

**Teoremas fundamentais  
para o caminho mais curto  
entre duas sequências**

Rodrigo Lambert

TESE APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
DOUTOR EM ESTATÍSTICA

Programa: Estatística  
Orientador: Prof. Dr. Miguel Natalio Abadi  
Orientador: Prof. Dr. Sandro Vaienti

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da  
CAPES/CNPq/Cnrs/BREUDS

São Paulo, 25 de fevereiro de 2015



**THÈSE DE DOCTORAT DE  
L'UNIVERSITÉ DE TOULON**

Spécialité

**Mathématiques**

École doctorale Mer et Sciences 548

Présentée par

**Rodrigo LAMBERT**

Pour obtenir le grade de

**DOCTEUR en MATHÉMATIQUES**

Sujet de la thèse :

**Teoremas fundamentais para o caminho  
mais curto entre duas sequências**

soutenue le 17 Juin 2015  
devant le jury composé de:

M. Sandro VAIENTI	Directeur de thèse
M. Miguel ABADI	Directeur de thèse
M. Nicolai HAYDN	Rapporteur
M. Aleksandro GALLO	Rapporteur
Mme. Nancy Lopes GARCIA	Examinatrice
M. Fábio Armando TAL	Examineur
M. Cesar Octavio MALDONADO	Examineur
M. Walter ASCHBACHER	Examineur



# Teoremas fundamentais para o caminho mais curto entre duas sequências

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 17/06/2015. 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. Sandro Vaienti (orientador) - UTLN - França
- Profa. Dra. Nancy Lopes Garcia - IMECC - UNICAMP
- Prof. Dr. Fábio Armando Tal - IME-USP
- Prof. Dr. Cesar Octavio Maldonado - CMM - UCHILE
- Prof. Dr. Walter Aschbacher - UTLN - França









# Agradecimentos

Ao Professor Miguel Abadi pelo apoio durante o Mestrado e o Doutorado. Por toda sua simplicidade, disposição, e vontade de passar conhecimentos e técnicas matemáticas para seus alunos. Além disso, ainda me encorajou a buscar novos horizontes em outros grupos de pesquisa, me mandando para trabalhar nos mais diversos lugares. Obrigado por me proporcionar experiências tão enriquecedoras. Além disso, muito obrigado pela sua amizade.

Ao professor Sandro Vaienti pelo apoio nos períodos que passei no Centre de Physique Théorique(CPT), e pelas dicas e sugestões sobre quais problemas explorar.

Ao professor Benoit Saussol, que esteve disponível para discutir matemática e apontar erros em minhas contas(pra depois corrigí-los), durante uma curta visita a Brest-França.

À Karina Yuriko, pelas discussões no NUMEC, e, em especial, pela ajuda no Lema 2.17.

À Alejandra Rada, pelas discussões (intermináveis), e por usar sua capacidade incontestável de fazer continhas complicadas para me ajudar.

A Cesar Maldonado e Nancy Lopes Garcia, pelas correções no texto e, principalmente, pelas correções na notação.

À Cordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Centre national de la recherche scientifique(Cnrs), e ao Brazilian-European partnership in Dynamical Systems(BREUDS) pelo apoio financeiro durante o desenvolvimento deste trabalho.

Ao Núcleo de Modelagem Estocástica e Complexidade (NUMEC IME-USP), e ao Centre de Physique Théorique(CPT), por cederem o espaço para desenvolvimento do presente trabalho.

À Universidade Federal de Uberlândia(UFU), por ter me cedido um período de afastamento para terminar a Tese.

Aos professores Fábio Machado, Vladimir, Antônio, Anatoli, Luiz Renato, Fábio Tal e Yuri pelas aulas e pelo conhecimento que (tentaram) me fizeram adquirir ao longo das disciplinas.

Aos amigos de república em São Paulo: Thiago, Pedro, Tiago, Evander, Sarmiento, e Marquinhos.

Aos meus amigos de Goiânia, muito bem representados por Felipe, Fábio, David, Erick, Joaquim e Tiago.

Aos meus amigos de IME/NUMEC: Manuel, Aldo, Mariana, Michel, Xavier, Wagner e Douglas.

Ao casal de gaúchos nota dez: Aline e Guilherme.

Aos amigos de república de Campinas Lucas, Lino, Welington, Anderson, Fábio, Marquinhos, Luciano, Elen, Vitão e Mineiro.

E finalmente, à minha família, representada por Thelma mãe), Elizabeth(tia), e Aristarcho(primo).

Esta tese foi desenvolvida como parte das atividades do *Centro de Neuromatemática* da **FAPESP**(grant#2013/07699 – 0 , S.Paulo Research Foundation).

As visitas ao CPT em Marselha foram parcialmente financiadas pelas bolsas **CNPq** SWE-236825/2012-7, e **BREUDS** FP7-PEOPLE-2012-IRSES 318999.

# Resumo

LAMBERT, R. **Teoremas fundamentais para o caminho mais curto entre duas sequências**. 2015. 57 f. Tese - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. Université du sud Toulon Var, Toulon, 2015.

Seja  $\mathcal{C}$  um alfabeto finito ou infinito enumerável, e considere  $\mathcal{C}^{\mathbb{N}} := \cup_{n=1}^{\infty} \mathcal{C}^n$  como o conjunto das sequências formadas por concatenação de símbolos de  $\mathcal{C}$ .

No presente trabalho, nós consideramos a função  $T_n^{(2)}$ , definida em  $\mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ , e que toma valores inteiros entre 1 e  $\infty$ .  $T_n^{(2)}$  será o primeiro tempo que demora para que uma sequência de tamanho  $n$ , digamos  $x_1^n$ , apareça sobre uma sequência infinita do processo que começa com  $y_1^n$ . A essa função daremos o nome de *caminho mais curto entre  $x_1^n$  e  $y_1^n$* .

A variável aleatória  $T_n^{(2)}$  é uma extensão do tempo de primeiro retorno  $T_n$ , que se define considerando  $x_1^n = y_1^n$ . Enquanto  $T_n$  considera o conjunto de partida e o conjunto-alvo iguais,  $T_n^{(2)}$  permite que eles tenham naturezas diferentes.

Em termos topológicos:  $T_n$  descreve uma propriedade local do processo, enquanto  $T_n^{(2)}$  trata uma propriedade global.

Saussol *et al.* [2002], Afraimovich *et al.* [2003] provaram que  $T_n/n$  converge quase certamente para 1, para processos ergódicos, com entropia positiva e que satisfazem a propriedade de especificação. Abadi e Vaienti [2008], Haydn e Vaienti [2010], Abadi e Cardeño [2015] estabeleceram um princípio de grandes desvios para  $T_n/n$ , e o relacionaram com a entropia de Rényi do processo. Abadi e Lambert [2013] e Rada [2014] provaram a convergência da distribuição de  $n - T^n$ . O primeiro considera as sequências escolhidas de acordo com a medida produto de  $n$  variáveis aleatórias independentes e identicamente distribuídas no alfabeto  $\mathcal{C}$ . O segundo generalizou o resultado para processos com uma condição de dependência  $\beta$ -mixing.

No presente trabalho, estudamos as propriedades de  $T_n^{(2)}$ , e apresentamos três resultados principais:

O primeiro diz que  $T_n^{(2)}$  se comporta assintoticamente como uma função linear de  $n$ .

O segundo apresenta um princípio de grandes desvios para  $T_n^{(2)}/n$ .

E o terceiro trata a convergência de uma versão re-escalada de  $T_n^{(2)}$ .

**Palavras-chave:** *Mixing*, Tempo de retorno, Caminho mais curto, Convergência.



# Abstract

LAMBERT, R. **Fundamental theorems for the shortest path between two cylinders**. 2015. 57 f. Phd Thesis - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. Université du sud Toulon Var, Toulon, 2015.

Let  $\mathcal{C}$  be a finite or countable alphabet, and consider  $\mathcal{C}^{\mathbb{N}} = \cup_{n=1}^{\infty} \mathcal{C}^n$  the set of all sequences with symbols in  $\mathcal{C}$ .

In this thesis, we consider the random variable  $T_n^{(2)}$ , defined over  $\mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ , and taking integer values between 1 and  $\infty$ .  $T_n^{(2)}(x_1^n, y_1^n)$  will be defined as the shortest time taken to a sequence  $x_1^n$  appears in an infinite sequence that starts with  $y_1^n$ . We will call it *the shortest path between  $x_1^n$  and  $y_1^n$* .

The random variable  $T_n^{(2)}$  is an extension of the *short return function*  $T_n(x_1^n)$ , and they coincide when  $x_1^n = y_1^n$ . While  $T_n$  considers start and target point being the same set,  $T_n^{(2)}$  allows them with different natures.

In topological terms:  $T_n$  describes a local property of the process, while  $T_n^{(2)}$  describes a global property.

Saussol *et al.* [2002], Afraimovich *et al.* [2003] proved that  $T_n/n$  converges 1 almost surely, for ergodic processes, with positive entropy, and satisfying the specification property. Abadi e Vaienti [2008], Haydn e Vaienti [2010], Abadi e Cardeno [2015] proved a large deviation principle for  $T_n/n$ , and related it with the Rényi's entropy of the process. Abadi e Lambert [2013], and Rada [2014] proved the convergence in distribution of  $n - T_n$ . First work considered processes generated by IID random variables, and the second one assumed the  $\beta$ -mixing dependence.

In this work we study the properties of  $T_n^{(2)}$ , and present three main results:

The first one is the asymptotic behavior of  $T_n^{(2)}$  as a linear function of  $n$ .

The second is a Large deviation principle for  $T_n^{(2)}/n$ .

And the third one is about the convergence in distribution of the re-scaled  $T_n^{(2)}$ .

**Keywords:** Mixing, Return times, Shortest path, Convergence.



# Résumé

LAMBERT, R. **Théorèmes fondamentaux pour le plus court chemin entre deux sequences.** 2015. 57 f. Thèse doctorale - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. Université du sud Toulon Var, Toulon, 2015.

Soit  $\mathcal{C}$  est un alphabet finie ou infinie dénombrable, et en prendre  $\mathcal{C}^{\mathbb{N}} := \cup_{n=1}^{\infty} \mathcal{C}^n$  comme l'ensemble des séquences formées par la concaténation des symboles de  $\mathcal{C}$ .

Dans cette thèse, nous considérons la fonction  $T_n^{(2)}$ , définis sur  $\mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ , et qui prend des valeurs entières entre 1 et  $\infty$ .  $T_n^{(2)}$  sera le nombre minimum d'étapes pour une séquence de taille  $n$ , dirons  $x_1^n$  apparaître sur une suite infinie du processus qui commence par  $y_1^n$ . Pour cette fonction en donnera le nom *le plus court chemin* entre  $x_1^n$  et  $y_1^n$ .

La variable aléatoire  $T_n^{(2)}$  est une extension du premier temps de retour  $T_n$ , définit quand  $x_1^n = y_1^n$ . Une fois que  $T_n$  concerne l'ensemble de départ et l'objectif fixé égale,  $T_n^{(2)}$  leur permet d'avoir différentes natures.

En termes topologiques:  $T_n$  décrit une propriété locale du processus, tandis que  $T_n^{(2)}$ , une propriété globale.

Saussol *et al.* [2002], Afraimovich *et al.* [2003] ont prouvé que  $T_n/n$  converge presque partout à 1, pour processus ergodiques avec entropie positif et, en plus, la propriété de specification.

Abadi e Vaienti [2008], Haydn e Vaienti [2010], Abadi e Cardeño [2015] établi un principe de grandes déviations pour  $T_n/n$ , et le rapport avec l'entropie de Rényi du processus.

Abadi e Lambert [2013] e Rada [2014] ont prouvé la convergence de la distribution de  $n - T_n$ . Le premier travail examine les séquences choisie par rapport au mesure-produit de  $n$  variables aléatoires indépendantes et identiquement distribuées dans l'alphabet  $\mathcal{C}$ . Le deuxième a généralisé le résultat pour des processus avec une condition de dépendance  $\beta$ -mélangeant.

Dans ce travail, nous étudions les propriétés de  $T_n^{(2)}$ , et en présente trois principaux résultats:

Le premier est le comportement asymptotique de  $T_n^{(2)}$  comme une fonction linéaire de  $n$ .

Le deuxième est un principe de grandes déviations pour  $T_n^{(2)}/n$ .

Et le troisième est de la convergence en distribution d'une version re-mise à l'échelle de  $T_n^{(2)}$ .

