition 3.2.5 implies that

$$\mathrm{Ext}(\mu P_1 | P_2) = \lim_{n\to\infty} \frac{E|V_n P_2|}{E|V_n|},$$

for any sequence of random words (V_n) converging to μP_1. Thus, since $X_n P_1 \to \mu P_1$, as seen in the beginning of the proof,

$$\mathrm{Ext}(\mu P_1 | P_2) = \lim_{n\to\infty} \frac{E|X_n P_1 P_2|}{E|X_n P_1|},$$

Therefore,

$$\lim_{n\to\infty} \frac{E|X_n P_1 P_2|}{E|X_n|} = \lim_{n\to\infty} \left(\frac{E|X_n P_1 P_2|}{E|X_n P_1|} \cdot \frac{E|X_n P_1|}{E|X_n|} \right) =$$

$$\lim_{n\to\infty} \frac{E|X_n P_1 P_2|}{E|X_n P_1|} \cdot \lim_{n\to\infty} \frac{E|X_n P_1|}{E|X_n|} = \mathrm{Ext}(\mu P_1 | P_2) \cdot \mathrm{Ext}(\mu | P_1).$$

The above expression implies that the extension of μ resulting from application of a superposition $P_1 P_2$ exists and equals

$$\mathrm{Ext}(\mu | P_1 P_2) = \mathrm{Ext}(\mu P_1 | P_2) \times \mathrm{Ext}(\mu | P_1).$$

Lemma 3.2.6 is proved.

3.3 Extension of a Substitution Operator

Now let us show that every measure has an extension under every substitution operator and provide an explicit expression for it.

Proposition 3.3.1. *If* $(G \xrightarrow{\rho} H)$ *is a substitution operator acting on random words, then the extension of any* $\mu \in \mathcal{M}$ *under this operator exists and equals*

$$\mathrm{Ext}(\mu | (G \xrightarrow{\rho} H)) = 1 + \rho \cdot (|H| - |G|) \cdot \mu(G), .$$

Proof: We know from Lemma 3.1.2 that:

$$E|X_n(G \xrightarrow{\rho} H)| = E|X_n| + \rho \cdot (|H| - |G|) \cdot E\#(G \ \text{ in } \ X_n)$$

Dividing the above expression by $E|X_n|$ yields

$$\frac{E|X_n(G \xrightarrow{\rho} H)|}{E|X_n|} = 1 + \rho(|H| - |G|) \cdot \text{freq}(G \ \text{ in } \ X_n) \frac{E|X_n| - |G| + 1}{E|X_n|}.$$

Therefore

$$\lim_{n \to \infty} \frac{E|V_n(G \xrightarrow{\rho} H)|}{|V_n|} = 1 + \rho(|H| - |G|) \cdot \mu(G).$$

Proposition 3.3.1 is proved.

3.4 General Definition of a Substitution Operator acting on Measures

Given a measure μ and a triple (G, ρ, H), a generic substitution operator acting on measures is also denoted by $(G \xrightarrow{\rho} H)$, where G and H are words, G is self-avoiding and ρ is a number in $[0, 1]$. Informally speaking, this operator substitutes every entrance of the word G by the word H with a probability ρ or leaves it unchanged with a probability $1 - \rho$ independently of states and fate of the other components.

The main task of this section is to define a general substitution operator acting on measures and to prove that it is always **consistent**. If both G and H are empty, our operator $(G \xrightarrow{\rho} H)$ leaves all measures unchanged by definition. Leaving this trivial case aside, we assume that at least one of the words G and H is non-empty.

The following attempt of definition seems the most natural to us. Given a measure μ, approximate it with a sequence of words (V_n), that is, let $(V_n) \to \mu$ as $n \to \infty$.

Then apply $(G \xrightarrow{\rho} H)$ to each V_n. Thus we get a sequence of random words

$$V_n(G \xrightarrow{\rho} H). \tag{3.1}$$

Then we would like to prove that the above sequence is Cauchy and therefore converges to a measure and that this measure is one and the same for all sequences (V_n) which tend to μ. All this proved, we might take this limit as the desired definition:

$$\mu(G \xrightarrow{\rho} H) = \lim_n V_n(G \xrightarrow{\rho} H).$$

To prove the statement above we will introduce several special operators, prove their consistency, and represent a general substitution operator as superposition of those operators.

These special classes will also help us in another way. Having described our work up to this point, we see that our approach is more general than the areas mentioned in the introduction. One might say that we generalize beyond necessity. The following special classes of operators help us to show that our theoretic constructions are not very far from possible applications.

3.4.1 Basic Substitution Operators

Let us define several small classes of operators, which we call *basic operators* and prove that all of them are consistent. In doing this we follow our previous setup of consistent operators to define how they act on measures (see Definition 3.2.3).

Basic operator 1: Conversion $(g \xrightarrow{\rho} h)$ is the only linear operator in our list. For any two different letters $g,\ h \in A$, we define *conversion from* g *to* h as a map from $\mathcal{M}$ to $\mathcal{M}$. Informally, conversion changes each occurrence of the letter g into the letter h

with probability $\rho \in [0,1]$ or doesn't change it with probability $1-\rho$ independently of the states of the other occurrences. In biological literature the analogous transformation is often called *substitution.*

Lemma 3.4.1. *The basic operator conversion is consistent.*

Proof: Let (V_n) be a sequence of words converging to μ. We know that the extension of this operator equals 1, that is

$$Ext(\mu|(g \xrightarrow{\rho} h)) = 1.$$

Therefore, it is sufficient to verify that the following limit exists:

$$\lim_{n\to\infty} \frac{E\#(W \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|},$$

since, from the expression of the extension of this operator, we have the identity

$$\lim_{n\to\infty} \frac{E\#(W \text{ in } V_n(g \xrightarrow{\rho} h))}{E|V_n(g \xrightarrow{\rho} h)| - |W| + 1} = \lim_{n\to\infty} \frac{E\#(W \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|}.$$

Indeed, denoting $m = \#(g \text{ in } V_n)$, it is easy to see that

$$\#(g \text{ in } V_n(g \xrightarrow{\rho} h)) \sim \mathrm{Bin}(m, 1-\rho),$$

and hence

$$E\#(g \text{ in } V_n(g \xrightarrow{\rho} h)) = (1-\rho)\#(g \text{ in } V_n).$$

This yields

$$\lim_{n\to\infty} \frac{E\#(g \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|} = (1-\rho) \lim_{n\to\infty} \frac{\#(g \text{ in } V_n)}{|V_n|} = (1-\rho)\mu(g).$$

After similar calculations we obtain

$$\lim_{n\to\infty} \frac{E\#(h \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|}.$$

We now note that it is easy to see that

$$\#(h \text{ in } V_n(g \xrightarrow{\rho} h)) = \#(h \text{ in } V_n) + K,$$

where $K \sim \text{Bin}(m, \rho)$ represents the letters g that turned into h, and $m = \#(g \text{ in } V_n)$.

Therefore

$$E\#(h \text{ in } V_n(g \xrightarrow{\rho} h)) = \#(h \text{ in } V_n) + \rho \cdot \#(g \text{ in } V_n),$$

whence

$$\lim_{n\to\infty} \frac{E\#(h \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|} = \lim_{n\to\infty} \frac{\#(h \text{ in } V_n)}{|V_n|} + \rho \lim_{n\to\infty} \frac{\#(g \text{ in } V_n)}{|V_n|} = \mu(h) + \rho \cdot \mu(g).$$

For any letter e different from g and h

$$\lim_{n\to\infty} \frac{E\#(e \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|} = \lim_{n\to\infty} \frac{\#(e \text{ in } V_n)}{|V_n|} = \mu(e).$$

Thus, we define how this operator acts on the measure μ as follows:

$$\mu(g \xrightarrow{\rho} h)(g) = \lim_{n\to\infty} \frac{E\#(g \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|},$$

$$\mu(g \xrightarrow{\rho} h)(h) = \lim_{n\to\infty} \frac{E\#(h \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|}$$

and

$$\mu(g \xrightarrow{\rho} h)(e) = \lim_{n\to\infty} \frac{E\#(e \text{ in } V_n(g \xrightarrow{\rho} h))}{|V_n|}.$$

Then, if we let $F(g|g) = 1 - \rho$, $F(h|g) = \rho$ and $F(h|h) = F(e|e) = 1$, we have

$$\mu(g \xrightarrow{\rho} h)(g) = F(g|g)\mu(g) = (1-\rho)\mu(g),$$

$$\mu(g \xrightarrow{\rho} h)(h) = F(h|h)\mu(h) + F(h|g)\mu(g) = \mu(h) + \rho \cdot \mu(g)$$

and

$$\mu(g \xrightarrow{\rho} h)(e) = F(e|e)\mu(e) = \mu(e).$$

More generally, given a word $W = (a_1, \ldots, a_k)$, we have by similar calculations:

$$\lim_{n\to\infty} \frac{E\#(W \ \text{ in } \ V_n(g \xrightarrow{\rho} h))}{|V_n|} = \sum_{b_1,\ldots,b_k \in \mathcal{A}} \left(\prod_{i=1}^{k} F(a_i|b_i) \times \mu(b_1, \ldots, b_k) \right),$$

where

$$F(a|b) = \begin{cases} 1-\rho & \text{if } b = g, a = g, \\ \rho & \text{if } b = g, a = h, \\ 0 & \text{if } b = g \text{ and } a \text{ is not } g \text{ nor } h, \\ 1 & \text{if } b \neq g \text{ and } a = b, \\ 0 & \text{if } b \neq g \text{ and } a \neq b. \end{cases}$$

Lemma 3.4.1 is proved.

Now we can use the consistency of this operator to define the conversion operator acting on any measure μ applied in any word $W = (a_1, \ldots, a_k)$:

$$\mu(g \xrightarrow{\rho} h)(W) = \sum_{b_1,\ldots,b_k \in \mathcal{A}} \left(\prod_{i=1}^{k} F(a_i|b_i) \times \mu(b_1, \ldots, b_k) \right).$$

Basic operator 2: Insertion $(\Lambda \xrightarrow{\rho} h)$. Informally, insertion of a letter $h \notin \mathcal{A}$ into a measure in the alphabet $\mathcal{A}$ with a rate $\rho \in [0,1]$ means that a letter h is inserted with probability ρ between every two neighbor letters independently from other places. This term is identical to the one used in biological literature and the meaning is similar.

