Title
The Doob-Martin compactification of Markov chains of growing words

Author
Choi, Hye Soo

Publication Date
2017

Peer reviewed|Thesis/dissertation

The Doob-Martin compactification of Markov chains of growing words

by

Hye Soo Choi

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Steven Neil Evans, Chair
Professor David Aldous
Professor Fraydoun Rezakhanlou

Fall 2017

The Doob-Martin compactification of Markov chains of growing words

Abstract

The Doob-Martin compactification of Markov chains of growing words

by

Hye Soo Choi

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Steven Neil Evans, Chair

We study the limiting behavior of Markov chains that iteratively generate a sequence of random finite words.

In the first part of thesis, we consider a Markov chain such that the n^{th} word is uniformly distributed over the set $\mathbb{W}_n$ of words of length $2n$ in which n letters are a and n letters are b: at each step an a and a b are shuffled in uniformly at random among the letters of the current word. We also consider a Markov chain such that the n^{th} word takes values in the set of words in $\mathbb{B}_n$ such that the number of letters a in the first k letters is at least the number of letters b in those positions for any $1 \leq k \leq 2n$: at each step an a and a b are shuffled uniformly at random into the existing word so that the a precedes the b. We obtain a concrete characterization of the respective Doob-Martin boundaries of these Markov chains and thereby delineate all the ways in which the Markov chains can be conditioned to behave at large times. We exhibit a bijective correspondence between the points in the respective boundaries of Markov chains and ergodic random total orders on the set $\{a_1, b_1, a_2, b_2, \ldots\}$ that have the specific properties determined by the Markov chains. We establish for the first Markov chain a further bijective correspondence between the set of such random total orders and the set of pairs (μ, ν) of diffuse probability measures on $[0, 1]$ such that $\frac{1}{2}(\mu+\nu)$ is Lebesgue measure: the restriction of the random total order to $\{a_1, b_1, \ldots, a_n, b_n\}$ is obtained by taking $X_1, \ldots, X_n$ (resp. $Y_1, \ldots, Y_n$) i.i.d. with common distribution μ (resp. ν), letting $(Z_1, \ldots, Z_{2n})$ be $\{X_1, Y_1, \ldots, X_n, Y_n\}$ in increasing order, and declaring that the k^{th} smallest element in the restricted total order is a_i (resp. b_j) if $Z_k = X_i$ (resp. $Z_k = Y_j$).

The second part of thesis focuses on the mixing time of a Markov chain of words in $\mathbb{W}_n$ that arises from removing a letter a and a letter b uniformly at random followed by inserting the letter a and the letter b uniformly at random back into one of the slots defined by the remaining letters. We present an upper bound of the form $n \log n + (\log 8)n$ and a lower bound of the form $(1 - \alpha)n \log n - c_\alpha n, \alpha > \frac{1}{2}$.

Contents

Acknowledgments

First, I would like to express sincere gratitude to my advisor, Professor Steven N. Evans, for his patient guidance throughout my time at Berkeley. His vast knowledge gave me the freedom to explore many research areas in whichever direction I choose and the opportunity to find the most interesting subject to work on. He has been always the first person to whom I looked for inspiration and guidance, and has been very patient throughout my extremely clumsy and slow path of learning. Without him, it would have been impossible for me to write this dissertation.

It has been such a pleasure getting to know many professors at Berkeley including David Aldous, Jim Pitman, Allan Sly, Elchanan Mossel, Nike Sun, Fraydoun Rezhakanlou, and Bin Yu. I was deeply impressed not only by their enthusiasm and achievement in research but also by the realization that they are truly humble people.

I would like to say thanks to many warm-hearted friends with a brilliant mind in our department for being such a great company. I cannot imagine my grad school years without Yumeng, Yuting, and Lisha. They are the wonderful friends with whom I shared excitements and frustrations both in research and in my personal life. Sujayam, Fu, and Christine are joyous people who brought me to a lot of great restaurants. I also want to say thanks to Wenpin and Miki for encouraging words.

I appreciate Berkeley Statistics department for providing such a great environment for Ph.D. students. Especially I would like to thank La Shana and Mary for their outstanding abilities in sorting out the many administrative problems and cheerful banters.

Last but not least, my deepest gratitude goes to my parents and friends who are not mentioned here. I would like to say thanks for always supporting me for who I am with faith and love.

Chapter 1

Introduction

This thesis studies the limiting behavior of Markov chains of words drawn from the alphabet $\{a, b\}$. We build Markov chains of growing words drawn from the alphabet $\{a, b\}$ in such a way that, starting with the empty word, a letter a and a letter b are added into random positions of the current word. By imposing different rules for the manner in which a letter a and a letter b are added, we construct two Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$ that arise from two different addition mechanisms. As the length of the words U_n and W_n in the Markov chains goes off to infinity as $n \to \infty$, the possible arrangements of the $2n$ letters of the words become more diversified over time. We delineate the wide variety of limiting behavior by considering the Doob-Martin compactification of the Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$(see Section 2.2 for a summary of the relevant general theory).

We also consider a Down-up Markov chain $(DU_t)_{t\in\mathbb{N}_0}$ for which the length of the word remains constant over time. The forward transition dynamics of $(DU_t)_{t\in\mathbb{N}_0}$ is built from the chain $(U_n)_{n\in\mathbb{N}_0}$ by composing the backward transition dynamics of $(U_n)_{n\in\mathbb{N}_0}$ with the forward transition dynamics of $(U_n)_{n\in\mathbb{N}_0}$: a letter a and a letter b are chosen at random, removed from the current word, and reinserted in random positions. This can be regarded as relative of random-to-random card shuffling where a card is chosen at random, removed from the deck, and reinserted in a random position. We study the limiting behavior of the chain $(DU_t)_{t\in\mathbb{N}_0}$ by investigating how fast the distribution of DU_t converges to its stationary distribution as $t \to \infty$ with respect to total variation distance. In doing so, we obtain bounds on the mixing time of the chain $(DU_t)_{t\in bN_0}$.

1.1 Motivation

As suggested by the existence [27] of more than 300 equivalent combinatorial interpretations of the Catalan number, many bijective correspondences exist between the families of discrete combinatorial ensembles counted by the Catalan numbers. Indeed, our investigation for Markov chains of words drawn from two-letter alphabet was initially motivated by such a bijection between trees and a certain type of words drawn from the two-letter alphabet $\{a, b\}$.

This relation enables every Markov chain that appears in this thesis to be transformed into a Markov chain of corresponding trees and the details will be discussed in later chapters. To explain the bijection, we recall some relevant terms and a good source for the material in this part is [1].

Definition 1.1.1. A *Dyck word* is defined iteratively as follows:

- The empty word $\emptyset$ is a Dyck word,
- All words w that can be factored as $w = aubv$ where the subwords u, v are Dyck words are Dyck words.

Definition 1.1.2. A *1-dominated word* with $n+1$ letters a and n letters b is a word w of form $w = au$ where u is a Dyck word of length $2n$.

Definition 1.1.3. A *tree* on the finite set $S \neq \emptyset$ is a pair $(r, (T_1, T_2, \ldots, T_k))$ with $k \geq 0$ such that

- r is an element of S,
- $T_1, \ldots, T_k$ are some trees defined on the sets $S_1, ..., S_k$,
- the sets $\{r\}, S_1, \ldots, S_k$ form a partition of S.

Definition 1.1.4. A *postfix labeling* of the nodes of the tree $T = (r, (T_1, \ldots, T_k))$ for an ordered set B with $\#B = \#S$, is a labeling such that the root r is labeled with $\max B$ and the subtrees $T_1, \ldots, T_k$ are given postfix labelings with the respective sets $B_1, \ldots, B_k$, where $B_1, \ldots, B_k$ is a partition of $B \backslash \{\max B\}$ and $b_i < b_j$ for $b_i \in B_i$ and $b_j \in B_j$ with $i < j$. When we refer to the postfix labeling of T we mean the postfix labeling with $B = [n]$, where $n = \#S$ is the number of nodes of T.

It is a well-known result from [26] (see also [1, Chapter 4]) that there exists a bijection between the set of trees with $n+1$ nodes and the set of 1-dominated words with $n+1$ letters a and n letters b for n in $\mathbb{N}_0$. While traversing nodes in a tree in the order of the postfix labeling, one can construct a 1-dominated word with $n+1$ letters a and n letters b as follows: starting from the empty word $\emptyset$, one can keep concatenating the word with a letter a followed by as many letters b as the node has children.

There is, moreover, a bijection between pairs (i, w), (see [1, Section 4.3.3] for details) where $i \in [2n-1]$ and w is a 1-dominated word with n letters a and $n-1$ letters b, and words v with n letters a and $n-1$ letters b: the bijection sends (i, w) to the word $v = v_1 \cdots v_{2n-1}$ given by $v_j = w_{i+j}$, where the addition is modulo $2n-1$.

For future reference in later chapters, we formally state these theorems.

Theorem 1.1.5. *There exists a bijection between the set of trees with $n+1$ nodes and the set of 1-dominated words with $n+1$ letters a and n letters b for n in $\mathbb{N}_0$.*

Theorem 1.1.6. *There exists a bijection between pairs (i, w), where $i \in [2n-1]$ and w is a 1-dominated word with n letters a and $n-1$ letters b, and words v with n letters a and $n-1$ letters b.*

1.2 Main results

To introduce our main results formally, we give the definition of the Markov chains $(U_n)_{n\in\mathbb{N}_0}$, $(W_n)_{n\in\mathbb{N}_0}$, and $(DU_t)_{t\in\mathbb{N}_0}$.

Definition 1.2.1. Define a Markov chain $(U_n)_{n\in\mathbb{N}_0}$ such that

- U_0 is the empty word $\emptyset$,
- conditional on $U_n = u_1 \dots u_{2n}$, the word U_{n+1} is constructed by choosing $1 \leq I_{n+1},\ J_{n+1} \leq 2n+2$ with $I_{n+1} \neq J_{n+1}$ uniformly at random, placing an a in position I_{n+1} and a b in position J_{n+1}, and placing the letters $u_1, \dots, u_{2n}$ in that order into the remaining $2n$ positions.

Definition 1.2.2. Define a Markov chain $(W_n)_{n\in\mathbb{N}_0}$ such that

- W_0 is the empty word $\emptyset$.
- conditional on $W_n = v_1 \dots v_{2n}$, the word W_{n+1} is constructed by choosing $1 \leq I_{n+1} < J_{n+1} \leq 2n+2$ uniformly at random (that is, all $\binom{2n+2}{2}$ possibilities are equally likely), placing an a in position I_{n+1} and a b in position J_{n+1}, and placing the letters $v_1, \dots, v_{2n}$ in that order into the remaining $2n$ positions.

Definition 1.2.3. Define a Markov chain $(DU_t)_{t\in\mathbb{N}_0}$ such that

- DU_0 is a random word of length $2n$ drawn from the alphabet $\{a, b\}$ that consists of n letters a and n letters b,
- conditional on $DU_t = u_1 u_2 \cdots u_{2n-1} u_{2n}$, the word DU_t is constructed by choosing I_t and J_t independently and uniformly at random from the index set $\{1 \leq i \leq 2n : u_i = a\}$ and the index set $\{1 \leq j \leq 2n : u_j = b\}$), respectively, and $1 \leq \widetilde{I}_t,\ \widetilde{J}_t \leq 2n$ with $\widetilde{I}_t \neq \widetilde{J}_t$ uniformly at random (that is, all $2n(2n-1)$ possibilities are equally likely) followed by placing an a in position $\widetilde{I}_t$ and a b in position $\widetilde{J}_t$, and placing the letters $u_{\sigma(1)} u_{\sigma(2)} \cdots u_{\sigma(2n-2)}$ in that order in the remaining $2n-2$ positions, where σ is the unique increasing bijection from $[2n-2]$ to $[2n] \setminus \{I_t, J_t\}$.

The following theorem collects together the results we establish in this dissertation. We will describe the bijections asserted in Theorem 1.2.4 and Theorem 1.2.5 explicitly as we proceed.

Theorem 1.2.4. *There are bijective correspondences between the following sets:*

(i) extremal elements of the Doob–Martin boundary associated with the transition probabilities of the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ *and the reference state* $\emptyset$*,*

(ii) extremal nonnegative harmonic functions for the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ *normalized to take the value* 1 *at the reference state* $\emptyset$*,*

(iii) pairs (μ, ν) *of probability measures on* $[0,1]$ *such that* $\frac{1}{2}(\mu+\nu) = \lambda$*, the Lebesgue measure on* $[0,1]$*.*

(iv) the distributions of extremal infinite bridges for the Markov chain $(U_n)_{n\in\mathbb{N}_0}$*,*

(v) the distributions of ergodic exchangeable random total orders $\prec$ *on the set* $\mathbb{I}_0 = \bigcup_{n\in\mathbb{N}}\{a_n, b_n\}$*.*

Theorem 1.2.5. *There are bijective correspondences between the following sets:*

(i) extremal elements of the Doob–Martin boundary associated with the transition probabilities of the Markov chain $(W_n)_{n\in\mathbb{N}_0}$ *and the reference state* $\emptyset$*,*

(ii) extremal nonnegative harmonic functions for the Markov chain $(W_n)_{n\in\mathbb{N}_0}$ *normalized to take the value* 1 *at the reference state* $\emptyset$*,*

(iii) probability measures μ *concentrated on* $\{(x,y) : 0 \le x \le y \le 1\}$ *such that* $\frac{1}{2}(\mu(\cdot \times \mathbb{R}) + \mu(\mathbb{R} \times \cdot)) = \lambda$*, the Lebesgue measure on* $[0,1]$ *and the labeled infinite bridge* $(L_n^\infty)_{n\in\mathbb{N}_0}$ *uniquely determined by* μ *satisfies that the conditional distribution of* L_n^∞ *given* $\Psi(L_n^\infty)$ *is uniform on the set of labeled associated admissible matchings of* $\Psi(L_n^\infty)$ *for all* $n \in \mathbb{N}_0$*.*

(iv) the distributions of extremal infinite bridges for the Markov chain $(W_n)_{n\in\mathbb{N}_0}$*,*

(v) the distributions of ergodic exchangeable random paired total orders $\prec$ *on the set* $\mathbb{I}_0 = \bigcup_{n\in\mathbb{N}}\{a_n, b_n\}$ *such that the labeled infinite bridge* $(L_n^\infty)_{n\in\mathbb{N}_0}$ *uniquely determined by* $\prec$ *satisfies that the conditional distribution of* L_n^∞ *given* $\Psi(L_n^\infty)$ *is uniform on the set of labeled associated admissible matchings of* $\Psi(L_n^\infty)$ *for all* $n \in \mathbb{N}_0$*,*

where $\Psi(L_n^\infty)$ *is the word drawn from the alphabet* $\{a, b\}$ *that consist of* n *letters* a *and* n *letters that is uniquely determined by* L_n^∞*,* $n \in \mathbb{N}_0$*.*

Theorem 1.2.6. *The mixing time* t_{mix} *for the Markov chain* $(DU_t)_{t\in\mathbb{N}_0}$ *has an upper bound of the form* $n\log n + (\ln 8)\, n$ *and a lower bound of the form* $(1-\alpha)n\log n - c_\alpha n, \alpha > \frac{1}{2}$*.*

1.3 Overview

The remainder of this thesis is organized into two additional chapters. In Chapter 2, we characterize the Doob-Martin compactification of the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$. This chapter is the combination of a joint work with Steven N. Evans [6] and a joint work with Steven N. Evans and Anton Wakolbinger. As the characterizations of the Doob-Martin compactification of the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$ have many common features due to their apparent similarity, we interleave the processes of identifying the Doob-Martin compactification of the Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$. However there are still some subtle difference in the two processes, so we often distinguish the part for the chain $(U_n)_{n\in\mathbb{N}_0}$ and the part for the chain $(W_n)_{n\in\mathbb{N}_0}$ by splitting the parts into subsections. In Chapter 3, we obtain upper and lower bounds on the mixing time of the down-up chain $(DU_t)_{t\in\mathbb{N}_0}$ using coupling methods.

Chapter 2

The Doob-Martin compactification of Markov chains of growing words

2.1 Introduction

Generating Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$ drawn from the alphabet $\{a, b\}$

There is a very simple way of producing a uniformly distributed random permutation of a set with n objects, say $[n] := \{1, \ldots, n\}$: we take the elements of $[n]$ in order and lay them down successively so that the k^{th} element goes into a uniformly chosen one of the k "slots" defined by the $k-1$ elements that have already been laid down (the slot before the first element, the slot after the last element, or one of the $k-2$ slots between elements). This sequential algorithm has the attractive feature that when the first k elements have been laid down they are in uniform random order; that is, the algorithm builds uniformly distributed random permutations of $[1], [2], \ldots, [n]$ in a sequential manner.

Suppose that we enumerate a standard deck of cards with the elements of the set [52]. If the deck is in some order, then the colors of the successive cards (**R**ed or **B**lack) define a word of length 52 from the two-letter alphabet $\{R, B\}$ in which 26 letters are R and 26 letters are B (recall that a word of length k from a finite alphabet $\mathcal{A}$ is just an element of the Cartesian product $\mathcal{A}^k$, although it is usual to write the word $(a_1, \ldots, a_k)$ more succinctly as $a_1 \cdots a_k$). Moreover, if the order of the deck is random and uniformly distributed, then the resulting word is uniformly distributed over the set of $\frac{52!}{26!26!}$ such words.

Unfortunately, our sequential randomization algorithm doesn't have the feature that at the $(2k)^{th}$ step for $1 \le k \le 26$ we have a random word from the alphabet $\{R, B\}$ that is uniformly distributed over the set of $\binom{2k}{k}$ words in which k letters are R and k letters are B.

However, there is a simple way of modifying our algorithm to produce the latter type of random words sequentially. We begin at step 0 with the empty word. Suppose that we have completed k steps and a word of length $2k$ has been produced. The first sub-step of step $k+1$ inserts the letter R uniformly at random into one of the $2k+1$ slots defined by these $2k$ letters to produce a word of length $2k+1$. The second sub-step inserts the letter B uniformly at random into one of the $2k+2$ slots defined by these $2k+1$ letters to produce a word of length $2k+2$ and thereby complete step $k+1$. It is not difficult to see that, despite the apparent dependence of this procedure on the ordering of the letters R and B, this procedure does indeed achieve what it is claimed to achieve.

From now on we will replace the alphabet $\{R, B\}$ by the alphabet $\{a, b\}$ and write $(U_n)_{n\in\mathbb{N}_0}$ for the Markov chain that arises from our random insertion procedure. The formal definition of $(U_n)_{n\in\mathbb{N}_0}$ is given as follows.

Definition 2.1.1. Define a Markov chain $(U_n)_{n\in\mathbb{N}_0}$ such that

- U_0 is the empty word $\emptyset$,
- conditional on $U_n = u_1 \ldots u_{2n}$, the word U_{n+1} is constructed by choosing $1 \le I_{n+1},\ J_{n+1} \le 2n+2$ with $I_{n+1} \ne J_{n+1}$ uniformly at random, placing an a in position I_{n+1} and a b in position J_{n+1}, and placing the letters $u_1, \ldots, u_{2n}$ in that order into the remaining $2n$ positions.

Thus, $U_n \in \mathbb{W}_n$, where $\mathbb{W}_n$ is the set words drawn from the alphabet $\{a, b\}$ that consist of n letters a and n letters b. Set $\mathbb{W} := \bigsqcup_{n\in\mathbb{N}_0} \mathbb{W}_n$ and denote $N(w) = n$ for $w \in \mathbb{W}_n$, $n \in \mathbb{N}_0$.

A *ballot sequence* is a string consisting of n letters a and n letters b such that for any $1 \le k \le 2n$ the number of letters a in the first k letters is at least the number of letters b in those positions. Note that a ballot sequence is just another expression for a Dyck word defined in Section 1.1. Write $\mathbb{B}_n$ for the set of ballot sequences of length $2n$, and set $\mathbb{B} := \bigsqcup_{n\in\mathbb{N}_0} \mathbb{B}_n$. The cardinality of the set $\mathbb{B}_n$ is the nth Catalan number C_n, where

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

See [27] for an enumeration of other combinatorial interpretations, generalization, refinements, variants of the Catalan number.

A natural way to generate ballot sequences in a sequential manner is to start with the empty word and insert an additional letter a and an additional letter b at each step in such a way that such a way that the additional letter a comes before the additional letter b in the word.

Definition 2.1.2. Define a Markov chain $(W_n)_{n\in\mathbb{N}_0}$ such that

- W_0 is the empty word $\emptyset$.
- conditional on $W_n = v_1 \ldots v_{2n}$, the word W_{n+1} is constructed by choosing $1 \le I_{n+1} < J_{n+1} \le 2n+2$ uniformly at random (that is, all $\binom{2n+2}{2}$ possibilities are equally likely), placing an a in position I_{n+1} and a b in position J_{n+1}, and placing the letters $v_1, \ldots, v_{2n}$ in that order into the remaining $2n$ positions.

Thus, W_n takes values in $\mathbb{B}_n$ for $n \in \mathbb{N}_0$. Denote $N(w) = n$ for $w \in \mathbb{B}_n$, $n \in \mathbb{N}_0$.

The Markov chains of trees that correspond to Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$

Markov chain $(U_n)_{n\in\mathbb{N}_0}$ that sequentially generates random words such that U_n is uniformly distributed over the set of words drawn from the alphabet $\{a, b\}$ that have n letters a and n letters b can be trivially modified to give a Markov chain $(\widetilde{U}_n)_{n\in\mathbb{N}_0}$ such that $\widetilde{U}_n$ is uniformly distributed over the set of words drawn from the alphabet $\{a, b\}$ that have $n+1$ letters a and n letters b: instead of starting with the empty word, one starts with the word consisting of a single letter a. By Theorem 1.1.5 and Theorem 1.1.6 there is a bijection that turns the latter Markov chain into a Markov chain $((I_n, \widetilde{S}_n))_{n\in\mathbb{N}_0}$, where I_n and $\widetilde{S}_n$ are independent for each n, I_n is uniformly distributed on $[2n+1]$, and $\widetilde{S}_n$ is uniformly distributed over the set of 1-dominated word with $n+1$ letters a and n letters b. Moreover, composing these two bijections turns the Markov chain $(\widetilde{U}_n)_{n\in\mathbb{N}_0}$ into a Markov chain $((I_n, S_n))_{n\in\mathbb{N}_0}$, where I_n and S_n are independent for each n, I_n is uniformly distributed on $[2n+1]$, and S_n is uniformly distributed over the set of trees with $n+1$ vertices that are equipped with the postfix labeling.

Theorem 1.1.5 also enables the Markov chain $(W_n)_{n\in\mathbb{N}_0}$ that sequentially generates a word $\mathbb{B}_n$ to be transformed into a Markov chain $(\widetilde{T}_n)_{n\in\mathbb{N}_0}$ such that $\widetilde{T}_n$ are trees with $n+1$ nodes, and vice versa. The stepwise procedure can be described as follows. By concatenating letter a with W_n, we can build a 1-dominated word with $n+1$ letters a and n letters b. This 1-dominated word can then form the corresponding tree with $n+1$ nodes under the bijection discussed above. One can compose these two transformations to turn the Markov chain $(W_n)_{n\in\mathbb{N}_0}$ into a Markov chain $(\widetilde{T}_n)_{n\in\mathbb{N}_0}$. Note that all the steps are invertible.

The Markov chains $((I_n, T_n))_{n\in\mathbb{N}_0}$ and $(\widetilde{T}_n)_{n\in\mathbb{N}_0}$ are certainly transient and have countable state spaces. We are interested in the manner in which (I_n, T_n) and $(\widetilde{T}_n)_{n\in\mathbb{N}_0}$ "escape to infinity" as $n \to \infty$. This question is made precise by considering the Doob-Martin compactification of the Markov chains (see Section 2.2 for a summary of the relevant general theory). One can obtain an explicit description of the transition dynamics of these Markov chains, but since $((I_n, T_n))_{n\in\mathbb{N}_0}$ and $(\widetilde{T}_n)_{n\in\mathbb{N}_0}$ are just the Markov chains $(V_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$, respectively, "in disguise" it is easier to work with the latter Markov chains (the limit space of the Markov chains within the Doob-Martin compactification remains the same up to isomorphism). Since $(V_n)_{n\in\mathbb{N}_0}$ is a minor modification of $(U_n)_{n\in\mathbb{N}_0}$, we go even further and investigate $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$.

Shuffle product and Hopf algebra

Let $\overline{\mathbb{W}}$ be the algebra that consists of the empty word $\emptyset$ as well as finite linear combinations of words drawn from the alphabet $\{a, b\}$ with scalars in $\mathbb{Q}$. For example, $ab + \frac{2}{3}ba$ and $\frac{2}{7}abbb + baabb + \frac{1}{5}bbbbaa$ are elements in the algebra $\overline{\mathbb{W}}$. Equivalently $\overline{\mathbb{W}}$ is a free $\mathbb{Q}$-module generated by finite words drawn from the alphabet $\{a, b\}$.

Definition 2.1.3. The *shuffle product* $\sqcup\!\sqcup$ is a bilinear map from $\overline{\mathbb{W}} \times \overline{\mathbb{W}}$ to $\overline{\mathbb{W}}$. It is uniquely determined by its value for each pair of words. The shuffle product $v \sqcup\!\sqcup w$ of two words $v = v_1 \cdots v_m$ and $w = w_1 \cdots w_n$ is

$$v \sqcup\!\sqcup w := \sum_{\substack{f:[m]\mapsto[m+n] \\ g:[n]\mapsto[m+n]}} u_1 \cdots u_{n+m},$$

where the sum runs over pairs of functions (f, g) that satisfy the conditions listed as follows:

- $f : [m] \mapsto [m + n]$ and $g : [n] \mapsto [m + n]$ are strictly increasing,
- $f([m]) \cap g([n]) = \emptyset$
- $f([m]) \cup g([n]) = [n + m]$,
- the letter $u_{f(i)}$ is the ith letter v_i in the word v for $i \in [m]$ and likewise the letter $u_{g(j)}$ is the jth letter w_j in the word w for $j \in [n]$.

For examples,

$$\begin{aligned} a \sqcup\!\sqcup \mathbf{b} &= a\mathbf{b} + \mathbf{b}a, \\ ab \sqcup\!\sqcup \mathbf{ab} &= \mathbf{ab}ab + \mathbf{a}a\mathbf{b}b + \mathbf{a}ab\mathbf{b} + a\mathbf{ab}b + a\mathbf{a}b\mathbf{b} + ab\mathbf{ab} \\ &= 2abab + 4aabb. \end{aligned}$$

In other words, the shuffle product of two words is the sum of all ways of interleaving the two words.

Up to a normalization, the coefficient of a word $u = u_1 \cdots u_{n+m}$ in the shuffle product of two words v and w is the probability that the shuffle of v and w results in u. Taking the previous example, the shuffle of a with b results in ab with probability $\frac{1}{2}$ and ba with probability $\frac{1}{2}$. It is noteworthy that the algebra $\overline{\mathbb{W}}$ equipped with the shuffle product and suitable associated coproduct and antipode constitutes a prime example of a Hopf algebra – see, for example, [4, 23, 21, 8, 5, 10]. The Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$ can be embedded into the algebra $\overline{\mathbb{W}}$, and their distributions can be characterized by

$$\sum_{w\in\mathbb{W}_n} \mathbb{P}\{U_n = w\}\, w = \frac{1}{(2n)!} \underbrace{a \sqcup\!\sqcup b \sqcup\!\sqcup a \sqcup\!\sqcup b \sqcup\!\sqcup \cdots \sqcup\!\sqcup a \sqcup\!\sqcup b}_{2n \text{ terms}},$$

$$\sum_{w\in\mathbb{B}_n} \mathbb{P}\{W_n = w\}\, w = \frac{2^n}{(2n)!} \underbrace{ab \sqcup\!\sqcup ab \sqcup\!\sqcup \cdots \sqcup\!\sqcup ab}_{n \text{ terms}}.$$

Infinite bridges

We investigate the infinite bridges (equivalently, the Doob h-transforms) for the Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$; that is, the Markov chains that have the same backwards-in-time transition dynamics as $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$, respectively. We thereby identify the

Doob-Martin compactification of the state space of each Markov chain. This enables us to characterize the nonnegative harmonic functions for the Markov chains and hence delineate all the ways that the Markov chains can be conditioned to "behave at infinity". Merging the processes of characterizing the Doob-Martin compactification of the Markov chains $(U_n)_{n\in\mathbb{N}_0}$ and $(W_n)_{n\in\mathbb{N}_0}$ may confuse the reader and thus we divide our exposition into two parts (the parts use concepts with similar names but use them in different ways).

Infinite bridges and exchangeable random total orders for $(U_n)_{n\in\mathbb{N}_0}$

More specifically, we show that a $\mathbb{W}$-valued Markov chain is an infinite bridge for the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ if and only if the backwards dynamics are given by removing one letter a and one letter b uniformly at random from the current word. We can enrich the state space of the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ by replacing $\mathbb{W}_n$ with the set $\tilde{\mathbb{W}}_n$ that consists of words made up from the letters $a_1, b_1, \dots, a_n, b_n$ written down in some order (each letter appearing once); that is, a word such as $aababb$ will be associated with a word such as $a_3a_1b_2a_2b_1b_3$ – a given $w \in \mathbb{W}_n$ has $(n!)^2$ associated words in $\tilde{\mathbb{W}}_n$. We can then enhance an infinite bridge $(U_n^\infty)_{n\in\mathbb{N}_0}$ to produce a Markov chain $(\tilde{U}_n^\infty)_{n\in\mathbb{N}_0}$ with values in $\tilde{\mathbb{W}} := \bigsqcup_{n\in\mathbb{N}_0} \tilde{\mathbb{W}}_n$ such that given $U_n^\infty = u$ the value of $\tilde{U}_n^\infty$ is uniformly distributed over all ways of "subscripting" the letters in u; for example, if $U_2^\infty = abba$, then $\tilde{U}_2^\infty$ is uniformly distributed over the four words $a_1b_1b_2a_2$, $a_2b_1b_2a_1$, $a_1b_2b_1a_2$, $a_2b_2b_1a_1$. Moreover, in going from $\tilde{U}_n^\infty$ to $\tilde{U}_{n-1}^\infty$ the letters a_n and b_n are deleted. We may view $\tilde{U}_n^\infty$ as a random total (that is, linear) order on the set $\{a_1, b_1, \dots, a_n, b_n\}$. As n varies, these orders are consistent in the sense that the order $\tilde{U}_n^\infty$ induces on $\{a_1, b_1, \dots, a_{n-1}, b_{n-1}\}$ is just the order given by $\tilde{U}_{n-1}^\infty$. Consequently, there is a total order on $\{a_1, b_1, a_2, b_2, \dots\}$ that induces each of the orders given by the $\tilde{U}_n^\infty$. This total order is exchangeable in the sense that finite permutations of the subscripts of the a's and b's separately leave its distribution unchanged.

The infinite bridge $(U_n^\infty)_{n\in\mathbb{N}_0}$ is extremal (that is, not a mixture of infinite bridges or, equivalently, has an almost surely trivial tail σ-field) if and only if the exchangeable random total order on $\{a_1, b_1, a_2, b_2, \dots\}$ is ergodic in the sense that if an event is unchanged by finite permutations of the subscripts of the a's and b's separately, then it has probability zero or one. By general Doob–Martin theory, extremal bridges correspond to extremal elements of the Doob–Martin boundary and, in general, some elements of the Doob–Martin boundary may not be extremal. We show that the latter phenomenon does not occur in our setting – all Doob–Martin boundary points of the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ are extremal.

We demonstrate that there is a bijective correspondence between ergodic exchangeable random total orders on $\{a_1, b_1, a_2, b_2, \dots\}$ and pairs (μ, ν) of diffuse probability measures on the unit interval $[0, 1]$ such that $\frac{\mu+\nu}{2} = \lambda$, where λ is Lebesgue measure on $[0, 1]$: let $V_1, V_2, \dots$ be i.i.d. with distribution μ and $W_1, W_2, \dots$ be independent and i.i.d. with distribution ν, then, writing $\prec$ for the total order we have $a_i \prec a_j$ (resp. $a_i \prec b_j$, $b_i \prec a_j$, $b_i \prec b_j$) if $V_i < V_j$ (resp. $V_i < W_j$, $W_i < V_j$, $W_i < W_j$). Another way of describing this construction is the following. We only need to describe the restriction of the random total order to $\{a_1, b_1, \dots, a_n, b_n\}$ for each $n \in \mathbb{N}_0$. Let $(Z_1, \dots, Z_{2n})$ be $\{V_1, W_1, \dots, V_n, W_n\}$ in increasing order and declare that the k^{th} smallest element of $\{a_1, b_1, \ldots, a_n, b_n\}$ in the restricted total order is a_i (resp. b_j) if $Z_k = X_i$ (resp. $Z_k = Y_j$).

