TOPICS ON THE LENGTH OF THE LONGEST COMMON SUBSEQUENCES, WITH BLOCKS, IN BINARY RANDOM WORDS

A Dissertation
Presented to
The Academic Faculty

By

Yuze Zhang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology

December 2019

TOPICS ON THE LENGTH OF THE LONGEST COMMON SUBSEQUENCES, WITH BLOCKS, IN BINARY RANDOM WORDS

Approved by:

Professor Christian Houdré, Advisor
School of Mathematics
Georgia Institute of Technology

Professor Robert D. Foley
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology

Professor Michael Damron
School of Mathematics
Georgia Institute of Technology

Professor Konstantin Tikhomirov
School of Mathematics
Georgia Institute of Technology

Professor Jack Hanson
Department of Mathematics
City College of New York

Date Approved: August 8, 2019

Dedication to Public Interests, Acquisition of All-round Capability, Aspiration for Progress with Each Passing Day.

Motto of Tianjin Nankai High School

To my family.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my greatest gratitude to my Ph.D. advisor Prof. Christian Houdré, for his consistent guidance and support throughout my Ph.D. study at Georgia Tech. I will always be grateful for all his inspiring insights on research topics, invaluable advice during the difficulties along the way, and great help on my academic writing. Nothing in this thesis could have been achieved without his extraordinary supervision over the years.

I would also like to record much appreciation to Prof. Robert Foley, Prof. Vladimir Koltchinskii, Prof. Michael Damron, Prof. Konstantin Tikhomirov and Prof. Jack Hanson for serving as my committee members, as well as to Dr. Ruoting Gong, for conversations from which I learned a lot of valuable suggestions on, and interesting possible extensions of, my research.

I am grateful for all the academic and financial support provided by the School of Mathematics at Georgia Tech over the years. I would like to express my sincere appreciation to every professor who taught me classes in the probability and statistics area. In addition, I also wanted to offer my appreciation to Ms. Cathy Jacobson and Dr. Burke Moreg, whose training classes on oral English and American culture at the very beginning of the program significantly helped me with my communication skills, teaching ability and career development. Also I would like to thank Ms. Klara Grodzinsky for all her help on teaching arrangements over the years.

I am fortunate to have made so many friends within the School of Mathematics and I want to thank them for their many beneficial advice and discussions on research, including Tongzhou Chen, Rundong Du, Juntao Duan, Dawei He, Qingqing Liu, Jinyong Ma, Chen Xu, Xin Wang, Fan Zhou and so many others that I am not able to enumerate them all. My thanks also go to Xu Du, Ruian Duan, Chenyou Fan, Chuanfeng Guo, Mengfan Jiang, Tong Jin and Zihao Qu. Our friendship has always been valued to me and has made all the days in

Atlanta more colorful. Besides, I also wanted to thank Di Wu and Rong Yuan from JD.com America, who served as my mentors during my summer internship in Mountain View, for their research guidance on quantile regression for demand forecasting I performed there.

Finally, I want to particularly express my great gratitude to my family. To my parents, Jingang Zhang and Lin Zhao, and my grandma Wenzhao Wang, for all their unconditional love and support since I was born. And to my wife, Yanan Li, for the tremendous love and happiness you brought into my life and for making me a better man. This thesis is dedicated to my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The study of LI_n, the length of the longest increasing subsequences, and of LCI_n, the length of the longest common and increasing subsequences in random words is classical in computer science and bioinformatics, and has been well explored over the last few decades. This dissertation studies a generalization of LCI_n for two binary random words, namely, it analyzes the asymptotic behavior of $LCbB_n$, the length of the longest common subsequences containing a fixed number, b, of blocks. We first prove that after proper centerings and scalings, $LCbB_n$, for two sequences of i.i.d. Bernoulli random variables with possibly two different parameters, converges in law towards limits we identify. This dissertation also includes an alternative approach to the one-sequence LbB_n problem, and Monte-Carlo simulations on the asymptotics of $LCbB_n$ and on the growth order of the limiting functional, as well as several extensions of the $LCbB_n$ problem to the Markov context and some connection with percolation theory.