We already know that the extension of any μ for this operator equals

$$\text{Ext}(\mu|(\Lambda \xrightarrow{\rho} h)) = 1 + \rho.$$

Lemma 3.4.2. *The basic operator insertion is consistent.*

Proof: Let (V_n) be a sequence of words converging to some measure μ. It is enough to prove that the left and right sides of the following equation exist:

$$\lim_{n\to\infty} \frac{E\#(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))}{E|V_n(\Lambda \xrightarrow{\rho} h)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(\Lambda \xrightarrow{\rho} h))} \lim_{n\to\infty} \frac{E\#(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))}{|V_n|}. \quad (3.2)$$

The fact that the above limits are equal comes from the definition and existence of extension of this operator. We begin by noting that for any word W in the alphabet $\mathcal{A}' = \mathcal{A} \cup \{h\}$:

$$E\#(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h))$$

$$= \sum_{i_j \in \{0,1\}} \#(W \ \text{ in } \ V_n(i_1, \ldots, i_{|V_n|+1})\rho^{\sum_j i_j}(1-\rho)^{|V_n|+1-\sum_j i_j},$$

where $V_n(i_1, \ldots, i_{|V_n|+1})$ is the word obtained from V_n after inserting the letter h in the positions where $i_j = 1$. It is clear that if $\sum_j i_j < \#(h \ \text{ in } \ W)$, then $\#(W \ \text{ in } \ V_n(i_1, \ldots, i_{|V_n|+1}) = 0$, using this, we may simplify the above expression to obtain

$$E\#(W \ \text{ in } \ V_n(\Lambda \xrightarrow{\rho} h)) = \#(W' \ \text{ in } \ V_n)\rho^{\#(h \ \text{ in } \ W)}(1-\rho)^M,$$

where M is the number of pairs of consecutive letters in W both of which are not h, and W' is the word obtained from W by deleting all the letter h. Therefore it is easy to conclude that the limits on 3.2 exist.

Lemma 3.4.2 is proved.

Then we use the consistency of this operator to define how the operator $(\Lambda \xrightarrow{\rho} h)$ acts on any measure μ. We define the result of its application to an arbitrary word W

by:

$$\mu(\Lambda \xrightarrow{\rho} h)(W)$$

$$= \frac{1}{\mathrm{Ext}(\mu|(\Lambda \xrightarrow{\rho} h))}\mu(W')\rho^{\#(h \text{ in } W)}(1-\rho)^M$$

$$= \frac{1}{1+\rho}\mu(W')\rho^{\#(h \text{ in } W)}(1-\rho)^M$$

Basic operator 3: Deletion $(g \xrightarrow{\rho} \Lambda)$. Informally, deletion of a letter $g \in \mathcal{A}$ with some probability $\rho \in [0,1)$ means that each occurrence of g disappears with probability ρ or remains unchanged with probability $1-\rho$ independently from the other occurrences. Like *insertion* operator, this operator is similar to what is called *deletion* in the biological literature.

We already know that the extension of any measure μ under this operator is

$$\mathrm{Ext}(\mu|(g \xrightarrow{\rho} \Lambda)) = 1 - \rho \cdot \mu(g).$$

Lemma 3.4.3. *The basic operator deletion is consistent.*

Proof: Let (V_n) be a sequence of words converging to a measure μ . We then must prove that the following limits exist:

$$\lim_{n\to\infty} \frac{E\#(W \text{ in } V_n(g \xrightarrow{\rho} \Lambda))}{E|V_n(g \xrightarrow{\rho} \Lambda)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(g \xrightarrow{\rho} \Lambda))} \lim_{n\to\infty} \frac{E\#(W \text{ in } V_n(g \xrightarrow{\rho} \Lambda))}{|V_n|}, \quad (3.3)$$

The equality in the above expression is due to the definition of extension. Let $W = (a_0, \ldots, a_m)$ be any word and $N_n = \#(g \text{ in } V_n)$, then, from Definition 2.0.9

$$E\#(W \text{ in } V_n(g \xrightarrow{\rho} \Lambda)) = \sum_{k;i_1,\ldots,i_k} \#(W \text{ in } V_n(k;i_1,\ldots,i_k))\rho^k(1-\rho)^{N_n-k},$$

where $M = \#(g \text{ in } W)$ and $V(k; i_1, \ldots, i_k)$ is the word obtained from V by the deletion of k letters g in positions $i_1, \ldots, i_k$.

Then, by a "change of variables", we may disregard the letters g deleted from V_n, and put them in the word W. With this the number k of deletions is now the total number of insertions of the letter g in the word W:

$$E\#(W \text{ in } V_n(g \overset{\rho}{\to} \Lambda)) = \sum_{k; i_1, \ldots, i_k}$$

$$= \sum_{n_1+\cdots+n_{m+1} \leq N_n} \#((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \text{ in } V_n)\rho^{n_1+\cdots+n_m+n_{m+1}}(1-\rho)^M,$$

Therefore, we have:

$$\frac{E\#(W \text{ in } V_n(g \overset{\rho}{\to} \Lambda))}{|V_n|}$$

$$= \sum_{n_1+\cdots+n_{m+1} \leq N_n} \frac{\#((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \text{ in } V_n)\rho^{n_1+\cdots+n_{m+1}}(1-\rho)^M}{|V_n|}$$

$$= \sum_{n_1, \ldots, n_{m+1}=0}^{\infty} \mathrm{I}_{\{n_1+\cdots+n_{m+1} \leq N_n\}} \frac{\#((g^{n_1}, a_1, g^{n_2}, \ldots, a_m, g^{n_{m+1}}) \text{ in } V_n)\rho^{n_1+\cdots+n_{m+1}}(1-\rho)^M}{|V_n|}$$

where I_A is the indicator function of the set A, and, in the last identity, the indicator function was used to avoid dependence of n on the index of summation.

Clearly, $\mathrm{I}_{\{n_1+\cdots+n_{m+1} \leq N_n\}} \overset{n\to\infty}{\longrightarrow} 1$ (i.e. converges to the function identically equal to 1).

Also

$$\left| \mathrm{I}_{\{n_1+\cdots+n_{m+1} \leq N_n\}} \frac{\#((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \text{ in } V_n)}{|V_n|} \right| \leq 1,$$

and, since

$$\sum_{n_1, \ldots, n_{m+1}=0}^{\infty} \rho^{n_1+\cdots+n_{m+1}}(1-\rho)^M < \infty \text{ if } \rho < 1,$$

from the dominated convergence theorem:

$$\lim_{n\to\infty} \frac{E\#(W \text{ in } V_n(g \stackrel{\rho}{\to} \Lambda))}{|V_n|}$$

$$= \lim_{n\to\infty} \sum_{n_1,\dots,n_{m+1}=0}^{\infty} \mathrm{I}_{\{n_1+\dots+n_{m+1}\le N_n\}} \frac{\#((g^{n_1},a_1,\dots,a_m,g^{n_{m+1}}) \text{ in } V_n)\rho^{n_1+\dots+n_{m+1}}(1-\rho)^M}{|V_n|}$$

$$= \sum_{n_1,\dots,n_{m+1}=0}^{\infty} \lim_{n\to\infty} \mathrm{I}_{\{n_1+\dots+n_{m+1}\le N_n\}} \frac{\#((g^{n_1},a_1,\dots,a_m,g^{n_{m+1}}) \text{ in } V_n)\rho^{n_1+\dots+n_{m+1}}(1-\rho)^M}{|V_n|}$$

$$= \sum_{n_1,\dots,n_{m+1}=0}^{\infty} \mu(g^{n_1},a_1,\dots,a_m,g^{n_{m+1}})\rho^{n_1+\dots+n_{m+1}}(1-\rho)^M$$

Now it is easy to conclude that both limits in equation 3.3 exist.

Lemma 3.4.3 is proved.

Then we use the consistency of this operator to define how the operator $(g \stackrel{\rho}{\to} \Lambda)$ acts on an arbitrary measure μ:

$$\mu(g \stackrel{\rho}{\to} \Lambda)(W) =$$

$$\frac{1}{\mathrm{Ext}(\mu|(g \stackrel{\rho}{\to} \Lambda))} \sum_{n_1,\dots,n_{m+1}=0}^{\infty} \mu(g^{n_1},a_1,\dots,a_m,g^{n_{m+1}})\rho^{n_1+\dots+n_{m+1}}(1-\rho)^M =$$

$$\frac{1}{1-\rho\cdot\mu(g)} \sum_{n_1,\dots,n_{m+1}=0}^{\infty} \mu(g^{n_1},a_1,\dots,a_m,g^{n_{m+1}})\rho^{n_1+\dots+n_{m+1}}(1-\rho)^M,$$

for all words $W = (a_1,\dots,a_m)$, and $M = \#(g \text{ in } W)$.

Basic operator 4: Compression $(G \stackrel{1}{\to} h)$. Given a non-empty self-avoiding word G in an alphabet $\mathcal{A}$ and a letter $h \notin \mathcal{A}$, *compression from* G *into* h is the following map from $\mathcal{M}_{\mathcal{A}}$ to $\mathcal{M}_{\mathcal{A}'}$, where $\mathcal{A}' = \mathcal{A} \cup \{h\}$: each occurrence of the word G is replaced by the letter h with probability 1.

We already know that the extension of any measure μ under this operator is

$$\mathrm{Ext}(\mu|(G \xrightarrow{1} h)) = 1 - (|G| - 1) \cdot \mu(G).$$

Lemma 3.4.4. *The basic operator compression is consistent.*

Proof: Let (V_n) be a sequence of words converging to μ. Then, it is enough to prove that the following limits exist.

$$\lim_{n\to\infty} \frac{\#(W \text{ in } V_n(G \xrightarrow{1} h))}{|V_n(G \xrightarrow{\rho} h)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(G \xrightarrow{1} h))} \lim_{n\to\infty} \frac{\#(W \text{ in } V_n(G \xrightarrow{1} h))}{|V_n|}. \tag{3.4}$$

The identity on the equation above follows from the definition of extension of an operator. First notice that for any word W in the alphabet $\mathcal{A}'$

$$\#(W \text{ in } V_n(G \xrightarrow{1} h)) = \#(W' \text{ in } V_n) - \#(W \text{ in } G)\#(G \text{ in } V_n),$$

where W' is the word obtained from W by replacing every letter h by the word G. Then

$$\begin{aligned}\lim_{n\to\infty} \frac{\#(W \text{ in } V_n(G \xrightarrow{1} h))}{|V_n|} &= \lim_{n\to\infty} \frac{\#(W' \text{ in } V_n) - \#(W \text{ in } G)\#(G \text{ in } V_n)}{|V_n|} \\ &= \mu(W') - \#(W \text{ in } G)\mu(G).\end{aligned}$$

It is easy to see that the limits in equation 3.4 exist.

Lemma 3.4.4 is proved.