We remark that, due to the relationship $\frac{\mu+\nu}{2} = \lambda$, the probability measure ν is uniquely determined by the probability measure μ and *vice versa* and hence we could have said that the ergodic exchangeable random total orders are in bijective correspondence with the probability measures μ on $[0,1]$ that satisfy $\mu \le 2\lambda$. However, we find the more symmetric description to be preferable.

In terms of the Doob–Martin topology, we show that a sequence $(y_k)_{k\in\mathbb{N}}$ with $y_k \in \mathbb{W}_{N(y_k)}$ and $N(y_k) \to \infty$ as $k \to \infty$ converges to the point in the Doob–Martin boundary corresponding to the pair of measures (μ, ν) if and only if for each $m \in \mathbb{N}$ the sequence of random words obtained by selecting m letters a and m letters b uniformly at random from y_k and maintaining their relative order converges in distribution as $k \to \infty$ to the random word that is obtained by writing $V_1, \ldots, V_m, W_1, \ldots, W_m$ in increasing order to make a list $(Z_1, \ldots, Z_{2m})$ as above and then putting a letter a (resp. b) in position ℓ of the word when $Z_\ell \in \{V_1, \ldots, V_m\}$ (resp. $Z_\ell \in \{W_1, \ldots, W_m\}$). Moreover, the convergence of $(y_k)_{k\in\mathbb{N}}$ to y is equivalent to the weak convergence of μ_k to μ and ν_k to ν, where μ_k (resp. ν_k) is the probability measure that places mass $\frac{1}{N(y_k)}$ at the point $\frac{\ell}{2N(y_k)}$ $1 \le \ell \le 2N(y_k)$, if the ℓ^{th} letter of the word y_k is the letter a (resp. b).

Infinite bridges and exchangeable random paired total orders for $(W_n)_{n\in\mathbb{N}_0}$

A *matching* of $[2n]$ is a partition of $[2n]$ into subsets (called blocks) of size 2. We can take any matching $\mathcal{M}$ of $[2n]$ and produce a ballot sequence $w = w_1 \ldots w_{2n} \in \mathbb{B}_n$: if $\{i, j\}$ is a block of $\mathcal{M}$ with $i < j$, then $w_i = a$ and $w_j = b$. Conversely, we say that a matching $\mathcal{M}$ of $[2n]$ is an *associated admissible matching* for the word $w = w_1 \ldots w_{2n}$ if for every block $\{i, j\}$ of $\mathcal{M}$ with $i < j$ we have $w_i = a$ and $w_j = b$; that is, it is possible to build w by successive shuffles in such a way that the indices in each block correspond to letters that were shuffled in together. A *labeled matching* of $[2n]$ is a matching in which the n blocks are labeled with distinct elements of $[n]$. A labeled associated admissible matching $\mathcal{L}$ for the word $w = w_1 \ldots w_{2n}$ describes a possible way of building w by successive shuffles: if $\{i, j\}$ is a block of $\mathcal{L}$ labeled by k, then i (respectively, j) is the location at step n of the letter a (respectively, b) shuffled in at step k. Given a labeled matching $\mathcal{L}$ of $[2n]$, let $\Psi(\mathcal{L}) \in \mathbb{B}_n$ be the corresponding word (that is, forget about the block labels and for each block $\{i, j\}$ with $i < j$ place a letter a in position i and a letter b in position j).

Suppose that $(W_n^\infty)_{n\in\mathbb{N}_0}$ is an infinite bridge. We show that there exists a Markov process $(L_n^\infty)_{n\in\mathbb{N}_0}$ with distribution uniquely specified by the requirements that:

- L_n^∞ is a random labeled matching of $[2n]$ for all $n \in \mathbb{N}$,
- the process $(\Psi(L_n^\infty))_{n\in\mathbb{N}_0}$ has the same distribution as $(W_n^\infty)_{n\in\mathbb{N}_0}$,
- the conditional distribution of L_n^∞ given $\Psi(L_n^\infty)$ is uniform on the set of labeled associated admissible matchings of $\Psi(L_n^\infty)$ for all $n \in \mathbb{N}_0$,

- the backward transition mechanism of $(L_n^\infty)_{n\in\mathbb{N}_0}$ is deterministic: in going from step $n+1$ to step n the pair $\{i,j\}$ labeled with $n+1$ is deleted and the increasing bijection from $[2(n+1)] \setminus \{i,j\}$ to $[2n]$ is applied to turn the resulting labeled matching of $[2(n+1)] \setminus \{i,j\}$ into a labeled matching of $[2n]$.

We can turn L_n^∞ into a word of length $2n$ in the alphabet $\bigcup_{k=1}^n \{a_k, b_k\}$ in which each letter appears exactly once as follows: place the letter a_p in position i if the block of L_n^∞ labeled p is of the form $\{i,j\}$ with $i<j$ and place the letter b_q in position ℓ if the block of L_n^∞ labeled q is of the form $\{k,\ell\}$ with $k<\ell$. This word defines a total order on $\bigcup_{k=1}^n \{a_k, b_k\}$ in the obvious way: $c \in \bigcup_{k=1}^n \{a_k, b_k\}$ precedes $d \in \bigcup_{k=1}^n \{a_k, b_k\}$ in the total order if the letter c comes before the letter d in the word. This total order is *paired*, by which we mean that a_r always precedes b_r. These paired total orders are consistent as n varies and hence define a paired total order on $\mathbb{I}_0 := \bigcup_{k\in\mathbb{N}} \{a_k, b_k\}$. This random paired total order on $\mathbb{I}_0$ is exchangeable in the obvious sense. Conversely, we can reverse this process, start with any exchangeable random paired total order on $\mathbb{I}_0$, and produce an infinite bridge. Moreover, this procedure is bijective: distinct infinite bridge distributions correspond to distinct exchangeable random paired total order distributions and *vice versa*. It therefore suffices to characterize the possible distributions of exchangeable random paired total orders.

A mixture of infinite bridges is an infinite bridge, so we actually want to characterize the infinite bridges that cannot be written as mixtures. The latter are the *extremal* infinite bridges and are the infinite bridges that have almost surely trivial tail σ-fields. Extremal infinite bridges correspond to exchangeable random paired total orders that are *ergodic* in the sense that events which are invariant under finite permutations of indices have probability zero or one. It therefore further suffices to characterize the possible distributions of ergodic exchangeable random paired total orders.

One way to produce an ergodic exchangeable random paired total order $\triangleleft$ on $\mathbb{I}_0$ is the following. Take a probability measure η on $\mathbb{R}^2$ that assigns all of its mass to the set $\{(s,t) \in \mathbb{R}^2 : s \le t\}$ and has diffuse marginals. Let $((S_n, T_n))_{n\in\mathbb{N}}$ be independent and identically distributed with common distribution η. Declare that

- $a_i \triangleleft a_j$ if $S_i < S_j$,
- $b_i \triangleleft b_j$ if $T_i < T_j$,
- $a_i \triangleleft b_j$ if $S_i < T_j$,
- $b_i \triangleleft a_j$ if $T_i < S_j$,
- $a_k \triangleleft b_k$ if $S_k = T_k$.

The exchangeability of $\triangleleft$ is clear. The ergodicity follows from the Hewitt–Savage zero–one law. Alternatively, observe first that we can encode $\triangleleft$ by a jointly exchangeable array $D = (D(i,j))_{i,j\in\mathbb{N}, i\neq j}$ that takes values in the 6-element set $\{IJIJ, IJJI, IIJJ, JIIJ, JIJI, JJII\}$

as follows: for $i, j \in \mathbb{N}$, $i \neq j$,

$$D(i,j) := \begin{cases} IJIJ, & \text{if } a_i \lhd a_j \lhd b_i \lhd b_j, \\ IJJI, & \text{if } a_i \lhd a_j \lhd b_j \lhd b_i, \\ IIJJ, & \text{if } a_i \lhd b_i \lhd a_j \lhd b_j, \\ JIIJ, & \text{if } a_j \lhd a_i \lhd b_i \lhd b_j, \\ JIJI, & \text{if } a_j \lhd a_i \lhd b_j \lhd b_i, \\ JJII, & \text{if } a_j \lhd b_j \lhd a_i \lhd b_i. \end{cases}$$

The ergodicity of $\lhd$ is equivalent to the ergodicity of D in the usual sense for such arrays and, by a result of Aldous (see Lemma 7.35 of [17]), this is equivalent to the array being *dissociated* – a condition which clearly holds here; that is, the ergodicity of $\lhd$ follows from the fact that if $H_1, \dots, H_s$ are disjoint finite subsets of $\mathbb{N}$, then the s subarrays of D consisting of entries indexed by the respective sets $\{(i,j) : i, j \in H_r, i \neq j\}$, $1 \leq r \leq s$, are independent.

One of our key results is that all ergodic exchangeable random paired total orders arise from the construction above for some measure η. Of course, η is not unique: if ν is the push-forward of η by $(s,t) \mapsto (g(s), g(t))$ for some function g such that $\frac{1}{2}(\eta(\cdot \times \mathbb{R}) + \eta(\mathbb{R} \times \cdot))$-a.e $u \in \mathbb{R}$ is a point of increase of g, then applying the above construction with ν instead of η gives an ergodic exchangeable random paired total order on $\mathbb{I}_0$ with the same distribution. Taking $g(u) = \frac{1}{2}(\eta((-\infty, u] \times \mathbb{R}) + \eta(\mathbb{R} \times (-\infty, u]))$, $u \in \mathbb{R}$, the push-forward ν of η is a measure on $\{(x,y) : 0 \leq x \leq y \leq 1\}$ that satisfies $\frac{1}{2}(\nu(\cdot \times [0,1]) + \nu([0,1] \times \cdot)) = \lambda$, where λ is Lebesgue measure on $[0,1]$. We obtain a bijective correspondence between probability measures μ on $\{(x,y) : 0 \leq x \leq y \leq 1\}$ with diffuse marginals and distributions of ergodic exchangeable random paired total orders (and hence between probability measures on $\{(x,y) : 0 \leq x \leq y \leq 1\}$ with diffuse marginals and infinite bridges with almost surely trivial tail σ-fields) if we impose the normalization that $\frac{1}{2}(\mu(\cdot \times \mathbb{R}) + \mu(\mathbb{R} \times \cdot))$ is the Lebesgue measure λ. Note that the normalization condition implies that μ has diffuse marginals.

The following theorem collects together the results we establish in the paper. We will describe the bijections asserted in the theorem explicitly as we proceed.

Theorem 2.1.4. *There are bijective correspondences between the following sets:*

(i) extremal elements of the Doob–Martin boundary associated with the transition probabilities of the Markov chain $(W_n)_{n \in \mathbb{N}_0}$ and the reference state state $\emptyset$,

(ii) extremal nonnegative harmonic functions normalized to take the value 1 at the reference state $\emptyset$,

(iii) probability measures μ concentrated on $\{(x,y) : 0 \leq x \leq y \leq 1\}$ such that $\frac{1}{2}(\mu(\cdot \times \mathbb{R}) + \mu(\mathbb{R} \times \cdot)) = \lambda$, the Lebesgue measure on $[0,1]$, and the labeled infinite bridge $(L_n^\infty)_{n \in \mathbb{N}_0}$ uniquely determined by μ satisfies that the conditional distribution of L_n^∞ given $\Psi(L_n^\infty)$ is uniform on the set of labeled associated admissible matchings of $\Psi(L_n^\infty)$ for all $n \in \mathbb{N}_0$.

(iv) the distributions of extremal infinite bridges for the Markov chain $(W_n)_{n\in\mathbb{N}_0}$*,*

(v) the distributions of ergodic exchangeable random paired total orders $\prec$ *on the set* $\mathbb{I}_0 = \bigcup_{n\in\mathbb{N}}\{a_n, b_n\}$ *such that the labeled infinite bridge* $(L_n^\infty)_{n\in\mathbb{N}_0}$ *uniquely determined by* $\prec$ *satisfies that the conditional distribution of* L_n^∞ *given* $\Psi(L_n^\infty)$ *is uniform on the set of labeled associated admissible matchings of* $\Psi(L_n^\infty)$ *for all* $n \in \mathbb{N}_0$*.*

Suitably composing the following bijections suffices to construct bijections between any of the sets (i)–(v) in Theorem 2.1.4. For each point ξ *in the Doob–Martin boundary the normalized nonnegative harmonic function* $K(\cdot,\xi)$ *is extremal and all extremal normalized nonnegative functions are of this form. A function* h *is a normalized nonnegative harmonic function if and only if* $h = h^\mu$ *for a unique probability measures* μ *concentrated on* $\{(x,y) \in [0,1]^2 : x < y\}$ *that has diffuse marginals. A random total order on* $\mathbb{I}_0$ *is an ergodic exchangeable random paired order if and only if it arises from the construction of Remark 2.5.2 for some unique probability measures* μ *concentrated on* $\{(x,y) \in [0,1]^2 : x < y\}$ *that has diffuse marginals. Lemma 2.4.6 describes a bijective correspondence between infinite bridges and exchangeable random paired total orders, and by Lemma 2.5.14 an infinite bridge is extremal if and only if the corresponding random order is ergodic.*

Example 2.1.5. The chain $(W_n)_{n\in\mathbb{N}_0}$ is itself an infinite bridge. The corresponding normalized nonnegative harmonic function is the constant function $h \equiv 1$. One way to construct $(W_n)_{n\in\mathbb{N}_0}$ is as follows. Let $(X_1,Y_1),(X_2,Y_2),\ldots$ be independent identically distributed random variables with common distribution $\mu(dx,dy) = 2\,dx\,dy$, $0 \le x \le y \le 1$; that is, μ is the distribution of a pair (X,Y) chosen uniformly at random from $\{(x,y) : 0 \le x \le y \le 1\}$. For $n \in \mathbb{N}_0$, let $Z_{n,1},\ldots,Z_{n,2n}$ be $X_1,Y_1,\ldots,X_n,Y_n$ listed in increasing order. Define the word $W_n = W_{n,1}\ldots W_{n,2n}$ by setting $W_{n,k} = a$ (resp. $W_{n,k} = b$) if $Z_{n,k} \in \{X_1,\ldots,X_n\}$ (resp. $Z_{n,k} \in \{Y_1,\ldots,Y_n\}$). It follows from the Hewitt–Savage zero–one law that the tail σ-field of the chain $(W_n)_{n\in\mathbb{N}_0}$ is almost surely trivial and hence this infinite bridge is extremal. In particular, the chain converges almost surely to an extremal element of the Doob–Martin boundary. Denote by $(L_n)_{n\in\mathbb{N}_0}$ the process of labeled associated admissible matchings corresponding to $(W_n)_{n\in\mathbb{N}_0}$; that is, $(L_n)_{n\in\mathbb{N}_0}$ is the particular instance of the object denoted above by $(L_n^\infty)_{n\in\mathbb{N}_0}$ in the special case where $(W_n^\infty)_{n\in\mathbb{N}_0}$ is $(W_n)_{n\in\mathbb{N}_0}$. We may construct L_n by matching $i,j \in [2n]$ and labeling this block with k when $X_k = Z_{n,i}$ and $Y_k = Z_{n,j}$. The corresponding ergodic exchangeable random paired total order $\prec$ on $\mathbb{I}_0 = \bigcup_{k\in\mathbb{N}}\{a_k,b_k\}$ is defined by declaring that

- $a_i \prec a_j$ if $X_i < Y_j$,
- $b_i \prec b_j$ if $Y_i < Y_j$,
- $a_i \prec b_j$ if $X_i < Y_j$,
- $b_i \prec a_j$ if $Y_i < X_j$.

2.2 Background on the Doob-Martin compactification

We review some parts of Doob-Martin compactification theory that are essential to understand further discussion. While the primary source for this material is [11], helpful reviews can be found in [18, Chapter 10], [29, Chapter 7], [24, Chapter 7], [25].

Let $(X_n)_{n\in\mathbb{N}_0}$ be a Markov chain with countable state space S and transition probabilities P. Assume that $(X_n)_{n\in\mathbb{N}_0}$ is transient, i.e. $\mathbb{P}^x\{\exists n > 0 \,|\, X_n = x\} < 1$ for $x \in S$. Suppose also that there is a reference state $o \in S$ with $F(o,x) > 0$ for $x \in S$, where

$$\begin{aligned} F(x,y) &:= \sum_{n=0}^{\infty} \mathbb{P}^x\{X_n = y, X_j \neq y \text{ for } 0 \leq j \leq n-1\} \\ &= \mathbb{P}^x\{X \text{ hits y}\}. \end{aligned}$$

Definition 2.2.1. The *Doob-Martin kernel* K with reference state o is

$$K(x,y) := \frac{F(x,y)}{F(o,y)} = \frac{G(x,y)}{G(o,y)} \quad \text{for } x,y \in S,$$

where $G(x,y) := \sum_{n=0}^{\infty} P^n(x,y)$ is the *Green kernel*. The kernel K is well-defined since the denominators in the definition are always positive by the assumption on the distinguished state o.

Two important properties of the Doob-Martin kernel are that $K(o,y) = 1$ for $y \in S$ and that, for $x,y \in S$ with $x \neq y$,

$$\begin{aligned} \sum_{z\in S} P(x,z)K(z,y) &= K(x,y), \\ \sum_{z\in S} P(y,z)K(z,y) &> K(y,y). \end{aligned}$$

It follows that the functions $K(\cdot,y)$, $y \in S$, are distinct superharmonic functions.

We recall the following general fact (see, for example, [29, Theorem 7.13]).

Theorem 2.2.2. *If $\mathcal{F}$ is a countable family of bounded real-valued functions on a countable set S, then there exists a unique (up to homeomorphic equivalence) compactification $\hat{S} = \hat{S_{\mathcal{F}}}$ of S such that*

(a) every function $f \in \mathcal{F}$ extends uniquely to a continuous function on $\hat{S}$ (which we still denote by f), and

(b) the family $\mathcal{F}$ separates the boundary points: if $\xi, \eta \in \hat{S}\backslash S$ are distinct, then there is $f \in \mathcal{F}$ with $f(\xi) \neq f(\eta)$.

The compactification $\hat{S}$ is metrizable.

Definition 2.2.3. Let $\mathcal{F}$ be the family of functions $\{K(x,\cdot) : x \in S\}$. Because $K(x,y) = \frac{F(x,y)}{F(o,y)} \leq \frac{F(x,y)}{F(o,x)F(x,y)} = \frac{1}{F(o,x)}$ for $x, y \in S$, the set $\{K(x,\cdot) : x \in S\}$ is a countable family of bounded functions on S. Therefore, $\mathcal{F}$ induces a compactification $\hat{S}$ of S via Theorem 2.2.2. This compactification is called the *Doob-Martin compactification* of S (with respect to P). The *Doob-Martin boundary* of S (with respect to P), written as ∂S, is the set $\hat{S} \backslash S$.

By construction, the Doob–Martin kernel extends to a continuous nonnegative function $K : S \times \hat{S} \mapsto \mathbb{R}$ called the *extended Doob–Martin kernel*. In general, $K(\cdot,\xi)$ is a nonnegative superharmonic function for $\xi \in \partial S$. However, if P has finite range, that is, the set $\{y \in S : P(x,y) > 0\}$ is finite for all $x \in S$, then for all $\xi \in \partial S$ the nonnegative function $K(\cdot,\xi)$ is harmonic.

Given any nonnegative harmonic function h, there exists a finite Borel measure ν on ∂S such that

$$h(x) = \int_{\partial S} K(x,\xi)\,\nu(d\xi).$$

In general, the measure ν is not unique, but it is unique if it is required to be concentrated on the subset of ∂S consisting of points ξ such that $K(\cdot,\xi)$ is an *extremal* nonnegative harmonic function (a nonnegative harmonic function is extremal if it cannot be written as a nontrivial convex combination of nonnegative harmonic functions).

There is a random variable X_∞ taking its values in ∂S such that, for $x \in S$,

$$\lim_{n\to\infty} X_n = X_\infty \quad \mathbb{P}^x\text{-almost surely}$$

in the topology of $\hat{S}$. The tail σ-field of $(X_n)_{n\in\mathbb{N}_0}$ coincides with the σ-field generated by X_∞, $\mathbb{P}^x$-almost surely.

For the rest of this section, we consider the special case where there exists a partition of $\bigsqcup_{n=0}^{\infty} S_n$ of S with $S_0 = \{o\}$ and S_n finite, $n \in \mathbb{N}_0$, such that $P(x,y) > 0$ only if $x \in S_n$ and $y \in S_{n+1}$ for some $n \in \mathbb{N}_0$. The chain $(U_n)_{n\in\mathbb{N}_0}$ (resp. $(W_n)_{n\in\mathbb{N}_0}$) is of this form with $S = \mathbb{W}$ (resp. $S = \mathbb{B}$), $S_n = \mathbb{W}_n$ (resp. $S_n = \mathbb{B}_n$), and $o = \emptyset$. For z in S, there exists a unique index $N_z \in \mathbb{N}_0$ such that $z \in S_{N_z}$. If the chain $(X_n)_{n\in\mathbb{N}_0}$ starts from the reference state o, then N_z is the only time when it can take the value of z with positive probability. We have $F(x,y) = G(x,y)$, $x, y \in S$, in this case.

Definition 2.2.4. Let $(X_0^y, \ldots, X_{N_y}^y)$ be the Markov chain $(X_0, \ldots, X_{N_y})$ under $\mathbb{P}^o$ conditioned on the event $\{X_{N_y} = y\}$. This conditioned process is called a *bridge*. It is a Markov chain with transition probabilities

$$\begin{aligned}\mathbb{P}\{X^y_{n+1} = z \,|\, X^y_n = x\} &= \frac{\mathbb{P}^o\{X_{n+1} = z, X_n = x, X_{N_y} = y\}}{\mathbb{P}^o\{X_n = x, X_{N_y} = y\}} \\ &= \frac{F(o,x)P(x,z)F(z,y)}{F(o,x)F(x,y)} \\ &= \frac{P(x,z)F(z,y)/F(o,y)}{F(x,y)/F(o,y)} \\ &= \frac{P(x,z)K(z,y)}{K(x,y)} \quad \text{for } x,z \in S.\end{aligned}$$

The backward transition probabilities of a bridge are calculated as follows:

$$\begin{aligned}\mathbb{P}\{X^y_n = x \,|\, X^y_{n+1} = z\} &= \frac{\mathbb{P}^o\{X_{n+1} = z, X_n = x, X_{N_y} = y\}}{\mathbb{P}^o\{X_{n+1} = z, X_{N_y} = y\}} \\ &= \frac{F(o,x)P(x,z)F(z,y)}{F(o,z)F(z,y)} \\ &= \frac{F(o,x)P(x,z)}{F(o,z)} \\ &= \mathbb{P}\{X_n = x \,|\, X_{n+1} = z\} \quad \text{for } x,z \in S.\end{aligned}$$

These backward transition probabilities are thus independent of the choice of y in S, and hence are common to all bridges.

Writing Q for the matrix of backward transition probabilities we have for $x \in S_m$ and $y \in S_{m+n}$

$$Q^n(y,x) = \frac{F(o,x)F(x,y)}{F(o,y)} = F(o,x)K(x,y). \tag{2.1}$$

Definition 2.2.5. An *infinite bridge* for $(X_n)_{n\in\mathbb{N}_0}$ is a Markov chain $(X^\infty_n)_{n\in\mathbb{N}_0}$ with $X^\infty_0 = o$ and the same backward transition probabilities as $(X_n)_{n\in\mathbb{N}_0}$.

An examination of the transitions probabilities stated in the Definition 2.2.4 shows that a sequence $(y_n)_{n\in\mathbb{N}_0}$ with $N_{y_n} \to \infty$ as $n \to \infty$ converges in the Doob-Martin topology if and only if for all $k \in \mathbb{N}_0$ the initial segment $(X^{y_n}_0, \dots, X^{y_n}_k)$ of the associated bridges converges in distribution as $n \to \infty$ (cf. [15]). The limiting distributions of the initial segments are consistent and form the distributions of initial segments of an infinite bridge $(X^\infty_n)_{n\in\mathbb{N}_0}$. Put $h(x) := \lim_{n\to\infty} K(x, y_n)$, $x \in S$. Then h is a nonnegative harmonic function and the transition probabilities of the infinite bridge can be written as

$$\mathbb{P}\{X^\infty_{n+1} = z \,|\, X^\infty_n = x\} = \frac{P(x,z)h(z)}{h(x)} \quad \text{for } x,z \in S^h,$$

where $S^h := \{x \in S : h(x) > 0\}$. That is, $(X^\infty_n)_{n\in\mathbb{N}_0}$ is the *Doob harmonic transform* of $(X_n)_{n\in\mathbb{N}_0}$ determined by the harmonic function h.

If $(X_n^\infty)_{n\in\mathbb{N}_0}$ is any infinite bridge, then

$$\begin{aligned}\mathbb{P}\{X_{n+1}^\infty = z \mid X_n^\infty = x\} &= \frac{\mathbb{P}^o\{X^\infty \text{ hits } z\}\mathbb{P}\{X_n^\infty = x \mid X_{n+1}^\infty = z\}}{\mathbb{P}^o\{X^\infty \text{ hits } x\}} \\ &= \frac{h(z)P(x,z)}{h(x)} \quad \text{for } z, x \in S^h,\end{aligned}$$

where h is the harmonic function defined by $h(x) := \frac{\mathbb{P}^o\{X^\infty \text{ hits } x\}}{F(o,x)}$ and S^h denotes its support $\{x \in S : h(x) > 0\}$. Conversely, any Doob harmonic transform of $(X_n)_{\mathbb{N}_0}$ is an infinite bridge. There is thus a bijective correspondence between infinite bridges and nonnegative harmonic functions normalized to have $h(o) = 1$.

An infinite bridge is *extremal* if its distribution cannot be written as a nontrivial mixture of distributions of infinite bridges. An infinite bridge is extremal if and only if its tail σ-field is almost surely trivial. Also, an infinite bridge is extremal if and only if the corresponding nonnegative harmonic function cannot be written as a nontrivial convex combination of nonnegative harmonic functions with $h(o) = 1$. (in which case we recall from [29] that we say that the nonnegative harmonic function is extremal or minimal). A necessary condition for a nonnegative harmonic function with $h(0) = 1$ to be extremal is that is of the form $K(\cdot, \xi)$ for some $\xi \in \partial S$, but in general this condition is not sufficient. Consequently, a necessary condition for an infinite bridge to be extremal is that it can be constructed as the limit in distribution of bridges $(X_0^{y_n}, \dots, X_{N_{y_n}}^{y_n})$ for some sequence $(y_n)_{n\in\mathbb{N}}$ with $N_{y_n} \to \infty$ as $n \to \infty$.

Example 2.2.6. The classical two color Pólya's urn scheme is a prime example of how Doob-Martin topology delineates the limiting behavior of a Markov chain. We refer to [3] for rigorous proofs and more details.

Imagine an urn containing a black ball and a white ball. Each time, another ball is added to the urn as follows:

1. a ball is drawn uniformly at random from the urn and its color is observed,
2. it is then replaced in the urn with an additional ball of the same color,
3. and the selection process is repeated.

Write B_n and H_n for the number of black balls and the number of white balls, respectively, in the urn after n additional balls are added, $n \in \mathbb{N}_0$. By definition, $(B_0, H_0) = (1, 1)$. It is easy to see that $(B_n)_{n\in\mathbb{N}_0}$ and $(H_n)_{n\in\mathbb{N}_0}$ are non-decreasing sequences and converge almost surely to infinity. In other words, with the usual discrete topology on the state space $\mathbb{N} \times \mathbb{N}$, the Markov chain $(B_n, H_n)_{n\in\mathbb{N}_0}$ converges to infinity as $n \to \infty$ in the sense of one point compactification of the state space.

However, we can embed the state space $\mathbb{N} \times \mathbb{N}$ into a richer topological space and delineate the asymptotic behaviors of the chain. The Doob-Martin compactification is such a topological space where the chain converges to a limit almost surely.

It is obvious that the initial position $(1,1)$ is a reference state of the Markov chain $(B_n, H_n)_{n\in\mathbb{N}_0}$. The transition probabilities and Green kernels for the chain $(B_n, H_n)_{n\in\mathbb{N}_0}$ are described as follows. For $b, h, \tilde{b}, \tilde{h} \in \mathbb{N}$ with $b \le \tilde{b}$, $h \le \tilde{h}$,

$$\begin{aligned} G((b,h),(\tilde{b},\tilde{h})) &= P^{(\tilde{b}+\tilde{h}-b-h)}((b,h),(\tilde{b},\tilde{h})) \\ &= \binom{\tilde{b}+\tilde{h}-b-h}{\tilde{b}-b} \frac{b\cdots(\tilde{b}-1)h\cdots(\tilde{h}-1)}{(b+h)\cdots(\tilde{b}+\tilde{h}-1)}, \end{aligned}$$

$$\begin{aligned} K((b,h),(\tilde{b},\tilde{h})) &= \frac{G((b,h),(\tilde{b},\tilde{h}))}{G((1,1),(\tilde{b},\tilde{h}))} \\ &= \frac{\frac{b\cdots(\tilde{b}-1)h\cdots(\tilde{h}-1)}{(b+h)\cdots(\tilde{b}+\tilde{h}-1)}\binom{\tilde{b}+\tilde{h}-(b+h)}{\tilde{b}-b}}{\frac{(\tilde{b}-1)!(\tilde{h}-1)!}{(\tilde{b}+\tilde{h}-1)!}\binom{\tilde{b}+\tilde{h}-2}{\tilde{b}-1}} \\ &= \frac{\frac{b\cdots(\tilde{b}-1)h\cdots(\tilde{h}-1)}{(b+h)\cdots(\tilde{b}+\tilde{h}-1)}\frac{(\tilde{b}+\tilde{h}-(b+h))!}{(\tilde{b}-b)!(\tilde{h}-b)!}}{\frac{(\tilde{b}-1)!(\tilde{h}-1)!}{(\tilde{b}+\tilde{h}-1)!}\frac{(\tilde{b}+\tilde{h}-2)!}{(\tilde{b}-1)!(\tilde{h}-1)!}} \\ &= \frac{(b+h-1)!}{(b-1)!(h-1)!}\frac{(\tilde{b}-b+1)\cdots(\tilde{b}-1)(\tilde{h}-h+1)\cdots(\tilde{h}-1)}{(\tilde{b}+\tilde{h}-b-h+1)\cdots(\tilde{b}+\tilde{h}-2)}. \end{aligned}$$

It follows that a sequence $(b_n, h_n)_{n\in\mathbb{N}_0}$ in $\mathbb{N}\times\mathbb{N}$ with $b_n \to \infty, h_n \to \infty$ as $n \to \infty$ converges in the Doob-Martin topology if and only if $\frac{b_n}{b_n+h_n}$ converges as $n \to \infty$. More precisely, if $\lim_{n\to\infty} \frac{b_n}{b_n+h_n} = c$ for $0 \le c \le 1$, then for $(b,h) \in \mathbb{N}\times\mathbb{N}$

$$\begin{aligned} &\lim_{n\to\infty} K((b,h),(b_n,h_n)) \\ &= \lim_{n\to\infty} \frac{(b+h-1)!}{(b-1)!(h-1)!}\frac{(b_n-b+1)}{b_n+h_n-b-h+1}\cdots\frac{b_n-1}{b_n+h_n-h-1}\frac{h_n-h+1}{b_n+h_n-b-h}\cdots\frac{h_n-1}{b_n+h_n-2} \\ &= \frac{(b+h-1)!}{(b-1)!(h-1)!}c^{b-1}(1-c)^{h-1}. \end{aligned}$$

Because $\left(\frac{B_n}{B_n+H_n}\right)_{n\in\mathbb{N}_0}$ is a bounded martingale, we have that $\left(\frac{B_n}{B_n+H_n}\right)_{n\in\mathbb{N}_0}$ converges almost surely and therefore the Markov chain $(B_n, H_n)_{n\in\mathbb{N}_0}$ converges almost surely in the Doob-Martin topology. Moreover, an induction establishes that $\frac{B_n}{B_n+H_n}$ is uniformly distributed over the set $\{\frac{1}{n+2}, \frac{2}{n+2}, \dots, \frac{n+1}{n+2}\}$, $n \in \mathbb{N}_0$. Consequently, $\left(\frac{B_n}{B_n+H_n}\right)_{n\in\mathbb{N}_0}$ converges in distribution to a $U[0,1]$ random variable and the Doob-Martin boundary can be identified with the unit interval $[0,1]$.

