Random Processes with Variable Length

Publicações Matemáticas

Random Processes with Variable Length

A. Toom
UFPE

A. Ramos
UFPE

A. Rocha
UFPE

A. Simas
UFPE

impa

28º Colóquio Brasileiro de Matemática

Impresso no Brasil / Printed in Brazil

Capa: Noni Geiger / Sérgio R. Vaz

28º Colóquio Brasileiro de Matemática

- Cadenas de Markov y Teoría de Potencial - Johel Beltrán
- Cálculo e Estimação de Invariantes Geométricos: Uma Introdução às Geometrias Euclidiana e Afim - M. Andrade e T. Lewiner
- De Newton a Boltzmann: o Teorema de Lanford - Sérgio B. Volchan
- Extremal and Probabilistic Combinatorics - Robert Morris e Roberto Imbuzeiro Oliveira
- Fluxos Estrela - Alexander Arbieto, Bruno Santiago e Tatiana Sodero
- Geometria Aritmética em Retas e Cônicas - Rodrigo Gondim
- Hydrodynamical Methods in Last Passage Percolation Models - E. A. Cator e L. P. R. Pimentel
- Introduction to Optimal Transport: Theory and Applications - Nicola Gigli
- Introdução à Aproximação Numérica de Equações Diferenciais Parciais Via o Método de Elementos Finitos - Juan Galvis e Henrique Versieux
- Matrizes Especiais em Matemática Numérica - Licio Hernanes Bezerra
- Mecânica Quântica para Matemáticos em Formação - Bárbara Amaral, Alexandre Tavares Baraviera e Marcelo O. Terra Cunha
- Multiple Integrals and Modular Differential Equations - Hossein Movasati
- Nonlinear Equations - Gregorio Malajovich
- Partially Hyperbolic Dynamics - Federico Rodriguez Hertz, Jana Rodriguez Hertz e Raúl Ures
- Random Process with Variable Length - A. Toom, A. Ramos, A. Rocha e A. Simas
- Um Primeiro Contato com Bases de Gröbner - Marcelo Escudeiro Hernandes

ISBN: 978-85-244-331-6

Distribuição: IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
E-mail: ddic@impa.br
http://www.impa.br

INTRODUÇÃO

Na maior parte da pesquisa teórica de sistemas aleatórios com partículas interagentes, o conjunto dos sítios, também chamado de espaço, não muda no processo de interação. Elementos deste espaço, também chamados de componentes, podem estar em estados diferentes, por exemplo 0 e 1, freqüentemente interpretados como ausência vs. presença de uma partícula, e podem ir de um estado para outro, o que pode ser interpretado como mudança, ou nascimento, ou morte de uma partícula. Mas, os sítios não aparecem, nem desaparecem no processo de funtionamento. Chamemos de *operadores e processos com comprimento fixo*, os processos onde sítios não podem ser criados ou eliminados. Vários processos aleatórios bem conhecidos são deste tipo. Por exemplo, processos de contato, processos de exclusão, modelo votante, etc.

Porém, na natureza existem muitas seqüências longas, cujo comprimento pode crescer ou decrescer durante o funcionamento. Por exemplo, muitas estruturas biológicas em vários níveis, macro, celular, molecular, são longas e finas, e por esta razão podem ser aproximadas por modelos uni-dimensionais, onde os componentes podem representar células ou microorganismos, os quais podem se dividir, ou morrer, ou sofrer mutações, ou pegar infeção uma de outra. Na informática e no desenvolvimento de linguagens também encontramos seqüências longas de símbolos, cujas transformações mudam o comprimeto da seqüência. Então, nós reconhecemos a necessidade de estudar *operadores e processos com comprimento variável.*

Porém, até agora, somente um pequeno grupo de matemáticos, em todo mundo, estudam estes novos processos. Este livro resume vários anos de trabalho de matemáticos brasileiros sobre estes assuntos. Concentramos nossa atenção numa classe de processos com comprimento variável chamada de *Processos de Substituição.* Onde a própria definição desta classe não é trivial. Após a sua definição, estudamos as propriedades destes processos. Depois passamos para o estudo teórico e computacional de exemplos, os quais têm um papel importante em nossos estudos.

CONTENT

Chapter 1

Theory

1.1 Informal Introduction.

The bulk of modern studies of locally interacting particle processes is based on the assumption that the set of sites, called the space, does not change in the process of interaction. Elements of this space, 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 various areas of knowledge we deal with long sequences of components, which are subject to some local random transformations, which may change their lengths. The simplest and the most well-known of such transformations are often called "insertion" and "deletion" and are widely discussed in informatics and molecular biology (see e. g. [29, 30], where one can find more references). Importance of these ideas in linguistics was emphasized lately in [31]. In such cases we use the phrase *variable-lenght processes* and our goal is to provide a rigorous definition of Substitution Operators, which are a class of variable-length processes with infinite space and study their properties.

[26, 27, 28, 18, 19, 21] have been published by our group; besides that we found only several works on similar processes (see e.g. [3, 4, 12, 13, 14, 16]), which did much to emphasize connections of such processes with modern physics; however these works seem to contain no attempts to define variable-length processes with infinite space. Our processes have discrete time and therefore can be defined in terms of operators acting on probability measures, which we call *substitution operators* or *SO* for short.

By $\#(S)$ we denote the cardinality of any finite set S. Throughout this article $\mathcal{A}$ is a non-empty finite set called *alphabet*. 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 denote by $\mathbb{Z}$ the set of integer numbers and $\mathcal{A}^{\mathbb{Z}}$ the set of bi-infinite (that is, infinite in both directions) sequences whose terms are elements of $\mathcal{A}$. We denote by $\mathcal{M}$ the set of translation-invariant probability measures on $\mathcal{A}^{\mathbb{Z}}$, that is on the σ-algebra generated by cylinders.

A generic Substitution Operator or SO acts from $\mathcal{M}$ to $\mathcal{M}$ roughly as follows. 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 (or leaves it unchanged with a probability $1 - \rho$) independently from what happens elsewhere. This rule 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.

Before going into formal details let us present a short synopsis of the first half of our work. Our first task is to define SO. We do it in several stages. In section 2 we define how measures may be approximated by words. Then we introduce random words, which also can approximate measures. In sections 3 and 4 we define how SO act on words and random words. This allows us to introduce *extension*, that is the coefficient, by which is multiplied the length of a typical word when a SO is applied to it. We do it in section 5. Extension, in its turn, allows us to define how SO act on measures. However, we found it too complicated to define directly how an arbitrary SO acts on measures. For this reason in section 6 we present a short list of *basic* SO (including insertion and deletion mentioned above) and define how they act on measures. Then in section 7 we represent an arbitrary SO as a composition of several basic SO and use this representation to define how an arbitrary SO acts on measures. Thus SO are completely defined. Based on this theoretical preparation, we study some properties of SO. A major difficulty in dealing with SO is that they are in general non-linear unlike the bulk of random processes studied till now. However, we found another property, which sometimes is as good as linearity: in section 8 we introduce *segment-preserving* operators and prove that all our SO have this property. In addition to this, in section 9 we prove that all our SO are continuous, which allows us to prove that each of them has at least one invariant measure. Using [27, 18], we prove that a certain operator has at least two invariant measures, which contributes to the study of one-dimensional non-ergodicity.

Now about the second half of our work. Sections 1-7 present a proof of several results, the most important of which is a proof of non-ergodicity of a special variable-length process with discrete time, which we call Flip-Annihilation. Sections 9-14 are devoted to Monte Crarlo and Chaos numerical approximations to the Flip-Annihilation and another variable-length process with continuous time called Annihilation-Flip-Mitosis processes.

1.2 Formal Introduction.

Let us denote by $\mathbb{A}$ the discrete topology on $\mathcal{A}$. We consider probability measures on the σ-algebra $\mathbb{A}^{\mathbb{Z}}$ on the product space $\mathcal{A}^{\mathbb{Z}}$ endowed with the topology - product of discrete topologies on all the copies of $\mathcal{A}$. Since $\mathcal{A}$ is finite, it is compact in the discrete topology, and by Tychonoff's compact theorem, $\mathcal{A}^{\mathbb{Z}}$ also is 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 the words and it is a probability, that is normalized measure if its value on the empty word equals 1. So we may define a measure in $\mathcal{M}$ by its values on words. 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 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.

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 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, $W^0 = \Lambda$.

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 [8].

We denote by $\mathrm{freq}\,(W \;\text{ in }\; V)$ the frequency of W in V, that is 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 *relative frequency* of a word W in a word V and denote by $\mathrm{rel.freq}\,(W \;\text{ in }\; V)$ the number of positions at which W enters V divided by the total number of positions of W in V, that is the fraction

$$\mathrm{rel.freq}\,(W \;\text{ in }\; V) = \frac{\mathrm{freq}\,(W \;\text{ in }\; V)}{|V| - |W| + 1}. \tag{1.1}$$

Notice that the relative 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 relative frequency of W in V is zero by definition.

We call a *pseudo-measure* any map $\mu : \mathrm{dic}(\mathcal{A}) \to \mathbb{R}$. In particular, any measure $\mu \in \mathcal{M}$ is a pseudo-measure if it is defined by its values on words.

Definition 1.2.1. For every word $V \in \mathrm{dic}(\mathcal{A})$ we define the corresponding pseudo-measure, denoted by meas^V and defined by the rule $\mathrm{meas}^V(W) = \mathrm{rel.freq}\,(W \;\text{ in }\; V)$ for every word W.

Definition 1.2.2. We say that a sequence (V_n) of words in $\mathrm{dic}(\mathcal{A})$ converges to a measure $\mu \in \mathcal{M}$ if for every word $W \in \mathrm{dic}(\mathcal{A})$ the relative frequency of W in V_n tends to $\mu(W)$ as $n \to \infty$, that is, if $\mathrm{meas}^{V_n}(W)$ tends to $\mu(W)$ as $n \to \infty$.

Remark 1.2.3. Notice that since the relative frequencies of all W in a given V are zeros for all W longer than V, the convergence in the definition 1.2.2 is possible only if the length of V_n tends to ∞ as $n \to \infty$.

Definition 1.2.4. 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 $W \in \mathrm{dic}(\mathcal{A})$,

$$|W| \le r \Rightarrow |\mathrm{rel.freq}\,(W \;\text{ in }\; V) - \mu(W)| \le \varepsilon.$$

Lemma 1.2.5. *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

$$\left.\begin{array}{c} \forall\ \varepsilon > 0\ \ \forall\ r \in N \quad \exists\ n_0\ \ \forall\ n \geq n_0,\ \forall\ W \in \mathrm{dic}(\mathcal{A}) : \\ |W| \leq r\ \Rightarrow\ |\mathrm{rel.freq}\,(W\ \text{ in }\ V_n) - \mu(W)| \leq \varepsilon. \end{array}\right\} \tag{1.2}$$

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

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

that is

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

Taking $\varepsilon' = \varepsilon$ and n_0 equal to the maximum of n_W over all those non-empty W, whose length does not exceed r, we obtain (1.2).

In the other direction the proof is straightforward. *Lemma 1.2.5 is proved.*

Theorem 1.2.6. *For any $\mu \in \mathcal{M}$, any $\varepsilon > 0$ and any natural r there is a word which $(\varepsilon,\ r)$-approximates μ.*

Proof: If $\#(\mathcal{A}) = 1$, the theorem is trivial. So let $\#(\mathcal{A}) > 1$. Let us introduce parameters

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

Hence

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

Notice that there are N words in $\mathrm{dic}(\mathcal{A})$, whose length equals s, and we denote them by $U_1,\ U_2, \ldots, U_N$.

Furthermore, for any position k of W in U_i, that is, for any $k \in [1, s - |W| + 1]$, we define

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

Then

$$\sum_{k=1}^{s-|W|+1} \operatorname{freq}(W \ \text{ in } \ U_i)_k = \operatorname{freq}(W \ \text{ in } \ U_i) \tag{1.3}$$

and

$$\sum_{i=1}^{N} (\operatorname{freq}(W \ \text{ in } \ U_i)_k \cdot \mu(U_i)) = \mu(W).$$

Summing this over k yields

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

Hence

$$\sum_{i=1}^{N} \left(\mu(U_i) \cdot \sum_{k=1}^{s-|W|+1} \operatorname{freq}(W \ \text{ in } \ U_i)_k \right) = (s - |W| + 1) \cdot \mu(W). \tag{1.4}$$

Replacing (1.3) by (1.4) gives

$$\sum_{i=1}^{N} (\operatorname{freq}(W \ \text{ in } \ U_i) \cdot \mu(U_i)) = (s - |W| + 1) \cdot \mu(W). \tag{1.5}$$

Further, for every $i = 1, \dots, N$ we denote

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

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

$$Q \cdot \mu(U_i) - 1 < p_i \leq Q \cdot \mu(U_i). \tag{1.6}$$

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

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

Let us check that the word V has the desired property, namely (ε, r)-approximates the measure μ. Since $\mu(U_1)+\dots+\mu(U_N) = 1$, summing (1.6) over $i = 1, \dots, N$ gives

$$Q - N < \sum_{i=1}^{N} p_i \leq Q. \tag{1.7}$$

Let us estimate the relative frequency of W in V. First from below: In the present case the numerator of that fraction is

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

and the denominator is

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

Hence, using equation (1.5), we obtain

$$\begin{aligned}
&\mathrm{rel.freq}\,(W \ \text{ in } \ V) \geq \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} (\mathrm{freq}\,(W \ \text{ in } \ U_i) \cdot p_i) \geq \\
&\frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} (\mathrm{freq}\,(W \ \text{ in } \ U_i) \cdot (Q \cdot \mu(U_i) - 1)) = \\
&\frac{1}{s} \cdot \sum_{i=1}^{N} (\mathrm{freq}\,(W \ \text{ in } \ U_i) \cdot \mu(U_i)) - \frac{1}{s \cdot Q} \cdot \sum_{i=1}^{N} \mathrm{freq}\,(W \ \text{ in } \ U_i) \geq \\
&\frac{s - |W| + 1}{s} \cdot \mu(W) - \frac{s \cdot N}{s \cdot Q} \geq \left(1 - \frac{r}{s}\right) \cdot \mu(W) - \frac{N}{Q} \geq \\
&\left(1 - \frac{\varepsilon}{4}\right) \cdot \mu(W) - \frac{\varepsilon}{4} \geq \mu(W) - \varepsilon. \qquad (1.8)
\end{aligned}$$

Now let us estimate the relative frequency of W in V from above. The numerator of the fraction (1.1) is not greater than

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

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 $\#(\mathcal{A}) \geq 2, \quad r \geq 1$ and we may assume that $\varepsilon \leq 1$,

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

Therefore

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

Thus, using (1.6), (1.7) and (1.8), we get

$$\text{rel.freq}\,(W \quad \text{in} \quad V) \leq$$

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

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

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

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

Theorem 1.2.6 is proved.

Corollary 1.2.7. *For any $\mu \in \mathcal{M}$ there is a sequence of words which converges to it.*

Proof: Due to the previous theorem 1.2.6, 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 1.2.7 is proved.*

Remark 1.2.8. One of our referees noticed that Corollary 1.2.7 should have been published somewhere, even in a more general form, but we found no reference and present our own proof.

We define a *random word* X in an alphabet $\mathcal{A}$ as a random variable on $\mathrm{dic}(\mathcal{A})$ which is concentrated on a finite subset of $\mathrm{dic}(\mathcal{A})$. A random word is determined by its components $P(X = V)$, that is probabilities that $X = V$, whose sum is 1, the set $\{V : P(X = V) > 0\}$ being finite. We denote by Ω the set of random words in the alphabet $\mathcal{A}$.

Definition 1.2.9. We define the mean length of any random word X as

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

Definition 1.2.10. We define the mean frequency of a word W in a random word X as

$$E[\mathrm{freq}\,(W \ \text{ in } \ X)] = \sum_{V \in \mathrm{dic}(\mathcal{A})} P(X = V) \cdot \mathrm{freq}\,(W \ \text{ in } \ V).$$

Definition 1.2.11. We define the mean relative frequency of a word W in a random word X as

$$\mathrm{rel.freq}_E(W \ \text{ in } \ X) = \frac{E[\mathrm{freq}\,(W \ \text{ in } \ X)]}{E|X| - |W| + 1}. \tag{1.9}$$

For any random word X we define the corresponding pseudo-measure meas^X by the rule

$$\mathrm{meas}^X(W) = \mathrm{rel.freq}_E(W \ \text{ in } \ X) \ \text{ for every word } \ W.$$

Definition 1.2.12. We say that a sequence (X_n) of random words $X_1, \ X_2, \ X_3, \cdots \in \Omega$ converges to a measure $\mu \in \mathcal{M}$ if for every word $W \in \mathrm{dic}(\mathcal{A})$ the mean relative frequency of W in X_n tends to $\mu(W)$ as $n \to \infty$, that is if $\mathrm{meas}^{X_n}(W)$ tends to $\mu(W)$ as $n \to \infty$.

Definition 1.2.13. 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 $W \in \mathrm{dic}(\mathcal{A})$,

$$|W| \le r \Rightarrow |\mathrm{rel.freq}_E(W \ \text{ in } \ X) - \mu(W)| \le \varepsilon.$$

Theorem 1.2.14. *For any $\mu \in \mathcal{M}$, any $\varepsilon > 0$ and any natural r there is a random word which $(\varepsilon, \ r)$-approximates μ.*

Proof: It could be obtained as a corollary of theorem 1.2.6 by considering the random word as a distribution concentrated on a single word. One could also adapt the proof of theorem 1.2.6 by considering a random word X with $P(X = U_i) = p_i$, with p_i and U_i being the same as in the proof of theorem 1.2.6. *Theorem 1.2.14 is proved.*

Corollary 1.2.15. *For any $\mu \in \mathcal{M}$ there is a sequence of random words which converges to it.*

Proof: Analogous to the proof of corollary 1.2.7. *Corollary 1.2.15 is proved.*

1.3 SO Act on Words

A generic Substitution Operator or SO for short 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 $(G \stackrel{\rho}{\to} H)$. 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 articles in this area, we write operators on the right side of objects (words, measures) on which they act.

Our goal in this section is to define a general SO, which is denoted by $(G \stackrel{\rho}{\to} H)$, acting on words. If G and H are empty, the operator $(G \stackrel{\rho}{\to} H)$ changes nothing. Now let G or H be non-empty. Let us define how $(G \stackrel{\rho}{\to} H)$ acts on an arbitrary word V. Let us denote $N = \mathrm{freq}\,(G \;\text{ in }\; V)$, 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 may, therefore, define the SO $(G \stackrel{\rho}{\to} H)$ 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

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

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 $P(X = V_j)$. We define $X(G \xrightarrow{\rho} H)$ as the random word which equals the words $V_j(i_1, \ldots, i_N)$ with probabilities

$$\mathbb{P}(X = V_j) \cdot \rho^{\sum_j i_j} \cdot (1-\rho)^{N - \sum_j i_j}.$$

Lemma 1.3.1. *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*

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

Proof: Let us evaluate the mean length (definition 1.2.9) 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 $N = \text{freq}\,(G \ \text{ in } \ V)$. Therefore

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

which is equal to

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

Lemma 1.3.1 is proved.

Lemma 1.3.2. *For any random word X and a number $\rho \in [0, 1]$ we can express the mean length of the random word $X(G \xrightarrow{\rho} H)$ in the following simple way*

$$E|X(G \xrightarrow{\rho} H)| = E|X| + \rho \cdot (|H| - |G|) \cdot E[\text{freq}\,(G \ \text{ in } \ X)].$$

Proof. Suppose that the possible values of X are the words $V_1, \ldots, V_n$. Then we note that

$$E|X(G \xrightarrow{\rho} H)| = \sum_{j=1}^{n} E|V_j(G \xrightarrow{\rho} H)| P(X = V_j).$$

Now we use lemma 1.3.1 to obtain

$$\begin{aligned} & E|X(G \xrightarrow{\rho} H)| \\ & = \sum_{j=1}^{n} [|V_j| + \rho \cdot (|H| - |G|) \text{freq } (G \text{ in } V_j)] P(X = V_j) \\ & = E|X| + \rho \cdot (|H| - |G|)\ E[\text{freq}\,(G \text{ in } X)]. \end{aligned}$$

Lemma 1.3.2 is proved.

1.4 SO Act on Random Words

Remember our notations: $\mathcal{A}$ is an alphabet, Ω is the set of random words on $\mathcal{A}$, $\mathcal{M}$ is the set of uniform probability measures on dic$(\mathcal{A})$.

Proposition 1.4.1. *Let X_n be a sequence of random words. If X_n converges to a pseudo-measure μ, then, μ is, in fact, a measure.*

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

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

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

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

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

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

Therefore $0 \leq \mu(W) \leq 1$. We still have to show that

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

To do 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 in $\mathrm{dic}(\mathcal{A})$. 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 c, at the right side of W, such that (W,c) still enters V 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 V. Second, if W enters V at a non-last position, then (W,c) enters V for some c, as explained before. Therefore

$$0 \;\leq\; \mathrm{freq}\,(W \;\text{ in }\; V) - \sum_a \mathrm{freq}\,((W,a) \;\text{ in }\; V) \;\leq\; 1$$

for any word V. Then, multiplying the above expression by $P(X_n = V)$, we get

$$\begin{aligned}
&0 \leq \\
&P(X_n{=}V)\cdot \mathrm{freq}\,(W \;\text{ in }\; V) - P(X_n{=}V)\cdot \sum_a \mathrm{freq}\,((W,a) \;\text{ in }\; V) \\
&\leq P(X_n = V).
\end{aligned}$$

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

$$0 \leq E[\mathrm{freq}\,(W \;\text{ in }\; X_n)] - \sum_a E[\mathrm{freq}\,((W,a) \;\text{ in }\; X_n)] \leq 1.$$

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

$$\begin{aligned}
0 \;&\leq\; \mathrm{rel.freq}_E(W \;\text{ in }\; X_n) - \sum_a \frac{E[\mathrm{freq}\,((W,a) \;\text{ in }\; X_n)]}{E|X_n| - |W| + 1} \\
&\leq\; \frac{1}{E|X_n| - |W| + 1},
\end{aligned}$$

which yields

$$
\begin{aligned}
0 \;\le\; & \operatorname{rel.freq}_E(W \;\text{ in }\; X_n) \\
& - \sum_a \operatorname{rel.freq}_E((W,a) \;\text{ in }\; X_n) \times \frac{E|X_n| - |W|}{E|X_n| - |W| + 1} \\
\le\; & \frac{1}{E|X_n| - |W| + 1},
\end{aligned}
$$

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

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

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

$$0 \le \lim_{n\to\infty} \operatorname{rel.freq}_E(W \;\text{ in }\; X_n) - \sum_a \lim_{n\to\infty} \operatorname{rel.freq}_E((W,a) \;\text{ in }\; X_n) \le 0$$

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. *Proposition 1.4.1 is proved.*

Definition 1.4.2. We say that a map $P : \Omega \to \Omega$ is *consistent* if the following condition holds: 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 $X_n \to \mu$.

Definition 1.4.3. Given any consistent map $P : \Omega \to \Omega$ and any $\mu \in \mathcal{M}$, we define μP , that is the result of application of P to μ, as the measure (see Proposition 1.4.1, note also that it is unique according to definition 1.4.2), to which $(X_n P)$ converges for all $(X_n) \to \mu$, and we call it *limit of consistent operators.*

Lemma 1.4.4. *Let $P_1,\ P_2 : \Omega \to \Omega$ be consistent operators. Then their composition 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 1.4.3), hence $Q_n\, P_2$ tends to $\mu\, P_1\, P_2$. *Lemma 1.4.4 is proved.*

1.5 Extension

Definition 1.5.1. 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 word approximating μ.