CHAPTER 1

INTRODUCTION AND BACKGROUND

Over the past few decades, the study of LI_n, the length of the longest increasing subsequences, and of LCI_n, the length of the longest common and increasing subsequences in random words has been well explored, starting with [31]. After proper centering and normalization, the limiting law of LI_n, in a finite totally ordered alphabet, can be expressed as the maximal eigenvalue of some Gaussian random matrix, which in turn can also be interpreted as the law of a Brownian functional, akin to a functional one introduced in Queuing Theory (see [31], [42], [28], [25], [26], [3], [18], [17], [22], [33], [9], [24]). More recently, LCI_n is also shown to have a limiting distribution which can be represented as a Brownian functional (see [20], [10], [14]).

It is straightforward to think of the LI_n problem in the following way: To find the longest increasing subsequences of $(X_n)_{n\geq 1}$ on a finite ordered alphabet $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < ... < \alpha_m\}$, is to first divide the whole sequence into consecutive (possibly empty) blocks, and to count the number of occurrences of α_i in the i-th block, for $1 \leq i \leq m$, then to take the sum, and finally to find the maximum over all the consecutive possible divisions of the word. A follow up idea, which is one of the main motivations for the present work, is to investigate loosening some of the increasing requirements above. To be more specific, first, in LI_n it is naturally assumed that the number of blocks b, is identical to the size of the alphabet m. However, following are worth exploring cases, for example, when b is not equal to but still depends on m, both could be finite or simultaneously grow to infinity with some order of n^α where α small, or b and m could be totally independent; we may also pick and count the occurrences of any letter in the i-th block, not necessarily the α_i. Moreover, any letter can be counted more than once in different blocks, which is also prohibited in

LI_n. Clearly, then, LI_n becomes a special case of the block problem. It is also clear that the LCI_n problem could also use these ideas.

Below, the aforementioned generalization of LCI_n is studied for binary random words, namely we analyze the asymptotic behavior of $LCbB_n$, the length of the longest common subsequence with a fixed number, b, of blocks. As it will become clear, when $b = 2$, and for binary words, say, starting with a zero, $LCbB_n = LCI_n$.

There are two insights that can be helpful to know what to expect for the limiting functionals of $LCbB_n$. First, in previous work on LI_n, e.g. [22], the limiting Brownian functionals obtained are being taking the maximum over the partition of $[0, 1]$, and m, the size of partition, comes from the number of blocks. On the other hand, the size of alphabet m, determines the dimension of resulting Brownian motions (where sometimes we have $m - 1$ dimensional Brownian motion instead via pointwise linear transformation). Another insight here is about some max-min functionals that obtained in previous work of LCI_n, e.g. [10]. We may consider finding the minima part is to compare between two sequences to reflect 'common'; and finding the maxima is to ensure the 'longest' length among all possible division of each sequence. Therefore in a binary setting $(m = 2)$, it is reasonable to expect and has shown that, in some cases the limiting behavior of $LCbB_n$ consists of some max-min functional, of one or two dimensional Brownian functionals over the partition of $[0, 1]$ of size b.

Let us start by giving a formal definition of sequences with blocks. Let $(X_n)_{n\geq 1}$ be a sequence of random variables, for any positive integers n_1 and n_2 such that $1 \leq n_1 \leq n_2 \leq n$, a block is said to be located at $[n_1, n_2]$ if the following three conditions hold:

(i) $X_{n_1} = X_{n_1+1} = ... = X_{n_2}$;

(ii) $X_{n_1-1} \neq X_{n_1}$ *or* $n_1 = 1$;

(iii) $X_{n_2} \neq X_{n_2+1}$ *or* $n_2 = n$.

