

O Problema Generalizado do Dominó

Kayo Douglas da Silva

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Estatística

Orientador: Prof. Dr. Miguel Natalio Abadi

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES

São Paulo, agosto de 2019

O Problema Generalizado do Dominó

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 17/02/2020. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Miguel Natalio Abadi (orientador) - IME-USP
- Prof. Dr. Jerome François Alain Jean Rousseau - UFBA
- Prof. Dr. Rodrigo Lambert - UFU

Agradecimentos

Primeiramente aos meus pais e minha irmã por me fornecerem as lições e modelo de integridade, justiça e dignidade que guiam a maioria das minhas decisões através da vida.

Aos meus professores, da educação regular à pós-graduação, pela inspiração e iluminação nos difíceis, porém libertadores, caminhos do conhecimento. Menção especial a Anatoli Iambartsev, Marcia Branco e Victor Fossaluzza, professores do IME-USP, pelos conselhos que me foram de grande importância nesses dois anos de curso.

Também seria um emprego exagerado da imaginação pensar em como o mestrado começaria sem o apoio de Rafael Grisi (que me ajudou a desenrolar algumas contas deste trabalho), Daniel Miranda (que foi o primeiro a saber que eu viria para o programa de mestrado do IME), Aleja Rada (que além de me aguentar como uma santa, foi quem me apresentou à área). Cito também Marcus Marrocos e Ana Boero, que me deram apoio nos momentos em que achei que não chegaria ao fim do curso. De forma geral, ao corpo docente da Universidade Federal do ABC, meu porto seguro.

Aos meus amigos da Universidade de São Paulo. Esther e Daiane, minhas atletas favoritas que me ensinaram o que era necessário para sobreviver nessa cidade dentro da cidade. Também a Gabriel, Danilo, Luise e o inesperadamente alto número de alunos de Ciências Biológicas com quem almocei nesses dois anos.

Aos meus colegas do Instituto de Matemática e Estatística, em especial a Julia Faria Codas, que não queria pegar uma matéria sozinha e acabou me dando um orientador. Não fosse a capacidade dela de acompanhar o raciocínio dos dois homens mais confusos dessa instituição, esse trabalho seria uma tragédia. Não fosse a capacidade dela de me acalmar e de ensinar todo tipo de truque útil, além das suas impressionantes demonstrações de amizade, a vida nesse dois anos teria muito menos graça. Menção Honrosa a José Augusto, Lucas Prates e Adriana Coutinho.

Aos meus amigos de fora da USP e celebro o apoio vital de cada um. Cabe aqui uma citação especial a Sarah Meneghetti, pela terapêutica companhia, pelo ótimo gosto para especiais de Stand Up da Netflix e a paciência *sui generis* quando eu faço coisas como botar um latim desnecessário no meio da frase. Incluo Marina, Djarrá, Beatriz Felix, Beatriz Nogueira, Aurora, Lucas, Raul, Carolina, Amanda, Daniela e Jorge. Espero que possam perdoar minha ausência e que eu consiga compensá-la e apoiá-los em cada desafio que a vida achar que pode impor a vocês.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, a CAPES, pelo apoio financeiro durante a realização desse trabalho. Espero que os leitores do futuro, que passarem por essa página pra entender se eu já fui um ser humano com emoções antes de virar um professor robô do mal,

se surpreendam ao saber que estes são tempos em que não se pode exagerar a importância dessa agência na vida dos pesquisadores.

À comissão Julgadora, Prof. Dr. Jerome Rousseau e Prof. Dr. Rodrigo Lambert, por examinarem esse trabalho e guiarem nos aspectos em que ele poderia melhorar. Finalmente, ao maior torcedor do Belgrano de Córdoba que eu conheço, o meu orientador Miguel Abadi. Por buscar entender cada uma das variáveis que eu criei nesse trabalho. Por me ajudar a eliminar pelo menos metade delas e, através de sua paciência e sabedoria, ter se tornado o maior apoiador do meu trabalho.

Nunca serei, em toda a minha existência, capaz de agradecê-los o suficiente, mas posso começar agora: muito obrigado!

Resumo

SILVA, K. D. **O Problema Generalizado do Dominó**. 2020. 40 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

Em um jogo de dominó clássico, com 7 símbolos possíveis e 2 símbolos por peça, qual a probabilidade de que duas peças escolhidas ao acaso e sem reposição se encaixem? A resolução desse problema prevê separar o conjunto de peças entre aquelas que possuem dois símbolos iguais (chamaremos de simétricas) e as que possuem símbolos distintos (não-simétricas) e calcular a probabilidade em cada caso. Numa generalização do problema, temos uma quantidade a de símbolos possíveis tomados de um conjunto finito \mathcal{A} qualquer e peças constituídas já não de 2, mas de n símbolos tomados desse conjunto. Estamos interessados em encontrar, dadas duas peças escolhidas sem reposição, o tamanho máximo do encaixe entre elas.

Palavras-chave: Sobreposição, Máximo Encaixe, Tempo de Chegada, Espaços Produto, Convergência.

Abstract

SILVA, K. D. **The Generalized Domino Problem**. 2020. 40 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

In a classic domino game with 7 possible symbols and 2 of them per piece, what is the probability that two randomly chosen pieces without replacement will fit together? The solution to this problem requires to separate the set of pieces between those that have two equal symbols (we will call them symmetrical) and those that have different ones (non-symmetrical) and calculate the probability in each case. In a generalization of the problem, we have a quantity a of possible symbols taken from any finite set \mathcal{A} and pieces no longer consisting of 2 but n symbols taken from that set. We are interested in finding, given two pieces chosen without replacement, the maximum fitting size between them.

Keywords: Overlap, maximum fitting size, hitting time, product spaces, convergence.

Sumário

Lista de Figuras	ix
1 A Construção da Medida	5
1.1 Alfabetos e palavras	5
1.2 Conjuntos e medidas	7
1.3 O espaço de pares	9
1.3.1 O problema do encaixe	10
2 A probabilidade de um encaixe	13
2.1 O que importa é a não-simetria	13
2.2 Modos de encaixe	15
2.3 Convergência	17
3 A Distribuição do Máximo Encaixe	21
3.1 Definições	21
3.2 Crescimento assintótico dos diferentes tipos de encaixe	23
3.3 Um minorante para a cauda da distribuição do máximo encaixe	28
3.4 Comparação de valores assintóticos	29
3.5 Análise do Termo dominante	31
Referências Bibliográficas	35

Lista de Figuras

- 3.1 Comparação dos termos de convergência a_k, c_k e d_k para tamanhos diferentes de \mathcal{A} . 30
- 3.2 Comparação entre b_k (em verde), c_k (em vermelho) e d_k (em azul) para tamanhos diferentes de \mathcal{A} 32

Introdução

Em um jogo de Dominó clássico, se duas peças são escolhidas ao acaso, qual a probabilidade de elas se encaixarem? A resolução deste problema prevê separar o conjunto de peças entre aquelas que possuem símbolos iguais e aquelas que possuem símbolos diferentes. Neste trabalho, consideramos a generalização deste problema. Um alfabeto \mathcal{A} finito qualquer e um tamanho n que representa o comprimento das peças que são formadas pela concatenação de n símbolos do alfabeto. Escolhidas duas peças quaisquer do conjunto sem reposição, qual a probabilidade de obter-se um encaixe de tamanho k ? Neste trabalho descobrimos que a distribuição exata é não-trivial e usamos de simulações para determinar o comportamento dela.

Este problema de caráter probabilístico e puramente combinatorial tem motivação em um problema vindo da Teoria de Recorrência de Poincaré e Processos Estocásticos para duas sequências quaisquer. O problema foi considerado por [Abadi e Lambert \(2019\)](#), que estudaram o tamanho da sobreposição entre duas sequências com orientação definida, ou seja, sem contar as inversões das sequências. Ainda mais, eles provaram um resultado baseado nas flutuações da variável

$$T_n^{(2)}(x, y) = \inf\{k \geq 1 : y_0^{n-1} \cap \sigma^{-k}(x_0^{n-1}) \neq \emptyset\}$$

isto é, provaram a convergência da distribuição de $T_n^{(2)}$ devidamente reescalada para uma distribuição limite não-trivial, especificamente, provaram que

$$\frac{T_n^{(2)}}{n} \xrightarrow{n \rightarrow \infty} 1$$

No mesmo artigo, os autores provaram um princípio de grandes desvios para $T_n^{(2)}$.

Estamos interessados em um problema que, em princípio, se reduz a um problema combinatorial mas que descreve uma espécie de reversibilidade do problema considerado pelos autores. Queremos dizer especificamente que consideramos tanto o menor caminho de x para y como o menor caminho de y para x . Ainda mais, se \tilde{x} e \tilde{y} as palavras reversas de x e y , respectivamente, consideramos o menor caminho de x para \tilde{y} e de y para \tilde{x} . Nos interessa, enfim, o mínimo entre essas quatro formas de ir de um estado a outro do processo. A primeira impressão nos diz que, se o tamanho das sequências x e y é grande, os possíveis encaixes são independentes, mas ficará claro que não é esse o caso. Quando comparamos o tamanho do encaixe entre x e y e entre x e \tilde{y} , por exemplo, estamos utilizando a mesma sequências x , de maneira que as quatro formas de encaixe apresentam correlações entre elas. O artigo de [Abadi e Lambert \(2019\)](#) tem seus resultados predecessores. Ainda no contexto da Teoria de Recorrência de Poincaré se considera este tipo de problema no caso em

que o ponto de partida e o de chegada correspondem à mesma sequência x , como estudado em [Abadi e Lambert \(2013\)](#). Consideramos, então, a seguinte definição:

$$T_n(x) = \min\{k \geq 1 | x_0^{n-k-1} = x_k^{n-1}\}$$

O primeiro resultado a respeito de $T_n(x)$ foi apresentado por [Afraimovich et al. \(2003\)](#) e concomitantemente por [Saussol et al. \(2002\)](#), em que eles provam que T_n/n converge para 1 quase certamente. Princípios de grandes desvios foram provados para $T_n(x)$ sucessivamente por [Abadi e Vaienti \(2008\)](#), logo depois por [Haydn e Vaienti \(2010\)](#) e, finalmente, por [Abadi e Cardeno \(2015\)](#).

A convergência da distribuição de T_n para uma distribuição limite foi provada primeiramente por [Abadi e Lambert \(2013\)](#) e posteriormente por [Abadi et al. \(2018\)](#). Essa convergência em distribuição foi provada ainda para o caso em que a medida é uma medida produto em [Abadi e Lambert \(2019\)](#) e logo estendido para medidas misturadoras em [Abadi et al. \(2018\)](#). Nesses textos as sequências x e y são escolhidas de forma independente. Nosso interesse é considerar a escolha sem reposição de maneira que a segunda peça dependa da primeira. Nos interessa representar a ideia de que a escolha de cada x é feita de forma uniforme entre o total de peças disponíveis de maneira que a descrição de uma medida que representa essa ideia é parte importante desde trabalho. A ideia da sutileza que devemos ter na escolhas das sequências simétricas e não-simétricas. Note que num espaço \mathcal{A}^n temos uma única representação de uma sequência simétrica mas temos duas representações de uma sequência não-simétrica. Devemos, portanto, definir uma medida que represente apenas o espaço de peças de tamanho n em que não exista ordenamento.