**Mots-clés:** *Mixing*, Temps de retour, plus court chemin, Convergence.





# Sumário

<b>Resumo</b>	<b>xi</b>
<b>Abstract</b>	<b>xiii</b>
<b>Résumé</b>	<b>xv</b>
<b>Lista de Figuras</b>	<b>xix</b>
<b>Introdução</b>	<b>xxi</b>
<b>1 O caminho de retorno mais curto</b>	<b>1</b>
1.1 Definição Combinatória . . . . .	2
1.2 Definição Dinâmica . . . . .	4
1.3 A propriedade de especificação . . . . .	5
1.4 Resultados anteriores: o estado da arte . . . . .	6
1.4.1 Convergência forte . . . . .	6
1.4.2 Princípio de grandes desvios . . . . .	7
1.4.3 Convergência fraca . . . . .	8
<b>2 O menor caminho entre duas sequências</b>	<b>13</b>
2.1 Definições preliminares . . . . .	13
2.1.1 Linguagem combinatória . . . . .	13
2.1.2 Linguagem dinâmica . . . . .	15
2.2 A divergência relativa entre duas medidas . . . . .	16
2.2.1 A divergência relativa . . . . .	16
2.2.2 Propriedades . . . . .	16
2.3 O crescimento linear . . . . .	17
2.4 Grandes Desvios . . . . .	20
2.5 A distribuição do caminho mais curto . . . . .	24
2.5.1 Não-convergência em probabilidade . . . . .	27
<b>3 Conclusões</b>	<b>29</b>
<b>Referências Bibliográficas</b>	<b>31</b>



# Lista de Figuras

1.1	Saindo de 000 e retornando nele mesmo. O menor caminho leva um passo. Logo, $T_3(000) = 1$ . . . . .	3
1.2	Saindo de 001 e retornando nele mesmo. O menor caminho possui três passos. Logo, $T_3(001) = 3$ . . . . .	3
1.3	Saindo de 010 e retornando nele mesmo. O menor caminho possui dois passos. Logo, $T_3(010) = 2$ . . . . .	4
1.4	Processo com $T_n$ maior do que $n$ . . . . .	4
1.5	Representação gráfica da entropia de Rényi. . . . .	8
1.6	Relação entre $T_n$ e $S_n$ . . . . .	8
1.7	Relação entre $T_n$ e $S_n$ : caso $T_n > n \implies S_n < 0$ . . . . .	9
1.8	Definição do processo $\beta$ -mixing. . . . .	10
2.1	Saindo de 011 e indo até 110. O menor caminho possui um passo. . . . .	14
2.2	Saindo de 010 e indo até 001. O menor caminho possui dois passos. . . . .	14
2.3	Saindo de 000 e indo até 111. O menor caminho possui três passos. . . . .	15



# Introdução

Na análise estatística da Recorrência de Poincaré, é clássico considerar o tempo que uma determinada realização do processo leva até atingir um conjunto de medida positiva. O famoso Teorema de Recorrência de Poincaré diz que quase toda realização atinge infinitas vezes esse conjunto. Desde então, muitos trabalhos têm considerado como conjuntos mensuráveis as sequências de tamanho  $n$ . Precisamente: Seja  $(\Omega, \mathcal{F})$  um espaço mensurável, onde  $\Omega = \mathcal{C}^{\mathbb{N}}$  é o conjunto das sequências infinitas formadas por concatenação de símbolos do alfabeto  $\mathcal{C}$ , e  $\mathcal{F} = \sigma(x_0^{n-1})$  é a sigma-álgebra gerada pelos conjuntos  $x_0^{n-1}$ , a quem chamaremos de cilindros, ou palavras, ou sequências. O *tempo de chegada* a  $x_0^{n-1}$  de uma realização  $y_0^\infty$  do processo é definido explicitamente como:

$$\tau_{x_0^{n-1}}(y_0^\infty) = \inf\{t \geq 1 : y_t^{t+n-1} = x_0^{n-1}\}$$

O comportamento da distribuição  $\mathbb{P}(\tau_{x_0^{n-1}} > t)$  é bem conhecido na literatura. É sabido que esse tempo de retorno pode ser aproximado por uma lei exponencial [Abadi \[2001\]](#), [Abadi \[2004\]](#), [Abadi e Saussol \[2010\]](#), [Galves e Schmitt \[1997\]](#).

Do ponto de vista empírico, usar esta quantidade para obter informação sobre o processo que gera os dados tem uma desvantagem: o tempo observado é exponencial em  $n$ , o tamanho da sequência-alvo  $x_0^{n-1}$ . Portanto, esse tempo é tipicamente grande. Isso implica em grandes amostras e grande tempo computacional.

Para evitar esses problemas, se fez necessário introduzir alguma quantidade que traga informação sobre o processo, mas que seja "mais fácil" de ser observada. Aqui, "mais fácil" significa levar menos tempo até ser observada.

Motivados pelos trabalhos [Abadi \[2004\]](#), [Abadi \[2001\]](#), [Abadi e Saussol \[2010\]](#), [Galves e Schmitt \[1997\]](#) sobre tempos de chegada, nós consideramos o primeiro retorno de uma sequência. O primeiro retorno de uma sequência  $x_0^{n-1}$  é definido como o ínfimo tomado entre todas as realizações  $y_0^\infty$  que começam com a palavra  $x_0^{n-1}$ . Ou seja:

$$T_n(x_0^{n-1}) = \inf_{y_0^\infty: y_0^{n-1} = x_0^{n-1}} \tau_{x_0^{n-1}}(y_0^\infty)$$

Assim,  $T_n$  é chamado de *Tempo de Primeiro Retorno* da sequência  $x_0^{n-1}$ .

A relevância da função  $T_n$  foi colocada em evidência na análise estatística da Recorrência de Poincaré. Para provar a convergência do número de ocorrência de palavras (digamos, de tamanho  $n$ ) quando  $n$  diverge, para a distribuição de Poisson, é necessário que a palavra não se sobreponha com ela mesma [Hirata \[1993\]](#). Ou, pelo menos, que a proporção das palavras que se sobrepõem seja pequena com relação a  $n$  [Abadi e Vienti \[2008\]](#). Também

temos na literatura algumas aproximações para esse limite em Reinert e Schbath [1998], Reinert e Schbath [2010] e Roquain e Schbath [2007].

Tal relevância também aparece quando consideramos o tempo decorrido até a primeira ocorrência da palavra. Esse tempo é conhecido como o Tempo de Ocorrência ou *Hitting time*. É sabido que o tempo de ocorrência pode ser aproximado por uma lei exponencial com parâmetro dado pela medida da palavra, no caso em que a palavra possui uma sobreposição com ela mesma ou quando tal sobreposição é "pequena". Mas, quando a palavra apresenta sobreposição (que não seja "pequena") com ela mesma, o parâmetro deve ser corrigido por um fator. Tal fator é a probabilidade da palavra não aparecer duas vezes consecutivas. E essa probabilidade é dada pelas propriedades de sobreposição da palavra, Abadi [2001] e Abadi [2008].

Uma situação similar acontece quando consideramos o tempo decorrido até a primeira ocorrência da palavra, colocando como condição inicial a ocorrência da própria palavra. Este tempo é chamado de Tempo de Retorno (*Return Time*). Neste caso, acontece o mesmo. É sabido que o tempo de retorno também pode ser bem aproximado por uma lei exponencial. Mas quando a palavra apresenta sobreposição com ela mesma, essa lei deixa de ser exponencial, passando a ser uma combinação convexa de uma medida degenerada na origem e uma exponencial Galves *et al.* [1999]. Neste caso, o mesmo fator citado no parágrafo anterior, aparece na combinação convexa e a lei exponencial.

Até onde sabemos, os primeiros a observarem que a medida de todas as palavras que possuem "grandes" sobreposições converge para zero foram Collet, Galves & Schmitt, em 1999 Galves *et al.* [1999]. Nesse trabalho, os autores provaram o decaimento exponencial dessa medida quando "grande" significa maior ou igual a  $n - n/3$ . Tal resultado vale para processos misturadores com função  $\psi$  de decaimento exponencial. Mais tarde, o mesmo resultado foi generalizado em Abadi [2001] para processos misturadores com função  $\phi$  de decaimento exponencial. No último caso, "grande" significa ser maior ou igual a  $n - cn$ , onde  $c$  é uma constante que depende da cardinalidade do alfabeto.

A distribuição de  $T_n$  é desconhecida. Ou seja, não há uma forma explícita para

$$F_{T_n}(k) := \mu(T_n \leq k).$$

Em 2002 usando a complexidade de Kolmogorov, Saussol, Troubetzkoy & Vienti Saussol *et al.* [2002] provaram, para um processo ergódico com entropia positiva, cumprindo propriedade de especificação Afraimovich *et al.* [2003], que

$$\lim_{n \rightarrow \infty} \frac{T_n(x_1^n)}{n} = 1 \quad \mu - q.t.p. . \quad (1)$$

Para quase toda realização do processo, esta expressão diz que, assintoticamente,  $T_n$  cresce na mesma velocidade que  $n$ . Isto sugere que a variável aleatória  $T_n$  apresenta um comportamento assintótico defectivo. Ou seja, temos então uma sequência de variáveis aleatórias não *estocasticamente limitadas*. Aí vemos a necessidade de definir convenientemente uma outra sequência de variáveis aleatórias que nos forneça informação sobre o *Primeiro Tempo de Retorno*.

Em 2003, usando o Teorema de Shannon, McMillan & Breiman, Afraimovich, Chazottes e Saussol provaram que vale o mesmo resultado que se apresenta em (1), Afraimovich *et al.* [2003], para a variável aleatória  $T_n^{(2)}$ . Este resultado também foi provado para uma classe

de funções não uniformemente expandidas do intervalo [Hirata \*et al.\* \[1999\]](#) no contexto de sistemas dinâmicos.

O princípio de grandes desvios para  $T_n$  tem sido e continua sendo motivo de estudo. [Abadi e Vaianti \[2008\]](#) encontraram a função de grandes desvios e relacionaram  $T_n$  com a entropia de Renyi do processo gerador de palavras. [Abadi e Cardeño, \[2015\]](#) estudaram a entropia de Renyi e a função de grandes desvios de  $T_n$ , para processos com condições de mistura. [Haydn e Vaianti \[2009\]](#) estudaram o mesmo princípio e sua relação com a entropia de Renyi, mas assumindo condições de decaimento exponencial. Logo o limite da função de grandes desvios esta relacionado com a entropia do processo. Então é natural perguntar pelas flutuações de  $T_n$ . Ou seja, perguntar pela existência da distribuição limite de  $T_n$  re-escalada.

Motivados pelo estudo da distribuição limite de  $n - T_n$  como uma fonte de informação para encontrar propriedades interessantes anunciadas anteriormente de  $T_n$ , [Lambert \[2010\]](#) e [Abadi e Lambert \[2013\]](#) encontraram a convergência da distribuição de  $n - T_n$  para uma distribuição limite cuja cauda decai exponencialmente, quando  $n$  diverge. Isto foi provado considerando um processo independente e identicamente distribuído. Além disso, eles apresentaram um limite superior para a velocidade dessa convergência. Tal limite superior é não-uniforme, porém permite obter como corolário a esperança e, em geral, todos os momentos finitos de ordem polinomial. O mesmo resultado foi obtido em [Rada \[2014\]](#) para processos com a condição de dependência  $\beta$ -mixing.

Agora note que  $T_n$  é uma função que considera condição inicial igual ao conjunto-alvo. Isso impõe um limite à informação que essa quantidade nos fornece sobre o espaço  $(\Omega, \mathcal{F}, \mathbb{P})$ . Como uma ilustração, considere uma cadeia de Markov redutível. Nesse caso, sempre podemos retornar de um estado nele mesmo. Isso evidencia o fato de que  $T_n$  descreve uma propriedade local do espaço topológico subjacente a  $(\Omega, \mathcal{F})$ .

Sendo assim, se faz necessário introduzir uma nova variável aleatória que trate propriedades globais. Pensando nisso, definimos o *caminho mais curto entre duas sequências*  $T_n^{(2)}$ , que é uma função definida sobre o conjunto de todos os pares de sequências  $(x_1^n, y_1^n)$ , e tomando valores inteiros positivos. Tal variável considera que conjunto de partida  $(y_1^n)$  e conjunto de chegada  $(x_1^n)$  podem ser diferentes. Precisamente, os dois conjuntos são gerados por dois processos independentes com medidas  $\mu$  e  $\nu$ . Tais medidas estão definidas sobre o mesmo espaço de probabilidades.

A variável aleatória  $T_n^{(2)}(x_1^n, y_1^n)$  define o menor tempo possível gasto para encontrarmos a palavra  $x_1^n$ , tomado observando entre todas as realizações do processo que se iniciam com a sequência  $y_1^n$ . Formalmente:

$$T_n^{(2)}(x_1^n, y_1^n) = \inf_{z_0^\infty : z_0^{n-1} = y_1^n} \{k \geq 1 : z_k^{n+k-1} = x_0^{n-1}\} .$$

A presente tese tem como principal objetivo estudar a variável aleatória  $T_n^{(2)}$  e suas propriedades, e é organizada da seguinte maneira: No primeiro capítulo, apresentamos algumas definições iniciais para o tempo de primeiro retorno  $T_n$ . Enunciamos, em primeiro lugar, um Teorema de concentração para  $T_n$ . Em segundo lugar, apresentamos um princípio de grandes desvios. Finalmente, apresentamos resultados sobre as flutuações da variável  $T_n$ . Nos dois últimos resultados, ver-se-á uma relação íntima entre as expressões obtidas e a Entropia de Rényi do processo gerador das palavras.

No capítulo 2, apresentamos os resultados obtidos durante o desenvolvimento dessa tese. Tal capítulo possui a mesma estrutura de seu antecessor.

No terceiro capítulo, ponderamos alguns problemas em aberto para trabalhos futuros.



# Capítulo 1

## O caminho de retorno mais curto

Quando mencionamos o termo *retorno*, associamo-lo a conceitos como volta, redundância, regresso ou repetição. Repetição no sentido de voltar a ocupar uma posição já antes ocupada(ou posição repetida), de não acrescentar nova informação(informação redundante), ou simplesmente reencontrar algo já anteriormente notado. A busca por entender os significados por detrás de tais ideias parece ser, por si só, interessante. Poderíamos até ir um pouco mais longe na exploração de tal conceito. Porém, sempre correríamos o risco de acabar *retornando* às ideias supracitadas.

Tentando tratar a noção de retorno de maneira mais objetiva, encontramos no ferramentário matemático uma forma mais precisa de descrevê-la. Assim, podemos tratá-lo como um objeto exato. Depois disso, podemos finalmente seguir nos perguntando sobre propriedades e resultados associados a tal conceito.

### O primeiro retorno de um observável

O objetivo da presente seção é introduzir a noção de primeiro retorno de uma sequência nela própria(ou simplesmente, *caminho de retorno mais curto*). Note que aqui estamos falando de ideias como *repetição de padrão* ou *comportamento cíclico*. Em outras palavras: estamos tratando de uma propriedade local do nosso espaço.

Seja  $\mathcal{C}$  um alfabeto finito ou infinito-enumerável. Defina nosso espaço amostral  $\Omega$  como sendo o conjunto de todas as sequências construídas por concatenação de símbolos do alfabeto  $\mathcal{C}$ . Ou seja

$$\Omega = \mathcal{C}^* = \bigcup_{n=1}^{\infty} \mathcal{C}^n ,$$

onde  $\mathcal{C}^n$  representa o conjunto de todas as sequências de tamanho  $n$  formada por símbolos do alfabeto  $\mathcal{C}$ (e  $\mathcal{C}^{\mathbb{N}}$  o conjunto de todas as sequências infinitas). Chamaremos tais sequências de *palavras*, *cilindros* ou simplesmente *sequências*, e as denotaremos por  $\omega$ ,  $x_1^n$ , ou  $\omega_n$ .

A seguir apresentaremos duas definições para o *Tempo de primeiro retorno*. A primeira, seguirá uma abordagem combinatória do assunto. A segunda tratará a função desde um ponto de vista dinâmico.