If $[n_1, n_2]$ is a block, then $n_2 - n_1 + 1$ is called the length of the block. For example

the finite sequence 001111100 has three blocks: the first one located at $[1, 2]$, the second located at $[3, 7]$ and the last one located at $[8, 9]$. In this example they have lengths 2, 5 and 2, respectively. On the other hand, the sequence 010101010101 has twelve blocks, each one of length one. Indeed, according to the definition, the vacuous blocks which are the ones of length zero, are not allowed; neither do we split a block into more blocks. However, it will be clear in the proof that the things before do not change the limiting behavior of $LCbB_n$ due to the almost surely continuity.

Let $b \in \mathbb{N}, b \geq 2$ be fixed, and let LbB_n be the length of the longest subsequence of $X_1, X_2, ..., X_n$ either starting with a zero or a one, and containing exactly b blocks. Here by a subsequence we mean either the sequence itself or any string obtained from $X_1, X_2, ..., X_n$ by removing any number of the X_i. For example, assume $X_1, ..., X_9$ takes on the value: 001111010. Clearly $L3B_9 = 8$ can be obtained from subsequence 00111100 or 00111110; also, $L4B_9 = 8$ from the subsequence 00111101.

Let us now extend our framework to two sequences $X = (X_i)_{i\geq 1}$ and $Y = (Y_i)_{i\geq 1}$. Define $LCbB_n$, the length of the longest common subsequences with b-blocks, $b \geq 2$, to be the maximal integer $k \in \{1, ..., n\}$, such that there exist $1 \leq i_1 < ... < i_k \leq n$ and $1 \leq j_1 < ... < j_k \leq n$ satisfying:

(i) $X_{i_s} = Y_{j_s}$, for $s = 1, 2, ..., k$,

(ii) $(X_{i_s})_{1\leq s\leq k}$ and $(Y_{i_s})_{1\leq s\leq k}$ both consist of b-blocks.

For example, let $X = 0010$ and $Y = 0101$. Then $LC2B_4 = 3$ with the longest common subsequence 001, and $LC3B_4 = 3$ with the longest common subsequence 010.

Below is the main result on the asymptotic behavior of $LCbB_n$,

Theorem 1.1. *Let $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ be two independent sequences of i.i.d. Bernoulli random variables with respective parameter p_1 and p_2, and without loss of generality $0 < p_1 \leq p_2 < 1$. Then,*

(i) For $p_1 = p_2 = 1/2$*,*

$$\frac{LCbB_n - n/2}{\sqrt{n}/2} \Longrightarrow \max_{\substack{0=t_0\le t_1\le \dots \\ \le t_{b-1}\le t_b=1}} \min_{k=1,2} \left(\sum_{i=1}^{b} (-1)^{i+1} \Big(B^k(t_i) - B^k(t_{i-1}) \Big) \right) \bigvee \max_{\substack{0=t_0\le t_1\le \dots \\ \le t_{b-1}\le t_b=1}} \min_{k=3,4} \left(\sum_{i=1}^{b} (-1)^{i+1} \Big(B^k(t_i) - B^k(t_{i-1}) \Big) \right), \tag{1.1}$$

where B^1, B^2*,* B^3*,* B^4 *are independent standard Brownian motions defined on* $[0,1]$*.*

(ii) For $p_1 = p_2 \neq 1/2$*,*

$$\frac{LCbB_n - n\max(p_1, 1-p_1)}{\sqrt{p_1(1-p_1)n}} \Longrightarrow \min(Z_1, Z_2), \tag{1.2}$$

where Z_1 *and* Z_2 *are two independent standard normal random variables.*

(iii) For $p_1 < p_2 < 1/2$*, or* $1/2 < p_1 < p_2$*,*

$$\frac{LCbB_n - n\max(p_1, 1-p_2)}{\sqrt{\max(p_1, 1-p_2)\min(1-p_1, p_2)n}} \Longrightarrow Z, \tag{1.3}$$