2.3 Transition probabilities and Labeled Infinite bridges

Transition probabilities and Labeled infinite bridges for $(U_n)_{n\in\mathbb{N}_0}$

Definition 2.3.1. For $n \in \mathbb{N}_0$ write $\mathbb{W}_n$ for the set of words from the alphabet $\{a, b\}$ that have n letters a and n letters b and put $\mathbb{W} := \bigsqcup_{n\in\mathbb{N}_0} \mathbb{W}_n$.

By definition, the Markov chain $(U_n)_{n\in\mathbb{N}_0}$ has state space $\mathbb{W}$ and one-step transition probabilities

$$\mathbb{P}\{U_{n+1} = w \,|\, U_n = v\} = \frac{M(v, w)}{(2n + 2)(2n + 1)}$$

for $v \in \mathbb{W}_n$ and $w \in \mathbb{W}_{n+1}$, where $M(v, w)$ is the number of ways to write $w = v_1xv_2yv_3$ in such a way that $\{x, y\} = \{a, b\}$ and v_1, v_2, v_3 are (possibly empty) words such that $v = v_1v_2v_3$. That is, $M(v, w)$ is the number of times that v appears inside w as a *sub-word*. (We recall that, in general, a word $c_1 \cdots c_p$ is a sub-word of a word $d_1 \cdots d_q$ if there is a map $f : [p] \to [q]$ such that $f(i) < f(j)$ for $1 \le i < j \le p$ and $d_{f(k)} = c_k$ for $1 \le k \le p$.)

In order to write down multi-step transition probabilities for the Markov chain $(U_n)_{n\in\mathbb{N}_0}$, it is convenient to introduce the following standard notation (see, for example, [20]).

Definition 2.3.2. Given two words w and v drawn from some finite alphabet, write $\binom{w}{v}$ for the number of times that v appears as a sub-word of w.

Example 2.3.3. For example, $\binom{abbaba}{bba} = 4$ because *bba* appears inside *abbaba* as a sub-word four times:

a**bba**ba a**bb**ab**a** a**b**ba**ba** ab**b**a**ba**.

Remark 2.3.4. Note that if our alphabet has only one letter, then $\binom{w}{v}$ is just the usual binomial coefficient $\binom{|w|}{|v|}$, where we use the notation $|u|$ for the length of the word u.

For a general finite alphabet $\mathcal{A}$, $\binom{w}{v}$ is uniquely determined by the following three properties, where we write $\mathcal{A}^*$ for the set of finite words with letters drawn from the alphabet $\mathcal{A}$ (see [20, Proposition 6.3.3]):

- $\binom{w}{\emptyset} = 1$ for all $w \in \mathcal{A}^*$, where $\emptyset$ is the empty word,
- $\binom{w}{v} = 0$ for all $v, w \in \mathcal{A}^*$ with $|w| < |v|$,
- $\binom{wy}{vx} = \binom{w}{vx} + \delta_{x,y}\binom{w}{v}$, for all $v, w \in \mathcal{A}^*$ and $x, y \in \mathcal{A}$, where δ is the usual Kronecker delta.

The counting involved in determining $\binom{w}{v}$ for general $v, w \in \mathcal{A}^*$ is handled by the following result from [7]. Define an infinite matrix $\mathcal{P}$ with entries indexed by $\mathcal{A}^*$ by setting the (v, w) entry to be $\binom{w}{v}$. If the row and column indices are ordered so that they are nondecreasing in word length, then $\mathcal{P}$ is an upper triangular matrix with 1 in every position on the diagonal.

Define another infinite matrix $\mathcal{H}$ indexed by $\mathcal{A}^*$ by setting the (v, w) entry to be $\binom{w}{v}$ if $|w| = |v| + 1$ and 0 otherwise. With the same ordering of the indices as for $\mathcal{P}$, the matrix $\mathcal{H}$ is upper triangular with 0 in every position on the diagonal. The matrix exponential $\exp(\mathcal{H})$ is well-defined and is equal to $\mathcal{P}$.

Using the above notation, we can express the transition probabilities of $(U_n)_{n\in\mathbb{N}_0}$ as follows.

Lemma 2.3.5. *For words $v \in \mathbb{W}_m$ and $w \in \mathbb{W}_{m+n}$*

$$\mathbb{P}\{U_{m+n} = w \,|\, U_m = v\} = \binom{w}{v} \frac{n!n!}{(2m+1)(2m+2)\cdots(2(m+n))}.$$

Proof. We proceed by induction. The result is certainly true when $n = 1$. Supposing it is true for some value of n, in order to show it is true for $n + 1$, we need to show that for $u \in \mathbb{W}_m$ and $w \in \mathbb{W}_{m+n+1}$ we have

$$\sum_{v\in\mathbb{W}_{m+1}} \binom{v}{u} \frac{1}{(2m+1)(2m+2)} \binom{w}{v} \frac{n!n!}{(2m+3)(2m+4)\cdots(2(m+n+1))}$$
$$= \binom{w}{u} \frac{(n+1)!(n+1)!}{(2m+1)(2m+2)\cdots(2(m+n+1))},$$

or, equivalently, that

$$\sum_{v\in\mathbb{W}_{m+1}} \binom{v}{u}\binom{w}{v} = \binom{w}{u}(n+1)^2.$$

This, however, is clear. The lefthand side counts the number of words $v \in \mathbb{W}_{m+1}$ such that u is subword of v and v is a subword of w. Any such v and its embedding in w arises by taking an embedding of u in w and then specifying which of the remaining $n + 1$ letters a in w and which of the remaining $n + 1$ letters b in w are used to build the word with its particular embedding, and this is what the righthand side counts. □

Corollary 2.3.6. *The Doob–Martin kernel of $(U_n)_{n\in\mathbb{N}_0}$ with distinguished state the empty word is, for $v \in \mathbb{W}_m$ and $w \in \mathbb{W}_{m+n}$,*

$$K(v, w) = \binom{w}{v} \frac{\binom{2m}{m}}{\binom{m+n}{m}^2}.$$

Proof. We have

$$\begin{aligned}
&K(v,w)\\
&\quad= \frac{\mathbb{P}\{U_{m+n} = w \mid U_m = v\}}{\mathbb{P}\{U_{m+n} = w \mid U_0 = \emptyset\}}\\
&\quad= \frac{\binom{w}{v}\frac{n!n!}{(2m+1)(2m+2)\cdots(2(m+n))}}{\binom{w}{\emptyset}\frac{(m+n)!(m+n)!}{(2(m+n))!}}\\
&\quad= \binom{w}{v}\frac{n!n!(2(m+n))!}{(m+n)!(m+n)!(2m+1)(2m+2)\cdots(2(m+n))}\\
&\quad= \binom{w}{v}\frac{\binom{2m}{m}}{\binom{m+n}{n}\binom{m+n}{n}}.
\end{aligned}$$

□

Remark 2.3.7. Up to the factor $\binom{2m}{m}$, the Doob–Martin kernel $K(v,w)$ is the probability that if we select m of the letters a and m of the letters b uniformly at random from w and list these letters in the same relative order that they appear in w the resulting word is v. Therefore, a sequence $(w_k)_{k\in\mathbb{N}}$ in $\mathbb{W}$ with $N(w_k) \to \infty$ as $k \to \infty$ converges in the Doob–Martin topology if and only if for every $m \in \mathbb{N}$ the sequence of random words in $\mathbb{W}_m$ obtained by selecting m letters a and m letters b from w_k (and maintaining their relative order) converges in distribution as $k \to \infty$.

Definition 2.3.8. For $w \in \mathbb{W}_k$, $k \in \mathbb{N}_0$, let $(U_0^w, \ldots, U_k^w)$ be the bridge for $(U_n)_{n\in\mathbb{N}_0}$ from the empty word to w.

Theorem 2.3.9. *The backward transition dynamics for all bridges for $(U_n)_{n\in\mathbb{N}_0}$ from the empty word are the same and consist of removing at each step one letter a and one letter b uniformly at random.*

Proof. Consider the bridge for $(U_n)_{n\in\mathbb{N}_0}$ from the empty word to $w \in \mathbb{W}_k$.

For $0 \le m \le k-1$, $v \in \mathbb{W}_{m+1}$, and $u \in \mathbb{W}_m$ we have

$$\begin{aligned}
&\mathbb{P}\{U_m^w = u \,|\, U_{m+1}^w = v\} \\
&= \frac{\mathbb{P}\{U_m = u,\, U_{m+1} = v \,|\, U_k = w\}}{\mathbb{P}\{U_{m+1} = v \,|\, U_k = w\}} \\
&= \frac{\mathbb{P}\{U_m = u,\, U_{m+1} = v,\, U_k = w\}}{\mathbb{P}\{U_{m+1} = v,\, U_k = w\}} \\
&= \frac{\mathbb{P}\{U_m = u\}\mathbb{P}\{U_{m+1} = v \,|\, U_m = u\}\mathbb{P}\{U_k = w \,|U_{m+1} = v\}}{\mathbb{P}\{U_{m+1} = v\}\mathbb{P}\{U_k = w \,|U_{m+1} = v\}} \\
&= \binom{v}{u} \frac{\frac{1}{(2m+1)(2m+2)} \times \frac{m!m!}{(2m)!}}{\frac{(m+1)!(m+1)!}{(2m+2)!}} \\
&= \frac{\binom{v}{u}}{(m+1)^2}.
\end{aligned}$$

In order to go backward from the word v of length $2(m+1)$ to the word u of length $2m$, we have to remove one a and one b. There are $\binom{v}{u}$ pairs of a and b such that the removal of the pair from v results in u, and there are a total of $(m+1)^2$ pairs of a and b in v, and so the result follows from the calculation above. □

Suppose that $(y_n)_{n\in\mathbb{N}}$ is a sequence of words in $\mathbb{W} := \bigsqcup_{n\in\mathbb{N}_0} \mathbb{W}_n$ that converges in the Doob–Martin topology and is such that $N(y_n) \to \infty$ as $n \to \infty$. Recall that $(U_0^{y_n}, \dots, U_{N(y_n)}^{y_n})$, $n \in \mathbb{N}$, is the associated bridge that starts from the empty word and is tied to being in state y_n at time $N(y_n)$. The finite dimensional distributions of $(U_0^{y_n}, \dots, U_{N(y_n)}^{y_n})$ converge as $n \to \infty$. Thus, there exists a process $(U_n^\infty)_{n\in\mathbb{N}_0}$ such that for every $k \in \mathbb{N}_0$ the random $(k+1)$-tuple $(U_0^{y_n}, \dots, U_k^{y_n})$ converges in distribution to $(U_0^\infty, \dots, U_k^\infty)$.

The forward evolution dynamics of the Markov chain $(U_n^\infty)_{n\in\mathbb{N}}$ depend on the sequence $(y_n)_{n\in\mathbb{N}}$, whereas from Section 2.2 and Theorem 2.3.9 the backward evolution is Markovian and doesn't depend on the sequence $(y_n)_{n\in\mathbb{N}}$; given U_{k+1}^∞, the word U_k^∞ is obtained by removing one letter a and one letter b uniformly at random from U_{k+1}^∞.

For each $n \in \mathbb{N}_0$ the distribution of U_n^∞ defines the distribution of a random element $\tilde{U}_{n,n}^\infty$ of the set $\tilde{\mathbb{W}}_n$ of words of length $2n$ drawn from the alphabet $\{a_1, b_1, \dots, a_n, b_n\}$ with each letter appearing once by assigning the labels $[n]$ uniformly at random to the letters a and to the letters b. More precisely, for $U_n^\infty = c_1 \dots c_{2n}$, let $A_n := \{i \in [n] : c_i = a\}$ and $B_n := \{j \in [n] : c_j = b\}$, let $\Sigma : A_n \to [n]$ and $T : B_n \to [n]$ be random bijections that are conditionally independent and uniformly distributed given U_n^∞, and define $\tilde{U}_{n,n}^\infty := \tilde{c}_1 \dots \tilde{c}_{2n}$ by

$$\tilde{c}_k := \begin{cases} a_{\Sigma(k)}, & k \in A_n, \\ b_{T(k)}, & k \in B_n. \end{cases}$$

For $0 \le p \le n$, define $\tilde{U}_{n,p}^\infty$ to be the word obtained by deleting $\{a_{p+1}, b_{p+1}, \dots, a_n, b_n\}$ from $\tilde{U}_{n,n}^\infty$. Observe that if $0 \le p \le m \wedge n$, then $\tilde{U}_{m,p}^\infty$, $\tilde{U}_{n,p}^\infty$ and $\tilde{U}_{p,p}^\infty$ have the same distribution.

Moreover, if for $0 \le p \le n$ we let $U_{n,p}^\infty$ be the result of removing the labels from $\tilde{U}_{n,p}^\infty$ (that is, $U_{n,p}^\infty$ is the element of $\mathbb{W}_p$ obtained by replacing the letters a_k, $1 \le k \le p$, by the letter a and the letters b_k, $1 \le k \le p$, by b), then $(U_{n,0}^\infty, \dots, U_{n,n}^\infty)$ has the same distribution as $(U_0^\infty, \dots, U_n^\infty)$.

By Kolmogorov's consistency theorem, there is a process $(\tilde{U}_n^\infty)_{n\in\mathbb{N}_0}$ such that $(\tilde{U}_0^\infty, \dots, \tilde{U}_m^\infty)$ has the same distribution as $(\tilde{U}_{n,0}^\infty, \dots, \tilde{U}_{n,m}^\infty)$ for any $m \le n$ and the result of removing the labels from $(\tilde{U}_n^\infty)_{n\in\mathbb{N}_0}$ has the same distribution as $(U_n^\infty)_{n\in\mathbb{N}_0}$. By the transfer theorem [16, Theorem 6.10], we may even suppose that $(\tilde{U}_n^\infty)_{n\in\mathbb{N}_0}$ is defined on an extension of the probability space on which $(U_n^\infty)_{n\in\mathbb{N}_0}$ is defined in such a way that $(U_n^\infty)_{n\in\mathbb{N}_0}$ is the result of removing the labels from $(\tilde{U}_n^\infty)_{n\in\mathbb{N}_0}$.

Transition probabilities and labeled infinite bridges for the Markov chain $(W_n)_{n\in\mathbb{N}_0}$

By definition, $(W_n)_{n\in\mathbb{N}_0}$ is a Markov chain with state space $\mathbb{B}$ and its transition probabilities are as follows: for $v = v_1 \dots v_{2n} \in \mathbb{B}_n$ and $w = w_1 \dots w_{2(n+1)} \in \mathbb{B}_{n+1}$,

$$\mathbb{P}\{W_{n+1} = w \,|\, W_n = v\} = \frac{R(v,w)}{\binom{2(n+1)}{2}}, \tag{2.2}$$

where $R(v,w)$ is the number of pairs (i,j) with $1 \le i < j \le 2(n+1)$ such that $w = v_1 \dots v_{i-1} a v_i \dots v_{j-2} b v_{j-1} \dots v_{2n}$. For example, $\mathbb{P}\{W_2 = aabb\} = \frac{2}{3}$ and $\mathbb{P}\{W_2 = abab\} = \frac{1}{3}$ corresponding to the respective sets of possibilities $\{a\mathbf{a}b\mathbf{b}, \mathbf{a}ab\mathbf{b}, a\mathbf{a}\mathbf{b}b, \mathbf{a}a\mathbf{b}b\}$ and $\{ab\mathbf{ab}, \mathbf{ab}ab\}$ for placing the new $\mathbf{a}$ and $\mathbf{b}$. Similar enumerations establish that the probabilities of the various transitions from states in $\mathbb{B}_2$ to states in $\mathbb{B}_3$ are

$$\begin{array}{c} \\ aabb \\ abab \end{array} \begin{array}{c} \begin{array}{ccccc} aaabbb & aababb & aabbab & abaabb & ababab \end{array} \\ \left(\begin{array}{ccccc} 9/15 & 4/15 & 1/15 & 1/15 & 0 \\ 0 & 4/15 & 4/15 & 4/15 & 3/15 \end{array}\right) \end{array}$$

and that the marginal probability distribution of W_3 is given by $\mathbb{P}\{W_3 = aaabbb\} = \frac{6}{15}$, $\mathbb{P}\{W_3 = aababb\} = \frac{4}{15}$, $\mathbb{P}\{W_3 = aabbab\} = \frac{2}{15}$, $\mathbb{P}\{W_3 = abaabb\} = \frac{2}{15}$, and $\mathbb{P}\{W_3 = ababab\} = \frac{1}{15}$.

Definition 2.3.10. A *matching* of $[2n]$ is a partition of $[2n]$ into subsets of size 2. Given a word $w = w_1 \dots w_{2n} \in \mathbb{B}_n$, we say that a matching $\mathcal{M}$ of $[2n]$ is an *associated admissible matching* of the word w if for every block $\{i,j\}$ of $\mathcal{M}$ with $i < j$ we have $w_i = a$ and $w_j = b$. For example, if $n = 3$ and $w = aababb$, then the associated admissible matchings of the word w are $\{\{1,3\},\{2,5\},\{4,6\}\}$, $\{\{1,3\},\{2,6\},\{4,5\}\}$, $\{\{1,5\},\{2,3\},\{4,6\}\}$, and $\{\{1,6\},\{2,3\},\{4,5\}\}$.

Any matching $\mathcal{M}$ of $[2n]$ defines a word $w \in \mathbb{B}_n$ for which $\mathcal{M}$ is an associated admissible matching: if $\{i,j\}$ is a block of $\mathcal{M}$ with $i < j$, then place a letter a in the position i and a letter

b in the position j. Denote this word by $\Phi(\mathcal{M})$; for example, $\Phi(\{\{1,3\},\{2,5\},\{4,6\}\}) = aababb$. Let $\Lambda(w) := \#\{\mathcal{M} : \Phi(\mathcal{M}) = w\}$ be the number of associated admissible matchings of a word w; for example, $\Lambda(aababb) = 4$.

Definition 2.3.11. For a word $w = w_1 \dots w_{2n} \in \mathbb{B}_n$, define its *height function*

$$H(t) := \#\{1 \le i \le t : w_i = a\} - \#\{1 \le j \le t : w_j = b\}, \quad 1 \le t \le 2n.$$

Lemma 2.3.12. *Given a word $w = w_1 \dots w_{2n} \in \mathbb{B}_n$, the number of associated admissible matchings is*

$$\Lambda(w) = \prod_{1 \le k \le 2n,\, w_k = a} H(k).$$

Proof. For $1 \le k \le n$, write i_k (respectively, j_k) for the position of the k^{th} letter a (respectively, letter b) in the word w; that is, $\#\{1 \le r \le i_k : w_r = a\} = k$, $\#\{1 \le s \le j_k : w_s = b\} = k$, $w_{i_k} = a$, and $w_{j_k} = b$. Then, for $1 \le t \le 2n$,

$$H(t) = \#\{1 \le p \le n : i_p \le t\} - \#\{1 \le q \le n : i_q \le t\}.$$

Note that $w_{i_n} = a$ and $w_k = b$ for $i_n + 1 \le k \le 2n$, and therefore i_n must be matched with one of $\{i_n + 1, \dots, 2n\}$. Observe that the cardinality of this set is

$$\begin{aligned} n - \#\{1 \le j \le i_n : w_j = b\} &= \#\{1 \le i \le i_n : w_i = a\} - \#\{1 \le j \le i_n : w_j = b\} \\ &= H(i_n). \end{aligned}$$

Removing the letter a in position i_n and the letter b in the position with which i_n is matched produces the same word $v = v_1 \dots v_{2n-2}$ regardless of which index is matched with i_n. The number of associated admissible matchings of the word w is therefore the product of $H(i_n)$ and the number of associated admissible matchings of the word v.

Let the function G be defined from v in the same manner that H was defined from w. Then $i_{n-1} \in [2n-2]$ is the position of the last letter a in w and, for $1 \le t \le i_{n-1}$,

$$\#\{1 \le i \le t : v_i = a\} = \#\{1 \le i \le t : w_i = a\}$$

and

$$\#\{1 \le j \le t : v_j = b\} = \#\{1 \le i \le t : w_j = b\}$$

so that $G(t) = H(t)$ for $1 \le t \le i_{n-1}$. An induction then establishes that the number of associated admissible matchings of w is

$$\prod_{p=1}^{n} H(i_p),$$

as required. □

Using the same randomness that went into the construction of $(W_n)_{n\in\mathbb{N}_0}$, it is possible to generate a Markov chain $(M_n)_{n\in\mathbb{N}_0}$ such that M_n is an associated admissible matching of W_n for n in $\mathbb{N}$. Recall that $I_{n+1} < J_{n+1}$ is the pair of indices uniformly picked from $\{(i,j) : 1 \le i < j \le 2n+2\}$ when constructing W_{n+1} from W_n. In going from M_n to M_{n+1}, we match I_{n+1} and J_{n+1}; more precisely, we make $\{I_{n+1}, J_{n+1}\}$ a block of the partition M_{n+1} and define the remaining blocks by taking each block $\{k,\ell\}$ of M_n with $k<\ell$ and transforming it into the block $\{p,q\}$ of M_{n+1}, where

- $p = k$ and $q = \ell$ if $k < \ell < I_{n+1} < J_{n+1}$,
- $p = k$ and $q = \ell+1$ if $k < I_{n+1} < \ell+1 < J_{n+1}$,
- $p = k+1$ and $q = \ell+1$ if $I_{n+1} < k+1 < \ell+1 < J_{n+1}$,
- $p = k+1$ and $q = \ell+2$ if $I_{n+1} < k+1 < J_{n+1} < \ell+2$,
- $p = k+2$ and $q = \ell+2$ if $I_{n+1} < J_{n+1} < k+2 < \ell+2$.

Proposition 2.3.13. *(i) For each $n \in \mathbb{N}_0$, the random matching M_n is uniformly distributed over the $\frac{1}{n!}\prod_{k=1}^{n}\binom{2k}{2} = (2n-1)(2n-3)\cdots 3\cdot 1 = (2n-1)!!$ matchings of $[2n]$.*

(ii) For each $n \in \mathbb{N}_0$ and $w \in \mathbb{B}_n$, the marginal distribution of W_n is

$$\mathbb{P}\{W_n = w\} = \mathbb{P}\{\Phi(M_n) = w\} = \frac{\Lambda(w)}{(2n-1)!!}.$$

(iii) For each $n \in \mathbb{N}_0$ and $w \in \mathbb{B}_n$, the conditional distribution of M_n given $W_n = w$ is uniform on the $\Lambda(w)$ associated admissible matchings of w.

(iv) For $v \in \mathbb{B}_n$ and $w \in \mathbb{B}_{n+1}$,

$$\mathbb{P}\{W_n = v \,|\, W_{n+1} = w\} = \frac{1}{n+1} R(v,w) \frac{\Lambda(v)}{\Lambda(w)}.$$

(v) The conditional distribution of M_n given $M_{n+1} = \mathcal{M}$ is the distribution of the random partition of $[2n]$ that is produced by first removing a block $\{i,j\}$ uniformly at random from the $n+1$ blocks of $\mathcal{M}$ to produce a matching of the set $[2n+2] \setminus \{i,j\}$ and then applying the unique increasing bijection from $[2n+2] \setminus \{i,j\}$ to $[2n]$ to turn this matching into a matching of $[2n]$.

(vi) Consider $w \in \mathbb{B}_{n+1}$ and construct a random matching R of $[2n]$ as follows. Let S be a uniform random admissible associated matching for w and let R be such that the conditional distribution of R given $S = \mathcal{S}$ coincides with the conditional distribution of M_n given $M_{n+1} = \mathcal{S}$ described in (v). Then, the distribution of R is the same as

the conditional distribution of M_n given $\{W_{n+1} = w\}$. Thus, the distribution of the random word $\Phi(R)$ coincides with the conditional distribution of W_n given $\{W_{n+1} = w\}$. Moreover, given $\Phi(R)$ the conditional distribution of the random matching R is uniform on the set of associated admissible matchings of $\Phi(R)$.

Proof. The proof of part (i) is essentially immediate from the definition of $(M_n)_{n\in\mathbb{N}_0}$, but we present a proof by induction for the sake of completeness. The result is certainly true when $n = 1$. Suppose that the result holds for some $n \in \mathbb{N}$. Let $\mathcal{M}$ be a matching of $[2n + 2]$. If $\{i, j\}$ is a block of $\mathcal{M}$, let $\mathcal{M}_{ij}$ be the matching of $[2n]$ that is obtained by removing the block $\{i, j\}$ from $\mathcal{M}$ to produce a matching of the set $[2n + 2] \setminus \{i, j\}$ and then applying the unique increasing bijection from $[2n+2]\setminus\{i, j\}$ to $[2n]$ to turn this matching into a matching of $[2n]$. By the inductive hypothesis,

$$\begin{aligned}
\mathbb{P}\{M_{n+1} = \mathcal{M}\} &= \sum_{1\le i<j\le 2n+2} \mathbb{P}\{M_{n+1} = \mathcal{M},\, (I_{n+1}, J_{n+1}) = (i, j)\} \\
&= \sum_{\{i,j\}\in\mathcal{M},\, i<j} \mathbb{P}\{M_n = \mathcal{M}_{ij},\, (I_{n+1}, J_{n+1}) = (i, j)\} \\
&= (n + 1)\frac{1}{(2n-1)!!}\frac{2}{(2n + 2)(2n + 1)} \\
&= \frac{1}{(2(n + 1) - 1)!!},
\end{aligned}$$

as required.

Part (ii) follows readily from part (i), and part (iii) can be derived from parts (i) and (ii) : for $\mathcal{M}$ such that $\Phi(\mathcal{M}) = w$,

$$\begin{aligned}
\mathbb{P}\{M_n = \mathcal{M} \,|\, W_n = w\} &= \frac{\mathbb{P}\{M_n = \mathcal{M}, W_n = w\}}{\mathbb{P}\{W_n = w\}} \\
&= \frac{\mathbb{P}\{M_n = \mathcal{M}\}}{\mathbb{P}\{W_n = w\}} \\
&= \frac{1}{(2n-1)!!}\Big/\frac{\Lambda(w)}{(2n-1)!!} \\
&= \frac{1}{\Lambda(w)}.
\end{aligned}$$

For part (iv) we have

$$\mathbb{P}\{W_n = v \,|\, W_{n+1} = w\} = \frac{\mathbb{P}\{W_n = v\}\mathbb{P}\{W_{n+1} = w \,|\, W_n = v\}}{\mathbb{P}\{W_{n+1} = w\}},$$

where

$$\mathbb{P}\{W_{n+1} = w \,|\, W_n = v\} = \frac{2R(v, w)}{(2n + 2)(2n + 1)},$$

$$\mathbb{P}\{W_{n+1} = w\} = \frac{\Lambda(w)}{(2n+1)!!},$$

and

$$\mathbb{P}\{W_n = v\} = \frac{\Lambda(v)}{(2n-1)!!}.$$

Thus the backward transition probability can be rewritten as follows:

$$\begin{aligned}\mathbb{P}\{W_n = v \,|\, W_{n+1} = w\} &= \frac{\Lambda(v)}{(2n-1)!!}\frac{2R(v,w)}{(2n+2)(2n+1)}\Big/\frac{\Lambda(w)}{(2n+1)!!} \\ &= \frac{1}{n+1}R(v,w)\frac{\Lambda(v)}{\Lambda(w)}.\end{aligned}$$

The proof of part (v) is very similar to that of part (i). Let $\mathcal{M}$ be a matching of $[2n]$ and $\mathcal{N}$ be a matching of $[2n+2]$. As above, if $\{i,j\}$ is a block of $\mathcal{N}$, write $\mathcal{N}_{ij}$ for the matching of $[2n]$ that is obtained by removing this block from $\mathcal{N}$ and then applying the unique increasing bijection from $[2n+2] \setminus \{i,j\}$ to $[2n]$ to turn this matching of $[2n+2] \setminus \{i,j\}$ into a matching of $[2n]$. Then

$$\begin{aligned}&\mathbb{P}\{M_n = \mathcal{M} \,|\, M_{n+1} = \mathcal{N}\} \\ &\quad= \frac{\mathbb{P}\{M_n = \mathcal{M}\}\mathbb{P}\{M_{n+1} = \mathcal{N} \,|\, M_n = \mathcal{M}\}}{\mathbb{P}\{M_{n+1} = \mathcal{N}\}} \\ &\quad= \frac{\mathbb{P}\{M_n = \mathcal{M}\}}{\mathbb{P}\{M_{n+1} = \mathcal{N}\}} \times \sum_{1 \le i < j \le 2n+2} \mathbb{1}\{\{i,j\} \in \mathcal{N},\, \mathcal{M} = \mathcal{N}_{ij}\}\frac{2}{(2n+2)(2n+1)} \\ &\quad= \frac{(2n+1)!!}{(2n-1)!!} \cdot \#\{\{i,j\} \in \mathcal{N} : \mathcal{M} = \mathcal{N}_{ij}\} \cdot \frac{2}{(2n+2)(2n+1)} \\ &\quad= \frac{\#\{\{i,j\} \in \mathcal{N} : \mathcal{M} = \mathcal{N}_{ij}\}}{n+1},\end{aligned}$$

as required.

For part (vi) we have that

$$\begin{aligned}&\mathbb{P}\{M_n = \mathcal{R} \,|\, W_{n+1} = w\} \\ &\quad= \sum_{\Phi(\mathcal{S})=w} \mathbb{P}\{M_n = \mathcal{R} \,|\, W_{n+1} = w,\, M_{n+1} = \mathcal{S}\}\, \mathbb{P}\{M_{n+1} = \mathcal{S} \,|\, W_{n+1} = w\} \\ &\quad= \sum_{\Phi(\mathcal{S})=w} \mathbb{P}\{M_n = \mathcal{R} \,|\, M_{n+1} = \mathcal{S}\}\, \mathbb{P}\{M_{n+1} = \mathcal{S} \,|\, W_{n+1} = w\} \\ &\quad= \sum_{S} \mathbb{P}\{R = \mathcal{R} \,|\, S = \mathcal{S}\}\, \mathbb{P}\{S = \mathcal{S}\} \\ &\quad= \mathbb{P}\{R = \mathcal{R}\}.\end{aligned}$$

Thus, the distribution of R is, as claimed, the same as the conditional distribution of M_n given $\{W_{n+1} = w\}$.

Moreover, because M_n given W_n is uniformly distributed on the set of associated admissible matchings for W_n by part (iii) and by construction W_{n+1} is conditionally independent of M_n given W_n, we have

$$\begin{aligned}
&\mathbb{P}\{R = \mathcal{R} \,|\, \Phi(R) = v\} \\
&\quad = \mathbb{P}\{M_n = \mathcal{R} \,|\, W_n = v,\, W_{n+1} = w\} \\
&\quad = \frac{\mathbb{P}\{M_n = \mathcal{R},\, W_n = v,\, W_{n+1} = w\}}{\mathbb{P}\{W_n = v,\, W_{n+1} = w\}} \\
&\quad = \frac{\mathbb{P}\{W_n = v\}\mathbb{P}\{M_n = \mathcal{R} \,|\, W_n = v\}\mathbb{P}\{W_{n+1} = w \,|\, M_n = \mathcal{R},\, W_n = v\}}{\mathbb{P}\{W_n = v,\, W_{n+1} = w\}} \\
&\quad = \frac{1}{\Lambda(v)} \frac{\mathbb{P}\{W_n = v\}\mathbb{P}\{W_{n+1} = w \,|\, W_n = v\}}{\mathbb{P}\{W_n = v,\, W_{n+1} = w\}} \\
&\quad = \frac{1}{\Lambda(v)}.
\end{aligned}$$

This shows that the conditional distribution of R given $\Phi(R)$ is indeed uniformly distributed on the set of associated admissible matchings of $\Phi(R)$. □

Remark 2.3.14. The above proof of Proposition 2.3.13 contains a proof of part (iv) of the result. However, it is informative to observe that part (iv) is, as follows, a consequence of part (vi) of the result. Part (vi) says that $\mathbb{P}\{W_n = v \,|\, W_{n+1} = w\} = \mathbb{P}\{\Phi(R) = v\}$ and

$$\begin{aligned}
\mathbb{P}\{\Phi(R) = v\} &= \sum_{\Phi(\mathcal{R}) = v} \mathbb{P}\{R = \mathcal{R}\} \\
&= \sum_{\Phi(\mathcal{R}) = v, \Phi(\mathcal{S}) = w} \mathbb{P}\{R = \mathcal{R},\, S = \mathcal{S}\}.
\end{aligned}$$

In the sum we only need to consider pairs $(\mathcal{R}, \mathcal{S})$ such that $\Phi(\mathcal{S}) = w$, there is a block $\{i, j\}$ of the partition $\mathcal{S}$ such that $1 \le i < j \le 2n+2$ and the removal of the the letters a and b from the positions in w given by the indices i, j produces v, and the partition $\mathcal{R}$ is obtained by removal of the block from $\mathcal{S}$ followed by the usual transformation of indices that transforms a matching of $[2n + 2] \setminus \{i, j\}$ into a matching of $[2n]$. The number of pairs of indices that can be matched is $R(v, w)$. The number of associated admissible matchings of w in which two particular indices i, j are matched is $\Lambda(v)$, because if we remove the matched pair $\{i, j\}$ of indices we obtain a matching of v after the usual transformation of indices. For each such associated admissible matching $\mathcal{S}$ of w there is probability $\frac{1}{\Lambda(w)}$ that S will take the value $\mathcal{S}$. Lastly, for such a pair $(\mathcal{R}, \mathcal{S})$, if S takes the value $\mathcal{S}$, then for each of the $n + 1$ blocks of $\mathcal{S}$ there is probability $\frac{1}{n+1}$ that this block will be removed so that R takes the value $\mathcal{R}$. Thus,

$$\mathbb{P}\{\Phi(R) = v\} = R(v, w)\Lambda(v) \frac{1}{\Lambda(w)} \frac{1}{n + 1},$$

as claimed.