### 1.1 Definição Combinatória

Nesta seção, iremos introduzir de forma simples e intuitiva o conceito de *primeiro retorno*. A definição será ilustrada com alguns exemplos.

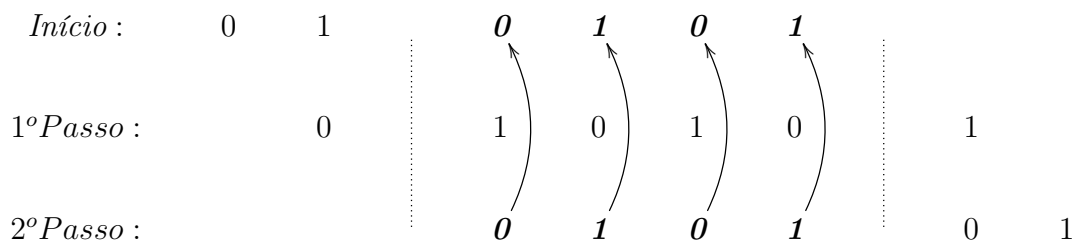
**Definição 1.1.** Chamamos de *primeiro retorno* a função  $T_n : \mathcal{C}^n \rightarrow \{1, 2, \dots, n\}$  definida da seguinte maneira:

$$T_n(x_1^n) = \inf\{k \geq 1 : x_1^{n-k} = x_{k+1}^n\},$$

e, no caso em que aquele ínfimo não é atingido, dizemos simplesmente que  $T_n(x_1^n) = n$ .

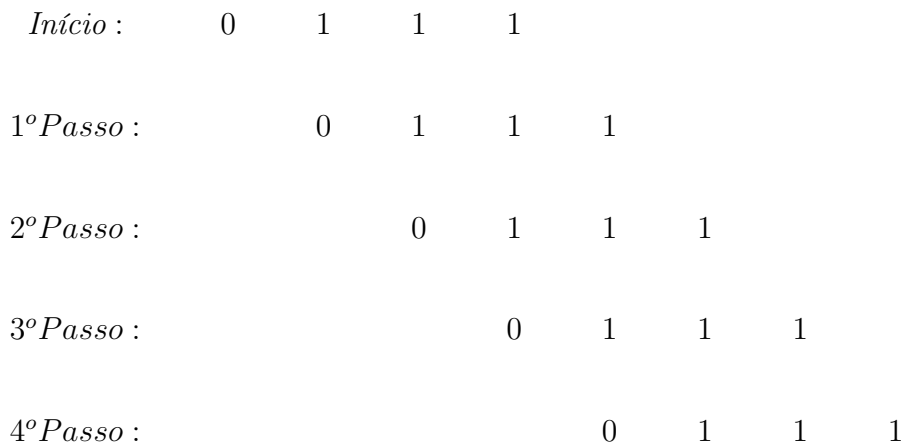
Agora, citamos dois exemplos de primeiro retorno de palavras no alfabeto binário  $\mathcal{C} = \{0, 1\}$ . No primeiro caso, o ínfimo é atingido. Já no segundo, isso não acontece.

**Exemplo 1.2.** Seja a palavra  $x_1^6 = 010101$ :



Portanto, como a palavra realizou um encaixe no segundo passo (ou segunda translação), temos que o tempo de encaixe de  $x_1^6$  é igual a 2.

**Exemplo 1.3.** Seja a palavra  $x_1^4 = 0111$ :



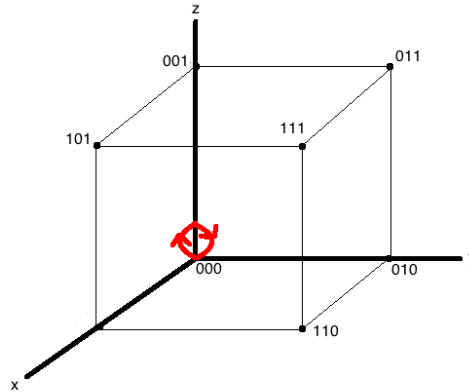
Portanto, como a palavra não realizou encaixe, temos que o tempo de encaixe de  $x_1^4$  é igual a 4, que é o tamanho da palavra.

Uma outra forma de imaginar o retorno de uma sequência nela mesma, seria imaginar um passeio no hipercubo  $n$ -dimensional, com ponto de partida igual ao ponto de chegada, definidos pela sequência  $x_1^n$ . O tempo de retorno fica então representado pela mínima quantidade de passos para realizar esse retorno.

Note que, no nosso caso, o caminho percorrido nas arestas do cubo não ocorre da mesma forma que no caso do passeio clássico no hibercubo(ver [Simonis \[1998\]](#), por exemplo). Aqui, são permitidos saltos de uma aresta para uma outra que não seja vizinha. Na verdade, no

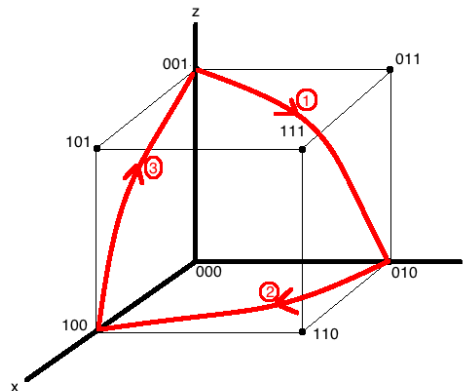
nosso caso, como o processo consiste em apagar a letra do começo, e concatenar uma letra no final, isso nem acontece. Os exemplos a seguir ilustram bem essa afirmação:

**Exemplo 1.4.** Se  $x_1^3 = 000$ , então ele retorna nele mesmo já na primeira tentativa. Ou seja,  $T_3(000) = 1$ .



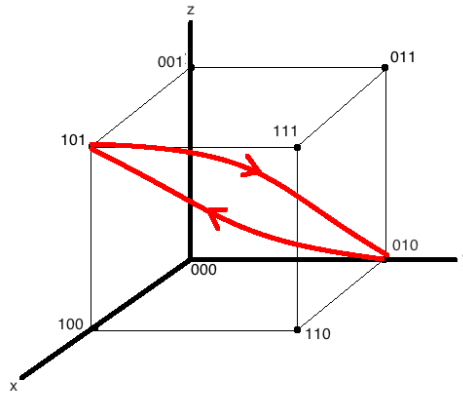
**Figura 1.1:** Saindo de 000 e retornando nele mesmo. O menor caminho leva um passo. Logo,  $T_3(000) = 1$ .

**Exemplo 1.5.** Seja a sequência  $x_1^3 = 001$ . Se nós concatenarmos símbolos do alfabeto  $\{0, 1\}$  de forma a vermos aparecer novamente a sequência 001 da forma mais rápida possível, construiremos a sequência: 001001, que nada mais é do que percorrer o caminho passando pelos pontos 001 – 010 – 100 – 001.



**Figura 1.2:** Saindo de 001 e retornando nele mesmo. O menor caminho possui três passos. Logo,  $T_3(001) = 3$ .

**Exemplo 1.6.** A mesma ideia vale para o exemplo 1.2, onde retornamos a 010 em dois passos. O caminho percorrido fica definido da seguinte maneira: 010 – 101 – 010.



**Figura 1.3:** Saindo de 010 e retornando nele mesmo. O menor caminho possui dois passos. Logo,  $T_3(010) = 2$ .

## 1.2 Definição Dinâmica

Embora a definição combinatória seja prática e de fácil entendimento, ela não abrange alguns tipos de processos interessantes. Quando definimos (de maneira arbitrária) a variável aleatória  $T_n$  com contra-domínio no conjunto  $\{1, 2, \dots, n\}$ , deixamos de contemplar casos em que a quantidade de passos até retornarmos a  $x_1^n$  é maior do que  $n$ . O problema é que esses casos são bem comuns na literatura (vide próximo exemplo). Isso nos motiva a buscar uma definição mais abrangente de  $T_n$ .

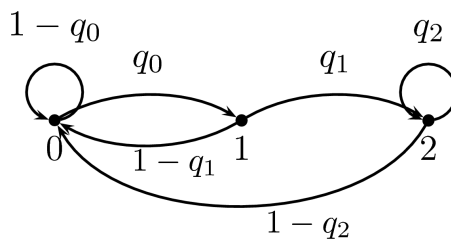
A seguir, daremos um exemplo onde  $T_n > n$ . Depois, daremos uma *definição mais abrangente* de  $T_n$ , a qual daremos o nome de *definição dinâmica*.

No seguinte exemplo mostramos que o *tempo de retorno* de uma palavra pode ser maior do que o tamanho da palavra.

**Exemplo 1.7.** Considere uma cadeia de Markov em  $\mathcal{C} = \{0, 1, 2\}$  tal que

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{pmatrix} 1 - q_0 & q_0 & 0 \\ 1 - q_1 & 0 & q_1 \\ 1 - q_2 & 0 & q_2 \end{pmatrix} \end{matrix}$$

onde  $0 < q_i < 1$ , para  $i = 0, 1, 2$ .



**Figura 1.4:** Processo com  $T_n$  maior do que  $n$ .

Seja  $x_1^4 = 1012$ . Note que nenhuma palavra  $y_1^\infty$  que comece com 1012 vai repeti-la imediatamente, pois  $p_{2,1} = 0$ . Logo,  $T_4(1012) > 4$ .

Agora, definimos o *tempo de chegada*, que é o tempo mínimo observado em uma realização infinita do processo, até vermos a primeira aparição da palavra  $x_1^n$ .

**Definição 1.8.** Dizemos que o tempo de chegada de uma realização  $z = z_1^\infty$  a  $x_1^n$  é o tempo mínimo (em quantidade de translações  $T$ ) no decorrer do processo, até encontrarmos pela primeira vez a sequência  $x_1^n$ . Ou seja

$$\tau_{x_1^n}(z) = \inf\{k \geq 1 : z_k^{k+n-1} = x_1^n\}$$

Uma vez definido o tempo de chegada, podemos definir o primeiro retorno, que é o mínimo entre todos os tempos de chegada a  $x_1^n$ , para processos que comecem em  $x_1^n$ .

**Definição 1.9.** Seja  $x_1^n$  uma palavra em  $\Omega$ . Defina

$$T_n(x_1^n) = \inf_{z: z_1^n = x_1^n} \tau_{x_1^n}(z) .$$

Uma vez definido o caminho de retorno mais curto, vamos seguir com o nosso texto. Na seção seguinte, definiremos uma condição que é usada como hipótese para um resultado que será apresentado na presente tese.

## 1.3 A propriedade de especificação

Aqui, apresentamos duas definições que serão bastante usadas ao longo do nosso texto, e que constituirão parte das hipóteses de nossos resultados.

**Definição 1.10.** Sejam  $\{X_n\}_{n \in \mathbb{N}}$  um processo estocástico definido sobre o alfabeto  $\mathcal{C}$  e  $\mu$  uma medida de probabilidade do processo em  $\mathcal{C}^{\mathbb{N}}$ . Diremos que o processo tem gramática completa se para todo  $x_1^n \in \mathcal{C}^n$ ,  $\mu(x_1^n) > 0$ .

Aqui definiremos a propriedade de especificação. Tal propriedade nos garante que podemos concatenar ao nosso cilindro  $x_1^n$  uma outra sequência  $x_{n+1}^\infty$  de tal forma que possamos encontrar novamente  $x_1^n$  em  $x_2^n x_{n+1}^\infty$ .

Primeiramente, vamos comentar algumas notações de dinâmica:

Seja  $(\Omega, T)$  um sistema dinâmico, onde  $(\Omega, d)$  é um espaço métrico completo e compacto, e  $T$  uma aplicação contínua de  $\Omega$  em si mesmo. Seja  $d$  a distância em  $(\Omega, d)$ . Denotamos por uma *corda*  $A = [a, b]$ , o conjunto finito de inteiros consecutivos  $\{a, a + 1, \dots, b\}$ , e por um *pedaço de órbita* o conjunto  $\{T^j x : x \in \Omega, j \in A\}$

**Observação 1.11.** Note que, se o processo é ergódico e, além disso,  $|\mathcal{C}| \leq \infty$ , então a condição de especificação é automaticamente satisfeita.

**Observação 1.12.** A notação  $T$  para a transformação de  $\Omega$  em si mesmo pode gerar confusão com a variável aleatória  $T_n$ . Pedimos ao leitor a atenção em observar o seguinte fato: sempre que houver um sub-índice em  $T$  (a saber:  $T_n$ ), estaremos nos referindo a uma variável aleatória, e não ao shift  $T$  no espaço  $(\Omega, T)$ .

**Definição 1.13.** *Sigmund [1974]* Dizemos que  $(\Omega, T)$  possui a propriedade de especificação  $((\Omega, T) \in \text{SPEC})$  se, para todo  $\epsilon > 0$  existir um inteiro  $M(\epsilon)$  tal que, para quaisquer dois

pontos  $x_1, x_2 \in \Omega$  e quaisquer cordas  $A_1 = [a_1, b_1]$  e  $A_2 = [a_2, b_2]$ , com  $a_2 - b_1 > M(\epsilon)$ , e qualquer inteiro  $p > b_2 - a_1 + M(\epsilon)$ , existir um ponto  $p$ -periódico  $x \in \Omega$  tal que

$$d(T^j x, T^j x_1) < \epsilon, \quad j \in A_1,$$

$$d(T^j x, T^j x_2) < \epsilon, \quad j \in A_2.$$

A motivação para tal definição vem do desejo de se aproximar simultaneamente, com precisão  $\epsilon$ , dois pedaços finitos de órbita  $\{T^j x_1 : j \in A_1\}$  e  $\{T^j x_2 : j \in A_2\}$  por uma órbita periódica.

Note aqui que  $A_1$  e  $A_2$  devem ser disjuntos. Gostaríamos que o salto entre  $A_1$  e  $A_2$ , que corresponde ao tempo de mudar de um pedaço de órbita ao outro, fosse o menor possível. Também gostaríamos que o comprimento da órbita periódica fosse o menor possível. Isso significa que o tempo de trocar de uma órbita pra outra também deve ser pequeno. Se  $(\Omega, T)$  possui a propriedade de especificação, tal aproximação é possível, com tempo de troca sendo maior do que  $M(\epsilon)$ . Tal número não depende das escolhas dos pontos  $x_1$  e  $x_2$ , tampouco da cardinalidade de  $A_1$  e  $A_2$ .