where Z *is a standard normal random variable.*

(iv) For $p_1 < p_2 = 1/2$*, or* $1/2 = p_1 < p_2$*,*

$$\frac{LCbB_n - n/2}{\sqrt{n}} \Longrightarrow \max_{\substack{0=t_0\le t_1\le \dots \\ \le t_{b-1}\le t_b=1}} \left(-\frac{1}{2}B(1) + \sum_{i=1}^{b-1} (-1)^{i-1} B\big(2\min(p_1, 1-p_2)t_i\big) \right), \tag{1.4}$$

where B *is a standard Brownian motion defined on* $[0,1]$*.*

(v) For $p_1 < 1/2 < p_2$*,*

$$\frac{LCbB_n - n(2p_1p_2 - p_1 - p_2 + 1)}{\sigma\sqrt{n}} \Longrightarrow \max(Z_1, Z_2), \tag{1.5}$$

where Z_1 and Z_2 are two independent standard normal random variables, and where $\sigma = \sqrt{p_1(p_1^2 - 2p_1p_2 + p_2)/(1-p_1)}$ *when* $p_1 + p_2 < 1$, *while* $\sigma = \sqrt{(1-p_2)(p_2^2 - 2p_1p_2 + p_1)/p_2}$ *when* $p_1 + p_2 \geq 1$.

Remark 1.2. It is noteworthy that Theorem 1.1 recovers the uniform LCI_n result for binary words obtained in [20], i.e. the case of only two blocks (see Remark 3.3). It also recovers a binary version of the conjectured limit, presented in [10, Concluding Remarks 2]. Furthermore, the one-sequence's version of Theorem 1.1 (i) and (ii), agree with the LI_n result achieved in [22] for both uniform and non-uniform case (see Chapter 5). Moreover, note that in case of two dependent sequences, the limiting behavior of $LCbB_n$ can also be derived accordingly along the rationale presented.

Remark 1.3. As a heuristic note, let us now take a moment to include some high level insights on the main theorem 1.1. First, in case (i), (ii) and (v), we need to take advantage of all information from both subsequences to obtain the $LCbB_n$. However, in (iii) and (iv), the optimal subsequences will be shown later that limited only by one sequence and consequentially only one component will appear in the limiting functionals of those two cases. Secondly, depending on whether the dominant letter existing in both X and Y or not, in the cases (i), (iv) and (v), it is necessary to take the maximum over two different orders of counting, but not for situation (ii) and (iii). Finally, the symmetry should occur in both pairs $(X, Y)\&(Y, X)$, and $(X, Y)\&(1 - X, 1 - Y)$, which indicates that the limiting results should be unchanged if p_1 and p_2 are switched, or p_1, p_2 are replaced by $1-p_1, 1-p_2$.

The dissertation is organized as follows. In Chapter 2, elementary combinatorics arguments first allow us to represent $LCbB_n$ as the maximum of two max-min functionals. Then we proceed with probabilistic developments and transform each max-min functional into the maximum over a random set of the minimum of random sums of random variables. Subsequently in Chapter 3, we complete the proof of the uniform case, which is the first

part of the main theorem. For the non-uniform cases ((ii), (iii), (iv), (v)), the proofs are done in Chapter 4. We also develop an alternative approach to the one-sequence problem LbB_n in Chapter 5. In Chapter 6 we summarize this dissertation and discuss some potential extensions and related problems of interest. Finally, the thesis is wrapped up with an appendix which consists of code snippet on Monte-Carlo simulation.