Lemma 1.5.2. *Suppose that $P_1,\ P_2 : \Omega \to \Omega$ have extensions for all measures and P_1 is consistent. Then their composition $P_1 P_2$ 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 1.4.3 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, definition 1.5.1 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 composition P_1P_2 exists and equals

$$\mathrm{Ext}(\mu|P_1P_2) = \mathrm{Ext}(\mu P_1|P_2) \times \mathrm{Ext}(\mu|P_1).$$

Lemma 1.5.2 is proved.

Now let us show that every measure in $\mathcal{M}$ has an extension under every SO and provide an explicit expression for it.

Proposition 1.5.3. *If $(G \xrightarrow{\rho} H)$ is a SO 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 1.3.2 that

$$E|X_n(G \xrightarrow{\rho} H)| = E|X_n| + \rho \cdot (|H| - |G|) \cdot E[\mathrm{freq}\,(G \ \text{ in } \ X_n)].$$

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

$$\frac{E|X_n(G \xrightarrow{\rho} H)|}{E|X_n|} =$$

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

Therefore

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

Proposition 1.5.3 is proved.

1.6 Basic SO Act on Measures

Given a measure μ and a triple $(G,\ \rho,\ H)$, a generic SO acting on measures is also denoted by $(G \xrightarrow{\rho} H)$, where G and H are words, G is self-avoiding and $\rho \in [0,\ 1]$. 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.

Now we want to define a general SO acting on measures. However, it is too difficult to do it in a straightforward way. Instead, we shall introduce several simple operators acting on random words, prove their consistency and represent a general SO acting on random words as a composition of those operators. Recall the definition of consistent operators in definition 1.4.2. 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.

Let us define several small classes of operators acting on random words, 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 1.4.3).

Basic operator 1: Conversion $(g \xrightarrow{\rho} h)$ is the only linear operator in our list. For any two different letters $g,\ h \in \mathcal{A}$, we define *conversion from g to h* as a map from Ω to Ω. The conversion operator changes each occurrence of the letter g into the letter h with probability $\rho \in [0,1]$ or does not change it with probability $1-\rho$ independently of the states of the other occurrences. In various sciences a similar transformation is often called *substitution*.

Lemma 1.6.1. *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 for all words W:

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

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

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

$$\lim_{n\to\infty} \frac{E[\operatorname{freq}(W \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|}.$$

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

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

whence

$$E[\operatorname{freq}(g \ \text{ in } \ V_n(g \xrightarrow{\rho} h))] = (1-\rho) \operatorname{freq}(g \ \text{ in } \ V_n).$$

This yields

$$\lim_{n\to\infty} \frac{E[\operatorname{freq}(g \ \text{ in } \ V_n(g \xrightarrow{\rho} h))]}{|V_n|} =$$

$$(1-\rho) \lim_{n\to\infty} \frac{\operatorname{freq}(g \ \text{ in } \ V_n)}{|V_n|} = (1-\rho)\cdot\mu(g).$$

After similar calculations we obtain

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

It is easy to see that

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

where $K \sim \operatorname{Bin}(m, \rho)$ represents the number of copies of the letter g that turned into h, and $m = \operatorname{freq}(g \ \text{ in } \ V_n)$. Therefore

$$E[\operatorname{freq}(h \ \text{ in } \ V_n(g \xrightarrow{\rho} h))] = \operatorname{freq}(h \ \text{ in } \ V_n) + \rho\cdot\operatorname{freq}(g \ \text{ in } \ V_n),$$

whence

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

$$= \lim_{n\to\infty} \frac{\text{freq}\,(h \ \text{ in } \ V_n)}{|V_n|} + \rho \lim_{n\to\infty} \frac{\text{freq}\,(g \ \text{ in } \ V_n)}{|V_n|}$$

$$= \mu(h) + \rho \cdot \mu(g).$$

For any letter e different from g and h

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

Thus we define how this operator acts on μ:

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

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

and

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

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

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

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

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

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

$$\lim_{n\to\infty} \frac{E[\text{freq}\,(W \quad \text{in} \quad V_n(g \xrightarrow{\rho} h))]}{|V_n|} =$$

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

where

$$F(a|b) = \begin{cases} 1-\rho & \text{if} \ \ b = g \ \ \text{and} \ \ a = g, \\ \rho & \text{if} \ \ b = g \ \ \text{and} \ \ a = h, \\ 0 & \text{if} \ \ b = g \ \ \text{and} \ \ a \ \ \text{is neither} \ \ g \ \ \text{nor} \ \ h, \\ 1 & \text{if} \ \ b \neq g \ \ \text{and} \ \ a = b, \\ 0 & \text{if} \ \ b \neq g \ \ \text{and} \ \ a \neq b. \end{cases}$$

Lemma 1.6.1 is proved.

Now we can use consistency of this operator to define the conversion operator acting on any measure μ applied to 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 $(\Lambda \xrightarrow{\rho} h)$. Insertion of a letter $h \notin \mathcal{A}$ into a random word 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 used in molecular biology and computer science with a similar meaning [30].

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

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

Lemma 1.6.2. *The basic operator insertion is consistent.*

Proof: Let (V_n) be a sequence of words converging to some measure μ. We need to prove the following equation for any word W:

$$\lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \quad \mathrm{in} \quad V_n(\Lambda \xrightarrow{\rho} h))]}{E|V_n(\Lambda \xrightarrow{\rho} h)| - |W| + 1} =$$

$$\frac{1}{\mathrm{Ext}(\mu|(\Lambda \xrightarrow{\rho} h))} \quad \times \lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \quad \mathrm{in} \quad V_n(\Lambda \xrightarrow{\rho} h))]}{|V_n|}. \quad (1.10)$$

First let us prove that the limits in the left and right sides of (1.10) exist.

Now, let a word W be in the alphabet $\mathcal{A}' = \mathcal{A} \cup \{h\}$. If W contains two consecutive appearances of h, then

$$E[\mathrm{freq}\,(W \quad \mathrm{in} \quad V_n(\Lambda \xrightarrow{\rho} h))] = 0,$$

otherwise:

$$E[\mathrm{freq}\,(W \quad \mathrm{in} \quad V_n(\Lambda \xrightarrow{\rho} h))] =$$

$$\sum_{i_j \in \{0,1\}} \mathrm{freq}\,(W \quad \mathrm{in} \quad V_n(i_1, \ldots, i_{|V_n|+1}))$$

$$\times\ \rho^{\sum_j i_j} \times (1-\rho)^{|V_n|+1-\sum_j i_j}, \quad (1.11)$$

where $V_n(i_1, \ldots, i_{|V_n|+1})$ is the word obtained from V_n after inserting the letter h in those positions where $i_j = 1$. Further, let M be the number of pairs of consecutive letters in W, both of which are not h. It is clear that if $\sum_j i_j < \mathrm{freq}\,(h \quad \mathrm{in} \quad W)$ or if $|V_n| + 1 - \sum_j i_j < M$, then $\mathrm{freq}\,(W \quad \mathrm{in} \quad V_n(i_1, \ldots, i_{|V_n|+1})) = 0$. Let also

$$R = \{x \in \mathbb{N}; \mathrm{freq}\,(h \quad \mathrm{in} \quad W) \le x \le |V_n| + 1 - M\},$$

W' being the word obtained from W by deleting all the entrances of the letter h, and let $f_h(W) = \mathrm{freq}\,(h \quad \mathrm{in} \quad W)$. Note also that

$$\sum_{\substack{i_j \in \{0,1\} \\ \sum_j i_j = f_h(W)}} \mathrm{freq}\,(W \quad \mathrm{in} \quad V_n(i_1, \ldots, i_{|V_n|+1})) \quad (1.12)$$

$$= \mathrm{freq}\,(W' \quad \mathrm{in} \quad V_n),$$

and, more generally, for $0 \leq k \leq |V_n| + 1 - M - f_h(W)$, we have

$$\sum_{\substack{i_j \in \{0,1\} \\ \sum_j i_j = f_h(W) + k}} \mathrm{freq}\,(W \;\text{ in }\; V_n(i_1, \ldots, i_{|V_n|+1})) =$$

$$\binom{|V_n| + 1 - M - f_h(W)}{k} \cdot \mathrm{freq}\,(W' \;\text{ in }\; V_n). \tag{1.13}$$

Now we can simplify the expression (1.11) to obtain

$$\begin{aligned}
E[\mathrm{freq}\,(W \;\text{ in }\; V_n(\Lambda \xrightarrow{\rho} h))] &= \rho^{f_h(W)} \cdot (1-\rho)^M \times \\
&\sum_{\substack{i_j \in \{0,1\} \\ \sum_j i_j \in R}} \mathrm{freq}\,(W \;\text{ in }\; V_n(i_1, \ldots, i_{|V_n|+1})) \\
&\times \rho^{\sum_j i_j - f_h(W)} \times (1-\rho)^{|V_n|+1-\sum_j i_j - M} \\
&= \rho^{f_h(W)} \cdot (1-\rho)^M \cdot \mathrm{freq}\,(W' \;\text{ in }\; V_n) \times \\
&\sum_{k=0}^{|V_n|+1-M-f_h(W)} \binom{|V_n| + 1 - M - f_h(W)}{k} \times \\
&\rho^k \times (1-\rho)^{|V_n|+1-M-k-f_h(W)} \\
&\mathrm{freq}\,(W' \;\text{ in }\; V_n) \times \rho^{\mathrm{freq}\,(h \text{ in } W)} \times (1-\rho)^M.
\end{aligned}$$

Hence it is easy to conclude that both limits in (1.10) exist. Now the fact that these limits are equal comes from the definition and existence of extension of this operator. *Lemma 1.6.2 is proved.*

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

$$\begin{aligned}
\mu(\Lambda \xrightarrow{\rho} h)(W) &= \frac{1}{\mathrm{Ext}(\mu|(\Lambda \xrightarrow{\rho} h))} \; \mu(W') \; \rho^{\mathrm{freq}\,(h \text{ in } W)} \; (1-\rho)^M \\
&= \frac{1}{1+\rho} \cdot \mu(W') \; \rho^{\mathrm{freq}\,(h \text{ in } W)} \cdot (1-\rho)^M.
\end{aligned}$$

Basic operator 3: Deletion $(g \xrightarrow{\rho} \Lambda)$. Deletion of a letter $g \in \mathcal{A}$ with some probability $\rho \in [0, 1)$ in a random word means that each occurrence of g disappears with probability ρ or remains unchanged with probability $1 - \rho$ independently from the other occurrences. This term is also used in sciences with a similar meaning [30]. The extension of any measure μ under this operator is

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

Lemma 1.6.3. *The basic operator deletion is consistent.*

Proof. Let (V_n) be a sequence of words converging to a measure μ. We need to prove that for any word W

$$\lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))]}{E|V_n(g \xrightarrow{\rho} \Lambda)| - |W| + 1} =$$

$$\frac{1}{\mathrm{Ext}(\mu|(g \xrightarrow{\rho} \Lambda))} \times \lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))]}{|V_n|}, \qquad (1.14)$$

but first we need to prove that both limits in (1.14) exist. Let $W = (a_0, \ldots, a_m)$ be any word with $|W| = m$, and $N_n = \mathrm{freq}\,(g \ \text{ in } \ V_n)$. Then from definition 1.2.10

$$E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))] =$$

$$\sum_{k; j_1, \ldots, j_k} \mathrm{freq}\,(W \ \text{ in } \ V_n(k; j_1, \ldots, j_k)) \cdot \rho^k \cdot (1-\rho)^{N_n - k},$$

where $V(k; j_1, \ldots, j_k)$ is the word obtained from V by deletion of k letters g from positions $j_1, \ldots, j_k$.

Let $M = \mathrm{freq}\,(g \ \text{ in } \ W)$ and note that

$$\text{if } \ M > N_n, \ \text{ then } \ E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))] = 0.$$

Fix some k in $\{0, \ldots, N_n - M\}$, and note that the equations (1.12) and (1.13), written in the context of the deletion operator, provide the following equation:

$$\sum_{j_1, \dots, j_k} \operatorname{freq}(W \text{ in } V_n(k; j_1, \dots, j_k))$$

$$= \sum_{n_1+\cdots+n_{m+1} \le k} \binom{N_n - M - (n_1 + \cdots + n_{m+1})}{k - (n_1 + \cdots + n_{m+1})}$$

$$\times \quad \operatorname{freq}((g^{n_1}, a_1, g^{n_2}, \dots, g^{n_m}, a_m, g^{n_{m+1}}) \text{ in } V_n).$$

Multiplying the above expression by $\rho^k \cdot (1-\rho)^{N_n - k}$, summing over k and inverting the order of summation on the right side of the equation yields

$$\sum_{k=0}^{N_n - M} \sum_{j_1, \dots, j_k} \operatorname{freq}(W \text{ in } V_n(k; j_1, \dots, j_k)) \times \rho^k \times (1-\rho)^{N_n - k}$$

$$= \sum_{n_1+\cdots+n_{m+1} \le N_n - M} \quad \sum_{k=n_1+\cdots+n_{m+1}}^{N_n - M}$$

$$\operatorname{freq}((g^{n_1}, a_1, g^{n_2}, \dots, g^{n_m}, a_m, g^{n_{m+1}}) \text{ in } V_n)$$

$$\times \quad \binom{N_n - M - (n_1 + \cdots + n_{m+1})}{(n_1 + \cdots + n_{m+1}) - k} \times \rho^k \times (1-\rho)^{N_n - k}$$

$$= \sum_{n_1+\cdots+n_{m+1} \le N_n - M} \quad \sum_{j=0}^{N_n - M - (n_1+\cdots+n_{m+1})}$$

$$\operatorname{freq}((g^{n_1}, a_1, g^{n_2}, \dots, g^{n_m}, a_m, g^{n_{m+1}}) \text{ in } V_n)$$

$$\times \binom{N_n - M - (n_1 + \cdots + n_{m+1})}{j}$$

$$\times \rho^{n_1+\cdots+n_{m+1}+j} \times (1-\rho)^{N_n - j - (n_1+\cdots+n_{m+1})}.$$

Then we can use the previous equation to obtain

$$E[\operatorname{freq}(W \text{ in } V_n(g \xrightarrow{\rho} \Lambda))] =$$

$$(1-\rho)^M \sum_{k=0}^{N_n-M} \sum_{j_1,\dots,j_k} \mathrm{freq}\,(W \ \text{ in } \ V_n(k;j_1,\dots,j_k))$$
$$\times \rho^k \times (1-\rho)^{N_n-k-M}$$
$$= \sum_{n_1+\cdots+n_{m+1}\le N_n-M} \mathrm{freq}\,((g^{n_1},a_1,g^{n_2},\dots,g^{n_m},a_m,g^{n_{m+1}}) \ \text{ in } \ V_n)$$
$$\times \quad \rho^{n_1+\cdots+n_m+n_{m+1}} \times (1-\rho)^M$$
$$\times \sum_{j=0}^{N_n-M-(n_1+\cdots+n_{m+1})} \binom{N_n-M-(n_1+\cdots+n_{m+1})}{j}$$
$$\times \quad \rho^j \times (1-\rho)^{N_n-M-j-(n_1+\cdots+n_{m+1})}$$
$$= \sum_{n_1+\cdots+n_{m+1}\le N_n-M} \mathrm{freq}\,((g^{n_1},a_1,g^{n_2},\dots,g^{n_m},a_m,g^{n_{m+1}}) \ \text{ in } \ V_n)$$
$$\times \quad \rho^{n_1+\cdots+n_m+n_{m+1}} \times (1-\rho)^M.$$

Therefore

$$\frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))]}{|V_n|} = \sum_{n_1+\cdots+n_{m+1}\le N_n-M}$$
$$\frac{\mathrm{freq}\,((g^{n_1},a_1,g^{n_2},\dots,g^{n_m},a_m,g^{n_{m+1}}) \ \text{ in } \ V_n)}{|V_n|}$$
$$\times \rho^{n_1+\cdots+n_{m+1}} \times (1-\rho)^M$$
$$\sum_{n_1,\dots,n_{m+1}=0}^{\infty} \mathrm{I}_{\{n_1+\cdots+n_{m+1}\le N_n-M\}} \times$$
$$\frac{\mathrm{freq}\,((g^{n_1},a_1,g^{n_2},\dots,a_m,g^{n_{m+1}}) \ \text{ in } \ V_n)}{|V_n|},$$
$$\times \rho^{n_1+\cdots+n_{m+1}} \times (1-\rho)^M,$$

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

$$\mathrm{I}_{\{n_1+\cdots+n_{m+1}\le N_n\}} \overset{n\to\infty}{\longrightarrow} 1$$

(i.e. it converges to the function, which is identically equal to 1). Also

$$\left| \frac{\mathrm{freq}\,((g^{n_1}, a_1, g^{n_2}, \ldots, g^{n_m}, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n)}{|V_n|} \right.$$

$$\times \quad \left. \mathrm{I}_{\{n_1+\cdots+n_{m+1} \le N_n - M\}} \right| \le 1.$$

Since

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

we may conclude from the dominated convergence theorem that

$$\lim_{n\to\infty} \frac{E[\mathrm{freq}\,(W \ \text{ in } \ V_n(g \xrightarrow{\rho} \Lambda))]}{|V_n|}$$

$$\lim_{n\to\infty} \sum_{n_1,\ldots,n_{m+1}=0}^{\infty} \mathrm{I}_{\{n_1+\cdots+n_{m+1} \le N_n\}}$$

$$\frac{\mathrm{freq}\,((g^{n_1}, a_1, \ldots, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n) \cdot \rho^{n_1+\cdots+n_{m+1}} \cdot (1-\rho)^M}{|V_n|}$$

$$\sum_{n_1,\ldots,n_{m+1}=0}^{\infty} \lim_{n\to\infty} \mathrm{I}_{\{n_1+\cdots+n_{m+1} \le N_n\}}$$

$$\frac{\mathrm{freq}\,((g^{n_1}, a_1, \ldots, a_m, g^{n_{m+1}}) \ \text{ in } \ V_n) \cdot \rho^{n_1+\cdots+n_{m+1}} \cdot (1-\rho)^M}{|V_n|}$$

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

Hence it is easy to conclude that both limits in (1.14) exist. In addition, we notice that the equality in (1.14) follows from the definition of extension. *Lemma 1.6.3 is proved.*

Now let us use consistency of this operator to define how the operator $(g \xrightarrow{\rho} \Lambda)$ acts on an arbitrary measure μ:

$$\mu(g \xrightarrow{\rho} \Lambda)(W) \ = \ \frac{1}{\mathrm{Ext}(\mu|(g \xrightarrow{\rho} \Lambda))} \times$$

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

$$= \frac{1}{1-\rho\cdot\mu(g)} \times$$

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

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

Basic operator 4: Compression $(G \xrightarrow{1} h)$. Given a non-empty self-avoiding word G in an alphabet $\mathcal{A}$ and a letter $h \notin \mathcal{A}$, *compression from* G *to* h is the following map from $\Omega(\mathcal{A})$ to $\Omega(\mathcal{A}')$, where $\mathcal{A}' = \mathcal{A} \cup \{h\}$ and $\Omega(\mathcal{A})$ is the set of random words on the alphabet $\mathcal{A}$: each occurrence of the word G is replaced by the letter h with probability 1. The extension of any measure μ under this operator is

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

Lemma 1.6.4. *The basic operator compression is consistent.*

Proof: Let (V_n) be a sequence of words converging to μ. We need to prove that

$$\lim_{n\to\infty} \frac{\text{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{1} h))}{|V_n(G \xrightarrow{\rho} h)| - |W| + 1} =$$

$$\frac{1}{\text{Ext}(\mu|(G \xrightarrow{1} h))} \ \lim_{n\to\infty} \frac{\text{freq}\,(W \ \text{ in } \ V_n(G \xrightarrow{1} h))}{|V_n|}. \qquad (1.15)$$

Let us first prove that both limits in (1.15) exist. Let W be a word in the alphabet $\mathcal{A}'$. If there exist words U and V, with $|U| < |G|$ and $|V| < |G|$, satisfying $\text{freq}\,(G \ \text{ in } \ W) < \text{freq}\,(G \ \text{ in } \ \text{concat}\,(U, W, V))$, then $\text{freq}\,(W \ \text{ in } \ V_n) = 0$. Otherwise, notice that

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

$$\text{freq}\,(W' \ \text{ in } \ V_n) - \text{freq}\,(W \ \text{ in } \ G)\cdot\text{freq}\,(G \ \text{ in } \ V_n),$$

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

$$\lim_{n\to\infty} \frac{\operatorname{freq}(W \ \text{ in } \ V_n(G \xrightarrow{1} h))}{|V_n|} =$$

$$\lim_{n\to\infty} \frac{\operatorname{freq}(W' \ \text{ in } \ V_n) - \operatorname{freq}(W \ \text{ in } \ G) \operatorname{freq}(G \ \text{ in } \ V_n)}{|V_n|}$$

$$= \mu(W') - \operatorname{freq}(W \ \text{ in } \ G) \cdot \mu(G).$$

Hence it is easy to see that both limits in (1.15) exist. Now the equation (1.15) follows from the definition of extension. *Lemma 1.6.4 is proved.*

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

$$\begin{aligned} \mu(G \xrightarrow{1} h)(W) &= \frac{\mu(W') - \operatorname{freq}(W \ \text{ in } \ G) \times \mu(G)}{\operatorname{Ext}(\mu|(G \xrightarrow{1} h))} \\ &= \frac{\mu(W') - \operatorname{freq}(W \ \text{ in } \ G) \times \mu(G)}{1 - (|G| - 1) \times \mu(G)}. \end{aligned}$$

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 to H* is the following map from $\Omega(\mathcal{A}')$ to $\Omega(\mathcal{A})$, where $\mathcal{A}' = \mathcal{A} \cup \{g\}$ and, again, $\Omega(\mathcal{A})$ is the set of random words on the alphabet $\mathcal{A}$: every occurrence of the letter g is replaced by the word H with probability 1. The extension of any measure μ for this operator is

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

Lemma 1.6.5. *The basic operator decompression is consistent.*

Proof: Let (V_n) be a sequence of words converging to the measure

μ. We need to prove that for any word W

$$\lim_{n\to\infty} \frac{\operatorname{freq}(W \ \text{ in } \ V_n(g \xrightarrow{1} H))}{|V_n(g \xrightarrow{\rho} H)| - |W| + 1} =$$

$$\frac{1}{\operatorname{Ext}(\mu|(g \xrightarrow{1} H))} \lim_{n\to\infty} \frac{\operatorname{freq}(W \ \text{ in } \ V_n(g \xrightarrow{1} H))}{|V_n|}. \qquad (1.16)$$

Let us first prove that the limit in the right side of (1.16) exists. 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

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

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

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

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

After that we turn each entrance of the word (h_1, h_2) in W' into the letter g and denote the resulting word by W''. (We may do it 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 $\operatorname{freq}(W \ \text{ in } \ V_n(g \xrightarrow{1} (h_1, h_2))) = 0$, and therefore the above limit equals zero. Otherwise,

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

and thus,

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

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

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

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

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 $\mathcal{A} \cup \{h_1, \ldots, h_k\}$

$$\operatorname{freq}(W \ \text{ in } \ V_n(g \xrightarrow{1} (h_1, \ldots, h_k))) =$$

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

and then we can prove by induction that the following limits exist and the following equality holds:

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

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

Now it is easy to see that the limits in the equation (1.16) exist.