A seguir, daremos outra definição de propriedade de especificação, devida a Bowen:

**Definição 1.14.** *Bowen [1971]* Para todo  $\epsilon > 0$ , existe um  $M(\epsilon)$  tal que, se  $x_1, x_2, \dots, x_k$  são pontos de  $\Omega$  e  $A_i = [a_i, b_i]$ ,  $i = 1, 2, \dots, k$  são cordas com  $a_i - b_{i-1} > M(\epsilon)$ ,  $i = 1, 2, \dots, k$  e se  $p > b_k - a_1 + M(\epsilon)$ , então existe um ponto periódico  $x \in \Omega$  tal que

$$d(T^j x, T^j x_i) < \epsilon \quad \forall j \in A_i, \quad i = 1, 2, \dots, k.$$

**Observação 1.15.** *Pedimos ao leitor para observar que, se um processo satisfaz a condição de gramática completa, então ele também satisfaz a propriedade de especificação. A recíproca, em geral, não é verdade. O exemplo 1.7 ilustra bem esse fato.*

## 1.4 Resultados anteriores: o estado da arte

O objetivo da presente seção é apresentar boa parte dos resultados conhecidos sobre o Caminho mais curto de retorno. Aqui, enunciaremos e comentaremos os principais resultados da literatura sobre os tempos de retornos curtos (ou tempo de primeiro retorno). A apresentação de tais resultados será feita de forma breve e sucinta. Convidamos o leitor mais interessado a ler as referências indicadas no texto. O primeiro resultado, diz sobre a concentração de massa da variável aleatória  $T_n$  para grandes valores (com respeito a  $n$ ). O segundo estabelece um princípio de grandes desvios para  $T_n/n$ , e o relaciona com a entropia de Rényi do processo. No terceiro, se prova um resultado sobre a versão re-escalada  $n - T_n$  da variável  $T_n$ .

### 1.4.1 Convergência forte

O resultado apresentado nessa seção trata de uma convergência quase certa. Ele essencialmente afirma que a variável aleatória  $T_n$  se concentra em valores da mesma ordem de  $n$ , quando  $n$  tende a infinito. Em outras palavras: assintoticamente, não há retornos curtos.

**Teorema 1.16.** *Saussol et al. [2002]*

Se as sequências  $x_1^n$  são geradas de acordo com um processo ergódico, de entropia positiva, e que satisfaz a propriedade de especificação, num alfabeto  $\mathcal{C}$  finito, então:

$$\lim_{n \rightarrow \infty} \frac{T_n(x_1^n)}{n} = 1,$$

$\mathbb{P}$ -quase certamente em  $\Omega$ .

Tal resultado foi provado usando *Complexidade de Kolmogorov* e *Teorema de Brudno*. Logo após, tal resultado foi generalizado em [Afraimovich et al. \[2003\]](#), para processos com a propriedade fraca de especificação, e alfabeto  $\mathcal{C}$  enumerável. A demonstração foi feita usando Teorema de Shannon, Mc-Millan e Breiman.

### 1.4.2 Princípio de grandes desvios

Uma vez provada a convergência quase certa de  $T_n/n$  quando  $n$  tende a infinito, naturalmente nos fazemos algumas perguntas do tipo:

- Qual a velocidade de tal convergência?
- Poderíamos provar a existência de um Princípio de Grandes Desvios para  $T_n/n$ , ou seja: o que dizer sobre o limite

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{T_n(x_1^n)}{n} < 1 - \epsilon \right) ?$$

Na busca por responder essas questões, três trabalhos foram publicados. A seguir, apresentamos o primeiro deles. Em seguida, vamos comentar as extensões que foram feitas.

**Teorema 1.17.** *Abadi e Vaienti [2008]*

Se o processo gerador das palavras satisfaz a propriedade  $\psi$ -misturadora, então o seguinte limite existe:

$$\Delta(\epsilon) = \lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{T_n(x_1^n)}{n} < 1 - \epsilon \right).$$

Além disso, ele possui uma forma explícita, que está relacionada com a Entropia de Rényi do processo.

Em [\[Haydn e Vaienti, 2010\]](#), o resultado acima foi estendido para processos fracamente  $\psi$ -misturadores. Finalmente, em [Abadi e Cardeño \[2015\]](#), o mesmo resultado foi provado para processos com medidas geradoras que satisfazem a condição  $\psi_g$ -regular. No caso do resultado acima, cada trabalho assumiu uma hipótese diferente sobre a medida geradora do processo. Em [\[Abadi, 2008\]](#), a medida tinha que ter a propriedade  $\psi$ -misturadora. Em [Haydn e Vaienti \[2010\]](#), fracamente  $\psi$ -misturadora e em [\[Abadi e Cardeño, 2015\]](#), o processo tinha que ser  $\psi_g$ -regular.

A Figura 1.5 apresenta o gráfico da entropia de Rényi apresentado em [Abadi e Cardeño \[2015\]](#)

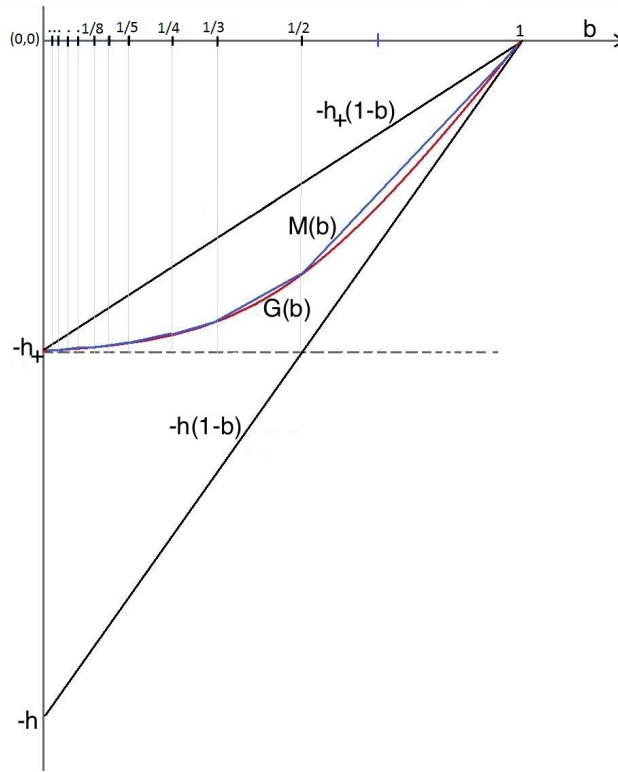


Figura 1.5: Representação gráfica da entropia de Rényi.

### 1.4.3 Convergência fraca

Como já foi apresentado anteriormente Saussol *et al.* [2002] e Afraimovich *et al.* [2003] provaram que, quando  $n$  tende a infinito,  $T_n$  tem a mesma ordem que  $n$ , em um conjunto de medida cheia.

Daí, surge uma pergunta natural: existe alguma forma explícita para relacionar  $T_n$  e  $n$ ? Seria  $T_n = n + k$ , onde  $k$  é uma constante real? Ou será que  $T_n = n + f(n)$ , para alguma função real  $f(n)$  que satisfaça  $\lim_{n \rightarrow \infty} f(n)/n = 0$ ?

Na tentativa de responder a tais perguntas, foi definida a *função de sobreposição*  $S_n$ , da seguinte maneira:

$$S_n(x_1^n) = n - T_n(x_1^n) .$$

Note que, se  $T_n \leq n$ , então  $S_n$  define o máximo tamanho de uma sobreposição que a palavra  $x_1^n$  possui com uma cópia(transladada) de si mesma.

Veja na figura seguinte uma interpretação ilustrativa:

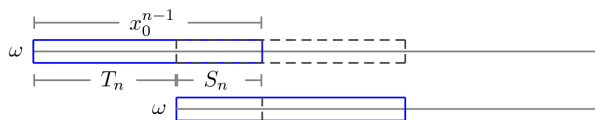
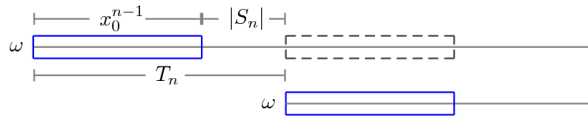


Figura 1.6: Relação entre  $T_n$  e  $S_n$ .

Porém, sempre que  $T_n > n$ ,  $S_n$  assumirá valores negativos, e não poderemos fazer a mesma interpretação do objeto  $S_n$ .



A figura a seguir ilustra o que acontece nesse caso.



**Figura 1.7:** Relação entre  $T_n$  e  $S_n$ : caso  $T_n > n \implies S_n < 0$

Observe que, quando o processo gerador das palavras satisfaz a propriedade de gramática completa (ver definição 1.10), temos  $S_n \geq 0$ .

A seguir, apresentamos o resultado sobre a convergência em distribuição da função de sobreposição  $S_n$ :

**Teorema 1.18.** *Abadi e Lambert [2013]*

Se as palavras são geradas de acordo com um processo IID no alfabeto  $\mathcal{C}$  finito ou infinito-enumerável, então existe uma variável aleatória não-degenerada  $S_n = n - T_n$  que converge em distribuição para uma variável aleatória não-degenerada  $S$ . Tal variável aleatória possui também esperança e variância finitas. Além disso,  $n - T_n$  não converge em probabilidade para  $S$  quando  $n$  tende a infinito.

Logo depois, o mesmo resultado foi generalizado para processos  $\beta$ -mixing em Rada [2014]. A seguir, apresentaremos o principal resultado desse trabalho, com alguns detalhes. A razão para uma apresentação tão detalhada é o fato de que precisaremos usá-los na seção 2.5, do próximo capítulo.

Antes, daremos algumas definições que nos permitirão entender melhor os resultados.

Seja  $\{X_n\}_{n \in \mathbb{Z}}$  um processo estocástico estacionário assumindo valores no alfabeto  $\mathcal{C}$ . O valor de  $X_n$  é interpretado como o estado do processo no tempo  $n$ . Para  $-\infty \leq J \leq L \leq \infty$ , definimos a  $\sigma$ -álgebra

$$\mathcal{F}_J^L := \sigma(X_k, J \leq k \leq L \ (k \in \mathbb{Z})).$$

Ou seja,  $\mathcal{F}_J^L$  é a  $\sigma$ -álgebra gerada pelas variáveis aleatórias  $X_k$  para  $k$  entre  $J$  e  $L$ .

**Definição 1.19.** Para quaisquer duas  $\sigma$ -álgebras  $\mathcal{F}_J^L$  e  $\mathcal{F}_M^N$ , definimos a medida de dependência  $\beta$ -mixing como

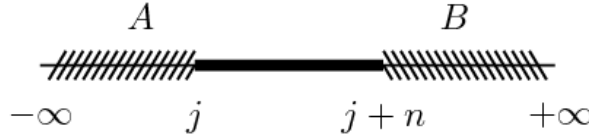
$$\beta(\mathcal{F}_J^L, \mathcal{F}_M^N) := \sup \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K |\mu(A_i \cap B_k) - \mu(A_i)\mu(B_k)|,$$

onde o supremo é tomado sobre todos os pares de partições finitas  $\{A_1, \dots, A_I\}$  e  $\{B_1, \dots, B_K\}$  de  $\mathcal{C}^{\mathbb{Z}}$  tais que  $A_i \in \mathcal{F}_J^L$  para cada  $i$  e  $B_k \in \mathcal{F}_M^N$  para cada  $k$ .

**Definição 1.20.** Para  $n \geq 1$ , definimos o coeficiente  $\beta$ -mixing por,

$$\beta(n) := \sup_{j \in \mathbb{Z}} \beta(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty}).$$

Na seguinte figura se mostra o conjunto  $A \in \mathcal{F}_{-\infty}^j$  e o conjunto  $B \in \mathcal{F}_{j+n}^{\infty}$ . O coeficiente  $\beta$  destes dois conjuntos está baseado na distância que existe entre eles.



**Figura 1.8:** Definição do processo  $\beta$ -mixing.

**Definição 1.21.** Diremos que o processo  $\{X_n\}_{n \in \mathbb{Z}}$  é  $\beta$ -mixing se

$$\lim_{n \rightarrow \infty} \beta(n) = 0.$$

Depois de todas estas definições, podemos finalmente enunciar o resultado:

**Teorema 1.22.** *Rada [2014]* Seja  $\mu$  uma medida de probabilidade definida em  $\mathcal{C}^{\mathbb{N}}$  para um processo  $\beta$ -mixing com gramática completa. Considere  $k \in \mathbb{N}$  tal que  $n \geq 4k$ . Então

$$|\mu(n - T_n \geq k) - \mathbb{E}_{\mu, \mu}(k) - a_k^{(1)}| \leq K_{\rho_\mu} \sum_{i=\lceil n/4 \rceil}^{\infty} \inf_{1 \leq m \leq i} (\rho_\mu^m + \tilde{\beta}_m(i)) .$$

Onde:

- $K_{\rho_\mu}$  é uma constante que só depende de  $\rho_\mu$ .
- $\tilde{\beta}_m(i) = 2\beta_m(i) = 2\beta(\lfloor i/m \rfloor - 1)$  .
- $\rho_\mu = \max_{x \in \mathcal{C}} \mu(x)$  .
- $\mathbb{E}_{\mu, \mu}(k) = \sum_{\omega \in \mathcal{C}^k} \mu(\omega)\mu(\omega) = \sum_{\omega \in \mathcal{C}^k} \mu^2(\omega)$  .
- $a_k^{(1)} = \sum_{i=k+1}^{\infty} \sum_{\omega \notin \cup_{j=k}^{i-1} R_i^{(1)}(j)} \mu^2(\omega)$  .

Se, além da hipótese acima definida, o processo  $\beta$ -mixing é tal que  $i\beta(i)$  é somável, então

$$\lim_{n \rightarrow \infty} \mu(n - T_n \geq k) = \mathbb{E}_{\mu, \mu}(k) + a_k^{(1)} .$$

**Observação 1.23.** No último somatório de cada item acima aparece a notação  $R_i^{(1)}(j)$ , que indica o conjunto de palavras de tamanho  $i$  que possuem alguma sobreposição de tamanho  $j$  consigo mesma, e que será definido no capítulo 2 (Definição 2.22).

No mesmo trabalho, a autora obtém limitantes para o valor  $a_k^{(1)}$ . Tal resultado é apresentado a seguir.