Now we use the consistency of this operator to define how the operator $(G \xrightarrow{1} h)$ acts on any measure μ:

$$\mu(G \xrightarrow{1} h)(W) = \frac{\mu(W') - \#(W \text{ in } G)\mu(G)}{\mathrm{Ext}(\mu|(G \xrightarrow{1} h))} = \frac{\mu(W') - \#(W \text{ in } G)\mu(G)}{1 - (|G| - 1)\mu(G)}.$$

Basic operator 5: Decompression $(g \xrightarrow{1} H)$. Given a non-empty self-avoiding word H in an alphabet $\mathcal{A}$ and a letter $g \notin \mathcal{A}$, *decompression of* g *into* H is the following map from $\mathcal{M}_{\mathcal{A}'}$ to $\mathcal{M}_{\mathcal{A}}$, where $\mathcal{A}' = \mathcal{A} \cup \{g\}$: every occurrence of the letter g is replaced by the word H with probability 1.

We already know that the extension of any measure μ for this operator is

$$\mathrm{Ext}(\mu|(g \xrightarrow{1} H)) = 1 + (|H| - 1) \cdot \mu(g).$$

Lemma 3.4.5. *The basic operator decompression is consistent.*

Proof: Let (V_n) be a sequence of words converging to the measure μ. Then it is enough to prove that the following limits exist.

$$\lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} H))}{|V_n(g \xrightarrow{\rho} H)| - |W| + 1} = \frac{1}{\mathrm{Ext}(\mu|(g \xrightarrow{1} H))} \lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} H))}{|V_n|}. \tag{3.5}$$

The equality in the above equation follows from the definition of extension. Let us prove that the limits in equation 3.5 exist. First let us consider the decompression of the letter g into the word (h_1, h_2) with probability 1, where the letters h_1 and h_2 are different and do not belong to the alphabet $\mathcal{A}$. The extension for this operator equals $1 + \mu(g)$. Now let us compute the following limit

$$\lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, h_2)))}{|V_n|}.$$

Let W be a word in the alphabet $\mathcal{A} \cup \{h_1, h_2\}$. We define a new word W' as the concatenation $W' = \mathrm{concat}(U, W, V)$, where

$$U = \begin{cases} h_1 & \text{if the first letter of } W \text{ is } h_2, \\ \Lambda & \text{otherwise.} \end{cases}$$

and

$$V = \begin{cases} h_2 & \text{if the last letter of } W \text{ is } h_1, \\ \Lambda & \text{otherwise.} \end{cases}$$

Afterwards, we turn each entrance of the word (h_1, h_2) in W' into the letter g and denote the resulting word by W''. (This is well-defined since the word (h_1, h_2) is self-avoiding.) Now, if W'' contains any entrance of h_1 or h_2 it means that $\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, h_2))) = 0$, and therefore the above limit equals zero. Otherwise,

$$\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, h_2))) = \#(W'' \text{ in } V_n),$$

and thus,

$$\lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, h_2)))}{|V_n|} = \lim_{n\to\infty} \frac{\#(W'' \text{ in } V_n)}{|V_n|} = \mu(W'').$$

Therefore, if W'' constrains any entrance of h_1 or h_2, we define $\mu(g \xrightarrow{1} (h_1, h_2))) = 0$, otherwise, we define:

$$\mu(g \xrightarrow{1} (h_1, h_2)))(W) = \frac{\mu W''}{\mathrm{Ext}(\mu|(g \xrightarrow{1} (h_1, h_2)))} = \frac{\mu(W'')}{1 + \mu(g)}.$$

Now, we will define the decompression of a letter g into the word $(h_1, \ldots, h_k)$ with probability 1, where the letters $h_1, \ldots, h_k$ are all different from each other and do not belong to the alphabet $\mathcal{A}$. Let us then define how the operator $(g \xrightarrow{1} (h_1, \ldots, h_k)))$ acts on the measure μ by induction in k. The case $k = 2$ was treated above. Now, let us take $k > 2$ and a letter s not belonging to $\mathcal{A}$. Then for any word W in the alphabet $\mathcal{A} \cup \{h_1, \ldots, h_k\}$

$$\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, \ldots, h_k))) = \#(W \text{ in } (V_n(g \xrightarrow{1} (h_1, s)))(s \xrightarrow{1} (h_2, \ldots, h_k))),$$

and then we can prove by induction that the following limit exists and is well-defined:

$$\lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, \ldots, h_k)))}{|V_n|} =$$

$$\lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k)))}{|V_n|}.$$

Now it is easy to see that the limits in equation 3.5 exist.

Lemma 3.4.5 is proved.

We will now use the consistency of this operator to define how it acts on an arbitrary measure. Let $V_n \to \mu$ when $n \to \infty$, and assume that $h_1, \ldots, h_k, s$ do not belong to $\mathcal{A}$. Then

$$
\begin{aligned}
& \operatorname{Ext}(\mu|(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k))) \\
= \quad & \lim_{n\to\infty} \frac{|V_n(g \xrightarrow{1} (h_1, s)(s \xrightarrow{1} (h_2, \ldots, h_k))))}{|V_n|} \\
= \quad & \lim_{n\to\infty} \frac{|V_n(g \xrightarrow{1} (h_1, s))| + (|H| - 2)\#(s \ \text{ in } \ V_n(g \xrightarrow{1} (h_1, s)))}{|V_n|} \\
= \quad & \lim_{n\to\infty} \frac{|V_n(g \xrightarrow{1} (h_1, s))| + (|H| - 2)\#(g \ \text{ in } \ V_n))}{|V_n|} \\
= \quad & \lim_{n\to\infty} \frac{|V_n| + (|H| - 1)\#(g \ \text{ in } \ V_n)}{|V_n|} \\
= \quad & \operatorname{Ext}(\mu|(g \xrightarrow{1} (h_1, \ldots, h_k))).
\end{aligned}
$$

After that, we define how the operator $(g \xrightarrow{1} (h_1, \ldots, h_k))$ acts on an arbitrary measure μ in the following inductive way:

$$\mu(g \xrightarrow{1} (h_1, \ldots, h_k)) = \mu(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k)),$$

we can check this claim by noting that:

$$\begin{aligned} & \mu(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k))(W) \\ = & \lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k))}{\mathrm{Ext}(\mu|(g \xrightarrow{1} (h_1, s))(s \xrightarrow{1} (h_2, \ldots, h_k)))} \\ = & \lim_{n\to\infty} \frac{\#(W \text{ in } V_n(g \xrightarrow{1} (h_1, \ldots, h_k))}{\mathrm{Ext}(\mu|(g \xrightarrow{1} (h_1, \ldots, h_k)))} \\ = & \mu(g \xrightarrow{1} (h_1, \ldots, h_k))(W). \end{aligned}$$

This is a superposition of the decompression from g to (h_1, s) and the decompression from s into $(h_2, \ldots, h_k)$.

Finally, we define the decompression operator, acting on a measure μ, from a letter g to an arbitrary word $H = (s_1, \ldots, s_k)$, with no restrictions on letters $s_1, \ldots, s_k$, that us they may belong to the alphabet and/or coincide with each other in the following way: first, we use the decompression from g to a word $(h_1, \ldots, h_k)$, where all the letters $h_1, \ldots, h_k$ are different from each other and do not belong to the alphabet $\mathcal{A}$. Further, we perform k conversions, each with probability 1, from h_i to s_i for all $i = 1, \ldots, k$.

3.4.2 Superposition of Basic Operators

The main goal of this subsection is to give a general definition of the substitution operator $(G \xrightarrow{\rho} H)$ acting on measures. We will do it by representing an arbitrary substitution operator as a superposition of several basic operators, which we have defined in the previous section.

This definition is based on the following facts.

Theorem 3.4.6. *Let* $(G \xrightarrow{\rho} H)$ *be a substitution operator acting on words. Let also* $(G \xrightarrow{\rho} \Lambda)$ *and* $(\Lambda \xrightarrow{\rho} H)$ *be substitution operators acting on words, where* Λ *is the empty word,* G *is self-avoiding,* s, g, h *are different letters not belonging to* $\mathcal{A}$*, and* $\rho \in [0,1]$ *(and* $\rho < 1$ *for the operator* $(G \xrightarrow{\rho} \Lambda)$ *). Then, for any word* V *and any* W *:*

$$\#(W \text{ in } V(G \xrightarrow{\rho} H)) = \#(W \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)), \quad (3.6)$$

$$\#(W \text{ in } V(G \xrightarrow{\rho} \Lambda)) = \#(W \text{ in } V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)), \quad (3.7)$$

and,

$$\#(W \text{ in } V(\Lambda \xrightarrow{\rho} H)) = \#(W \text{ in } V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)). \quad (3.8)$$

Also

$$E|V(G \xrightarrow{\rho} H)| = E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)|, \quad (3.9)$$

$$E|V(G \xrightarrow{\rho} \Lambda)| = E|V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)|, \quad (3.10)$$

and finally,

$$E|V(\Lambda \xrightarrow{\rho} H)| = E|V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)|. \quad (3.11)$$

We postpone the proof of this theorem to the end of this Section.
Before that we shall state several corollaries of it.

Corollary 3.4.7. *Let the substitution operators*

$$(G \xrightarrow{\rho} H), \quad (G \xrightarrow{\rho} \Lambda), \quad (\Lambda \xrightarrow{\rho} H)$$

act on words. Here G *is a self-avoiding word,* $s, g, h \notin \mathcal{A}$*, and* $\rho \in [0,1]$ *(and* $\rho < 1$ *for the operator* $(G \xrightarrow{\rho} \Lambda)$ *). Then, for any words* V*,* W

$$\text{freq}(W \text{ in } V(G \xrightarrow{\rho} H)) = \text{freq}(W \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)), \quad (3.12)$$

$$\text{freq}(W \ \text{ in } \ V(G \xrightarrow{\rho} \Lambda)) = \text{freq}(W \ \text{ in } \ V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)), \tag{3.13}$$

and

$$\text{freq}(W \ \text{ in } \ V(\Lambda \xrightarrow{\rho} H)) = \text{freq}(W \ \text{ in } \ V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)). \tag{3.14}$$

Therefore, if (V_n) *is a sequence of words, then*

$$V_n(G \xrightarrow{\rho} H) \ \ \textit{converges} \iff V_n(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G) \ \ \textit{converges}, \tag{3.15}$$

$$V_n(G \xrightarrow{\rho} \Lambda) \ \ \textit{converges} \iff V_n(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G) \ \ \textit{converges}, \tag{3.16}$$

and,

$$V_n(\Lambda \xrightarrow{\rho} H) \ \ \textit{converges} \iff V_n(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H) \ \ \textit{converges}. \tag{3.17}$$

Proof: Equation (3.12) follows directly from (3.6) and (3.9);

equation (3.13) follows directly from (3.7) and (3.10); and

equation (3.14) follows directly from (3.8) and (3.11).

Moreover,

equation (3.15) follows from (3.12);

equation (3.16) follows from (3.13); and

equation (3.17) follows from (3.14).

Corollary 3.4.7 is proved.

To imitate the $(G \xrightarrow{\rho} H)$ operator we compress each entrance of the word G into a letter h which does not belong to the alphabet. Then, we turn each letter h with probability ρ into a letter $s \neq h$ which does not belong to the alphabet. We then decompress the letter s into a word H and decompress the letter h into a word G.