CHAPTER 2

COMBINATORICS AND PROBABILISTIC DEVELOPMENT

Let us get started by introducing some notation and methods from [10]. First, let $N_0(X)$ be the number of zeros in $X_1, X_2, ..., X_n$, i.e.,

$$N_0(X) = \#\{i = 1, ..., n : X_i = 0\} = \sum_{i=1}^{n} \mathbb{1}_{\{X_i=0\}}, \tag{2.1}$$

and let $N_1(X)$ be the number of ones in $X_1, X_2, ..., X_n$, and similarly define $N_0(Y)$ and $N_1(Y)$. Then, let $N_0^{s,t}(X)$ be the number of zeros in $X_{s+1}, X_{s+2}, ..., X_t$, i.e.,

$$N_0^{s,t}(X) = \#\{i = s+1, ..., t : X_i = 0\} = \sum_{i=s+1}^{t} \mathbb{1}_{\{X_i=0\}}, \tag{2.2}$$

and similarly define $N_1^{s,t}(X)$, $N_0^{s,t}(Y)$, and $N_1^{s,t}(Y)$.

Next, let $T_0^j(X)$ be the location of the j^{th} zero in the infinite sequence $X_1, X_2, ...$ For $j = 1, 2, ..., T_0^j(X)$ is recursively defined via

$$T_0^j(X) = \min\left\{s \in \mathbb{N} : s > T_0^{j-1}(X),\ X_s = 0\right\}, \tag{2.3}$$

with the convention $T_0^0(X) = 0$. Similarly define $T_1^j(X)$, $T_0^j(Y)$, and $T_1^j(Y)$.

From now on, throughout the paper if a property is valid for both sequences X and Y, then the symbol X or Y is omitted. Clearly, there are two ways of counting blocks, starting with block of zeros, or block of ones, the corresponding length of the longest common subsequences with b blocks are then respectively denoted by $LCbB_n^0$ and $LCbB_n^1$, and

$$LCbB_n = \max\left(LCbB_n^0, LCbB_n^1\right).$$

Since $LCbB_n^0$ and $LCbB_n^1$ are highly similar, next we will assume the block starts with a zero and concentrate on the combinatorial expression of $LCbB_n^0$ first. Also, assume that the number of blocks, b, is even. The proof for b odd could be analogously achieved and will be provided at the end.

Now let $H_X(k_1, k_2, ..., k_{b-1})$ be the maximal number of ones contained in a subsequence of $X_1X_2...X_n$, after $b-1$ blocks with k_1 zeros, k_2 ones, k_3 zeros, ..., and k_{b-1} zeros, already drawn in that order. (All the k_is are positive integers with the convention that $k_0 = 0$, and $H_X(k_1, k_2, ..., k_{b-1})$ is assumed to be negative infinity if such subsequence does not exist.)

Replacing X by Y it is then clear that

$$\min\left(k_1+...+k_{b-1}+H_X(k_1,...,k_{b-1}), k_1+...+k_{b-1}+H_Y(k_1,...,k_{b-1})\right), \tag{2.4}$$

is the length of the longest common subsequences of X and Y made of b blocks both containing k_1 zeros, k_2 ones, k_3 zeros, ..., k_{b-1} zeros and $\min(H_X(k_1, ..., k_{b-1}), H_Y(k_1, ..., k_{b-1}))$ ones in this exact order, and thus $LCbB_n^0$ will be the maximum of (2.4) over all possible choices of $k_1, k_2, ..., k_{b-1}$.

Next, let us study the constraints. First, clearly, $0 \leq k_1 \leq N_0$. Now for k_2, its maximum should be the total number of ones, namely N_1, minus the number of ones occurring in the first block with k_1 zeros. Let N_2^* be the number of these inadmissible ones in picking the second block. Also by convention we set $N_1^* = 0$.