Definition 2.3.15. A *labeled matching* of $[2n]$ is a matching in which the n blocks are labeled with distinct elements of $[n]$.

Given a labeled matching $\mathcal{L}$ of $[2n]$, let $\Psi(\mathcal{L}) \in \mathbb{B}_n$ be the corresponding word produced by ignoring the labels and placing a letter a in position i and a letter b in position j for each block $\{i, j\}$ of $\mathcal{L}$ with $i < j$. Using the same randomness that was used to construct $(W_n)_{n\in\mathbb{N}_0}$ and $(M_n)_{n\in\mathbb{N}_0}$, it is possible to build a Markov chain $(L_n)_{n\in\mathbb{N}_0}$ such that L_n is a labeled matching of $[2n]$ for $n \in \mathbb{N}$: the blocks of L_n are the same as the blocks of M_n and in going from L_n to L_{n+1} the newly created block $\{I_{n+1}, J_{n+1}\}$ is labeled with $n + 1$ whilst the blocks that arise by transforming blocks already present in M_n keep their labels. Thus, $\Psi(L_n) = W_n$.

The following result is immediate from Proposition 2.3.13.

Corollary 2.3.16. *(i) For each $n \in \mathbb{N}$, the random matching L_n is uniformly distributed over the $n!(2n-1)!!$ labeled matchings of $[2n]$.*

(ii) For each $n \in \mathbb{N}$, the conditional distribution of L_n given M_n is uniform over the $n!$ labelings of M_n and the conditional distribution of L_n given W_n is uniform over the $n!\Lambda(W_n)$ labeled associated admissible matchings of W_n.

(iii) The labeled matching L_n is obtained from the labeled matching L_{n+1} by removing the block labeled $n + 1$ and if this block contains the indices $\{i, j\}$ applying the unique increasing bijection from $[2n + 2] \setminus \{i, j\}$ to $[2n]$ to turn this labeled matching of $[2n + 2]\setminus\{i, j\}$ into a labeled matching of $[2n]$ (in particular, the backward transition dynamics of $(L_n)_{n\in\mathbb{N}_0}$ are deterministic).

(iv) Consider $w \in \mathbb{B}_{n+1}$ and construct a random labeled matching C of $[2n]$ as follows. Let D be a uniform random labeled associated admissible matching for w and C be such that the conditional distribution of C given $\{D = \mathcal{D}\}$ coincides with the conditional distribution of L_n given $\{L_{n+1} = \mathcal{D}\}$ described in (iii). Then, the distribution of C is the same as the conditional distribution of L_n given $\{W_{n+1} = w\}$. Thus, the distribution of the random word $\Psi(C)$ coincides with the conditional distribution of W_n given $\{W_{n+1} = w\}$. Moreover, given $\Psi(C)$, the the conditional distribution of the random labeled matching C is uniform on the set of labeled associated admissible matchings of $\Psi(C)$.

Recall that an infinite bridge for the Markov chain $(W_n)_{n\in\mathbb{N}_0}$ is a Markov chain $(W_n^\infty)_{n\in\mathbb{N}_0}$ such that $W_0^\infty = \emptyset$ and $\mathbb{P}\{W_n^\infty = v \,|\, W_{n+1}^\infty = w\} = \mathbb{P}\{W_n = v \,|\, W_{n+1} = w\}$ for all $v \in \mathbb{B}_n$, $w \in \mathbb{B}_{n+1}$, and $n \in \mathbb{N}_0$.

Corollary 2.3.17. *Suppose that $(W_n^\infty)_{n\in\mathbb{N}_0}$ is an infinite bridge for $(W_n)_{n\in\mathbb{N}_0}$. Then there exists a Markov process $(L_n^\infty)_{n\in\mathbb{N}_0}$ with distribution uniquely specified by the requirements that:*

- *L_n^∞ is a random labeled matching of $[2n]$ for all $n \in \mathbb{N}$,*

- *the process* $(\Psi(L_n^\infty))_{n\in\mathbb{N}_0}$ *has the same distribution as* $(W_n^\infty)_{n\in\mathbb{N}_0}$*,*
- *the conditional distribution of* L_n^∞ *given* $\Psi(L_n^\infty)$ *is uniform on the set of labeled associated admissible matchings of* $\Psi(L_n^\infty)$ *for all* $n \in \mathbb{N}_0$*,*
- *the conditional distribution of* L_n^∞ *given* $\{L_{n+1}^\infty = \mathcal{L}\}$ *is the same as the conditional distribution of* L_n *given* $\{L_{n+1} = \mathcal{L}\}$*.*

Proof. Fix $n \in \mathbb{N}_0$ and $w \in \mathbb{B}_n$. Define a Markov process $(L_0^{n,w}, \dots, L_n^{n,w})$ by requiring that

- $L_n^{n,w}$ is a random labeled matching of $[2n]$ that is uniformly distributed over the labeled associated admissible matchings for w,
- the conditional probabilities for $L_m^{n,w}$ given $L_{m+1}^{n,w}$ are the same as the conditional probabilities for L_m given L_{m+1} for $1 \le m \le n-1$. In other words, $L_m^{n,w}$ is obtained from $L_{m+1}^{n,w}$ by removing the block labeled $m+1$ and transforming the remaining blocks without changing their labels by applying the unique increasing bijection to $[2n]$.

It follows from Corollary 2.3.16 that $(\Psi(L_0^{n,w}), \dots, \Psi(L_n^{n,w}))$ has the same distribution as $(W_0, \dots, W_n)$ conditional on the event $\{W_n = w\}$. Moreover, the latter conditional distribution is the same as the distribution of $(W_0^\infty, \dots, W_n^\infty)$ conditional on the event $\{W_n^\infty = w\}$.

We can therefore construct a Markov process $(L_0^n, \dots, L_n^n)$ with a distribution that is uniquely specified by the requirements that

- $(\Psi(L_0^n), \dots, \Psi(L_n^n))$ has the same distribution as $(W_0^\infty, \dots, W_n^\infty)$,
- conditional probabilities for L_m^n given L_{m+1}^n are the same as the conditional probabilities for L_m given L_{m+1} for $1 \le m \le n-1$,
- the conditional distribution of L_m^n given $\Psi(L_m^n)$ is uniform on the set of labeled associated admissible matchings of $\Psi(L_m^n)$ for $1 \le m \le n$.

It is clear that $(L_0^{n'}, \dots, L_{n'}^{n'})$ has the same distribution as $(L_0^{n''}, \dots, L_{n'}^{n''})$ when $n' < n''$ and hence, using Kolmogorov's extension theorem, we can construct a process $(L_n^\infty)_{n\in\mathbb{N}_0}$ such that $(L_0^\infty, \dots, L_n^\infty)$ has the same distribution as $(L_0^n, \dots, L_n^n)$ for any $n \in \mathbb{N}_0$. □

2.4 The exchangeable random total order associated with a labeled infinite bridge

The exchangeable random total order associated with an infinite bridge $(\widetilde{U}_n^\infty)_{n\in\mathbb{N}_0}$

A state of a labeled infinite bridge is a word of length $2n$ from the alphabet $\{a_1, b_1, \dots, a_n, b_n\}$ in which each letter appears once. Another way to think of such an object is as a total order on the set $\bigcup_{k=1}^n \{a_k, b_k\}$. Because the labeled infinite bridge evolves by slotting in the letters a_{n+1} and b_{n+1} at the $(n+1)^{\text{th}}$ step while leaving the relative positions of $\{a_1, b_1, \ldots, a_n, b_n\}$ unchanged, these successive total orders are consistent: the total order on $\{a_1, b_1, \ldots, a_n, b_n\}$ given by the state of the infinite bridge at step n is the same as the total order obtained by taking the state of the infinite bridge at step $n+1$ (a total order on $\{a_1, b_1, \ldots, a_n, b_n, a_{n+1}, b_{n+1}\}$) and looking at the corresponding induced total order on $\{a_1, b_1, \ldots, a_n, b_n\}$.

This projective structure means that we can associate any path of a labeled infinite bridge with a unique total order on $\mathbb{I}_0 := \bigcup_{n\in\mathbb{N}} \{a_n, b_n\}$ such that the induced total order on $\{a_1, b_1, \ldots, a_n, b_n\}$ coincides with the state of the labeled infinite bridge at step n.

We now introduce some general notions about random total orders.

Definition 2.4.1. A *random total order* $\prec$ on $\mathbb{I}_0$ is a map from the underlying probability space to the collection of total orders on $\mathbb{I}_0$ such that the indicator $\mathbb{1}\{x \prec y\}$ is a random variable for every $x, y \in \mathbb{I}_0$. A random total order $\prec$ is *exchangeable* if for every $n \in \mathbb{N}$ the induced total order $\prec^n$ on $\bigcup_{k=1}^n \{a_k, b_k\}$ has the same distribution as the random total order $\prec^n_{\sigma,\tau}$ for any permutations σ, τ of $\{1, 2, \ldots, n\}$, where $\prec^n_{\sigma,\tau}$ is defined as follows:

- $a_{\sigma(i)} \prec^n_{\sigma,\tau} b_{\tau(j)}$ iff $a_i \prec^n b_j$
- $b_{\tau(i)} \prec^n_{\sigma,\tau} a_{\sigma(j)}$ iff $b_i \prec^n a_j$
- $a_{\sigma(i)} \prec^n_{\sigma,\tau} a_{\sigma(j)}$ iff $a_i \prec^n a_j$
- $b_{\tau(i)} \prec^n_{\sigma,\tau} b_{\tau(j)}$ iff $b_i \prec^n b_j$.

Remark 2.4.2. The distribution of a random total order $\prec$ is determined by the joint distribution of the random variables $\{\mathbb{1}\{x \prec y\} : x, y \in \bigcup_{k=1}^n \{a_k, b_k\}\}$ for arbitrary $n \in \mathbb{N}$ (of course, since $\mathbb{1}\{x \prec x\} = 0$ and $\mathbb{1}\{x \prec y\} = 1 - \mathbb{1}\{y \prec x\}$ for $x \neq y$, it suffices to take an appropriate subset of size $\binom{2n}{2}$ of these $(2n)^2$ random variables for each $n \in \mathbb{N}$).

Remark 2.4.3. If $\prec$ is an exchangeable random total order, then the induced random total orders $\prec^n$, $n \in \mathbb{N}$, are consistent in the sense that if we take the random total order $\prec^{n+1}$ on $\bigcup_{k=1}^{n+1} \{a_k, b_k\}$ and remove $\{a_{n+1}, b_{n+1}\}$, then the induced random total order on $\bigcup_{k=1}^n \{a_k, b_k\}$ is $\prec^n$.

Furthermore, $\bigcup_{k=1}^{n+1} \{a_k\}$ and $\bigcup_{k=1}^{n+1} \{b_k\}$ under the total orders induced by $\prec^{n+1}$ are each order isomorphic to $[n+1]$ with the usual total order, and, moreover, if we let I and J be the images of a_{n+1} and b_{n+1}, then I and J are independent and uniformly distributed on $[n+1]$.

Conversely, suppose for each $n \in \mathbb{N}$ that there is a random total order $\prec^n$ on $\bigcup_{k=1}^n \{a_k, b_k\}$, these random total orders have the property that $\prec^n$ has the same distribution as $\prec^n_{\sigma,\tau}$ for any permutations σ, τ of $[n]$ for all $n \in \mathbb{N}$, and these total orders are consistent. Then there is an exchangeable random order $\prec$ on $\mathbb{I}_0$ such that $\prec^n$ is the corresponding induced total order on $\bigcup_{k=1}^n \{a_k, b_k\}$.

In terms of these general notions, if we let $\prec^n$, $n \in \mathbb{N}$, be the random total order on $\bigcup_{k=1}^n \{a_k, b_k\}$ corresponding to $\tilde{U}_n^\infty$, then these total orders are consistent and there is an exchangeable random total order $\prec$ on $\mathbb{I}_0$ such that the restriction of $\prec$ to $\bigcup_{k=1}^n \{a_k, b_k\}$ is $\prec^n$.

The exchangeable random paired total order associated with an infinite bridge $(\widetilde{W}_n^\infty)_{n\in\mathbb{N}_0}$

Definition 2.4.4. A *random total order* $\prec$ on $\mathbb{I}_0 := \bigcup_{n\in\mathbb{N}} \{a_n, b_n\}$ is a map from the underlying probability space to the collection of total orders on $\mathbb{I}_0$ such that $\mathbb{1}\{x \prec y\}$ is a random variable for every $x, y \in \mathbb{I}_0$. The distribution of $\prec$ is determined by the joint distribution of the random variables $\{\mathbb{1}\{x \prec y\} : x, y \in \bigcup_{k=1}^n \{a_k, b_k\}\}$ for arbitrary $n \in \mathbb{N}$. A random total order $\prec$ on $\mathbb{I}_0$ is *paired* if $a_n \prec b_n$ for all $n \in \mathbb{N}$. A random paired total order $\prec$ on $\mathbb{I}_0$ is *exchangeable* if for every $n \in \mathbb{N}$ the induced random total order $\prec^n$ on $\bigcup_{k=1}^n \{a_k, b_k\}$ has the same distribution as the random total order $\prec_\sigma^n$ for any permutation σ of $[n] := \{1, 2, \ldots, n\}$, where $\prec_\sigma^n$ is defined as follows:

- $a_{\sigma(i)} \prec_\sigma^n b_{\sigma(j)}$ iff $a_i \prec^n b_j$,
- $b_{\sigma(i)} \prec_\sigma^n a_{\sigma(j)}$ iff $b_i \prec^n a_j$,
- $a_{\sigma(i)} \prec_\sigma^n a_{\sigma(j)}$ iff $a_i \prec^n a_j$,
- $b_{\sigma(i)} \prec_\sigma^n b_{\sigma(j)}$ iff $b_i \prec^n b_j$.

Remark 2.4.5. If $\prec$ is an exchangeable random total order, then the induced random total orders $\prec^n$, $n \in \mathbb{N}$, are consistent in the sense that if we take the random total order $\prec^{n+1}$ on $\bigcup_{k=1}^{n+1} \{a_k, b_k\}$ and remove $\{a_{n+1}, b_{n+1}\}$, then the induced random total order on $\bigcup_{k=1}^n \{a_k, b_k\}$ is $\prec^n$.

Conversely, suppose that for each $n \in \mathbb{N}$ there is a random total order $\lhd^n$ on $\bigcup_{k=1}^n \{a_k, b_k\}$. Suppose that these random total orders are paired in the obvious sense and have the property that $\lhd^n$ has the same distribution as $\lhd_\sigma^n$ for any permutation σ of $[n]$ for all $n \in \mathbb{N}$. Suppose further that these paired total orders are consistent. Then, by Kolmogorov's extension theorem there is an exchangeable random paired total order $\lhd$ on $\mathbb{I}_0$ such that $\lhd^n$ is the corresponding induced total order on $\bigcup_{k=1}^n \{a_k, b_k\}$.

We can turn L_n^∞ (from Corollary 2.3.17) into a word of length $2n$ in the alphabet $\bigcup_{k=1}^n \{a_k, b_k\}$ in which each letter appears exactly once as follows: place the letter a_p in position i if the block of L_n^∞ labeled p is of the form $\{i, j\}$ with $i < j$ and place the letter b_q in position ℓ if the block of L_n^∞ labeled q is of the form $\{k, \ell\}$ with $k < \ell$. This is equivalent to replacing a letter a (respectively, b) in the word $\Psi(L_n^\infty)$ appearing in position i (respectively, ℓ) with the letter a_p (respectively, b_q) if the block of L_n^∞ containing i (respectively, ℓ) is labeled p (respectively, q). The resulting word in the alphabet $\bigcup_{k=1}^n \{a_k, b_k\}$ then defines a paired total order on that alphabet in the obvious way: $c \in \bigcup_{k=1}^n \{a_k, b_k\}$ precedes

$d \in \bigcup_{k=1}^{n}\{a_k, b_k\}$ in the total order if the letter c comes before the letter d in the word. These paired total orders are consistent as n varies and hence define a paired total order on $\mathbb{I}_0$. The following result defines this paired total order a little more formally, records the fact that it is exchangeable, and lays out how to go in the reverse direction and produce an infinite bridge from an exchangeable random total order. We omit the straightforward proof.

Lemma 2.4.6. *Suppose that $(W_n^\infty)_{n\in\mathbb{N}_0}$ is an infinite bridge for $(W_n)_{n\in\mathbb{N}_0}$ and $(L_n^\infty)_{n\in\mathbb{N}_0}$ is the corresponding process of labeled associated admissible matchings guaranteed by Corollary 2.3.17. Define a random paired total order $\prec$ on $\mathbb{I}_0$ by declaring for $p, q \in \mathbb{N}$ that*

- *$a_p \prec a_q$ if for any $n \in \mathbb{N}$ with $1 \le p, q \le n$ the block labeled p (respectively, q) in L_n^∞ is of the form $\{i, j\}$ with $i < j$ (respectively, $\{k, \ell\}$ with $k < \ell$) where $i < k$,*
- *$b_p \prec b_q$ if for any $n \in \mathbb{N}$ with $1 \le p, q \le n$ the block labeled p (respectively, q) in L_n^∞ is of the form $\{i, j\}$ with $i < j$ (respectively, $\{k, \ell\}$ with $k < \ell$) where $j < \ell$,*
- *$a_p \prec b_q$ if for any $n \in \mathbb{N}$ with $1 \le p, q \le n$ the block labeled p (respectively, q) in L_n^∞ is of the form $\{i, j\}$ with $i < j$ (respectively, $\{k, \ell\}$ with $k < \ell$) where $i < \ell$,*
- *$b_p \prec a_q$ if for any $n \in \mathbb{N}$ with $1 \le p, q \le n$ the block labeled p (respectively, q) in L_n^∞ is of the form $\{i, j\}$ with $i < j$ (respectively, $\{k, \ell\}$ with $k < \ell$) where $j < k$.*

Then the random paired total order $\prec$ is exchangeable.

Conversely, suppose that $\prec$ is an exchangeable random total order on $\mathbb{I}_0$. Define a process of labeled matchings $(L_n^\infty)_{n\in\mathbb{N}_0}$ as follows: if $c_1, \ldots, c_{2n}$ is a listing of $\bigcup_{k=1}^{n}\{a_k, b_k\}$ with $c_1 \prec \cdots \prec c_{2n}$, then make $\{i, j\} \subset [2n]$ a block of L_n^∞ labeled with $k \in [n]$ whenever $c_i = a_k$ and $c_j = b_k$. Then $(W_n^\infty)_{n\in\mathbb{N}_0} := (\Psi(L_n^\infty))_{n\in\mathbb{N}_0}$ is an infinite bridge and $(L_n^\infty)_{n\in\mathbb{N}_0}$ is the corresponding process of labeled associated admissible matchings guaranteed by Corollary 2.3.17 if and only if the conditional distribution of L_n^∞ given $\Psi(L_n^\infty)$ is uniform on the set of labeled associated admissible matchings of $\Psi(L_n^\infty)$ for all $n \in \mathbb{N}_0$.

2.5 Characterization of exchangeable random total orders

Characterization of exchangeable random total orders associated with infinite bridges $(\widetilde{U}_n)_{n\in\mathbb{N}_0}$

The results of the previous sections indicate that if we want to understand the Doob–Martin compactification for $(U_n)_{n\in\mathbb{N}_0}$ we need to understand infinite bridges for $(U_n)_{n\in\mathbb{N}_0}$, and this boils down to understanding exchangeable random total orders on $\mathbb{I}_0$.

A mixture of two exchangeable random total orders is also an exchangeable random total order, so we are interested in exchangeable random total orders $\prec$ that are extremal in the sense that their distributions cannot be written as a nontrivial mixture of the distributions of two other exchangeable random total orders. This is equivalent to requiring that if A is a measurable subset of the space of total orders on $\mathbb{I}_0$ with the property that $\prec \in A$ if and only if $\prec^{\sigma,\tau} \in A$ for all finite permutations σ, τ, then $\mathbb{P}\{\prec \in A\} \in \{0,1\}$. We say that an exchangeable random total order with this property is *ergodic*.

The following result can be established using essentially the same argument as in Proposition 5.19 of [13], and we omit the details.

Lemma 2.5.1. *The tail σ-field of an infinite bridge $(U_n^\infty)_{n\in\mathbb{N}_0}$ is almost surely trivial if and only if the exchangeable random total order induced by the corresponding labeled infinite bridge $(\tilde{U}_n^\infty)_{n\in\mathbb{N}_0}$ is ergodic.*

Remark 2.5.2. There is one obvious way to produce an ergodic exchangeable random total order. Let ζ and η be two diffuse probability measures on $\mathbb{R}$. Let $(V_n)_{n\in\mathbb{N}}$ be i.i.d. with common distribution ζ, let $(W_n)_{n\in\mathbb{N}}$ be i.i.d. with common distribution η, and suppose that these two sequences are independent. The total order $\prec$ on $\mathbb{I}_0$ defined by declaring that

- $a_i \prec a_j$ if $V_i < V_j$,
- $b_i \prec b_j$ if $W_i < W_j$,
- $a_i \prec b_j$ if $V_i < W_j$,
- $b_i \prec a_j$ if $W_i < V_j$,

is exchangeable and ergodic; exchangeability is obvious and ergodicity is immediate from the Hewitt–Savage zero–one law applied to the i.i.d. sequence $((V_n, W_n))_{n\in\mathbb{N}}$.

We will show that all ergodic exchangeable random total orders arise this way. Note that many pairs of probability measures can give rise to random total orders with the same distribution: replacing ζ and η by their push-forwards by some common strictly increasing function does not change the distribution of the resulting random total order.

Definition 2.5.3. Given an exchangeable random total order $\prec$ on $\mathbb{I}_0$, define $d : \mathbb{I}_0 \times \mathbb{I}_0 \to [0,1]$ by requiring that $d(x,x) = 0$ for all $x \in \mathbb{I}_0$, $d(x,y) = d(y,x)$ for all $x, y \in \mathbb{I}_0$, and

$$\begin{aligned} d(x,y) := &\limsup_{n\to\infty} \frac{1}{2n} \#\{1 \le k \le n : x \prec a_k \prec y\} \\ &+ \limsup_{n\to\infty} \frac{1}{2n} \#\{1 \le \ell \le n : x \prec b_\ell \prec y\} \end{aligned}$$

for $x \prec y$. It follows from exchangeability, de Finetti's theorem, and the strong law of large numbers that in the above the superior limits are actually limits almost surely.

Remark 2.5.4. It is clear that by redefining d on a $\mathbb{P}$-null set we may assume for every $x, y, z \in \mathbb{I}_0$ that

- $d(x,y) \ge 0$,

- $d(x, y) = d(y, x)$,
- $d(x, z) \le d(x, y) + d(y, z)$,
- $d(x, y) = 0$ if $x = y$.

Remark 2.5.5. For distinct $x, y, z \in \mathbb{I}_0$ the triangle inequality $d(x, z) \le d(x, y) + d(y, z)$ can be sharpened to a statement that for all x, y, z

- $d(x, z) = d(x, y) + d(y, z)$ if $x \prec y \prec z$,
- $d(x, z) = d(x, y) - d(y, z)$ if $x \prec z \prec y$,
- $d(x, z) = d(y, z) - d(x, y)$ if $y \prec x \prec z$,

and three analogous equalities when $z \prec x$.

Proposition 2.5.6. *If $x, y \in \mathbb{I}_0$ with $x \neq y$, then $d(x, y) > 0$ almost surely. Therefore almost surely d is a metric.*

Proof. We need to show for $k, \ell \in \mathbb{N}$ with $k \neq \ell$ that $d(a_k, a_\ell) > 0$ and $d(b_k, b_\ell) > 0$, and, furthermore, for arbitrary $k, \ell \in \mathbb{N}$ that $d(a_k, b_\ell) > 0$.

Consider $d(a_k, a_\ell)$. Set

$$I_m := \mathbb{1}(\{a_k \prec a_m \prec a_\ell\} \cup \{a_\ell \prec a_m \prec a_k\}), \quad m \notin \{k, \ell\}.$$

Suppose that Π_n, $n \in \mathbb{N}$ is a uniform random permutation of $[n]$. By exchangeability of the total order, if $k \vee \ell \le n$, then

$$\begin{aligned}
&\mathbb{P}\{I_m = 0,\, 1 \le m \le n,\, m \notin \{k, \ell\}\} \\
&\quad = \mathbb{P}(\{\Pi_n(\ell) = \Pi_n(k) + 1\} \cup \{\Pi_n(k) = \Pi_n(\ell) + 1\}) \\
&\quad = 2(n-1)\frac{1}{n(n-1)} \\
&\quad = \frac{2}{n}
\end{aligned}$$

and the random variables $\{I_m : m \in \mathbb{N},\, m \notin \{k, \ell\}\}$ are exchangeable. It follows from de Finetti's theorem and the strong law of large numbers that

$$\lim_{n\to\infty} \frac{1}{n} \#\{1 \le m \le n : a_k \prec a_m \prec a_\ell\} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I_m > 0$$

almost surely and hence $d(a_k, a_\ell) > 0$. A similar argument shows that $d(b_k, b_\ell) > 0$.

It remains to show that $d(a_k, b_\ell) > 0$. Set $M := \{m \in \mathbb{N} : a_k \prec b_m\}$. It follows from exchangeability that on the event $\{M \neq \emptyset\} \supseteq \{a_k \prec b_\ell\}$ we have $\#M = \infty$ almost surely and indeed that $\lim_{n\to\infty} \frac{1}{n}\#(M \cap [n]) > 0$. Write $M = \{m_1, m_2, \ldots\}$ with $m_1 < m_2 < \ldots$. Fix $p \in \mathbb{N}$ and set

$$J_q := \mathbb{1}\{b_{m_q} \prec b_{m_p}\}, \quad q \neq p.$$

By exchangeability of the total order, if $p \vee q \leq r$, then

$$\mathbb{P}\{J_q = 0,\, 1 \leq q \leq r,\, q \neq p \mid M \neq \emptyset\} = \mathbb{P}\{\Pi_r(p) = 1\} = \frac{1}{r}$$

and the random variables $\{J_q : q \in \mathbb{N},\, q \neq p\}$ are conditionally exchangeable given $\{M \neq \emptyset\}$. It follows from de Finetti's theorem that on the event $\{M \neq \emptyset\}$

$$\lim_{n\to\infty} \frac{1}{n}\#\{q : m_q \in [n],\, a_k \prec b_{m_q} \prec b_{m_p}\} > 0$$

almost surely and hence $d(a_k, b_\ell) > 0$ almost surely on the event $\{a_k \prec b_\ell\}$. A similar argument shows that $d(a_k, b_\ell) > 0$ almost surely on the event $\{b_\ell \prec a_k\}$. □

Definition 2.5.7. Given an ergodic exchangeable random total order $\prec$ on $\mathbb{I}_0$, denote by $\mathbb{I}$ the completion of $\mathbb{I}_0$ with respect to the metric d.

Definition 2.5.8. Define $f : \mathbb{I}_0 \to [0, 1]$ by

$$f(y) := \sup\{d(x, y) : x \in \mathbb{I}_0,\, x \prec y\}.$$

Remark 2.5.9. It follows from Remark 2.5.5 that

$$\begin{aligned} f(y) = &\limsup_{n\to\infty} \frac{1}{2n}\#\{1 \leq k \leq n : a_k \prec y\} \\ &+ \limsup_{n\to\infty} \frac{1}{2n}\#\{1 \leq \ell \leq n : b_\ell \prec y\}, \end{aligned}$$

$$|f(x) - f(y)| = d(x, y), \quad x, y \in \mathbb{I}_0,$$

and

$$f(x) < f(y) \iff x \prec y, \quad x, y \in \mathbb{I}_0,$$

so that f is an order-preserving isometry from $\mathbb{I}_0$ into $[0, 1]$. Thus the function f extends by continuity to an isometry from $\mathbb{I}$ into $[0, 1]$ and if $\prec$ is extended to $\mathbb{I}$ by declaring that $x \prec y \iff f(x) < f(y)$, then $\prec$ is a total order on $\mathbb{I}$ and f is an order-preserving isometry from $\mathbb{I}$ into $[0, 1]$ and hence an order-preserving isometric bijection from $\mathbb{I}$ to the image set $\mathbb{J} := f(\mathbb{I}) \subseteq [0, 1]$. Because $\mathbb{I}$ is complete, $\mathbb{J}$ is complete. Because $\mathbb{J}$ is a complete subset of $[0, 1]$ it is closed and hence compact, and therefore $\mathbb{I}$ itself is compact. It follows from the ergodicity of $\prec$ that $\mathbb{J}$ is almost surely constant. We will see below that $\mathbb{J} = [0, 1]$.

Remark 2.5.10. Define a sequence $((X_n, Y_n))_{n\in\mathbb{N}}$ of $\mathbb{J}^2$-valued random variables by setting $X_n := f(a_n)$ and $Y_n := f(b_n)$. The exchangeability of $\prec$ implies that if σ and τ are two finite permutations of $\mathbb{N}$, then $((X_{\sigma(n)}, Y_{\tau(n)}))_{n\in\mathbb{N}}$ has the same distribution as $((X_n, Y_n))_{n\in\mathbb{N}}$. In particular, the sequence $((X_n, Y_n))_{n\in\mathbb{N}}$ is exchangeable. It is a consequence of de Finetti's theorem and the ergodicity of $\prec$ that this sequence is i.i.d. with common distribution some probability measure π on $\mathbb{J}^2$. It follows from the next result that $\pi = \mu\otimes\nu$ for two probability measures μ and ν on $\mathbb{J}$ that we call the *canonical pair*. Because $X_m \neq X_n$ and $Y_m \neq Y_n$ almost surely for $m \neq n$, the probability measures μ and ν must be diffuse.

Lemma 2.5.11. *Suppose that the random variables X', Y', X'', Y'' are such that*

1. $(X', Y') \stackrel{d}{=} (X'', Y'')$
2. $((X', Y'), (X'', Y'')) \stackrel{d}{=} ((X', Y''), (X'', Y'))$
3. $(X', Y') \perp\!\!\!\perp (X'', Y'')$.

Then X', X'', Y', Y'' are independent.

Proof. For Borel sets A', A'', B', B'' we have

$$\begin{aligned}
&\mathbb{P}\{X' \in A',\ X'' \in A'',\ Y' \in B',\ Y'' \in B''\} \\
&= \mathbb{P}\{X' \in A',\ Y' \in B'\}\mathbb{P}\{X'' \in A'',\ Y'' \in B''\} && \text{by (3)} \\
&= \mathbb{P}\{X' \in A',\ Y'' \in B'\}\mathbb{P}\{X'' \in A'',\ Y' \in B''\} && \text{by (2)} \\
&= \mathbb{P}\{X' \in A'\}\mathbb{P}\{Y'' \in B'\}\mathbb{P}\{X'' \in A''\}\mathbb{P}\{Y' \in B''\} && \text{by (3)} \\
&= \mathbb{P}\{X' \in A'\}\mathbb{P}\{Y' \in B'\}\mathbb{P}\{X'' \in A''\}\mathbb{P}\{Y'' \in B''\} && \text{by (1).}
\end{aligned}$$

□

Theorem 2.5.12. *Any ergodic exchangeable random total order $\prec$ has the same distribution as one given by the construction in Remark 2.5.2 for some pair of diffuse probability measures (ζ, η) on $\mathbb{R}$. The canonical pair of diffuse probability measures (μ, ν) on $[0, 1]$ are uniquely determined by the moment formulae*

$$\begin{aligned}
&\int_{[0,1]} x^n\,\mu(dx) \\
&\quad = \left(\frac{1}{2}\right)^n \sum_{c\in\prod_{k=1}^n\{a_k,b_k\}} \mathbb{P}\{c_1 \prec a_{n+1},\ \ldots,\ c_n \prec a_{n+1}\}
\end{aligned}$$

and

$$\begin{aligned}
&\int_{[0,1]} y^n\,\nu(dy) \\
&\quad = \left(\frac{1}{2}\right)^n \sum_{c\in\prod_{k=1}^n\{a_k,b_k\}} \mathbb{P}\{c_1 \prec b_{n+1},\ \ldots,\ c_n \prec b_{n+1}\}.
\end{aligned}$$

The probability measure $\frac{1}{2}(\mu+\nu)$ *is Lebesgue measure on* $[0,1]$ *and, in particular,* $\mathbb{J}=[0,1]$*. Moreover,* μ *and* ν *are the respective push-forwards of* ζ *and* η *by the function* $z \mapsto \frac{1}{2}(\zeta+\eta)((-\infty,z])$

Proof. We have already shown that an ergodic exchangeable random total order has the same distribution as one built from an arbitrary pair (ζ,η) of diffuse probability measures on $\mathbb{R}$ using the construction in Remark 2.5.2.