O trabalho será estruturado da seguinte forma:

No capítulo 1 construiremos três espaços: um para a primeira peça, um para a segunda (dado que conhecemos a primeira) e um para o par de peças. Definiremos também as respectivas medidas que se pretendem uniformes sobre esses espaço. Neste mesmo capítulo, definiremos o que chamaremos de "modos de encaixe";

No capítulo 2 calcularemos a probabilidade de um par de palavras no espaço de pares possuir um encaixe de tamanho k positivo, baseando-nos nos quatro modos de encaixe. Mostraremos que essa possibilidade converge, quando n diverge, para um certo a_k dado por

$$a_k = 4a^{-k} - 6a^{-2k} + 3a^{-3k} \quad (1)$$

Além disso, mostraremos que só precisamos nos preocupar com pares em que as duas peças são não simétricas;

No capítulo 3 definiremos $M_n(x, y)$ o comprimento do maior encaixe possível entre duas peças de um par admissível (x, y) . Mostraremos que a probabilidade de M_n assumir um valor k pode ser descrita, quando n tende a infinito, pela soma $a_k + b_k$, em que a_k é descrito na equação 1 e b_k tem como minorante:

$$d_k = 4a^{-k-1} - (6a^{-2} + 16a^{-1})a^{-2k} + (24a^{-1} + 16a^{-2} + 3a^{-3})a^{-3k} - (12a^{-1} + 16a^{-2} + 3a^{-3})a^{-4k} \quad (2)$$

e como majorante:

$$c_k = (1 - a^{-k}) \left(\frac{4a^{-k}}{a-1} - \frac{6a^{-2k}}{a^2-1} + \frac{3a^{-3k}}{a^3-1} \right) \quad (3)$$

daremos ainda evidências de que, sob as condições do trabalho,

$$a_k > b_k;$$

Finalmente, no capítulo ?? faremos simulações para discutirmos acerca dos resultados encontrados nos capítulos anteriores;

Capítulo 1

A Construção da Medida

Neste capítulo introduziremos conceitos, conjuntos e ferramentas que serão importantes para o desenvolvimento de todo o trabalho. Primeiramente definiremos um alfabeto sobre o qual construiremos palavras, estudando suas propriedades e sua decomposição em partes menores. Além disso, construiremos um espaço do qual poderemos escolher, através de uma medida, um par de palavras que não sejam iguais ou o inverso uma da outra. Ao fim, definiremos quatro combinações das palavras e seus inversos que podem formar sobreposição. Essas combinações gerarão conjuntos que chamaremos de "Modos de Encaixe" em alusão a um jogo de dominó clássico.

1.1 Alfabetos e palavras

Seja \mathcal{A} um conjunto finito de símbolos, de cardinalidade $\#\mathcal{A} = a$. Dado $n \in \mathbb{N}^+$, chamamos

$$x_1^n = x_1 x_2 \dots x_n, x_i \in \mathcal{A} \forall i \in \{1, 2, \dots, n\}$$

de palavra de tamanho n . Quando não houver ambiguidade a respeito de n , escreveremos apenas x .

Uma noção importante neste trabalho é a de palavra inversa. Dessa forma, dada a palavra $x = x_1^n$, definimos

$$\tilde{x} = x_n x_{n-1} x_{n-2} \dots x_2 x_1 = x_n^1$$

o inverso de x . Observe que $\tilde{\tilde{x}} = x$.

Denotaremos por α , β , γ e δ palavras de comprimento finito. $|\alpha|$ representará o comprimento da palavra α . Se

$$\alpha = a_1 a_2 \dots a_{n_1}, a_i \in \mathcal{A}$$

e

$$\beta = a_{n_1+1} a_{n_1+2} \dots a_{n_1+n_2}, a_i \in \mathcal{A}$$

então a concatenação $\alpha\beta$ é dada por:

$$\alpha\beta = a_1a_2 \dots a_{n_1}a_{n_1+1} \dots a_{n_1+n_2}$$

Algumas definições importantes com relação à estrutura das palavras são:

- x é simétrica se $x = \tilde{x}$
ex.: 010010 é uma palavra simétrica;
- p_x é um prefixo de x se $x = p_x\alpha$
ex.: 011 é prefixo de 011010101;
- s_x é um sufixo de x se $x = \alpha s_x$
ex.: 111 é sufixo de 100110111;
- α é borda de x se α é simultaneamente prefixo e sufixo de x
ex.: 010 é borda de 0101101010;
- ξ denota a palavra nula, ou seja, uma sequência de comprimento zero;
- $x = \alpha^k$ se $x = \underbrace{\alpha \dots \alpha}_{k \text{ vezes}}$ e $k \in \mathbb{N}$; se $k = 0$, então $\alpha^k = \xi$
ex.: $(0110)^4 = 0110011001100110$;
- x periódica se $x = \alpha^k p_\alpha$ com $k \in \mathbb{N}^+$ e p_α prefixo de α
ex.: 10010010010 é uma palavra periódica;
- $x = x_1 \dots x_n$ tem período p se $x_i = x_{i+p}$ para $1 \leq i \leq n-p$; note que x sempre tem período n
ex.: 1010101 tem períodos 2,4,6 e 7;
- p período (ou *menor período*) de x se $p = \min\{q \in \mathbb{N} : x \text{ tem período } q\}$
ex.: o período de 00100100100100 é 3;
- $y = y_1 \dots y_m$ é uma subpalavra de $x = x_1 \dots x_n$ se existe k em $[0, n-m] \cap \mathbb{N}^+$ tal que $y_i = x_{i+k} \forall i \in [1, m] \cap \mathbb{N}^+$
ex.: 010 é uma possível subpalavra de 111001011000;
- y é o máximo fator sem borda de x se y é a subpalavra de maior comprimento de x que não possui borda

ex.: o máximo fator sem borda de 111110111 é 111110 (note que qualquer subpalavra iniciada ou terminada por 0, exceto a própria sequência 0, é um fator sem borda);

Para evitar o conflito entre as bordas de uma palavra ou subpalavra, lançamos mão da seguinte proposição:

Proposição 1.1 *Dado \mathcal{A}^n , com $\#\mathcal{A} = a \geq 2$. O comprimento esperado do máximo fator sem borda de uma palavra em \mathcal{A}^n com probabilidade uniforme para cada símbolo é $n - O(a^{-1})$.*

A demonstração pode ser encontrada em [Cording e Knudsen \(2016\)](#), embora seja fácil entender como podemos usá-la para trabalhar apenas com palavras cujas bordas não se encontram.

1.2 Conjuntos e medidas

Diremos que $x \sim y$ se $x = y$ ou $x = \tilde{y}$

Proposição 1.2 *A relação \sim é uma relação de equivalência em \mathcal{A}^n*

Demonstração:

1. reflexividade:

$x = x \Rightarrow x \sim x$ imediatamente

2. simetria:

$x \sim y$

- Se $x = y$, então $y = x$.
Logo $y \sim x$

- Se $x \neq y$, então $x = \tilde{y}$
Logo $\tilde{\tilde{y}} = \tilde{x}$
Assim $y = \tilde{x}$ e, portanto, $y \sim x$

3. transitividade:

$x \sim y$ e $y \sim z$

- Se $x = y$, então $x \sim z$ imediatamente

- Se $x \neq y$, então $x = \tilde{y}$

– Se $y = z$, então $x = \tilde{z}$

– Se $y \neq z$, então $y = \tilde{z}$
portanto $x = \tilde{\tilde{z}} = z$

□

Podemos definir, portanto, a classe de equivalência $\bar{x} = \{x, \tilde{x}\}$ para $x \in \mathcal{A}^n$. Perceba que se x for simétrica, \bar{x} é um conjunto unitário.

O espaço que nos interessa é o quociente $\Omega_n = \mathcal{A}^n / \sim$ e a medida que nos interessa é a uniforme sobre Ω_n , portanto usaremos a σ -álgebra 2^{Ω_n} . Por exemplo, se $\mathcal{A} = \{0, 1\}$, então temos $\mathcal{A}^2 = \{00, 01, 10, 11\}$ e $\Omega_2 = \{00, 01, 11\}$. Será mais fácil trabalhar com as palavras em \mathcal{A}^n , então precisaremos definir uma medida μ_n sobre \mathcal{A}^n cuja medida induzida em Ω_n seja a uniforme. Para obtermos a medida μ_n de cada elemento, particionamos \mathcal{A}^n nos seguintes conjuntos:

- $\mathcal{S}^n = \{x \in \mathcal{A}^n : x = \tilde{x}\}$, o conjunto das palavras simétricas
- $\mathcal{N}^n = \{x \in \mathcal{A}^n : x \neq \tilde{x}\}$, o conjunto das palavras não-simétricas

É fácil ver que o conjunto de palavras simétricas (\mathcal{S}^n) tem cardinalidade

$$\#\mathcal{S}^n = a^{\lceil \frac{n}{2} \rceil}$$

em que, para um $t \in \mathbb{R}$, $\lceil t \rceil$ representa o menor inteiro maior ou igual a t . Como

$$\#\mathcal{A}^n = a^n,$$

temos que

$$\#\mathcal{N}^n = a^n - a^{\lceil \frac{n}{2} \rceil}.$$

Como cada classe de Ω_n com elementos não-simétricos está relacionada a dois elementos de \mathcal{A}^n e cada classe de elemento simétrico tem identificação unívoca em \mathcal{A}^n , temos que a medida μ_n vale um certo t_n entre as sequências não-simétricas e $2t_n$ para as simétricas. Para que valha

$$2t_n \#\mathcal{S}^n + t_n \#\mathcal{N}^n = 1$$

temos $t_n = (a^n + a^{\lceil \frac{n}{2} \rceil})^{-1}$. Em resumo:

$$\mu_n(x) = \begin{cases} t_n & x \in \mathcal{N}^n \\ 2t_n & x \in \mathcal{S}^n \end{cases} \quad (1.1)$$

1.3 O espaço de pares

Queremos escolher dois elementos em Ω_n sem reposição. Será mais fácil pensar em sequências ordenadas. Para isso, definimos o conjunto:

$$\Delta_n = \{(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \mid \bar{x} = \bar{y}\} \quad (1.2)$$

O problema de escolher duas palavras sem reposição em Ω_n equivale a escolher um par em $\mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n$. Uma medida para um par neste espaço é a seguinte

$$\mu_n^{(2)}(x, y) = \mu_n(x) \cdot \bar{\mu}_n(y|x)$$

em que $\bar{\mu}_n(\cdot|x)$ é a medida μ_n de \mathcal{A}^n , excluindo \bar{x} . Como $\mu_n(\bar{x}) = \mu_n(\{x, \tilde{x}\}) = 2t_n$, independente de x ser ou não simétrico, daí $\bar{\mu}_n(\mathcal{A}^n \setminus \{x, \tilde{x}\}|x) = 1 - 2t_n$ e, como resultado, para algum $y \in \mathcal{A}^n \setminus \{x, \tilde{x}\}$:

$$\bar{\mu}_n(y|x) = \begin{cases} \frac{t_n}{1 - 2t_n}, & y \in \mathcal{N}^n \\ \frac{2t_n}{1 - 2t_n}, & y \in \mathcal{S}^n \end{cases} \quad (1.3)$$