**Proposição 1.24.** *Rada [2014]*

Sob as hipóteses do Teorema 1.22, valem os seguintes limitantes para  $a_k^{(1)}$ :

$$(a) \ a_k^{(1)} \leq \sum_{i=k+1}^{\infty} \mathbb{E}_{\mu, \mu}(i) - \sum_{i=k+1}^{\infty} \sum_{\omega \in R_i^{(1)}(k)} \mu^2(\omega) .$$

$$(b) a_k^{(1)} \geq \sum_{i=k+1}^{\infty} \mathbb{E}_{\mu, \mu}(i) - \sum_{i=k+1}^{\infty} \sum_{j=k}^{i-1} \sum_{\omega \in R_i^{(1)}(j)} \mu^2(\omega) .$$



# Capítulo 2

## O menor caminho entre duas sequências

Dados os conceitos e resultados para o primeiro retorno, passamos a uma abordagem global do problema. Aqui, ao invés de considerarmos o tempo que uma sequência leva para voltar em si mesma, consideraremos duas sequências, e o tempo que se leva para percorrer um caminho que leva de uma à outra. Este capítulo compreende a Tese propriamente dita. É nele que iremos enunciar e provar os resultados obtidos. Organizaremos o capítulo da seguinte maneira: Na primeira seção, definiremos  $T_n^{(2)}$ , que é o nosso objeto de estudo. Abordaremos a definição do ponto de vista combinatório num primeiro momento. Depois, definiremos sob um ponto de vista dinâmico. Além disso, apresentamos também algumas ferramentas básicas que serão usadas durante o capítulo. Na segunda parte, apresentamos uma quantidade que aparece de forma recorrente no nosso texto, explicando algumas propriedades. Tal quantidade recebe o nome de *divergência relativa entre duas medidas*. A seguir, provamos uma convergência quase certa. Na seção seguinte, apresentamos um princípio de grandes desvios para  $T_n^{(2)}$ . Depois, provamos uma convergência em distribuição para uma forma re-escalada de  $T_n^{(2)}$ . E, finalmente, apresentaremos alguns exemplos.

### 2.1 Definições preliminares

O objetivo da presente seção é apresentar as definições básicas, e também alguma notação utilizada no capítulo. Além disso, exibimos alguns exemplos, para esclarecê-las.

#### 2.1.1 Linguagem combinatória

Seguindo a mesma estrutura do capítulo anterior, abordaremos aqui a notação combinatória do problema.

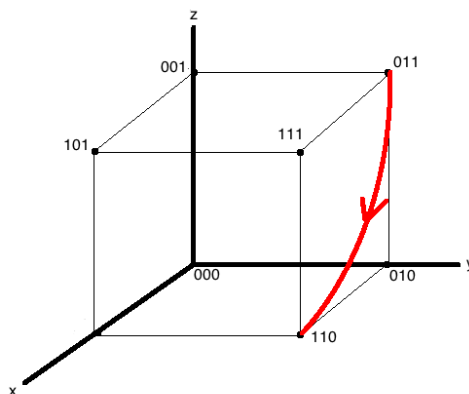
**Definição 2.1.** *Seja  $(x_1^n, y_1^n) \in \mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ . Definimos a variável aleatória  $T_n^{(2)} : \mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}} \rightarrow \{1, 2, \dots, n\}$  da seguinte maneira:*

$$T_n^{(2)}(x_1^n, y_1^n) = \inf\{k \geq 1 : x_1^{n-k} = y_{k+1}^n\},$$

e  $n$ , caso contrário.

A seguir, daremos alguns exemplos de  $T_n^{(2)}$ , ilustrando com o hipercubo.

**Exemplo 2.2.** *Sejam duas sequências  $x_1^3 = 110$  e  $y_1^3 = 011$ . Para sairmos de 011 e chegarmos em 110, temos que percorrer o caminho: 011 – 110. Logo,  $T_3^{(2)}(110, 011) = 1$ .*

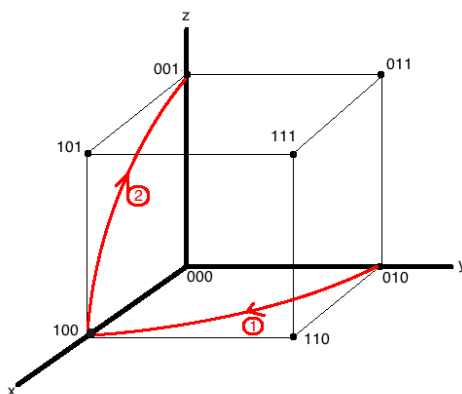


**Figura 2.1:** *Saindo de 011 e indo até 110. O menor caminho possui um passo.*

A figura 2.2 ilustra o caminho percorrido no hipercubo.

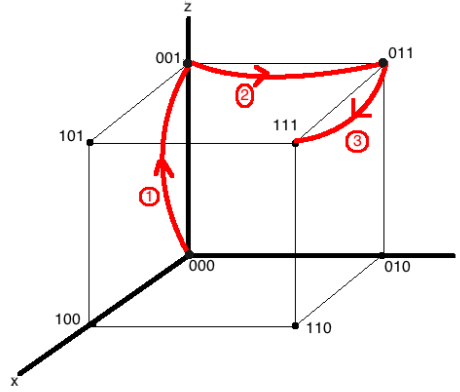
**Exemplo 2.3.** *Sejam duas sequências  $x_1^n = 001$  e  $y_1^n = 010$ . Para sairmos de 010 e chegarmos em 001, temos que percorrer o caminho: 010 – 100 – 001. Logo,  $T_3^{(2)}(001, 010) = 2$ .*

A figura 2.2 ilustra o caminho percorrido no hipercubo.



**Figura 2.2:** *Saindo de 010 e indo até 001. O menor caminho possui dois passos.*

**Exemplo 2.4.** *Se nesse caso, queremos sair de  $y_1^n = 000$  e chegar em  $x_1^n = 111$ , então o menor caminho é: 000 – 001 – 011 – 111, e possui 3 passos. Portanto,  $T_3^{(2)}(111, 000) = 3$ .*



**Figura 2.3:** *Saindo de 000 e indo até 111. O menor caminho possui três passos.*

### 2.1.2 Linguagem dinâmica

Seguindo a sequência proposta no início do texto, vamos agora dar uma definição dinâmica para  $T_n^{(2)}$ .

Primeiramente, vamos recordar a definição de *Tempos de Espera* ou *Waiting Times*, que aparece pela primeira vez na literatura em Wyner e Ziv [1989], mas pode ser encontrada também em Shields [1993], Kontoyiannis [1998], Wyner [1999] e Chazottes [2000].

**Definição 2.5.** *Para o par de realizações  $(x, y) \in \mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ , definimos a função Tempo de espera  $W_n(x, y)$  por:*

$$W_n(x, y) = \min\{k : y_k^{k+n-1} = x_1^n\} .$$

Agora, vamos definir  $T_n^{(2)}$  em função de  $W_n$ :

**Definição 2.6.** *Para o par de palavras  $(x_1^n, y_1^n) \in \mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ , definimos a função caminho mais curto entre  $x_1^n$  e  $y_1^n$  como:*

$$T_n^{(2)}(x_1^n, y_1^n) = \inf_{z: z_1^n = y_1^n} W_n(x, z) .$$

A seguir, vamos dar uma definição para todos os conjuntos de pares  $(x_1^n, y_1^n)$  que possuem alguma sobreposição do começo de uma das palavras com o fim da outra.

**Definição 2.7.** *Dizemos que  $R_n^{(2)}(k)$  é o conjunto de pares  $(x_1^n, y_1^n)$  que verificam a seguinte condição: os primeiros  $k$  símbolos da sequência  $x_1^n$  coincidem com os  $k$  últimos símbolos da sequência  $y_1^n$ . De maneira formal:*

*Para  $k \in \{1, 2, \dots, n-1\}$ , defina*

$$R_n^{(2)}(k) = \{(x, y) \in \mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}} : y_{n-k+1}^n = x_1^k\} .$$

A seção seguinte trata de uma quantidade definida por  $\mathbb{P}(R_n^{(2)}(k))$ , onde  $\mathbb{P}(A \times B) = (\mu \times \nu)(A \times B)$ , para todo  $A, B \in \mathcal{F}$  é a medida-produto de  $\mu$  e  $\nu$ .

## 2.2 A divergência relativa entre duas medidas

A presente seção se dedicará a uma quantidade que aparece recorrentemente no nosso trabalho. Tal quantidade é vista nas demonstrações que seguem. A ela daremos o nome de *divergência relativa entre duas medidas*, e é, em sua essência, o produto interno entre dois vetores de medidas. Primeiramente, apresentamos sua definição. A seguir, enunciaremos e provamos algumas propriedades.

### 2.2.1 A divergência relativa

Quando relacionamos duas medidas no espaço produto, algumas quantidades conhecidas podem aparecer, como *entropia relativa* ou *divergência de Kullback-Leibler* (Cover e Thomas [1991]).

Por vezes, nos deparamos com algumas quantidades que não são conhecidas na literatura. A próxima definição trata de uma dessas quantidades.

**Definição 2.8.** *Sejam  $\mu$  e  $\nu$  duas medidas definidas sobre  $(\Omega, \mathcal{F})$ , onde  $\Omega = \mathcal{C}^{\mathbb{N}}$ , e  $\mathcal{F}$  é a sigma-álgebra gerada pelos cilindros de tamanho  $n$  em  $\mathcal{C}^{\mathbb{N}} \subset \mathcal{C}^{\mathbb{N}}$ .*

*Dizemos que a divergência relativa de ordem  $k$  entre  $\mu$  e  $\nu$  é dada por:*

$$\mathbb{E}_{\mu, \nu}(k) = \sum_{\omega \in \mathcal{C}^k} \mu(\omega) \nu(\omega) .$$

Note que tal medida nada mais é do que a esperança da variável aleatória  $\nu(Z)$ , quando  $Z$  é a variável aleatória  $k$ -dimensional que assume valores em  $\mathcal{C}^k$  com a seguinte lei:

$$P(Z = \omega) = \mu(\omega) .$$

Portanto, estamos tratando da esperança de uma medida  $\nu$  com respeito à medida  $\mu$ . Formalmente:

$$\mathbb{E}_{\mu, \nu}(k) := \mathbb{E}_{\mu}(\nu(Z)) .$$

### 2.2.2 Propriedades

Aqui, apresentamos propriedades extremamente úteis de  $\mathbb{E}_{\mu, \nu}(k)$ . Usaremos tais resultados nas demonstrações e nos exemplos que aparecerão em nosso trabalho.

As provas para as propriedades a seguir são obtidas imediatamente da definição de  $\mathbb{E}_{\mu, \nu}(k)$ .

**Propriedades 2.9.** *A divergência relativa  $\mathbb{E}_{\mu, \nu}(k)$  satisfaz:*

- (a)  $\mathbb{E}_{\mu, \nu}(k)$  é simétrica. Ou seja:  $\mathbb{E}_{\mu, \nu}(k) = \mathbb{E}_{\nu, \mu}(k)$  .
- (b)  $\mathbb{E}_{\mu, \nu}(k) \neq 0 \Leftrightarrow \text{supp}(\mu) \cap \text{supp}(\nu) \neq \emptyset$ , onde

$$\text{supp}(\mu) := \{\omega \in \Omega : \mu(\omega) > 0\}$$

*é o suporte da medida  $\mu$  (respectivamente  $\nu$ ).*

Agora, vamos apresentar um resultado que será de grande utilidade para provar um princípio de grandes desvios, na seção 2.8:



**Proposição 2.10.** *Se  $\mu$  e  $\nu$  são independentes, então  $\mathbb{E}_{\mu,\nu}(k)$  é uma função não-crescente em  $k$ .*

*Demonstração.* Para todo  $\omega \in \mathcal{C}^k$ , seja  $\omega a$  a palavra de tamanho  $k + 1$  construída pela concatenação da letra  $a$  ao final da palavra  $\omega$ . Defina também os vetores de medidas:

$$\mu_\omega = (\mu(\omega a_1), \mu(\omega a_2), \dots, \mu(\omega a_{\#\mathcal{C}})) ,$$

da mesma maneira:

$$\nu_\omega = (\nu(\omega a_1), \nu(\omega a_2), \dots, \nu(\omega a_{\#\mathcal{C}})) ,$$

onde  $\#\mathcal{C}$  é a cardinalidade do alfabeto  $\mathcal{C}$ .

Pela desigualdade de Cauchy-Schwartz e pela desigualdade  $\|\cdot\|_p \leq \|\cdot\|_q$ , se  $p > q$ , temos que

$$\begin{aligned} \sum_{a \in \mathcal{C}} \mu(\omega a) \nu(\omega a) &= \langle \mu_\omega, \nu_\omega \rangle \\ &= \|\mu_\omega\|_2 \|\nu_\omega\|_2 \cos \theta(\mu_\omega, \nu_\omega) \\ &\leq \|\mu_\omega\|_2 \|\nu_\omega\|_2 \\ &\leq \|\mu_\omega\|_1 \|\nu_\omega\|_1 \\ &= \sum_{a \in \mathcal{C}} \mu(\omega a) \sum_{a \in \mathcal{C}} \nu(\omega a) \end{aligned}$$

Logo, para todo  $k$ , temos

$$\begin{aligned} \mathbb{E}_{\mu,\nu}(k) &= \sum_{\omega \in \mathcal{C}^k} \mu(\omega) \nu(\omega) \\ &= \sum_{\omega \in \mathcal{C}^k} \left[ \sum_{a \in \mathcal{C}} \mu(\omega a) \sum_{a \in \mathcal{C}} \nu(\omega a) \right] \\ &\geq \sum_{\omega \in \mathcal{C}^k} \left[ \sum_{a \in \mathcal{C}} \mu(\omega a) \nu(\omega a) \right] \\ &= \sum_{\omega \in \mathcal{C}^{k+1}} \mu(\omega) \nu(\omega) \end{aligned}$$

Como o último termo da expressão acima é exatamente  $\mathbb{E}_{\mu,\nu}(k + 1)$ , finalizamos a prova.  $\square$

## 2.3 O crescimento linear

O principal resultado da presente seção mostra que, para grandes valores de  $n$ ,  $T_n^{(2)}$  se comporta como uma função linear de  $n$ . O resultado é condizente com o que foi feito em [Saussol \*et al.\* \[2002\]](#), e depois em [Afraimovich \*et al.\* \[2003\]](#). A diferença é que, ao invés de considerar repetições, estaremos buscando coincidências. Em termos matemáticos: enquanto a abordagem anterior enxerga o problema de um ponto de vista mais local, nós consideramos uma forma global de tratá-lo.