Define $((X_n,Y_n))_{n\in\mathbb{N}}$ as in Remark 2.5.10. It follows from Remark 2.5.9 that

$$X_n = \frac{1}{2}\mu((-\infty,X_n]) + \frac{1}{2}\nu((-\infty,X_n])$$

and

$$Y_n = \frac{1}{2}\mu((-\infty,Y_n]) + \frac{1}{2}\nu((-\infty,Y_n])$$

for any $n\in\mathbb{N}$. Let $(I_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables that is independent of $((X_n,Y_n))_{n\in\mathbb{N}}$ with $\mathbb{P}\{I_n=0\}=\mathbb{P}\{I_n=1\}=\frac{1}{2}$ and set $Z_n := I_nX_n+(1-I_n)Y_n$ so that the sequence $(Z_n)_{n\in\mathbb{N}}$ is i.i.d. with common distribution $\frac{1}{2}(\mu+\nu)$. We have

$$Z_n = \frac{1}{2}(\mu+\nu)((-\infty,Z_n]),$$

and so $\frac{1}{2}(\mu+\nu)$ is Lebesgue measure on $[0,1]$. Thus, for any $n\in\mathbb{N}$

$$\begin{aligned}\int_{[0,1]} x^n\,\mu(dx) &= \mathbb{P}\{Z_1 < X_{n+1},\,\ldots,\,Z_n < X_{n+1}\}\\ &= \left(\frac{1}{2}\right)^n \sum_{c\in\prod_{k=1}^n\{a_k,b_k\}} \mathbb{P}\{c_1 \prec a_{n+1},\,\ldots,\,c_n \prec a_{n+1}\}\end{aligned}$$

and

$$\begin{aligned}\int_{[0,1]} y^n\,\nu(dy) &= \mathbb{P}\{Z_1 < Y_{n+1},\,\ldots,\,Z_n < Y_{n+1}\}\\ &= \left(\frac{1}{2}\right)^n \sum_{c\in\prod_{k=1}^n\{a_k,b_k\}} \mathbb{P}\{c_1 \prec b_{n+1},\,\ldots,\,c_n \prec b_{n+1}\},\end{aligned}$$

as claimed.

The proof of the final claim is straightforward and we omit it. □

Remark 2.5.13. We haven't shown that if $(y_k)_{k\in\mathbb{N}}$ is a sequence of points of $\mathbb{W}$, where $y_k\in\mathbb{W}_{N(y_k)}$, $N(y_k)\to\infty$ as $k\to\infty$, and $\lim_{k\to\infty} y_k = y$ in the Doob–Martin topology for some arbitrary y in the Doob–Martin boundary that the harmonic function $K(\cdot,y)$ is extremal. This is equivalent to showing that if the infinite bridge $(U_n^\infty)_{n\in\mathbb{N}_0}$ is the limit of the bridges $(U_0^{y_k},\ldots,U_{N(y_k)}^{y_k})$, then $(U_n^\infty)_{n\in\mathbb{N}_0}$ has an almost surely trivial tail σ-field. This is, in turn, equivalent to showing that the corresponding labeled infinite bridge induces an ergodic exchangeable random order. The latter, however, can be established along the lines of [13, Corollary 5.21] and [14, Corollary 7.2], so we omit the details.

Characterization of exchangeable random paired total orders associated with infinite bridges $(\widetilde{W}_n)_{n\in\mathbb{N}_0}$

We see from previous sections that understanding the Doob-Martin boundary is equivalent to understanding infinite bridges, and that this further boils down to understanding the corresponding exchangeable random paired total orders on $\mathbb{I}_0$. A mixture of two exchangeable random paired total orders is also an exchangeable random paired total order, so we are interested in exchangeable random paired total orders $\prec$ that are extremal in the sense that they cannot be written as a mixture of two other exchangeable random paired total orders. This is equivalent to requiring that if A is a measurable subset of the space of total orders on $\mathbb{I}_0$ with the property that $\lhd \in A$ if and only if $\lhd^\sigma \in A$ for all paired total orders $\lhd$ and all finite permutations σ of $\mathbb{N}$, then $\mathbb{P}\{\prec \in A\} \in \{0, 1\}$. We say that an exchangeable random paired total order with this property is *ergodic*.

The pertinence of ergodic random total orders to the distribution of infinite bridges $(W_n^\infty)_{n\in\mathbb{N}_0}$ is due to the following lemma which follows from essentially the same arguments as the proof of [13, Proposition 5.19].

Lemma 2.5.14. *The tail σ-field of an infinite bridge $(W_n^\infty)_{n\in\mathbb{N}_0}$ is almost surely trivial if and only if the exchangeable random total order induced by the corresponding labeled infinite bridge $(L_n^\infty)_{n\in\mathbb{N}_0}$ is ergodic.*

Remark 2.5.15. Let η be a probability measure on $\mathbb{R}^2$ that assigns all of its mass to the set $\{(s,t) : s \le t\}$ and has diffuse marginals. Let $((S_n, T_n))_{n\in\mathbb{N}}$ be i.i.d. with common distribution η. A random total order $\lhd$ on $\mathbb{I}_0$ may be constructed by declaring that

- $a_i \lhd a_j$ if $S_i < S_j$,
- $b_i \lhd b_j$ if $T_i < T_j$,
- $a_i \lhd b_j$ if $S_i < T_j$,
- $b_i \lhd a_j$ if $T_i < S_j$,
- $a_k \lhd b_k$ if $S_k \le T_k$,

The is random total order is paired, exchangeable, and ergodic. The properties of being paired and exchangeable are obvious from the construction. As remarked in the Introduction, ergodicity follows from the Hewitt-Savage zero-one law or by suitably encoding the exchangeable random total order as a jointly exchangeable array and checking that this array is ergodic because it is dissociated. We now proceed to show that all ergodic exchangeable random paired total orders are of this form.

Proposition 2.5.16. *Suppose that $\prec$ is an ergodic exchangeable random paired total order on $\mathbb{I}_0$. Define $f : \mathbb{I}_0 \to [0,1]$ by*

$$f(c) = \limsup_{n\to\infty} \frac{1}{2n} \#\{1 \le k \le n : a_k \prec c\} + \limsup_{n\to\infty} \frac{1}{2n} \#\{1 \le \ell \le n : b_\ell \prec c\}.$$

If $c, d \in \mathbb{I}_0$ with $c \prec d$, then $f(c) < f(d)$ almost surely with a possible exception when $c = a_k$ and $d = b_k$ for some $k \in \mathbb{N}$.

Proof. It follows from exchangeability, de Finetti's theorem, and the strong law of large numbers that limit superiors in the definition are actually limits almost surely.

Consider first the proof that $f(a_k) < f(a_\ell)$ on the event $\{a_k \prec a_\ell\}$. Set

$$I_m := \mathbb{1}\{a_k \prec a_m \prec a_\ell\}, \quad m \notin \{k, \ell\}.$$

By exchangeability of the random total order, if $k \vee \ell \le n$, then

$$\mathbb{P}\{I_m = 0,\, 1 \le m \le n,\, m \notin \{k,\ell\} \,|\, a_k \prec a_\ell\} = (n-1)\frac{1}{n(n-1)}\Big/\frac{1}{2} = \frac{2}{n}$$

because

$$\mathbb{P}\{I_m = 0,\, 1 \le m \le n,\, m \notin \{k,\ell\},\, a_k \prec a_\ell\}$$

is the probability $(n-1)\frac{1}{n(n-1)}$ that a uniform random permutation σ of $[n]$ is such that $\sigma(p) = k$ and $\sigma(p+1) = \ell$ for some p with $1 \le p \le n-1$. Also, the random variables $\{I_m : m \in \mathbb{N},\, m \notin \{k,\ell\}\}$ are conditionally exchangeable given the event $\{a_k \prec a_\ell\}$. It follows from de Finetti's theorem and the strong law of large numbers that

$$\lim_{n\to\infty} \frac{1}{n} \#\{1 \le m \le n : a_k \prec a_m \prec a_\ell\} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I_m > 0$$

almost surely on the event $\{a_k \prec a_\ell\}$ and hence $f(a_k) < f(a_\ell)$ almost surely on the event $\{a_k \prec a_\ell\}$.

A similar argument shows that $f(b_k) < f(b_\ell)$ almost surely on the event $\{b_k \prec b_\ell\}$.

It remains to show for $k \ne \ell$ that $f(a_k) < f(b_\ell)$ almost surely on the event $\{a_k \prec b_\ell\}$ (respectively, $f(b_k) < f(a_\ell)$ almost surely on the event $\{b_k \prec a_\ell\}$). Consider the first case and denote the random set $\{m \in \mathbb{N} : a_k \prec b_m\}$ by M. It follows from exchangeability that $\#M = \infty$ almost surely and indeed $\lim_{n\to\infty} \frac{1}{n}\#(M \cap [n]) > 0$ on the event $\{M \ne \emptyset\}$. For this case, write $M = \{m_1, m_2, \ldots\}$ with $m_1 < m_2 < \ldots$ and for fixed $p \in \mathbb{N}$ define

$$J_q := \mathbb{1}\{b_{m_q} \prec b_{m_p}\}, \quad q \ne p.$$

By exchangeability of the total order, if $p \vee q \leq r$, then

$$\mathbb{P}\{J_q = 0,\, 1 \leq q \leq r,\, q \neq p \mid M \neq \emptyset\} = \frac{1}{r}$$

and the random variables $\{J_q : q \in \mathbb{N},\, q \neq p\}$ are conditionally exchangeable given $\{M \neq \emptyset\}$. It follows from de Finetti's theorem that on the event $\{M \neq \emptyset\}$

$$\lim_{n\to\infty} \frac{1}{n} \#\{q : m_q \in [n],\, a_k \prec b_{m_q} \prec b_{m_p}\} > 0$$

almost surely and hence $f(a_k) < f(b_\ell)$ almost surely on the event $\{a_k \prec b_\ell\}$. A similar argument shows that $f(b_k) < f(a_\ell)$ almost surely on the event $\{b_k \prec a_\ell\}$. □

Theorem 2.5.17. *Suppose that $\prec$ is an ergodic exchangeable random paired total order on $\mathbb{I}_0$. Let $f : \mathbb{I}_0 \to [0,1]$ be as in Proposition 2.5.16. Define a sequence $((X_n, Y_n))_{n\in\mathbb{N}}$ of $[0,1]^2$-valued random variables by setting $X_n := f(a_n)$ and $Y_n := f(b_n)$. Then $((X_n, Y_n))_{n\in\mathbb{N}}$ is i.i.d. with common distribution a probability measure μ that assigns all of its mass to the set $\{(x,y) : 0 \leq x \leq y \leq 1\}$ and has the property that $\frac{1}{2}(\mu(\cdot \times \mathbb{R}) + \mu(\mathbb{R} \times \cdot))$ is Lebesgue measure on $[0,1]$. The probability measure μ is uniquely determined by the moment formulae*

$$\begin{aligned}
&\int_{[0,1]^2} x^m y^n \,\mu(dx, dy) \\
&\quad = \left(\frac{1}{2}\right)^{m+n} \sum_{c \in \prod_{k=1}^{m+n}\{a_k, b_k\}} \mathbb{P}\{c_1 \prec a_{m+n+1},\, \ldots,\, c_m \prec a_{m+n+1}, \\
&\qquad\qquad c_{m+1} \prec b_{m+n+1},\, \ldots,\, c_{m+n} \prec b_{m+n+1}\}.
\end{aligned}$$

The ergodic exchangeable random paired total order $\prec$ on $\mathbb{I}_0$ can be recovered from the random sequence $((X_n, Y_n))_{n\in\mathbb{N}}$ by declaring

- $a_i \prec a_j$ *if* $X_i < X_j$,
- $b_i \prec b_j$ *if* $Y_i < Y_j$,
- $a_i \prec b_j$ *if* $X_i < Y_j$,
- $b_i \prec a_j$ *if* $Y_i < X_j$,
- $a_k \prec b_k$ *if* $X_k = Y_k$.

Proof. The exchangeability of the random paired total order $\prec$ implies that the sequence $((X_n, Y_n))_{n\in\mathbb{N}}$ is exchangeable. It is a consequence of de Finetti's theorem that in order to show that this sequence is i.i.d. it suffices to show that it has an almost surely trivial tail σ-field. For any $m, n \in \mathbb{N}$ we have that

$$\begin{aligned}
X_n &= f(a_n) \\
&= \limsup_{p\to\infty} \frac{1}{2p} \#\{k : m \le k \le p,\, a_k \prec a_n\} \\
&\quad + \limsup_{p\to\infty} \frac{1}{2p} \#\{\ell : m \le \ell \le p,\, b_\ell \prec a_n\}, \\
Y_n &= f(b_n) \\
&= \limsup_{p\to\infty} \frac{1}{2p} \#\{k : m \le k \le p,\, a_k \prec b_n\} \\
&\quad + \limsup_{p\to\infty} \frac{1}{2p} \#\{\ell : m \le \ell \le p,\, b_\ell \prec b_n\}.
\end{aligned}$$

Thus, given a real-valued random variable V_m that is measurable with respect to $\sigma\{(X_n, Y_n) : n \ge m\}$ there exists a measurable real-valued function Υ_m defined on the paired total orders on $\mathbb{I}_0$ such that $V_m = \Upsilon_m(\prec)$ almost surely and $\Upsilon_m(\prec) = \Upsilon_m(\prec_\pi)$ almost surely for any permutation π of $\mathbb{N}$ that leaves the elements of $\{m, m+1, \ldots\}$ fixed. Consequently, given a real-valued random variable V that is measurable with respect to the tail σ-field $\bigcap_{m\in\mathbb{N}} \sigma\{(X_n, Y_n) : n \ge m\}$ there exists a measurable real-valued function Υ defined on the paired total orders on $\mathbb{I}_0$ such that $V = \Upsilon(\prec)$ almost surely and $\Upsilon(\prec) = \Upsilon(\prec_\pi)$ almost surely for any permutation π of $\mathbb{N}$ that leaves all but finitely many of the elements of $\mathbb{N}$ fixed. It follows from the ergodicity of $\prec$ that any tail measurable random variable V is constant almost surely and hence the sequence $((X_n, Y_n))_{n\in\mathbb{N}}$ is i.i.d. with common distribution some probability measure μ on $[0,1]^2$. By Proposition 2.5.16, μ is concentrated on the set $\{(x, y) : 0 \le x \le y \le 1\}$, and $\mathbb{P}\{X_m = X_n\} = \mathbb{P}\{Y_m = Y_n\} = 0$ for $m \ne n$, hence the marginal distributions of μ must be diffuse.

Let $(I_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables that is independent of $((X_n, Y_n))_{n\in\mathbb{N}}$ with $\mathbb{P}\{I_n = 0\} = \mathbb{P}\{I_n = 1\} = \frac{1}{2}$ and set $U_n := I_n X_n + (1 - I_n) Y_n$ so that the sequence $(U_n)_{n\in\mathbb{N}}$ is i.i.d. with common distribution $\frac{1}{2}(\alpha + \beta)$, where α and β are the distributions of X_n and Y_n. By construction,

$$\begin{aligned}
X_n &= f(a_n) \\
&= \lim_{m\to\infty} \frac{1}{2m} \#\{1 \le k \le m : a_k \prec a_n\} \\
&\quad + \lim_{m\to\infty} \frac{1}{2m} \#\{1 \le \ell \le m : b_\ell \prec a_n\} \\
&= \lim_{m\to\infty} \frac{1}{2m} \#\{1 \le k \le m : X_k \prec X_n\} \\
&\quad + \lim_{m\to\infty} \frac{1}{2m} \#\{1 \le \ell \le m : Y_\ell \prec X_n\} \\
&= \frac{1}{2}(\alpha + \beta)([0, X_n]),
\end{aligned}$$

and a similar argument shows that

$$Y_n = \frac{1}{2}(\alpha + \beta)([0, Y_n]).$$

Therefore,

$$U_n = \frac{1}{2}(\alpha + \beta)([0, U_n]).$$

Now for any diffuse probability measure γ on $[0,1]$, the distribution of $\gamma([0,Z])$ for Z a random variable with distribution γ is uniform on $[0,1]$. Therefore, U_n has a uniform distribution on $[0,1]$ and $\frac{1}{2}(\alpha+\beta)$ is the Lebesgue measure on $[0,1]$. Thus,

$$\begin{aligned}
&\int_{[0,1]^2} x^m y^n \, \mu(dx, dy) \\
&= \mathbb{P}\{U_1 < X_{m+n+1}, \ldots, U_m < X_{m+n+1}, \\
&\qquad U_{m+1} < Y_{m+n+1}, \ldots, U_{m+n} < Y_{m+n+1}\} \\
&= \left(\frac{1}{2}\right)^{m+n} \sum_{c \in \prod_{k=1}^{m+n}\{a_k, b_k\}} \mathbb{P}\{c_1 \prec a_{m+n+1}, \ldots, c_m \prec a_{m+n+1}, \\
&\qquad c_{m+1} \prec b_{m+n+1}, \ldots, c_{m+n} \prec b_{m+n+1}\}.
\end{aligned}$$

□

Remark 2.5.18. Let μ be a probability measure on $[0,1]^2$ that assigns all of its mass to the set $\{(s,t) : 0 \le s \le t \le 1\}$ and has the property that $\frac{1}{2}(\mu(\cdot \times \mathbb{R}) + \mu(\mathbb{R} \times \cdot))$ is Lebesgue measure on $[0,1]$. Let $((S_n, T_n))_{n\in\mathbb{N}}$ be i.i.d. with common distribution μ. Define a process of labeled matchings $(L_n^\infty)_{n\in\mathbb{N}_0}$ as follows: if $C_1, \ldots, C_{2n}$ are the order statistics of $\bigcup_{k=1}^n \{S_k, T_k\}$ with $C_1 \le \cdots \le C_{2n}$, then make $\{i,j\} \subset [2n]$ a block of L_n^∞ labeled with $k \in [n]$ whenever $C_i = S_k$ and $C_j = T_k$. Then $(W_n^\infty)_{n\in\mathbb{N}_0} := (\Psi(L_n^\infty))_{n\in\mathbb{N}_0}$ is an extremal infinite bridge and $(L_n^\infty)_{n\in\mathbb{N}_0}$ is the corresponding process of labeled associated admissible matchings guaranteed by Corollary 2.3.17 if and only if the conditional distribution of L_n^∞ given $\Psi(L_n^\infty)$ is uniform on the set of labeled associated admissible matchings of $\Psi(L_n^\infty)$ for all $n \in \mathbb{N}_0$.

2.6 Identification of extremal harmonic functions

Idenfitication of extremal harmonic functions for the chain $(U_n)_{n\in\mathbb{N}_0}$

Any extremal infinite bridge $(U_n^\infty)_{n\in\mathbb{N}_0}$ is the h-transform of the original Markov chain $(U_n)_{n\in\mathbb{N}_0}$ with an extremal harmonic function h. We know from the above that such a process arises as follows in terms of the canonical pair (μ, ν) of diffuse probability measures associated with the corresponding point in the Doob–Martin boundary.

We first require some notation. Given $(x_1, \dots, x_n, y_1, \dots, y_n) \in \mathbb{R}^{2n}$ with distinct entries, let $z_1 < \dots < z_{2n}$ be a listing of $\{x_1, \dots, x_n, y_1, \dots, y_n\}$ in increasing order. Define

$$\mathcal{W}((x_1, \dots, x_n, y_1, \dots, y_n)) = u_1 \dots u_{2n} \in \mathbb{W}_n$$

by

$$u_i = \begin{cases} a, & \text{if } z_i \in \{x_1, \dots, x_n\}, \\ b, & \text{if } z_i \in \{y_1, \dots, y_n\}. \end{cases}$$

Given $v \in \mathbb{W}_n$, set

$$\mathcal{S}(v) := \mathcal{W}^{-1}(\{v\}) \subset \mathbb{R}^{2n}.$$

For example,

$$\mathcal{S}(abba) = \bigsqcup_{\sigma,\tau} \{(x_1, x_2, y_1, y_2) \in \mathbb{R}^4 : x_{\sigma(1)} < y_{\tau(1)} < y_{\tau(2)} < x_{\sigma(2)}\},$$

where the union is over all pairs of permutations σ, τ of the set $\{1, 2\}$. In general, $\mathcal{S}(v)$ is the disjoint union of $(n!)^2$ connected open sets that all have boundaries of zero Lebesgue measure.

Now take independent sequences of real-valued random variables $(X_k)_{k\in\mathbb{N}}$ and $(Y_k)_{k\in\mathbb{N}}$, where the X_k are i.i.d. with common distribution μ and the Y_k are i.i.d. with common distribution ν and set

$$U_n^\infty = \mathcal{W}((X_1, \dots, X_n, Y_1, \dots, Y_n)).$$

We have

$$\mathbb{P}\{U_n^\infty = u\} = \mu^{\otimes n} \otimes \nu^{\otimes n}(\mathcal{S}(u))$$

We also know that

$$\mathbb{P}\{U_n^\infty = u \,|\, U_{n+1}^\infty = v\} = \frac{\binom{v}{u}}{(n+1)^2}.$$

It follows that

$$\begin{aligned}&\mathbb{P}\{U_{n+1}^\infty = v \,|\, U_n^\infty = u\} \\ &\quad = \mu^{\otimes(n+1)} \otimes \nu^{\otimes(n+1)}(\mathcal{S}(v)) \frac{\binom{v}{u}}{(n+1)^2} \Big/ \mu^{\otimes n} \otimes \nu^{\otimes n}(\mathcal{S}(u)).\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{P}\{U_{n+1}^\infty = v \,|\, U_n^\infty = u\} &= \frac{1}{h(u)} \mathbb{P}\{U_{n+1} = v \,|\, U_n = u\} h(v) \\ &= \frac{h(v)}{h(u)} \frac{\binom{v}{u}}{(2n+2)(2n+1)}.\end{aligned}$$

Thus,

$$\frac{h(v)}{h(u)} = \frac{\mu^{\otimes(n+1)} \otimes \nu^{\otimes(n+1)}(\mathcal{S}(v))}{\mu^{\otimes n} \otimes \nu^{\otimes n}(\mathcal{S}(u))} \frac{(2n+2)(2n+1)}{(n+1)^2}$$

and, up to an arbitrary multiplicative constant,

$$h(w) = \binom{2m}{m} \mu^{\otimes m} \otimes \nu^{\otimes m}(\mathcal{S}(w))$$

for $w \in \mathbb{W}_m$.

Since $h(\emptyset) = 1$, this normalization is the extended Doob–Martin kernel $w \mapsto K(w, y)$, where y is the point in the Doob–Martin boundary that corresponds to the pair of diffuse probability measures (μ, ν).

Remark 2.6.1. The constant harmonic function $h \equiv 1$ arises from the above construction with μ and ν both being the Lebesgue measure λ on $[0, 1]$. Therefore the process $(U_n)_{n\in\mathbb{N}_0}$ is itself the extremal bridge associated with the pair (λ, λ). In particular, $(U_n)_{n\in\mathbb{N}_0}$ converges almost surely to this point in the Doob–Martin boundary associated with this pair.

We observed in Remark 2.3.7 that a sequence $(y_k)_{k\in\mathbb{N}}$ with $N(y_k) \to \infty$ as $k \to \infty$ converges in the Doob–Martin topology if and only if for every $m \in \mathbb{N}$ the sequence of random words in $\mathbb{W}_m$ obtained by selecting m letters a and m letters b uniformly at random from y_k and maintaining their relative order converges in distribution as $k \to \infty$. We can now enhance that result as follows.

Proposition 2.6.2. *Consider a sequence $(y_k)_{k\in\mathbb{N}}$ in $\mathbb{W}$, where $y_k \in \mathbb{W}_{N(y_k)}$, $k \in \mathbb{N}$, and $N(y_k) \to \infty$ as $k \to \infty$. If y is the point in the Doob–Martin boundary that corresponds to the pair of diffuse probability measures (μ, ν) with $\frac{1}{2}(\mu + \nu) = \lambda$, then $\lim_{k\to\infty} y_k = y$ in the Doob–Martin topology if and only if*

$$\lim_{k\to\infty} \frac{\binom{y_k}{w}}{\binom{N(y_k)}{m}^2} = \mu^{\otimes m} \otimes \nu^{\otimes m}(\mathcal{S}(w))$$

for all $w \in \mathbb{W}_m$ for all $m \in \mathbb{N}$. That is, $\lim_{k\to\infty} y_k = y$ if and only if for each $m \in \mathbb{N}$ the sequence of random words in $\mathbb{W}_m$ obtained by selecting m letters a and m letters b uniformly at random from y_k and maintaining their relative order converges in distribution as $k \to \infty$ to the random word $U_m^\infty = \mathcal{W}(X_1, \ldots, X_m, Y_1, \ldots, Y_m)$ defined above.

Given a sequence $(y_k)_{k\in\mathbb{N}}$ in $\mathbb{W}$, where $y_k \in \mathbb{W}_{N(y_k)}$, $k \in \mathbb{N}$, and $N(y_k) \to \infty$ as $k \to \infty$, define a sequence of pairs of discrete probability measures $((\mu_k, \nu_k))_{k\in\mathbb{N}}$ on $[0, 1]$ as follows. For $k \in \mathbb{N}$ the two probability measure μ_k and ν_k both assign all of their mass to the set $\{\frac{\ell}{2N(y_k)} : 1 \le \ell \le 2N(y_k)\}$. For $1 \le i \le 2N(y_k)$, $\mu_k(\frac{i}{2N(y_k)}) = \frac{1}{N(y_k)}$ if the i^{th} letter of y_k is the letter a, otherwise $\mu_k(\frac{i}{2N(y_k)}) = 0$. Similarly, for $1 \le j \le 2N(y_k)$, $\nu_k(\frac{j}{2N(y_k)}) = \frac{1}{N(y_k)}$ if the j^{th} letter of y_k is the letter b, otherwise $\nu_k(\frac{j}{2N(y_k)}) = 0$. In particular, $\frac{1}{2}(\mu_k + \nu_k)$ is the uniform probability measure on $\{\frac{\ell}{2N(y_k)} : 1 \le \ell \le 2N(y_k)\}$. Observe that if $w \in \mathbb{W}_m$, then, for $w \in \mathbb{W}_m$,

$$(N(y_k)^m)^2 \mu_k^{\otimes m} \otimes \nu_k^{\otimes m}(\mathcal{S}(w)) = (m!)^2 \binom{y_k}{w}$$

so that

$$\frac{\binom{y_k}{w}}{\binom{N(y_k)}{m}^2} = \left(\frac{N(y_k)^m}{N(y_k)(N(y_k)-1)\cdots(N(y_k)-m+1)}\right)^2 \mu_k^{\otimes m} \otimes \nu_k^{\otimes m}(\mathcal{S}(w)).$$

One direction of the following corollary is now immediate.

Corollary 2.6.3. *Suppose that $(y_k)_{k\in\mathbb{N}}$ and $((\mu_k,\nu_k))_{k\in\mathbb{N}}$ are as above. If $(y_k)_{k\in\mathbb{N}}$ converges in the Doob–Martin topology to the point y in the Doob–Martin boundary that corresponds to the pair of probability distributions (μ,ν), then $(\mu_k)_{k\in\mathbb{N}}$ converges weakly to μ and $(\nu_k)_{k\in\mathbb{N}}$ converges weakly to ν. Conversely, if $(\mu_k)_{k\in\mathbb{N}}$ converges weakly to μ and $(\nu_k)_{k\in\mathbb{N}}$ converges weakly to ν, then $\frac{1}{2}(\mu+\nu)=\lambda$, and if y is the point in the Doob–Martin boundary that corresponds to the pair (μ,ν), then $(y_k)_{k\in\mathbb{N}}$ converges in the Doob–Martin topology to y.*

Proof. As we have already remarked, if $(\mu_k)_{k\in\mathbb{N}}$ converges weakly to μ and $(\nu_k)_{k\in\mathbb{N}}$ converges weakly to ν then, since the boundary of $\mathcal{S}(w)$ is Lebesgue null for any word $w\in\mathbb{W}_m$, $m\in\mathbb{N}$, we have that $\mu_k^{\otimes m}\otimes\nu_k^{\otimes m}(\mathcal{S}(w))$ converges to $\mu^{\otimes m}\otimes\nu^{\otimes m}(\mathcal{S}(w))$ so that

$$\lim_{k\to\infty}\frac{\binom{y_k}{w}}{\binom{N(y_k)}{m}^2} = \mu^{\otimes m}\otimes\nu^{\otimes m}(\mathcal{S}(w)),$$

and it follows from Proposition 2.6.2 that $(y_k)_{k\in\mathbb{N}}$ converges to the point y in the Doob–Martin boundary that corresponds to the pair (μ,ν).

Conversely, suppose that $(y_k)_{k\in\mathbb{N}}$ converges to the point y in the Doob–Martin boundary corresponding to the pair (μ,ν). Given any subsequence of $\mathbb{N}$ there is, by the compactness in the weak topology of probability measures on $[0,1]$, a further subsequence such that along this further subsequence μ_k converges weakly to some probability measure μ' and ν_k converges weakly to some probability measure ν'. Note that $\frac{1}{2}(\mu'+\nu')=\lambda$. From the other direction of the corollary, this implies that along the subsubsequence y_k converges to the point y' in the Doob–Martin boundary corresponding to the pair of probability measures (μ',ν'). Because $y'=y$ it must be the case $(\mu',\nu')=(\mu,\nu)$. Thus, from any subsequence of $\mathbb{N}$ we can extract a further subsequence along which μ_k converges weakly to μ and ν_k converges weakly to ν, and this implies that $(\mu_k)_{k\in\mathbb{N}}$ converges weakly to μ and $(\nu_k)_{k\in\mathbb{N}}$ converges weakly to ν. ☐

Identification of extremal harmonic functions for $(W_n)_{n\in\mathbb{N}_0}$

As discussed in Section 2.4, an extremal infinite bridge $(W_n^\infty)_{n\in\mathbb{N}_0}$ arises as follows when the corresponding point in the Doob-Martin boundary is identified with the probability measure μ constructed in Theorem 2.5.17 that is concentrated on $\{(x,y): 0\le x\le y\le 1\}$.

We start with some notation. Given $((x_1,y_1)\ldots,(x_n,y_n))\in\{(x,y):0\le x\le y\le 1\}^n$ with $\{x_i,y_i\}\cap\{x_j,y_j\}=\emptyset$, $i\neq j$, let $z_1\le\cdots\le z_{2n}$ be a listing of $\{x_1,y_1,\ldots,x_n,y_n\}$ in nondecreasing order. Define

$$\mathcal{W}((x_1,y_1),\ldots,(x_n,y_n)) = u_1\ldots u_{2n}\in\mathbb{B}_n$$

by

$$u_i = \begin{cases} a, & \text{if } z_i \in \{x_1, \ldots, x_n\} \text{ and } z_i \notin \{y_1, \ldots, y_n\}, \\ a, & \text{if } z_i = z_{i+1}, \\ b, & \text{if } z_i \in \{y_1, \ldots, y_n\} \text{ and } z_i \notin \{x_1, \ldots, x_n\}, \\ b, & \text{if } z_i = z_{i-1}. \end{cases}$$

Now take a sequence $((X_k, Y_k))_{k \in \mathbb{N}}$ that is i.i.d. with common distribution μ. Then

$$W_n^\infty = \mathcal{W}(((X_1, Y_1), \ldots, (X_n, Y_n)).$$

The extremal infinite bridge $(W_n^\infty)_{n \in \mathbb{N}_0}$ is the Doob harmonic transform of the chain $(W_n)_{n \in \mathbb{N}_0}$ with a unique extremal harmonic function h normalized so that $h(\emptyset) = 1$, and we now identify this function.