Clearly, $N_2^* = N_1^{0,T_0^{k_1+N_1^*}}$, and generally, we let N_r^* designate the number of zeros, for r odd; or the number of ones, for r even, occurring before the r-th block. Roughly speaking, the process of counting letters is first to put $b-1$ partitions in the sequence, and then to count the number of zeros in the first block, the number of ones in the second block, and so on. Obviously it is never optimal to divide the sequence between two consecutive 00 or 11. In other words, except for the first block, the block to be counted zeros (resp. ones)

within it should both start and end with a zero (resp. one). So by the very definition of T_i^j, we have

$$T_0^{1+N_i^*} = 1 + T_1^{k_{i-1}+N_{i-1}^*},$$

and

$$T_1^{1+N_i^*} = 1 + T_0^{k_{i-1}+N_{i-1}^*},$$

for $i = 2, 3, .., b$. So the ranges for the constraints are as follows:

$$0 \leq k_i \leq N_0 - N_i^*, \textit{where } N_i^* = N_0^{0,T_1^{k_{i-1}+N_{k_{i-1}}^*}} \textit{ for i odd},$$

$$0 \leq k_i \leq N_1 - N_i^*, \textit{where } N_i^* = N_1^{0,T_0^{k_{i-1}+N_{k_{i-1}}^*}} \textit{ for i even}.$$

Finally, throughout the paper let $r_i = 1/2 + (-1)^i/2$ represent the parity of index i. To date we have our first combinatorial expression of $LCbB_n^0$:

$$LCbB_n^0 = \max_{\mathcal{C}_n} \min\left(\sum_{i=1}^{b-1} k_i + H_X(k_1, ..., k_{b-1}), \sum_{i=1}^{b-1} k_i + H_Y(k_1, ..., k_{b-1}) \right), \tag{2.5}$$

$$where\, \mathcal{C}_n = \{(k_1, ..., k_{b-1}) : k_1 \in \mathcal{C}_{n,1}, k_2 \in \mathcal{C}_{n,2}(k_1), ..., k_{b-1} \in \mathcal{C}_{n,b-1}(k_1, ..., k_{b-2})\},$$

$$where\, \mathcal{C}_{n,1} = \{0 \leq k_1 \leq (N_0(X)) \wedge (N_0(Y))\}, and\ for\ i = 2, ..., b-1,$$

$$\mathcal{C}_{n,i}(k_1, ..., k_i) = \{0 \leq k_i \leq (N_{r_i}(X) - N_i^*(X)) \wedge (N_{r_i}(Y) - N_i^*(Y))\},$$

$$and\ r_i = 1/2 + (-1)^i/2.$$

Our next goal is to identify the function $H(k_1, k_2, ..., k_{b-1})$. Notice that $H(k_1, k_2, ..., k_{b-1})$ is equal to N_1 minus the number of ones occurring in the first $b-1$ blocks. As mentioned before, we may assume the vacuous blocks are allow to happen. For i odd, let us consider the number of ones occurring between different zeros in the i-th block. Take the third block, for example, then the number of ones in this block is

$$N_1^{T_1^{k_2+N_2^*}, T_0^{N_3^*+1}} + N_1^{T_0^{N_3^*+1}, T_0^{N_3^*+2}} + N_1^{T_0^{N_3^*+2}, T_0^{N_3^*+3}} + ... + N_1^{T_0^{N_3^*+k_3-1}, T_0^{N_3^*+k_3}},$$

where the first term above, which is the term left when $k_3 = 1$, is actually zero. Note that all the odd blocks share this form, except for the first block, where the first term of it is instead $T_0^1 - 1$ since the first block is not necessarily starting with a zero. For i even, clearly the number of ones in the i-th block is exactly k_i, which can be expressed as a summation similar to the above one. For instance, in the second block, the number of ones in it equals

$$N_1^{T_0^{k_1+N_1^*}, T_1^{N_2^*+1}} + N_1^{T_1^{N_2^*+1}, T_1^{N_2^*+2}} + N_1^{T_1^{N_2^*+2}, T_1^{N_2^*+3}} + ... + N_1^{T_1^{N_2^*+k_2-1}, T_1^{N_2^*+k_2}},$$

where every terms above equals 1, making the summation equal to k_2. In particular, the first term is the one left when $k_2 = 1$.