To imitate the $(G \xrightarrow{\rho} \Lambda)$ operator we compress the word G into the letter g which does not belong to the alphabet $\mathcal{A}$. Then we delete the letter g with probability ρ, and lastly we decompress the letter g into the word G with probability 1.

Finally, to imitate the $(\Lambda \xrightarrow{\rho} H)$ operator we insert the letter h which does not belong to the alphabet, between every two consecutive letters with probability ρ independently from other places. After that we decompress the letter h into the word H with probability 1.

Corollary 3.4.8. *For any words G, H, where G is self-avoiding, and $\rho \in [0,1]$ (where $\rho < 1$ if $H = \Lambda$), the operator $(G \xrightarrow{\rho} H)$ acting on words is consistent.*

Proof: The identities (3.12), (3.13) and (3.14) yield that the substitution operator is the superposition of basic operators described in the last subsection, and each basic operator is consistent. Thus, by Lemma 3.2.4, their superposition is also consistent. Thus $(G \xrightarrow{\rho} H)$ is consistent.

Corollary 3.4.8 is proved.

The next Corollary shows how to define substitution operators acting on measures.

Corollary 3.4.9. *We define $\mu(G \xrightarrow{\rho} H)$, that is, the result of application of the operator $(G \xrightarrow{\rho} H)$ to a measure $\mu \in \mathcal{M}$ (following Definition 3.2.3) by*

$$\mu(G \xrightarrow{\rho} H)) = \lim_{n\to\infty} V_n(G \xrightarrow{\rho} H)),$$

where V_n is a sequence converging to μ. Corollary 3.4.9 is an obvious consequence of Corollary 3.4.8, and therefore we omitted the proof. Note also that the above definition is similar to the one we stated without proof in page 34.

Thus, we have succeeded to define an arbitrary substitution operator $(G \stackrel{\rho}{\to} H)$ acting on measures.

Corollary 3.4.10. *Consider the operator* $(G \stackrel{\rho}{\to} H)$ *acting on measures, for* G *self-avoiding and* $\rho \in [0,1]$ *(*$\rho < 1$ *if* $H = \Lambda$*). Then, for* $s, g, h \notin \mathcal{A}$*, the following identities hold:*

$$\mu(G \stackrel{\rho}{\to} H)) = \mu(G \stackrel{1}{\to} h)(h \stackrel{\rho}{\to} s)(s \stackrel{1}{\to} H)(h \stackrel{1}{\to} G),$$

$$\mu(G \stackrel{\rho}{\to} \Lambda)) = \mu(G \stackrel{1}{\to} g)(g \stackrel{\rho}{\to} \Lambda)(g \stackrel{1}{\to} G),$$

and

$$\mu(\Lambda \stackrel{\rho}{\to} H)) = \mu(\Lambda \stackrel{\rho}{\to} h)(h \stackrel{1}{\to} H).$$

Proof: It is a straightforward consequence of Corollary 3.4.7.

Corollary 3.4.10 is proved.

Proof of Theorem 3.4.6.

Now we start to prove theorem 3.4.6, which was stated on page 46.

Lemma 3.4.11. *For* $s, h \notin \mathcal{A}$*, we have:*

$$E\#(W \text{ in } V(G \stackrel{\rho}{\to} H)) = E\#(W \text{ in } V(G \stackrel{1}{\to} h)(h \stackrel{\rho}{\to} s)(s \stackrel{1}{\to} H)(h \stackrel{1}{\to} G)).$$

Proof:

$$E\#(W \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G))$$

$$= E\#(W' \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H))$$

$$= E\#(W'' \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s))$$

$$= \sum_{i_j \in \{0,1\}} E\#(W'' \text{ in } (V(G \xrightarrow{1} h)))(i_1, \dots, i_N)\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j}$$

$$= \sum_{i_j \in \{0,1\}} E\#(W \text{ in } (V(G \xrightarrow{1} h)(s \xrightarrow{1} H)(h \xrightarrow{1} G))(i_1, \dots, i_N))\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j}$$

$$= \sum_{i_j \in \{0,1\}} E\#(W \text{ in } V(i_1, \dots, i_N))\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j}$$

$$= E\#(W \text{ in } V(G \xrightarrow{\rho} H)),$$

where the word W' is obtained from W by replacing each entry of G by h, and the word W'' is obtained from W' by replacing each entry of H by s.

Lemma 3.4.11 is proved.

Lemma 3.4.12. *Suppose that $h, s \notin \mathcal{A}$, and that G and H are non-empty words, G is self-avoiding and μ is any measure. Then*

$$E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)| = E|V(G \xrightarrow{\rho} H)|.$$

Proof: First notice that for any different letters $g, h \in \mathcal{A}$ and any word V it follows from the definition of the conversion operator that

$$E\#(g \text{ in } V(g \xrightarrow{\rho} h)) = (1-\rho)\#(g \text{ in } V)$$

and

$$E\#(h \text{ in } V(g \xrightarrow{\rho} h)) = \#(h \text{ in } V) + \rho \cdot \#(g \text{ in } V).$$

Therefore

$$E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)|$$

$$= E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)|$$

$$+ (|G|-1)E\#(h \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H))$$

$$= E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)| + (|G|-1)E\#(h \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s))$$

$$= E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)| + (|G|-1)(1-\rho)\#(G \text{ in } V) \tag{3.18}$$

$$= E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)| + (|H|-1)E\#(s \text{ in } V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s))$$

$$+ (|G|-1)(1-\rho)\#(G \text{ in } V)$$

$$= E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)| + (|H|-1)\rho\#(G \text{ in } V) \tag{3.19}$$

$$+ (|G|-1)(1-\rho)\#(G \text{ in } V).$$

Hence

$$
\begin{aligned}
& E|V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)| \\
= \ & |V(G \xrightarrow{1} h)| + (|H|-1)\rho\#(G \ \text{ in } \ V) \\
+ \ & (|G|-1)(1-\rho)\#(G \ \text{ in } \ V) \\
= \ & |V| + (1-|G|)\#(G \ \text{ in } \ V) + (|H|-1)\rho\#(G \ \text{ in } \ V) \\
+ \ & (|G|-1)(1-\rho)\#(G \ \text{ in } \ V) \\
= \ & |V| + \#(G \ \text{ in } \ V)(1-|G|+\rho|H|-\rho+|G|-1-\rho|G|+\rho) \\
= \ & |V| + \rho(|H|-|G|)\#(G \ \text{ in } \ V) \\
= \ & E|V(G \xrightarrow{\rho} H)|
\end{aligned}
$$

Here (3.18) follows from (3.4.2), (3.19) follows from (3.4.2) and (3.20) follows from the fact that the conversion operator has extension one.

Lemma 3.4.12 is proved.

Lemma 3.4.13. *For any* $g \notin \mathcal{A}$*:*

$$E\#(W \ \text{ in } \ V(G \xrightarrow{\rho} \Lambda)) = E\#(W \ \text{ in } \ V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)).$$

Proof:

$$E\#(W \ \text{ in } \ V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G))$$

$$= \quad E\#(W' \text{ in } V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda))$$

$$= \quad \sum_{i_j \in \{0,1\}} E\#(W' \text{ in } (V(G \xrightarrow{1} g)))(i_1, \ldots, i_N)\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j}$$

$$= \quad \sum_{i_j \in \{0,1\}} E\#(W \text{ in } (V(G \xrightarrow{1} g)(g \xrightarrow{1} G))(i_1, \ldots, i_N))\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j}$$

$$= \quad \sum_{i_j \in \{0,1\}} E\#(W \text{ in } V(i_1, \ldots, i_N))\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j}$$

$$= \quad E\#(W \text{ in } V(G \xrightarrow{\rho} \Lambda)),$$

where W' is the word obtained from W by replacing each entry of G by g.

Lemma 3.4.13 is proved.

Lemma 3.4.14. *Given $g \in \mathcal{A}$ and a non-empty word V in the alphabet, we have*

$$E\#(g \text{ in } V(g \xrightarrow{\rho} \Lambda)) = (1-\rho)\#(g \text{ in } V).$$

Proof: First notice that for any $h \notin \mathcal{A}$

$$E\#(g \text{ in } V(g \xrightarrow{\rho} \Lambda)) = E\#(g \text{ in } V(g \xrightarrow{\rho} h)).$$

But we already know that

$$E\#(g \text{ in } V(g \xrightarrow{\rho} h)) = (1-\rho)\#(g \text{ in } V).$$

Lemma 3.4.14 is proved.

Lemma 3.4.15. *Suppose that $g \notin \mathcal{A}$ and that G is a self-avoiding non-empty word. Then for any measure μ*

$$E|V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)| = E|V(G \xrightarrow{\rho} \Lambda)|.$$

Proof: Let V be a word. Then, using Lemma 3.4.14 and Lemma 3.1.1, we have:

$$
\begin{aligned}
& E|V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)| \\
= \; & E|V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)| + (|G|-1)E\#(g \ \text{ in } \ V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)) \\
= \; & E|V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)| + (|G|-1)(1-\rho)\#(G \ \text{ in } \ V) \qquad (15) \\
= \; & |V(G \xrightarrow{1} g)| - \rho\#(g \ \text{ in } \ V(G \xrightarrow{1} g)) \\
+ \; & (|G|-1)(1-\rho)\#(G \ \text{ in } \ V) \\
= \; & |V| + (1-|G|)\#(G \ \text{ in } \ V) - \rho\#(G \ \text{ in } \ V) \\
+ \; & (|G|-1)(1-\rho)\#(G \ \text{ in } \ V) \\
= \; & |V| - \rho|G|\#(G \ \text{ in } \ V) = E|V(G \xrightarrow{\rho} \Lambda)|.
\end{aligned}
$$

Lemma 3.4.15 is proved.

Lemma 3.4.16. *For $g \notin \mathcal{A}$, we have:*

$$E\#(W \ \text{ in } \ V(\Lambda \xrightarrow{\rho} H)) = E\#(W \ \text{ in } \ V(\Lambda \xrightarrow{1} h)(h \xrightarrow{\rho} H)).$$

Proof:

$$
\begin{aligned}
E\#(W \text{ in } V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)) &= \#(W' \text{ in } V(\Lambda \xrightarrow{\rho} h)) \\
&= \sum_{i_j \in \{0,1\}} \#(W' \text{ in } V(i_1, \dots, i_N))\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j} \\
&= \sum_{i_j \in \{0,1\}} \#(W \text{ in } \tilde{V}(i_1, \dots, i_N))\rho^{\sum_j i_j}(1-\rho)^{N-\sum_j i_j} \\
&= E\#(W \text{ in } V(\Lambda \xrightarrow{1} H)),
\end{aligned}
$$

where W' is the word obtained from W by replacing each entry of H by h, and for $i_j \in \{0,1\}$, we define the word $\tilde{V}(i_1, \dots, i_N)$ by the word obtained from V by inserting W at the positions in which $i_j = 1$ and leaving it unchanged in the positions in which $i_j = 0$.