Given $v \in \mathbb{B}_n$, set

$$\mathcal{S}(v) := \mathcal{W}^{-1}(\{v\}) \subset \{(x, y) : 0 \le x \le y \le 1\}^n.$$

Then for $u \in \mathbb{B}_n$,

$$\mathbb{P}\{W_n^\infty = u\} = \mu^{\otimes n}(\mathcal{S}(u)).$$

Also, for $v \in \mathbb{B}_{n+1}$, by Proposition 2.3.13 (iv),

$$\mathbb{P}\{W_n^\infty = u \,|\, W_{n+1}^\infty = v\} = \frac{1}{n+1} R(u, v) \frac{\Lambda(u)}{\Lambda(v)}.$$

It follows that

$$\begin{aligned} &\mathbb{P}\{W_{n+1}^\infty = v \,|\, W_n^\infty = u\} \\ &\quad = \mu^{\otimes(n+1)}(\mathcal{S}(v)) \frac{1}{n+1} R(u, v) \frac{\Lambda(u)}{\Lambda(v)} \Big/ \mu^{\otimes n}(\mathcal{S}(u)). \end{aligned}$$

On the other hand, it follows from (2.2) that

$$\begin{aligned} \mathbb{P}\{W_{n+1}^\infty = v \,|\, W_n^\infty = u\} &= \frac{1}{h(u)} \mathbb{P}\{W_{n+1} = v \,|\, W_n = u\} h(v) \\ &= \frac{h(v)}{h(u)} \frac{R(u, v)}{\binom{2(n+1)}{2}}. \end{aligned}$$

Thus,

$$\frac{h(v)}{h(u)} = \frac{\mu^{\otimes(n+1)}(\mathcal{S}(v))}{\mu^{\otimes n}(\mathcal{S}(u))} \frac{\Lambda(u)}{\Lambda(v)} \frac{(2n+2)(2n+1)}{2(n+1)} = \frac{\mu^{\otimes(n+1)}(\mathcal{S}(v))}{\mu^{\otimes n}(\mathcal{S}(u))} \frac{\Lambda(u)}{\Lambda(v)} (2n+1),$$

and so

$$h(w) = (2m-1)!! \mathbb{P}\{W_m^\infty = w\} / \Lambda(w) = (2m-1)!! \mu^{\otimes m}(\mathcal{S}(w)) / \Lambda(w) \tag{2.3}$$

for $w \in \mathbb{B}_m$, $m \ge 1$, when we set $h(\emptyset) = 1$. Put $h^\mu := h$ to record the dependence on μ.

Remark 2.6.4. The foregoing has established Theorem 2.1.4.

Example 2.6.5. Consider the normalized harmonic function $h \equiv 1$. We have that $h = h^\mu$ where

$$\mu(dx, dy) = 2\,dx\,dy, \quad 0 \le x \le y \le 1.$$

The chain $(W_n)_{n\in\mathbb{N}_0}$ is therefore itself an extremal bridge that converges also in the Doob–Martin topology to the extremal element of the Doob–Martin boundary corresponding to μ.

Example 2.6.6. Consider the sequence $(w_n)_{n\in\mathbb{N}}$ where $w_n \in \mathbb{B}_n$ is the word $abab\ldots ab$. Recall that Q is the matrix of backward transition probabilities. It is clear that if $v \in \mathbb{B}_m$, $m < n$, then $Q(w_n, v) = 1$ if $v = abab\ldots ab$ and $Q(w_n, v) = 0$ otherwise. From (2.1), if $v = abab\ldots ab \in \mathbb{B}_m$, then

$$K(v, w_n) = \frac{Q^{n-m}(w_n, v)}{P^m(\emptyset, v)} = \frac{(2m-1)!!}{\Lambda(v)},$$

where $\Lambda(v) = 1$. It follows that $(w_n)_{n\in\mathbb{N}}$ converges in the Doob–Martin topology to the point y in the Doob–Martin boundary for which the extended Doob–Martin kernel is, for $v \in \mathbb{B}_m$,

$$K(v, y) = \begin{cases} (2m-1)!!, & \text{if } v = abab\ldots ab, \\ 0, & \text{otherwise.} \end{cases}$$

Thus $K(\cdot, y)$ is the extremal harmonic function h^μ for $\mu(dx, dy) = dx\,\delta_x(dy)$, $0 \le x \le y \le 1$.

Example 2.6.7. Consider the sequence $(w_n)_{n\in\mathbb{N}}$ where $w_n \in \mathbb{B}_n$, $n \in \mathbb{N}$, is the word consisting of C runs of the letter a of successive lengths $A_n^1, \ldots, A_n^C$ interleaved with C runs of the letter b of successive lengths $B_n^1, \ldots, B_n^C$. Here $\sum_{i=1}^C A_n^i = \sum_{j=1}^C B_n^j = n$ and $\sum_{i=1}^c A_n^i \ge \sum_{j=1}^c B_n^j$ for $1 \le c \le C$.

Suppose that

$$\lim_{n\to\infty} \left(\frac{A_n^1}{n}, \ldots, \frac{A_n^C}{n}\right) = (p^1, \ldots, p^C)$$

and

$$\lim_{n\to\infty} \left(\frac{B_n^1}{n}, \ldots, \frac{B_n^C}{n}\right) = (q^1, \ldots, q^C).$$

It will be convenient to set

$$\gamma^0 = 0,\ \gamma^1 = \frac{p^1}{2},\ \gamma^2 = \frac{p^1 + q^1}{2},\ \gamma^3 = \frac{p^1 + q^1 + p^2}{2}, \ldots,$$
$$\gamma^{2C} = \frac{p^1 + \cdots + p^C + q^1 + \cdots + q^C}{2} = 1,$$

Consider a uniform labeled associated admissible matching $L_n = \{((i,j),k)\}$ of the word w_n; the presence of the element $((i,j),k)$ indicates that a letter a with index $i \in [2n]$ is matched with a letter b with index $j \in [2n]$ and that this pair is labeled by $k \in [n]$.

We start by observing that $A_n^C \leq B_n^C$. Following lines similar to our proof of the formula for $\Lambda(w)$ in Lemma 2.3.12, all the letters a in the final run of length A_n^C must be matched with a b in the final run of length B_n^C. This leaves $B_n^C - A_n^C$ letters b in the final run unmatched. Assigning mass $\frac{1}{n}$ to each letter, there is thus asymptotically a mass $q^C - p^C$ of letters b unmatched in the final run. The matching arises by choosing uniformly at random without replacement the letters b that will be matched with the successive letters a. More explicitly, denote by $1 \leq k_1 < \cdots < k_{A_n^C} \leq n$ the labels for which the corresponding matched pairs (i,j) of indices are such that i indexes a letter a in the final run and j indexes a letter b in the final run. Denote by (S_{nh}, T_{nh}) the matched pair of indices labeled by k_h, $1 \leq h \leq A_n^C$. The sequence $(S_{n1}, \ldots, S_{nA_n^C})$ (resp. $(T_{n1}, \ldots, T_{nA_n^C})$ consists of A_n^C uniform picks without replacement from the range $A_n^1 + B_n^1 + \cdots + A_n^{C-1} + B_n^{C-1} + [A_n^C]$ (resp. $A_n^1 + B_n^1 + \cdots + A_n^{C-1} + B_n^{C-1} + A_n^C + [B_n^C]$) and these two sequences are independent. The finite-dimensional marginals of $((\frac{S_{n1}}{2n}, \frac{T_{n1}}{2n}), \ldots, (\frac{S_{nA_n^C}}{2n}, \frac{T_{nA_n^C}}{2n}))$ converge in distribution as $n \to \infty$ to an i.i.d. infinite sequences with common distribution ν given by

$$\nu(dx, dy) = \begin{cases} \frac{2^2}{p^C q^C}\, dx\, dy, & (x,y) \in (\gamma^{2C-2}, \gamma^{2C-1}) \times (\gamma^{2C-1}, \gamma^{2C}), \\ 0, & \text{otherwise.} \end{cases}$$

Now $A_n^{C-1} \leq B_n^{C-1} - A_n^C + B_n^C$. All of the letters a in the $(C-1)^{\text{st}}$ run of length A_n^{C-1} must be matched either with the letters b in the $(C-1)^{\text{st}}$ run of length B_n^{C-1} or with the remaining $B_n^C - A_n^C$ unmatched letters b in the C^{th} run. The matching arises by choosing uniformly without replacement the letters b in the pool of size $B_n^{C-1} - A_n^C + B_n^C$ that will be matched with the successive letters a. Asymptotically, proportions $\frac{q^{C-1}}{q^{C-1}+q^C-p^C}$ and $\frac{q^C-p^C}{q^{C-1}+q^C-p^C}$ of the mass p^{C-1} of letters a in the $(C-1)^{\text{st}}$ run will be matched with, respectively, the letters b in the $(C-1)^{\text{st}}$ run and the remaining letters b in the C^{th} run. Thus, asymptotically, the mass of letters b in the $(C-1)^{\text{st}}$ and C^{th} runs is reduced to, respectively, $q^{C-1} - p^{C-1}\frac{q^{C-1}}{q^{C-1}+q^C-p^C}$ and $q^C - p^C - p^{C-1}\frac{q^C-p^C}{q^{C-1}+q^C-p^C}$.

Continuing in this way, we see that for $1 \leq i \leq j \leq C$ there is asymptotically a constant amount of mass r^{ij} matched from the i^{th} run of letters a to the j^{th} run of letters b. The masses r^{ij} satisfy $p^i = \sum_{i \leq j} r^{ij}$, $1 \leq i \leq C$, and $q^j = \sum_{i \leq j} r^{ij}$, $1 \leq j \leq C$, and are determined inductively as follows. Firstly, $r^{CC} = p^C$. Secondly, suppose that r^{ij} has been determined for $k+1 \leq i \leq j \leq C$. For $k+1 \leq j \leq C$ the amount of mass left unmatched in the j^{th} run of letters b is $q^j - \sum_{\ell=k+1}^{j} r^{\ell j}$. It follows that r^{kj}, $k \leq j \leq C$, is determined by

$$r^{kk} = p^k \frac{q^k}{q^k + \sum_{j=k+1}^{C} \left(q^j - \sum_{\ell=k+1}^{j} r^{\ell j}\right)}$$

and

$$r^{kh} = p^k \frac{q^h - \sum_{\ell=k+1}^{h} r^{\ell h}}{q^k + \sum_{j=k+1}^{C} \left(q^j - \sum_{\ell=k+1}^{j} r^{\ell j}\right)}, \quad k+1 \leq h \leq C.$$

Given the uniform random labeled associated admissible matching L_n, write $1 \leq U_{nk} < V_{nk} \leq 2n$, for the pair of indices such that $((U_{nk}, V_{nk}), k) \in L_n$, $1 \leq k \leq n$. That is, k labels a matched pair of indices (U_{nk}, V_{nk}) such that U_{nk} is the index of a letter a and V_{nk} is the index of a letter b. Set $X_{nk} = \frac{U_{nk}}{2n}$ and $Y_{nk} = \frac{V_{nk}}{2n}$, $1 \leq k \leq n$. The finite-dimensional distributions of $((X_{n1}, Y_{n1}), \ldots, (X_{nn}, Y_{nn}))$ converge as $n \to \infty$ to those of the sequence $((X_1, Y_1), (X_2, Y_2), \ldots)$ that is i.i.d. with common distribution the probability measure μ concentrated on $\{(x, y) : 0 \leq x \leq y \leq 1\}$ that is defined as follows:

$$\mu(dx, dy) = \begin{cases} \frac{2^2 r^{ij}}{p^i q^j}\, dx\, dy, & (x, y) \in (\gamma^{2i-2}, \gamma^{2i-1}) \times (\gamma^{2j-1}, \gamma^{2j}),\ 1 \leq i \leq j \leq C, \\ 0, & \text{otherwise.} \end{cases}$$

This tells us that we can build an infinite bridge using μ and that the infinite bridge is the limit of the bridges going to the w_n.

Remark 2.6.8. In Example 2.6.7, for $1 \leq j_1 \leq j_2 \leq C$, define

$$c^{j_1 j_2} := \frac{r^{j_1 j_2}}{r^{j_1 j_1}},$$

then we have that for $1 \leq j_1 = j_2 \leq 2n$

$$c^{j_1 j_2} = 1,$$

and for $1 \leq j_1 < j_2 \leq C$

$$c^{j_1 j_2} = \frac{q^{j_2} - \sum_{l=j_1+1}^{j_2} r^{l j_2}}{q^{j_1}}.$$

We also have that for $1 \leq j_1 < j_2 \leq C$

$$\begin{aligned} \frac{r^{(j_1-1)j_2}}{r^{(j_1-1)j_1}} &= \frac{q^{j_2} - \sum_{l=j_1}^{j_2} r^{l j_2}}{q^{j_1} - r^{j_1 j_1}} \\ &= \frac{q^{j_2} - \sum_{l=j_1+1}^{j_2} r^{l j_2} - r^{j_2 j_2}}{q^{j_1} - r^{j_1 j_1}} \\ &= c^{j_1 j_2}. \end{aligned}$$

Continuing in this way, a backward induction establishes that for $1 \leq i \leq j_1 \leq j_2$ the value of $\frac{r^{i j_2}}{r^{i j_1}}$ does not depend on i and

$$\frac{r^{i j_2}}{r^{i j_1}} = c^{j_1 j_2}.$$

It follows that

$$r^{ij} = r^{i1} c^{1j}, 1 \leq i \leq j \leq n.$$

Consequently,

$$\mu(dx, dy) = \begin{cases} \frac{2r^{i1}}{p^i} \frac{2c^{1j}}{q^j}\, dx\, dy, & (x, y) \in (\gamma^{2i-2}, \gamma^{2i-1}) \times (\gamma^{2j-1}, \gamma^{2j}),\ 1 \leq i \leq j \leq C, \\ 0, & \text{otherwise.} \end{cases}$$

Observe that the measures μ presented in Example 2.6.5 and Example 2.6.7 have the common features that

$$\mu(\{(x,y) : 0 \le x = y \le 1\}) = 0,$$

and that there exist functions $f, g : [0,1] \mapsto \mathbb{R}$ such that

$$\mu(dx, dy) = f(x)g(y)dxdy$$

on the set $\{(x,y) : 0 \le x < y \le 1\}$.

We will see that having these common features is a sufficient condition for the chain $(W_n^\infty)_{n\in\mathbb{N}_0}$ that arises via the construction of Remark 2.5.18 from μ to be an extremal infinite bridge. Let $(L_n^\infty)_{n\in\mathbb{N}_0}$ be the labeled infinite bridge that arises via the construction of Remark 2.5.18 from such measure μ. Given a labeled matching $\mathcal{L}$ of $[2n]$, define $Z_i^{\mathcal{L}} = S_k$, $Z_j^{\mathcal{L}} = T_k$ for each block $\{i,j\}$ labeled with $k \in [n]$ of $\mathcal{L}$ with $i < j$. Then

$$\begin{aligned}
&\mathbb{P}\{L_n^\infty = \mathcal{L}\} \\
&= \mathbb{P}\{Z_1^{\mathcal{L}} < Z_2^{\mathcal{L}} < \ldots < Z_{2n}^{\mathcal{L}}\} \\
&= \int \cdots \int_{z_1^{\mathcal{L}} < z_2^{\mathcal{L}} < \ldots < z_{2n}^{\mathcal{L}}} \prod_{k=1}^{n} f(s_k)g(t_k)ds_1 \cdots ds_n dt_1 \cdots dt_n,
\end{aligned}$$

where the letter $z_i^{\mathcal{L}}$ is substituted by s_k and the letter $z_j^{\mathcal{L}}$ is substituted by t_k for each block $\{i,j\}$ labeled with $k \in [n]$ of $\mathcal{L}$ with $i < j$. Since $\prod_{k=1}^n f(s_k)g(t_k)$ is invariant under permutation of each index set; that is, for permutations σ, τ of $[n]$

$$\prod_{k=1}^{n} f(s_{\sigma(k)})g(t_{\tau(k)}) = \prod_{k=1}^{n} f(s_k)g(t_k),$$

if there are labeled matchings $\mathcal{L}_1, \mathcal{L}_2$ of $[2n]$ such that

$$\{1 \le i \le 2n : (i,j) \text{ is a block of } \mathcal{L}_1\} = \{1 \le i \le 2n : (i,j) \text{ is a block of } \mathcal{L}_2\},$$

or equivalently,

$$\Psi(\mathcal{L}_1) = \Psi(\mathcal{L}_2),$$

then

$$\mathbb{P}\{L_n^\infty = \mathcal{L}_1\} = \mathbb{P}\{L_n^\infty = \mathcal{L}_2\}.$$

This establishes that the conditional distribution of L_n^∞ given $\Psi(L_n^\infty)$ is uniform on the set of labeled associated admissible matchings of $\Psi(L_n^\infty)$ for all $n \in \mathbb{N}_0$ and thus $(W_n^\infty)_{n\in\mathbb{N}_0} := (\Psi(L_n^\infty))_{n\in\mathbb{N}_0}$ is an extremal infinite bridge.

2.7 Extremal harmonic functions from exponential distributions

The Plackett-Luce model for the chain $(U_n)_{n\in\mathbb{N}_0}$

In general, there is no simple closed form expression for the transition probabilities of an infinite bridge $(U_n^\infty)_{n\in\mathbb{N}_0}$ associated with a pair of (not necessarily canonical) diffuse probability measures ζ, η and hence the associated harmonic function h. However, it is possible to obtain such expressions in the special case where ζ is the exponential distribution with rate parameter α and η is the exponential distribution with rate parameter β. Given $u \in \mathbb{W}_n$ and $1 \le i \le 2n$, set

$$\mathbf{A}_i^n(u) := \#\{i \le j \le 2n : u_j = a\}$$

and

$$\mathbf{B}_i^n(u) := \#\{i \le j \le 2n : u_j = b\}.$$

By the reasoning that goes into the analysis of the Plackett-Luce or vase model of random permutations (see, for example, [22]),

$$\mathbb{P}\{U_n^\infty = u\} = (n!)^2\alpha^n\beta^n \prod_{i=1}^{2n} \frac{1}{\mathbf{A}_i^n(u)\alpha + \mathbf{B}_i^n(u)\beta}$$

– this is essentially just repeated applications of the elementary result usually called *competing exponentials*: if S and T are independent exponentially distributed random variables with rate parameters λ and θ, then the probability of the event $\{S < T\}$ is $\frac{\lambda}{\lambda+\theta}$ and conditional on this event the random variables S and $T - S$ are independent and exponentially distributed with rate parameters $\lambda + \theta$ and θ. (As a check, note that when $\alpha = \beta = \gamma$, say, this probability is, as expected, $1/\binom{2n}{n}$.) We also know that

$$\mathbb{P}\{U_n^\infty = u \,|\, U_{n+1}^\infty = v\} = \frac{\binom{v}{u}}{(n+1)^2}.$$

It follows that

$$\begin{aligned}
&\mathbb{P}\{U_{n+1}^\infty = v \,|\, U_n^\infty = u\} \\
&= \frac{\binom{v}{u}}{(n+1)^2}((n+1)!)^2\alpha^{n+1}\beta^{n+1} \prod_{i=1}^{2(n+1)} \frac{1}{\mathbf{A}_i^{n+1}(v)\alpha + \mathbf{B}_i^{n+1}(v)\beta} \\
&\qquad \Big/ (n!)^2\alpha^n\beta^n \prod_{i=1}^{2n} \frac{1}{\mathbf{A}_i^n(u)\alpha + \mathbf{B}_i^n(u)\beta} \\
&= \binom{v}{u}\alpha\beta \frac{\prod_{i=1}^{2n}(\mathbf{A}_i^n(u)\alpha + \mathbf{B}_i^n(u)\beta)}{\prod_{i=1}^{2(n+1)}(\mathbf{A}_i^{n+1}(v)\alpha + \mathbf{B}_i^{n+1}(v)\beta)}.
\end{aligned}$$

As a check, when $\alpha = \beta = \gamma$, say, this transition probability is

$$\binom{v}{u}\frac{(2n)!}{(2(n+1))!} = \frac{\binom{v}{u}}{(2n+2)(2n+1)},$$

as expected.

The corresponding harmonic function h satisfies

$$\binom{v}{u}\alpha\beta\frac{\prod_{i=1}^{2n}(\mathbf{A}_i^n(u)\alpha + \mathbf{B}_i^n(u)\beta)}{\prod_{i=1}^{2(n+1)}(\mathbf{A}_i^{n+1}(v)\alpha + \mathbf{B}_i^{n+1}(v)\beta)}$$
$$= \frac{h(v)}{h(u)}\frac{\binom{v}{u}}{(2n+2)(2n+1)}.$$

We conclude from this that, up to an arbitrary positive constant,

$$h(w) = \frac{(2m)!\alpha^m\beta^m}{\prod_{i=1}^{2m}(\mathbf{A}_i^m(w)\alpha + \mathbf{B}_i^m(w)\beta)}$$

for $w \in \mathbb{W}_n$.

The Plackett-Luce model for the chain $(W_n)_{n\in\mathbb{N}_0}$

We know that any extremal infinite bridge and hence any extremal harmonic function arises via the construction of Remark 2.5.2 from a probability measure η that is concentrated on $\{(x, y) : x \le y\}$. Here we obtain the explicit form of the extremal harmonic function when η is the probability distribution of (U, V) conditional on $U < V$, where U and V are independent exponential random variables with respective rate parameters α and β. It is a familiar fact (sometimes referred to as the property of "competing exponentials") that $\mathbb{P}\{U < V\} = \frac{\alpha}{\alpha+\beta}$ and conditional on the event $\{U < V\}$, the random variables U and $V - U$ are independent with distributions that are exponential with respective parameters $\alpha + \beta$ and β. Therefore,

$$\eta(dx, dy) = (\alpha + \beta)\beta e^{-\alpha x}e^{-\beta y}\,dx\,dy, \quad 0 \le x < y.$$

Suppose that $Z_1, \dots, Z_m$ are independent exponential random variables with respective rate parameters $\gamma_1, \dots, \gamma_m$. For $i \in [m]$ put $\Sigma(i) = j$ if $\#\{k \in [m] : Z_k \le Z_j\} = i$; that is, $\Sigma(i)$ is the index of the i^{th} smallest of the values in a realization of $Z_1, \dots, Z_m$. It follows from the competing exponentials fact that

$$\mathbb{P}\{\Sigma(1) = \sigma(1), \dots, \Sigma(m) = \sigma(m)\}$$
$$= \frac{\gamma_{\sigma(1)}}{\gamma_{\sigma(1)} + \cdots + \gamma_{\sigma(m)}}\frac{\gamma_{\sigma(2)}}{\gamma_{\sigma(2)} + \cdots + \gamma_{\sigma(m)}}\cdots\frac{\gamma_{\sigma(m-1)}}{\gamma_{\sigma(m-1)} + \gamma_{\sigma(m)}}.$$

This is a standard formula for the Plackett-Luce model of random permutations – see [22, Section 5.6].

Now consider independent random variables $U_1, \dots, U_n, V_1, \dots, V_n$, where each U_i has an exponential distribution with parameter α and each V_i has an exponential distribution with parameter β. For $i \in [n]$ and $j \in [n]$ put

$$I(i) := \#\{k \in [n] : U_k \le U_i\} + \#\{\ell \in [n] : V_\ell \le U_i\}$$

and

$$J(j) := \#\{k \in [n] : U_k \le V_j\} + \#\{\ell \in [n] : V_\ell \le V_j\}.$$

A labeled associated admissible matching of a word $w \in \mathbb{B}_n$ is described uniquely by giving the indices $f_1 < f_2 < \cdots < f_n \in [2n]$ of the letters a in the word, the distinct indices $g_1, \dots, g_n \in [2n]$ of letters b with $g_i > f_i$ being the index matched with the index f_i, and a bijection $\pi : [n] \to [n]$, where $\pi(i)$ is the label of the block $\{f_i, g_i\}$ in the matching. Writing $\theta = \frac{\alpha}{\beta}$, we have from the above that

$$\begin{aligned}
&\mathbb{P}\{I(\pi(1)) = f_1,\ J(\pi(1)) = g_1, \dots, I(\pi(n)) = f_n,\ J(\pi(n)) = g_n \,|\, U_1 < V_1, \dots, U_n < V_n\} \\
&\quad = \frac{\alpha^n \beta^n}{\prod_{k=1}^{2n} \left(\#\{s \in [n] : f_s \ge k\}\alpha + \#\{t \in [n] : g_t \ge k\}\beta\right)} \Big/ \left(\frac{\alpha}{\alpha+\beta}\right)^n \\
&\quad = \frac{(1+\theta)^n}{\prod_{k=1}^{2n} \left(\#\{s \in [n] : f_s \ge k\}\theta + \#\{t \in [n] : g_t \ge k\}\right)}.
\end{aligned}$$

Recall from Definition 2.3.11 the height function H defined by the word w and, adopting the convention $H(0) = 0$, note that

$$2n - k + 1 = \#\{s \in [n] : f_s \ge k\} + \#\{t \in [n] : g_t \ge k\}$$

and

$$H(k-1) = (n - \#\{s \in [n] : f_s \ge k\}) - (n - \#\{t \in [n] : g_t \ge k\}).$$

Thus,

$$\#\{s \in [n] : f_s \ge k\} = \frac{1}{2}(2n - k + 1 - H(k-1))$$

and

$$\#\{t \in [n] : g_t \ge k\} = \frac{1}{2}(2n - k + 1 + H(k-1)).$$

Therefore, if $(W_n^\infty)_{n \in \mathbb{N}_0}$ is the infinite bridge associated with the probability measure η, we have by Corollary 2.3.16 (ii)

$$\begin{aligned}
&\mathbb{P}\{W_n^\infty = w\} \\
&\quad = n!\Lambda(w)\mathbb{P}\{I(\pi(1)) = f_1, J(\pi(1)) = g_1, \dots, I(\pi(n)) = f_n, J(\pi(n)) = g_n \\
&\qquad\qquad | \, U_1 < V_1, \dots, U_n < V_n\} \\
&\quad = n!\Lambda(w) \frac{(1+\theta)^n}{\prod_{k=1}^{2n} \left(\frac{1}{2}(2n-k+1-H(k-1))\theta + \frac{1}{2}(2n-k+1+H(k-1))\right)}.
\end{aligned}$$

As a check, consider the case when $\alpha = \beta$ and hence $\theta = 1$. In this case, the righthand side is

$$n!\Lambda(w)\frac{2^n}{(2n)!} = \frac{\Lambda(w)}{(2n-1)!!},$$

as expected from Examples 2.1.5 and 2.6.5 and Proposition 2.3.13.

Recalling the formula we obtained for $\Lambda(w)$ from Lemma 2.3.12, we have

$$\mathbb{P}\{W_n^\infty = w\} = n!\frac{(1+\theta)^n \prod_{i=1}^n H(f_i)}{\prod_{k=1}^{2n} \left(\frac{1}{2}(2n-k+1-H(k-1))\theta + \frac{1}{2}(2n-k+1+H(k-1))\right)}.$$

It follows from (2.3) that

$$h(w) = (2m-1)!!m!\frac{(1+\theta)^m}{\prod_{k=1}^{2m} \left(\frac{1}{2}(2m-k+1-H(k-1))\theta + \frac{1}{2}(2m-k+1+H(k-1))\right)}$$

for $w \in \mathbb{B}_m$. Again as expected, $h \equiv 1$ when $\theta = 1$.

Chapter 3

Mixing time of the Down-up Markov chain of words

3.1 Introduction

In a standard deck of cards, the colors of successive cards (**R**ed or **B**lack) defines a word of length 52 drawn from the two-letter alphabet $\{R, B\}$ in which 26 letters are R and 26 letters are B. Consider a card shuffle that consists of removing a card in the red suits and a card in the black suits independently and uniformly at random followed by inserting the two cards back into uniformly chosen positions in the deck of the remaining 50 cards independently. The resulting transition in the color of successive cards defines a new word of length 52 drawn from the two-letter alphabet $\{R, B\}$ in which 26 letters are R and 26 letters are B.

We can generalize this shuffling mechanism to any number of letters R and letters B. From now on, we will replace the two-letter alphabet $\{R, B\}$ by the two-letter alphabet $\{a, b\}$. Choose a word uniformly at random from the set of words drawn from the two-letter alphabet $\{a, b\}$ that consists of n letters a and n letters b, and denote the set of such words by $\mathbb{W}_n$, $n \in \mathbb{N}_0$. We shuffle the word in $\mathbb{W}_n$ by removing a letter a and a letter b from the word uniformly at random, followed by reinserting the letter a into a uniformly chosen one of $2n-1$ slots defined by the other remaining $2n-2$ letters, then reinstating the letter b into a uniformly chosen one of $2n$ slots defined by the $2n-1$ letters. Write $(DU_t)_{t\in\mathbb{N}_0}$ for a Markov chain with state space $\mathbb{W}_n$ that arises from this removal and reinsertion procedure.

Definition 3.1.1. Define a Markov chain $(DU_t)_{t\in\mathbb{N}_0}$ such that

- DU_0 takes values in $\mathbb{W}_n$,
- conditional on $DU_t = u_1u_2 \cdots u_{2n-1}u_{2n}$, the word DU_t is constructed by choosing I_t and J_t independently and uniformly at random from the index set $\{1 \le i \le 2n : u_i = a\}$ and from the index set $\{1 \le i \le 2n : u_i = b\}$), respectively, choosing $1 \le \widetilde{I}_t, \widetilde{J}_t \le 2n$ with $\widetilde{I}_t \ne \widetilde{J}_t$ uniformly at random (that is, all $2n(2n-1)$ possibilities are equally

likely), placing an a in position $\widetilde{I}_t$ and a b in position $\widetilde{J}_t$, and placing the letters $u_{\sigma(1)}u_{\sigma(2)}\cdots u_{\sigma(2n-2)}$ in that order in the remaining $2n-2$ positions, where σ is the unique increasing bijection from $[2n-2]$ to $[2n]\setminus\{I_t, J_t\}$.

Our shuffling technique is related to random-to-random card shuffling where a card is chosen at random, removed from the deck, and reinserted in a random position. More specifically, random-to-random card shuffle on the deck of n cards consists of picking a card and a position of the deck independently, uniformly at random and moving the chosen card to the chosen position. Bernstein and Nestoridi [2] proved that the random-to-random card shuffle exhibits cutoff at $\frac{3}{4}n\log n$: they proved an upper bound for the mixing time for the random-to-random insertion shuffle of the form $\frac{3}{4}n\log n + cn$ and Subag [28] proved a lower bound for the mixing time of the form $(\frac{3}{4} - o(1))n\log n$.

The Markov chain $(DU_t)_{t\in\mathbb{N}_0}$ that sequentially generates random words such that DU_t is asymptotically uniformly distributed over $\mathbb{W}_n$ can be trivially modified to give a Markov chain $(\widetilde{DU}_t)_{t\in\mathbb{N}_0}$ such that $\widetilde{DU}_t$ is asymptotically uniformly distributed over the set of words drawn from the alphabet $\{a, b\}$ that have $n+1$ letters a and n letters b: adding a letter a into the slot before the first letter of the word $(DU_t)_{t\in\mathbb{N}_0}$. By Theorem 1.1.5 and Theorem 1.1.6 there is a bijection that turns the Markov chain $(\widetilde{DU}_t)_{t\in\mathbb{N}_0}$ into a Markov chain $((K_t, \widetilde{T}_t))_{n\in\mathbb{N}_0}$, where K_t and $\widetilde{T}_t$ are independent for each t, K_t is uniformly distributed on $[n+1]$, and $\widetilde{T}_t$ is uniformly distributed over the set of 1-dominated word with $n+1$ letters a and n letters b as there are $n+1$ cyclic permutations which transforms a 1-dominated word with $n+1$ letters a and n letters b into a word with $n+1$ letters a and n letters b that starts with a. Moreover, composing these two bijections turns the Markov chain $(\widetilde{DU}_t)_{t\in\mathbb{N}_0}$ into a Markov chain $((K_t, T_t))_{t\in\mathbb{N}_0}$, where K_t and T_t are independent for each t, K_t is uniformly distributed on $[n+1]$, and T_t is uniformly distributed over the set of trees with $n+1$ vertices that are equipped with the postfix labeling. In fact, removing a letter a and a letter b from DU_t is a equivalent of deleting a vertex and an edge from the tree that corresponds to the word DU_t, and reinserting those two letters a and b back to the word is a equivalent of adding a vertex and an edge to the tree.