Finalmente, para $(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n$, podemos ter escolhido duas palavras simétricas, duas não-simétricas ou uma simétrica e uma não-simétrica. Dessa maneira, temos

$$\mu_n^{(2)}(x, y) = \begin{cases} \frac{t_n^2}{1 - 2t_n}, & (x, y) \in \mathcal{N}^n \times \mathcal{N}^n \setminus \Delta_n \\ \frac{4t_n^2}{1 - 2t_n}, & (x, y) \in \mathcal{S}^n \times \mathcal{S}^n \setminus \Delta_n \\ \frac{2t_n^2}{1 - 2t_n}, & \text{caso contrário} \end{cases} \quad (1.4)$$

1.3.1 O problema do encaixe

Na alegoria do dominó, duas peças $a - b$ e $c - d$ se "encaixam" se $a = c$, se $a = d$, se $b = c$ ou se $b = d$, como por exemplo as peças $1 - 4$ e $5 - 1$ se encaixam pois existe uma coincidência entre as extremidades. Suponha a seguinte alteração sobre o jogo: ao invés de dois símbolos por peça, passamos a ter 4 símbolos. Considere as seguintes peças:

$$\begin{array}{r}
 2 - 4 - 5 - 6 \\
 6 - 3 - 2 - 4 \\
 1 - 6 - 5 - 4 \\
 4 - 5 - 6 - 1
 \end{array} \tag{1.5}$$

Note que, entre a segunda peça e a primeira, podemos formar a sequência $6 - 3 - 2 - 4 - 5 - 6$, dado que existe a coincidência do $2 - 4$ no final de uma e no começo da outra. Também é possível formar $2 - 4 - 5 - 6 - 3 - 2 - 4$, pois encontramos um 6 no final da primeira e no início da segunda. Dessa forma, podemos dizer que existe um encaixe de tamanho 2 e um de tamanho 1 entre as duas. Olhando para a primeira e a terceira, podemos formar $2 - 4 - 5 - 6 - 1$. Isso é possível pois podemos inverter a terceira peça de forma a obter $4 - 5 - 6 - 1$. Assim, a primeira e a terceira peças têm um encaixe de tamanho 3. O leitor pode reparar que a terceira e a quarta peça são a inversa uma da outra, dessa forma parece lógico supor que o encaixe entre as duas é de tamanho 4, mas não existem peças iguais ou co-inversas no jogo de dominó, portanto esse par não faria parte do espaço sobre o qual construiremos a função tamanho de encaixe.

Partindo de [Abadi e Lambert \(2019\)](#), a seguinte definição:

$$R_n^{(2)}(k) = \{(x, y) \in \mathcal{A}^n \times \mathcal{A}^n : x_{n-k}^{n-1} = x_0^{k-1}\}$$

isto é, $R_n^{(2)}(k)$ representa o conjunto de pares (x, y) em $\mathcal{A}^n \times \mathcal{A}^n$ tais que os últimos k símbolos de x são iguais aos k primeiros de y , definimos:

Definição 1.3 Dado $n \in \mathbb{N}^*$, temos que duas palavras em \mathcal{A}^n podem se encaixar de quatro modos diferentes. Se $k \in [1, n-1] \cap \mathbb{N}$, chamamos de modos de encaixe de tamanho k os seguintes conjuntos:

- $\Lambda_1^n(k) = \{(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n \mid (x, y) \in R_n^{(2)}(k)\}$
- $\Lambda_2^n(k) = \{(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n \mid (x, \tilde{y}) \in R_n^{(2)}(k)\}$
- $\Lambda_3^n(k) = \{(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n \mid (\tilde{x}, \tilde{y}) \in R_n^{(2)}(k)\}$
- $\Lambda_4^n(k) = \{(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n \mid (\tilde{x}, y) \in R_n^{(2)}(k)\}$

Definição 1.4 Dizemos que um par de palavras $(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n$ tem um encaixe de tamanho k se ele pertence a um dos quatro modos de encaixe deste tamanho. Em outras palavras (x, y) tem algum encaixe de tamanho k se o par está em:

$$\bar{R}_n^{(2)}(k) = \bigcup_{i=1}^4 \Lambda_i^n(k)$$

Note que a noção de encaixe está profundamente conectada à de sobreposição entre duas palavras. Enquanto o conjunto $R_n^{(2)}(k)$ admite apenas um modo de encaixe, $\bar{R}_n^{(2)}(k)$, permite contar os quatro modos. Por exemplo, se

$$x = 010001$$

e

$$y = 111100$$

temos que $(x, y) \notin R_6^{(2)}(3)$, mas $(x, y) \in \bar{R}_6^{(2)}(3)$, pois $(x, \tilde{y}) \in \mathbb{R}_6^{(2)}(3)$ e, dessa forma, $(x, y) \in \Lambda_2^6(3)$.

Capítulo 2

A probabilidade de um encaixe

Neste capítulo, provaremos que a probabilidade de um par de palavras em $\mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n$ ter um encaixe de tamanho k converge para $4a^{-k} - 6a^{-2k} + 3a^{-3k}$ quando n diverge. Para isso, usaremos a fórmula da inclusão-exclusão, dado que $\bar{R}_n^{(2)}(k)$ é a união dos quatro modos de encaixe de tamanho k . Veremos que o número de termos na fórmula pode ser reduzido dadas as equivalências entre esses conjuntos e entre intersecções deles.

2.1 O que importa é a não-simetria

Ao tomarmos duas palavras em \mathcal{A}^n , podemos obter um par de palavras não-simétricas, duas palavras simétricas, uma palavra simétrica e outra não-simétrica e até mesmo duas palavras iguais. Como, segundo a equação 1.4, em cada um desses casos a medida $\mu_n^{(2)}$ do par possui um valor diferente, deveríamos dividir cada probabilidade em casos e exceções. Os lemas a seguir nos garantem que um par em $\mathcal{A}^n \times \mathcal{A}^n \setminus \Delta_n$ está quase certamente em $\mathcal{N}^n \times \mathcal{N}^n$ quando n tende a infinito.

Lema 2.1 $t_n \xrightarrow{n \rightarrow \infty} 0$

Demonstração: Como $t_n = (a^n + a^{\lceil \frac{n}{2} \rceil})^{-1}$, então

$$\lim_{n \rightarrow \infty} t_n = 0$$

□

Lema 2.2 $\mathbb{P}(\mathcal{N}^n \times \mathcal{S}^n \setminus \Delta_n) \xrightarrow{n \rightarrow \infty} 0$

Demonstração:

$$\begin{aligned}
\mathbb{P}(\mathcal{N}^n \times \mathcal{S}^n \setminus \Delta_n) &= \sum_{x \in \mathcal{N}^n} \sum_{y \in \mathcal{S}^n} \mu_n^{(2)}(x, y) \\
&= \sum_{x \in \mathcal{N}^n} \sum_{y \in \mathcal{S}^n} \frac{2t_n^2}{1 - 2t_n} \\
&= \frac{2(a^n + a^{\lceil \frac{n}{2} \rceil})^{-2}}{1 - 2t_n} \bar{\mu}_n(\mathcal{N}^n) \mu_n^*(\mathcal{S}^n | \mathcal{N}^n) \\
&= \frac{2(a^n + a^{\lceil \frac{n}{2} \rceil})^{-2}}{1 - 2t_n} ((a^n - a^{\lceil \frac{n}{2} \rceil}) a^{\lceil \frac{n}{2} \rceil}) \\
&= \frac{2(a^n + a^{\lceil \frac{n}{2} \rceil})^{-2} (a^{n + \lceil \frac{n}{2} \rceil} - a^{2\lceil \frac{n}{2} \rceil})}{1 - 2t_n}
\end{aligned} \tag{2.1}$$

Note que a primeira igualdade vem do fato de $\mathcal{N}^n \times \mathcal{S}^n \setminus \Delta = \mathcal{N}^n \times \mathcal{S}^n$, pois um elemento simétrico não pode ser igual a um elemento não-simétrico ou à sua inversão. Dado que, pelo lema 2.1, o denominador tende a zero, podemos analisar apenas o termo dominante de cada parte do produto acima, no caso a^{-2n} e $a^{n + \lceil \frac{n}{2} \rceil}$. Dessa forma, podemos dizer que $\mathbb{P}(\mathcal{N}^n \times \mathcal{S}^n \setminus \Delta_n) \sim O(a^{-\lceil \frac{n}{2} \rceil}) \xrightarrow{n \rightarrow \infty} 0$
□

Lema 2.3 $\mathbb{P}(\mathcal{S}^n \times \mathcal{S}^n \setminus \Delta_n) \xrightarrow{n \rightarrow \infty} 0$

Demonstração: Usando argumentação parecida ao lema anterior, devemos apenas tomar o cuidado para não escolher a mesma palavra duas vezes (esse problema é evitado no lema 2.2, porque $\mathcal{N}^n \cap \mathcal{S}^n = \emptyset$)

$$\begin{aligned}
\mathbb{P}(\mathcal{S}^n \times \mathcal{S}^n \setminus \Delta_n) &= \sum_{x \in \mathcal{S}^n} \sum_{y \in \mathcal{S}^n \setminus \{x\}} \mu_n^{(2)}(x, y) \\
&= \sum_{x \in \mathcal{S}^n} \sum_{y \in \mathcal{S}^n \setminus \{x\}} \frac{4t_n^2}{1 - 2t_n} \\
&= \frac{4(a^n + a^{\lceil \frac{n}{2} \rceil})^{-2} (a^{\lceil \frac{n}{2} \rceil} (a^{\lceil \frac{n}{2} \rceil} - 1))}{1 - 2t_n} \\
&= \frac{4(a^n + a^{\lceil \frac{n}{2} \rceil})^{-2} (a^{2\lceil \frac{n}{2} \rceil} - a^{\lceil \frac{n}{2} \rceil})}{1 - 2t_n}
\end{aligned} \tag{2.2}$$

Observe que o argumento $a^{\lceil \frac{n}{2} \rceil} - 1$ se refere à todas as palavras simétricas menos àquela que foi escolhida no primeiro somatório. Podemos supor $a^{\lceil \frac{n}{2} \rceil} \sim a^{\frac{n}{2}}$, assim, a análise de termo dominante nos dá

$$O(a^{-2n})O(a^n) = O(a^{-n}) \xrightarrow{n \rightarrow \infty} 0$$

□

Os lemas 2.2 e 2.3 nos garantem que só precisamos nos preocupar com $\mathcal{N}^n \times \mathcal{N}^n \setminus \Delta_n$ quando n vai ao infinito. Podemos ainda desconsiderar o Δ_n para fins de facilidade. Isso nos é garantido pelo seguinte lema:

Lema 2.4 $\mathbb{P}(\Delta_n) \xrightarrow{n \rightarrow \infty} 0$

Demonstração: Observe que, se um par (x, y) escolhido está em Δ_n , então não poderemos usar a medida $\mu_n^{(2)}$. Uma alternativa é usar μ_n , dado que a definição de elemento determina imediatamente o outro, a menos de inversão no caso não simétrico. Dessa forma

$$\begin{aligned}
\mathbb{P}(\Delta_n) &= \sum_{x \in \mathcal{N}^n} \sum_{\{x, \tilde{x}\}} \mu_n(x)^2 + \sum_{x \in \mathcal{S}^n} \sum_{\{x\}} \mu_n(x)^2 \\
&= \sum_{x \in \mathcal{N}^n} \sum_{\{x, \tilde{x}\}} (t_n)^2 + \sum_{x \in \mathcal{S}^n} \sum_{\{x\}} (2t_n)^2 \\
&= \sum_{x \in \mathcal{N}^n} 2t_n^2 + \sum_{x \in \mathcal{S}^n} 4t_n^2 \\
&= 2t_n^2(a^n - a^{\lceil \frac{n}{2} \rceil}) + 4t_n^2 a^{\lceil \frac{n}{2} \rceil} \\
&= 2t_n^2(a^n + a^{\lceil \frac{n}{2} \rceil}) \\
&= 2t_n^2 t_n^{-1} = 2t_n \xrightarrow{n \rightarrow \infty} 0
\end{aligned} \tag{2.3}$$

□

2.2 Modos de encaixe

Inicialmente, iremos descrever a forma das palavras, segundo os quatro modos de encaixe que elas podem apresentar:

Dados $n \in \mathbb{N}^*$ e $k \in [1, n-1] \cap \mathbb{N}$

$$(x, y) \in \Lambda_1^n(k) \iff \begin{cases} x = \alpha_1 \gamma \beta \\ y = \beta \delta \alpha_2 \end{cases} \tag{2.4}$$

$$(x, y) \in \Lambda_2^n(k) \iff \begin{cases} x = \alpha_1 \gamma \beta \\ y = \alpha_2 \delta \tilde{\beta} \end{cases} \tag{2.5}$$

$$(x, y) \in \Lambda_3^n(k) \iff \begin{cases} x = \alpha \gamma \beta_1 \\ y = \beta_2 \delta \alpha \end{cases} \tag{2.6}$$

$$(x, y) \in \Lambda_4^n(k) \iff \begin{cases} x = \alpha \gamma \beta_1 \\ y = \tilde{\alpha} \delta \beta_2 \end{cases} \tag{2.7}$$

para alguma escolha de $\alpha, \beta, \alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathcal{A}^k$ e $\gamma, \delta \in \mathcal{A}^{n-2k}$. Segue diretamente que

$$\mathbb{P}(\Lambda_1^n(k)) = \mathbb{P}(\Lambda_2^n(k)) = \mathbb{P}(\Lambda_3^n(k)) = \mathbb{P}(\Lambda_4^n(k))$$

Estamos também interessados nas intersecções dois a dois entre esses conjuntos. De maneira que, similarmente à descrição acima, temos que, dados $n \in \mathbb{N}^*$ e $k \in [1, n-1] \cap \mathbb{N}$, valem:

$$(x, y) \in \Lambda_1^n(k) \cap \Lambda_2^n(k) \iff \begin{cases} x = \alpha\gamma\beta \\ y = \beta\delta\tilde{\beta} \end{cases} \quad (2.8)$$

$$(x, y) \in \Lambda_3^n(k) \cap \Lambda_4^n(k) \iff \begin{cases} x = \alpha\gamma\beta \\ y = \tilde{\alpha}\delta\alpha \end{cases} \quad (2.9)$$

$$(x, y) \in \Lambda_1^n(k) \cap \Lambda_4^n(k) \iff \begin{cases} x = \alpha\gamma\tilde{\alpha} \\ y = \tilde{\alpha}\delta\beta \end{cases} \quad (2.10)$$

$$(x, y) \in \Lambda_2^n(k) \cap \Lambda_3^n(k) \iff \begin{cases} x = \alpha\gamma\tilde{\alpha} \\ y = \beta\delta\alpha \end{cases} \quad (2.11)$$

$$(x, y) \in \Lambda_1^n(k) \cap \Lambda_3^n(k) \iff \begin{cases} x = \alpha\gamma\beta \\ y = \beta\delta\alpha \end{cases} \quad (2.12)$$

$$(x, y) \in \Lambda_2^n(k) \cap \Lambda_4^n(k) \iff \begin{cases} x = \alpha\gamma\beta \\ y = \tilde{\alpha}\delta\tilde{\beta} \end{cases} \quad (2.13)$$

para alguns $\alpha, \beta \in \mathcal{A}^k$ e $\gamma, \delta \in \mathcal{A}^{n-2k}$. Disso, pode-se concluir que

$$\mathbb{P}(\Lambda_1^n(k) \cap \Lambda_2^n(k)) = \mathbb{P}(\Lambda_3^n(k) \cap \Lambda_4^n(k)) = \mathbb{P}(\Lambda_1^n(k) \cap \Lambda_4^n(k)) = \mathbb{P}(\Lambda_2^n(k) \cap \Lambda_3^n(k))$$

enquanto

$$\mathbb{P}(\Lambda_1^n(k) \cap \Lambda_2^n(k)) = \mathbb{P}(\Lambda_3^n(k) \cap \Lambda_4^n(k))$$

Podemos, enfim, descrever na forma de concatenação de subpalavras as intersecções de três dos conjuntos e a intersecção de todos eles:

Dados $n \in \mathbb{N}^*$ e $k \in [1, n-1] \cap \mathbb{N}$, vale:

$$\begin{aligned} (x, y) \in \Lambda_1^n(k) \cap \Lambda_2^n(k) \cap \Lambda_3^n(k) &\iff (x, y) \in \Lambda_1^n(k) \cap \Lambda_2^n(k) \cap \Lambda_4^n(k) \\ &\iff (x, y) \in \Lambda_1^n(k) \cap \Lambda_3^n(k) \cap \Lambda_4^n(k) \\ &\iff (x, y) \in \Lambda_2^n(k) \cap \Lambda_3^n(k) \cap \Lambda_4^n(k) \\ &\iff (x, y) \in \Lambda_1^n(k) \cap \Lambda_2^n(k) \cap \Lambda_3^n(k) \cap \Lambda_4^n(k) \\ &\iff \begin{cases} x = \alpha\gamma\tilde{\alpha} \\ y = \tilde{\alpha}\delta\alpha \end{cases} \end{aligned} \quad (2.14)$$

para $\alpha \in \mathcal{A}^k$ e $\delta, \gamma \in \mathcal{A}^{n-2k}$

2.3 Convergência

Nesta secção encontraremos o valor para o qual converge a probabilidade de duas peças do nosso jogo terem algum encaixe de tamanho k . Para isso lançaremos mão dos lemas da secção 2.1 e dos quatro lemas abaixo. Note que, pelas igualdades obtidas na secção 2.2, teremos uma relevante redução no número de cálculos. A partir de agora, usaremos a seguinte notação:

$$\Lambda_{i_1, \dots, i_l}^n = \Lambda_{i_1}^n \cap \dots \cap \Lambda_{i_l}^n, i_1, \dots, i_l \in \{1, 2, 3, 4\}$$

Lema 2.5 Para $k \in \mathbb{N}^*$, temos que

$$\mathbb{P}(\Lambda_1^n(k)) \xrightarrow{n \rightarrow \infty} a^{-k}$$

Demonstração: Observe que, dados os lemas 2.2, 2.3 e 2.4, podemos escrever:

$$\mathbb{P}(\Lambda_1^n(k)) = \left(\sum_{x \in \mathcal{N}^n} \sum_{y \in \mathcal{N}_{x,k}^n} \mu_n^{(2)}(x, y) \right) + r$$

em que $r \xrightarrow{n \rightarrow \infty} 0$ e, se x tem a forma $\alpha_1 \gamma \beta$, então

$$\mathcal{N}_{x,k}^n = \{y \in \mathcal{A}^n : y = \beta \delta \alpha_2, \text{ para alguma escolha de } \delta \in \mathcal{A}^{n-2k}, \alpha_2 \in \mathcal{A}^k \text{ e } \beta \text{ conhecido}\}$$

Observe que r é uma quantidade residual ligada a pares em $\mathcal{N}^n \times \mathcal{S}^n$, $\mathcal{S}^n \times \mathcal{S}^n$ e Δ_n . Note também que é possível que $y \in \mathcal{N}_{x,k}^n$ pertença a \mathcal{S}^n ou até mesmo que o par pertença a Δ_n , mas essas exceções também tornam-se negligenciáveis à medida que n cresce. Enfim, podemos dizer que:

$$\mathbb{P}(\Lambda_1^n(k)) = \frac{t_n^2}{1 - 2t_n} (\#\mathcal{N}_{x,k}^n \#\mathcal{N}^n) + r$$

Como $y \in \mathcal{N}_{x,k}^n$ implica que y é da forma $\beta \delta \alpha_2$ com β conhecido, só podemos escolher $\delta \in \mathcal{A}^{n-2k}$ e $\alpha_2 \in \mathcal{A}^k$ e temos $a^{n-2k+k} = a^{n-k}$ formas de fazê-lo, logo $\#\mathcal{N}_{x,k}^n = a^{n-k}$ e, dado que $\#\mathcal{N}^n = a^n - a^{\lfloor \frac{n}{2} \rfloor}$,

$$\mathbb{P}(\Lambda_1^n(k)) = \frac{t_n^2}{1 - 2t_n} (a^{n-k} (a^n - a^{\lfloor \frac{n}{2} \rfloor})) + r = \frac{t_n^2}{1 - 2t_n} (a^{2n-k} - a^{n+\lfloor \frac{n}{2} \rfloor - k}) + r$$

Dado que o t_n^2 é da ordem de a^{-2n} e o fator dominante do segundo termo é da ordem de a^{2n-k} ,

$$\mathbb{P}(\Lambda_1^n(k)) \xrightarrow{n \rightarrow \infty} a^{-k}$$

□ .