Lemma 3.4.16 is proved.

Lemma 3.4.17. *Let $h \notin \mathcal{A}$. Then*

$$E\#(h \text{ in } V(\Lambda \xrightarrow{\rho} h)) = (|V| - 1) \cdot \rho.$$

Proof: We have that

$$\#(h \text{ in } V(\Lambda \xrightarrow{\rho} h)) \sim \text{Bin}(|V| - 1, \rho),$$

and hence

$$E\#(h \text{ in } V(\Lambda \xrightarrow{\rho} h)) = (|V| - 1) \cdot \rho.$$

Lemma 3.4.17 is proved.

Lemma 3.4.18. *Let $h \notin \mathcal{A}$, H be a non-empty word in $\mathcal{A}$ and μ any measure on $\mathcal{A}^{\mathbb{Z}}$. Then*

$$E|V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)| = E|V(\Lambda \xrightarrow{\rho} H)|.$$

Proof: By Lemma 3.4.17 and Lemma 3.1.1, we have

$$
\begin{aligned}
& E|V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)| \\
= \;& E|V(\Lambda \xrightarrow{\rho} h)| + (|H|-1)E\#(h \;\text{ in }\; V(\Lambda \xrightarrow{\rho} h)) \\
= \;& |V| + \rho\#(\Lambda \;\text{ in }\; V) + (|H|-1)\#(\Lambda \;\text{ in }\; V)\rho \\
= \;& |V| + \rho|H|\#(\Lambda \;\text{ in }\; V) = E|V(\Lambda \xrightarrow{\rho} H)|
\end{aligned}
$$

Lemma 3.4.18 is proved.

Proof of Theorem 3.4.6: It is a simple consequence of Lemmas 3.4.11, 3.4.13, 3.4.16, 3.4.12, 3.4.15 and 3.4.18.

Theorem 3.4.6 is proved.

Chapter 4

Convexity and Fine Operators

For any two measures $\mu,\ \nu$ we denote by convex$\,(\mu,\nu)$ their convex hull , that is

$$\text{convex}(\mu,\nu) = \{k\mu + (1-k)\nu \,|\, 0 \leq k \leq 1\}. \tag{4.1}$$

Lemma 4.0.19. *Let (V_n) and (W_n) be sequences of words converging to measures μ and ν respectively. Let the following limit exist*

$$\lim_{n\to\infty} \frac{|V_n|}{|V_n| + |W_n|} = L.$$

Then the sequence $\text{concat}(V_n,\ W_n)$ *converges to the measure $L\cdot\mu + (1-L)\cdot\nu$ when $n \to \infty$.*

Proof: We clearly have $|\text{concat}(V_n,\ W_n)| = |V_n| + |W_n|$ and also

$$\begin{aligned}
\#(W \text{ in } V_n) + \#(W \text{ in } W_n) &\leq \#(W \text{ in } \text{concat}(V_n, W_n)) \\
&\leq \#(W \text{ in } V_n) + \#(W \text{ in } W_n) + |W|.
\end{aligned}$$

Hence,

$$\frac{\#(W \ \text{ in } \ V_n)}{|V_n| + |W_n|} + \frac{\#(W \ \text{ in } \ W_n)}{|V_n| + |W_n|} \leq \text{freq}(W \ \text{ in } \ \text{concat}(V_n, W_n))$$

$$\leq \quad \left(\frac{\#(W \ \text{ in } \ V_n)}{|V_n| + |W_n|} + \frac{\#(W \ \text{ in } \ W_n)}{|V_n| + |W_n|} + \frac{|W|}{|V_n| + |W_n|} \right) \cdot \frac{|V_n| + |W_n|}{|V_n| + |W_n| - |W| + 1}.$$

Therefore

$$\frac{|V_n| - |W| + 1}{|V_n| + |W_n|} \cdot \text{freq}(W \ \text{ in } \ V_n) + \frac{|W_n| - |W| + 1}{|V_n| + |W_n|} \cdot \text{freq}(W \ \text{ in } \ W_n)$$

$$\leq \text{freq}(W \ \text{ in } \ \text{concat}(V_n, W_n))$$

$$\leq \left(\frac{|V_n| - |W| + 1}{|V_n| + |W_n|} \cdot \text{freq}(W \ \text{ in } \ V_n) + \frac{|W_n| - |W| + 1}{|V_n| + |W_n|} \cdot \text{freq}(W \ \text{ in } \ W_n) \frac{|W|}{|V_n| + |W_n|} \right)$$

$$\cdot \frac{|V_n| + |W_n|}{|V_n| + |W_n| - |W| + 1}.$$

But

$$\frac{|V_n| - |W| + 1}{|V_n| + |W_n|} \cdot \text{freq}(W \ \text{ in } \ V_n) + \frac{|W_n| - |W| + 1}{|V_n| + |W_n|} \cdot \text{freq}(W \ \text{ in } \ W_n)$$

$$\to L \cdot \mu(W) + (1 - L) \cdot \nu(W) \ \text{ as } \ n \to \infty$$

and

$$\frac{|W|}{|V_n| + |W_n|} \to 0 \ \text{ as } \ n \to \infty,$$

also

$$\frac{|V_n| + |W_n|}{|V_n| + |W_n| - |W| + 1} \to 1 \ \text{ as } \ n \to \infty.$$

Thus the left and right sides of the above inequality tend to $L \cdot \mu(W) + (1 - L) \cdot \nu(W)$.

Then

$$\lim_{n \to \infty} \text{freq}(W \ \text{ in } \ \text{concat}(V_n, W_n)) = L \cdot \mu + (1 - L) \cdot \nu.$$

That is, the sequence of words $\operatorname{concat}(V_n,\ W_n)$ converges to the measure $L\mu+(1-L)\nu$.

Lemma 4.0.19 is proved.

Lemma 4.0.20. *Let (V_n) be a sequence of words such that $(V_n) \to \mu$ and (k_n) a sequence of natural (therefore positive) numbers. Then*

$$V_n^{k_n} \to \mu, \quad \text{as} \quad n \to \infty.$$

Proof: Follows immediately from the inequalities:

$$k_n \cdot \#(W \ \text{ in } \ V_n) \leq \#(W \ \text{ in } \ V_n^{k_n}) \leq k_n \cdot \#(W \ \text{ in } \ V_n) + k_n \cdot |W|$$

and from the fact that $|V_n^{k_n}| = k_n|V_n|$.

Lemma 4.0.20 is proved.

The following Theorem tells us that we can get approximations to any convex combination of measures by words. This will be very useful when we apply the substitution operators on any convex combinations of two measures, because in order to see how it behaves we will need only to check how it behaves on certain sequences of words.

Theorem 4.0.21. *Given $L \in [0,1]$ and any measures, μ and ν, there is a sequence of words (V_n) converging to μ and another sequence of words (W_n) converging to ν, such that* $\operatorname{concat}(V_n,\ W_n)$ *converges to $L \cdot \mu + (1-L) \cdot \nu$.*

Proof: Take any sequences of words (V_n) and (W_n) such that $(V_n) \to \mu$ and $(W_n) \to \nu$ as $n \to \infty$. Then we construct a sequence of pairs $(\widetilde{V}_n,\ \widetilde{W}_n)$, such that $\widetilde{V}_n \to \mu,\ \widetilde{W}_n \to \nu$ and $|\widetilde{V}_n| = |\widetilde{W}_n|$. Indeed, let us consider

$$\widetilde{V}_n = V_n^{t_n}, \quad \text{where} \quad t_n = |W_n|, \quad \text{and} \quad \widetilde{W}_n = W_n^{u_n}, \quad \text{and} \quad u_n = |V_n|.$$

Then we get $|\widetilde{V}_n| = |V_n| \cdot |W_n| = |\widetilde{W}_n|$.

Now we need to obtain a new sequence of pairs $(\widehat{V}_n,\ \widehat{W}_n)$, such that $\widehat{V}_n \to \mu$, $\widehat{W}_n \to \nu$ and

$$\mathrm{concat}(\widehat{V}_n, \widehat{W}_n) \to L \cdot \mu + (1 - L) \cdot \nu.$$

To do so, let $r \geq 0$ be given as $r = 1/L - 1$, and $r = +\infty$ if $L = 0$. Further, consider $r_n > 0$ a sequence of positive rational numbers such that $r_n \to r$. Let us write r_n in a more convenient way: $r_n = p_n/q_n$, where $p_n,\ q_n > 0$ are natural numbers. Then we take

$$\widehat{V}_n = \widetilde{V}_n^{q_n} \quad \text{and} \quad \widehat{W}_n = \widetilde{W}_n^{p_n}.$$

Noting that $|\widehat{V}_n| = q_n \cdot |\widetilde{V}_n|$, $|\widehat{W}_n| = p_n \cdot |\widetilde{W}_n|$ and $|\widetilde{V}_n| = |\widetilde{W}_n|$, thus, $\widetilde{V}_n = V_n^{t_n + q_n}$, $\widetilde{W}_n = W_n^{u_n + p_n}$, and therefore, by Lemma 4.0.20, $\widehat{V}_n \to \mu$ and $\widehat{W}_n \to \nu$. We also get that

$$\frac{|\widehat{V}_n|}{|\widehat{V}_n| + |\widehat{W}_n|} = \frac{1}{1 + p_n/q_n} \to \frac{1}{1 + r} = L,$$

and by Lemma 4.0.19, we have that

$$\mathrm{concat}(\widehat{V}_n, \widehat{W}_n) \to \frac{1}{1 + r} \cdot \mu + \frac{r}{1 + r} \cdot \nu = L\mu + (1 - L)\nu.$$

Theorem 4.0.21 is proved.

The following definition gives us a remarkable property, which all substitution operators have, as we shall prove. This property is trivially satisfied for all linear operators, but is not true in general.

Definition 4.0.22. An operator $P : \mathcal{M}_\mathcal{A} \to \mathcal{M}_\mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are alphabets, is called *fine* if:

$$\forall \mu, \nu \in \mathcal{M}_\mathcal{A} \quad \lambda \in \mathrm{convex}(\mu,\ \nu) \Rightarrow \lambda P \in \mathrm{convex}(\mu P,\ \nu P),$$

where $\text{convex}(\mu, \nu)$ was defined in (4.1).