There is a clear relationship between the transition dynamics of the Markov chains $(DU_t)_{t\in\mathbb{N}_0}$ and $(U_n)_{n\in\mathbb{N}_0}$. Recall from Theorem 2.3.9 that the backward transition dynamics of $(U_n)_{n\in\mathbb{N}_0}$ consists of removing an a and a b independently and uniformly at random at each step and that the forward transition dynamics of $(U_n)_{n\in\mathbb{N}_0}$ consists of adding an a and a b independently into uniformly chosen one of slots defined by the letters of the word at each step. Therefore, the forward transition dynamics of the Markov chain $(DU_t)_{t\in\mathbb{N}_0}$ is the same as the backward transition dynamics of $(U_n)_{n\in\mathbb{N}_0}$ followed by the forward transition dynamics of $(U_n)_{n\in\mathbb{N}_0}$.

The Markov chain $(DU_t)_{t\in\mathbb{N}_0}$ has an attractive feature that if DU_t is uniformly distributed over the set $\mathbb{W}_n$, then DU_{t+1} is also uniformly distributed over $\mathbb{W}_n$: in other words, the stationary distribution is the uniform distribution over $\mathbb{W}_n$. Using total variation distance as a metric on the space of probabilities on $\mathbb{W}_n$, we study the mixing time for the chain $(DU_t)_{t\in\mathbb{N}_0}$.

3.2 Background on Mixing times

We review some parts of general theory for mixing times that are essential to understand further discussion. The primary reference for this section is [19, Chapter 4] and [9] where the reader may find the original source of strong stationary times and a wide range of methods to estimate the rate of convergence of a Markov chain to its stationary distribution.

Consider a discrete Markov chain $(X_t)_{t\in\mathbb{N}_0}$ with the finite state space Ω and transition matrix P that is irreducible and aperiodic.

Definition 3.2.1. A probability π on Ω is a *stationary distribution* for the Markov chain $(X_t)_{t\in\mathbb{N}_0}$ if

$$\pi(y) = \sum_{x\in\Omega} \pi(x)P(x,y), y \in \Omega.$$

It is a classical result that Markov chain $(X_t)_{t\in\mathbb{N}_0}$ has a unique stationary distribution π, and the distribution of X_t converges to its stationary distribution π as $t \to \infty$ with respect to the metric known as the *total variation distance*.

Definition 3.2.2. *Total variation distance* between two probability distributions μ and ν on a measurable space $(\Omega, \mathscr{A})$ is defined by

$$||\mu - \nu||_{TV} := \sup_{A\in\mathscr{A}} |\mu(A) - \nu(A)|.$$

If the space Ω is a finite set and the σ-algebra $\mathscr{A}$ is the maximal σ-algebra on Ω consisting of all subsets of Ω, then Definition 3.2.2 can be re-formulated:

$$||\mu - \nu||_{TV} := \max_{A\subseteq\Omega} |\mu(A) - \nu(A)|.$$

There are other probabilistic interpretations of total variation distance for a finite set Ω. More specifically, Proposition 3.2.3 reduces total variation distance to a simple sum over the state space Ω. Proposition 3.2.4 suggests a different probabilistic approach to total variation distance using coupling: $||\mu - \nu||_{TV}$ measures how close to identical two random variables realizing μ and ν can be forced to be.

Proposition 3.2.3. *Let μ and ν be two probability distributions on a finite set Ω. Then*

$$||\mu - \nu||_{TV} = \frac{1}{2} \sum_{x\in\Omega} |\mu(x) - \nu(x)|.$$

Proposition 3.2.4. *Let μ and ν be two probability distributions on Ω. A coupling of probability distributions μ and ν is a pair of random variables (X, Y), defined on a single probability space, such that the marginal distribution of X is μ and the marginal distribution of Y is ν. Then*

$$||\mu - \nu||_{TV} = \inf\{\mathbb{P}\{X \neq Y\} : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}.$$

Definition 3.2.5. Let $\mathcal{P}$ be the set of all probability distributions on the state space of the Markov chain. Denote the supremum of the total variation distance to its stationary distribution π over all possible initial distributions by

$$d(t) := \sup_{\mu \in \mathcal{P}} \left\| \mu P^t - \pi \right\|_{TV}.$$

Note that $d(t)$ is a non-increasing function of t.

It is useful to introduce a parameter which measures the time required by $(X_t)_{t \in \mathbb{N}_0}$ for the total variation distance to its stationarity distribution π to be small.

Definition 3.2.6. The mixing time $t_{\text{mix}}(\epsilon)$, $\epsilon > 0$, of a Markov chain is defined by

$$t_{\text{mix}}(\epsilon) := \min\{t : d(t) \leq \epsilon\}.$$

Set

$$t_{\text{mix}} := t_{\text{mix}}\big(1/4\big).$$

Along with coupling, *strong stationary times* are effective methods for obtaining upper bounds on the total variation distance to the stationary distributoin.

Definition 3.2.7. A random variable τ taking values in $\mathbb{N}_0 \cup \{\infty\}$ is said to be a *stopping time relative to a filtration* $(\mathscr{F}_t)_{t \in \mathbb{N}_0}$ if

$$\{\tau = t\} \in \mathscr{F}_t, \; t \in \mathbb{N}_0.$$

Especially, a stopping time τ relative to a filtration $(\mathscr{F}_t)_{t \in \mathbb{N}_0}$ is said to be a *randomized stopping time* for the Markov chain $(X_t)_{t \in \mathbb{N}_0}$ if the filtration $(\mathscr{F}_t)_{t \in \mathbb{N}_0}$ is larger than the canonical filtration $(\sigma(X_0, X_1, \ldots, X_t))_{t \in \mathbb{N}_0}$; that is,

$$\sigma(X_0, X_1, \ldots, X_t) \subseteq \mathscr{F}_t, \; t \in \mathbb{N}_0.$$

Definition 3.2.8. A *stationary time* τ for $(X_t)_{t \in \mathbb{N}_0}$ is a randomized stopping time such that the distribution of X_τ is the stationary distribution π;

$$\mathbb{P}_x\{X_\tau \in A\} = \pi(A), \; A \subseteq \Omega.$$

Definition 3.2.9. A *strong stationary time* for a Markov chain $(X_t)_{t \in \mathbb{N}_0}$ is a randomized stopping time τ such that

$$\mathbb{P}_x\{\tau = t, X_\tau \in A\} = \mathbb{P}_x\{\tau = t\}\pi(A), \; A \subseteq \Omega,$$

that is, τ is a stationary time for $(X_t)_{t \in \mathbb{N}_0}$ and is independent of X_τ.

Remark 3.2.10. In fact, since Ω is a finite space, Definition 3.2.9 and Definition 3.2.8 can be simplified: a randomized stopping time τ is a stationary time if

$$\mathbb{P}_x\{X_\tau = y\} = \pi(y),\ x, y \in \Omega.$$

Moreover, a randomized stopping time τ is a strong stationary time if

$$\mathbb{P}_x\{\tau = t, X_\tau = y\} = \mathbb{P}_x\{\tau = t\}\pi(y),\ x, y \in \Omega.$$

Theorem 3.2.11 shows that a strong stationary time provides an upper bound on $d(t)$, and thus an upper bound on the mixing time t_{mix}.

Theorem 3.2.11. *If τ is a strong stationary time for $(X_t)_{t\in\mathbb{N}_0}$, then*

$$d(t) = \max_x ||\mathbb{P}_x(X_t \in \cdot) - \pi(\cdot)||_{TV} \le \max_x \mathbb{P}_x\{\tau > t\}.$$

3.3 The stationary distribution for $(DU_t)_{t\in\mathbb{N}_0}$

Definition 3.3.1. Let $\mathbb{W}_n$ be the set of words drawn from two-letter alphabet $\{a, b\}$ that consist of n letters a and n letters b.

In order to write down the transition probabilities for the Markov chain $(DU_t)_{t\in\mathbb{N}_0}$, it is convenient to introduce the following standard notation (see, for example, [20]).

Definition 3.3.2. Given two words w and v drawn from some finite alphabet, write $\binom{w}{v}$ for the number of times that v appears as a sub-word of w.

Example 3.3.3. For example, $\binom{abbaba}{abb} = 3$ because *abb* appears inside *abbaba* as a sub-word three times:

abbaba **ab**ba**b**a **a**b**b**a**b**a.

By definition, the Markov chain $(DU_t)_{t\in\mathbb{N}_0}$ has state space $\mathbb{W}_n$ and one-step transition probabilities

$$\begin{aligned}\mathbb{P}\{DU_{t+1} = v \mid DU_t = u\} &= \sum_{\substack{w\in\mathbb{W}_{n-1}\\ w|u, w|v}} \binom{u}{w}\frac{1}{n^2}\cdot\binom{v}{w}\frac{1}{(2n-1)(2n)}\\ &= \sum_{\substack{w\in\mathbb{W}_{n-1}\\ w|u, w|v}} \frac{1}{2n^3(2n-1)}\binom{u}{w}\binom{v}{w}, u, v \in \mathbb{W}_n,\end{aligned}$$

where $w|u$ (resp. $w|v$) means that w is a subword of u (resp. v).

Definition 3.3.4. Write π_n for the uniform distribution over $\mathbb{W}_n$.

Theorem 3.3.5. *The stationary distribution of the Down-up Markov chain* $(DU_t)_{t\in\mathbb{N}}$ *is the uniform distribution* π_n *over* $\mathbb{W}_n$.

Proof. As the transition probability $\mathbb{P}\{DU_{t+1} = v \mid DU_t = u\}$ is the same as the transition probability $\mathbb{P}\{DU_{t+1} = u \mid DU_t = v\}$, the detailed balance equations are satisfied by the uniform distribution π_n on $\mathbb{W}_n$; that is,

$$\mathbb{P}\{DU_{t+1} = v \mid DU_t = u\}\pi_n(u) = \mathbb{P}\{DU_{t+1} = u \mid DU_t = v\}\pi_n(v),\ u, v \in \mathbb{W}_n.$$

□

3.4 Coupling

We will construct a Markov chain $(M_t)_{t\in\mathbb{N}_0}$ that has the same distribution as the Markov chain $(DU_t)_{t\in\mathbb{N}_0}$. Consider a random word M_0 in $\mathbb{W}_n$, and write M_0^k for the k^{th} letter of M_0; that is,

$$M_0 = M_0^1 M_0^2 \cdots M_0^{2n}.$$

Definition 3.4.1. Define two index sets $\mathbb{I}$ and $\mathbb{J}$ for the letters a and the letters b, respectively, in the word M_0 such that

$$\mathbb{I} := \{1 \le k \le 2n : M_0^k = a\},$$

$$\mathbb{J} := \{1 \le k \le 2n : M_0^k = b\}.$$

Definition 3.4.2. Construct a Markov chain $(V_t)_{t\in\mathbb{N}_0}$ with the state space $[0,1]^{2n}$ as follows

- $V_0 = (V_0^1, V_0^2, \ldots, V_0^{2n})$, where $V_0^1, V_0^2, \ldots, V_0^{2n}$ are i.i.d. $U[0,1]$ random variables,
- sort the realizations of $V_0^1, V_0^2, \ldots, V_0^{2n}$ in increasing order, and write σ for the unique permutation of $[2n]$ such that

$$V_0^{\sigma(1)} < V_0^{\sigma(2)} < \cdots < V_0^{\sigma(2n-1)} < V_0^{\sigma(2n)},$$

- conditional on $V_t = (V_t^1, V_t^2, \ldots, V_t^{2n})$, the sequence $V_{t+1} = (V_{t+1}^1, V_{t+1}^2, \ldots, V_{t+1}^{2n})$ is constructed by taking i.i.d $U[0,1]$ variables V_t^a, V_t^b, choosing I_t, J_t independently and uniformly at random from the index sets $\mathbb{I}$ and $\mathbb{J}$, respectively, and putting

$$V_{t+1}^{\sigma(I_t)} = V_t^a,$$

$$V_{t+1}^{\sigma(J_t)} = V_t^b,$$

$$V_{t+1}^k = V_t^k,\ 1 \le k \le 2n \text{ with } k \ne \sigma(I_t), \sigma(J_t).$$

Definition 3.4.3. Define a random sequence $(M_t)_{t\in\mathbb{N}_0}$ in $\mathbb{W}_n$ such that M_t is built from V_t by laying down $V_t^1, V_t^2, \ldots, V_t^{2n}$ in increasing order and replacing $V_t^{\sigma(k)}$ by M_0^k for $1 \leq k \leq 2n$.

At each time $t \in \mathbb{N}_0$ we associate the random variable $V_t^{\sigma(k)}$ with the k^{th} letter M_0^k of M_0, $1 \leq k \leq 2n$. Thus, $V_t^{\sigma(k)}$ is associated with a letter a (resp. b) in M_0 if $k \in \mathbb{I}$ (resp. $k \in \mathbb{J}$). Updating the random variables $V_{t+1}^{\sigma(I_t)}$ and $V_{t+1}^{\sigma(J_t)}$ relocates the letter a associated with $V_t^{\sigma(I_t)}$ and the letter b associated with $V_t^{\sigma(J_t)}$ in the word M_t into uniformly chosen one of the slots defined by the other $2n-2$ letters while constructing M_{t+1} from M_t. Moreover, the associated letter a (resp. b) is uniformly chosen from the n letters a (resp. b) of M_t because I_t (resp. J_t) is chosen uniformly at random from $\mathbb{I}$ (resp. $\mathbb{J}$). This establishes that the forward transition dynamics of $(M_t)_{t\in\mathbb{N}_0}$ is the same as those of $(DU_t)_{t\in\mathbb{N}_0}$. Consequently, $(M_t)_{t\in\mathbb{N}_0}$ is also a Markov chain and the mixing time for the chain $(M_t)_{t\in\mathbb{N}_0}$ is the mixing time for the chain $(DU_t)_{t\in\mathbb{N}_0}$.

Theorem 3.4.4. *The mixing time for the chain $(M_t)_{t\in\mathbb{N}_0}$ is the mixing time for the chain $(DU_t)_{t\in\mathbb{N}_0}$.*

We will therefore study the mixing time for the chain $(M_t)_{t\in\mathbb{N}_0}$ from now on.

3.5 An upper bound on the mixing time

Definition 3.5.1. Define a randomized stopping time τ_{ab} for the chain $(M_t)_{t\in\mathbb{N}_0}$ by

$$\tau_{ab} := \min\left\{t : \bigcup_{1\leq s\leq t} \{I_s, J_s\} = [2n]\right\}.$$

In other words, τ_{ab} is the first time when all $2n$ letters of M_0 have had their associated $U[0,1]$ random variable updated at least once. Write $\mathscr{G}_t$ for the σ-algebra $\sigma(I_s, J_s, M_s, 1 \leq s \leq t)$, then $\{\tau_{ab} = t\} \in \mathscr{G}_t$, $t \in \mathbb{N}_0$. Thus, τ_{ab} is a randomized stopping time for the Markov chain $(M_t)_{t\in\mathbb{N}_0}$ relative to the augmented filtration $(\mathscr{G}_t)_{t\in\mathbb{N}_0}$.

Lemma 3.5.2. *The randomized stopping time τ_{ab} is a strong stationary time for the Markov chain $(M_t)_{t\in\mathbb{N}_0}$.*

Proof. By construction, conditional on τ_{ab}, for $1 \leq k \leq 2n$ the random variable $V_0^{\sigma(k)}$ has been updated to a new $U[0,1]$ random variable $V_0^{\sigma(k)}$ at some time $1 \leq t \leq \tau_{ab}$. Let t_k be the most recent time before the time τ_{ab} such that $V_{t_k}^{\sigma(k)}$ was updated to a new $U[0,1]$ random variable. By definition,

$$V_{t_k}^{\sigma(k)} = V_{\tau_{ab}}^{\sigma(k)},\ 1 \leq k \leq 2n.$$

Since $V_{t_1}^{\sigma(1)}, V_{t_2}^{\sigma(2)}, \ldots, V_{t_{2n}}^{\sigma(2n)}$ are i.i.d. $U[0,1]$ random variables, $V_{\tau_{ab}}^{\sigma(1)}, V_{\tau_{ab}}^{\sigma(2)}, \ldots, V_{\tau_{ab}}^{\sigma(2n)}$ are also conditionally i.i.d. $U[0,1]$ random variables conditional on τ_{ab}. Consequently, the order of $V_{\tau_{ab}}^{\sigma(1)}, V_{\tau_{ab}}^{\sigma(2)}, \ldots, V_{\tau_{ab}}^{\sigma(2n)}$ is uniformly distributed over the $(2n)!$ possible orders given τ_{ab} so that the conditional distribution of $M_{\tau_{ab}}$ given τ_{ab} is the uniform distribution π_n over $\mathbb{W}_n$. Therefore,

$$\begin{aligned} &\mathbb{P}\{M_t = w, \tau_{ab} = t\} \\ &= \mathbb{P}\{M_t = w | \tau_{ab} = t\} \mathbb{P}\{\tau_{ab} = t\} \\ &= \pi_n(w) \mathbb{P}\{\tau_{ab} = t\}, w \in \mathbb{W}_n, \end{aligned}$$

as required. □

Definition 3.5.3. Write τ_a (resp. τ_b) for the first time when the n letters a (resp. b) have had their associated $U[0,1]$ random variable updated at least once; that is,

$$\tau_a := \min\{t : \{I_s : 1 \leq s \leq t\} = \mathbb{I}\},$$

$$\tau_b := \min\{t : \{J_s : 1 \leq s \leq t\} = \mathbb{J}\}.$$

Observe that, by definition,

$$\tau_{ab} = \max\{\tau_a, \tau_b\}.$$

Remark 3.5.4. Moreover, recall that each time t, I_t and J_t are chosen independently and uniformly at random from $\mathbb{I}$ and $\mathbb{J}$, respectively. The dynamics of choosing an index I_t (resp. J_t) from $\mathbb{I}$ (resp. $\mathbb{J}$) is just the same as the dynamics of choosing a coupon from an urn containing n different coupons in the coupon collector's game. Thus, the distribution of τ_a (resp. τ_b) corresponds to the distribution of the coupon collector's time when he first collects every one of the n types of coupons.

As a result, we can apply the following general bound on the tail of the coupon collector time (see, for example, [19] for the proof) to τ_a and τ_b.

Theorem 3.5.5. *Consider a collector attempting to collect a complete set of n coupons. Assume that each time a new coupon is chosen uniformly and independently from the set of n possible types, and let τ be the first time when the set contains every type. Then, for $c > 0$,*

$$\mathbb{P}\{\tau > \lfloor n \log n + cn \rfloor\} \leq e^{-c},\ n \in \mathbb{N}.$$

Corollary 3.5.6. *For $c > 0$,*

$$\mathbb{P}\{\tau_a > \lfloor n \log n + cn \rfloor\} \leq e^{-c}, n \in \mathbb{N}.$$

$$\mathbb{P}\{\tau_b > \lfloor n \log n + cn \rfloor\} \leq e^{-c}, n \in \mathbb{N}.$$

Lemma 3.5.7. *For $c > 0$,*

$$\mathbb{P}\{\tau_{ab} > \lfloor n \log n + cn \rfloor\} \leq 2e^{-c}.$$

Proof. As $\tau_{ab} = \max(\tau_a, \tau_b)$, for $c > 0$

$$\begin{aligned}\mathbb{P}\{\tau_{ab} > \lfloor n\log n + cn\rfloor\} &= \mathbb{P}\{\max\{\tau_a, \tau_b\} > \lfloor n\log n + cn\rfloor\}\\ &\leq \mathbb{P}\{\tau_a > \lfloor n\log n + cn\rfloor\} + \mathbb{P}\{\tau_b > \lfloor n\log n + cn\rfloor\}\\ &\leq e^{-c} + e^{-c}\\ &= 2e^{-c}.\end{aligned}$$

□

Theorem 3.5.8.

$$t_{\text{mix}} \leq n\log n + (\log 8)n.$$

Proof. It follows from Theorem 3.2.11 that the strong stationary time τ_{ab} gives

$$d(t) \leq \max_{w\in\mathbb{W}_n} \mathbb{P}_w\{\tau_{ab} > t\},\ t > 0.$$

Applying this inequality with $\lfloor n\log n + cn\rfloor$ instead of t, together with Lemma 3.5.7, establishes that, for $c > 0$,

$$\begin{aligned}&d(\lfloor n\log n + cn\rfloor)\\ &\leq \max_{w\in\mathbb{W}_n} \mathbb{P}_w\{\tau_{ab} > \lfloor n\log n + cn\rfloor\}\\ &\leq 2e^{-c}.\end{aligned}$$

Consequently, for $c = \log 8$,

$$d(\lfloor n\log n + (\log 8)n\rfloor) \leq 2e^{-\log 8} = \frac{1}{4},$$

as desired. □

3.6 A lower bound on the mixing time

Definition 3.6.1. For a word $w = w_1 \dots w_{2n} \in \mathbb{W}_n$, define its *height function*

$$H_w(t) := \#\{1 \leq i \leq t : w_i = a\} - \#\{1 \leq j \leq t : w_j = b\}, \quad 1 \leq t \leq 2n.$$

Definition 3.6.2. Define functions $H_{\max}, H_{\min}, H_{\text{extr}}$ on $\mathbb{W}_n$ such that

$$\begin{aligned}H_{\max}(w) &:= \max_{1\leq t\leq 2n} H_w(t), w \in \mathbb{W}_n,\\ H_{\min}(w) &:= \min_{1\leq t\leq 2n} H_w(t), w \in \mathbb{W}_n,\\ H_{\text{extr}}(w) &:= \max_{1\leq t\leq 2n} |H_w(t)|, w \in \mathbb{W}_n.\end{aligned}$$

If we represent a word w in $\mathbb{W}_n$ by a polygonal line with segments joining $(t-1, H_w(t-1))$ and $(t, H_w(t)), 1 \leq t \leq 2n$, then $H_{\text{extr}}(w)$ is the maximum vertical distance between the resulting path and the straight line joining $(0,0)$ and $(2n, 0)$. By definition,

$$H_{\text{extr}}(w) = \max\{H_{\max}(w), -H_{\min}(w)\}.$$

Definition 3.6.3. Write $\mathbb{W}_n^\alpha$ for the set of words w in $\mathbb{W}_n$ such that $H_{\text{extr}}(w) > \frac{1}{3}n^\alpha$.

Lemma 3.6.4. *For $0 < \alpha < 1$, there exists a constant $K_\alpha > \frac{3e^2\pi}{2}$ such that*

$$\pi_n(\mathbb{W}_n^\alpha) \leq K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}, n \in \mathbb{N}_0.$$

Proof. By union bound,

$$\begin{aligned}
&\pi_n(\mathbb{W}_n^\alpha) \\
&= \pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) > \frac{1}{3}n^\alpha\}) \\
&= \pi_n(\{w \in \mathbb{W}_n : \max\{H_{\max}(w), -H_{\min}(w)\} > \frac{1}{3}n^\alpha\}) \\
&\leq \pi_n(\{w \in \mathbb{W}_n : H_{\max}(w) > \frac{1}{3}n^\alpha\}) + \pi_n(\{w \in \mathbb{W}_n : H_{\min}(w) < -\frac{1}{3}n^\alpha\}).
\end{aligned}$$

Represent a word w in $\mathbb{W}_n$ by a polygonal graph with segments joining $(t-1, H_w(t-1))$ and $(t, H_w(t)), 1 \leq t \leq 2n$. Let a *path* be a polygonal graph that can arise by this geometric representation of a word in $\mathbb{W}_n$. Then the number of words w such that $H_{\max}(w) > \frac{1}{3}n^\alpha$ is the same as the number of paths from $(0,0)$ and $(2n, 0)$ whose height is equal to $\lfloor \frac{1}{3}n^\alpha \rfloor + 1$ at some time $1 \leq t \leq 2n$. By the reasoning that goes into reflection principle for a simple random walk (see, for example, [12]) the number of such paths corresponds to the number of paths from $(0,0)$ to $(2n, 2\lfloor \frac{1}{3}n^\alpha \rfloor + 2)$. Thus,

$$\begin{aligned}
\pi_n(\{w \in \mathbb{W}_n : H_{\max}(w) > \frac{1}{3}n^\alpha\}) &= \binom{2n}{n + \lfloor \frac{1}{3}n^\alpha \rfloor + 1} \Big/ \binom{2n}{n} \\
&= \frac{n!n!}{(n + \lfloor \frac{1}{3}n^\alpha \rfloor + 1)!(n - \lfloor \frac{1}{3}n^\alpha \rfloor - 1)!}.
\end{aligned}$$

We recall Stirling's formula:

$$\sqrt{2\pi n}(\frac{n}{e})^n \leq n! \leq e\sqrt{n}(\frac{n}{e})^n, n \in \mathbb{N}.$$

It follows from Stirling's formula that

$$\begin{aligned}
&\pi_n(\{w \in \mathbb{W}_n : H_{\max}(w) > \frac{1}{3}n^\alpha\}) \\
&\leq \left(e^2 n^{2n+1}\right) \Big/ \left(2\pi \left(n + \lfloor \frac{1}{3}n^\alpha \rfloor + 1\right)^{n + \lfloor \frac{1}{3}n^\alpha \rfloor + \frac{3}{2}} \left(n - \lfloor \frac{1}{3}n^\alpha \rfloor - 1\right)^{n - \lfloor \frac{1}{3}n^\alpha \rfloor - \frac{1}{2}}\right).
\end{aligned}$$

Observe that

$$\frac{\partial\left(\log\left((n+x)^{n+x+\frac{1}{2}}(n-x)^{n-x+\frac{1}{2}}\right)\right)}{\partial x} = \log(n+x) + \frac{1}{2(n+x)} - \log(n-x) - \frac{1}{2(n-x)} \\ \geq 0$$

because

$$\frac{\partial(\log y + \frac{1}{2y})}{\partial y} = \frac{1}{y} - \frac{1}{2y^2} \geq 0,\ y > \frac{1}{2}.$$

Thus, $(n+x)^{n+x+\frac{1}{2}}(n-x)^{n-x+\frac{1}{2}}$ is an increasing function of x on the set $\{x \in \mathbb{R} : 0 < x < n - \frac{1}{2}\}$. Since $\lfloor \frac{1}{3}n^\alpha \rfloor + 1 \geq \frac{1}{3}n^\alpha$,

$$\begin{aligned}
&\pi_n(\{w \in \mathbb{W}_n : H_{\max}(w) > \frac{1}{3}n^\alpha\}) \\
&\leq \left(e^2 n^{2n+1}\right) \Big/ \left(2\pi\left(n + \frac{1}{3}n^\alpha\right)^{n+\frac{1}{3}n^\alpha+\frac{1}{2}}\left(n - \frac{1}{3}n^\alpha\right)^{n-\frac{1}{3}n^\alpha+\frac{1}{2}}\right) \\
&= e^2 \Big/ \left(2\pi\left(1 + \frac{1}{3}n^{\alpha-1}\right)^{n+\frac{1}{3}n^\alpha+\frac{1}{2}}\left(1 - \frac{1}{3}n^{\alpha-1}\right)^{n-\frac{1}{3}n^\alpha+\frac{1}{2}}\right) \\
&= e^2 \Big/ \left(2\pi\sqrt{1 - \frac{1}{9n^{2(1-\alpha)}}}\left(\left(1 - \frac{1}{9n^{2(1-\alpha)}}\right)^{-9n^{2(1-\alpha)}}\right)^{-\frac{1}{9}n^{2\alpha-1}+\frac{1}{27}n^{3\alpha-2}}\left(\left(1 + \frac{1}{3n^{1-\alpha}}\right)^{3n^{1-\alpha}}\right)^{\frac{2}{9}n^{2\alpha-1}}\right).
\end{aligned}$$

Since $(1 + \frac{1}{3n^{1-\alpha}})^{3n^{1-\alpha}}$ is an increasing function of n that converges to e as $n \to \infty$, and $(1 - \frac{1}{9n^{2(1-\alpha)}})^{-9n^{2(1-\alpha)}}$ is a decreasing function of n that converges to e as $n \to \infty$, there exists $N_\alpha \in \mathbb{N}$ such that for $n \geq N_\alpha$

$$\begin{aligned}
\sqrt{1 - \frac{1}{9n^{2(1-\alpha)}}} &\geq \frac{2}{3}, \\
\left(1 + \frac{1}{3n^{1-\alpha}}\right)^{3n^{1-\alpha}} &\geq e^{\frac{3}{4}}, \\
\left(1 - \frac{1}{9n^{2(1-\alpha)}}\right)^{-9n^{2(1-\alpha)}} &\leq e^{\frac{5}{4}}.
\end{aligned}$$

Consequently, for $n \geq N_\alpha$

$$\begin{aligned}
\pi_n(\{w \in \mathbb{W}_n : H_{\max}(w) > \frac{1}{3}n^\alpha\}) &\leq e^2 \Big/ \left(\frac{4}{3}\pi\left(e^{\frac{5}{4}}\right)^{-n^{2\alpha-1}+n^{3\alpha-2}}\left(e^{\frac{3}{4}}\right)^{2n^{2\alpha-1}}\right) \\
&\leq \frac{3\pi}{4}e^{(-\frac{1}{4}n^{2\alpha-1}+2)}.
\end{aligned}$$

This establishes that there exists a constant $K_\alpha := \max\{\frac{3e^2\pi}{2}, 2e^{\frac{1}{4}N_\alpha^{2\alpha-1}}\}$ such that

$$\pi_n(\{w \in \mathbb{W}_n : H_{\max}(w) > \frac{1}{3}n^\alpha\}) \le \frac{K_\alpha}{2}e^{-\frac{1}{4}n^{2\alpha-1}}, n \in \mathbb{N}_0.$$

By symmetry, a similar argument shows that

$$\pi_n(\{w \in \mathbb{W}_n : H_{\min}(w) < -\frac{1}{3}n^\alpha\}) \le \frac{K_\alpha}{2}e^{-\frac{1}{4}n^{2\alpha-1}}, n \in \mathbb{N}_0.$$

It follows that

$$\begin{aligned}\pi_n(\mathbb{W}_n^\alpha) &\le \pi_n(\{w \in \mathbb{W}_n : H_{\max}(w) > \frac{1}{3}n^\alpha\}) + \pi_n(\{w \in \mathbb{W}_n : H_{\min}(w) < -\frac{1}{3}n^\alpha\}) \\ &\le K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}, n \in \mathbb{N}_0.\end{aligned}$$

□

Definition 3.6.5. Define randomized stopping times τ_a^α and τ_b^α for the Markov chain $(M_t)_{t\in\mathbb{N}_0}$ by

$$\begin{aligned}\tau_a^\alpha &:= \min\{t : |\{I_s : 1 \le s \le t\}| = n - \lfloor n^\alpha \rfloor\}, \\ \tau_b^\alpha &:= \min\{t : |\{J_s : 1 \le s \le t\}| = n - \lfloor n^\alpha \rfloor\}.\end{aligned}$$

Definition 3.6.6. Define a randomized stopping time τ^α for the Markov chain $(M_t)_{t\in\mathbb{N}_0}$ by

$$\tau_{ab}^\alpha := \min\{\tau_a^\alpha, \tau_b^\alpha\}.$$

By definition, τ_a^α (resp. τ_b^α) is the first time when $(n - \lfloor n^\alpha \rfloor)$ letters a (resp. b) have had their associated $U[0,1]$ random variables updated, and τ_{ab}^α is the first time when either $(n - \lfloor n^\alpha \rfloor)$ letters a or $(n - \lfloor n^\alpha \rfloor)$ letters b have had their associated $U[0,1]$ random variables updated in the construction of $(M_t)_{t\in\mathbb{N}_0}$. It is guaranteed that until time τ_{ab}^α there are at least $\lfloor n^\alpha \rfloor$ letters a and at least $\lfloor n^\alpha \rfloor$ letters b that have not been relocated and have kept their relative order.

As explained in Remark 3.5.4, taking an index I_t and an index J_t uniformly at random from $\mathbb{I}$ and $\mathbb{J}$, respectively, is just the same as choosing a coupon from the urn containing n types of coupons. Consequently, τ_a^α and τ_b^α have the same distribution with the first time when $(n - \lfloor n^\alpha \rfloor)$ types of coupons have been collected in the coupon collector's game.