**Teorema 2.11.** *Sejam  $\mu$  e  $\nu$  duas medidas estacionárias definidas em  $\mathcal{C}^{\mathbb{N}}$ , independentes uma da outra e que admitam a propriedade de gramática completa. Se  $\mu$  for ergódica, e com entropia positiva, então*

$$\lim_{n \rightarrow \infty} \frac{T_n^{(2)}(x, y)}{n} = 1, \quad (2.1)$$

para  $\mu \times \nu$ -quase todo ponto  $(x, y) \in \mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ .

A demonstração do teorema 2.11 será dividida em duas partes. Na proposição 2.12, mostramos que  $\limsup_{n \rightarrow \infty} T_n^{(2)}/n \leq 1$ . No teorema 2.13, provamos que  $\liminf_{n \rightarrow \infty} T_n^{(2)}/n \geq 1$ . Juntando os dois resultados, concluímos a prova do teorema.

A seguir, apresentaremos o primeiro resultado. Sua demonstração é imediata.

**Proposição 2.12.** *Se  $\mu$  e  $\nu$  satisfazem a propriedade de gramática completa, então:*

$$\limsup_{n \rightarrow \infty} \frac{T_n^{(2)}}{n} \leq 1.$$

*Demonstração.* Se a condição de gramática completa é satisfeita, então  $T_n^{(2)} \in \{1, 2, \dots, n\}$ , o que implica em  $T_n^{(2)}/n \leq 1$ , e portanto,  $\limsup_{n \rightarrow \infty} T_n^{(2)}/n \leq 1$ .  $\square$

**Teorema 2.13.** *Sejam  $\mu$  e  $\nu$  duas medidas estacionárias definidas em  $\mathcal{C}^{\mathbb{N}}$ , independentes uma da outra. Se  $\mu$  for ergódica, e com entropia positiva, então*

$$\liminf_{n \rightarrow \infty} \frac{T_n^{(2)}(x, y)}{n} \geq 1. \quad (2.2)$$

*Demonstração.* A estratégia da prova será provar que, para todo  $\epsilon > 0$ , o evento

$$\{T_n^{(2)} \leq (1 - \epsilon)n\}$$

ocorre apenas um número finito de vezes. Para isso, usaremos o Lema de Borel-Cantelli.

Se  $\mu$  é ergódica com entropia positiva  $h > 0$ , o Teorema de Shannon-McMillan-Breiman diz que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu(x_1^n) = -h,$$

$\mu$ -quase certamente em  $\Omega$ .

Portanto, dado  $\epsilon > 0$ , existe uma constante inteira positiva  $k_0 = k_0(\epsilon, x_1^k)$ , tal que para quase todo conjunto  $x_1^k$

$$\left| \frac{1}{k} \log \mu(x_1^k) + h \right| < \epsilon, \quad (2.3)$$

sempre que  $k > k_0$ .

Note que aqui  $k_0$  depende do ponto  $x \in x_1^k$  escolhido. Para eliminar essa dependência (mesmo que em um subconjunto de  $\Omega \times \Omega$ ), vamos invocar o Teorema de Egorov.

Pelo Teorema de Egorov, para todo  $\epsilon > 0$ , sabemos que existe um "subconjunto de medida grande" de  $\Omega \times \Omega$ , tal que a convergência seja uniforme dentro dele.

Formalmente, existe um conjunto  $\Omega_\epsilon \times \Omega$ , tal que

$$(a) \quad (\mu \times \nu)(\Omega_\epsilon \times \Omega) \geq 1 - \epsilon.$$

(b) Existe um  $k_0(\epsilon)$ , tal que a desigualdade (2.3) vale para todo  $k > k_0(\epsilon)$ , e para quase todo ponto  $x_1^k \in \Omega_\epsilon$ .

Portanto, para quase todo ponto  $x \in \Omega_\epsilon$ , e para todo  $k > k_0$ , temos que:

$$e^{-k(h+\epsilon)} < \mu(x_1^k) < e^{-k(h-\epsilon)}. \quad (2.4)$$

Para continuarmos com a prova, note que

$$\{T_n^{(2)}(x_1^n, y_1^n) \leq k\} = \{(x_1^n, y_1^n) \in \bigcup_{i=n-k}^{n-1} R_n^{(2)}(i)\}.$$

Para algum  $0 < \delta < 1$ , temos que

$$\begin{aligned} \mathbb{P}(\{T_n^{(2)}(x, y) \leq \delta n\} \cap \Omega_\epsilon) &= \mathbb{P}\left(\bigcup_{j=n-\delta n}^{n-1} R_n^{(2)}(j) \cap \Omega_\epsilon\right) \\ &\leq \sum_{j=n-\delta n}^{n-1} \mathbb{P}(R_n^{(2)}(j) \cap \Omega_\epsilon) \\ &\leq \sum_{j=n-\delta n}^{n-1} \mathbb{E}_{\mu, \nu}(j) \\ &= \sum_{j=n-\delta n}^{n-1} \sum_{x_1^j} \mu(x_1^j) \nu(x_1^j) \\ &\leq \sum_{j=n-\delta n}^{n-1} e^{-j(h-\epsilon)} \end{aligned}$$

Onde a última desigualdade acima foi obtida usando (2.4). Portanto:

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(T_n^{(2)}(x, y) \leq \delta n) &\leq \sum_{n=1}^{\infty} \sum_{j=n-\delta n}^{n-1} e^{-j(h-\epsilon)} \\ &= \sum_{n=1}^{\infty} \frac{e^{-(n-\delta n)(h-\epsilon)} - e^{-n(h-\epsilon)}}{1 - e^{-(h-\epsilon)}} \\ &= \left(\frac{1}{1 - e^{-h+\epsilon}}\right)^2 (e^{-(h-\epsilon)(1-\delta)} - e^{-(h-\epsilon)}) \end{aligned}$$

O termo acima é um número finito. Pelo Lema de Borel Cantelli, temos que o evento  $\{T_n^{(2)}(x, y) \leq \delta n\}$  ocorre apenas um número finito de vezes. Tomando  $\delta = 1 - \epsilon$ , temos que

$$\liminf_{n \rightarrow \infty} \frac{T_n^{(2)}}{n} \geq 1 - \epsilon, \quad \mu \times \nu - \text{q.t.p. em } \Omega_\epsilon \times \Omega. \quad (2.5)$$

Como esse resultado vale para todo  $\epsilon$ , e o último valor em (2.5) converge para 1 quando  $\epsilon$  vai para zero, concluímos a prova. □

Agora, basta juntar a proposição 2.12 ao teorema 2.13, para concluir o principal resultado

dessa parte (Teorema 2.11). Assim, encerramos a seção.

## 2.4 Grandes Desvios

A seção passada prova que, para  $\mu \times \nu$ - quase toda escolha  $(x_1^n, y_1^n) \in \mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$ , a quantidade  $T_n^{(2)}/n$  converge para 1 sempre que  $n$  diverge. Como convergência quase certa implica na convergência em medida, temos que, sob as hipóteses do teorema 2.11, e para todo  $\epsilon > 0$ , vale

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \left| \frac{T_n^{(2)}(x, y)}{n} - 1 \right| > \epsilon \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{T_n^{(2)}(x, y)}{n} < 1 - \epsilon \right) = 0, \quad (2.6)$$

onde a primeira igualdade acima foi obtida pela proposição 2.12.

Sendo assim, duas perguntas que podemos nos fazer são:

- (a) Qual a velocidade da convergência apresentada em 2.6?
- (b) Além disso, como se comportaria essa convergência para cada valor de  $\epsilon$ ?

Para responder a essas perguntas, vamos calcular a função taxa dos grandes desvios para  $\frac{T_n^{(2)}}{n}$ . Ou seja, queremos descobrir qual o valor de

$$-\Delta(\epsilon) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{T_n^{(2)}(x, y)}{n} < 1 - \epsilon \right),$$

A seguir, apresentamos um lema muito útil na prova do nosso resultado:

**Lema 2.14.** *Se  $\{a_n\}_{n \in \mathbb{M}_n}$  uma sequência de termos não-negativos, indexada em um conjunto  $\mathbb{M}_n = \{m_n, m_n + 1, \dots, n - 1\}$ , para alguma função  $m_n < n$ , então:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left[ \sum_{j \in \mathbb{M}_n} a_j \right] = \lim_{n \rightarrow \infty} \frac{1}{n} \log \left[ \max_{j \in \mathbb{M}_n} a_j \right]$$

*Demonstração.* Note que

$$\max_{j \in \mathbb{M}_n} a_n \leq \sum_{j \in \mathbb{M}_n} a_n \leq (n - m_n) \max_{j \in \mathbb{M}_n} a_n.$$

Aplicando o logaritmo e dividindo por  $n$ , temos:

$$\frac{1}{n} \log \left[ \max_{j \in \mathbb{M}_n} a_n \right] \leq \frac{1}{n} \log \sum_{j \in \mathbb{M}_n} a_n \leq \frac{\log(n - m_n)}{n} + \frac{1}{n} \log \left[ \max_{j \in \mathbb{M}_n} a_n \right].$$

Como  $\lim_{n \rightarrow \infty} (\log(n - m_n))/n = 0$ , temos que os limites inferior e superior da expressão acima convergem para o mesmo valor (se existir), quando  $n$  tende ao infinito. □

O próximo resultado fornece o cálculo da função taxa dos grandes desvios.

**Proposição 2.15.**

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{T_n^{(2)}(x, y)}{n} < 1 - \epsilon \right) = \lim_{m \rightarrow \infty} \frac{\epsilon}{m} \log \mathbb{E}_{\mu, \nu}(m).$$

*Demonstração.*

$$\mathbb{P}\left(\frac{T_n^{(2)}}{n} < 1 - \epsilon\right) = \mathbb{P}\left(\bigcup_{j=n-[n-n\epsilon]}^{n-1} R_n^{(2)}(j)\right)$$

Tome  $n^* := n - [n - n\epsilon]$ , e note que:

$$\max_{j \in \{n^*, \dots, n-1\}} \mathbb{P}(R_n^{(2)}(j)) \leq \mathbb{P}\left(\bigcup_{j=n-[n-n\epsilon]}^{n-1} R_n^{(2)}(j)\right) \leq \sum_{j=n^*}^{n-1} \mathbb{P}(R_n^{(2)}(j))$$

Pelo Lema 2.14, temos que os dois limitantes da equação acima convergem para o mesmo limite (se esse limite existir). Portanto, estamos interessados em:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \max_{j \in \{n^*, \dots, n-1\}} \mathbb{P}(R_n^{(2)}(j)) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \left[ \sum_{\omega \in \mathcal{C}^{n^*}} \mu(\omega) \nu(\omega) \right] \quad (2.7)$$

Pela proposição 2.10, temos  $\mathbb{E}_{\mu, \nu}(n)$  atinge seu ponto de máximo em  $n = n^*$ , o que nos levou à (2.7).

Fazendo a mudança  $m = [n\epsilon]$ , temos que  $m = n^*$ , e o último termo da expressão anterior fica:

$$\lim_{m \rightarrow \infty} \frac{\epsilon}{m} \log \left[ \sum_{\omega \in \mathcal{C}^m} \mu(\omega) \nu(\omega) \right] = \lim_{m \rightarrow \infty} \frac{\epsilon}{m} \log \mathbb{E}_{\mu, \nu}(m). \quad (2.8)$$

Desta forma, finalizamos a demonstração da proposição.  $\square$

O próximo resultado garante a existência do limite (2.8), no caso em que as medidas  $\nu$  e  $\mu$  satisfazem à *condição de mistura de Pittel*, que será definida a seguir:

**Definição 2.16.** *Pittel [1985]*

*Dizemos que uma medida  $\mu$  satisfaz à condição de mistura de Pittel, se existem constantes positivas  $c_{1\mu}$  e  $c_{2\mu}$  tais que, para todo  $1 \leq a \leq a + b_\mu \leq b$ , valem as desigualdades:*

$$c_{1\mu} \mu(A) \mu(B) \leq \mu(A) \mu(B) \leq c_{2\mu} \mu(A) \mu(B),$$

para todo  $A \in \mathcal{F}_1^a$  e  $B \in \mathcal{F}_{a+b_\mu}^b$ , onde  $\mathcal{F}_m^n$  é a  $\sigma$ -álgebra gerada pelos cilindros  $x_m^n$ .

O resultado a seguir constitui a ferramenta fundamental para demonstrar a existência da função de Grandes Desvios de  $T_n^{(2)}/n$ .

Ele é uma versão de um lema apresentado em Pittel [1985] e também em Haydn e Vaienti [2010], que por sua vez é uma versão do famoso *Lema de Fekete para sequências subaditivas*.