Theorem 4.0.23. *Every substitution operator* $(G \xrightarrow{\rho} H)$ *is fine and*

$$(L \cdot \mu + (1 - L) \cdot \nu)(G \xrightarrow{\rho} H) = \widetilde{L} \cdot \mu(G \xrightarrow{\rho} H) + (1 - \widetilde{L}) \cdot \nu(G \xrightarrow{\rho} H)$$

for any measures μ, ν, *where*

$$\widetilde{L} = \frac{L \cdot \text{Ext}(\mu|(G \xrightarrow{\rho} H))}{L \cdot \text{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1 - L) \cdot \text{Ext}(\nu|(G \xrightarrow{\rho} H))}. \tag{4.2}$$

Proof: The proof will be done by first obtaining a similar result for words, then taking the limit as $n \to \infty$ and finally proving it for measures. The key tool is Theorem 4.0.21. Let $L \in (0, 1)$. Due to Theorem 4.0.21 we can take two sequences of words (V_n) and (W_n) converging to μ and ν, respectively, such that

$$\text{concat}(V_n, W_n) \to L \cdot \mu + (1 - L) \cdot \nu, \quad \text{as} \quad n \to \infty.$$

We want to show that the substitution operator $(G \xrightarrow{\rho} H)$ satisfies:

$$(L \cdot \mu + (1 - L) \cdot \nu)(G \xrightarrow{\rho} H) = \widetilde{L} \cdot \mu(G \xrightarrow{\rho} H) + (1 - \widetilde{L}) \cdot \nu(G \xrightarrow{\rho} H).$$

Now we need Corollary 3.4.8.

Let us choose any word W. Then

$$|E\#(W \text{ in } \text{concat}(V_n, W_n)(G \xrightarrow{\rho} H)) - \\ -(E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H)))| < K,$$

where $K = (\#(W \text{ in } H) + \#(W \text{ in } G) + 4 + |W|)$. Moreover, note that

$$\frac{K}{E|\text{concat}(V_n, \ W_n)(G \xrightarrow{\rho} H)|} = \frac{(\#(W \text{ in } H) + \#(W \text{ in } G) + 4 + |W|)}{E|\text{concat}(V_n, \ W_n)(G \xrightarrow{\rho} H)|} \to 0.$$

Therefore, to prove the convergence of the sequence of the words $\text{concat}(V_n,\ W_n)(G \xrightarrow{\rho} H)$, it is sufficient to look at the limit values of

$$\frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H))}{E|\text{concat}(V_n,\ W_n)(G \xrightarrow{\rho} H)|}.$$

Notice further that

$$E|\text{concat}(V_n,\ W_n)(G \xrightarrow{\rho} H)| = |V_n| + |W_n| + \rho(|H| - |G|) \cdot \#(G \text{ in } \text{concat}(V_n,\ W_n))$$

and furthermore that

$$\begin{aligned}\#(G \text{ in } V_n) + \#(G \text{ in } W_n) &\leq \#(G \text{ in } \text{concat}(V_n, W_n)) \\ &\leq \#(G \text{ in } V_n) + \#(G \text{ in } W_n) + |G|,\end{aligned}$$

Due to the analogies between several parts of our argument, we will examine in detail only some cases and omit the others since they are analogous to those studied below. For instance, if we sandwich (in terms of inequality) the quantity

$$\frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H))}{E|\text{concat}(V_n, W_n)(G \xrightarrow{\rho} H)|}$$

between two values, say a_n and b_n, such that

$$a_n \leq \frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H))}{E|\text{concat}(V_n, W_n)(G \xrightarrow{\rho} H)|} \leq b_n \tag{4.3}$$

we will prove one part of the inequality considered in equation (4.3) (i.e., we will find a quantity b_n, such that the equation 4.3 holds), since the other one is entirely analogous (i.e., the strategy to obtain a quantity a_n such that equation (4.3) holds is entirely analogous). Thus, we obtain the inequality by taking the explicit values of the denominator in the above expression:

$$\frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H))}{E|\text{concat}(V_n, W_n)(G \xrightarrow{\rho} H)|}$$

$$= \frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H))}{|V_n| + |W_n| + \rho(|H| - |G|)\#(G \text{ in } \text{concat}(V_n, W_n))}$$

$$\leq \frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H))}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \text{ in } V_n) + \#(G \text{ in } W_n))},$$

that is, if we set

$$b_n = \frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H)) + E\#(W \text{ in } W_n(G \xrightarrow{\rho} H))}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \text{ in } V_n) + \#(G \text{ in } W_n))}, \tag{4.4}$$

equation (4.3) holds. Further, as mentioned before, the other inequality can be proved analogously.

We want to check the limiting behavior of b_n. Therefore, we begin by checking the limiting behavior of:

$$\frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H))}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \text{ in } V_n) + \#(G \text{ in } W_n))},$$

the limit behavior of the other part is obtained by simply replacing each entry of V_n by W_n, and each entry of W_n by V_n in the above expression. Therefore for the second case we will just give the resulting expression. Thus, we begin with:

$$\frac{E\#(W \text{ in } V_n(G \xrightarrow{\rho} H))}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \text{ in } V_n) + \#(G \text{ in } W_n))}$$

$$= \frac{|V_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \text{ in } V_n) + \#(G \text{ in } W_n))} \times S_n \times M_n$$

$$= \frac{|V_n| + |W_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \text{ in } V_n) + \#(G \text{ in } W_n))} \times L_n \times S_n \times M_n$$

where

$$L_n = \frac{|V_n|}{(|V_n| + |W_n|)} \quad \rightarrow \quad L,$$

$$S_n = \frac{E|V_n(G \xrightarrow{\rho} H)|}{|V_n|} \quad \rightarrow \quad \mathrm{Ext}(\mu|G \xrightarrow{\rho} H)$$

$$M_n = \frac{E\#(W \ \text{ in } \ V_n(G \xrightarrow{\rho} H))}{E|V_n(G \xrightarrow{\rho} H)|} \quad \rightarrow \quad \mu(G \xrightarrow{\rho} H)(W).$$

Going on, we have

$$\frac{|V_n| + |W_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \ \text{ in } \ V_n) + \#(G \ \text{ in } \ W_n))} \times L_n \times S_n \times M_n$$

$$= \frac{L_n S_n M_n}{L_n \left[1 + \rho(|H| - |G|)\frac{\#(G \ \text{ in } \ V_n)}{|V_n|}\right] + (1 - L_n)\left[1 + \rho(|H| - |G|)\frac{\#(G \ \text{ in } \ W_n)}{|W_n|}\right]}$$

$$\xrightarrow{n\to\infty} \frac{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H))}{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \xrightarrow{\rho} H))} \mu(G \xrightarrow{\rho} H)(W)$$

By means of the same calculations, one obtains:

$$\frac{E\#(W \ \text{ in } \ W_n(G \xrightarrow{\rho} H))}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \ \text{ in } \ V_n) + \#(G \ \text{ in } \ W_n))}$$

$$\xrightarrow{n\to\infty} \frac{(1 - L) \cdot \mathrm{Ext}(\nu|(G \xrightarrow{\rho} H))}{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \xrightarrow{\rho} H))} \nu(G \xrightarrow{\rho} H)(W)$$

Therefore, we have obtained the limiting behavior of b_n in equation (4.4):

$$b_n \xrightarrow{n\to\infty} \frac{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H))}{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \xrightarrow{\rho} H))} \times \mu(G \xrightarrow{\rho} H)(W)$$

$$+ \frac{(1 - L) \cdot \mathrm{Ext}(\nu|(G \xrightarrow{\rho} H))}{L \cdot \mathrm{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1 - L) \cdot \mathrm{Ext}(\nu|(G \xrightarrow{\rho} H))} \times \nu(G \xrightarrow{\rho} H)(W).$$

Analogously, it is possible to find an a_n satisfying equation (4.3) such that it has the same limit as b_n, that is

$$a_n \overset{n\to\infty}{\longrightarrow} \frac{L\cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H))}{L\cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1-L)\cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))} \times \mu(G \overset{\rho}{\to} H)(W)$$

$$+ \quad \frac{(1-L)\cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))}{L\cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)) + (1-L)\cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H))} \times \nu(G \overset{\rho}{\to} H)(W).$$

For instance, choose

$$a_n = \frac{E\#(W \ \text{ in } \ V_n(G \overset{\rho}{\to} H)) + E\#(W \ \text{ in } \ W_n(G \overset{\rho}{\to} H))}{|V_n| + |W_n| + \rho(|H| - |G|)(\#(G \ \text{ in } \ V_n) + \#(G \ \text{ in } \ W_n) + |G|)}.$$

Finally, we conclude that

$$\mathrm{concat}(V_n, W_n)(G \overset{\rho}{\to} H)$$

$$\overset{n\to\infty}{\longrightarrow} \frac{L\cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H)}{L\cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H) + (1-L)\cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H)} \times \mu(G \overset{\rho}{\to} H)$$

$$+\frac{(1-L)\mathrm{Ext}(\nu|(G \overset{\rho}{\to} H)}{L\cdot \mathrm{Ext}(\mu|(G \overset{\rho}{\to} H) + (1-L)\cdot \mathrm{Ext}(\nu|(G \overset{\rho}{\to} H)} \times \nu(G \overset{\rho}{\to} H).$$

Thus, applying Theorem 4.0.21, since

$$\mathrm{concat}(V_n,\ W_n) \to L\cdot\mu + (1-L)\cdot\nu,$$

we have

$$(L\cdot\mu + (1-L)\cdot\nu)(G \overset{\rho}{\to} H) = \widetilde{L}\cdot\mu(G \overset{\rho}{\to} H) + (1-\widetilde{L})\cdot\nu(G \overset{\rho}{\to} H),$$

for all $L \in (0,1)$, where $\widetilde{L}$ was defined in (4.2).

Theorem 4.0.23 is proved.

Chapter 5

Continuity and Invariant Measures

We say that an operator $P : \mathcal{M}' \to \mathcal{M}'$ is *continuous* if whenever a sequence $\mu_n \in \mathcal{M}'$ tends to $\lambda \in \mathcal{M}'$, the sequence $\mu_n P$ tends to λP (the well-known sequential continuity), where $\mathcal{M}'$ is a subset of $\mathcal{M}$.

Definition 5.0.24. A measure μ is called *invariant* for an operator P if $\mu P = \mu$.

Toom [15] proved the following result:

Theorem 5.0.25. *For any non-empty compact convex $\mathcal{M}' \subset \mathcal{M}$, any continuous operator $P : \mathcal{M}' \to \mathcal{M}'$ has an invariant measure.*

We now state an important result regarding continuity of consistent operators.

Theorem 5.0.26. *Every consistent operator $P : \mathcal{M}' \to \mathcal{M}'$ is continuous.*

Proof: Let μ be a measure in $\mathcal{M}'$ and let (μ_n) be a sequence of measures in $\mathcal{M}'$ converging to μ. Then $\mu_n(W) \to \mu(W)$ as $n \to \infty$ for every word W.