We will now use 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

$h_1, \ldots, h_k$ and s do not belong to $\mathcal{A}$. Then

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

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:

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

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

Finally, we can define the decompression operator acting on a measure μ. It transforms a letter g into an arbitrary word $H = (s_1, \ldots, s_k)$ with no restrictions on letters $s_1, \ldots, s_k$. 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 $i = 1, \ldots, k$. *Lemma 1.6.5 is proved.*

1.7 Compositions of Basic SO

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

Theorem 1.7.1. *Let $(G \overset{\rho}{\to} H)$, where G is self-avoiding, be a SO acting on words. Let also $(G \overset{\rho}{\to} \Lambda)$ and $(\Lambda \overset{\rho}{\to} H)$ be SO acting on words, where Λ is the empty word, s, g, h are different letters not belonging to $\mathcal{A}$, and $\rho \in [0, 1]$ (and $\rho < 1$ for the operator $(G \overset{\rho}{\to} \Lambda)$). Then, for any words V and W:*

$$E[\mathrm{freq}\,(W \;\; \mathrm{in} \;\; V(G \overset{\rho}{\to} H))] =$$

$$E[\mathrm{freq}\,(W \;\; \mathrm{in} \;\; V(G \overset{1}{\to} h)(h \overset{\rho}{\to} s)(s \overset{1}{\to} H)(h \overset{1}{\to} G))],$$

$$E[\mathrm{freq}\,(W \;\; \mathrm{in} \;\; V(G \overset{\rho}{\to} \Lambda))] =$$

$$E[\mathrm{freq}\,(W \;\; \mathrm{in} \;\; V(G \overset{1}{\to} g)(g \overset{\rho}{\to} \Lambda)(g \overset{1}{\to} G))],$$

$$E[\mathrm{freq}\,(W \;\; \mathrm{in} \;\; V(\Lambda \overset{\rho}{\to} H))] =$$

$$E[\mathrm{freq}\,(W \;\; \mathrm{in} \;\; V(\Lambda \overset{\rho}{\to} h)(h \overset{1}{\to} H))].$$

Also

$$E|V(G \overset{\rho}{\to} H)| = E|V(G \overset{1}{\to} h)(h \overset{\rho}{\to} s)(s \overset{1}{\to} H)(h \overset{1}{\to} G)|, \qquad (1.23)$$

$$E|V(G \xrightarrow{\rho} \Lambda)| = E|V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G)|, \tag{1.24}$$

and finally,

$$E|V(\Lambda \xrightarrow{\rho} H)| = E|V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)|. \tag{1.25}$$

Proof of theorem 1.7.1: Observe that for any word V, the distributions of the random words

$$V(G \xrightarrow{\rho} H) \ \text{ and } \ V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)$$

are one and the same. Therefore the mean frequency and the mean length are the same for both random words. The same argument holds for the other cases. *Theorem 1.7.1 is proved.*

Now let us state several corollaries, which will allow us to define SO on measures.

Corollary 1.7.2. *Let the SO*

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

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

$$\text{rel.freq}_E(W \ \text{ in } \ V(G \xrightarrow{\rho} H)) = \tag{1.26}$$

$$\text{rel.freq}_E(W \ \text{ in } \ V(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G)),$$

$$\text{rel.freq}_E(W \ \text{ in } \ V(G \xrightarrow{\rho} \Lambda)) = \tag{1.27}$$

$$\text{rel.freq}_E(W \ \text{ in } \ V(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G))$$

and

$$\text{rel.freq}_E(W \ \text{ in } \ V(\Lambda \xrightarrow{\rho} H)) = \tag{1.28}$$

$$\text{rel.freq}_E(W \ \text{ in } \ V(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H)). \tag{1.29}$$

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

$$V_n(G \xrightarrow{\rho} H) \ \text{ converges } \iff \tag{1.30}$$

$$V_n(G \xrightarrow{1} h)(h \xrightarrow{\rho} s)(s \xrightarrow{1} H)(h \xrightarrow{1} G) \ \text{ converges,} \tag{1.31}$$

$$V_n(G \xrightarrow{\rho} \Lambda) \quad converges \quad \Longleftrightarrow V_n(G \xrightarrow{1} g)(g \xrightarrow{\rho} \Lambda)(g \xrightarrow{1} G) \quad converges \tag{1.32}$$

and

$$V_n(\Lambda \xrightarrow{\rho} H) \quad converges \quad \Longleftrightarrow V_n(\Lambda \xrightarrow{\rho} h)(h \xrightarrow{1} H) \quad converges. \tag{1.33}$$

Proof: straightforward. *Corollary 1.7.2 is proved.*

To imitate an arbitrary operator $(G \xrightarrow{\rho} H)$, where G, H are words in an alphabet $\mathcal{A}$ and G is self-avoiding we first compress (with probability 1) each entrance of the word G into a letter h, which is introduced especially for this purpose and does not belong to $\mathcal{A}$. Then with probability ρ we turn each letter h into a letter $s \neq h$ which also does not belong to $\mathcal{A}$. After that we decompress (with probability 1) the letter s into a word H and decompress the letter h into a word G. We proceeded analogously to imitate the other operators.

Corollary 1.7.3. *For any words G, H, where G is self-avoiding, and any $\rho \in [0,1]$ (where $\rho < 1$ if $H = \Lambda$), the operator $(G \xrightarrow{\rho} H)$ acting on words is consistent.*

Proof: The identities (1.26), (1.27) and (1.29) yield that the SO is the composition of basic operators described in the last section, and each basic operator is consistent. Thus, by lemma 1.4.4, their composition is also consistent. Thus $(G \xrightarrow{\rho} H)$ is consistent. *Corollary 1.7.3 is proved.*

In view of the above corollary, we have the following definition:

Definition 1.7.4. 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 1.4.3) 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 1.7.5. *Consider the operator $(G \xrightarrow{\rho} H)$ acting on measures, where G is self-avoiding and $\rho \in [0,1]$ ($\rho < 1$ if $H = \Lambda$).*

Then the following identities hold for any $s, g, h \notin \mathcal{A}$:

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

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

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

Proof: It is a straightforward consequence of corollary 1.7.2. *Corollary 1.7.5 is proved.*

1.8 Segment-Preserving Operators

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

$$\mathrm{convex}\,(\mu, \nu) = \{k\mu + (1-k)\nu \,|\, 0 \le k \le 1\}. \tag{1.34}$$

Lemma 1.8.1. *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 concat $(V_n,\ W_n)$ *converges to the measure* $L \cdot \mu + (1-L) \cdot \nu$ *when* $n \to \infty$.

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

$$\begin{aligned} & \mathrm{freq}\,(W \quad \mathrm{in} \quad V_n) + \mathrm{freq}\,(W \quad \mathrm{in} \quad W_n) \\ \le \quad & \mathrm{freq}\,(W \quad \mathrm{in} \quad \mathrm{concat}\,(V_n, W_n)) \\ \le \quad & \mathrm{freq}\,(W \quad \mathrm{in} \quad V_n) + \mathrm{freq}\,(W \quad \mathrm{in} \quad W_n) + 1. \end{aligned}$$

Hence,

$$\frac{\operatorname{freq}(W \ \text{ in } \ V_n)}{|V_n|+|W_n|} + \frac{\operatorname{freq}(W \ \text{ in } \ W_n)}{|V_n|+|W_n|}$$

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

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

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

Therefore

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

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

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

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

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

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

But

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

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

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

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

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

That is, the sequence of words concat $(V_n,\ W_n)$ converges to the measure $L\mu + (1 - L)\nu$. *Lemma 1.8.1 is proved.*

Lemma 1.8.2. *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 $V_n^{k_n} \to \mu$ as $n \to \infty$.*

Proof follows immediately from the inequalities:

$$k_n \cdot \text{freq}\,(W \ \text{ in } \ V_n) \le \text{freq}\,(W \ \text{ in } \ V_n^{k_n}) \le k_n \cdot \text{freq}\,(W \ \text{ in } \ V_n) + k_n \cdot |W|$$

and from the fact that $|V_n^{k_n}| = k_n |V_n|$. *Lemma 1.8.2 is proved.*

Theorem 1.8.3. *For any $L \in [0, 1]$ 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* 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

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

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

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

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 $|\widehat{V}_n| = q_n \cdot |\widetilde{V}_n|$, we conclude that $|\widehat{W}_n| = p_n \cdot |\widetilde{W}_n|$ 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 1.8.2, $\widehat{V}_n \to \mu$ 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 1.8.1, we have that

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

Theorem 1.8.3 is proved.

The following definition gives us a useful property, which all SO have. This property is trivially satisfied for linear operators, but is not true in general.

Definition 1.8.4. An operator $P : \mathcal{M}_{\mathcal{A}} \to \mathcal{M}_{\mathcal{B}}$, where $\mathcal{A}$ and $\mathcal{B}$ are alphabets, is called *segment-preserving* if

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

where $\operatorname{convex}(\mu,\ \nu)$ was defined in (1.34).

Theorem 1.8.5. *Every SO $(G \xrightarrow{\rho} H)$ is segment-preserving 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 $\mu,\ \nu$, where

$$\widetilde{L} = \frac{L \cdot \operatorname{Ext}(\mu|(G \xrightarrow{\rho} H))}{L \cdot \operatorname{Ext}(\mu|(G \xrightarrow{\rho} H)) + (1-L) \cdot \operatorname{Ext}(\nu|(G \xrightarrow{\rho} H))}. \tag{1.35}$$

Proof: The proof will be done by first obtaining a similar result for words, then going to the limit and finally proving it for measures. The key tool is theorem 1.8.3. Let $L \in (0,1)$. Due to theorem 1.8.3 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 SO $(G \xrightarrow{\rho} H)$ satisfies this:

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

Let us choose any word W. Then

$$E[\text{freq}\,(W \quad \text{in} \quad V_n(G \xrightarrow{\rho} H))] + E[\text{freq}\,(W \quad \text{in} \quad W_n(G \xrightarrow{\rho} H))] \leq$$

$$E[\text{freq}\,(W \quad \text{in} \quad \text{concat}\,(V_n, W_n)(G \xrightarrow{\rho} H))] \leq$$

$$E[\text{freq}\,(W \quad \text{in} \quad V_n(G \xrightarrow{\rho} H))] + E[\text{freq}\,(W \quad \text{in} \quad W_n(G \xrightarrow{\rho} H))] + 1.$$

Note that

$$\frac{1}{E|\text{concat}\,(V_n,\ W_n)(G \xrightarrow{\rho} H)|} \overset{n\to\infty}{\longrightarrow} 0.$$

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

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

Notice further that

$$E|\text{concat}\,(V_n,\ W_n)(G \xrightarrow{\rho} H)| \ =$$

$$|V_n| + |W_n| + \rho(|H| - |G|) \cdot \text{freq}\,(G \quad \text{in} \quad \text{concat}\,(V_n,\ W_n))$$

and furthermore that

$$\begin{aligned} & \text{freq}\,(G \quad \text{in} \quad V_n) + \text{freq}\,(G \quad \text{in} \quad W_n) \\ \leq \quad & \text{freq}\,(G \quad \text{in} \quad \text{concat}\,(V_n, W_n)) \\ \leq \quad & \text{freq}\,(G \quad \text{in} \quad V_n) + \text{freq}\,(G \quad \text{in} \quad W_n) + 1. \end{aligned}$$

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. We want to sandwich the middle part of (1.36) between two values, say a_n and b_n, which we shall choose in an appropriate way:

$$a_n \leq$$

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

$$\leq b_n.$$

First let us care about the right inequality in (1.36). To choose appropriate values of b_n we use the following inequalities:

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

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

$$\leq \frac{E[\text{freq}\,(W \;\text{ in }\; V_n(G \xrightarrow{\rho} H))] + E[\text{freq}\,(W \;\text{ in }\; W_n(G \xrightarrow{\rho} H))]}{|V_n| + |W_n| + \rho(|H| - |G|) \cdot (\text{freq}\,(G \;\text{ in }\; V_n) + \text{freq}\,(G \;\text{ in }\; W_n))}.$$

These inequalities suggest one to choose

$$b_n = \tag{1.36}$$

$$\frac{E[\text{freq}\,(W \;\text{ in }\; V_n(G \xrightarrow{\rho} H))] + E[\text{freq}\,(W \;\text{ in }\; W_n(G \xrightarrow{\rho} H))]}{|V_n| + |W_n| + \rho(|H| - |G|) \cdot (\text{freq}\,(G \;\text{ in }\; V_n) + \text{freq}\,(G \;\text{ in }\; W_n))}.$$

It is evident that with these b_n the right inequality (1.36) holds. Analogously we can choose a_n to satisfy the left inequality in (1.36).

Now let us check the limiting behavior of b_n. We begin by checking the limiting behavior of the following quantity:

$$\frac{E[\text{freq}\,(W \;\text{ in }\; V_n(G \xrightarrow{\rho} H))]}{|V_n| + |W_n| + \rho(|H| - |G|) \cdot (\text{freq}\,(G \;\text{ in }\; V_n) + \text{freq}\,(G \;\text{ in }\; W_n))}.$$

The limit behavior of the other part included in b_n 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:

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

$$= \frac{|V_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\text{freq}\,(G \;\text{ in }\; V_n) + \text{freq}\,(G \;\text{ in }\; W_n))}$$

$$\times S_n \times M_n$$

$$= \frac{|V_n| + |W_n|}{|V_n| + |W_n| + \rho(|H| - |G|)(\text{freq}\,(G \;\text{ in }\; V_n) + \text{freq}\,(G \;\text{ in }\; W_n))}$$

$$\times L_n \times S_n \times M_n,$$

where

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

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

$$M_n = \frac{E[\text{freq}\,(W \;\text{ in }\; V_n(G \xrightarrow{\rho} H))]}{E|V_n(G \xrightarrow{\rho} H)|} \quad \to \quad \mu(G \xrightarrow{\rho} H)(W).$$

Starting here for the next three pages our text is abbreviated to fit into the format of this book. You can find the full text in [22]. Going on, we have

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

which tends to

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

as $n \to \infty$. By means of the analogous calculations, one can obtain:

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

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

Thus we have obtained the limiting behavior of b_n in the equation (1.36):

$$b_n \xrightarrow{n\to\infty}$$

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

$$\times \mu(G \xrightarrow{\rho} H)(W)$$

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

$$\times \nu(G \xrightarrow{\rho} H)(W).$$

Analogously, it is possible to find a sequence (a_n) satisfying equation (1.36) such that it has the same limit as (b_n) . We conclude that

$$\text{concat}\,(V_n, W_n)(G \xrightarrow{\rho} H) \xrightarrow{n\to\infty}$$

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

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

Thus, applying theorem 1.8.3, since

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

we obtain

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

for all $L \in (0,1)$, where $\widetilde{L}$ was defined in (1.35). *Theorem 1.8.5 is proved.*

1.9 All SO Are Continuous

For any $\mathcal{M}' \subset \mathcal{M}$ we say that an operator $P : \mathcal{M}' \to \mathcal{M}'$ is *continuous* if whenever a sequence $\mu_n \in \mathcal{M}'$ tends to $\lambda \in \mathcal{M}'$ (in the weak topology, i.e., convergence separately on every word), the sequence $\mu_n P$ tends to λP (the well-known sequential continuity).

Definition 1.9.1. A measure μ is called *invariant* for an operator P if $\mu P = \mu$.

The article [28] indicated the following corollary of the well-known fixed point theorems:

Theorem 1.9.2. *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 a general result about continuity of consistent operators.

Theorem 1.9.3. *Let $P : \Omega \to \Omega$ be a consistent operator. Given any non-empty compact convex $\mathcal{M}' \subset \mathcal{M}$, let $P : \mathcal{M}' \to \mathcal{M}'$ be the limit operator defined on measures (see definition 1.4.3). Then P (defined on measures) 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 $k \to \infty$. Let $\varepsilon > 0$ be any positive number. We can choose k such that

$$|\text{meas}^{V_{k_k}}(W) - \mu_k(W)| < \varepsilon/2 \;\text{ and }\; |\mu_k(W) - \mu(W)| < \varepsilon/2.$$

Then

$$|\text{meas}^{V_{k_k}}(W)-\mu(W)| \leq |\text{meas}^{V_{k_k}}(W)-\mu_k(W)|+|\mu_k(W)-\mu(W)| \leq \varepsilon.$$

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

Now, since P is consistent (see definition 1.4.2), we have $V_{n_k}P \to \mu_n P$ as $k \to \infty$ and $V_{k_k}P \to \mu P$ as $k \to \infty$. Therefore for any fixed $\varepsilon > 0$ and large enough k we have

$$|\mu_k P(W)-\text{meas}^{V_{k_k}P}(W)| < \varepsilon/2 \quad \text{and} \quad |\mu P(W)-\text{meas}^{V_{k_k}P}(W)| < \varepsilon/2.$$

Therefore

$$|\mu_k P(W) - \mu P(W)| \leq$$

$$|\mu_k P(W) - \text{meas}^{V_{k_k}P}(W)| + |\mu P(W) - \text{meas}^{V_{k_k}P}(W)|$$

$$\leq \varepsilon.$$

Theorem 1.9.3 is proved.

Now we note that $\mathcal{M}$ is convex and compact and apply theorem 1.9.2 to conclude the following:

Corollary 1.9.4. *Let $P : \mathcal{M}' \to \mathcal{M}'$ be the limit of consistent operators (see definition 1.4.3), where $\mathcal{M}'$ is a closed and convex subset of $\mathcal{M}$. Then P has at least one invariant measure.*

Proof: Since $\mathcal{M}'$ is closed and $\mathcal{M}$ is compact, $\mathcal{M}'$ also is compact. Further, by theorem 1.9.3, the operator P is continuous, threfore by theorem 1.9.2 P has an invariant measure. *Corollary 1.9.4 is proved.*

The next corollary applies these results to SO:

Corollary 1.9.5. *Every SO $(G \xrightarrow{\rho} H)$ (where $\rho < 1$ if $H = \Lambda$) is continuous and has an invariant measure.*

Proof: Take any $(G \xrightarrow{\rho} H)$. By corollary 1.7.3, it is consistent. Therefore by theorem 1.9.3, it is continuous. Then, by corollary 1.9.4, $(G \xrightarrow{\rho} H)$ has an invariant measure. *Corollary 1.9.5 is proved.*

Remark 1.9.6. We note that in [28], the proof of continuity of the SO is different from ours, since it proves that the basic operators are quasi-local and therefore continuous, and further, that any composition of continuous operators is continuous.

1.10 A Large Class of Operators

We now introduce a large class of stochastic processes which contains as particular cases, for instance, the process defined in [27]. For all these processes we prove existence of at least one invariant measure.

Definition 1.10.1. Let $\mu \in \mathcal{M}$ and let $(G \xrightarrow{\rho} H)$ (where $\rho < 1$ if $H = \Lambda$) be a SO. We define the *discrete substitution process* (μ_n) starting at μ, by

$$\mu_n(W) = \mu(G \xrightarrow{\rho} H)^n(W) \text{ for every word } W.$$

Definition 1.10.2. Let $\nu \in \mathcal{M}$, and $P_1, \ldots, P_j$ be a finite sequence of SO. Then we define the *generalized discrete substitution process* (ν_n), where $\nu_0 = \nu$, as follows:

$$\nu_n(W) = \nu(P_1 P_2 \cdots P_j)^n(W) \text{ for every word } W.$$

It is easy to see that the process defined in [27] is a special case of our 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 an even more general class of processes, which we will call the *consistent processes* and define them as follows:

Definition 1.10.3. Let $P : \mathcal{M}' \to \mathcal{M}'$ be the limit of consistent operators (see definition 1.4.3 for the definition of limit of consistent operators), and let μ be a measure in $\mathcal{M}'$. Then we say that (μ_n) is a *consistent process* starting at μ if

$$\mu_n(W) = \mu P^n(W) \text{ for all words } W.$$

Since every SO is consistent (see corollary 1.7.3), and a composition of several consistent operators is also consistent, any generalized discrete substitution process is also consistent.

Theorem 1.10.4. *Let $P : \mathcal{M}' \to \mathcal{M}'$ be the limit of consistent operators (see definition 1.4.3), where $\mathcal{M}'$ is a convex and closed subset of $\mathcal{M}$. Then P has an invariant measure.*

Proof: a straightforward application of corollary 1.9.4. *Theorem 1.10.4 is proved.*

Remark 1.10.5. Since any generalized discrete substitution process is a special case of consistent processes, every generalized discrete substitution process has at least one invariant measure.

Theorem 1.10.6. *Let us consider a generalized discrete substitution process* $\nu_n = \nu_0 P^n$, *where* $P = P_1 P_2 \cdots P_j$ *as in definition 1.10.2. Let* $S \subset \mathbb{A}^{\mathbb{Z}}$ *be some subset of the* σ*-algebra* $\mathbb{A}^{\mathbb{Z}}$*. Then, if* $\nu_n(c) \leq \delta$ *(respectively* $\nu_n(c) \geq \varepsilon$*) for all* $c \in S$*, then* P *has an invariant measure* μ *such that* $\mu(c) \leq \delta$ *(respectively* $\mu(c) \geq \varepsilon$*) 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 $\nu_0,\ \nu_1, \ldots$. 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 1.9.5 to note that P is continuous. Further, the continuity together with theorem 1.8.5 yields that if $\tau \in \mathcal{M}'$, then τP also belongs to $\mathcal{M}'$. Therefore, by theorem 1.9.2 the operator P has an invariant measure μ in $\mathcal{M}'$. Since for every n and every $c \in S$, $\nu_n(c) \leq \delta$ (respectively $\nu_n(c) \geq \varepsilon$), a simple calculation shows that for every $\tau \in \mathcal{M}'$ and every $c \in S$ we have $\tau(c) \leq \delta$ (respectively $\tau(c) \geq \varepsilon$). Therefore the invariant measure μ satisfies $\mu(c) \leq \delta$ (respectively $\mu(c) \geq \varepsilon$) for every $c \in S$. *Theorem 1.10.6 is proved.*

1.11 Application to the FA Case

We now consider the process studied in [27], which is a special case of the generalized substitution process defined in section 1.10. In this case our alphabet is $\mathcal{A} = \{\oplus, \ominus\}$, whose elements are called *plus* and *minus*. We consider 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* (see section 1.6). More precisely, Flip_β is $(\ominus \xrightarrow{\beta} \oplus)$, which turns every minus into plus with probability β independently from the fate of other components. 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 sequence of measures

$$\mu_n = \delta_{\ominus}(\mathrm{Flip}_\beta \mathrm{Ann}_\alpha)^n, \tag{1.37}$$

where δ_0 is the measure concentrated in the configuration, all of whose components are zeros. [27, 18] have proved the following:

Theorem 1.11.1. *For all $\beta \in [0, 1]$ and $\alpha \in (0, 1)$ the relative frequency of pluses in the measure μ_n does not exceed $250 \cdot \beta/\alpha^2$ for all n.*

Now, based on these results, we can prove more:

Theorem 1.11.2. *For all $\beta \in [0,\ 1]$ and $\alpha \in (0,\ 1)$ the operator $\mathrm{Flip}_\beta \mathrm{Ann}_\alpha$ has an invariant measure, whose relative frequency of pluses does not exceed $250 \cdot \beta/\alpha^2$.*

Proof: We may use theorem 1.10.6 with $S = \{\oplus\}$ and $\delta = 250 \cdot \beta/\alpha^2$ since by theorem 1.11.1

$$\mu_n(\oplus) < 250 \cdot \beta/\alpha^2 \quad \text{for all} \quad n.$$

Therefore, by theorem 1.10.6 the operator $\mathrm{Flip}_\beta \mathrm{Ann}_\alpha$ has an invariant measure ν such that

$$\nu(\oplus) \leq 250 \cdot \beta/\alpha^2.$$

Theorem 1.11.2 is proved.