On the other hand, by definition, for $i > 1$ it is clear that

$$N_1^{T_{1-r_i}^{N_{i-1}^*+k_{i-1}}, T_r^{N_i^*+1}} = r_i.$$

Combining these facts and take the vacuous blocks into account, we have

$$H(k_1, k_2, ..., k_{b-1}) = N_1 - D,$$

where

$$D = \sum_{i=1}^{b-1} \left(r_i \mathbb{1}_{\{k_i>0\}} + \sum_{j=N_i^*+2}^{N_i^*+k_i} N_1^{T_{r_i}^{j-1}, T_{r_i}^{j}} \right) + (T_0^1 - 1)\mathbb{1}_{\{k_1>0\}},$$

where, again $r_i = 1/2 + (-1)^i/2$, and moreover where the inner summation is not present if $k_i < 2$.

We thus arrived at our final combinatorial expression for $LCbB_n^0$, namely,

Lemma 2.1.

$$LCbB_n^0 = \max_{\mathcal{C}_n} \min \left(\sum_{i=1}^{b-1} k_i + N_1(X) - D(X), \sum_{i=1}^{b-1} k_i + N_1(Y) - D(Y) \right), \qquad (2.6)$$

where

$$D=\sum_{i=1}^{b-1}\left(r_i\mathbb{1}_{\{k_i>0\}}+\sum_{j=N_i^*+2}^{N_i^*+k_i}N_1^{T_{r_i}^{j-1},T_{r_i}^j}\right)+(T_0^1-1)\mathbb{1}_{\{k_1>0\}},$$

with the constraint $\mathcal{C}_n$ are as defined after (2.5), and where the inner summation in D is not present if $k_i<2$.

Let us now concentrate on the random variables $N_1^{T_0^{j-1},T_0^j}$ in (2.6), assuming a sequence of Bernoulli random variables with parameter $0<p<1$. By definition, $N_1^{T_0^{j-1},T_0^j}$ is the number of ones between T_0^{j-1} and T_0^j, namely, $T_0^j-T_0^{j-1}-1$. Note that T_0^j is a negative Binomial (Pascal) random variable with parameters j and $1-p$, and $T_0^j-T_0^{j-1}$ is a geometric random variable on $\{1,2,...\}$ with parameter $1-p$. Thus, $N_1^{T_0^{j-1},T_0^j}$ follows a geometric distribution on $\mathbb{N}=\{0,1,2,...\}$ with parameter $1-p$ (and so with mean $p/(1-p)$ and variance $p/(1-p)^2$). Moreover, according to [10, Prop 3.1], the random variables $\left(N_1^{T_0^{j-1},T_0^j}\right)_{j\geq 1}$ are i.i.d. This is summarized as:

Lemma 2.2. *The random variables $\left(N_1^{T_0^{j-1},T_0^j}\right)_{j\geq 1}$ are i.i.d. geometric, on $\mathbb{N}$, with parameter $1-p$.*

Continuing and from Lemma 2.1 it follows that:

$$LCbB_n^0=\max_{\mathcal{C}_n}\min\left(\sum_{i=1}^{b-1}k_i+N_1(X)-D_{b,n}(X),\sum_{i=1}^{b-1}k_i+N_1(Y)-D_{b,n}(Y)\right),\qquad(2.7)$$

where,

$$D_{b,n}=\sum_{i=1}^{b-1}\left(r_i\mathbb{1}_{\{k_i>0\}}+\sum_{j=N_i^*+2}^{N_i^*+k_i}\left(\left(\frac{N_1^{T_{r_i}^{j-1},T_{r_i}^j}-\frac{p}{1-p}}{\sqrt{\frac{pn}{(1-p)^2}}}\right)\sqrt{\frac{pn}{(1-p)^2}}+\frac{p}{1-p}\right)\right)$$
$$+(T_0^1-1)\mathbb{1}_{\{k_1>0\}},$$

and the inner summation in $D_{b,n}$ is not present if $k_i<2$.