Let (V_{n_k}) be a sequence of words converging to μ_n as $k \to \infty$. We claim that the sequence (V_{k_k}) converges to μ as $k \to \infty$. In fact, let $\varepsilon > 0$ be any small positive number and choose k large such that

$$|V'_{k_k}(W) - \mu_k(W)| < \varepsilon/2 \text{ and } |\mu_k(W) - \mu(W)| < \varepsilon/2,$$

then

$$|V'_{k_k}(W) - \mu(W)| \leq |V'_{k_k}(W) - \mu_k(W)| + |\mu_k(W) - \mu(W)| \leq \varepsilon.$$

Hence (V_{k_k}) converges to μ as $k \to \infty$.

Now, since P is consistent (see Definition 3.2.1), we have that $V_{n_k}P \to \mu_n P$ as $k \to \infty$ and $V_{k_k}P \to \mu P$ as $k \to \infty$. Therefore, for large k, and $\varepsilon > 0$ any fixed positive real number, we have

$$|\mu_k P(W) - (V_{kk}P)'(W)| < \varepsilon/2 \text{ and } |\mu P(W) - (V_{kk}P)'(W)| < \varepsilon/2,$$

therefore,

$$|\mu_k P(W) - \mu P(W)| \leq |\mu_k P(W) - (V_{kk}P)'(W)| + |\mu P(W) - (V_{kk}P)'(W)| \leq \varepsilon.$$

Theorem 5.0.26 is proved.

We now note that $\mathcal{M}$ is convex and compact and apply Theorem 5.0.25 to conclude the following:

Corollary 5.0.27. *Let $P : \mathcal{M}' \to \mathcal{M}'$ be a consistent operator, where $\mathcal{M}'$ is a closed and convex subset of $\mathcal{M}$. Then P has an invariant measure.*

Proof: Since $\mathcal{M}'$ is closed and $\mathcal{M}$ is compact, $\mathcal{M}'$ is also compact. Further, by Theorem 5.0.26, the operator P is continuous, then by Theorem 5.0.25 P has an invariant measure.

Corollary 5.0.27 is proved.

The next corollary applies these results to substitution operators:

Corollary 5.0.28. *Every substitution operator* $(G \overset{\rho}{\to} H)$ *(where* $\rho < 1$ *if* $H = \Lambda$*) is continuous and has an invariant measure.*

Proof: By Corollary 3.4.8, the substitution operator $(G \overset{\rho}{\to} H)$ is consistent. Therefore by Theorem 5.0.26, it is continuous. Then, by Corollary 5.0.27, there exists an invariant measure μ.

Corollary 5.0.28 is proved.

Remark 5.0.29. We note that in Toom [15], the proof of continuity of the substitution operator is different, since he proves there that the basic operators are quasi-local and therefore continuous, and further, that any superposition of continuous operators is continuous.

Chapter 6

The Substitution Process

6.1 The discrete substitution process

We now introduce a large class of stochastic processes that contains as particular cases, for instance, the process defined by Toom [14]. For this class we prove the existence of invariant measures, and also that, in general, this invariant measure is not unique, and thus, in general the process is not ergodic.

Definition 6.1.1. Let μ be a measure, and let $(G \xrightarrow{\rho} H)$ (where $\rho < 1$ if $H = \Lambda$) be a substitution operator. We define the *discrete substitution process* μ_n starting initially at μ, by

$$\mu_n(W) = \mu(G \xrightarrow{\rho} H)^n(W),$$

for every word W. Accordingly, let $P_1, \ldots, P_j$ be a finite quantity of substitution operators and ν a measure. Then we define the *generalized discrete substitution process* (ν_n) starting initially at ν, by

$$\nu_n(W) = \nu(P_1 P_2 \cdots P_j)^n(W),$$

for all words W.

It is easy to see that the process defined in Toom [14] is a special case of the generalized discrete substitution process, and further, that the substitution process itself is a special case of the generalized substitution process.

We will now apply the results of the last section to these processes. In fact, we will apply them to even more general class of processes, which we will call the *consistent processes* and define as follows:

Definition 6.1.2. Let $P : \mathcal{M}' \to \mathcal{M}'$ be a consistent operator and let μ be a measure in $\mathcal{M}'$. Then we say that (μ_n) is a *consistent process* starting at μ if

$$\mu_n(W) = \mu P^n(W)$$

for all words W.

Since the substitution operator is consistent (see Corollary 3.4.8), and superposition of several consistent operators is also consistent the generalized discrete substitution process is a consistent process.

The next result regards the existence of invariant measures for such processes.

Theorem 6.1.3. *Let $P : \mathcal{M}' \to \mathcal{M}'$ be a consistent operator, where $\mathcal{M}'$ is a convex and closed subset of $\mathcal{M}$. Then P has an invariant measure.*

Proof: Follows from a straightforward application of Corollary 5.0.27.

Theorem 6.1.3 is proved.

Remark 6.1.4. Since any generalized discrete substitution process is a special case of consistent processes, every generalized discrete substitution process has at least one invariant measure.

The following result is very useful, for instance, to check ergodicity of processes.

Theorem 6.1.5. *Let us consider a generalized discrete substitution process* (μ_n) *and let* $S \subset \mathbb{A}^{\mathbb{Z}}$ *be some subset of the* σ*-algebra* $\mathbb{A}^{\mathbb{Z}}$. *Then, if for every* $c \in S$, $\mu_n(c) < \delta$ *(respectivelly* $\mu_n(c) > \varepsilon$*), then, there exists an invariant measure* μ, *such that* $\mu(c) \leq \delta$ *(respectivelly* $\mu(c) \geq \varepsilon$*) for every* $c \in S$, *where* $\delta, \varepsilon > 0$ *are some positive constants.*

Proof: Let $\mathcal{M}'$ denote the closure in $\mathcal{M}$ of the convex hull of the measures μ_0, $\mu_1, \ldots$. Therefore, $\mathcal{M}'$ is a non-empty convex closed subset of $\mathcal{M}$. Since $\mathcal{M}$ is compact, $\mathcal{M}'$ is also compact.

We now apply Corollary 5.0.28 to note that P is continuous. Further, the continuity together with Theorem 4.0.23 yields that if $\nu \in \mathcal{M}'$, then νP also belongs to $\mathcal{M}'$. Therefore, by Theorem 5.0.25 the operator P has an invariant measure μ in $\mathcal{M}'$. Since for every n and every $c \in S$, $\mu_n(c) < \delta$ (respectivelly $\mu_n(c) > \varepsilon$), a simple calculation shows that for every $\nu \in \mathcal{M}'$ and every $c \in S$ we have $\nu(c) < \delta$ (respectivelly $\nu(c) > \varepsilon$). Therefore, the invariant measure μ satisfies $\mu(c) < \delta$ (respectivelly $\mu(c) > \varepsilon$) for every $c \in S$.

Theorem 6.1.5 is proved.

Remark 6.1.6. We conclude this Section by showing that neither the discrete substitution process and thus nor the generalized substitution process need to be ergodic in general. Let $\mathcal{A} = \{a, b\}$, let $h \notin \mathcal{A}$, and let us consider the operator $((a,b) \xrightarrow{\rho} h) : \mathcal{M}_{\mathcal{A}} \to \mathcal{M}_{\mathcal{A}'}$, where $\mathcal{A}' = \mathcal{A} \cup \{h\}$, for $0 < \rho < 1$. Then it is clear that both $\delta_{\{a\}}$ and $\delta_{\{b\}}$ are invariant for this operator and thus, are invariant for the discrete substitution processes induced by $((a,b) \xrightarrow{\rho} h)$ starting at $\delta_{\{a\}}$ and $\delta_{\{b\}}$. Therefore this process is not ergodic, which implies that, in general, the discrete substitution process is not ergodic.

The next Section contains an application of these results to the process defined in Toom [14] and, as a byproduct, we get another proof that the generalized discrete substitution process is not ergodic. In fact, the proof we will give there is more interesting since it is not entirely obvious as the example give in Remark 6.1.6.

6.2 Applications to Toom's Process

We now consider a process studied by Toom [14], which is a special case of the generalized substitution process defined in last Section. In fact, consider the alphabet $\mathcal{A} = \{\oplus, \ominus\}$, whose elements are called *plus* and *minus*, and two specific operators: *flip* denoted by Flip_β and *annihilation* denoted by Ann_α. Flip_β is a special case of the basic operator which we called *conversion*. More precisely, Flip_β is $(\ominus \xrightarrow{\beta} \oplus)$, which turns every occurrence of minus into plus with probability β independently from the fate of other components. Also, Ann_α is $((\oplus, \ominus) \xrightarrow{\alpha} \Lambda)$, which makes every entrance of the self-avoiding word $(\oplus, \ominus)$ disappear with probability $\alpha < 1$ independently from fates of the other components. We therefore consider the process (μ_n), where

$$\mu_n = \delta_\ominus(\text{Flip}_\beta \text{Ann}_\alpha)^n. \tag{6.1}$$

Note that the process (μ_n) is a special case of the generalized substitution process, and thus, by Remark 6.1.4, it has at least one invariant measure. But going further, Toom [14] proved the following result:

Theorem 6.2.1. *For all $\beta \in [0,1]$ and $\alpha \in (0,1)$ the frequency of pluses in the measure μ_n does not exceed $250 \cdot \beta/\alpha^2$ for all n.*

Therefore, we have as simple corollary of the above theorem and of the results in last

Section the following result:

Theorem 6.2.2. *For all* $\beta \in [0,1]$ *and* $\alpha \in (0,1)$ *the operator* $\mathrm{Flip}_\beta \mathrm{Ann}_\alpha$ *has an invariant measure, whose frequency of pluses does not exceed* $250 \cdot \beta/\alpha^2$ *.*

Proof: Let us use Theorem 6.1.5 with $S = \{\oplus\}$ and $\delta = 250 \cdot \beta/\alpha^2$, since by Theorem 6.2.1, $\mu_n(\oplus) < 250 \cdot \beta/\alpha^2$ for all n. Therefore, by Theorem 6.1.5 there exists an invariant measure ν, such that $\nu(\oplus) < 250 \cdot \beta/\alpha^2$.

Theorem 6.2.2 is proved. An immediate corollary is that this process is not ergodic on some conditions:

Corollary 6.2.3. *For* $\beta < \alpha^2/250$*, the process 6.1 is not ergodic.*

Proof: On one hand, the measure $\delta_\oplus$ concentrated in "all pluses" is invariant for the operator $\mathrm{Flip}_\beta\ \mathrm{Ann}_\alpha$. On the other hand, by Theorem 6.2.2 above, this operator has an invariant measure, in which the frequency of pluses does not exceed $250 \cdot \beta/\alpha^2$. Thus, with appropriate α and β this operator has at least two different invariant measures.

Corollary 6.2.3 is proved.

Notice also that this variable length process is a non-trivial example of non-ergodic process, which is a special case of the general substitution process. This also shows the usefulness of Theorem 4.0.23, since it was the main tool to prove Theorem 6.1.5.