Corollary 1.11.3. *For $\beta < \alpha^2/250$, the process (1.37) has at least two different invariant measures.*

Proof: On one hand the measure $\delta_\oplus$ concentrated in "all pluses" is invariant for the operator Flip_β Ann_α. On the other hand, by theorem 1.11.2 above, this operator has an invariant measure, in which the relative 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 1.11.3 is proved.*

1.12 Exercices

Exercise 1.12.1. *Explain why the main results proved here would also be valid if in the equation* (1.1)*, the normalizing factor were* $\mid V \mid$ *instead of* $\mid V \mid - \mid W \mid +1$.

Exercise 1.12.2. *Let* $\mathcal{A}$ *be an alphabet. Given a letter* $a \in \mathcal{A}$*, denote by* δ_a *the measure concentrated in the configuration, all components of which are* a*. Propose a sequence of words approximating* δ_a.

Exercise 1.12.3. *Fill the details in the proofs of Theorem 1.2.14 and Corollary 1.2.15.*

Exercise 1.12.4. *We say that a sequence of random words* (X_n) *is* Cauchy *if for any word* W *and any* $\varepsilon > 0$ *there is a* k *such that*

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

Show that a sequence of random words is Cauchy if and only if it converges to some measure.

Exercise 1.12.5. *Let* a *and* b *be letters. Describe the measure* $\delta_a(a \xrightarrow{\frac{1}{2}} b)$ *explicitly. Let* $\delta_2 = \delta_a(a \xrightarrow{\frac{1}{2}} b)$*. Obtain* $\delta_2(a \xrightarrow{\frac{1}{2}} b)$ *explicitly. Generalize to obtain* δ_n *explicitly.*

Exercise 1.12.6. *Let us call a process* ergodic *if, starting from any initial condition, it tends to one and the same invariant measure. In corollary 1.11.3 we presented a process with at least two different invariant measures, whence certainly non-ergodic. Let* $\mathcal{A} = \{a, b\}$*, and let* $h \notin \mathcal{A}$*. Use the operator*

$$((a, b) \xrightarrow{\rho} h) : \mathcal{M}_{\mathcal{A}} \to \mathcal{M}_{\mathcal{A}'},$$

where $\mathcal{A}' = \mathcal{A} \cup \{h\}$*, and* $0 < \rho < 1$ *to construct a non-ergodic substitution process.*

Chapter 2

Examples.

2.1 FA Revisited

Here we study in more detail the *Flip-Annihilation process*, which was mentioned in the previous section. We shall prove that it shows some form of non-ergodicity, similar to contact processes, but more unexpected. This process deserves attention because for some positive values of parameters it is non-ergodic and has at least two different invariant measuares.

As before, we have a finite set $\mathcal{A}$ called *alphabet*, whose elemens are called *letters* and the configuration space is $\mathcal{A}^Z$. Throughout this and next section $\mathcal{A}$ has only two elements *plus* and *minus* denoted by $\oplus$ and $\ominus$.

As in the previous sections, the time of our process is discrete and our process is a sequence of normalized uniform measures μ, μP, $\mu P^2, \ldots$, where μ is initial condition and P is transition operator. Informally speaking, this operator acts as follows. At every time step two transformations occur.

The first one, which we call *flip* and denote by Flip_β, turns every minus into plus with probability β independently from what happens at other places.

Evidently, flip is a special case of *conversion* studied in the previous sections and generally well-known. It is constant-length and linear.

The second operator can be informally described as follows: Under its action, whenever a plus is a left neighbor of a minus, both disappear with probability α independently from what happens at other places. It does not fit into that short list of basic operators, which were introduced in the previous sections, but it can be easily represented as a composition of some of them. This operator seems to be "impartial".

Now we shall declare our main results for $\alpha < 1$. Let us denote by $\delta_\ominus$ and $\delta_\oplus$ the degenerate measures concentrated in the configurations "all minuses" and "all pluses" respectively. For all natural t we denote

$$\mu_t \;=\; \delta_\ominus\,(\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha)^{\,t}. \tag{2.1}$$

Theorem 2.1.1. *For all natural t the frequency of pluses in the measure μ_t does not exceed $250 \cdot \beta/\alpha^2$.*

Theorem 2.1.1 was proved in [27] and improved in [18]. We must admit that we are not satisfied with this estimation. We would be much happier to have $const \cdot \beta/\alpha$ instead of it, because in this case we could hope to go to the limit in which α and β tended to zero with a constant proportion and thus define a process with a continuous time. Regretfully, we have to be content with what we have.

Since $\delta_\oplus$ is invariant for $\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ with any α and β, theorem 2.1.1 implies that the operator $\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ cannot be ergodic whenever $\beta < \alpha^2/250$ because in this case μ_t cannot tend to $\delta_\oplus$.

Theorem 2.1.2. *If $2\,\beta > \alpha$, the measures μ_t tend to $\delta_\oplus$ when $t \to \infty$.*

Taken together, theorems 2.1.1 and 2.1.2 show that the sequence of measures μ_t has at least two different modes of behavior. In one mode ($\beta > \alpha/2$) these measures tend to $\delta_\oplus$ when $t \to \infty$ and in the other mode ($\beta < \alpha^2/250$) they do not tend to $\delta_\oplus$.

Theorem 2.1.3. *Take any $\mu \in \mathcal{M}_{\{\ominus,\oplus\}}$ and suppose that $\beta > 0$ and $(1-\beta)\cdot\mu(\ominus) \le 1/2$. Then the measures $\mu\,(\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha)^{\,t}$ tend to $\delta_\oplus$ when $t \to \infty$.*

Let us denote by $s(\alpha,\ \beta)$ the supremum of density of $\oplus$ in measure μ_t for all natural t.

Theorem 2.1.4. *For every $\alpha \in (0,1)$, $s(\alpha,\ \beta)$ is not continuous as a function of β.*

These theorems show similarity and difference between our process and the well-known contact processes (see e.g. [9, 10]). Since our time is discrete, it is easier to compare our process with the well-known Stavskaya process, a discrete-time version of contact processes. (See [23] or example 1.2 on pp.8-10 of [2] or [24] or section 6.2 on p. 139 in [25].) Using our notations, Stavskaya process is a sequence of measures $\delta_\ominus\,(\mathsf{Flip}_\beta\ \text{Stav}\,)^t$, where the deterministic constant-length operator

$$\text{Stav}\ : \{\ominus, \oplus\}^{\mathbb{Z}} \to \{\ominus, \oplus\}^{\mathbb{Z}}$$

is defined by the rule

$$\forall\, x \in \{\ominus, \oplus\}^{\mathbb{Z}}, \quad k \in \mathbb{Z} : (x\ \text{Stav}\,)_k \ = \begin{cases} \oplus & \text{if } \ x_k = x_{k+1} = \oplus, \\ \ominus & \text{otherwise.} \end{cases}$$

The operator Stav favors minuses against pluses, because it turns any plus into a minus whenever its right neighbor is minus, but never turns minuses into pluses. The operator Flip_β, on the contrary, turns minuses into pluses with a rate β. So it is natural that their composition behaves in different ways for large vs. small β, namely, when β is large, minuses die out and when β is small, they do not. Contact processes behave in a similar way. In our case behavior is more unexpected: Flip_β favors pluses for any $\beta > 0$, annihilation is "impartial", but still minuses survive for β/α^2 small enough. Of course, "impartiality" of annihilation should be taken with a tongue in the cheek. In fact, it favors that state, which already prevails - see our lemma 2.3.1.

Theorem 2.1.3 shows another way in which our process is different from Stavskaya process, which does not tend to $\delta_\oplus$ from initial measures in which minuses and pluses are mixed at random in any proportion, provided initial density of minuses is positive and β is small enough. What about our function $s(\alpha,\ \beta)$, for some situations (contact processes, percolation) its analogs have been proved to be continuous. Theorem 2.1.4 shows that our process is different. In this respect (lack of continuity) its behavior may be classified as a first order phase transition.

2.2 Another Definition of FA

Now we need to define our operators formally. We shall represent them using independent auxiliary variables. Let us define Flip_β, denoting by $x_i \in \{\ominus, \oplus\}$ for all $i \in \mathbb{Z}$ the coordinates of $\{\ominus, \oplus\}^{\mathbb{Z}}$. Also, we use mutually independent variables F_i for all $i \in \mathbb{Z}$, each taking two values called *move* and *stay*, distributed according to a product-measure π, defined as follows:

$$F_i = \begin{cases} \textit{move} & \text{with probability} \quad \beta, \\ \textit{stay} & \text{with probability} \quad 1-\beta. \end{cases}$$

Finally, we have a third set of variables $y_i \in \{\ominus, \oplus\}$ for all $i \in \mathbb{Z}$, on which the measure $\mu\,\mathsf{Flip}_\beta$ is induced by the product of the measures μ and π with the map

$$y_i = \begin{cases} \ominus & \text{if } x_i = \ominus \text{ and } F_i = \textit{stay}, \\ \oplus & \text{in all the other cases.} \end{cases} \tag{2.2}$$

Clearly, operator Flip_β can be applied to any $\mu \in \mathcal{M}_{\{\ominus,\oplus\}}$ and produces a measure in $\mathcal{M}_{\{\ominus,\oplus\}}$, preserving the set of uniform normed measures.

The *Annihilation* operator $\mathsf{Ann}_\alpha : \mathcal{M}_{\{\ominus,\oplus\}} \to \mathcal{M}_{\{\ominus,\oplus\}}$. It is because of this operator we have to restrict our attention only to translation-invariant measures. We shall define Ann_α as a composition of two operators: $\mathsf{Ann}_\alpha = \text{Duel}_\alpha \text{ Clean}$ (first Duel_α, then Clean). You may imagine that when Duel_α is applied, a duel occurs between every pair of $\oplus$ and $\ominus$ occupying i-th and $(i+1)$-th sites respectively (in this order only). If the command *fire!* is given, which occurs for every such pair independently with a probability α, the duellists kill each other. Otherwise the command *stop!* is given and nothing happens. When Clean is applied, the dead bodies are disposed of and the live sites close ranks.

Now let us define operator Duel_α, a linear constant-length operator transforming any measure on $\{\ominus, \oplus\}^{\mathbb{Z}}$ into a measure on $\{\ominus, \oplus, \odot\}^{\mathbb{Z}}$, where $\odot$ is a third state introduced especially for this occasion and called *dead*. States different from dead, that is minus and plus, are called *live*. Let us call $x_i \in \{\ominus, \oplus\},\ i \in \mathbb{Z}$ the coordinates of the space $\{\ominus, \oplus\}^{\mathbb{Z}}$, where the original measure μ is defined. Also, we use mutually independent variables A_i for all $i \in \mathbb{Z}$, each taking two values called *fire* and *stop*, distributed according to a product-measure π, defined as follows:

$$A_i = \begin{cases} fire & \text{with probability} \quad \alpha, \\ stop & \text{with probability} \quad 1-\alpha \end{cases} \tag{2.3}$$

for any $i \in \mathbb{Z}$ independently of all the other components and of the measure μ. We denote by $y_i \in \{\ominus, \oplus, \odot\}$ the coordinates of the space, where the measure $\mu\,\mathrm{Duel}_\alpha$ is induced by the product of μ and π, with the following map:

$$y_i = \begin{cases} \odot & \text{if } \ x_i \ = \oplus, \ \ x_{i+1} = \ominus \ \text{ and } \ A_{i+1} = fire, \\ \odot & \text{if } \ x_{i-1} = \oplus, \ \ x_i \ = \ominus \ \text{ and } \ A_i \ = fire, \\ x_i & \text{in all the other cases.} \end{cases}$$

Also notice that $\mu(\oplus, \ominus) \le 1/2$ for any $\mu \in \mathcal{M}_{\{\ominus,\oplus\}}$, because $\mu(\oplus, \ominus) = \mu(\ominus, \oplus)$ and their sum does not exceed 1. Hence, since $\alpha < 1$ and from the definition of Duel_α

$$\mu\,\mathrm{Duel}_\alpha(\odot) \ = \ 2\,\alpha \cdot \mu(\oplus, \ominus) \ < \ 1. \tag{2.4}$$

Now let us define a variable-length operator

$$\mathrm{Clean} : \mathcal{M}_{\{\ominus,\oplus,\odot\}} \to \mathcal{M}_{\{\ominus,\oplus\}}.$$

For any $\mu \in \mathcal{M}_{\{\ominus,\oplus,\odot\}}$ we directly express the values of μ Clean on all words in the alphabet $\{\ominus, \oplus\}$ in terms of the values of μ on all words in the alphabet $\{\ominus, \oplus, \odot\}$. By definition we set μ Clean on the empty word to be 1. For any non-empty word $W = (a_0, \ldots, a_k) \in dict(\ominus, \oplus)$ we define μ Clean (W) as follows:

$$\mu \ \mathrm{Clean}\,(a_0, \ldots, a_k) \ = \frac{1}{1-\mu(\odot)} \ \times \tag{2.5}$$

$$\sum_{n_1,\ldots,n_k=0}^{\infty} \quad \mu(a_0 \ \odot^{n_1} \ a_1 \ \odot^{n_2} \ a_2 \ \ldots \ \odot^{n_{k-1}} \ a_{k-1} \ \odot^{n_k} \ a_k),$$

where $\odot^n$ means the word consisting of n letters, everyone of which is $\odot$ (in fact, the empty word if $n = 0$). So

$$a_0 \ \odot^{n_1} \ a_1 \ \odot^{n_2} \ a_2 \ \ldots \ \odot^{n_{k-1}} \ a_{k-1} \ \odot^{n_k} \ a_k$$

means the word, which starts with letter a_0, then go n_1 letters $\odot$ (in fact, none if $n_1 = 0$), then goes letter a_1, then n_2 letters $\odot$, then letter a_2, and so on till n_{k-1} letters $\odot$, then letter a_{k-1}, then n_k letters $\odot$, and finally letter a_k and the summing is done over all $n_1, \ldots, n_k$ from zero to infinity. Notice that the formula (2.5) is non-linear, whence the well-developed theory of linear operators cannot be applied here, which adds much to the difficulty of dealing with variable-length processes. Notice also that in the case $k = 1$ the formula (2.5) turns into

$$\mu\,\mathrm{Clean}\,(\ominus) = \frac{\mu(\ominus)}{1 - \mu(\odot)}, \qquad \mu\,\mathrm{Clean}\,(\oplus) = \frac{\mu(\oplus)}{1 - \mu(\odot)}. \tag{2.6}$$

2.3 Proof of Theorems 2.1.2, 2.1.3, 2.1.4

Lemma 2.3.1. *For any $\mu \in \mathcal{M}_{\{\ominus,\oplus\}}$,*

$$\text{if} \ \ \mu(\ominus) \le 1/2, \quad \text{then} \ \ \mu\,\mathsf{Ann}_\alpha\,(\ominus) \le \mu(\ominus).$$

Proof. From the definition of Duel_α

$$\mu \ \mathrm{Duel}_\alpha(\ominus) \ = \ \mu(\ominus) - \alpha \cdot \mu(\oplus, \ominus).$$

Hence, from (2.4) and from the definition of Clean

$$\mu\,\mathsf{Ann}_\alpha(\ominus) \ = \ \frac{\mu(\ominus) - \alpha \cdot \mu(\oplus, \ominus)}{1 - 2\alpha \cdot \mu(\oplus, \ominus)},$$

where the denominator is positive since $\alpha < 1$. Now, assuming that $\mu(\ominus) \le 1/2$,

$$\mu(\ominus) - \mu\,\mathsf{Ann}_\alpha(\ominus) \ = \ \mu(\ominus) - \frac{\mu(\oplus) - \alpha \cdot \mu(\oplus, \ominus)}{1 - 2\alpha \cdot \mu(\oplus, \ominus)} \ =$$

$$\frac{\alpha \cdot \mu(\oplus, \ominus)(1 - 2\mu(\ominus))}{1 - 2\alpha \cdot \mu(\oplus, \ominus)} \ \ge \ 0.$$

Lemma 2.3.1 is proved.

Proof of theorem 2.1.3. Since $\mu\,\mathsf{Flip}_\beta(\ominus) = (1-\beta)\cdot\mu(\ominus)$ and $(1-\beta)\cdot\mu(\ominus) \le 1/2$, the frequency of minuses in $\mu\,\mathsf{Flip}_\beta$ does not exceed $1/2$. Then, from lemma 2.3.1, the frequency of minuses in $\mu\,\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ also does not exceed $1/2$. Arguing in this way, we can prove by induction that the frequency of minuses in $\mu\,(\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha)^t$ does not exceed $(1-\beta)^{t-1}/2$ for all $t \ge 1$, and therefore tends to zero when $t \to \infty$, whence the measure tends to $\delta_\oplus$. *Theorem 2.1.3 is proved.*

Proof of theorem 2.1.2.. Here the case $\beta = 0$ is impossible and the case $\beta = 1$ is trivial, so let $0 < \beta < 1$. If there is t such that $(1-\beta)\cdot\mu_t(\ominus) \le 1/2$, theorem 2.1.2 follows from theorem 2.1.3. It remains to examine the case when $(1-\beta)\cdot\mu_t(\ominus) > 1/2$ for all t. We shall prove that this case is impossible. Notice that

$$\mu_{t+1}(\ominus) \;=\; \frac{(1-\beta)\cdot\mu_t(\ominus) - \alpha\cdot p}{1-2\alpha\cdot p}, \tag{2.7}$$

where we have denoted $p = \mu_t\,\mathsf{Flip}_\beta\,(\oplus,\ominus)$. It is easy to prove that the expression (2.7) is a growing function of p under our conditions. Since $\mu(\oplus,\ominus) \le 1/2$, whence $\alpha\cdot\mu(\oplus,\ominus) \le \alpha/2$, this implies that

$$\mu_{t+1}(\ominus) \;\le\; \frac{(1-\beta)\cdot x - \alpha/2}{1-\alpha},$$

where $x = \mu_t(\ominus)$. Therefore

$$\mu_{t+1}(\ominus) - \mu_t(\ominus) \;\le\; -\frac{(\beta-\alpha)x + \alpha/2}{1-\alpha}.$$

Here the right side is a linear function of x, which equals $-\alpha/(2-2\alpha)$ at $x = 1/(2-2\beta)$ and $-(\beta-\alpha/2)/(1-\alpha)$ at $x = 1$. Both of these values are negative, so

$$\mu_{t+1}(\ominus) - \mu_t(\ominus) \;\le\; m,$$

where m is a negative constant, whence $\mu_t(\ominus)$ tends to $-\infty$ when $t \to \infty$, which is impossible, because a probability cannot be negative. *Theorem 2.1.2 is proved.*

Proof of theorem 2.1.4 assuming that theorem 2.1.1 is already proved. Notice that $s(\alpha,\ \beta)$ cannot take values in $(1/2,\ 1)$, because if it does, then there is t such that $\mu_t(\oplus) > 1/2$. But then, due to theorem 2.1.3, $\mu_t(\oplus)$ tends to 1 when $t \to \infty$, whence $s(\alpha,\ \beta) = 1$. Thus, for any $\alpha \in (0,1)$ the value of $s(\alpha,\ \beta)$: (a) equals 1 if $\beta > \alpha/2$ due to theorem 2.1.2, (b) tends to 0 when $\beta \to 0$ due to theorem 2.1.1 and (c) cannot take values in $(1/2,\ 1)$ due to theorem 2.1.3; so it cannot be continuous. *Theorem 2.1.4 is reduced to theorem 2.1.1.*

Now we start to prove theorem 2.1.1. Henceforth, we assume that $\beta < \alpha^2/250$ because otherwise theorem 2.1.1 is obvious. Our proof of theorem 2.1.1 it is based on two well-known ideas: Peierls' contour method and duality of planar graphs. We present this proof only partially. Some parts of it are presented as exercises and you can find a complete argument in [27, 18].

2.4 Another Representation of FA

It is not necessary to clean out the dead particles out at every time step. We may leave them where they are, but in this case we have to sacrifice locality, namely we must organize interaction of live particles as if the dead particles were removed. Following this idea, we introduce a process ν, which differs from our original process in the following respect. Starting now, we denote by $x \in \mathbb{Z}$ the space coordinate. We shall also use a natural parameter y, which equals zero at the beginning and increases by one after every application of Flip_β or Ann_α. Thus y increases by two when t in the formula (2.1) increases by one. Accordingly, we denote by $F\,(x,t)$ and $A\,(x,t)$ and call *basic variables* those variables F_i and A_i, which participate in the $(t+1)$-th application of $\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$. Thus our basic space is

$$\Omega \;=\; \Big(\{move, stay\} \times \{fire, stop\}\Big)^{\mathbb{Z}\cdot\mathbb{Z}_+}$$

with coordinates

$$\Big(F\,(x,t)),\ A\,(x,t)\Big),\quad \text{where}\ \ x \in \mathbb{Z},\ t \in \mathbb{Z}_+ \tag{2.8}$$

and with a product measure π, according to which for all x, t

$$F(x,t) = \begin{cases} move & \text{with probability} \quad \beta, \\ stay & \text{with probability} \quad 1-\beta, \end{cases} \tag{2.9}$$

$$A(x,t) = \begin{cases} fire & \text{with probability} \quad \alpha, \\ stop & \text{with probability} \quad 1-\alpha. \end{cases} \tag{2.10}$$

Let us denote $V = \{(x,y),\ x \in \mathbb{Z},\ y \in \mathbb{Z}_+\}$. The sets of pairs $(x,y) \in V$ with a given y are called *y-levels* or just levels. Every pair $(x,y) \in V$ has a state denoted by $state(x,y)$, which equals $\ominus$, $\oplus$ or $\odot$ and all their states are functions of $\omega \in \Omega$ defined in the following inductive way.

Base of induction: $state(x,\ 0) = \ominus$ for all $x \in \mathbb{Z}$.