Lema 2.6 *Se $k \in \mathbb{N}^*$, então*

$$\mathbb{P}(\Lambda_{1,2}^n(k)) \xrightarrow{n \rightarrow \infty} a^{-2k}$$

Demonstração: Tome um inteiro positivo n e $k \in [1, n-1] \cap \mathbb{Z}$. Pelos lemas da secção 2.1

$$\mathbb{P}(\Lambda_{1,2}^n(k)) = \left(\sum_{x \in \mathcal{N}^n} \sum_{y \in \mathcal{N}_{x,k}^{n,2}} \mu_n^{(2)}(x, y) \right) + r$$

com $r \xrightarrow{n \rightarrow \infty} 0$ e, se x é igual a $\alpha \gamma \beta$, então

$$\mathcal{N}_{x,k}^{n,2} = \{y \in \mathcal{A}^n : y = \beta \delta \tilde{\beta}, \text{ para algum } \delta \in \mathcal{A}^{n-2k} \text{ e } \beta \text{ conhecido}\}$$

Temos então:

$$\mathbb{P}(\Lambda_{1,2}^n(k)) = \frac{t_n^2}{1-2t_n} (\#\mathcal{N}^n \#\mathcal{N}_{x,k}^{n,2}) + r$$

Dado que, na construção de y , β já é determinado em x , podemos escolher apenas δ . Dessa forma, $\#\mathcal{N}_{x,k}^{n,2} = a^{n-2k}$ e, como $\mathcal{N}^n = a^n - a^{\lfloor \frac{n}{2} \rfloor}$,

$$\mathbb{P}(\Lambda_{1,2}^n(k)) = \frac{t_n^2}{1-2t_n} (a^{-n-2k} (a^n - a^{\lfloor \frac{n}{2} \rfloor})) + r = \frac{t_n^2}{1-2t_n} (a^{2n-2k} - a^{n-2k+\lfloor \frac{n}{2} \rfloor}) + r$$

Comparando novamente os termos dominantes, vemos que a fração é da ordem de a^{-2n} e o termo dominante do fator que o multiplica tem ordem a^{2n-2k} , então esse produto converge para a^{-2k} . \square

Lema 2.7 Para algum $k \in \mathbb{N}^*$, vale

$$\mathbb{P}(\Lambda_{1,3}^n(k)) \xrightarrow{n \rightarrow \infty} a^{-2k}$$

Demonstração: De modo análogo à demonstração do lema anterior. No entanto, note que ao invés de escolher y da forma $\beta\delta\tilde{\beta}$, escolhamos y da forma $\beta\delta\alpha$, tendo a^{n-2k} escolhas para $\delta \in \mathcal{A}^{n-2k}$. \square

Daí, basta calcular o valor para a intersecção tripla:

Lema 2.8 Dado $k \in \mathbb{N}^*$, temos que

$$\mathbb{P}(\Lambda_{1,2,3}^n(k)) \xrightarrow{n \rightarrow \infty} a^{-3k}$$

Demonstração: Novamente, utilizando os lemas do início do capítulo,

$$\mathbb{P}(\Lambda_{1,2,3}^n(k)) = \left(\sum_{x \in \tilde{\mathcal{N}}_{n,k}} \sum_{y \in \mathcal{N}_{x,k}^{n,2}} \mu_n^{(2)}(x, y) \right) + r$$

com $r \xrightarrow{n \rightarrow \infty} 0$, $\mathcal{N}_{x,k}^{n,2}$ como definido na demonstração do lema 2.6. Note que as duas palavras do par possuem restrições. Então, definimos:

$$\tilde{\mathcal{N}}_{n,k} = \{x \in \mathcal{A}^n : x = \alpha\gamma\tilde{\alpha}, \text{ para algum } \alpha \in \mathcal{A}^k \text{ e } \gamma \in \mathcal{A}^{n-2k}\}$$

e segue que:

$$\mathbb{P}(\Lambda_1^n(k) \cap \Lambda_2^n(k) \cap \Lambda_3^n(k)) = \frac{t_n^2}{1-2t_n} (\#\tilde{\mathcal{N}}_{n,k} \#\mathcal{N}_{x,k}^{n,2}) + r$$

Já sabemos que $\#\mathcal{N}_{x,k}^{n,2} = a^{n-2k}$, então precisamos apenas contar as escolhas possíveis em $\tilde{\mathcal{N}}_{n,k}$. Como, ao definir α , também determinamos $\tilde{\alpha}$, para um x em $\tilde{\mathcal{N}}_{n,k}$, podemos escolher somente α e

γ , ou seja, temos a^{n-k} formas. Logo, $\#\tilde{\mathcal{N}}_{n,k} = a^{n-k}$. Portanto:

$$\mathbb{P}(\Lambda_{1,2,3}^n(k)) = \frac{t_n^2}{1-2t_n}(a^{n-k}a^{n-2k}) + r = \frac{t_n^2}{1-2t_n}a^{2n-3k} + r$$

E, pela comparação de termos dominantes, temos que essa probabilidade converge a a^{-3k} \square

Finalmente,

Teorema 2.9 *Seja \mathcal{A} um alfabeto finito com cardinalidade $\#\mathcal{A} = a$. Dado um inteiro positivo k , temos a seguinte convergência para a probabilidade de um encaixe de tamanho k entre duas palavras:*

$$\mathbb{P}(\bar{R}_n^{(2)}(k)) \xrightarrow{n \rightarrow \infty} 4a^{-k} - 6a^{-2k} + 3a^{-3k}$$

Demonstração: Precisamos calcular o valor para o qual $\mathbb{P}(\bar{R}_n^{(2)}(k))$ converge. Como $\bar{R}_n^{(2)}$ se trata de uma união de conjuntos, calcularemos sua probabilidade através da fórmula de inclusão-exclusão:

$$\begin{aligned} \mathbb{P}(\bar{R}_n^{(2)}(k)) &= \mathbb{P}(\Lambda_1^n(k) \cup \Lambda_2^n(k) \cup \Lambda_3^n(k) \cup \Lambda_4^n(k)) \\ &= \mathbb{P}(\Lambda_1^n(k)) + \mathbb{P}(\Lambda_2^n(k)) + \mathbb{P}(\Lambda_3^n(k)) + \mathbb{P}(\Lambda_4^n(k)) \\ &\quad - \mathbb{P}(\Lambda_{1,2}^n(k)) - \mathbb{P}(\Lambda_{3,4}^n(k)) \\ &\quad - \mathbb{P}(\Lambda_{1,4}^n(k)) - \mathbb{P}(\Lambda_{2,3}^n(k)) \\ &\quad - \mathbb{P}(\Lambda_{1,3}^n(k)) - \mathbb{P}(\Lambda_{2,4}^n(k)) \\ &\quad + \mathbb{P}(\Lambda_{1,2,3}^n(k)) + \mathbb{P}(\Lambda_{1,2,4}^n(k)) \\ &\quad + \mathbb{P}(\Lambda_{1,3,4}^n(k)) + \mathbb{P}(\Lambda_{2,3,4}^n(k)) \\ &\quad - \mathbb{P}(\Lambda_{1,2,3,4}^n(k)) \end{aligned} \tag{2.15}$$

Podemos simplificar essa soma usando as igualdades na secção 2.2, enfim, os lemas 2.5, 2.6, 2.7 e 2.8

$$\begin{aligned} \mathbb{P}(\bar{R}_n^{(2)}(k)) &= 4\mathbb{P}(\Lambda_1^n(k)) - 4\mathbb{P}(\Lambda_{1,2}^n(k)) \\ &\quad - 2\mathbb{P}(\Lambda_{1,3}^n(k)) + 3\mathbb{P}(\Lambda_{1,2,3}^n(k)) \\ &\quad \xrightarrow{n \rightarrow \infty} 4a^{-k} - 6a^{-2k} + 3a^{-3k} \end{aligned} \tag{2.16}$$

\square

Capítulo 3

A Distribuição do Máximo Encaixe

Após obtermos o valor assintótico para a probabilidade de duas palavras de tamanho n terem *algum* encaixe de tamanho k , estamos interessados em descobrir o maior encaixe possível entre elas. Um problema que surge é a possibilidade de, para um mesmo par de palavras e para um mesmo modo de encaixe, encontrarmos sobreposições de vários tamanhos.

3.1 Definições

Tendo em mente a definição 1.4

Definição 3.1 Dado $(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta$, o maior comprimento de sobreposição entre x e y para cada modo de encaixe é dado por:

$$S_i^n(x, y) = \max\{k \in [1, n-1] \cap \mathbb{Z} : (x, y) \in \Lambda_i^n(k)\}$$

ou 0, se esse conjunto for vazio

Daí, o comprimento máximo de encaixe (ou apenas máximo encaixe) entre x e y é definido como:

$$\begin{aligned} M_n(x, y) &= \max\{S_1^n(x, y), S_2^n(x, y), S_3^n(x, y), S_4^n(x, y)\} \\ &= \max\{k \in [0, n-1] \cap \mathbb{Z} : (x, y) \in \bar{R}_n^{(2)}(k)\} \end{aligned} \tag{3.1}$$

A partir de agora denotaremos $\{M^n \geq k\}$ o conjunto de pares cujo máximo encaixe entre suas palavras é maior ou igual a um inteiro k . Observe que podemos dividir o conjunto em duas partes disjuntas a partir da seguinte argumentação:

$$\begin{aligned}
\{M_n \geq k\} &= \{(x, y) \in \mathcal{A}^n \times \mathcal{A}^n \setminus \Delta : M^n(x, y) \geq k\} \\
&= \bigcup_{i=k}^{n-1} \bar{R}_n^{(2)}(i) \\
&= \bar{R}_n^{(2)}(k) \dot{\cup} \left(\bigcup_{i=k+1}^{n-1} \bar{R}_n^{(2)}(i) \right) \setminus \bar{R}_n^{(2)}(k) \\
&= \bar{R}_n^{(2)}(k) \dot{\cup} \bigcup_{i=k+1}^{n-1} \left(\bar{R}_n^{(2)}(i) \setminus \bar{R}_n^{(2)}(k) \right)
\end{aligned} \tag{3.2}$$

Daí, em termos de probabilidade:

Teorema 3.2 $\mathbb{P}(\{M^n \geq k\}) = a_{k,n} + b_{k,n}$, com

$$a_{k,n} = \mathbb{P}\left(\bar{R}_n^{(2)}(k)\right)$$

e

$$b_{k,n} = \mathbb{P}\left(\bigcup_{i=k+1}^{n-1} \left(\bar{R}_n^{(2)}(i) \setminus \bar{R}_n^{(2)}(k)\right)\right)$$

Além disso, dado que $a_{k,n} \xrightarrow{n \rightarrow \infty} a_k$ e $b_{k,n} \xrightarrow{n \rightarrow \infty} b_k$, temos

$$\mathbb{P}(\{M_n \geq k\}) \xrightarrow{n \rightarrow \infty} a_k + b_k$$

Investigaremos os casos em que cada um desses dois termos se torna dominante em relação ao outro. O valor assintótico a_k já foi encontrado no teorema 2.9. Buscaremos, então, um majorante para b_k . Note que

$$\begin{aligned}
b_{k,n} &= \mathbb{P}\left(\bigcup_{i=k+1}^{n-1} (\bar{R}_n^{(2)}(i) \setminus \bar{R}_n^{(2)}(k))\right) \\
&\leq \sum_{i=k+1}^{n-1} \mathbb{P}\left(\bar{R}_n^{(2)}(i) \setminus \bar{R}_n^{(2)}(k)\right) \\
&= \sum_{i=k+1}^{n-1} \left[\mathbb{P}\left(\bar{R}_n^{(2)}(i)\right) - \mathbb{P}\left(\bar{R}_n^{(2)}(i) \cap \bar{R}_n^{(2)}(k)\right)\right] \\
&\leq \sum_{i=k+1}^{n-1} \left[\mathbb{P}\left(\bar{R}_n^{(2)}(i)\right) - \mathbb{P}\left(\bar{R}_n^{(2)}(i) \cap \Lambda_1^n(k)\right)\right]
\end{aligned} \tag{3.3}$$

e, pelo teorema 2.9

$$a_k = 4a^{-k} - 6a^{-2k} + 3a^{-3k} \tag{3.4}$$

3.2 Crescimento assintótico dos diferentes tipos de encaixe

Lema 3.3 *Seja x uma palavra de tamanho $i > k$, se $x \in R_i(k)$, então x é periódica com período $i - k$*

Demonstração: Essa demonstração será dividida em três casos:

- Se $i > 2k$, então

$$x_1^i = x_1^k x_{k+1}^{i-k-1} x_1^k$$

ou seja, x_1^k é borda de x . Como x_1^k é prefixo de $x_1^{i-k-1} = x_1^k x_{k+1}^{i-k-1}$, temos que x tem período $i - k$