Chapter 7

Discussion and Expectations for the Future

The notion of variable length interaction processes appeared in the literature only a dozen years ago, so the theory to which this thesis contributes is still very young. The publications in this area, according to our knowledge, may be counted on fingers and toes. (We think that a good deal of them are mentioned in our list of references.) Our approach was close to that of [13, 14, 15], but the idea to approximate a measure with a sequence of words has never been mentioned in print, according to our knowledge.

As we have mentioned above, substitution operators in general are not linear, which is a major difficulty in dealing with them. However, we have shown that they possess the property of being fine, which seems to be almost as good in our setting as linearity. Our study was motivated by a general feeling of its relevance, which we have tried to exemplify by short references to some sciences, including biology, information theory, computer science, among others, as discussed in the introduction.

Moving to expectations for the future, we have several ideas which we plan to explore. Let us shortly describe them in the order of accessibility. First, the easiest one: we propose to study the k-superposed substitution operators, which means the following: let $G_1, G_2, \ldots, G_k$ be self-avoiding words and let $H_1, \ldots, H_k$ be any words. Let also $\rho_1, \ldots, \rho_k \in [0,1]$. Then we may define the corresponding operator, which we call the k-superposed substitution operator, as a superposition of k operators treated above. In short,

$$\mu P_{\mathrm{SP}} = \mu(G_1 \xrightarrow{\rho_1} H_1) \cdots (G_k \xrightarrow{\rho_k} H_k).$$

We believe that this operator has the following extension:

$$Ext(\mu|P_{\mathrm{SP}}) = 1 + \sum_{i=1}^{k} \rho_i(|H_i| - |G_i|) \left(\mu \prod_{j=1}^{i} (G_j \xrightarrow{\rho_j} H_j) \right) (G_i), \tag{7.1}$$

where

$$\mu \prod_{j=1}^{i} (G_j \xrightarrow{\rho_j} H_j)$$

means that we are considering the superposition of the first i substitution operators. The substitution operator defined in Chapter 3 is a special case of this operator with $k = 1$. Note also that the operators defined in Chapter 3 do not commute in general, whence the operator introduced here is not symmetric in the sense that changing the order of the operators changes the resulting measure.

Now let us mention another hypothetical class of operators, which also generalizes the operator defined in Chapter 3, but in a different way. Let us say that two different words U and V overlap if there is a word W, such that $|W| < |U| + |V|$ and both U and V enter W. Consider k different self-avoiding words $G_1, \ldots, G_k$, every two of which do not overlap. Also consider k words $H_1, \ldots, H_k$ and k numbers $\rho_1, \ldots, \rho_k \in [0,1]$. We propose to define a "k-joint substitution operator" acting on words, which may be denoted by

$$(G_1 \xrightarrow{\rho_1} H_1, \ldots, G_k \xrightarrow{\rho_k} H_k),$$

Informally speaking, each entrance of G_i is substituted by H_i with probability ρ_i for all $i = 1, \ldots, k$ simultaneously. We expect to successfully define this operator and prove that it is consistent by representing it as a superposition of several basic operators in a convenient way. Also we expect to prove that its extension is

$$Ext(\mu|(G_1 \xrightarrow{\rho_1} H_1, \ldots, G_k \xrightarrow{\rho_k} H_k)) = 1 + \sum_{i=1}^{k} \rho_i(|H_i| - |G_i|)\mu(G_i). \qquad (7.2)$$

We expect this operator (unlike the previous one) to be symmetric in the sense that permutations of the list of the triples $(G_i,\ \rho_i,\ H_i)$ do not change the result.

Finally, there is also another type of substitution operators that we think is interesting. We will call it a full substitution operator, or shortly, f-substitution operator. Let G be a self-avoiding word and $H_1, \ldots, H_k$ be k words, also, let $\rho_1, \ldots, \rho_k \in [0, 1]$ be such that $\sum_{i=1}^{k} \rho_i = 1$, then, the f-substitution operator, denoted by $(G \xrightarrow{\rho_1,\ldots,\rho_k} H_1, \ldots, H_k)$ acts on a word V by replacing each entry of the word G in V by H_i with probability ρ_i. Formally speaking in this situation the word G is always replaced by some word H_i, there is no way of letting an entry of G unchanged. However, it is easy to circumvent by assuming that one of H_i is in fact G. In particular, it is easy to see that the substitution operator defined in Chapter 3 is a special case of this one if we set $(G \xrightarrow{\rho,1-\rho} H, G)$. We believe that the extension of this operator is given by

$$Ext(\mu|(G \xrightarrow{\rho_1,\ldots,\rho_k} H_1, \ldots, H_k)) = 1 + \left(\sum_{i=1}^{k} \rho_i(|H_i| - |G|)\right) \cdot \mu(G).$$

We also would like to study a larger class of operators which includes all the operators mentioned above as special cases. We want to call operators of this class *local operators.* We do not know yet how to define the property of being local for an operator acting on random words, but we have a tentative definition of local deterministic operators, which turn any word into a word.

Definition 7.0.4. Let us take some alphabet $\mathcal{A}$ and define some metric $d(\cdot,\cdot)$ on $\mathrm{dic}(\mathcal{A})$, for instance the Levenshtein distance [4, 17]. Let us call an operator

$$P : \mathrm{dic}(\mathcal{A}) \to \mathrm{dic}(\mathcal{A}),$$

local if

$$\lim_{|V|+|W|\to\infty} \frac{d(\mathrm{concat}(V,\ W)P,\ \ \mathrm{concat}(VP,\ WP))}{|V|+|W|} = 0.$$

As mentioned above, we expect to prove that deterministic versions of all the operators mentioned above, that is k-superposed substitution operators, k-joint substitution operators and the f-substitution operators are local operators in this sense. We also would like to generalize this concept of local operators for operators acting on random words.

We believe that the following generalization of one of our results is also true:

Conjecture 7.0.5. *Let $P : \Omega \to \Omega$ be an operator, where Ω is the set of random words. Denote also by P its version acting on measures, which exists since the former P is consistent. Then, if*

$$\frac{E|\mathrm{concat}(U,V)P|}{E|UP| + E|VP|} \to 1 \ \text{as}\ E|V|, E|U| \to \infty,$$

and if

$$\frac{E\#(W\ \ \mathrm{in}\ \ \mathrm{concat}(U,\ V)P)}{E\#(W\ \ \mathrm{in}\ \ UP) + E\#(W\ \ \mathrm{in}\ \ VP)} \to 1 \ \text{as}\ E|V|, E|U| \to \infty,$$

then $P : \mathcal{M} \to \mathcal{M}$ is fine.

This conjecture being true, we would expect that the local operators satisfy the above conditions and hence are fine. Further, we expect that it would not be difficult to prove that the k-superposed substitution operator, the k-joint substitution operator and the f-substitution operator satisfy the above conditions, and hence that all of them are fine, which would help us a lot.

Finally, we shall try to use our constructions to define a substitution process in continuous time. This is a tentative definition: let μ be a measure, let G be a self-avoiding word and let H be any word. Then

$$\mu_t = \lim_{n\to\infty} \mu(G \stackrel{t/n}{\to} H)^n,$$

where the limit exists as we expect. We are able prove that the sequence $(\mu(G \stackrel{t/n}{\to} H)^n)$ has at least one convergent subsequence, but it remains to prove the uniqueness of limit points. This being proved, the next definition we think would be useful is the derivative of μ_t at time zero:

$$\mu L = \lim_{t\downarrow 0} \frac{\mu_t - \mu}{t}.$$

We expect L to behave like the well-known infinitesimal generator. In particular, we expect a measure μ to be invariant for the substitution process with continuous time if and only if $\mu L = 0$.

Another issue that arises from the study of the continuous time substitution processes is that we expect to remove the need of G to be self-avoiding, since the probability of two changes in neighboring words at the same time is zero.

We believe that the theory developed in this thesis deserves attention and further development.

Index of Symbols

Index of Terms

Bibliography

[1] Janson, Svante. Private communication.

[2] Gairat, A., Malyshev, V. Menshikov, M. and Pelikh, K. (1995): Classification of Markov chains describing the evolution of random strings. Russ. Math. Surv. **50** 237–255.

[3] Gairat A. , Iasnogorodski, R. and Malyshev, V. (1996): Null recurrent string. Markov Process. Related Fields, **2**, 427–460.

[4] Levenshtein, V. I. (1966): Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady, **10** 707–710.

[5] Malyshev, V. (1997): Interacting strings of symbols. Uspekhi Mat. Nauk, **52** (1997), 59–86; English transl. Russian Math. Surveys, **52** (1997), 299–326.

[6] Malyshev, V. (1998a): Random grammars. Russ. Math. Surv. **53** 345–370.

[7] Malyshev, V. (1998b): Stochastic Evolution via Graph Grammars. Diskret. Mat., **10** (1998), 30–44; English transl. Discrete Math. Appl., **8** (1998), 247–262.

[8] Malyshev, V. (2002): Quantum evolution of words. Theoretical Computer Science **273** 263–269.

[9] Ramos, A. D. (2007): Processos de particulas com comprimento variável. Doctoral thesis, UFPE. Available at
`http://www.de.ufpe.br/~toom/ensino/doutorado/alunos/index.htm`

[10] Ramos, A. D. and Toom, A. (2008a) An error correction. Letter to the editor. Journal of Stat. Physics, vol. 131, n. 1, april 2008, pp. 167-168.

[11] Ramos, A. D. and Toom, A. (2008b): Chaos and Monte Carlo Approximations of the Flip-Annihilation process. J. Stat. Phys. **133**, 761–771.

[12] Sousa, C. S. (2007): Processos de particulas sem colições. Doctoral thesis, UFPE. Available at
`http://www.de.ufpe.br/~toom/ensino/doutorado/alunos/index.htm`

[13] Toom, A. (2002) : Particle systems with variable length. Bulletin of Brasilian Math Society, v. 33, n. 3, Nov 2002, pp. 419-425.

[14] Toom, A. (2004): Non-ergodicity in a 1-D particle process with variable length. J. Stat. Phys. **115**, 895–924.

[15] Toom, A. (2007): Every continuous operator has an invariant measure. J. Stat. Phys. **129**, 555–566.

[16] Wikipedia. "Error detection and correction".
`http://en.wikipedia.org/wiki/Error_correction`.

[17] Wikipedia. "Levenshtein distance".
`http://en.wikipedia.org/wiki/Levenshtein_distance`

[18] Wikipedia. "Molecular biology".
`http://en.wikipedia.org/wiki/Molecular_biology`.

[19] Zalizniak, Andrei Anatolyevich. "On professional and amateur linguistics". A lecture on the festival of science. October 11, 2008. (In Russian.) Available at:
`http://sclon.livejournal.com/12265.html`.