Induction step when y is even, say $y = 2t$, where $t \in \mathbb{Z}_+$ (imitating the action of Flip_β). For all $x \in \mathbb{Z}$:

$$state\Big(x, 2t+1\Big) =$$
$$\begin{cases} \oplus & \text{if } \ state(x,\ 2t) = \ominus \ \text{ and } \ F(x,\ t) = move, \\ state(x,\ 2t) & \text{in all the other cases.} \end{cases}$$

Induction step when y is odd, say $y = 2t+1$, where $t \in \mathbb{Z}_+$ (imitating the action of Ann_α, but without locality). For all $x \in \mathbb{Z}$:

$$
state(x,\ 2t+2) \ = \ \begin{cases} \odot \quad \text{if } \ state(x,\ 2t+1) = \ominus \ \text{ and} \\ \quad A\,(x,\ t) = fire \ \text{ and there is } \ x' < x \\ \quad \text{such that } \ state(x',\ 2t+1) = \oplus \\ \quad \text{and } \ \forall\, x'' \in \mathbb{Z} : x' < x'' < x \Rightarrow \\ \quad \Rightarrow state(x'',\ 2t+1) = \odot \ ; \\ \\ \odot \quad \text{if } \ state(x,\ 2t+1) = \oplus \\ \quad \text{and there is } \ x' > x \ \text{ such that} \\ \quad state(x',\ 2t+1) = \ominus \\ \quad \text{and } \ A\,(x',\ t) = fire \\ \quad \text{and } \ \forall\, x'' \in \mathbb{Z} : x < x'' < x' \Rightarrow \\ \quad \Rightarrow state(x'',\ 2t+1) = \odot \ ; \\ \\ state(x,\ 2t+1) \qquad \text{in all the other cases.} \end{cases}
$$

Informally speaking, in this process the particles never disappear and keep the same integer indices, which they had at the beginning. If a particle annihilates, it goes to the dead state $\odot$ and remains in this state forever. Live particles interact as if dead components did not exist. Thus we have an inductionally defined map from Ω to $\{\ominus, \oplus, \odot\}^V$. We denote by ν the measure on $\{\ominus, \oplus, \odot\}^V$ induced by the distribution π of the basic variables (2.9) and (2.10) with this map and ν_y the distribution of states on the y-th level. The process ν is useful for us because

$$
\nu_{2t}\,\text{Clean} \ = \ \mu_t \ \text{ for all } \ t. \tag{2.11}
$$

We fix an arbitrary natural number T. Our overall goal is to estimate $\mu_T(\oplus)$ uniformly in T. Due to exercise 2.8.3, $\mu_T(\ominus)$ is positive, so the fraction $\mu_T(\oplus)/\mu_T(\ominus)$ makes sense and it is sufficient to estimate this fraction. Using exercise 2.8.7, we get

$$\mu_T(\oplus) \;\leq\; \frac{\mu_T(\oplus)}{\mu_T(\ominus)} \;=\; \sum_{k=1}^{\infty} \frac{\mu_T(\ominus, \oplus^k)}{\mu_T(\ominus)}. \tag{2.12}$$

To reduce our task further, we concentrate our attention on Ω_0, the set of those $\omega \in \Omega$, for which $state(0,\, 2T) = \ominus$. For any $\omega \in \Omega_0$ we denote by $x_{max}(\omega)$ the smallest positive x such that $state(x,\, 2T) = \ominus$. Due to item b) of exercise 2.8.3, $x_{max}(\omega)$ exists a.s. Let us call by *flowers* all those pairs $(x,\, 2T)$, where $0 < x < x_{max}(\omega)$, for which $state(x,\, 2T) = \oplus$. We denote by $\phi(\omega)$ the number of flowers. Since $x_{max}(\omega)$ exists a.s., $\phi(\omega)$ is finite a.s. For any $k = 1, 2, 3, \ldots$ we denote by Ω_k the set of those $\omega \in \Omega_0$ for which $\phi(\omega) \geq k$. Notice that $\Omega_0 \supseteq \Omega_1 \supseteq \Omega_2 \supseteq \ldots$ Let us prove for all k that

$$\frac{\pi(\Omega_k)}{\pi(\Omega_0)} \;=\; \frac{\mu_T(\ominus, \oplus^k)}{\mu_T(\ominus)}. \tag{2.13}$$

Notice that $\pi(\Omega_0) = \nu_{2T}\,(\ominus)$. But from (2.11) and (2.6)

$$\mu_T(\ominus) \;=\; \nu_{2T}\ \text{Clean}\,(\ominus) \;=\; \frac{\nu_{2T}\,(\ominus)}{1 - \nu_{2T}\,(\odot)}\,,$$

whence

$$\pi(\Omega_0) \;=\; \nu_{2T}\,(\ominus) \;=\; \mu_T(\ominus)\Big(1 - \nu_{2T}\,(\odot)\Big). \tag{2.14}$$

On the other hand, Ω_k is the set of those $\omega \in \Omega_0$, for which the configuration at the level $2T$ contains one of the words

$$\ominus\ \odot^{n_1}\ \oplus \odot^{n_2}\ \ldots \oplus\ \odot^{n_{k-1}}\ \oplus \odot^{n_k}\ \oplus$$

starting at the 0-th component. Therefore

$$\pi(\Omega_k) \;=\; \sum_{n_1,\ldots,n_k=0}^{\infty} \nu_{2T}\,(\ominus\ \odot^{n_1}\ \oplus\ \ldots, \odot^{n_k} \oplus).$$

But from (2.11) and (2.5)

$$\mu_T(\ominus, \oplus^k) \;=\; \nu_{2T}\ \text{Clean}\,(\ominus, \oplus^k) \;=$$

$$\frac{1}{1 - \nu_{2T}\,(\odot)} \quad \sum_{n_1,\ldots,n_k=0}^{\infty} \nu_{2T}\,(\ominus\ \odot^{n_1}\ \oplus\ \ldots, \odot^{n_k} \oplus).$$

Thus

$$\pi(\Omega_k) \;=\; \mu_T(\ominus, \oplus^k) \cdot \Big(1 - \nu_{2T}\,(\odot) \Big)\,.$$

Dividing this by (2.14) , we get (2.13) . Now we can sum (2.13) over k and use (2.12) to obtain

$$\frac{\mu_T(\oplus)}{\mu_T(\ominus)} \;=\; \sum_{k=1}^{\infty} \frac{\mu_T(\ominus, \oplus^k)}{\mu_T(\ominus)} \;=\; \sum_{k=1}^{\infty} \frac{\pi(\Omega_k)}{\pi(\Omega_0)}. \tag{2.15}$$

2.5 A Graphical Representation

Now we go to a graphical representation of the process ν. In the following text we shall ignore some events, whose probability is zero. So, reading it, you should mentally insert "almost", "almost all" or "almost sure" whenever necessary. For any $\omega \in \Omega$ we define a graph G. Along with describing the graph G, we shall describe how to draw it in a plane, representing vertices by points and edges by curves (in fact, straight segments). The set of vertices of G is

$$V_G \;=\; \{(x,y) \in V,\; \mathit{state}(x,y) \neq \odot\}$$

and every vertex (x,y) is placed at the point (x,y) of the plane, where x and y are the usual orthogonal coordinates, the axis x is horizontal and the axis y is vertical. Graph G has two kinds of edges, which we call *vertical* and *horizontal*. Let us describe them.

Vertical edges: Any two vertices (x,y_1), (x,y_2) of G, where $y_2 - y_1 = 1$, are connected with a vertical edge. Direction of this edge from (x,y_1) to (x,y_2) is called *north*, the other direction is called *south*. We call (x,y_1) the south neighbor of (x,y_2) and (x,y_2) the north neighbor of (x,y_1).

Horizontal edges: Any two vertices (x_1,y), (x_2,y) of G, where $x_1 < x_2$, are connected with a horizontal edge if

$$\forall x \in \mathbb{Z} : x_1 < x < x_2 \Rightarrow \mathit{state}(x,y) = \odot.$$

Direction of this edge from (x_1,y) to (x_2,y) is called *east*, the opposite direction is called *west*. We call (x_1,y) the west neighbor of (x_2,y) and (x_2,y) the east neighbor of (x_1,y). Thus G, which has only those edges, which are specified above, is defined. Its edges are represented by straight segments connecting the points representing ends of the edge.

A vertex of G, whose level y is even, always has exactly one west neighbor, exactly one east neighbor and exactly one north neighbor. Also it has exactly one south neighbor, except the case $y = 0$, when it has no south neighbor. A vertex of G, whose level is odd, always has exactly one west neighbor, exactly one east neighbor and exactly one south neighbor. Also it has at most one north neighbor. Due to the definition of G, every vertex of it is in a state $\oplus$ or $\ominus$; in the former case we call it a $\oplus$-vertex, in the latter a $\ominus$-vertex.

It is evident that different edges G do not intersect except common ends. We shall call the *picture* of G its representation in the plane just described. This picture cuts the plane into parts, which we call *faces*. We assume that all the faces are closed. We call two faces neighbors if they have a common edge. Our picture of G has exactly one unbounded face, namely the bottom half of the plane. All the other faces of G are bounded and we call them *boxes*. Every box has the form of a rectangle, sanwiched between two parallel lines at levels y_1 and $y_1 + 1$, where y_1 is natural, so it may be denoted

$$\{(x, y) \in \mathbb{R}^2 \, : \, x_1 \le x \le x_2, \quad y_1 \le y \le y_1 + 1\}. \tag{2.16}$$

For every natural y_1 the boxes sandwiched between the parallel lines at the levels y_1 and $y_1 + 1$ form a bi-infinite sequence in which every two next terms have a common side and which we call a *horizontal corridor* at sub-$(y_1 + 1)$ level. Any box has at least four vertices placed at its corners and has no more vertices on its west, east and north walls, so it has exactly one west neighbor, one east neighbor and one north neighbor. If y_1 is even, the box (2.16) has no more vertices at its south wall, whence it has exactly one south neighbor. if y_1 is odd, this box (2.16) has $2k + 1$ south neighbors, where k is the number of annihilations, which occured at the $(y_1 + 1)/2$-th application of the operator Ann_α between sites x_1 and x_2.

Like in [24], we use the well-known duality of pictures of graphs. Let us describe a graph, which we denote by $\overline{G}$, and its picture, which will be dual of the picture of G. We place that vertex of $\overline{G}$, which is dual of the box (2.16), at the point

$$\left(\frac{x_1 + x_2}{2}, \quad y_1 + 1 - \varepsilon \right), \tag{2.17}$$

where $\varepsilon > 0$ is chosen for different boxes fifferently, but should be small enough in every case; how small, we shall explain. We shall say that the vertex (2.17) has a sub-$(y_1 + 1)$ level. We say that it has a *sub-even* level if $y_1 + 1$ is even and has a *sub-odd* level if $y_1 + 1$ is odd. There is just one subtlety: that vertex of $\overline{G}$, which is dual of the only unbounded face of the picture of G, is placed "infinitely far" in the negative direction of the axis y and the edges leading to it are rays with the same direction. All the other edges of $\overline{G}$ are straight segments connecting the points representing their ends. Thus the graph $\overline{G}$ and its picture are defined. It is easy to see that for any box the corresponding ε can be chosen so small that the usual conditions of dual pictures be fulfilled.

We shall call *horizontal* those edges of $\overline{G}$, which are dual of vertical edges of G and *vertical* those edges of $\overline{G}$, which are dual of horizontal edges of G. Notice that horizontal edges of $\overline{G}$ are approximatedly horizontal because values of ε for all vertices of $\overline{G}$ are approximatedly equal to zero. For any natural y the vertices of $\overline{G}$, which are at sub-$(y + 1)$ level, and horizontal edges, connecting them, form a bi-infinite path, which we call a *horizontal path* at sub-$(y + 1)$ level and which is dual of the sub-$(y + 1)$ corridor. Any bounded face of $\overline{G}$ is sandwiched between horizontal paths at the levels sub-y and sub-$(y + 1)$. Unbounded faces of $\overline{G}$ are dual of vertices of G at the level zero. They are unbounded half-strips, which fill all the halfplane below the horizontal path at the sub-1 level. A face of $\overline{G}$ is called a west (respectively east, north or south) neighbor of another face of $\overline{G}$ if their corresponding vertices of G are in the same relation.

According to what we said about vertices of G at even levels, any face of $\overline{G}$ at an even level has exactly one west neighbor, exactly one east neighbor and exactly one north neighbor. Also it has exactly one south neighbor, except the case $y = 0$, when it has no south neighbor. Whenever $y > 0$, we call these faces of $\overline{G}$ *rectangles*. In fact, all of them approximatedly are rectangles. According to what we said about vertices of G at odd levels, any face of $\overline{G}$ at an odd level has at most one north neighbor. If it has one, we call it a *trapezium*, otherwise we call it a *triangle.* Indeed, these faces approximatedly are trapeziums and triangles.

Let us take any $\omega \in \Omega_1$ and call a path in G *north-west* if its every step goes north or west. Let us call a vertex of G a *root* if there is a north-west path from this vertex to some flower, all the vertices of this path having a state $\oplus$. In particular, all the flowers are roots. Vertices of G, which are not roots, are called *non-roots.* The set of roots is finite a.s. for the same reason why the set of flowers is finite a.s., namely because T is fixed and therefore $x_{max}(\omega)$ exists a.s. Our estimation is based on building a "contour" around all the roots.

Let us call a set S of vertices of a graph *connected* in this graph if for any two elements of this set there is a path in this graph connecting them, in which all the vertices belong to S.

Let us call *dual-roots* those faces of $\overline{G}$, which are dual of roots, and denote by U the union of dual-roots. Since every dual-root is bounded, U is also bounded and closed since we assume all faces to be closed. Hence from lemma 2.5.1, U is homeomorphic to a closed disk (provided the set of roots is finite). Then the boundary of U is a closed curve, which includes the east side of the rectangle dual of the vertex $(0, 2T)$. So this closed curve includes V_0, the north end of this side, and we may assume that it starts and ends at V_0 and surrounds U in the counter-clockwise direction. This curve can be represented as a path in $\overline{G}$, which we denote by $tour(\omega)$ because it is determined by ω. Figure 2.1 illustrates our constructions.

Let us use figure 2.1 to explain the funcioning of our process. The transformation from y to $y+1$ is done by $Flip_\beta$ if y is even and by Ann_α if y is odd. The figure includes six instances of minuses turning into pluses due to the action of $Flip_\beta$ (for the values 1, 2, 3, 4, 6, 7 of x) and two instances of annihilation due to the action of Ann_α: the plus at $(7, 1)$ annihilates with the minus at $(8, 1)$ and the plus at $(4, 3)$ annihilates with the minus at $(5, 3)$. The leftmost and rightmost columns are filled with zeros since these zeros never were disturbed by our operators. For the leftmost column it displays the fact that our configuration belongs to Ω_0. The rightmost column with this property exists a.s. There are four flowers between these columns, namely $(1, 4)$, $(2, 4)$, $(3, 4)$, $(6, 4)$, marked with the letter F. The path $tour(\omega)$ surrounding the union of dual-roots is shown with thick vectors. The vertex V_0 is in its left upper corner. The vertices inside

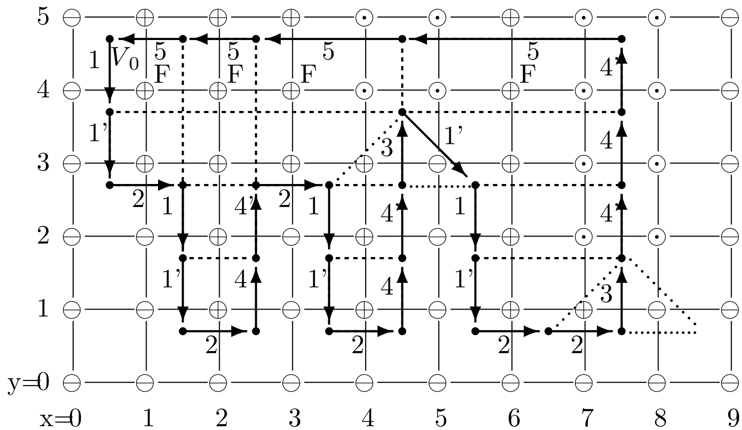

Figure 2.1: *A possible (that is, having a positive probability) fragment of our process ν.*

this path (all marked with pluses) are roots. Dual-roots, that is faces inside $tour(\omega)$, are separated from each other by dotted lines. To clarify the action of annihilation, boundaries of two triangles outside $tour(\omega)$ (dual of $(8,1)$ and $(5,3)$) are shown by dotted lines also. Types of steps of $tour(\omega)$ are shown near each step. These types form the code of $tour(\omega)$, which is

$$11'211'244'211'244'31'11'2234'44'5555.$$

This code includes all the types except 2' and 2" and all possible combinations of any two of these types. (We especially chose a configuration with this property.) The code of $bag(\omega)$ is the same without fives and short (code($bag(\omega)$)) is 121242124312234.

Lemma 2.5.1. *For any $\omega \in \Omega_1$: a) The set of roots is non-empty, finite and connected in G. b) The set of non-roots is infinite and connected in G.*

Proof is easy.

Now let us classify all the possible forms of $tour(\omega)$. To this end we need to classify all the steps which $tour(\omega)$ may include, that is some steps in $\overline{G}$. We shall start by classifying some steps in the original graph G. Let us call *types* elements of the set

$$\{1,\ 1',\ 2,\ 2',\ 2'',\ 3,\ 4,\ 4',\ 5\}. \tag{2.18}$$

We shall attribute types to those and only those steps in G, which start at $\oplus$-vertices. All the cases, which may occur, are listed in table 1.

Step in G starting at a $\oplus$-vertex	Type	Associated event	Associated variable
step west at an even level	1	trivial	none
step west at an odd level	1'	trivial	none
step from $(x,\ 2t+1)$ to $(x,\ 2t)$ if $F(x,\ t) = move$	2	$F(x,t)$ $= move$	$F(x,t)$
step from $(x,\ 2t+1)$ to $(x,\ 2t)$ if $F(x,\ t) = stay$	2'	$F(x,t)$ $= stay$	$F(x,t)$
step south from an even to an odd level	2"	trivial	none
step from $(x,\ 2t+1)$ to its east neighbor if $A(x,\ t) = fire$	3	$A(x,t)$ $= fire$	$A(x,t)$
step from $(x,\ 2t+1)$ to its east neighbor if $A(x,\ t) = stop$	4	$A(x,t)$ $= stop$	$A(x,t)$
step east at an even level	4'	trivial	none
step north	5	trivial	none

Table 1

Steps, having the word "trivial" in the third column, are called *trivial*, other steps are called *non-trivial*. To every step in G, which has a type, we attribute an *associated event*. For every trivial step the associated event is Ω and is also called trivial. Non-trivial events are represented in the table 1 by their conditions. For every non-trivial step we also define an *associated basic variable*, which is shown in the last column. Also every step in G, which has a type, has a *chance*. For typographical reasons chances are shown in the next table, but you can easily infer them right now because chance always equals the probability of the associated event. We shall use the same 1-to-1 correspondence between steps in G and steps in $\overline{G}$ as was defined in [24] and in more detail in [5]. Here it is:

$$\left.\begin{array}{l}\textit{If an edge } \bar{e} \textit{ of } \overline{G} \textit{ is dual of an edge } e \textit{ of } G\textit{, then for each}\\ \textit{direction of } e \textit{ the dual direction of } \bar{e} \textit{ is the direction from}\\ \textbf{right} \textit{ to } \textbf{left} \textit{ when we go along } e \textit{ in the given direction.}\end{array}\right\} \quad (2.19)$$

Type, event and chance, attributed to a step in G, are attributed to its dual step in $\overline{G}$ also. Since a step in G has a type if and only if it starts from a $\oplus$-vertex, a step in $\overline{G}$ has a type if and only if it has a $\oplus$-face on its left side.

step in $\overline{G}$, a $\oplus$-face on its left side	Type	Chance	Shift
step south across an even level	1	1	$(0, -1)$
step south across an odd level	1'	1	$(0, -1)$
"move" step east at a sub-odd level	2	β	$(1, 0)$
"stay" step east at a sub-odd level	2'	$1-\beta$	$(1, 0)$
step east at a sub-even level	2"	1	$(1, 0)$
"fire" step north across an odd level	3	α	$(-1, 1)$
"stop" step north across an odd level	4	$1-\alpha$	$(0, 1)$
step north across an even level	4'	1	$(0, 1)$
step west	5	1	$(-1, 0)$

Table 2

You may imagine tables 1 and 2 as one table, which is cut into two parts for typographical reasons. The last column of table 2 shows *shifts* defined for all types. Shift is a two-dimensional vector, whose components are called *Hshift* and *Vshift* (abbreviations for horizontal shift and vertical shift). The first column of table 2 is formally redundant because it follows from what was said in the first column of table 1; however, it helps to understand why shifts are defined in this way. Chances shown in the third column equal probabilities of events shown in the previous table.

Lemma 2.5.2. *For any $\omega \in \Omega_1$: a) all the steps of the path tour(ω) have types and b) the path tour(ω) is a concatenation of two paths, which we denote by bag(ω) and lid(ω), with the following properties: all the steps of bag(ω) have types different from 5; lid(ω) has $\phi(\omega)$ steps, all of which have type 5.*

Let us examine $bag(\omega)$. We start by two observations:

$$\left.\begin{array}{l}\text{a) If } bag(\omega) \text{ includes a step type 2, then this step has} \\ \text{a } \ominus\text{-face on its right (that is, south) side.} \\ \\ \text{b) If } bag(\omega) \text{ includes a step type 3, 4 or 4', then this} \\ \text{step has a } \ominus\text{-face on its right (that is, east) side.}\end{array}\right\} \qquad (2.20)$$

Indeed, in both cases, if there were a $\oplus$-face there, it would be a dual-root, but the contour surrounding U cannot separate dual-roots from each other.

Any sequence of types is called a *code*. By shift of a code we mean the sum of shifts of its terms and by chance of a code we mean the product of chances of its terms. If all the steps of a path p have types, we denote code(p) and call the *code* of p the sequence of types of steps of p. By shift and chance of such a path we mean shift and chance of its code. ¿From lemma 2.5.2, $bag(\omega)$ has a code and we need to study it. Let us call a path in $\overline{G}$ *well-placed* if it starts at V_0, all its steps have types and all the basic variables associated with its steps are independent from each other and from Ω_0. Given some $\omega \in \Omega_0$ and a code C, we say that ω *realizes* C if the graph $\overline{G}$ contains a well-placed path p, such that the code of p equals C.

Lemma 2.5.3. *Every $\omega \in \Omega_1$ realizes the code of bag(ω).*

For any code C we denote by $real(C)$ the set of those $\omega \in \Omega_0$, which realize C. It is easy to prove for any code C by induction in the length of codes that

$$\frac{\pi(real(C))}{\pi(\Omega_0)} \leq \text{chance}(C). \tag{2.21}$$

Hence, due to lemma 2.5.3, for any k

$$\frac{\pi(\Omega_k)}{\pi(\Omega_0)} \leq \frac{\sum \pi(real(\text{code}(bag(\omega))))}{\pi(\Omega_0)} \leq \sum \text{chance}(bag(\omega)), \tag{2.22}$$

where both sums are taken over all different code($bag(\omega)$) for $\omega \in \Omega_k$. To estimate the last sum, for every natural k we define a set of codes, which we denote LC_k and whose elements we call *k-legal codes.* A code $C = (c_1, \ldots, c_n)$ belongs to LC_k if it satisfies the following conditions:

$$\begin{cases} \text{LC-a) } c_1 = 1 \text{ and } c_n = 4'. \\ \text{LC-b) All the terms of } C \text{ belong to the list} \\ 1,\ 1',\ 2,\ 3,\ 4,\ 4'. \\ \text{LC-c) All the pairs } (c_i, c_{i+1}) \text{ belong to the list} \\ \\ 11',\ 1'1,\ 1'2,\ 21,\ 22,\ 23,\ 24,\ 31',\ 34',\ 44',\ 4'2,\ 4'3,\ 4'4. \\ \\ \text{LC-d) Hshift}(C) \geq k \text{ and Vshift}(C) = 0. \end{cases} \tag{2.23}$$

Since $\mathrm{LC}_1 \supseteq \mathrm{LC}_2 \supseteq \mathrm{LC}_3 \supseteq \ldots$, we denote $\mathrm{LC} = \mathrm{LC}_1$ and call elements of LC just *legal codes.* Of course, the definition of legal codes is chosen to fit codes of $bag(\omega)$ as the following shows.

Lemma 2.5.4. *For all $\omega \in \Omega_k$ the code of $bag(\omega)$ belongs to* LC_k.