- Se $i = 2k$, então

$$x_1^i = x_1^k x_1^k$$

isto é, x é a concatenação das duas bordas, implicando no período proposto

- Se $i < 2k$, note que

$$x_1^k = x_{i-k}^i \iff x_1^k = x_{i-k+1}^k x_{k+1}^i$$

Como $|x_{i-k+1}^k| = 2k - i$, temos $x_{i-k+1}^i = x_1^{2k-i}$ então

$$x_1^i = x_1^{i-k} x_{i-k+1}^i = x_1^{i-k} x_1^k = x_1^{i-k} x_1^{i-k} x_{i-k}^k = x_1^{i-k} x_1^{i-k} x_1^{2k-i}$$

Como x_1^{2k} é prefixo de x_1^{i-k} temos uma palavra de período $i - k$

□

Lema 3.4 *Suponhamos $i > k$. Se $(x, y) \in \Lambda_1^n(i) \cap \Lambda_1^n(k)$, então $x = \omega p_{ik}$, em que $\omega \in \mathcal{A}^{n-i}$ e p_{ik} é uma subpalavra em \mathcal{A}^i período $i - k$*

Demonstração: Suponha $x = \omega_1 \gamma_1 \beta_1$ e $y = \beta_2 \gamma_2 \omega_2$ com $\omega_1, \omega_2 \in \mathcal{A}^{n-i}$, $\gamma_1, \gamma_2 \in \mathcal{A}^{i-k}$ e $\beta_1, \beta_2 \in \mathcal{A}^k$. Como $(x, y) \in \Lambda_1^n(j)$ então $\beta_1 = \beta_2 = \beta$. Também temos que, se $(x, y) \in \Lambda_1^n(i)$ então $\gamma_1 \beta = \beta \gamma_2$, ou seja, $\gamma_1 \beta \in R_n(k)$. Daí, basta usar o lema 3.3 para mostrar que esta subpalavra tem período $i - k$. □

A partir de agora, denotaremos por p_{ik} uma palavra periódica de período $i - k$. Os lemas a seguir dão conta da convergência do majorante definido na equação 3.3.

Lema 3.5 *Para $i, k \in \mathbb{N}^*$ com $i > k$, temos*

$$\mathbb{P}(\Lambda_1^n(i) \cap \Lambda_1^n(k)) \xrightarrow{n \rightarrow \infty} a^{-i-k}$$

Demonstração: Se $(x, y) \in \Lambda_1^n(i) \cap \Lambda_1^n(k)$ então as palavras são da seguinte forma

$$x = \omega_1 p_{ik} \text{ e } y = p_{ik} \omega_2$$

Com $\omega_1, \omega_2 \in \mathcal{A}^{n-i}$.

Dessa forma, existem a^{n-i} escolhas para ω_1 , assim como para ω_2 . Como p_{ik} , tem período $i - k$, é imediato que é definida pelos seus $i - k$ primeiros termos, logo, há a^{i-k} escolhas. Assim,

$$\begin{aligned} \mathbb{P}(\Lambda_1^n(i) \cap \Lambda_1^n(k)) &= \frac{t_n^2}{1 - 2t_n} (a^{n-i} a^{n-i} a^{i-k}) + r \\ &= \frac{(a^n + a^{\lceil \frac{n}{2} \rceil})^{-2}}{1 - 2(a^n + a^{\lceil \frac{n}{2} \rceil})^{-1}} (a^{2n-i-k}) + r \\ &\xrightarrow{n \rightarrow \infty} a^{-i-k} \end{aligned} \tag{3.5}$$

Em que r é uma quantidade residual, ou seja, $r \xrightarrow{n \rightarrow \infty} 0$ □

Lema 3.6 *Para $i, k \in \mathbb{N}^*$, com $i > k$, vale*

$$\mathbb{P}(\Lambda_2^n(i) \cap \Lambda_1^n(k)) \xrightarrow{n \rightarrow \infty} a^{-i-k}$$

Demonstração: Se $(x, y) \in \Lambda_2^n(i) \cap \Lambda_1^n(k)$, então o par é da forma

$$x = \omega_1 \gamma \beta \text{ e } y = \beta \omega_2 \tilde{\beta} \tilde{\gamma}$$

com $\omega_1 \in \mathcal{A}^{n-i}$, $\omega_2 \in \mathcal{A}^{n-i-k}$, $\gamma \in \mathcal{A}^{i-k}$ e $\beta \in \mathcal{A}^k$. Os números de escolhas possíveis para cada subpalavra nos dá

$$\begin{aligned} \mathbb{P}(\Lambda_2^n(i) \cap \Lambda_1^n(k)) &= \frac{t_n^2}{1-2t_n} (a^{n-i} a^{n-i-k} a^{i-k} a^k) + r \\ &= \frac{t_n^2}{1-2t_n} (a^{2n-i-k}) + r \\ &\xrightarrow{n \rightarrow \infty} a^{-i-k} \end{aligned} \quad (3.6)$$

□

De maneira análoga, é possível obter que:

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_2^n(i) \cap \Lambda_1^n(k)) = \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_3^n(i) \cap \Lambda_1^n(k)) = \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_4^n(i) \cap \Lambda_1^n(k)) = a^{-i-k}$$

Lema 3.7 Para $i, k \in \mathbb{N}^*$, com $i > k$, vale

$$\mathbb{P}(\Lambda_{1,2}^n(i) \cap \Lambda_1^n(k)) \rightarrow a^{-2i-k}$$

Demonstração: Temos que um par (x, y) em $\Lambda_1^n(i) \cap \Lambda_2^n(k) \cap \Lambda_1^n(k)$ tem forma

$$x = \omega_1 p_{ik} \text{ e } y = p_{ik} \omega_2 \tilde{p}_{ik}$$

com $\omega_1 \in \mathcal{A}^{n-i}$, $\omega_2 \in \mathcal{A}^{n-2i}$. Logo

$$\begin{aligned} \mathbb{P}(\Lambda_{1,2}^n(i) \cap \Lambda_1^n(k)) &= \frac{t_n^2}{1-2t_n} (a^{n-i} a^{n-2i} a^{i-k}) + r \\ &= \frac{t_n^2}{1-2t_n} (a^{2n-2i-k}) \\ &\xrightarrow{n \rightarrow \infty} a^{-2i-k} \end{aligned} \quad (3.7)$$

□

Analogamente,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,2}^n(i) \cap \Lambda_1^n(k)) &= \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,3}^n(i) \cap \Lambda_1^n(k)) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,4}^n(i) \cap \Lambda_1^n(k)) \\ &= a^{-2i-k} \end{aligned} \quad (3.8)$$

Lema 3.8 Dado um $k \in \mathbb{N}^*$ e um natural $i > k$, temos

$$\mathbb{P}(\Lambda_{3,4}^n(i) \cap \Lambda_1^n(k)) \xrightarrow{n \rightarrow \infty} a^{-2i-k}$$

Demonstração: Se $(x, y) \in \Lambda_{3,4}^n(i) \cap \Lambda_1^n(k)$, é mandatório que o par seja da forma:

$$x = \alpha\gamma\omega_1\tilde{\gamma}\tilde{\alpha} \text{ e } y = \tilde{\gamma}\tilde{\alpha}\omega_2\alpha$$

com $\alpha \in \mathcal{A}^k$, $\gamma \in \mathcal{A}^{i-k}$, $\omega_1 \in \mathcal{A}^{n-2i}$ e $\omega_2 \in \mathcal{A}^{n-i-k}$.

Dessa forma:

$$\begin{aligned} \mathbb{P}(\Lambda_{3,4}^n(i) \cap \Lambda_1^n(k)) &= \frac{t_n^2}{1-2t_n} (a^k a^{i-k} a^{n-2i} a^{n-i-k}) + r \\ &= \frac{t_n^2}{1-2t_n} (a^{2n-2i-k}) + r \\ &\xrightarrow{n \rightarrow \infty} a^{-2i-k} \end{aligned} \quad (3.9)$$

□

Analogamente,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,2}^n(i) \cap \Lambda_1^n(k)) &= \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,3}^n(i) \cap \Lambda_1^n(k)) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,4}^n(i) \cap \Lambda_1^n(k)) \\ &= a^{-2i-k} \end{aligned} \quad (3.10)$$

Lema 3.9 Se $i, k \in \mathbb{N}^*$ e $i > k$, então

$$\mathbb{P}(\Lambda_{1,2,3}^n(i) \cap \Lambda_1^n(k)) \xrightarrow{n \rightarrow \infty} a^{-3i-k}$$

Demonstração: Dado que $(x, y) \in \Lambda_{1,2,3}^n(i) \cap \Lambda_1^n(k)$ implica que o par deve ter a seguinte forma:

$$x = p_{ik}\omega_1\tilde{p}_{ik} \text{ e } y = \tilde{p}_{ik}\omega_2p_{ik}$$

com $\omega_1, \omega_2 \in \mathcal{A}^{n-2i}$, temos que

$$\begin{aligned} \mathbb{P}(\Lambda_{1,2,3}^n(i) \cap \Lambda_1^n(k)) &= \frac{t_n^2}{1-2t_n} (a^{n-2i} a^{n-2i} a^{i-k}) + r \\ &= \frac{t_n^2}{1-2t_n} (a^{2n-3i-k}) + r \\ &\xrightarrow{n \rightarrow \infty} a^{-3i-k} \end{aligned} \quad (3.11)$$

□

Lema 3.10 Dados i e k naturais não nulos com $i > k$, vale a seguinte convergência

$$\mathbb{P}(\bar{R}_n^{(2)}(i) \cap \Lambda_1^n(k)) \xrightarrow{n \rightarrow \infty} a^{-k} (4a^{-i} - 6a^{-2i} + 3a^{-3i})$$

Demonstração:

Como $\bar{R}_n^{(2)}(i)$ é uma união de conjuntos, usamos a fórmula de inclusão-exclusão condicionada à intersecção com $\Lambda_1^n(k)$:

$$\begin{aligned}
\mathbb{P}(\bar{R}_n^{(2)}(i) \cap \Lambda_1^n(k)) &= \mathbb{P}([\Lambda_1^n(i) \cup \Lambda_2^n(i) \cup \Lambda_3^n(i) \cup \Lambda_4^n(i)] \cap \Lambda_1^n(k)) \\
&= \mathbb{P}(\Lambda_1^n(i) \cap \Lambda_1^n(k)) + \mathbb{P}(\Lambda_2^n(i) \cap \Lambda_1^n(k)) \\
&\quad + \mathbb{P}(\Lambda_3^n(i) \cap \Lambda_1^n(k)) + \mathbb{P}(\Lambda_4^n(i) \cap \Lambda_1^n(k)) \\
&\quad - \mathbb{P}(\Lambda_{1,2}^n(i) \cap \Lambda_1^n(k)) - \mathbb{P}(\Lambda_{3,4}^n(i) \cap \Lambda_1^n(k)) \\
&\quad - \mathbb{P}(\Lambda_{1,4}^n(i) \cap \Lambda_1^n(k)) - \mathbb{P}(\Lambda_{2,3}^n(i) \cap \Lambda_1^n(k)) \\
&\quad - \mathbb{P}(\Lambda_{1,3}^n(i) \cap \Lambda_1^n(k)) - \mathbb{P}(\Lambda_{2,4}^n(i) \cap \Lambda_1^n(k)) \\
&\quad + \mathbb{P}(\Lambda_{1,2,3}^n(i) \cap \Lambda_1^n(k)) + \mathbb{P}(\Lambda_{1,2,4}^n(i) \cap \Lambda_1^n(k)) \\
&\quad + \mathbb{P}(\Lambda_{1,3,4}^n(i) \cap \Lambda_1^n(k)) + \mathbb{P}(\Lambda_{2,3,4}^n(i) \cap \Lambda_1^n(k)) \\
&\quad - \mathbb{P}(\Lambda_{1,2,3,4}^n(i) \cap \Lambda_1^n(k))
\end{aligned} \tag{3.12}$$