**Lema 2.17.** *Seja  $\{f(n)\}_{n \in \mathbb{N}}$  uma sucessão que satisfaz:*

$$f(j + k + b_0) \leq f(j) + f(k) + c_{b_0}, \quad (2.9)$$

onde  $j, k$  e  $b_0$  são inteiros não-negativos, e  $c_{b_0}$  é uma constante que depende de  $b_0$ . Então  $f(n)/n$  converge, quando  $n$  tende a infinito. Mais ainda:

$$\lim_{n \rightarrow \infty} \frac{f(n)}{n} = \inf_{n \in \mathbb{N}} \frac{f(n)}{n}.$$

*Demonstração.* Seja  $n_t$  uma sequência crescente e não limitada de números inteiros tal que:

$$\liminf_{n \rightarrow \infty} \frac{f(n)}{n} = \lim_{t \rightarrow \infty} \frac{f(n_t)}{n_t} . \quad (2.10)$$

Agora, para  $n > b_0$ , escreva:  $n - b_0 = qn_t + r$ , para algum  $0 \leq q$  e  $0 \leq r < n_t$ . Pela propriedade 2.9, temos que:

$$f(n) \leq f(qn_t) + f(r) + c_{b_0} . \quad (2.11)$$

Observando que  $qn_t = ((q-1)n_t) + (n_t - b_0) + b_0$ , e aplicando recursivamente 2.11, obtemos

$$f(n) \leq f(n_t) + (q-1)f(n_t - b_0) + (q-1)c_{b_0} + f(r) .$$

Dividindo os dois lados por  $n$ , chegamos a:

$$\begin{aligned} \frac{f(n)}{n} &\leq \frac{f(n_t) + (q-1)f(n_t - b_0) + (q-1)c_{b_0} + f(r)}{qn_t + r + b_0} \\ &\leq \frac{f(n_t)}{qn_t} + \frac{(q-1)f(n_t - b_0)}{qn_t} + \frac{(q-1)c_{b_0}}{qn_t} + \frac{f(r)}{qn_t} \end{aligned}$$

Como  $q \rightarrow \infty$  implica  $n \rightarrow \infty$ , temos

$$\begin{aligned} \limsup_{q \rightarrow \infty} \frac{f(n)}{n} &= \limsup_{n \rightarrow \infty} \frac{f(n)}{n} \\ &\leq \limsup_{q \rightarrow \infty} \left[ \frac{f(n_t)}{qn_t} + \frac{(q-1)f(n_t - b_0)}{qn_t} + \frac{(q-1)c_{b_0}}{qn_t} + \frac{f(r)}{qn_t} \right] \\ &= \frac{f(n_t - b_0)}{n_t} + \frac{c_{b_0}}{n_t} \end{aligned}$$

Como a desigualdade acima vale para todo  $t$ , basta fazer  $t \rightarrow \infty$  para obter finalmente

$$\limsup_{n \rightarrow \infty} \frac{f(n)}{n} \leq \limsup_{t \rightarrow \infty} \frac{f(n_t - b_0)}{n_t} = \lim_{t \rightarrow \infty} \frac{f(n_t)}{n_t} = \liminf_{n \rightarrow \infty} \frac{f(n)}{n} ,$$

e concluir assim a demonstração do lema.  $\square$

O resultado a seguir nos garante que existe a função taxa dos grandes desvios. Tal função é uma reta em  $\epsilon$ , com coeficiente linear nulo e coeficiente angular dado pelo limite :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{\omega \in \mathcal{C}^n} \mu(\omega) \nu(\omega) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_{\mu, \nu}(n) .$$

**Proposição 2.18.** *Se as medidas  $\mu$  e  $\nu$  satisfazem à condição de mistura de Pittel (2.16), então o limite (2.8) existe.*

*Demonstração.* Defina:

$$f(n) = \log \sum_{\omega \in \mathcal{C}^n} \mu(\omega) \nu(\omega) ,$$

e tome  $n = j + k + b_\mu$ . Se  $\omega_n$  é uma sequência genérica com  $n$  símbolos, note que:

$$\mu(\omega_n) = \mu(\omega_j \omega_k \omega_{b_\mu}) = \mu(\omega_j \tilde{\omega}_{b_\mu} \tilde{\omega}_k) ,$$

onde  $\tilde{\omega}_{b_\mu} \tilde{\omega}_k = \omega_j \omega_k \omega_{b_\mu}$ . Usando o teorema da probabilidade total e a condição de Pittel, chegamos a

$$\mu(\omega_n) \leq c_{2_\mu} \mu(\omega_j) \mu(\omega_k) .$$

Repetindo o argumento para a medida  $\nu$ , obtemos:

$$\log \sum_{\omega \in \mathcal{C}^n} \mu(\omega) \nu(\omega) \leq \log \left[ c_{2_\mu} c_{2_\nu} \sum_{\omega \in \mathcal{C}^j} \mu(\omega) \nu(\omega) \sum_{\omega \in \mathcal{C}^k} \mu(\omega) \nu(\omega) \right] \quad (2.12)$$

$$= f(j) + f(k) + c_{b_0} , \quad (2.13)$$

onde  $c_{b_0} = \log c_{2_\mu} c_{2_\nu}$ .

O argumento acima nos diz então que  $\{f(n)\}_{n \in \mathbb{N}}$  satisfaz à condição (2.9), e portanto  $f(n)/n$  é uma sucessão convergente. Assim, provamos a existência do limite (2.8).  $\square$

A seguir, apresentamos alguns exemplos para elucidar os resultados supracitados.

**Exemplo 2.19.** *Sejam  $\mu$  e  $\nu$  duas medidas com suportes disjuntos. Ou seja:*

$$\text{supp}(\mu) \cap \text{supp}(\nu) = \emptyset .$$

*Pelo item (b) das propriedades 2.9, temos que para todo  $n > 1$  vale*

$$\mathbb{E}_{\mu, \nu}(n) = 0 .$$

*Portanto, a taxa de grandes desvios nesse caso é dada por*

$$-\Delta(\epsilon) = \epsilon \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_{\mu, \nu}(n) = -\infty .$$

**Exemplo 2.20.** *Nesse caso, tome as duas medidas  $\mu$  e  $\nu$  se concentrando em uma única realização do processo  $x_1^\infty$ . Ou seja:*

$$\mu(x_1^\infty) = \nu(x_1^\infty) = 1 .$$

*Portanto, para todo cilindro  $y_1^n \in \mathcal{C}^n$ , temos que  $\mu(y_1^n) \nu(y_1^n) = 1$  se, e somente se  $x_1^\infty \in y_1^n$ . Assim*

$$\mathbb{E}_{\mu, \nu}(n) = \mu(x_1^\infty) \nu(x_1^\infty) = 1 ,$$

*para todo  $n > 1$ . Isso nos leva à seguinte função taxa:*

$$-\Delta(\epsilon) = \epsilon \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_{\mu, \nu}(n) = 0 .$$

**Exemplo 2.21.** *(Medidas IID) Se  $\mu$  e  $\nu$  são medidas definidas em  $(\mathcal{C}^{\mathbb{N}}, \mathcal{F})$ , com marginais independentes e identicamente distribuídas, com leis*

$$\mu_i = \mu(X = i) \quad , \quad \nu_i = \nu(Y = i) \quad ,$$

*para todo  $i \in \mathcal{C}$ . Nesse caso, teremos então:*

$$\mathbb{E}_{\mu, \nu}(n) = \sum_{\omega \in \mathcal{C}^n} \mu(\omega) \nu(\omega) = \left[ \sum_{i \in \mathcal{C}} \mu_i \nu_i \right]^n .$$

Então, nesse caso, a nossa função taxa será dada por:

$$-\Delta(\epsilon) = \epsilon \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_{\mu, \nu}(n) = \epsilon \sum_{i \in \mathcal{C}} \mu_i \nu_i .$$

## 2.5 A distribuição do caminho mais curto

O objetivo da presente seção é calcular a distribuição de  $T_n^{(2)}$  re-escalada pelo tamanho das sequências. Primeiramente, vamos fazer breves comentários a respeito de algumas ferramentas a serem utilizadas. Em seguida, apresentamos os resultados principais.

Quando nos deparamos com um problema na matemática, é natural tentarmos relacioná-lo, de alguma forma, a um problema conhecido. Assim, estamos tentando trazer o problema para um ambiente em que possamos tratá-lo. E é exatamente isso que faremos a seguir.

O próximo lema possui uma demonstração bem simples, porém uma ideia poderosa: ele nos mostra que, em alguns casos, podemos tratar um subconjunto de  $\mathcal{C}^{\mathbb{N}} \times \mathcal{C}^{\mathbb{N}}$  simplesmente como um subconjunto de  $\mathcal{C}^{\mathbb{N}}$ . Em palavras: levamos um "problema em duas dimensões" para um "problema em uma dimensão", onde já existe um ferramentário conhecido para atacá-lo.

Antes de apresentar tal resultado, vamos aqui definir uma versão unidimensional dos conjuntos  $R_n^{(2)}(k)$ . Tal definição aparece em Lambert [2010], Abadi e Lambert [2013] e Rada [2014].

**Definição 2.22.** Para cada  $k < n$ , defina o conjunto de todas as palavras que possuem alguma sobreposição de tamanho  $k$  com uma cópia dela, transladada á direita. Formalmente:

$$R_n^{(1)}(k) = \{x_1^n \in \mathcal{C}^{\mathbb{N}} : x_{n-k+1}^n = x_1^k\} .$$

Agora, vamos ao lema:

**Lema 2.23.** para todo  $k < m < n$ , vale a seguinte igualdade de conjuntos:

$$R_n^{(2)}(m) \setminus \bigcup_{j=k}^{m-1} R_n^{(2)}(j) = \left\{ (\omega, \omega) \in \mathcal{C}^m \times \mathcal{C}^m : \omega \notin \bigcup_{j=k}^{m-1} R_m^{(1)}(j) \right\} .$$

*Demonstração.* Se  $(x_1^n, y_1^n)$  está em  $R_n^{(2)}(m)$ , então temos:

$$x_1^n = \omega x_{m+1}^n \quad \text{e} \quad y_1^n = y_1^{n-m} \omega ,$$

onde  $\omega$  é uma palavra em  $\mathcal{C}^m$ . Como aqui só estamos interessados em  $\omega$ , podemos jogar fora o começo de  $x_1^n$  e o final de  $y_1^n$ , que são, respectivamente,  $y_1^{m-n}$  e  $x_{m+1}^n$ .

Agora, note que  $(x_1^n, y_1^n)$  não está no conjunto  $\bigcup_{j=k}^{m-1} R_n^{(2)}(j)$  se, e somente se  $\omega$  não possui nenhuma sobreposição de tamanho  $j$ , com  $k \leq j \leq m-1$ . Ou seja,

$$\omega \notin \bigcup_{j=k}^{m-1} R_m^{(1)}(j) .$$

Desta maneira, concluímos a demonstração. □

Antes de apresentar o resultado principal desse capítulo, vamos definir algumas quantidades que nos ajudarão a dar uma forma mais limpa aos enunciados.



**Definição 2.24.** Seja  $\{a_{n,k}^{(2)}\}_{k,n \in \mathbb{N}}$  uma sequência com  $k < n$ , tal que

- $a_{n-1,n}^{(2)} = 0$ .

Além disso, para todo  $1 \leq k \leq n-2$ , defina:

- $a_{k,n}^{(2)} = \sum_{m=k+1}^{n-1} \sum_{\omega \notin \cup_{j=k}^{m-1} R_m^{(1)}(j)} \mu(\omega) \nu(\omega)$  .
- $a_k^{(2)} = \lim_{n \rightarrow \infty} a_{k,n}^{(2)} = \sum_{m=k+1}^{\infty} \sum_{\omega \notin \cup_{j=k}^{m-1} R_m^{(1)}(j)} \mu(\omega) \nu(\omega)$  ,

onde a demonstração do Teorema 2.25 a seguir garante a existência desse limite.

**Teorema 2.25.** Sejam  $\mu$  e  $\nu$  medidas independentes entre si e que admitam a propriedade de gramática completa (ver definição 1.10). Então, para todo  $1 \leq k \leq n-1$ , vale:

(a)  $\mathbb{P}(n - T_n^{(2)} \geq k) = \mathbb{E}_{\mu,\nu}(k) + a_{k,n}^{(2)}$  .

(b)  $\lim_{n \rightarrow \infty} \mathbb{P}(n - T_n^{(2)} \geq k) = \mathbb{E}_{\mu,\nu}(k) + a_k^{(2)}$  .

O resultado acima não nos diz o valor das probabilidades no caso  $k = 0$ . Isso é o que analisaremos a seguir.

Note que, se  $n - T_n^{(2)}(x_1^n, y_1^n) = 0$ , então  $x_1^{n-k} \neq y_k^{n-1}$ , para todo  $k = 1, 2, \dots, n-1$ . Mais precisamente:

$$\bigcup_{j=1}^{n-1} \{x_1^j = y_{n-j+1}^n\} = \emptyset . \quad (2.14)$$

Quando, além da condição 2.14 satisfeita, tivermos  $x_1^n = y_1^n$ , dizemos que  $x_1^n$  (ou  $y_1^n$ ) é uma *sequência auto-evitante*.

O conjunto das sequências auto-evitantes foi estudado primeiramente em Rocha [2009]. Neste trabalho, foi considerada a medida de contagem. Sendo assim, as palavras de tamanho  $n$  eram geradas segundo uma medida uniforme em  $\mathcal{C}^n$ , onde  $|\mathcal{C}| \leq \infty$ . Usando um argumento atribuído a Svante Janson, os autores calcularam a proporção de palavras auto-evitantes.

Além disso, estimaram a velocidade de convergência dessa proporção para uma proporção limite, quando  $n$  tende a infinito.

Posteriormente, o resultado foi generalizado em Abadi e Lambert [2013], e novamente generalizado em Rada [2014]. No primeiro trabalho, foram consideradas palavras geradas segundo uma medida com marginais independentes e identicamente distribuídas (porém não necessariamente uniformes) no alfabeto  $\mathcal{C}$ , e no último, foi considerada a condição de dependência  $\beta$ -mixing para a medida geradora das sequências. Nos dois casos, os resultados valem para  $\mathcal{C}$  finito ou infinito-enumerável.

Seguindo naturalmente essa denominação, chamaremos de *pares evitantes* às duplas  $(x_1^n, y_1^n)$  tais que  $n - T_n^{(2)}(x_1^n, y_1^n) = 0$  (ou, equivalentemente, às duplas que satisfazem a condição 2.14).

O próximo resultado segue imediatamente do teorema 2.25, e nos fornece uma forma de calcular a probabilidade de tal conjunto.

**Corolário 2.26.** *A medida do conjunto de pares evitantes é dada por:*

$$(a) \mathbb{P}(n - T_n^{(2)} = 0) = 1 - \mathbb{E}_{\mu, \nu}(1) - a_{1, n}^{(2)}.$$

$$(b) \lim_{n \rightarrow \infty} \mathbb{P}(n - T_n^{(2)} = 0) = 1 - \mathbb{E}_{\mu, \nu}(1) - a_1^{(2)}.$$

A seguir, provaremos o teorema 2.25:

*Demonstração do Teorema 2.25:*

$$\mathbb{P}(n - T_n^{(2)} \geq k) = \mathbb{P}\left(\bigcup_{j=k}^{n-1} R_n^{(2)}(j)\right)$$

Decompondo o último termo da expressão acima, temos que

$$\begin{aligned} \mathbb{P}\left(\bigcup_{j=k}^{n-1} (R_n^{(2)}(j))\right) &= \mathbb{P}(R_n^{(2)}(k)) + \mathbb{P}(R_n^{(2)}(k+1) \setminus R_n^{(2)}(k)) + \cdots + \mathbb{P}(R_n^{(2)}(n-1) \setminus \bigcup_{j=k}^{n-2} R_n^{(2)}(j)) \\ &= \mathbb{P}(R_n^{(2)}(k)) + \sum_{m=k+1}^{n-1} \mathbb{P}(R_n^{(2)}(m) \setminus \bigcup_{j=k}^{m-1} R_n^{(2)}(j)) \end{aligned}$$

Agora, note que  $\mathbb{P}(R_n^{(2)}(k)) = \sum_{\omega \in \mathcal{C}^k} \mu(\omega)\nu(\omega)$ .