It follows from lemma 2.5.4 and (2.22) that for all natural k

$$\frac{\pi(\Omega_k)}{\pi(\Omega_0)} \leq \sum_{C \,\in\, LC_k} \text{chance}(C). \tag{2.24}$$

2.6 Proof of Theorem 2.1.1 for $\alpha < 1$

To finish our argument, we need to make a numerical estimation, but it will be too cumbersome to do with so many types. To reduce their number to four, we call *main types* the elements of the set $\{1,\ 2,\ 3,\ 4\}$. Every main type is a type, so all the quantities defined for types are valid for main types. In particular, every main type has a shift and a chance listed in table 2 and shown again in table 3. Also every main type has a *rate*, which is shown in the same table:

main type	shift	chance	rate
1	$(\ 0,\ -1)$	1	1
2	$(\ 1,\ \ 0)$	β	2β
3	$(-1,\ \ 1)$	α	α
4	$(\ 0,\ \ 1)$	$1-\alpha$	$1-\alpha$

Table 3

A *main code* is a finite sequence, all the terms of which are main types. Its *rate* is the product of rates of its terms. For any code C we denote by $\operatorname{short}(C)$ the main code obtained from C by deleting all its non-main terms. We shall simplify our task by dealing with $\operatorname{short}(\operatorname{code}(bag(\omega)))$ instead of $\operatorname{code}(bag(\omega))$. For every natural k we define the set LMC_k, whose elements are called *k-legal main codes.* By definition, a k-legal main code is a main code $C=(c_1,\ldots,c_n)$, which satisfies the following conditions:

$$\begin{cases} \text{LMC-a) } c_1 = 1 \\ \text{LMC-b) For every } i=1,\ldots,n-1 \text{ it is impossible that} \\ (c_i=1,\ c_{i+1}=3) \text{ or } (c_i=1,\ c_{i+1}=4) \text{ or } (c_i=4,\ c_{i+1}=1). \\ \text{LMC-c) } c_n \text{ equals 3 or 4.} \\ \text{LMC-d) } \operatorname{Hshift}(C)\ \geq\ k. \\ \text{LMC-e) } \operatorname{Vshift}(C)\ \geq\ 0. \end{cases} \tag{2.25}$$

Since $\mathrm{LMC}_1 \supseteq \mathrm{LMC}_2 \supseteq \mathrm{LMC}_3 \supseteq \ldots$, we denote $\mathrm{LMC} = \mathrm{LMC}_1$ and call elements of LMC just *legal main codes.* You may notice also that from LMC-a), LMC-b) and LMC-c) any legal main code has length at least three, so in fact LMC = LMC_3, but we shall not use it. For any legal main code C let us denote by $\text{Long}\,(C)$ the set of legal codes C' such that $C = \text{short}\,(C')$. It is easy to observe that if $C' \in \text{Long}\,(C)$, then C' can be obtained from C by the following procedure:

$$\begin{cases} \text{a) We start with } C. \\ \text{b) After every 1 we insert 1'.} \\ \text{c) After every 4 we insert 4'.} \\ \text{d) If 3 is followed by 1, we insert 1' between them.} \\ \text{e) If 3 is followed by 2, we insert 1' or 4' between them.} \\ \text{f) If 3 is followed by 3, 4 or 5, we insert 4' between them.} \end{cases} \tag{2.26}$$

Also it is easy to prove that for any main code C and any $C' \in \text{Long}\,(C)$

$$\text{Hshift}(C') = \text{Hshift}(C), \tag{2.27}$$

$$\text{Vshift}(C') \leq 2 \cdot \text{Vshift}(C). \tag{2.28}$$

Here (2.27) is true because C' is obtained from C by inserting only types 1' and 4', both of which have Hshift = 0. To prove (2.28) , let us classify main types into *horizontal*, namely 2, whose Vshift is zero, and *vertical*, namely all the others. Due to the rule (2.26) we can establish a 1-to-1 correspondence between vertical terms of C and the terms inserted after them in the course of this procedure. Then Vshift of every newly inserted term is not greater than Vshift of the corresponding vertical term of C. Hence (2.28) immediately follows.

Lemma 2.6.1. *For any k, if $C \in LC_k$, then* $\text{short}\,(C) \in LMC_k$.

Now we can estimate the sum in the right side of (2.24) . Due to lemma 2.6.1 we can represent this sum as

$$\sum_{C' \in LC_k} \text{chance}(C') = \sum_{C \in LMC_k} \sum_{C' \in Long(C)} chance(C'). \tag{2.29}$$

Let us estimate the right side. Due to the item e), the result of the procedure (2.26) is not unique, generally speaking. However, the number of different possible outcomes, that is cardinality of $\mathrm{Long}\,(C)$, does not exceed 2^m, where m is the number of those terms of C, which equal 2. Also notice that $\mathrm{chance}(C') = \mathrm{chance}(C)$ whenever $C' \in \mathrm{Long}\,(C)$ because $\mathrm{chance}(1') = \mathrm{chance}(4') = 1$. Thus for any $C \in LMC_k$

$$\sum_{C' \in \mathrm{Long}\,(C)} \mathrm{chance}(C') \;\le\; 2^m \cdot \mathrm{chance}(C) \;\le\; \mathrm{rate}(C),$$

where m has the same meaning. Substituting this into (2.29) we obtain

$$\sum_{C' \in LC_k} \mathrm{chance}(C')) \;\le\; \sum_{C \in LMC_k} \mathrm{rate}(C). \tag{2.30}$$

It remains to prove this:

$$\sum_{k=1}^{\infty} \sum_{C \in \mathrm{LMC}_k} \mathrm{rate}(C) \;\le\; \frac{250 \cdot \beta}{\alpha^2}. \tag{2.31}$$

Instead we shall prove this:

$$\sum_{C \in \mathrm{LMC}} \mathrm{Hshift}(C) \cdot \mathrm{rate}(C) \;\le\; \frac{250 \cdot \beta}{\alpha^2}. \tag{2.32}$$

This is sufficient because the left sides of (2.31) and (2.32) are equal.

For any integer x and y, natural z and $k \in \{1, 2, 3, 4\}$ we denote by $S_k(x, y, z)$ the sum of rates of main codes satisfying conditions LMC-a) and LMC-b) of the definition of legal main codes, whose Hshift equals x, whose Vshift equals y and which have z terms, the last of which is k. It follows from the definition of $S_k(x, y, z)$ and conditions LMC-c), LMC-d) and LMC-e) of (2.25) that

$$\sum_{C \in \mathrm{LMC}} \mathrm{Hshift}(C) \cdot \mathrm{rate}(C) \;\le$$

$$\sum_{x=1}^{\infty} \sum_{y=0}^{\infty} \sum_{z=1}^{\infty} x \cdot \Big(S_3(x, y, z) + S_4(x, y, z) \Big). \tag{2.33}$$

Due to condition LMC-a) of (2.25) , the numbers $S_k(x,y,z)$ satisfy the initial condition

$$S_k(x,y,1) = \begin{cases} 1 & \text{if } \ x=0, \ y=-1 \ \text{ and } \ k=1, \\ 0 & \text{in all the other cases} \end{cases}$$

and due to condition LMC-b) they satisfy the transition equations

$$\begin{cases} S_1(x,y,z+1) = \\ \Big(S_1(x,y+1,z) + S_2(x,y+1,z) + S_3(x,y+1,z) \Big)\,, \\ S_2(x,y,z+1) = 2\beta \quad \times \\ \Big(S_1(x-1,y,z) + S_2(x-1,y,z) + S_3(x-1,y,z) + S_4(x-1,y,z) \Big)\,, \\ S_3(x,y,z+1) = \alpha \quad \times \\ \Big(S_2(x+1,y-1,z) + S_3(x+1,y-1,z) + S_4(x+1,y-1,z) \Big)\,, \\ S_4(x,y,z+1) = (1-\alpha) \quad \times \\ \Big(S_2(x,y-1,z) + S_3(x,y-1,z) + S_4(x,y-1,z) \Big)\,. \end{cases}$$

To estimate (2.33) , let us use sums

$$\begin{cases} S_1(z) \;=\; \sum\limits_{x=-\infty}^{\infty} \sum\limits_{y=-\infty}^{\infty} p^{-x}\, q^{-y}\, S_1(x,y,z), \\ S_2(z) \;=\; \sum\limits_{x=-\infty}^{\infty} \sum\limits_{y=-\infty}^{\infty} p^{-x}\, q^{-y}\, S_2(x,y,z), \\ S_3(z) \;=\; \sum\limits_{x=-\infty}^{\infty} \sum\limits_{y=-\infty}^{\infty} p^{-x}\, q^{-y}\, S_3(x,y,z), \\ S_4(z) \;=\; \sum\limits_{x=-\infty}^{\infty} \sum\limits_{y=-\infty}^{\infty} p^{-x}\, q^{-y}\, S_4(x,y,z), \end{cases} \tag{2.34}$$

where $p,\ q$ are positive parameters, which we need to choose. The following values are sufficient to obtain our estimations:

$$p = 1/3 \quad \text{and} \quad q = 1 - \alpha/6. \tag{2.35}$$

However, it is convenient to keep using letters p and q for a while. Due to our choice of p and q and since $x < 3^x$ for all integer x, the sum (2.33) is estimated by

$$\sum_{z=1}^{\infty} \Big(S_3(z) + S_4(z) \Big), \tag{2.36}$$

so it remains to estimate the sum (2.36) .

The quantities (2.34) satisfy the initial conditions

$$S_1(1) \;=\; q, \quad S_2(1) \;=\; S_3(1) \;=\; S_4(1) \;=\; 0$$

and recurrence conditions

$$\begin{cases} S_1(z+1) = & q\,\Big(S_1(z) + S_2(z) + S_3(z) \Big), \\ S_2(z+1) = & 2\beta/p\,\Big(S_1(z) + S_2(z) + S_3(z) + S_4(z) \Big), \\ S_3(z+1) = & p\alpha/q\,\Big(S_2(z) + S_3(z) + S_4(z) \Big), \\ S_4(z+1) = & (1-\alpha)/q\,\Big(S_2(z) + S_3(z) + S_4(z) \Big). \end{cases}$$

Notice that $S_3(z)$ and $S_4(z)$ are proportional, namely for every z they relate as $p \cdot \alpha$ to $(1-\alpha)$, so we may go to other quantities

$$S_1^*(z) \;=\; S_1(z), \quad S_2^*(z) \;=\; S_2(z), \quad S_3^*(z) \;=\; S_3(z) + S_4(z)$$

with initial conditions

$$S_1^*(1) \;=\; q, \quad S_2^*(1) \;=\; S_3^*(1) \;=\; 0 \tag{2.37}$$

and recurrence conditions

$$\begin{cases} S_1^*(z+1) = & q\,\Big(S_1^*(z) + S_2^*(z) \Big) + p \cdot \alpha/r\, S_3^*(z), \\ S_2^*(z+1) = & 2\beta/p\,\Big(S_1^*(z) + S_2^*(z) + S_3^*(z) \Big), \\ S_3^*(z+1) = & r\,\Big(S_2^*(z) + S_3^*(z) \Big), \end{cases}$$

where we have denoted

$$r = \frac{(1-\alpha) + p \cdot \alpha}{q}. \tag{2.38}$$

Introducing a vector $S^*(z) = (S_1^*(z),\ S_2^*(z),\ S_3^*(z))$, we can write these recurrence conditions as $S^*(z+1) = S^*(z) \cdot M$, whence $S^*(z) = S^*(1) \cdot M^{z-1}$, where M is the matrix

$$M = \left(\begin{array}{ccc} q & 2\beta/p & 0 \\ q & 2\beta/p & r \\ p \cdot \alpha/r & 2\beta/p & r \end{array} \right).$$

Notice that in the spirit of our article we write matrices on the right side of vectors, so vectors are horizontal. Eigen-vectors of M are roots of the equation

$$|M - \lambda_{max} \cdot E| \;=\; 0$$

(where E is the identity matrix), which can be simplified to

$$2\beta \cdot (\lambda^2 - (1-\alpha)) \;=\; p\,\lambda\,(\lambda - q)\,(\lambda - r). \tag{2.39}$$

Let us first consider the case $\beta = 0$. In this case all the eigen-values of M can be written explicitly: they equal q, r and zero and it is easy to show that $q > r > 0$ for all α, so q is the greatest eigen-value.

Now let $\beta > 0$. Remember that $\beta \leq 1/250$. ¿From Perron-Frobenius theorem, M has the "maximal" eigen-value λ_{max}, which is real and positive and which is not less than absolute values of all the other eigen-values of M. If $\beta = 0$, $\lambda_{max} = q$ and it is strictly greater than all the other eigen-values (which are real and non-negative in this case, as we have seen). When β grows from zero to $1/250$, λ_{max} also grows and still exceeds absolute values of all the other eigen-values.

All the components of the eigen-vector V corresponding to λ_{max} can be chosen real and non-negative. In the present case the first component of V is not zero, so we may assume that $V = (V_1,\ V_2,\ V_3)$ is normed in such a way that $V_1 = 1$. Then all the components of our initial vector (2.37) are not greater that the corresponding components of the vector V multiplied by 5/6, because

$$S_1^*(1) \;=\; q \;\leq\; \frac{5}{6}\, V_1, \quad S_2^*(1) \;=\; 0 \;\leq\; \frac{5}{6}\, V_2, \quad S_3^*(1) \;=\; 0 \;\leq\; \frac{5}{6}\, V_3.$$

Hence and from non-negativity of all elements of M,

$$S_i(z) \;\leq\; \frac{5}{6}\, V_i \cdot \lambda_{max}^z \;\text{ for all } z \text{ and } i \quad .$$

Therefore $S_3^*(z) \;\leq\; 5/6 \cdot V_3 \cdot \lambda_{max}^z$, whence we can estimate the sum (2.33) as well as the sum (2.36) as follows:

$$\sum_{z=1}^{\infty} \Big(S_3(z) + S_4(z) \Big) \;=\; \sum_{z=1}^{\infty} S_3^*(z) \;\leq$$

$$\frac{5}{6} \cdot V_3 \cdot \sum_{z=1}^{\infty} \lambda_{max}^z \;=\; \frac{5}{6} \cdot \frac{V_3}{1-\lambda_{max}}. \tag{2.40}$$

To estimate this expression, we need to estimate V_3 from above and $1-\lambda_{max}$ from below. Let us first estimate $1-\lambda_{max}$, for which we need to estimate λ_{max}. From (2.39)

$$\frac{\lambda_{max} - q}{2\beta} \;=\; \frac{\lambda_{max}^2 - (1-\alpha)}{p\,\lambda_{max}\,(\lambda_{max} - r)},. \tag{2.41}$$

To estimate λ_{max} we need to estimate the left side of this expression. First we estimate the numerator of the right side:

$$\lambda_{max}^2 - (1-\alpha) \;\leq\; 1 - (1-\alpha) \;=\; \alpha.$$

Now to estimate the denominator. Since $p \;=\; 1/3$ and $\lambda_{max} \geq q = 1 - \alpha/6 \geq 5/6$,

$$p \cdot \lambda_{max} \;\geq\; 5/18. \tag{2.42}$$

Also notice that $q - r \geq \alpha/3$, whence $\lambda_{max} - r \;\geq\; q - r \;\geq\; \alpha/3$. So we can conclude that

$$p\,\lambda_{max}\,(\lambda_{max} - r) \;\geq\; \frac{5}{18} \cdot \frac{\alpha}{3} \;=\; \frac{5\alpha}{54}.$$

Now we can estimate the right side and therefore the left side of (2.41) :

$$\frac{\lambda_{max} - q}{2\beta} \;\leq\; \frac{\alpha}{5\alpha/54} \;=\; \frac{5}{54}.$$

Since $\beta \leq \alpha^2/250$,

$$\lambda_{max} - q \;\leq\; \frac{2\alpha^2}{250} \cdot \frac{54}{5} \;=\; \frac{54\alpha}{625}.$$

Remember that $q = 1 - \alpha/6$. Therefore

$$1 - \lambda_{max} \;=\; (1-q) - (\lambda_{max} - q) \;\geq\; \frac{\alpha}{6} - \frac{54\alpha}{625} \;=\; \frac{301\alpha}{3750}. \quad (2.43)$$

Thus the denominator of (2.40) is estimated. Now let us estimate the numerator, i.e. V_3, using its explicit representation:

$$V_3 \;=\; \frac{2\beta \cdot r}{p\,\lambda_{max}\Big(\lambda_{max} - (2\beta/p + r)\Big)}. \quad (2.44)$$

It is easy to show that $r \leq 1 - \alpha/2$. Therefore the numerator of (2.44) does not exceed 2β. To estimate the denominator, remember that $\lambda_{max} \geq q = 1 - \alpha/6$ and $2\beta/p = 6\beta \leq 3\alpha^2/125$. Therefore

$$2\beta/p + r \;\leq\; \frac{3\alpha}{125} + 1 - \frac{\alpha}{2}.$$

Using (2.42) , we estimate the denominator:

$$p\,\lambda_{max}\Big(\lambda_{max} - (2\beta/p + r)\Big) \;\geq$$

$$\frac{5}{18} \cdot \left(1 - \frac{\alpha}{6} - \frac{3\alpha}{125} - 1 + \frac{\alpha}{2}\right) \;=\; \frac{58\alpha}{675}.$$

Thus

$$V_3 \;\leq\; \frac{2\beta}{58\alpha/625} \;=\; \frac{675\,\beta}{29\alpha}.$$

Hence and from (2.43) ,

$$\frac{5}{6} \times \frac{V_3}{1 - \lambda_{max}} \;\leq\; \frac{5}{6} \times \frac{675\beta}{29\alpha} \times \frac{3750}{301\alpha} \;\leq\; \frac{242\,\beta}{\alpha^2} \leq \frac{250\beta}{\alpha^2}.$$

The inequality (2.31) is proved. Collecting together the equality (2.15), the inequalities (2.24) and (2.30) summed over k, and (2.31), we prove theorem 2.1.1.

2.7 Proof of Our Theorems for $\alpha = 1$

Now let us prove theorems 2.1.1, 2.1.2, 2.1.3 and 2.1.4 for $\alpha = 1$. It is sufficient to prove that the process μ_t is defined when $\alpha = 1$. Let us denote by μ_{chess} the (unique) measure in $\mathcal{M}$ defined by the condition

$$\mu_{chess}(\ominus,\ \oplus) = \mu_{chess}(\oplus,\ \ominus) = 1/2. \tag{2.45}$$

The operator Ann_1 cannot be applied to μ_{chess}, which was the reason why [27] excluded the case $\alpha = 1$ from consideration. However, Ann_1 can be applied to all the other measures in $\mathcal{M}$. Thus, to include the case $\alpha = 1$, it is sufficient to prove that we never have to apply Ann_α to μ_{chess} in the course of inductive generation of measures μ_t. According to (2.1), Ann_α is always applied after Flip_β. It is evident that

$$\mu(\oplus,\ \oplus) \geq \beta^2 \tag{2.46}$$

for any measure μ, which is a result of application of operator Flip_β. We may exclude the trivial case $\beta = 0$. Then the right side of (2.46) is positive, whence the left side is positive, which is incompatible with the conditions (2.45).

Figure 2.2 ilustrates our mains results in this respect. It shows that in the case $\alpha = 1$ our estimations are tighter than in the case $\alpha < 1$. In more detail:

(1) If $\alpha = 1$, then the frequency of $\oplus$ in the measure μ_t does not exceed $150 \cdot \beta$. for all natural t.

(2) If $\alpha = 1$ and $\beta \geq 0.36$, the measure μ_t tends to $\delta_\oplus$ when $t \to \infty$.

The technical details may be found in [27, 18].

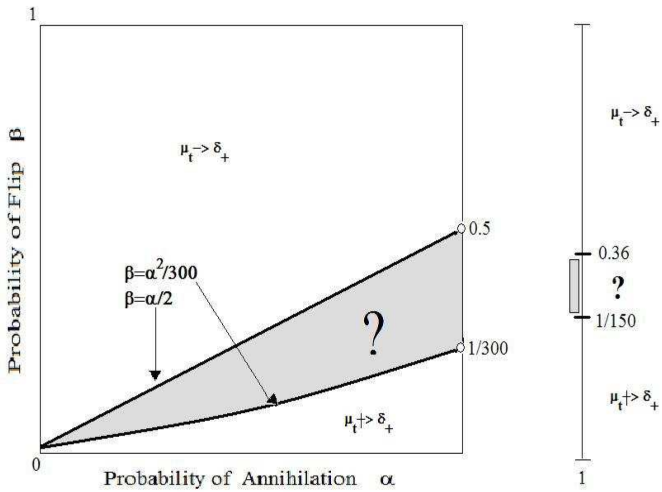

Figure 2.2: *A scheme of our main results about non-ergodicity of Flip-Annihilation process in the cases $\alpha < 1$ and $\alpha = 1$. The gray area is that, where we have no result.*

2.8 Exercices

Exercise 2.8.1. *Prove that for any $\alpha < 1$ the operator* $\mathsf{Ann}_\alpha = \text{Duel}_\alpha \text{ Clean}$ *can be applied to any measure in $\mathcal{M}_{\{\ominus,\oplus\}}$ and turns it into a measure in $\mathcal{M}_{\{\ominus,\oplus\}}$.*

Exercise 2.8.2. *Prove by induction that $\nu_{2t}\,\text{Clean} = \mu_t$ for all t.*

Exercise 2.8.3. *Prove the following items:*

$$
\begin{cases}
\text{a) } \nu(state(x,y) = \ominus) \ > \ 0 \text{ for all } (x,y) \in V. \\
\text{b) For any integer } x_0 \text{ and any natural } y \\
\qquad \nu\Big(\forall\ x \geq x_0 : state(x,y) \neq \ominus\Big) \ = \\
\qquad \nu\Big(\forall\ x \leq x_0 : state(x,y) \neq \ominus\Big) \ = \ 0. \\
\text{c) } \mu_t(\ominus) \ > \ 0 \text{ for all natural } t. \\
\text{d) For any integer } x_0 \text{ and any natural } t \\
\qquad \mu_t\Big(\forall\ x \geq x_0 : s_x \neq \ominus\Big) \ = \\
\qquad \mu_t\Big(\forall\ x \leq x_0 : s_x \neq \ominus\Big) \ = \ 0.
\end{cases}
\tag{2.47}
$$

Exercise 2.8.4. *Prove that for all $\mu \in \mathcal{M}$,*

$$
\begin{aligned}
&(a) \quad \mu(\oplus,\ominus) = \mu(\ominus,\oplus) \leq 1/2; \\
&(b) \quad \mu(\oplus,\ominus) = 1/2 \ \text{ if and only if } \ \mu = \mu_{chess}.
\end{aligned}
$$

Exercise 2.8.5. *Let $k \geq 2$ and $\mu \in \mathcal{M}$. If $\mu(\ominus^k) > 0$ and $\beta > 0$. Then:*

$$
\begin{aligned}
&(a) \quad \mu(\ominus^j) > 0 \ \text{ if } \ 0 < j < k; \\
&(b) \quad \mu(\oplus,\ominus) < 1/2; \\
&(c) \quad \mu\mathsf{Flip}_\beta(\ominus^l) > 0 \ \text{ if } \ 0 < j \leq k.
\end{aligned}
$$

Exercise 2.8.6. *Let $k \geq 2$ and $\mu \in \mathcal{M}$. Prove that*

$$\mu\,\mathrm{Duel}_\alpha(\ominus^{k-1}) \geq \mu(\ominus^k).$$

Exercise 2.8.7. *For any T prove:*

$$\begin{cases} (a)\ \mu_T(\oplus) \ = \ \sum_{k=1}^{\infty} \mu_T(\ominus, \oplus^k). \\ (b)\ \mu_T(\oplus) \ \le \ \dfrac{\mu_T(\oplus)}{\mu_T(\ominus)} = \sum_{k=1}^{\infty} \dfrac{\mu_T(\ominus, \oplus^k)}{\mu_T(\ominus)}. \end{cases}$$

Exercise 2.8.8. *Let $\alpha,\ \beta_1,\ \beta_2 \in [0,1]$ and $\beta_1 < \beta_2$. Prove that*

$$\delta_\ominus \mathsf{Flip}_{\beta_1} \mathsf{Ann}_\alpha(\oplus) \le \delta_\ominus \mathsf{Flip}_{\beta_2} \mathsf{Ann}_\alpha(\oplus).$$

Unsolved problem. Prove an analog of theorem 2.1.1 for a more symmetric process, in which pluses and minuses turn into each other independently with one and the same rate β.