Usando os lemas de 3.5 a 3.9, obtemos:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathbb{P}(\bar{R}_n^{(2)}(i) \cap \Lambda_1^n(k)) &= \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_1^n(i) \cap \Lambda_1^n(k)) + 3 \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_2^n(i) \cap \Lambda_1^n(k)) \\
&\quad - 3 \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,2}^n(i) \cap \Lambda_1^n(k)) - 3 \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{3,4}^n(i) \cap \Lambda_1^n(k)) \\
&\quad + 3 \lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{1,2,3}^n(i) \cap \Lambda_1^n(k)) \\
&= a^{-i-k} + 3a^{i-k} - 3a^{-2i-k} - 3a^{-2i-k} + 3a^{-3i-k} \\
&= 4a^{-i-k} - 6a^{-2i-k} + 3a^{-3i-k} \\
&= a^{-k} (4a^{-i} - 6a^{-2i} + 3a^{-3i})
\end{aligned} \tag{3.13}$$

□

Definição 3.11

$$c_{k,n} = \sum_{i=k+1}^{n-1} \left[\mathbb{P}(\bar{R}_n^{(2)}(i)) - \mathbb{P}(\bar{R}_n^{(2)}(i) \cap \Lambda_1^n(k)) \right]$$

$$c_k = \lim_{n \rightarrow \infty} c_{k,n}$$

Daí, segue a seguinte proposição:

Proposição 3.12 *Para qualquer $k \in \mathbb{N}^*$, temos que*

$$c_k = (1 - a^{-k}) \left(\frac{4a^{-k}}{a-1} - \frac{6a^{-2k}}{a^2-1} + \frac{3a^{-3k}}{a^3-1} \right)$$

Demonstração: Dado que

$$c_{k,n} = \sum_{i=k+1}^{n-1} \left[\mathbb{P}(\bar{R}_n^{(2)}(i)) - \mathbb{P}(\bar{R}_n^{(2)}(i) \cap \Lambda_1^n(k)) \right]$$

precisamos verificar a convergência da segunda parte da equação. Usando o teorema 2.9 e o lema 3.10:

$$\begin{aligned} c_{k,n} &= \sum_{i=k+1}^{n-1} \left[\mathbb{P}(\bar{R}_n^{(2)}(i)) - \mathbb{P}(\bar{R}_n^{(2)}(i) \cap \Lambda_1^n(k)) \right] \\ &\xrightarrow{n \rightarrow \infty} \sum_{i=k+1}^{\infty} \left[(4a^{-i} - 6a^{-2i} + 3a^{-3i}) - a^{-k}(4a^{-i} - 6a^{-2i} + 3a^{-3i}) \right] \\ &= (1 - a^{-k}) \sum_{i=k+1}^{\infty} (4a^{-i} - 6a^{-2i} + 3a^{-3i}) \\ &= (1 - a^{-k}) \left(4 \sum_{i=k+1}^{\infty} a^{-i} - 6 \sum_{i=k+1}^{\infty} a^{-2i} + 3 \sum_{i=k+1}^{\infty} a^{-3i} \right) \\ &= (1 - a^{-k}) \left(\frac{4a^{-k}}{a-1} - \frac{6a^{-2k}}{a^2-1} + \frac{3a^{-3k}}{a^3-1} \right) \end{aligned} \tag{3.14}$$

□

3.3 Um minorante para a cauda da distribuição do máximo encaixe

Encontrado um majorante para b_k , é preciso estabelecer um minorante para o mesmo a fim de estreitar nosso conhecimento sobre o comportamento da variável. Observe que:

$$\begin{aligned} b_{k,n} &= \mathbb{P}(\cup_{i=k+1}^{n-1} (\bar{R}_n^{(2)}(i) \setminus \bar{R}_n^{(2)}(k))) \\ &\geq \mathbb{P}(\bar{R}_n^{(2)}(k+1) \setminus \bar{R}_n^{(2)}(k)) \\ &= \mathbb{P}(\bar{R}_n^{(2)}(k+1)) - \mathbb{P}(\bar{R}_n^{(2)}(k+1) \cap \bar{R}_n^{(2)}(k)) \end{aligned} \tag{3.15}$$

Utilizando, o teorema 2.9, temos que

$$\mathbb{P}(\bar{R}_n^{(2)}(k+1)) \xrightarrow{n \rightarrow \infty} 4a^{-k-1} - 6a^{-2k-2} + 3a^{-3k-3}$$

Portanto, temos que obter o comportamento assintótico de $\mathbb{P}(\bar{R}_n^{(2)}(k+1) \cap \bar{R}_n^{(2)}(k))$, o que será apresentado no seguinte lema:

Lema 3.13 *Dado $k \in \mathbb{N}$, vale:*

$$\mathbb{P}(\bar{R}_n^{(2)}(k+1) \cap \bar{R}_n^{(2)}(k)) \xrightarrow{n \rightarrow \infty} 16a^{-2k-1} - 24a^{3k-1} - 16a^{-3k-2} + 12^{-4k-1} + 16a^{-4k-2} + 3a^{-4k-3}$$

Demonstração:

Utilizando as equivalências apresentadas na secção anterior e a fórmula de inclusão-exclusão é fácil notar que

$$\begin{aligned}
\mathbb{P}(\bar{R}_n^{(2)}(k+1) \cap \bar{R}_n^{(2)}(k)) &= 4\left(4\mathbb{P}(\Lambda_1^n(k+1) \cap \Lambda_1^n(k)) - \right. \\
&\quad - 6\mathbb{P}(\Lambda_{1,2}^n(k+1) \cap \Lambda_1^n(k)) + 3\mathbb{P}(\Lambda_{1,2,3}^n(k+1) \cap \Lambda_1^n(k))\left. \right) - \\
&\quad - 4\left(4\mathbb{P}(\Lambda_1^n(k+1) \cap \Lambda_{1,2}^n(k)) - 4\mathbb{P}(\Lambda_{1,3}^n(k+1) \cap \Lambda_{1,2}^n(k)) + \right. \\
&\quad - 2\mathbb{P}(\Lambda_{1,2}^n(k+1) \cap \Lambda_{1,2}^n(k)) + 3\mathbb{P}(\Lambda_{1,2,3}^n(k+1) \cap \Lambda_{1,2}^n(k))\left. \right) - \\
&\quad - 2\left(4\mathbb{P}(\Lambda_1^n(k+1) \cap \Lambda_{1,3}^n(k)) - 4\mathbb{P}(\Lambda_{1,2}^n(k+1) \cap \Lambda_{1,3}^n(k)) + \right. \\
&\quad - 2\mathbb{P}(\Lambda_{1,3}^n(k+1) \cap \Lambda_{1,3}^n(k)) + 3\mathbb{P}(\Lambda_{1,2,3}^n(k+1) \cap \Lambda_{1,3}^n(k))\left. \right) + \\
&\quad + 3\left(4\mathbb{P}(\Lambda_1^n(k+1) \cap \Lambda_{1,2,3}^n(k)) - 4\mathbb{P}(\Lambda_{1,2}^n(k+1) \cap \Lambda_{1,2,3}^n(k)) + \right. \\
&\quad - 2\mathbb{P}(\Lambda_{1,3}^n(k+1) \cap \Lambda_{1,2,3}^n(k)) + 3\mathbb{P}(\Lambda_{1,2,3}^n(k+1) \cap \Lambda_{1,2,3}^n(k))\left. \right) \xrightarrow{n \rightarrow \infty} \\
&\quad \xrightarrow{n \rightarrow \infty} 4\left(4a^{-2k-1} - 6a^{-3k-2} + 3a^{-4k-3}\right) \\
&\quad - 4\left(4a^{-3k-1} - 4a^{-4k-2} - 2a^{-3k-2} + 3a^{-4k-3}\right) - \\
&\quad - 2\left(4a^{-3k-1} - 4a^{-4k-2} - 2a^{-4k-2} + 3a^{-4k-3}\right) + \\
&\quad + 3\left(4a^{-4k-1} - 4a^{-4k-2} - 2a^{-4k-3} + 3a^{-4k-3}\right) = \\
&\quad = 16a^{-2k-1} - 24a^{-3k-1} - 16a^{-3k-2} + 12^{-4k-1} + 16a^{-4k-2} + 3a^{-4k-3}
\end{aligned} \tag{3.16}$$

□

Portanto, se $d_{k,n}$ representa o minorante encontrado na equação 3.15 e $d_{k,n} \xrightarrow{n \rightarrow \infty} d_k$, temos:

$$d_k = 4a^{-k-1} - (6a^{-2} + 16a^{-1})a^{-2k} + (24a^{-1} + 16a^{-2} + 3a^{-3})a^{-3k} - (12a^{-1} + 16a^{-2} + 3a^{-3})a^{-4k} \tag{3.17}$$

3.4 Comparação de valores assintóticos

Nesta secção, provaremos que $a_k > b_k$. Começaremos por definir

Teorema 3.14 $a_k > b_k \forall k \in N^*$

Demonstração: Demonstraremos pela transitividade, ou seja, provaremos a seguinte desigualdade:

$$a_k > c_k > b_k$$

Como, por definição, $c_{n,k} \geq b_{k,n}$ para qualquer n , vale que

$$c_k = \lim_{n \rightarrow \infty} c_{k,n} \geq \lim_{n \rightarrow \infty} b_{k,n} = b_k$$

Dessa forma, basta provar que $a_k > c_k$. Podemos fazê-lo ao verificar as funções em k que resultam do teorema 2.9 e da proposição 3.12. Podemos ver, pela figura 3.1, que, para todo k , essa desigualdade vale:

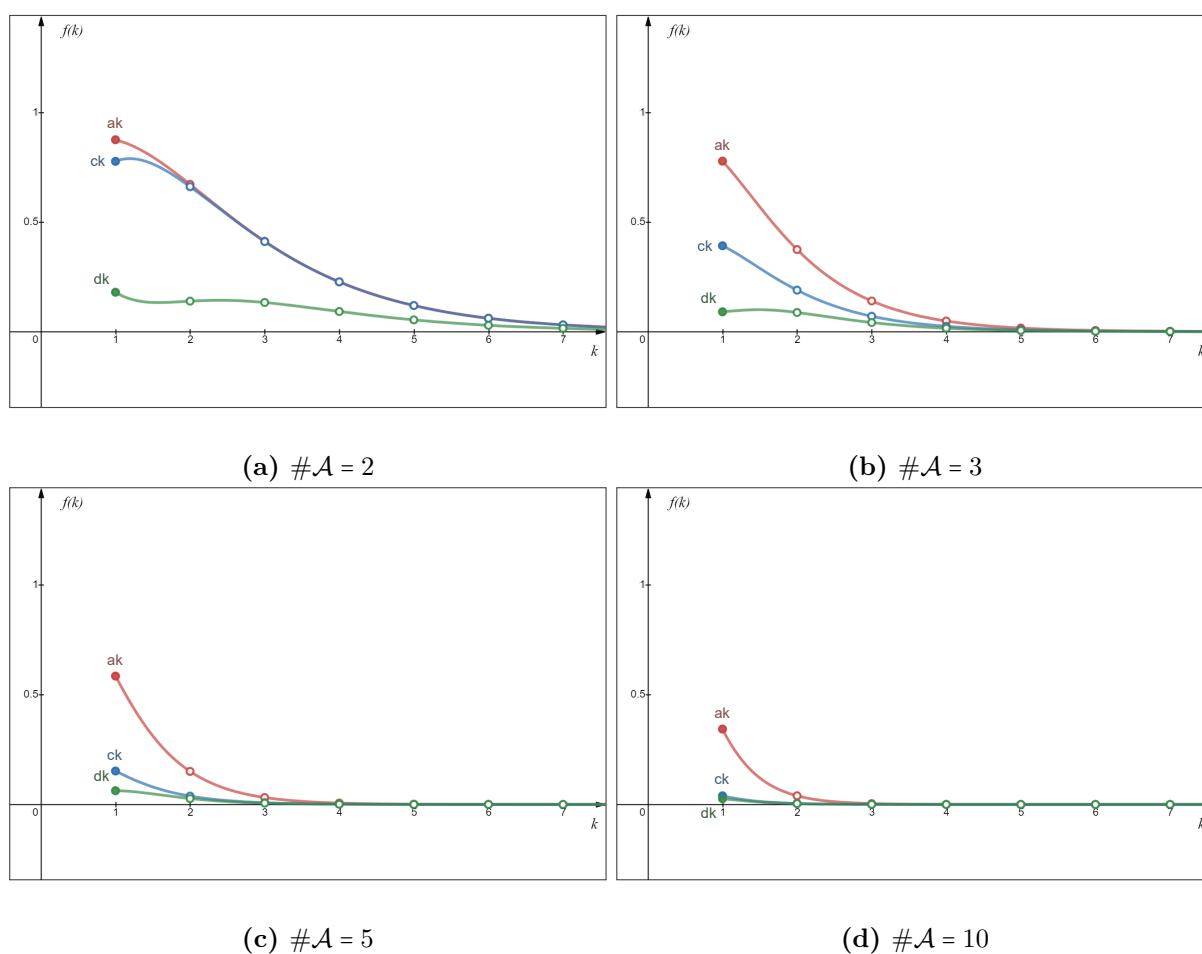


Figura 3.1: Comparação dos termos de convergência a_k, c_k e d_k para tamanhos diferentes de \mathcal{A}

□

3.5 Análise do Termo dominante

Uma forma de comparar a_k , c_k e d_k é a partir do termo dominante de cada variável. Observe que

- $a_k = 4a^{-k} + o(a^{-k})$

- $c_k = \frac{4a^{-k}}{a-1} + o(a^{-k})$

- $d_k = \frac{4a^{-k}}{a} + o(a^{-k})$

Em outras palavras, as três variáveis podem ser escritas como

$$c \cdot a^{-k} + o(a^{-k})$$

em que c é uma constante que depende apenas do tamanho do alfabeto e a notação " $o(a^{-k})$ " representa um resíduo que tende a zero mais rápido que a^{-k} , ou seja, todas as variáveis vão a zero com a mesma taxa, salvo por uma constante. Perceba que no caso em que $a = 2$, como $a - 1 = 1$, c_k e a_k têm o mesmo termo dominante, o que explica o comportamento semelhante entre as duas curvas. Quando $a > 2$, temos:

$$|(a-1) - a| < |(a-1) - 1|$$

isto é, c_k começa a se aproximar d_k e se afastar de a_k quando a cresce, o que pode ser observado na figura 3.1.

Como b_k não pode ser obtida analiticamente, realizamos simulações para obter seu ajuste aos limitantes. Utilizamos amostras de tamanho 50000 e $n = 70$, observando

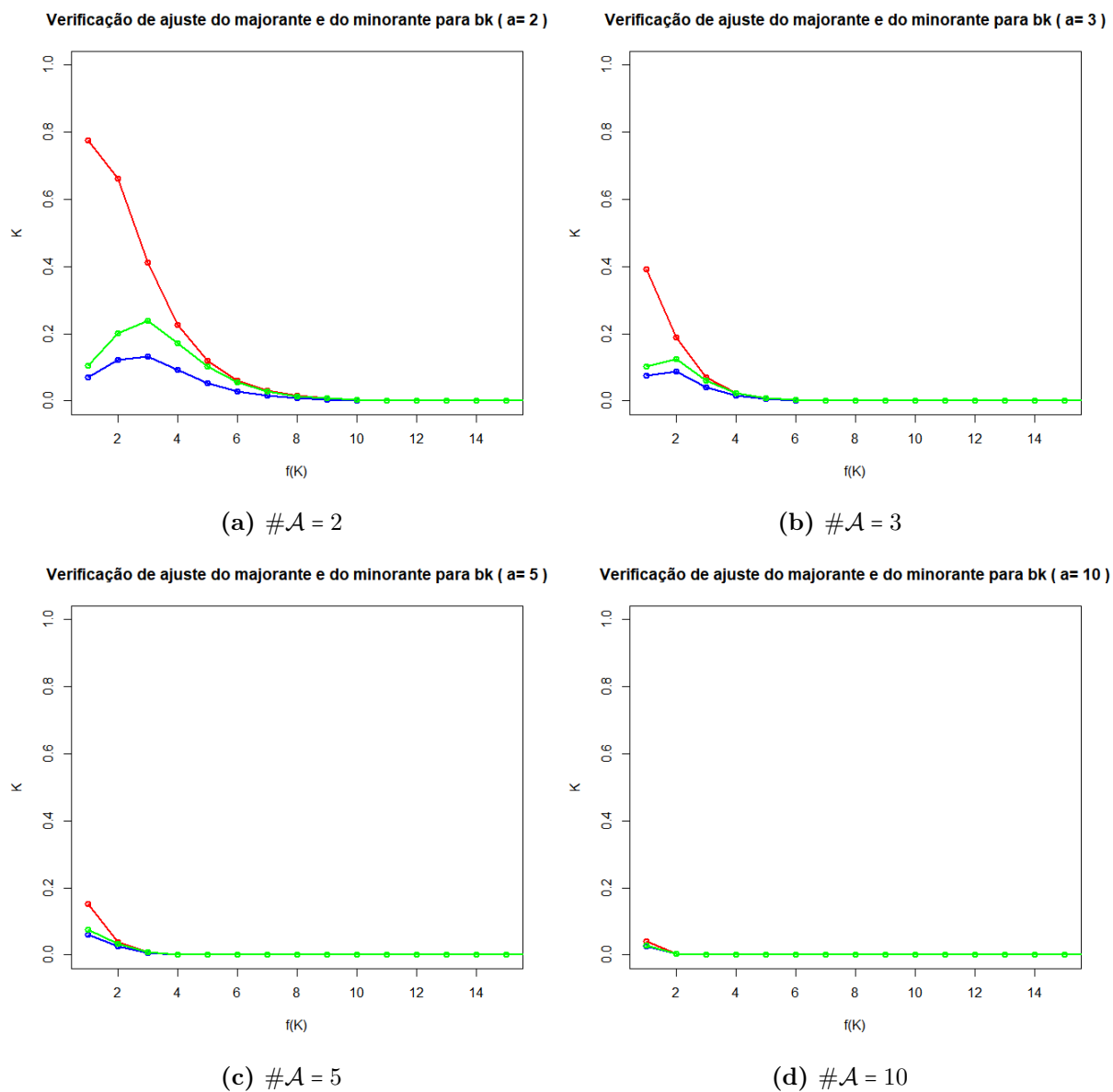


Figura 3.2: Comparação entre b_k (em verde), c_k (em vermelho) e d_k (em azul) para tamanhos diferentes de \mathcal{A}

Conclusões

No trabalho apresentado, foi possível transformar um problema da Teoria de Recorrência de Poincaré em um problema combinatório. Isso foi feito usando uma medida conveniente e hipóteses fortes, principalmente na escolha dos pares e na construção das palavras. Também obtivemos o fato de que uma medida sobre palavras de tamanho n qualquer não é uma medida marginal em relação à medida sobre palavras de tamanho $n + 2$, por exemplo. Alguns resultados interessantes podem seguir do relaxamento de uma ou mais dessas hipóteses, incluindo a construção de uma medida que satisfaça essa condição de marginalidade. Os trabalhos futuros podem tratar das relações entre M_n e n , procedendo como em [Abadi e Lambert \(2019\)](#) e [Abadi e Lambert \(2013\)](#), em que os autores chegaram a

$$\frac{T_n}{n} \xrightarrow{n \rightarrow \infty} 1$$

e

$$\frac{T_n^{(2)}}{n} \xrightarrow{n \rightarrow \infty} 1$$

Referências Bibliográficas

- Abadi e Lambert(2013)** Miguel Abadi e Rodrigo Lambert. The distribution of the short-return function. *Nonlinearity*, 26(5):1143. Citado na pág. 2, 33
- Abadi e Lambert(2019)** Miguel Abadi e Rodrigo Lambert. From the divergence between two measures to the shortest path between two observables. *Ergodic Theory and Dynamical Systems*, 39:1729–1744. Citado na pág. 1, 2, 10, 33
- Abadi e Vaienti(2008)** Miguel Abadi e Sandro Vaienti. Large deviations for short recurrence. *Discrete and Continuous Dynamical Systems-Series A*, 21(3):729–747. Citado na pág. 2
- Abadi et al.(2018)** Miguel Abadi, Sandro Gallo e Erika Alejandra Rada-Mora. The shortest possible return time of β -mixing processes. *IEEE Transactions on Information Theory*, 64(7):4895–4906. Citado na pág. 2
- Abadi e Cardeno(2015)** Miguel Natalio Abadi e Liliam Cardeno. Rényi entropies and large deviations for the first match function. *IEEE Transactions on Information Theory*, 61(4):1629–1639. Citado na pág. 2
- Afraimovich et al.(2003)** Valentin Afraimovich, Jean René Chazottes e Benoît Saussol. Pointwise dimensions for poincaré recurrences associated with maps and special flows. *Discrete and Continuous Dynamical Systems*, 9(2):263–280. Citado na pág. 2
- Cording e Knudsen(2016)** Patrick Hagge Cording e Mathias Bæk Tejs Knudsen. Maximal unbordered factors of random strings. Em *International Symposium on String Processing and Information Retrieval*, páginas 93–96. Springer. Citado na pág. 7
- Haydn e Vaienti(2010)** Nicolai Haydn e Sandro Vaienti. The rényi entropy function and the large deviation of short return times. *Ergodic Theory and Dynamical Systems*, 30(1):159–179. Citado na pág. 2
- Saussol et al.(2002)** B Saussol, S Troubetzkoy e S Vaienti. Recurrence, dimensions, and lyapunov exponents. *Journal of Statistical Physics*, 106(3-4):623–634. Citado na pág. 2