Lemma 3.6.7. *Consider a collector attempting to collect a set of coupons from an urn contain one of each of n types of coupons. Assume that each time a new coupon is chosen uniformly and independently from the the set of n possible types, and let τ^α be the first time when the collector's set contains $(n - \lfloor n^\alpha \rfloor)$ distinct types. Then*

$$|\mathbb{E}[\tau^\alpha] - (1-\alpha)n\log n| \le n \log\left(\frac{n^{\alpha+1}}{(n^\alpha - 1)(n+1)}\right),$$

$$\mathrm{Var}(\tau^\alpha) \le n^{2-\alpha}.$$

Proof. For $1 \leq m \leq n$, write T_m for the number of coupons collected after the set first contains $m-1$ distinct types until the set first contains m distinct types. Then $(T_m)_{m\in N}$ is a sequence of independent geometric random variables such that T_m has the geometric distribution with the probability of success $(n-m+1)/n$ for $1 \leq m \leq n$. Thus,

$$\begin{aligned}\mathbb{E}[\tau^\alpha] &= \mathbb{E}[T_1 + T_2 + \ldots + T_{n-\lfloor n^\alpha \rfloor}] \\ &= \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{n}{n-m+1} \\ &= n \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{1}{n-m+1}.\end{aligned}$$

Observe that

$$n \int_{\lfloor n^\alpha \rfloor +1}^{n+1} \frac{1}{x}\, dx \;\leq n \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{1}{n-m+1} \;\leq n \int_{\lfloor n^\alpha \rfloor}^{n} \frac{1}{x}\, dx,$$

or equivalently,

$$n\big(\log(n+1) - \log(\lfloor n^\alpha \rfloor + 1)\big) \;\leq n \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{1}{n-m+1} \;\leq n\big(\log n - \log(\lfloor n^\alpha \rfloor)\big).$$

As $\frac{1}{x}$ is a decreasing function of x, we have

$$n \int_{\lfloor n^\alpha \rfloor +1}^{n+1} \frac{1}{x}\, dx \;\leq n \int_{n^\alpha}^{n+(n^\alpha - \lfloor n^\alpha \rfloor)} \frac{1}{x}\, dx \;\leq n \int_{\lfloor n^\alpha \rfloor}^{n} \frac{1}{x}\, dx,$$

equivalently,

$$n\big(\log(n+1) - \log(\lfloor n^\alpha \rfloor + 1)\big) \;\leq n\big(\log(n + (n^\alpha - \lfloor n^\alpha \rfloor)) - \log(n^\alpha)\big) \;\leq n\big(\log n - \log(\lfloor n^\alpha \rfloor)\big).$$

Thus, the distance between $n \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{1}{n-m+1}$ and $n\big(\log(n + (n^\alpha - \lfloor n^\alpha \rfloor)) - \log(n^\alpha)\big)$ is upper bounded as follows:

$$\begin{aligned}&\left| n\big(\log(n + (n^\alpha - \lfloor n^\alpha \rfloor) - \log(n^\alpha)\big) - n \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{1}{n-m+1} \right| \\ &\leq n\big(\log n - \log(\lfloor n^\alpha \rfloor)\big) - n\big(\log(n+1) - \log(\lfloor n^\alpha \rfloor + 1)\big).\end{aligned}$$

It follows from the triangle inequality that

$$\begin{aligned}
&\left| n\big(\log n - \log(n^\alpha)\big) - n \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{1}{n-m+1} \right| \\
&\leq \left| n\big(\log(n + (n^\alpha - \lfloor n^\alpha \rfloor) - \log(n^\alpha)\big) - n \sum_{m=1}^{n-\lfloor n^\alpha \rfloor} \frac{1}{n-m+1} \right| + n\log(n + (n^\alpha - \lfloor n^\alpha \rfloor) - n\log n \\
&\leq n\big(\log n - \log(\lfloor n^\alpha \rfloor)\big) - n\big(\log(n+1) - \log(\lfloor n^\alpha \rfloor + 1)\big) + n\big(\log(n + (n^\alpha - \lfloor n^\alpha \rfloor) - \log n\big) \\
&= n\log\left(\frac{\lfloor n^\alpha \rfloor + 1}{\lfloor n^\alpha \rfloor} \cdot \frac{n + (n^\alpha - \lfloor n^\alpha \rfloor}{n+1}\right) \\
&\leq n\log\left(\frac{n(n^\alpha)}{(n^\alpha - 1)(n+1)}\right) \\
&= n\log\left(\frac{n^{\alpha+1}}{(n^\alpha - 1)(n+1)}\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathrm{Var}(\tau^\alpha) &= \mathrm{Var}(T_1 + T_2 + T_3 + \cdots + T_n) \\
&= n \sum_{m=2}^{n-\lfloor n^\alpha \rfloor} \frac{(m-1)}{(n-m+1)^2} \\
&\leq n \int_1^{n-n^\alpha} \frac{t}{(n-t)^2}\, dt \\
&= n\left[\log(n-t) + n\left(\frac{1}{n-t}\right)\right]_1^{n-n^\alpha} \\
&= n\left(\alpha \log\frac{n}{n-1} - \frac{n}{n-1} + n^{1-\alpha}\right) \\
&\leq n^{2-\alpha}.
\end{aligned}$$

□

It follows from Chebychev's inequality that the stopping time τ^α is highly likely to be concentrated near its expectation within the distance of order n. More specifically

$$\begin{aligned}
&\mathbb{P}\{\tau^{\alpha} \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n\} \\
&= \mathbb{P}\{\tau^{\alpha} - E[\tau^{\alpha}] \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n - E[\tau^{\alpha}]\} \\
&\leq \mathbb{P}\{|\tau^{\alpha} - E[\tau^{\alpha}]| \geq cn\} \\
&\leq \frac{\mathrm{Var}(\tau^{\alpha})}{c^2 n^2} \\
&\leq \frac{1}{c^2 n^{\alpha}}.
\end{aligned}$$

Corollary 3.6.8. *For $c > 0$,*

$$\mathbb{P}\{\tau_a^{\alpha} \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n\} \leq \frac{1}{c^2 n^{\alpha}},$$

$$\mathbb{P}\{\tau_b^{\alpha} \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n\} \leq \frac{1}{c^2 n^{\alpha}},$$

$$\mathbb{P}\{\tau_{ab}^{\alpha} \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n\} \leq \frac{2}{c^2 n^{\alpha}}.$$

Proof. The first two inequalities are immediate from Lemma 3.6.7. We have

$$\tau_{ab}^{\alpha} = \min\{\tau_a^{\alpha}, \tau_b^{\alpha}\}.$$

It follows from the preceding inequalities and a union bound that

$$\begin{aligned}
&\mathbb{P}\{\tau_{ab}^{\alpha} \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n\} \\
&= \mathbb{P}\{\tau_a^{\alpha} \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n\} \\
&\quad + \mathbb{P}\{\tau_b^{\alpha} \leq (1-\alpha)n\log n - \big(\log\left(\frac{n^{\alpha+1}}{(n^{\alpha}-1)(n+1)}\right) + c\big)n\} \\
&\leq \frac{2}{c^2 n^{\alpha}}.
\end{aligned}$$

□

Lemma 3.6.9. *Define $w_* := \underbrace{a\cdots a}_{n}\underbrace{b\cdots b}_{n}$ in $\mathbb{W}_n$ to be the word that consists of n successive letters a followed by n successive letters b. Then, for $\alpha \geq \frac{1}{2}$,*

$$\mathbb{P}_{w_*}\{H_{\mathrm{extr}}(M_t) > \frac{2}{3}n^{\alpha} \mid \tau_{ab}^{\alpha} \geq t\} \geq 1 - K_{\alpha}e^{-\frac{1}{4}n^{2\alpha-1}},$$

where $K_{\alpha} > \frac{3e^2\pi}{2}$ is the constant from Lemma 3.6.4.

Proof. Assume that the Markov chain $(M_t)_{t\in\mathbb{N}_0}$ starts with the word w_*, then by definition of τ_{ab}^{α}, there exist at least $\lfloor n^\alpha \rfloor$ letters a and $\lfloor n^\alpha \rfloor$ letters b that have not had their associated $U[0,1]$ random variables updated by the time τ_{ab}^{α}. Conditional on $\{\tau_{ab}^{\alpha} \geq t\}$, take $\lfloor n^\alpha \rfloor$ letters a and $\lfloor n^\alpha \rfloor$ letters b from the set of such letters to build a subword u_t^α of M_t and construct a subword v_t^α of M_t with the remaining $(n - \lfloor n^\alpha \rfloor)$ letters a and the remaining $(n - \lfloor n^\alpha \rfloor)$ letters b of M_t. As the $2\lfloor n^\alpha \rfloor$ letters of u_t^α have kept their original relative order, u_t^α is almost surely $\lfloor n^\alpha \rfloor$ letters a followed by $\lfloor n^\alpha \rfloor$ letters b. Consequently, conditional on $\{\tau_{ab}^{\alpha} \geq t\}$,

$$H_{\min}(u_t^\alpha) = 0,$$
$$H_{\max}(u_t^\alpha) = \lfloor n^\alpha \rfloor.$$

Suppose that the word v_t^α contains n_a letters a and n_b letters b with $n_a, n_b \geq 0$ that have not had their associated $U[0,1]$ random variables updated before time t. Let $k_1, \ldots, k_{2n}$ be a listing of $[2n]$ such that

- $V_t^{\sigma(k_s)}$, $1 \leq s \leq n_a$, is associated to one of the n_a letters a,
- $V_t^{\sigma(k_s)}$, $n_a + 1 \leq s \leq n_a + n_b$, is associated to one of the n_b letters b,
- $\{k_{n_a+n_b+1}, \ldots, k_{n+n_b}\} = [n] \setminus \{k_1, \ldots, k_{n_a}\}$,
- $\{k_{n+n_b+1}, \ldots, k_{2n}\} = \{n+1, \ldots, 2n\} \setminus \{k_{n_a+1}, \ldots, k_{n_a+n_b}\}$.

Since $M_0 = w_*$, the indices $k_1, \ldots, k_{n_a}$ are chosen uniformly without replacement from the set $\mathbb{I} = [n]$ and likewise the indices $k_{n_a+1}, \ldots, k_{n_a+n_b}$ are chosen uniformly without replacement from the set $\mathbb{J} = \{n+1, \ldots, 2n\}$. By definition, either $V_t^{\sigma(k_s)}$ is associated with a letter a for $1 \leq s \leq n_a$ or $n_a + n_b + 1 \leq s \leq n + n_b$. Similarly, either $V_t^{\sigma(k_s)}$ is associated with a letter b for $n_a + 1 \leq s \leq n_a + n_b$ or $n + n_b + 1 \leq s \leq 2n$. Moreover, for $1 \leq s \leq n_a + n_b$, the associated uniform variables have not been updated and thus $V_t^{\sigma(k_s)} = V_0^{\sigma(k_s)}$ is the k_s-th order statistics of the $2n$ i.i.d. $U[0,1]$ random variables $V_0^1, \ldots, V_0^{2n}$.

Using a coupling argument, we will prove that

$$\mathbb{P}\{H_{\min}(v_t^\alpha) < x \mid \tau_{ab}^{\alpha} \geq t\} \leq \pi_n\{w \in \mathbb{W}_{n-\lfloor n^\alpha \rfloor} : H_{\min}(w) < x\}, x \in \mathbb{R}.$$

Take i.i.d $U[0,1]$ random variables $U_1, \ldots, U_{4n-n_a-n_b}$. Write $\widetilde{U}_1 < \ldots < \widetilde{U}_{2n}$ for the order statistics of $U_{2n-n_a-n_b+1}, \ldots, U_{4n-n_a-n_b}$. Pick $l_1, \ldots, l_{n_a+n_b}$ uniformly at random without replacement from the index set $[2n]$ and construct a sequence $\tilde{k}_1, \ldots, \tilde{k}_{n_a+n_b}$ as follows.

- put $\tilde{k}_s := l_s$ for $s \in \mathbb{S}_a$, where $\mathbb{S}_a := \{1 \leq s \leq n_a : l_s \in [n]\}$,
- put $\tilde{k}_s := l_s$ for $s \in \mathbb{S}_b$, where $\mathbb{S}_b := \{n_a + 1 \leq s \leq n_a + n_b : l_s \in \{n+1, \ldots, 2n\}\}$,
- choose $\tilde{k}_s, s \in \{1, \ldots, n_a\} \setminus \mathbb{S}_a$, uniformly at random without replacement from $[n] \setminus \{l_s : s \in \mathbb{S}_a\}$,

- choose $\tilde{k}_s, s \in \{n_a+1,\dots,n_a+n_b\} \setminus \mathbb{S}_b$, uniformly at random without replacement from $\{n+1,\dots,2n\} \setminus \{l_s : s \in \mathbb{S}_b\}$.

Define

$$\begin{aligned}
\widetilde{V}_t^{\sigma(k_s)} &:= \widetilde{U}_{\tilde{k}_s}, 1 \le s \le n_a+n_b,\\
\widetilde{V}_t^{\sigma(k_s)} &:= U_{s-n_a-n_b}, n_a+n_b+1 \le s \le 2n.\\
\overline{V}_t^{\sigma(k_s)} &:= \widetilde{U}_{l_s}, 1 \le s \le n_a+n_b,\\
\overline{V}_t^{\sigma(k_s)} &:= U_{s-n_a-n_b}, n_a+n_b+1 \le s \le 2n.
\end{aligned}$$

Because $l_1,\dots,l_{n_a+n_b}$ are uniformly chosen without replacement from the index set $[2n]$, $\overline{V}_t^{\sigma(k_s)}, 1 \le s \le 2n$, are i.i.d. $U[0,1]$ random variables. Consequently, if we construct a word $\tilde{w}_t^\alpha$ by laying down $\overline{V}_t^1,\dots,\overline{V}_t^{2n}$ in order and replacing $\overline{V}_t^{\sigma(k_s)}$ by a (resp. b) for $1 \le s \le n$ (resp. $n+1 \le s \le 2n$), the n letters a and the n letters b are in the uniform order and $\tilde{w}_t^\alpha$ is uniformly distributed over $\mathbb{W}_{n-\lfloor n^\alpha \rfloor}$. Moreover, $\widetilde{V}_t^{\sigma(k_s)}, 1 \le s \le n_a+n_b$, is the $\tilde{k}_s$-th order statistics of i.i.d. $U[0,1]$ random variables $U_{2n-n_a-n_b+1},\dots,U_{4n-n_a-n_b}$. By construction we have

$$(k_1,\dots,k_{n_a+n_b}) \stackrel{d}{=} (\tilde{k}_1,\dots,\tilde{k}_{n_a.+n_b}),$$

and therefore

$$\left(\widetilde{V}_t^1,\dots,\widetilde{V}_t^{2n}\right) \stackrel{d}{=} \left(V_t^1,\dots,V_t^{2n}\right).$$

It follows that if we construct a word $\tilde{v}_t^\alpha$ by laying down $\widetilde{V}_t^1,\dots,\widetilde{V}_t^{2n}$ in order and replacing $\widetilde{V}_t^{\sigma(k_s)}$ by a (resp. b) for $1 \le s \le n$ (resp. $n+1 \le s \le 2n$), then $\tilde{v}_t^\alpha \stackrel{d}{=} v_t^\alpha$.

By definition, we also have that

$$\begin{aligned}
&\tilde{k}_s \le l_s,\ 1 \le s \le n_a,\\
&\tilde{k}_s \ge l_s,\ n_a+1 \le s \le n_a+n_b,
\end{aligned}$$

and thus

$$\begin{aligned}
\widetilde{V}_t^{\sigma(k_j)} &\le \overline{V}_t^{\sigma(k_j)},\ 1 \le j \le n_a,\\
\widetilde{V}_t^{\sigma(k_j)} &\ge \overline{V}_t^{\sigma(k_j)},\ n_a+1 \le j \le n_a+n_b,\\
\widetilde{V}_t^{\sigma(k_j)} &= \overline{V}_t^{\sigma(k_j)},\ n_a+n_b+1 \le j \le 2n.
\end{aligned}$$

Consequently, each letter a is located closer to the beginning in the word $\tilde{v}_t^\alpha$ than in the word $\tilde{w}_t^\alpha$ and that each letter b is located closer to the end in the word $\tilde{v}_t^\alpha$ than in the word $\tilde{w}_t^\alpha$, and therefore

$$H_{\min}(\tilde{v}_t^\alpha) \ge H_{\min}(\tilde{w}_t^\alpha).$$

For $x \in \mathbb{R}$

$$\begin{aligned}
\mathbb{P}\{H_{\min}(v_t^\alpha) < x |\ \tau_{ab}^\alpha \ge t\} &= \mathbb{P}\{H_{\min}(\tilde{v}_t^\alpha) < x\}\\
&\le \mathbb{P}\{H_{\min}(\tilde{w}_t^\alpha) < x\}\\
&= \pi_n\{w \in \mathbb{W}_{n-\lfloor n^\alpha \rfloor} : H_{\min}(w) < x\}.
\end{aligned}$$

In other words, the minimum of the height of v_t^α stochastically dominates the minimum of the height of a word of the same length when all of the associated uniform random variables have been updated.

By an argument similar to that in the proof for Lemma 3.6.4, if we represent a word in $\mathbb{W}_{n-\lfloor n^\alpha \rfloor}$ by a path, then the number of words w in $\mathbb{W}_{n-\lfloor n^\alpha \rfloor}$ such that $H_{\min}(w) < -\frac{1}{3}n^\alpha$ is the same as the number of paths from $(0,0)$ to $(2n-2\lfloor n^\alpha \rfloor, -2\lfloor \frac{1}{3}n^\alpha \rfloor - 2)$. Therefore,

$$\begin{aligned}
\pi_{n-\lfloor n^\alpha \rfloor}(\{w \in \mathbb{W}_{n-\lfloor n^\alpha \rfloor} : H_{\min}(w) < -\frac{1}{3}n^\alpha\}) &= \binom{2n-2\lfloor n^\alpha \rfloor}{n-2\lfloor \frac{1}{3}n^\alpha \rfloor - 1} \Big/ \binom{2n-2\lfloor n^\alpha \rfloor}{n-\lfloor n^\alpha \rfloor} \\
&\leq \binom{2n}{n-\lfloor \frac{1}{3}n^\alpha \rfloor - 1} \Big/ \binom{2n}{n} \\
&= \pi_n(\{w \in \mathbb{W}_n : H_{\min}(w) < -\frac{1}{3}n^\alpha\}).
\end{aligned}$$

It follows from Lemma 3.6.4 and the stochastic dominance that, for $\alpha \geq \frac{1}{2}$,

$$\begin{aligned}
\mathbb{P}_{w_*}\{H_{\min}(v_t^\alpha) < -\frac{1}{3}n^\alpha \mid \tau_{ab}^\alpha \geq t\} &\leq \pi_{n-\lfloor n^\alpha \rfloor}(\{w \in \mathbb{W}_{n-\lfloor n^\alpha \rfloor} : H_{\min}(w) < -\frac{1}{3}n^\alpha\}) \\
&\leq \pi_n(\{w \in \mathbb{W}_n : H_{\min}(w) < -\frac{1}{3}n^\alpha\}) \\
&\leq \pi_n(\{w \in \mathbb{W}_n : H_{\mathrm{extr}}(w) > \frac{1}{3}n^\alpha\}) \\
&\leq K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}.
\end{aligned}$$

Since the word M_t^α can be obtained by interleaving the words u_t^α and v_t^α,

$$\begin{aligned}
H_{\max}(M_t^\alpha) &\geq H_{\max}(u_t^\alpha) + H_{\min}(v_t^\alpha) \\
&= n^\alpha + H_{\min}(v_t^\alpha).
\end{aligned}$$

It follows that, for $\alpha \geq \frac{1}{2}$,

$$\begin{aligned}
\mathbb{P}_{w_*}\{H_{\mathrm{extr}}(M_t) \geq \frac{2}{3}n^\alpha \mid \tau_{ab}^\alpha \geq t\} &\geq \mathbb{P}_{w_*}\{H_{\max}(M_t) \geq \frac{2}{3}n^\alpha \mid \tau_{ab}^\alpha \geq t\} \\
&\geq \mathbb{P}_{w_*}\{n^\alpha + H_{\min}(v_t^\alpha) \geq \frac{2}{3}n^\alpha \mid \tau_{ab}^\alpha \geq t\} \\
&\geq \mathbb{P}_{w_*}\{H_{\min}(v_t^\alpha) \geq -\frac{1}{3}n^\alpha \mid \tau_{ab}^\alpha \geq t\}. \\
&\geq 1 - K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}.
\end{aligned}$$

□

Theorem 3.6.10. *For $\alpha > \frac{1}{2}$, there exists a constant $c_\alpha > 0$ such that*

$$t_{\mathrm{mix}} \geq (1-\alpha)n\log n - c_\alpha n,$$

where $K_\alpha > \frac{3e^2\pi}{2}$ is the constant from Lemma 3.6.4.

Proof.

$$\begin{aligned}
&\mathbb{P}_{w_*}\{H_{\text{extr}}(M_t) \geq \frac{2}{3}n^\alpha\} - \pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) \geq \frac{2}{3}n^\alpha\}) \\
&\geq \Big(\mathbb{P}_{w_*}\{H_{extr}(M_t) \geq \frac{2}{3}n^\alpha,\ \tau_{ab}^\alpha \geq t\} - \pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) \geq \frac{2}{3}n^\alpha\})\mathbb{P}\{\tau_{ab}^\alpha \geq t\}\Big) \\
&\quad - \Big|\mathbb{P}_{w_*}\{H_{\text{extr}}(M_t) \geq \frac{2}{3}n^\alpha,\ \tau_{ab}^\alpha < t\} - \pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) \geq \frac{2}{3}n^\alpha\})\mathbb{P}\{\tau_{ab}^\alpha < t\}\Big|.
\end{aligned}$$

It follows from Lemma 3.6.9 and Lemma 3.6.4 that there exists $K_\alpha > \frac{3e^2\pi}{2}$ such that

$$\begin{aligned}
&\mathbb{P}_{w_*}\{H_{\text{extr}}(M_t) \geq \frac{2}{3}n^\alpha,\ \tau_{ab}^\alpha \geq t\} - \pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) \geq \frac{2}{3}n^\alpha\})\mathbb{P}_{w_*}\{\tau_{ab}^\alpha \geq t\} \\
&\geq (1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}})\mathbb{P}_{w_*}\{\tau_{ab}^\alpha \geq t\}.
\end{aligned}$$

We also have

$$\begin{aligned}
&\Big|\mathbb{P}_{w_*}\{H_{\text{extr}}(M_t) \geq \frac{2}{3}n^\alpha,\ \tau_{ab}^\alpha < t\} - \pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) \geq \frac{2}{3}n^\alpha\})\mathbb{P}\{\tau_{ab}^\alpha < t\}\Big| \\
&\leq \Big|\mathbb{P}_{w_*}\{H_{\text{extr}}(M_t) \geq \frac{2}{3}n^\alpha,\ \tau_{ab}^\alpha < t\}\Big| + \Big|\pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) \geq \frac{2}{3}n^\alpha\})\mathbb{P}\{\tau_{ab}^\alpha < t\}\Big| \\
&\leq 2\mathbb{P}\{\tau_{ab}^\alpha < t\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
d(t) &\geq \mathbb{P}_{w_*}\{H_{\text{extr}}(M_t) \geq \frac{2}{3}n^\alpha\} - \pi_n(\{w \in \mathbb{W}_n : H_{\text{extr}}(w) \geq \frac{2}{3}n^\alpha\}) \\
&\geq (1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}})\mathbb{P}_{w_*}\{\tau_{ab}^\alpha \geq t\} - 2\mathbb{P}_{w_*}\{\tau_{ab}^\alpha < t\} \\
&= (1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}})(1 - \mathbb{P}_{w_*}\{\tau_{ab}^\alpha < t\}) - 2\mathbb{P}_{w_*}\{\tau_{ab}^\alpha < t\} \\
&= 1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}} - (3 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}})\mathbb{P}_{w_*}\{\tau_{ab}^\alpha < t\} \\
&\geq 1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}} - \Big|3 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}\Big|\mathbb{P}_{w_*}\{\tau_{ab}^\alpha < t\}.
\end{aligned}$$

Put $t = (1-\alpha)n\log n - \left(\log\left(\frac{n^{\alpha+1}}{(n^\alpha-1)(n+1)}\right) + c\right)n$ for $c > 0$. It follows from Corollary 3.6.8 that

$$\begin{aligned}
&d\Big((1-\alpha)n\log n - \Big(\log\Big(\frac{n^{\alpha+1}}{(n^\alpha-1)(n+1)}\Big) + c\Big)n\Big) \\
&\geq 1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}} - \Big|3 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}\Big|\mathbb{P}_{w_*}\Big\{\tau_{ab}^\alpha < (1-\alpha)n\log n - \Big(\log\Big(\frac{n^{\alpha+1}}{(n^\alpha-1)(n+1)}\Big) + c\Big)n\Big\} \\
&\geq 1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}} - \Big|3 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}\Big|\frac{2}{c^2n^\alpha}.
\end{aligned}$$

Since for $\alpha > \frac{1}{2}$

$$\lim_{n\to\infty} 1 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}} - \left|3 - 2K_\alpha e^{-\frac{1}{4}n^{2\alpha-1}}\right| \frac{1}{c^2 n^\alpha} = 1,$$

$$\lim_{n\to\infty} \log\left(\frac{n^{\alpha+1}}{(n^\alpha - 1)(n+1)}\right) = 0,$$

there exists a constant $c_\alpha > 0$ such that

$$t_{\text{mix}} \geq (1-\alpha) n \log n - c_\alpha n.$$

□

Corollary 3.6.11.

$$\frac{1}{2} \leq \liminf_{n\to\infty} \frac{t_{\text{mix}}}{n \log n} \leq \limsup_{n\to\infty} \frac{t_{\text{mix}}}{n \log n} \leq 1.$$

Proof. It is a immediate result of Theorem 3.6.10 and Theorem 3.5.8 that for $\alpha > \frac{1}{2}$

$$(1-\alpha) + c_\alpha \frac{1}{\log n} \leq \frac{t_{\text{mix}}}{n \log n} \leq 1 + \frac{\log 8}{\log n}.$$

Thus, by sending $n \to \infty$, we have for any $\alpha > \frac{1}{2}$

$$(1-\alpha) \leq \liminf_{n\to\infty} \frac{t_{\text{mix}}}{n \log n} \leq \limsup_{n\to\infty} \frac{t_{\text{mix}}}{n \log n} \leq 1.$$

It is established by sending $\alpha \to \infty$ that

$$\frac{1}{2} \leq \liminf_{n\to\infty} \frac{t_{\text{mix}}}{n \log n} \leq \limsup_{n\to\infty} \frac{t_{\text{mix}}}{n \log n} \leq 1.$$

□

Bibliography

[1] Laurent Alonso and René Schott. *Random generation of trees*. Random generators in computer science. Kluwer Academic Publishers, Boston, MA, 1995, pp. x+208. ISBN: 0-7923-9528-X. URL: https://doi.org/10.1007/978-1-4757-6353-9.

[2] M. Bernstein and E. Nestoridi. "Cutoff for random to random card shuffle". In: *ArXiv e-prints* (2017). arXiv: 1703.06210 [math.PR].

[3] David Blackwell and David Kendall. "The martin boundary of Pólya's urn scheme, and an application to stochastic population growth". In: *J. Appl. Probability* 1 (1964), pp. 284–296. ISSN: 0021-9002.

[4] Jonathan M. Borwein et al. "Combinatorial aspects of multiple zeta values". In: *Electron. J. Combin.* 5 (1998), Research Paper 38, 12. ISSN: 1077-8926. URL: http://www.combinatorics.org/Volume_5/Abstracts/v5i1r38.html.

[5] Pierre Cartier. "A primer of Hopf algebras". In: *Frontiers in number theory, physics, and geometry. II*. Springer, Berlin, 2007, pp. 537–615. URL: https://doi.org/10.1007/978-3-540-30308-4_12.

[6] Hye Soo Choi and Steven N. Evans. "Doob-Martin compactification of a Markov chain for growing random words sequentially". In: *Stochastic Process. Appl.* 127.7 (2017), pp. 2428–2445. ISSN: 0304-4149. URL: https://doi.org/10.1016/j.spa.2016.11.006.

[7] A. Claesson. "Subword counting and the incidence algebra". In: *ArXiv e-prints* (2015). arXiv: 1502.03065 [math.CO].

[8] M. D. Crossley. "Some Hopf algebras of words". In: *Glasg. Math. J.* 48.3 (2006), pp. 575–582. ISSN: 0017-0895. URL: https://doi.org/10.1017/S0017089506003302.

[9] Persi Diaconis and James Allen Fill. "Strong stationary times via a new form of duality". In: *Ann. Probab.* 18.4 (1990), pp. 1483–1522. ISSN: 0091-1798. URL: http://links.jstor.org/sici?sici=0091-1798(199010)18:4%3C1483:SSTVAN%3E2.0.CO;2-#&origin=MSN.

[10] Persi Diaconis, C. Y. Amy Pang, and Arun Ram. "Hopf algebras and Markov chains: two examples and a theory". In: *J. Algebraic Combin.* 39.3 (2014), pp. 527–585. ISSN: 0925-9899. URL: https://doi.org/10.1007/s10801-013-0456-7.

[11] J. L. Doob. "Discrete potential theory and boundaries". In: *J. Math. Mech.* 8 (1959), 433–458, erratum 993.

[12] Rick Durrett. *Probability: theory and examples*. Fourth. Vol. 31. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010, pp. x+428. ISBN: 978-0-521-76539-8. URL: https://doi.org/10.1017/CBO9780511779398.

[13] Steven N. Evans, Rudolf Grübel, and Anton Wakolbinger. "Doob-Martin boundary of Rémy's tree growth chain". In: *Ann. Probab.* 45.1 (2017), pp. 225–277. ISSN: 0091-1798. URL: https://doi.org/10.1214/16-AOP1112.

[14] Steven N. Evans and Anton Wakolbinger. "Radix sort trees in the large". In: *Electron. Commun. Probab.* 22 (2017), 13 pp. DOI: 10.1214/17-ECP77. URL: https://doi.org/10.1214/17-ECP77.

[15] Hans Föllmer. "Phase transition and Martin boundary". In: (1975), 305–317. Lecture Notes in Math., Vol. 465.

[16] Olav Kallenberg. *Foundations of modern probability*. Second. Probability and its Applications (New York). Springer-Verlag, New York, 2002, pp. xx+638. ISBN: 0-387-95313-2. URL: https://doi.org/10.1007/978-1-4757-4015-8.

[17] Olav Kallenberg. *Probabilistic symmetries and invariance principles*. Probability and its Applications (New York). Springer, New York, 2005, pp. xii+510. ISBN: 978-0387-25115-8; 0-387-25115-4.

[18] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp. *Denumerable Markov chains*. Second. With a chapter on Markov random fields, by David Griffeath, Graduate Texts in Mathematics, No. 40. Springer-Verlag, New York-Heidelberg-Berlin, 1976, pp. xii+484.

[19] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. With a chapter by James G. Propp and David B. Wilson. American Mathematical Society, Providence, RI, 2009, pp. xviii+371. ISBN: 978-0-8218-4739-8.

[20] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. With a foreword by Roger Lyndon and a preface by Dominique Perrin, Corrected reprint of the 1983 original, with a new preface by Perrin. Cambridge University Press, Cambridge, 1997, pp. xviii+238. ISBN: 0-521-59924-5. URL: https://doi.org/10.1017/CBO9780511566097.

[21] Jean-Gabriel Luque and Jean-Yves Thibon. "Pfaffian and Hafnian identities in shuffle algebras". In: *Adv. in Appl. Math.* 29.4 (2002), pp. 620–646. ISSN: 0196-8858. URL: https://doi.org/10.1016/S0196-8858(02)00036-2.

[22] John I. Marden. *Analyzing and modeling rank data*. Vol. 64. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1995, pp. xiv+329. ISBN: 0-412-99521-2.

[23] Christophe Reutenauer. *Free Lie algebras*. Vol. 7. London Mathematical Society Monographs. New Series. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1993, pp. xviii+269. ISBN: 0-19-853679-8.

[24] D. Revuz. *Markov chains*. North-Holland Mathematical Library, Vol. 11. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, 1975, pp. x+336.

[25] Stanley A. Sawyer. "Martin boundaries and random walks". In: *Harmonic functions on trees and buildings (New York, 1995)*. Vol. 206. Contemp. Math. Amer. Math. Soc., Providence, RI, 1997, pp. 17–44. URL: https://doi.org/10.1090/conm/206/02685.

[26] M. P. Schützenberger. "On context-free languages and push-down automata". In: *Information and Control* 6 (1963), pp. 246–264. ISSN: 0890-5401.

[27] Richard P. Stanley. *Catalan numbers*. Cambridge University Press, New York, 2015, pp. viii+215. ISBN: 978-1-107-42774-7; 978-1-107-07509-2. URL: https://doi.org/10.1017/CBO9781139871495.

[28] Eliran Subag. "A lower bound for the mixing time of the random-to-random insertions shuffle". In: *Electron. J. Probab.* 18 (2013), no. 20, 20. ISSN: 1083-6489. URL: https://doi.org/10.1214/EJP.v18-1950.

[29] Wolfgang Woess. *Denumerable Markov chains*. EMS Textbooks in Mathematics. Generating functions, boundary theory, random walks on trees. European Mathematical Society (EMS), Zürich, 2009, pp. xviii+351. ISBN: 978-3-03719-071-5. URL: https://doi.org/10.4171/071.