Unsolved problem Let $\alpha \in (0,1]$ be fixed and the values β_1 and β_2 belongs to $(0,1)$. If $\beta_1 < \beta_2$, then

$$\delta_\ominus (\mathsf{Flip}_{\beta_1} \mathsf{Ann}_\alpha)^t(\oplus) \le \delta_\ominus (\mathsf{Flip}_{\beta_2} \mathsf{Ann}_\alpha)^t(\oplus).$$

2.9 M. C. and Chaos Approximate FA

In this and two next sections we use Monte Carlo simulation and Chaos approximations for a numerical study of the process defined in the previous section and called Flip-Annihilation or FA for short. As in previous sections, the *alphabet* $\mathcal{A}$ is a finite non-empty set, whose elements are called *letters*. Throughout this section, $\mathcal{A} = \{\oplus,\ \ominus\}$, that is we have only two letters $\oplus$ and $\ominus$, called *plus* and *minus*.

In the previous section we studied the Flip-Annihilation operator $\mathsf{Flip}_\beta \mathsf{Ann}_\alpha$ depending on two real parameters $\alpha,\ \beta \in [0,\ 1]$. We proved several properties of this operator including the following ones, which we shall call *rigorous estimations*:

a) if $\beta > \alpha/2$, then $\mathsf{Flip}_\beta \mathsf{Ann}_\alpha$ is ergodic.

b) if $\beta < \alpha^2/250$, then $\mathsf{Flip}_\beta \mathsf{Ann}_\alpha$ is not ergodic and has at least two different invariant measures. For the case $\alpha < 1$ these estimations were proved in [27, 28, 18] and for the case $\alpha = 1$ they were proved in [17].

Now we go to computer approximations of this process. In addition to the known facts about it, we assume that whenever β increases, the operator $\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ cannot pass from ergodicity to non-ergodicity. Under this assumption, for every $\alpha \in [0,1]$ there is a value of β in $[0,\ 1]$, which we denote by $\mathsf{true}(\alpha)$, such that the operator $\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ is ergodic for $\beta > \mathsf{true}(\alpha)$ and non-ergodic for $\beta < \mathsf{true}(\alpha)$. We assume also that the function $\mathsf{true}(\alpha)$ is continuous, which allows us to speak about the curve $\beta = \mathsf{true}(\alpha)$, which serves as the boundary between the regions of ergodicity $\beta > \mathsf{true}(\alpha)$ and non-ergodicity $\beta < \mathsf{true}(\alpha)$. We call the set $\{(\alpha,\beta) : \beta = \mathsf{true}(\alpha)\}$ the *true curve.* Of course, the true curve (if it exists) is sandwiched between the rigorous estimations, but they are pretty far from each other; we want better numerical estimations and obtain them using Monte Carlo simulation and Chaos approximation.

2.10 Monte Carlo Approximates FA

Let us describe in detail the Monte Carlo approximation of the FA (Flip-Annihilation) process. Since there are no infinite computers, the Monte Carlo method by its very nature always refers to some finite space case, even when the ultimate motivation is to make conclusions about the infinite case. With this provision, we may approximate any infinite-space process with an auxiliary process, whose space is finite at every step of time but may change as time goes on. We might use *words* (that is finite sequences of letters) for that purpose, but this would necessitate special definitions at the ends when we transform them. That is why we use *circulars* as elements of our countable set Ω of states. By circulars we mean finite sequences of components (like words), but the indices of their components are remainders modulo $|C|$ where $|C|$ is the number of components in a circular C. See Figure 2.3, where $|C| = n$.

For $i = 0, \ldots, n-1$, we say that a word $W = (a_1, a_2, \ldots, a_k)$ *appears* at a place i in a circular $C = (c_1, \ldots, c_n)$ if

$$c_{i+1} = a_1,\ c_{i+2} = a_2, \ldots, c_{i+k} = a_k.$$

We denote by $\mathsf{quant}(W|C)$ the quantity of different places where the word W appears in a circular C. After that we define the *frequency* of W in C as

$$\text{freq}\,(W|C) = \frac{\mathsf{quant}(W|C)}{|C|}. \tag{2.48}$$

We denote by $\mathcal{M}$ the set of probability distributions, that is the set of normalized measures on Ω, the set of circulars. Thus the set, on which the measures $\mu^t \in \mathcal{M}$ are defined, is the set of circulars. In every single computer experiment we obtained a randomly generated sequence of circulars indexed by the values of parameter $t = 0, 1, 2, \ldots$ The circular obtained at time t is denoted by C^t and its components are denoted by C_i^t, where $i = 0, \ldots, |C^t| - 1$. We construct a sequence $C^0, C^1, C^2 \ldots$, which we call a *trajectory* of our process μ^t, where $t = 0, 1, 2, \ldots$. The integer time t grow from zero to N, where N is the maximal number of experiments.

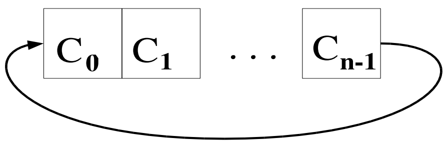

Figure 2.3: *A circular C with $|C| = n$.*

We call a measure $\mu \in \mathcal{M}$ *local* if it is in fact concentrated on a finite subset of Ω. In fact we shall deal only with local measures. For any $\mu \in \mathcal{M}$ we define the *frequency* of the word W according to μ as

$$\text{freq}\,(W|\mu) = \sum_{C \in \Omega} \text{freq}\,(W|C) \cdot \mu(C). \tag{2.49}$$

The main goal of our Monte Carlo simulaion is to decide (that is, to make an educated guess), for which α, β our process is ergodic and for which it is not. We want to make these conclusions based on behavior of the quantities $\text{freq}\,(\oplus|\mu^t)$ when $t \to \infty$. However, due to limitations of our computer facilities, we cannot estimate these quantities directly. So we approximate them by the following quantities:

$$\overline{\mathrm{freq}\,(\oplus|\mu^t)} \stackrel{def}{=} \frac{1}{t}\sum_{k=1}^{t} \mathrm{freq}\,(\oplus|C^k). \tag{2.50}$$

Figure 2.4 shows that the Chaos and Monte Carlo (abbreviated to M.C.) curves are closer to each other than to the rigorous estimations and we conjecture that they are closer to the true curve also.

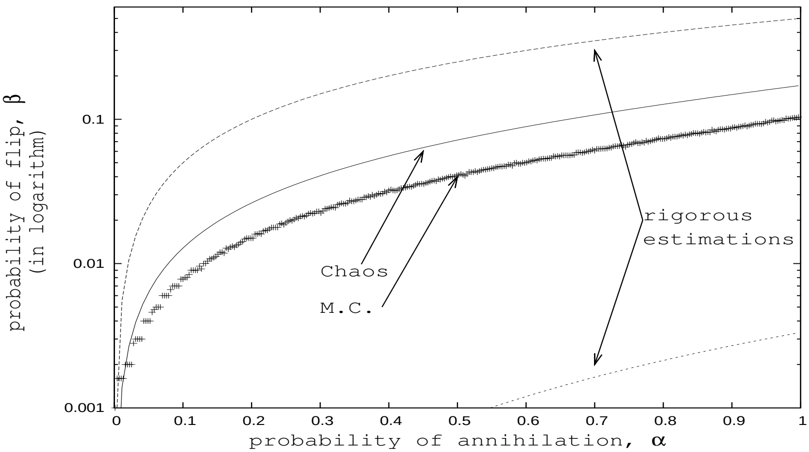

Figure 2.4: *This graph shows both rigorous estimations and the two approximations of* $\mathsf{true}(\alpha)$*: the Chaos approximation (Chaos) and the Monte Carlo approximation (M. C.).*

Our method is based on a procedure, which we call *Imitation.* This procedure generates a sequence of circulars in the following inductive way.

Base of induction. The initial circular C^0 consists of 1000 minuses.

t-th induction step. Given a circilar C^t, where $t = 0,\ 1,\ 2, \dots,$ we perform three procedures:

First procedure imitating the action of flip: every component of C^t, which is a minus, becomes a plus with a probability β

independently from other components. In more technical detail, for every minus in C^t we generate a new random variable distributed uniformly in $(0,\ 1)$ and change this minus into plus if this variable is less than β. We denote the resulting circular by $(C')^t$.

Second procedure imitating the action of annihilation: whenever a component of $(C')^t$, which is a plus, is a left neighbor of a component, which is a minus, both are eliminated from the circular with a probability α independently from other components. In more technical details, for every such pair we generate a new random variable distributed uniformly in $(0,\ 1)$ and perform this elimination if this variable is less than α. We denote the resulting circular by $(C'')^t$.

Third procedure, which allows us to overcome the limitations of computer memory: given $(C'')^t$, we generate a new circular, namely C^{t+1}, in the following way:
if $|(C'')^t| < N_{min}$, where $N_{min} = 500$,, then C^{t+1} is obtained from $(C'')^t$ by concatenating it with its copy and thereby duplicating its length; otherwise we keep $C^{t+1} = (C'')^t$.

When we stop: given a constant $T = 100000$, we stop when $t = T$ or there is none minus in the circular C^t.

Let us explain why we need the third procedure. Remember that under the action of our operator, components can disappear, but not appear; so for any $\beta > 0$ the length of any finite circular decreases in the average and finally degenerates. The third procedure allows us to avoid this and thereby helps us to make our simulation more similar to the infinity process. Thus the procedure Imitation is described.

We use Imitation to attribute an appropriate value to a Boolean variable denoted by E (ergodicity), namely E is set *yes* if the last circular C^t contains none minus; otherwise E is set *no*. If $E = yes$, we interpret this as a suggestion that the process with the given values of α and β is ergodic; the result $E = no$ is taken as a suggestion that our process is non-ergodic.

In fact we used Imitation within a cycle with growing β: we started with $\beta = 0$ and then iteratively performed Imitation and increased β by 0.001 and repeated this until β reached the value 1 or E got the value *yes*, that is ergodicity was suggested. Thus we obtained a certain value of β. In fact, we performed this cycle 5 times and recorded the arithmetical average of the 5 values of β thus obtained.

Remember that all this was done with a certain value of α. In fact we considered 1000 values of α, namely the values $\alpha_i = 0.001 \cdot i$ for $i = 1, \ldots, 1000$. The corresponding recorded value of β was denoted by β_i. Thus we obtained 1000 pairs $(\alpha_i,\ \beta_i)$. The graph called M. C. on Figure 2.4 consists of these pairs plotted.

In [27] we define a function $s(\alpha,\ \beta)$ as the supremum of density of pluses in the measure μ_t over all natural t. For every α the function $s(\alpha,\ \beta)$ has been proved not to be continuous as a function of β. Thus we got a first order phase transition, which is different from contact processes and Stavskaya processes, where the analogous phase transition is second order.

To get a better picture of our first order phase transition, we wanted to estimate $s(\alpha,\ \beta)$ numerically, but to estimate it directly was difficult, so, instead of that, we estimated

$$\overline{s(\alpha,\ \beta)} = \max\{\text{freq}\,(\oplus | C^t) : t = 0, \ldots, 100\,000\}.$$

Figure 2.5 shows the values of $\overline{s(\alpha,\ \beta)}$ in the following way. The area of ergodicity, that is the area, where $\overline{s(\alpha,\ \beta)} = 1$, is white. Other values of $\overline{s(\alpha,\ \beta)}$ are represented by colors according to the rule shown in the color box on the right side. There our approximation suggests non-ergodicity. All the values of $\overline{s(\alpha,\ \beta)}$, which we obtained for all the non-ergodic area, were less or equal to 0.14, which concretizes the non-continuity of $s(\alpha,\ \beta)$ as a function of β.

2.11 Chaos Approximates FA

Let us denote by $\mathcal{M}$ the set of normalizad translation-invariant measures on $\mathcal{A}^{\mathbb{Z}}$. We denote by $\mathcal{C} : \mathcal{M} \to \mathcal{M}$, the well-known *chaos operator* (also called mean-field approximation). Its action amounts to mixing randomly all the components. In other words, for each $\mu \in \mathcal{M}$ the measure $\mu\mathcal{C}$ is a product-measure with the same frequencies of all the letters as μ has. (Like in the previous sections, we write operators on the right side of measures.) The chaos operator allows us to approximate a given process μP^t on the configuration space $\mathcal{A}^{\mathbb{Z}}$ by another process $\mu(\mathcal{C}P)^t$ on the same space: at every time step we apply first $\mathcal{C}$, then P. Thus, instead of the original process, whose set of parameters is infinite or very large, we deal

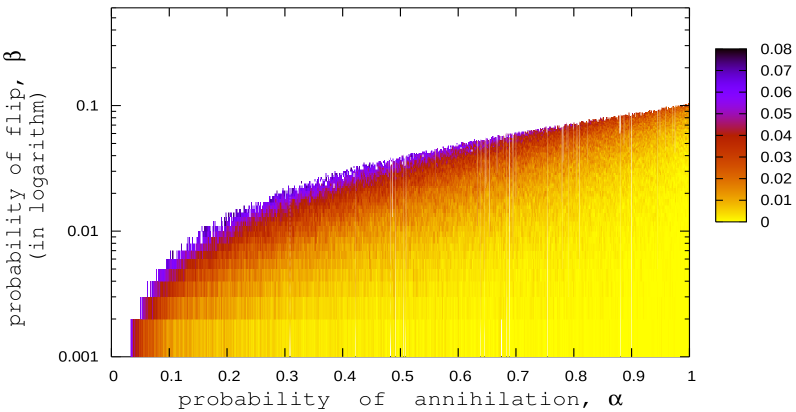

Figure 2.5: *Here we used colors to represent the values of* $\overline{s(\alpha,\ \beta)}$ *in the area, where the process is suggested to be non-ergodic. The color box on the right side shows how colors from yellow to black represent the values of* $\overline{s(\alpha,\ \beta)}$. *For better visualization, we excluded the values greater than* 0.08, *which constitute less than* 1% *of all data.*

with the evolution of densities of letters, that is we deal with a finite set of parameters. Since densities of the letters sum up to one, the number of independent parameters in the chaos approximation equals the number of letters in the alphabet minus one. In our case, with only two letters, we deal with only one parameter: as such we choose the density of pluses. Thus the density of pluses in the measure $\mu\mathcal{C}\,\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ depends only on the density of pluses in the measure μ and this dependence may be expressed by the formula

$$f(x) = \frac{b - \alpha \cdot b(1-b)}{1 - 2\alpha \cdot b(1-b)}, \tag{2.51}$$

where x denotes the frequency of pluses in the measure μ, $f(x)$ denotes the density of pluses in the measure $\mu\mathcal{C}\,\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ and $b = x + (1-x)\beta$. (b is the density of pluses in the measure $\mu\mathcal{C}\,\mathsf{Flip}_\beta$.) Since the maximal value of $b(1-b)$ is $1/4$, the denominator of (2.51) is not less than $1/2$. So $f(t)$ is defined and continuous for $x,\ \alpha,\ \beta \in [0,\ 1]$. Thus the study of the operator $\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ is substituted by a study of the operator $\mathcal{C}\,\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$, which boils down to the study of the one-dimensional dynamical system $f : [0,1] \to [0,1]$ with parameters $\alpha,\ \beta \in [0,\ 1]$. As usual, we call a *fixed point* of this system a value of $x \in [0,\ 1]$ such that $f(x) = x$. We call our dynamical system *ergodic* if it has a unique fixed point x_{fixed} and

$$\forall\ x \in [0,\ 1] : \lim_{t\to\infty} f^t(x) = x_{\text{fixed}},$$

where f^t means the t-th iteration of f.

We conclude that the chaos approximation $\mathcal{C}\,\mathsf{Flip}_\beta\,\mathsf{Ann}_\alpha$ is ergodic if $\beta > \beta^*(\alpha)$ and is not ergodic if $\beta \le \beta^*(\alpha)$, where

$$\beta^*(\alpha) = \begin{cases} \dfrac{4 - \alpha - 2\sqrt{4 - 2\alpha}}{\alpha} & \text{if } \ \alpha > 0, \\ 0 & \text{if } \ \alpha = 0. \end{cases} \tag{2.52}$$

Thus for the chaos approximation we know exactly the curve $\beta = \beta^*(\alpha)$ dividing ergodicity and non-ergodicity: it is continuous, starts at the origin with the slope $1/8$, grows smoothly and reaches $3 - 2\sqrt{2} \approx 0.17$ at $\alpha = 1$. The graph of this curve is labeled "Chaos" in the Figure 2.4.

In fact, we can describe completely the limit behavior of this dynamical system. Let us denote

$$p_1 = \frac{\alpha - 3\alpha\beta - \sqrt{\Delta}}{4\alpha(1-\beta)}, \qquad p_2 = \frac{\alpha - 3\alpha\beta + \sqrt{\Delta}}{4\alpha(1-\beta)}, \qquad p_3 = 1, \quad (2.53)$$

where

$$\Delta = \alpha^2\beta^2 + 2\alpha^2\beta + \alpha^2 - 8\alpha\beta. \quad (2.54)$$

Notice that $p_1,\ p_2,\ p_3$ are real and belong to $[0,\ 1]$. Finally

$$\text{If}\ \ \Delta < 0, \ \ \text{then}\ \ \lim_{t\to\infty} f^t(x_0) = p_3 = 1 \ \ \text{for all}\ \ x_0.$$

$$\text{If}\ \ \Delta = 0, \ \ \text{then}\ \ \lim_{t\to\infty} f^t(x_0) = \begin{cases} p_1 = p_2 & \text{if}\ \ x_0 \le p_1 = p_2, \\ p_3 = 1 & \text{if}\ \ x_0 > p_1 = p_2. \end{cases}$$

$$\text{If}\ \ \Delta > 0, \ \ \text{then}\ \ \lim_{t\to\infty} f^t(x_0) = \begin{cases} p_1 & \text{if}\ \ x_0 < p_2, \\ p_2 & \text{if}\ \ x_0 = p_2, \\ p_3 = 1 & \text{if}\ \ x_0 > p_2. \end{cases}$$

2.12 Annihilation-Flip-Mitosis aka AFM

One great disadvantage of the FA process is that its finite analog shrinks and thereby becomes useless. But we have a remedy: to introduce *Mitosis*, which compensates for shrinking. With this idea in mind, we shall imagine and approximate an infinite process with the infinite space. As in the other cases, we might choose discrete or continuous time and we chose the latter. Keep in mind that the process, which we are going to (informally) describe, has never been defined rigorously, but we are convinced that it can be defined and therefore our work makes sense.

So let us start our (intuitive) description. As continuous time goes on, our sequence (finite or infinite) undergoes the following types of transformation:

- **Annihilation:** $(\oplus, \ominus) \to \Lambda$ and $(\ominus, \oplus) \to \Lambda$. If the states of the components with indices x and $x+1$ are different, both disappear with a rate α independently of the other components. The components $x-1$ and $x+2$ become neighbours. In the finite case the length of the circular decreases by two.

- **Flip:** $\oplus \to \ominus$ and $\ominus \to \oplus$. This changes the state of one component with a rate β independently of the other components. In the finite case the length of the circular does not change.

- **Mitosis:** $\oplus \to \oplus\oplus$ and $\ominus \to \ominus\ominus$. This duplicates one component with a rate γ independently of other components. In the finite case the length of the cicular increases by one.

(2.55)

In the finite case the text presented above may be accepted as a definition, as it was actually done in [12, 13, 14, 15, 16]. In the infinite case we need a definition (which we do not yet have) similar to that for the discrete time, which was provided by the previous sections of this book.

2.13 Monte Carlo Approximates AFM

Now, we shall describe the Monte Carlo approximation of (2.55) . Here we define a procedure, which we call *Imitation*. This procedure generates a sequence of circulars in the following inductive way. (Forget imprefections of computer generated random numbers.)

Base of induction. The initial circular C^0 consists of 1000 minuses.

t-th induction step. Given a circular C^t, where $t = 0,\ 1,\ 2,\ldots$ we performed these three procedures:

The first procedure imitated the random choise of a place where to perform a transformation: a random integer number x distributed uniformly in $\{0, 1, \ldots, |C^t|-1\}$ was generated to identify the position, where the transformation would occur.

The second procedure imitated (2.55): first it generated a real random number ξ distributed uniformly in $(0, 1)$. Then:

- if $\xi \in \left[0, \dfrac{\alpha}{\alpha+\beta+\gamma}\right)$ and $C^t_x \neq C^t_{x+1}$, these components

  annihilated, that is both of them disappeared.

- If $\xi \in \left[\dfrac{\alpha}{\alpha+\beta+\gamma}, \dfrac{\alpha+\beta}{\alpha+\beta+\gamma}\right)$, the component C^t_x changes its

  state from $\ominus$ to $\oplus$ or from $\oplus$ to $\ominus$.

- If $\xi \in \left[\dfrac{\alpha+\beta}{\alpha+\beta+\gamma}, 1\right]$, this component underwent mitosis, that

  is turned into two components in the same state.

If we denote the resulting circular by C^{t+1}, we obtain the induction step of that process, which we have in mind. However, this process cannot yet be implemented on a real computer. That is why we denote the resulting circular by $(C')^t$. Due to presence of annihilation and mitosis, the length of $(C')^t$ may be different from the length of C^t; so in the course of the process the length of our circular changed randomly and usually had a tendency either to shrink most of the time or to grow most of the time. To prevent our process from shrinking to degeneration or growing beyond our computer limitations, we used the third procedure.

The third procedure which helped to keep $C^{\,t}$ within range: given $(C')^t$, we generated a new circular, namely $C^{\,t+1}$, in one of the following ways.

Duplication: if $|(C')^t| < N_{min}$, where $N_{min} = 500$, then $C^{\,t+1}$ was obtained from $(C')^t$ by concatenating it with its copy and thereby duplicating its length.

Cut: if $|(C')^t| > N_{max}$, where $N_{max} = 15,000$, then $C^{\,t+1}$ was obtained from $(C')^t$ deleting half of it.

Otherwise, that is in the case $N_{min} \leq |(C')^t| \leq N_{max}$, we kept $C^{\,t+1} = (C')^t$.

When we stop: We stop our simulation when each one of the three transformations (2.55) occured at least $100,000$ times. Now the procedure Imitation is described.

Now about ergodicity. Since the rules of our process do not change when we swap plus and minus, the ergodicity of our infinite process implies that the frequency of pluses tends to $1/2$. For its finite analog we suggest that this frequency tends to $1/2$ with a probability, which tends to 1 when the length of the initial condition (consisting only of minuses) tends to ∞.