Adicionando o fato acima ao lema 2.23, concluímos

$$\begin{aligned} \mathbb{P}(n - T_n^{(2)} \geq k) &= \sum_{\omega \in \mathcal{C}^k} \mu(\omega)\nu(\omega) + \sum_{m=k+1}^{n-1} \sum_{\omega \notin \bigcup_{j=k}^{m-1} R_m(j)} \mu(\omega)\nu(\omega) \\ &= \mathbb{E}_{\mu, \nu}(k) + a_{k, n}^{(2)}. \end{aligned}$$

Note que os termos da penúltima linha são limitados superiormente por 1. Como  $\mathbb{E}_{\mu, \nu}(k)$  não depende de  $n$ , o limite quando  $n$  tende a infinito só depende do segundo termo, que é uma função não-decrescente em  $n$ . Logo, temos uma sequência monótona e limitada e, portanto, convergente.

Assim, concluímos a demonstração. □

Vamos agora enunciar um corolário que segue imediatamente da Proposição 1.24, e que nos fornece limitantes para o valor de  $a_k^{(2)}$ :

**Corolário 2.27.** *Sob as hipóteses do Teorema 2.25, além das hipóteses da Proposição 1.24 para as medidas  $\mu$  e  $\nu$ , valem os seguintes limitantes para  $a_k^{(2)}$ :*

$$(a) a_k^{(2)} \leq \sum_{i=k+1}^{\infty} \mathbb{E}_{\mu, \nu}(i) - \sum_{i=k+1}^{\infty} \sum_{\omega \in R_i^{(1)}(k)} \mu(\omega)\nu(\omega) .$$

$$(b) a_k^{(1)} \geq \sum_{i=k+1}^{\infty} \mathbb{E}_{\mu, \nu}(i) - \sum_{i=k+1}^{\infty} \sum_{j=k}^{i-1} \sum_{\omega \in R_i^{(1)}(j)} \mu(\omega)\nu(\omega) .$$

**Observação 2.28.** *Note que, se  $\mu = \nu$ , então o corolário 2.27 e a proposição 1.24 se tornam o mesmo resultado.*

### 2.5.1 Não-convergência em probabilidade

Aqui mostramos que a convergência de  $n - T_n^{(2)}$  é exclusivamente fraca, sempre que a divergência relativa de ordem 1 entre as duas medidas for positiva.

**Proposição 2.29.** *Sejam  $\mu$  e  $\nu$  duas medidas estacionárias definidas sobre  $\mathcal{C}^{\mathbb{N}}$ . Uma condição suficiente para que  $n - T_n^{(2)}$  não convirja em probabilidade para uma variável aleatória  $S$  é:*

$$\mathbb{E}_{\mu,\nu}(1) > 0 .$$

*Demonstração.* Se  $\{n - T_n^{(2)}\}_{n \in \mathbb{N}}$  converge em probabilidade, então ela é uma *sequência de Cauchy em probabilidade*. Isso significa dizer que, para todo  $\epsilon > 0$ , vale

$$\lim_{n \rightarrow \infty} \mathbb{P}(|(n+1 - T_{n+1}^{(2)}) - (n - T_n^{(2)})| > \epsilon) = 0 . \quad (2.15)$$

Tome  $0 < \epsilon < 1$ , e note que

$$\{T_{n+1}^{(2)} = T_n^{(2)}\} \subset \{|T_n^{(2)} - T_{n+1}^{(2)} + 1| > \epsilon\} .$$

Agora nos resta provar que  $\mathbb{P}(T_{n+1}^{(2)} = T_n^{(2)}) > \delta > 0$ , para contradizer 2.15 e demonstrar por absurdo que  $\{n - T_n^{(2)}\}_{n \in \mathbb{N}}$  não é uma sequência convergente.

Observe que, para termos  $T_{n+1}^{(2)} = T_n^{(2)}$ , basta que a sequência  $y_1^n$  seja concatenada com o "símbolo conveniente"  $y_{n+1}$ . Ou seja, se  $y_{n+1} = x_r$ , onde  $r = n - T_n^{(2)} + 1$ , teremos

$$y_1^{n+1} = y_1^n y_{n+1} = y_1^n x_r . \quad (2.16)$$

Portanto, teremos que  $T_{n+1}^{(2)} = T_n^{(2)}$ , se e somente se 2.16 for satisfeita. Assim

$$\begin{aligned} \mathbb{P}(T_{n+1}^{(2)} = T_n^{(2)}) &= \sum_{\alpha \in \mathcal{C}} \mathbb{P}(y_{n+1} = x_r | x_r = \alpha) \mathbb{P}(x_r = \alpha) \\ &= \sum_{\alpha \in \mathcal{C}} \mu(\alpha) \nu(\alpha) \\ &= \mathbb{E}_{\mu,\nu}(1) . \end{aligned}$$

Por hipótese, o último termo acima é estritamente positivo. Dessa maneira, finalizamos a prova.  $\square$

A seguir, apresentamos um caso onde a hipótese da Proposição 2.29 não é satisfeita.

**Exemplo 2.30.** *(Medidas concentradas em órbitas 1-periódicas)*

*Considere  $\mu$  e  $\nu$  duas medidas independentes e concentradas nas realizações  $x_1^\infty = 0000 \dots$  e  $y_1^\infty = 1111 \dots$ , respectivamente. Ou seja:*

$$\mu(0000 \dots) = \nu(1111 \dots) = 1 .$$

*Pela definição 2.1, temos que, para todo par de cilindros  $(x_1^n, y_1^n)$  contendo  $(x_1^\infty, y_1^\infty)$ , temos que:*

$$T_n^{(2)}(x_1^n, y_1^n) = n .$$

*Sendo assim, temos:*

$$n - T_n^{(2)} = 0, \mu \times \nu - q.t.p. .$$

Portanto,  $n - T_n^{(2)}$  converge em probabilidade.

# Capítulo 3

## Conclusões

No presente trabalho, conseguimos reproduzir boa quantidade dos resultados clássicos de tempo de primeiro retorno para menor caminho entre duas sequências. O fato é que as hipóteses ainda são muito restritivas. Em muitos casos, acreditamos que podem ser menos restritas. Ainda fica faltando saber, por exemplo, qual a relação entre o valor esperado de  $n - T_n^{(2)}$  e a entropia relativa de Rényi do processo. Tal fato foi levantado em [Lambert \[2010\]](#) e [Rada \[2014\]](#). Outra pergunta interessante é sobre as hipóteses da proposição [2.18](#). Em [Abadi e Cardeño \[2015\]](#), a existência da Entropia de Rényi foi provada sob a hipótese de  $\Psi_g$ -regularidade, hipótese essa que generaliza a condição de Pittel (ver [2.16](#)). Portanto, seria interessante tentar generalizar também esse resultado aqui obtido.



# Referências Bibliográficas

- Abadi (2001)** M. Abadi. Exponential approximation for hitting times in mixing processes. *Mathematical Physics Electronic Journal*, 7(2). Citado na pág. [xxi](#), [xxii](#)
- Abadi (2004)** M. Abadi. Sharp error terms and necessary conditions for exponential hitting times in mixing processes. *The Annals of Probability*, 32(1A):243–264. Citado na pág. [xxi](#)
- Abadi (2008)** M. Abadi. Poisson approximations via chen-stein for non-markov processes. *In and out of equilibrium. 2. Progress in Probability*, 60(2):1–19. Citado na pág. [xxii](#), [7](#)
- Abadi e Cardeño (2015)** M. Abadi e L. Cardeño. Renyi entropies and large deviations for the first match function. *IEEE Trans. Inf. Theory*, 61(4):1629–1639. Citado na pág. [xi](#), [xiii](#), [xv](#), [xxiii](#), [7](#), [29](#)
- Abadi e Lambert (2013)** M. Abadi e R. Lambert. The distribution of the short-return function. *Nonlinearity*, 26(5):1143–1162. Citado na pág. [xi](#), [xiii](#), [xv](#), [xxiii](#), [9](#), [24](#), [25](#)
- Abadi e Saussol (2010)** M. Abadi e B. Saussol. Hitting and returning into rare events for all alpha-mixing processes. *Stochastic process and theirs applications*, 121(2):314–323. Citado na pág. [xxi](#)
- Abadi e Vaienti (2008)** M. Abadi e S. Vaienti. Large deviations for short recurrence. *Discrete and Continuous Dynamical Systems - Series A (DCDS-A)*, 21(3):729 – 747. Citado na pág. [xi](#), [xiii](#), [xv](#), [xxi](#), [xxiii](#), [7](#)
- Afraimovich et al. (2003)** V. Afraimovich, J. Chazottes e B. Saussol. Pointwise dimensions for poincaré recurrence associated with maps and special flows. *Discrete and Continuous Dynamical Systems*, 9:263–280. Citado na pág. [xi](#), [xiii](#), [xv](#), [xxii](#), [7](#), [8](#), [17](#)
- Bowen (1971)** R. Bowen. Periodic points and measures for axiom a diffeomorphisms. *Trans. Amer. Math. Soc.*, 154:377–397. Citado na pág. [6](#)
- Chazottes (2000)** J. R. Chazottes. Dimensions and waiting times for gibbs measures. *Journal of Statistical Physics*, 98(1/2):305–320. Citado na pág. [15](#)
- Cover e Thomas (1991)** T. Cover e J. A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications,, 3th ed. Citado na pág. [16](#)
- Galves e Schmitt (1997)** A. Galves e B. Schmitt. Inequalities for hitting times in mixing dynamical systems. *Random Comput. Dyn*, 5:337–348. Citado na pág. [xxi](#)
- Galves et al. (1999)** A. Galves, P. Collet e B. Schmitt. Repetition times for gibbsian sources. *Nonlinearity*, 12:1225–1237. Citado na pág. [xxii](#)

- Haydn e Vaienti (2010)** N. Haydn e S. Vaienti. The renyi entropy function and the large deviation of short return times. *Ergodic Theory and Dynamical Systems*, 30:159–179. Citado na pág. [xi](#), [xiii](#), [xv](#), [7](#), [21](#)
- Haydn e Vaienti (2009)** N. Haydn e S. Vaienti. The compound poisson distribution and return times in dynamical systems. *Prob. Theory Related Fields*, 144(1):159–179. Citado na pág. [xxiii](#)
- Hirata (1993)** M. Hirata. Poisson law for axiom a diffeomorphisms. *Ergodic Theory and Dynamical Systems*, 13(3):533–556. Citado na pág. [xxi](#)
- Hirata et al. (1999)** M. Hirata, B. Saussol e S. Vaienti. Statistics of returns times: a general framework and new applications. *Comm. Math. Phys.*, 206(3):33–55. Citado na pág. [xxiii](#)
- Kontoyiannis (1998)** I. Kontoyiannis. Asymptotic recurrence and waiting times for stationary processes. *Journal of Theoretical Probability*, 11:79517811. Citado na pág. [15](#)
- Lambert (2010)** Rodrigo Lambert. Comportamento assintótico do primeiro retorno de uma sequência gerada por variáveis aleatórias independentes e identicamente distribuídas. Dissertação de Mestrado, Instituto de Matemática, Estatística e Computação, Universidade Estadual de Campinas, Brasil. Citado na pág. [xxiii](#), [24](#), [29](#)
- Pittel (1985)** B. Pittel. Asymptotical growth of a class of random trees. *The Annals of Probability*, 13(2):414–427. Citado na pág. [21](#)
- Rada (2014)** M. E. A. Rada. *Primeiro Tempo de Retorno para Processos  $\beta$ -Mixing*. Tese de Doutorado. Citado na pág. [xi](#), [xiii](#), [xv](#), [xxiii](#), [9](#), [10](#), [24](#), [25](#), [29](#)
- Reinert e Schbath (1998)** G. Reinert e S. Schbath. Compound poisson and poisson process approximations for occurrences of multiple words in markov chains. *Journal of Computational Biology*, 5:223–254. Citado na pág. [xxii](#)
- Reinert e Schbath (2010)** G. Reinert e S. Schbath. Compound poisson approximations for occurrences of multiple words. *Ergodic Theory and Dynamical Systems*, 30:159–179. Citado na pág. [xxii](#)
- Rocha (2009)** A. V. Rocha. *Substitution Operators*. Tese de Doutorado. Citado na pág. [25](#)
- Roquain e Schbath (2007)** E. Roquain e S. Schbath. Improved compound poisson approximation for the number of occurrences of multiple words in a stationary markov chain. *Advances in Applied Probability*, 39:1–13. Citado na pág. [xxii](#)
- Saussol et al. (2002)** B. Saussol, S. Troubetzkoy e S. Vaienti. Recurrence, dimensions and lyapunov exponents. *Journal of Statistical Physics*, 106:623–634. Citado na pág. [xi](#), [xiii](#), [xv](#), [xxii](#), [6](#), [8](#), [17](#)
- Shields (1993)** P. Shields. Waiting times: Positive and negative results on the wyner-ziv problem. *Journal of Theoretical Probability*, 6(3):499–519. Citado na pág. [15](#)
- Sigmund (1974)** K. Sigmund. On dynamical systems with the specification property. *Trans. Amer. Math. Soc.*, 190:285–299. Citado na pág. [5](#)
- Simonis (1998)** A. Simonis. Filling the hypercube in the supercritical contact process in equilibrium. *Markov Process. Related Fields*, 4(1):113–130. Citado na pág. [2](#)



**Wyner e Ziv (1989)** A. Wyner e J. Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Transactions on Information Theory*, 35(6):1250–1258. Citado na pág. 15

**Wyner (1999)** A. J. Wyner. More on recurrence and waiting times. *The Annals of Applied Probability*, 9(3):780–796. Citado na pág. 15