To obtain the small squares on Figure 2.6, approximating the boundary between the regions of ergodicity and non-ergodicity, we used Imitation to attribute an appropriate value to a Boolean variable denoted by E (ergodicity) as follows: if at the end of iteration the quantity $\overline{\text{freq}\,(\oplus|\mu^t)}$ was in the range $(0.45, 0.55)$, we set $E = yes$; otherwise we set $E = no$. We interpreted the result $E = yes$ as a suggestion that the infinite process with the triple $(\alpha,\ \beta,\ \gamma)$ is ergodic; the result $E = no$ was interpreted as a suggestion that this triple produces a non-ergodic process.

To obtain the small circles on Figure 2.6, approximating the boundary between the regions of growth and shrinking, we used Imitation within a cycle with a fixed α/β and growing γ/β: we started with $\gamma/\beta = 0.1$ and then iteratively performed Imitation and increased γ/β by 0.1 and repeated this until γ/β reached the value 8 or there was none duplication in the course of performing Imitation. Thus we obtained a certain value of γ/β. In fact, we performed this cycle 5 times and recorded the arithmetical average of the 5 values of

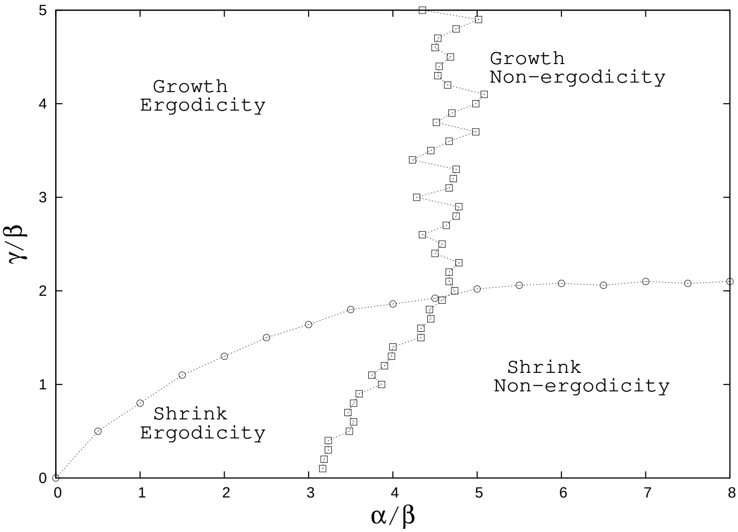

Figure 2.6: *White squares approximate the boundary between suggested ergodicity and suggested non-ergodicity. White balls approximate the boundary between suggested shrinking and suggested growth. Compare this figure with figure 2.10.*

γ/β thus obtained. All this was done with 17 values of α/β, namely the values $0.5 \times i$ with $i = 0, \ldots, 16$. Thus we obtained 17 pairs $(\alpha/\beta, \gamma/\beta)$ represented by small circles on Figure 2.6.

Notice that if we multiply α, β and γ by one and the same positive number, the process does not change. So only ratios of these three numbers to each other are important for us. We used Imitation within a cycle with a fixed γ/β and growing α/β: we started with $\alpha/\beta = 0.1$ and then iteratively performed Imitation and increased α/β by 0.1 and repeated this until α/β reached the value 8 or E got the value *no*. Thus we obtained a certain value of α/β. In fact we performed this cycle 5 times and recorded the arithmetical average of the 5 values of α/β thus obtained. All this was done for 50 values of γ/β, namely the values $0.1 \times i$ with $i = 1, \ldots, 50$. Thus we obtained 50 pairs $(\alpha/\beta, \gamma/\beta)$ represented by small squares on Figure 2.6.

2.14 Chaos Approximates AFM

Remember our procedure of Imitation. Let us imagine that at every step of this procedure, in addition to the operations described above (before or after them - it does not matter), all the components of C^t are randomly permuted. This is chaos approximation, which loses any spacial structure, so when we use it, we deal only with quantities of particles in every possible state. In the present case we have only two possible states for any particle: *plus* or *minus*. So behavior of the resulting process essentially has only two parameters: quantity of pluses and quantity of minuses at time t, which we denote by $X(t)$ and $Y(t)$. When $X(t)$ and $Y(t)$ are large, we may approximatedly treat them as if they were real. In this approximation, we obtain a deterministic process described by differential equations.

Let us examine the behavior of the densities x and y as a limit of the original process with $X(t)$ and $Y(t)$ tending to infinity. Let us consider the vector of densities, $u = (x, y)$ and its increment $\Delta u = (\Delta x, \Delta y)$.

The vector $v = u + \Delta u$ describes the new amounts after a time step Δt. Since thus obtained v is not (generally) normalized, we normalize v, thus obtaining a normalized vector w:

$$w = \frac{v}{|v|} = \left(\frac{x + \Delta x}{1 + \Delta x + \Delta y}, \quad \frac{y + \Delta y}{1 + \Delta x + \Delta y} \right) \tag{2.56}$$

where $|.|$ denotes the sum of componentes. Figure 2.7 illustrates that special case of normalization, which we have just described.

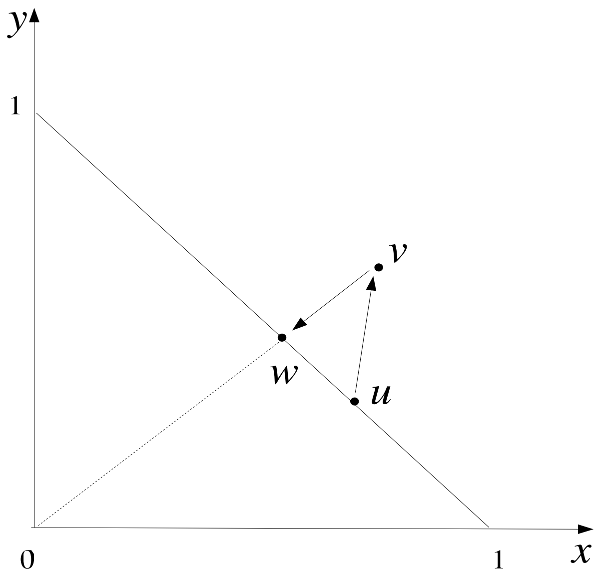

Figure 2.7: *An illustration of normalizaion.*

We assume that $\Delta t \to 0$, $\Delta x = O(\Delta t)$ and $\Delta y = O(\Delta t)$. Thus, $o(\Delta x)$ and $o(\Delta y)$ are $o(\Delta t)$. Therefore

$$
\begin{aligned}
& \frac{x+\Delta x}{1+\Delta x+\Delta y}-x \\
= \quad & \frac{x+\Delta x-x-x\Delta x-x\Delta y}{1+\Delta x+\Delta y} \\
= \quad & \frac{(1-x)\Delta x-x\Delta y}{1+\Delta x+\Delta y} \\
= \quad & \frac{y\Delta x-x\Delta y}{1+\Delta x+\Delta y} \\
= \quad & (y\Delta x-x\Delta y)\cdot(1-\Delta x-\Delta y+o(\Delta x+\Delta y)). \qquad (2.57)
\end{aligned}
$$

But, by our assumption

$$o(\Delta x+\Delta y)=o(\Delta x)+o(\Delta y)=o(\Delta t).$$

Also, Δx and Δy are $O(\Delta t)$. Thus

$$\Delta x\cdot\Delta y=\Delta x\cdot\Delta x=\Delta y\cdot\Delta y=O(\Delta t^2)=o(\Delta t).$$

So, the product in (2.57) can be rewritten as:

$$
\begin{aligned}
& (y\Delta x-x\Delta y)-(\Delta x+\Delta y)\cdot(y\Delta x-x\Delta y)+o(\Delta t)\cdot(y\Delta x-x\Delta y)= \\
& =(y\Delta x-x\Delta y)-O(\Delta t^2)\cdot(y-x)+o(\Delta t)\cdot O(\Delta t)\cdot(y-x) \\
& =(y\Delta x-x\Delta y)-o(\Delta t)\cdot(y-x)+o(\Delta t^2)\cdot(y-x) \\
& =y\Delta x-x\Delta y+o(\Delta t).
\end{aligned}
$$

Dividing the expression (2.58) by Δt and after that making $\Delta t\to 0$, we get

$$\left.\frac{dx}{dt}\right|_{x+y=1}=y\cdot\frac{dx}{dt}-x\cdot\frac{dy}{dt} \qquad (2.58)$$

Analogously, we compute $\left.\frac{dy}{dt}\right|_{x+y=1}$ so that,

$$\left.\frac{dx}{dt}\right|_{x+y=1} + \left.\frac{dy}{dt}\right|_{x+y=1} = 0.$$

In this approximation, we obtain a deterministic process described by the differential equations (2.59):

$$\left.\begin{aligned} \frac{dX(t)}{dt} &= -\beta \cdot X(t) + \beta \cdot Y(t) + \gamma \cdot X(t) - \alpha \cdot \frac{X(t)Y(t)}{X(t)+Y(t)}, \\ \frac{dY(t)}{dt} &= -\beta \cdot Y(t) + \beta \cdot X(t) + \gamma \cdot Y(t) - \alpha \cdot \frac{X(t)Y(t)}{X(t)+Y(t)}. \end{aligned}\right\} \tag{2.59}$$

The last term in each formula is based on our assumption that all the components are mixed all the time, whence the neighbor components are independent from each other all the time. Also notice that in this case multiplying α, β and γ by one and the same positive number does change the process, but in a very special way: it only slows it down or speeds it up. This allows us to use only ratios α/β and γ/β in the Figure 2.10.

Since the process (2.59) is homogeneous, we deal in fact with a two-dimensional analog of the theorem on p. 7 of [1]. So we may go to other variables

$$S(t) = X(t) + Y(t) \quad \text{and} \quad B(t) = \frac{X(t) - Y(t)}{X(t) + Y(t)} \tag{2.60}$$

For simplicity, sometimes we shall denote $X(t), Y(t), S(t)$ and $B(t)$ by X, Y, S and B respectively. The following system of equations is equivalent to (2.59):

$$\frac{dS}{dt} = S \cdot \left(\gamma - \frac{\alpha}{2}\left(1 - B^2\right)\right), \tag{2.61}$$

$$\frac{dB}{dt} = B \cdot \left(\frac{\alpha}{2}\left(1 - B^2\right) - 2\beta\right). \tag{2.62}$$

The last equation is easy to solve explicitly, but we shall get all we need by qualitative arguments. Since we are especially interested in the proportion of each type of particles, we consider also another process, which we call *normalized chaos approximation*:

$$X_{\text{norm}}(t) = \frac{X(t)}{X(t)+Y(t)}, \quad Y_{\text{norm}}(t) = \frac{Y(t)}{X(t)+Y(t)}. \tag{2.63}$$

Then

$$X_{\text{norm}}(t) = \frac{1+B(t)}{2} \quad \text{and} \quad Y_{\text{norm}}(t) = \frac{1-B(t)}{2}.$$

Thus, all we need to do is to study the behavior of $B(t)$. Let us treat the equation (2.62) as a deterministic dynamical system with a space $[-1,\ 1]$ and continuous time t. We call a number $B^* \in [-1,\ 1]$ a *fixed point* of this system if (2.62) equals zero at $B = B^*$. We say that a fixed point $B^* \in [-1,\ 1]$ attracts a point $B \in [-1,\ 1]$ if the process (2.62) starting at $B(0) = B$ tends to B^* when $t \to \infty$. Given a fixed point, we call its *basin of attraction* or just *basin* the set of points attracted by it. We call the process *ergodic* if there is only one fixed point B^* and $B(t)$ tends to B^* for any initial value when $t \to \infty$. Otherwise, we call the process *non-ergodic*. It is easy to describe completely fixed points and their basins for (2.62) . The right side of (2.62) equals zero at three (generally complex) values of B, which we denote by

$$B_1^* = -\sqrt{1 - \frac{4\beta}{\alpha}}, \qquad B_2^* = 0, \qquad B_3^* = \sqrt{1 - \frac{4\beta}{\alpha}}. \tag{2.64}$$

Hence follows our classification:

If $\alpha/\beta < 4$, then B_1^* and B_3^* are not real and the right side of (2.62) is

$$\begin{cases} \text{positive when} & B \in [-1,\ 0), \\ \text{zero when} & B = 0, \\ \text{negative when} & B \in (0,\ 1]. \end{cases}$$

Figure 2.8 illustrates this.

Therefore, in this case $B(t)$ tends to zero from any initial value when $t \to \infty$.

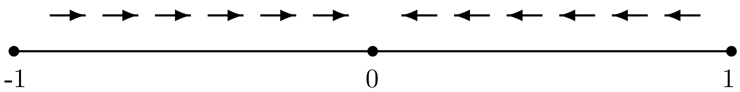

Figure 2.8: *Behavior of $B(t)$ when $\alpha/\beta < 4$.*

If $\alpha/\beta = 4$, then B_1^* and B_3^* are real and equal to zero. The signs of the right side of (2.62) are the same as in the previous case and $B(t)$ also tends to zero from any initial condition when $t \to \infty$.

If $\alpha/\beta > 4$, then B_1^* and B_3^* are real and

$$-1 < B_1^* < B_2^* = 0 < B_3^* < 1$$

(remember that $\beta > 0$). So the right side of (2.62) is

$$\begin{cases} \text{positive when} & B \in [-1,\ B_1^*), \\ \text{zero when} & B = B_1^*, \\ \text{negative when} & B \in (B_1^*,\ B_2^*), \\ \text{zero when} & B = B_2^* = 0, \\ \text{positive when} & B \in (B_2^*,\ B_3^*), \\ \text{zero when} & B = B_3^*, \\ \text{negative when} & B \in (B_3^*,\ 1]. \end{cases}$$

Figure 2.9 illustrates this case.

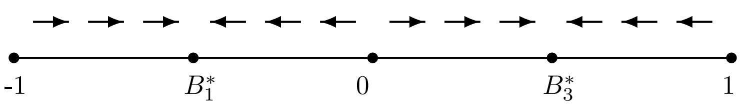

Figure 2.9: *Behavior of $B(t)$ when $\alpha/\beta > 4$.*

Therefore in this case $B(t)$ tends to B_1^* or B_2^* or B_3^* from any initial condition when $t \to \infty$. So $[-1,\ 1]$ is a union of these three basins:

$$basin(B_1^*) = [-1,\ 0), \quad basin(B_2^*) = \{0\}, \quad basin(B_3^*) = (0,\ 1].$$

Thus the normalized chaos approximation is ergodic if $\alpha/\beta \leq 4$ and non-ergodic if $\alpha/\beta > 4$.

Now we are ready to study the chaos approximation (2.59) . Let us remember that $X(t) + Y(t) = S(t)$ and say that our dinamical system:

- *grows* if $S(t)$ tends to infinity when $t \to \infty$.
- *shrinks* if $S(t)$ tends to zero when $t \to \infty$.

Let us find out when it grows and when it shrinks.

Notice that we may rewrite (2.61) as

$$\frac{d \ln S}{dt} = \gamma - \frac{\alpha}{2}(1 - B^2). \tag{2.65}$$

Let us denote by $G(B)$ the right side of (2.65) .

Given two positive functions f_1 and f_2 of $t \geq 0$, let us write $f_1 \asymp f_2$ if $f_1 = O(f_2)$ and $f_2 = O(f_1)$.

Lemma 2.14.1. *Let $B(0) \in basin(B_i^*)$, where $i \in \{1, 2, 3\}$. Then:*

If $G(B_i^*) > 0$, then $\ln S(t) \asymp t$.

If $G(B_i^*) = 0$, then $|\ln S(t)| = o(t)$.

If $G(B_i^*) < 0$, then $-\ln S(t) \asymp t$.

Proof: evident.

Figure 2.10 resumes our findings.

In the special case when $X(0) = Y(0)$ we have $B(t) = 0$ for all t. But zero is a fixed point and $B(t) \to 0$ when $t \to \infty$ for all initial values. So the process is ergodic.

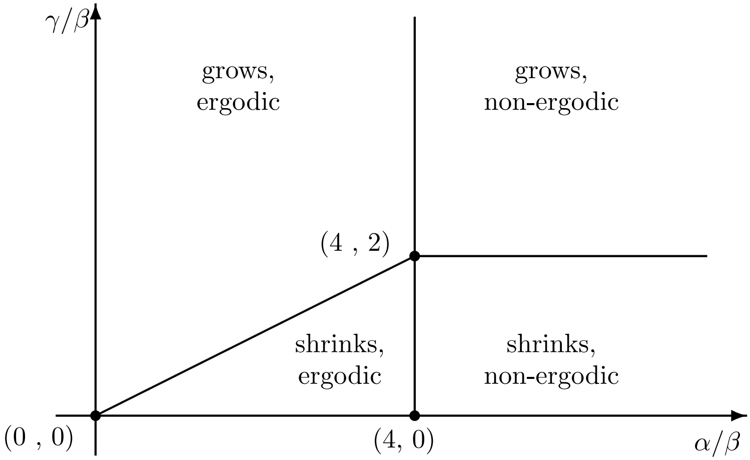

Figure 2.10: *Classification for $X(0) \neq Y(0)$. Compare this figure with figure 2.6*

.

2.15 Exercices

Exercise 2.15.1. *Let $v \in \mathbb{Z}$ and $s_v \in \{0,\ 1\}$. The Stavskaya process is a discrete-time version of contact processes. Stavskaya operator is defined as a composition of two operators: the first one is deterministic and is defined by the rule $(sD)_v = min(s_v,\ s_{v+1})$. The other one, R_α, is random and transforms any zero into one with probability α independently from others.*

Write and study the chaos approximation of Stavskaya process. Show that the chaos approximation presents a kind of phase transition between ergodicity and non-ergodicity.

Exercise 2.15.2. *Let $v \in \mathbb{Z}^2$ and $s_v \in \{0,\ 1\}$. Consider the following process with discrete time: the deterministic operator D is defined by the formula*

$$(Ds)_{(0,0)} = med(s_{(0,0)},\ s_{(0,1)},\ s_{(0,2)})$$

and the random operator R_α transform every zero into one with probability α independently from each other.

(a) Show that this process is ergodic whenever $\alpha > 0$.

(b) Show that the chaos approximation of this process presents a phase transition between ergodicity and non-ergodicity.

Exercise 2.15.3. *Let us consider the following process with variable length. Every particle has two possible states, minus and plus. At every step of the discrete time two transformations occur. The first one turns every minus into plus with probability β independently from what happens at other places and thereby favors pluses against minuses. The second one: whenever a plus is a left side neighbor of a minus, the plus disappears with probability α independently from what happens at other places.*

Write a chaos approximation to this process and show that this chaos approximation presents a kind of phase transition between ergodicity and non-ergodicity. Propose a computational model for this process.

We cordially thank the organizers of the 28-o Colóquio Brasileiro de Matemática, whose invitation stimulated this work.

Bibliography

[1] Vladimir I. Arnold. *Geometrical Methods in the Theory of Ordinary Differential Equations.* Springer-Verlag, (1983) New York Inc.

[2] Discrete local Markov systems. A. Toom, N. Vasilyev, O. Stavskaya, L. Mityushin, G. Kurdyumov and S. Pirogov. *Stochastic Cellular Systems : ergodicity, memory, morphogenesis.* Ed. by R. Dobrushin, V. Kryukov and A. Toom. Nonlinear Science: theory and applications, Manchester University Press, 1990, pp. 1-182.

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

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

[5] Pedro Ferreira de Lima and André Toom. *Dualities Useful in Bond Percolation. Cubo,* vol. 10, n. 3 (2008), pp. 93-102.

[6] Peter Gács. *Reliable cellular automata with self-organization.* Journal of Stat. Physics, Vol. 103 (2001),pp. 45-267.

[7] Lawrence F. Gray. *A Reader's guide to Gac's "Positive Rates" paper.* Journal of Stat. Physics,Vol. 103 (2001),pp. 1-44.

[8] Svante Janson (2007). Private communication.

[9] Thomas M. Liggett. *Interacting Particle Systems.* (1985) Springer.

[10] Thomas M. Liggett. *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes.* Springer, 1999.

[11] Christian Maes. *New Trends in Interacting Particle Systems.* Markov Proc. Rel. Fields. vol. 11 (2005), n. 2, pp. 283-288.

[12] Vadim Malyshev, (1997): *Interacting strings of symbols.* Uspekhi Mat. Nauk, **52**, 59–86 (in Russian). English translation: Russian Math. Surveys, **52**, 299–326.

[13] Vadim Malyshev, (1998): *Stochastic Evolution via Graph Grammars.* Diskret. Mat., **10**, 30–44 (in Russian). English translation: Discrete Math. Appl., **8**, 247–262.

[14] Vadim Malyshev, (1998): *Random grammars.* Russ. Math. Surv. **53** 345–370.

[15] Vadim Malyshev. *Quantum Grammars. Part 4.1: KMS states on Quantum Grammars.* Preprint INRIA, 1999, no. 3702, pp. 1-20 and Journal of Mathem. Physics, v. 41 (2000), n. 7, pp. 4508-4520.

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

[17] Alex D. Ramos. *Particle process with variable length.* Ph.D. Thesis. Federal University of Pernambuco, Departament of Statistics, Recife, Pernambuco, Brazil, 2007. (In Portuguese with an abstract in English). Avaliable at `http://de.ufpe.br/~toom/ensino/doutorado/alunos/index.htm`.

[18] Alex D. Ramos and André Toom (2008). *An error correction. Letter to the editor.* Journal of Stat. Physics, **131**, n. 1, april 2008, pp. 167-168.

[19] Alex D. Ramos and André Toom. *Chaos and Monte Carlo Approximations of the Flip-Annihilation process.* Journal of Stat. Physics, v. 133, 2008, p. 761-771.

[20] Alex D. Ramos and André Toom. *Nonergodicity and growth are compatible for 1D local interaction.* Brazilian Journal of Probability and Statistics, v. 24, p. 400-412, 2010.

[21] Andréa V. Rocha (2009): *Substitution operators.* Doctoral thesis, UFPE. In English. Available at `http://www.de.ufpe.br/~toom/alunos/doutorado/andrea/`

[22] Andréa Rocha, Alexandre Simas and André Toom (2011). *Substitution operators.* Journal of Stat. Physics, vol. 142, n. 3, pp. 565-618.

[23] Olga N. Stavskaya and Ilya I. Piatetski-Shapiro. *On homogeneous nets of spontaneously active elements.* Systems Theory Res., 1971, v. 20, pp. 75-88.

[24] André Toom. *A Family of Uniform Nets of Formal Neurons. Soviet Math. Doklady*, 1968, v.9, n.6, pp. 1338-1341.

[25] André Toom. *Cellular Automata with Errors: Problems for Students of Probability.* Topics in Contemporary Probability and its Applications. Ed. J. Laurie Snell. Series *Probability and Stochastics* ed. by Richard Durrett and Mark Pinsky. CRC Press, 1995, pp. 117-157.

[26] André Toom. *Particle systems with variable length.* Bulletin of the Brazilian Mathematical Society, v. 33, n. 3, november 2002, pp. 419-425.

[27] André Toom. *Non-Ergodicity in a 1-D Particle Process with variable Length.* Journal of Stat. Physics, vol. 115, nn. 3/4, 2004, pp. 895-924.

[28] André Toom. *Every continuous operator has an invariant measure.* Journal of Stat. Physics, vol. 129, n. 2007, pp. 555-566.

[29] Wikipedia (2010): *Damerau-Levenshtein distance.*

[30] Wikipedia (2010): *Indel.*

[31] Andrei A. Zalizniak. *On professional and amateur linguistics.* A lecture at the festival of science. October 11, 2009. (In Russian.) Available at
`http://lingvofreaks.narod.ru/zaliznyak.htm`