It is also clear that when $k_1>0$, $(T_0^1-1)/\sqrt{n}\xrightarrow{\mathbb{P}}0$, as $n\to\infty$. Indeed, for any

$\epsilon > 0$,

$$\mathbb{P}\left(\frac{T_0^1 - 1}{\sqrt{n}} > \epsilon\right) = \sum_{k=[\sqrt{n}\epsilon]+1}^{n} \mathbb{P}\left(T_0^1 - 1 = k\right)$$
$$= p^{[\sqrt{n}\epsilon]+1} - p^{n+1} \to 0,$$

as $n \to \infty$.

Next, let $S_{0,1}^{(n)}$ denote the number of ones after the occurrence of the last zero. The next lemma, modified from [10, Prop 3.2], to the block context, aims at representing N_1 in terms of the same random variables as above, with a remainder, which after been divided by $\sqrt{n}$, converges to zero in probability.

Lemma 2.3.

$$N_1 = np + (1-p)\sqrt{\frac{pn}{(1-p)^2}} \sum_{j=1}^{N_0} \frac{N_1^{T_0^{j-1}, T_0^j} - \frac{p}{1-p}}{\sqrt{\frac{pn}{(1-p)^2}}} + (1-p)S_{0,1}^{(n)} \tag{2.8}$$

where $S_{0,1}^{(n)}/\sqrt{n}$ converges to 0 in probability.

Since $N_1^{T_1^{j-1}, T_1^j} = 1$, with the help of Lemma 2.3, Lemma 2.1 then rewrites as:
(henceforth the symbols X and Y in the N_1's are omitted for the ease of notation)

$$LCbB_n^0 = \max_{\mathcal{C}_n} \min_{k=1,2} \Bigg(\sum_{i=1}^{b-1} k_i + p_k n + (1-p_k) S_{0,1}^{(n)} - (T_0^1 - 1)\mathbb{1}_{\{k_1>0\}}$$
$$+ (1-p_k)\sqrt{\frac{p_k n}{(1-p_k)^2}} \sum_{j=1}^{N_0} \frac{N_1^{T_0^{j-1}, T_0^j} - \frac{p_k}{1-p_k}}{\sqrt{\frac{p_k n}{(1-p_k)^2}}}$$
$$- \sum_{i=1}^{b-1} \Bigg(\sum_{j=N_i^*+2}^{N_i^*+k_i} \Bigg(\Bigg(\frac{N_1^{T_{r_i}^{j-1}, T_{r_i}^j} - \frac{p_k}{1-p_k}}{\sqrt{\frac{p_k n}{(1-p_k)^2}}} \Bigg) \sqrt{\frac{p_k n}{(1-p_k)^2}} + \frac{p_k}{1-p_k} \Bigg)$$
$$+ r_i \mathbb{1}_{\{k_i>0\}} \Bigg) \Bigg)$$

$$= \max_{\mathcal{C}_n} \min_{k=1,2} \left(\frac{1-2p_k}{1-p_k}(k_1+k_3+...+k_{b-1}) + \frac{p_k}{1-p_k}\frac{b}{2} \right.$$

$$+ (1-p_k)S_{0,1}^{(n)} - (T_0^1 - 1)\mathbb{1}_{\{k_1>0\}} + p_k n$$

$$+ (1-p_k)\sqrt{\frac{p_k n}{(1-p_k)^2}} \sum_{j=1}^{N_0} \frac{N_1^{T_0^{j-1},T_0^j} - \frac{p_k}{1-p_k}}{\sqrt{\frac{p_k n}{(1-p_k)^2}}}$$

$$\left. - \sum_{1\le i\le b-1, i\ odd} \left(\sum_{j=N_i^*+2}^{N_i^*+k_i} \left(\frac{N_1^{T_0^{j-1},T_0^j} - \frac{p_k}{1-p_k}}{\sqrt{\frac{p_k n}{(1-p_k)^2}}} \right) \sqrt{\frac{p_k n}{(1-p_k)^2}} \right) \right). \qquad (2.9)$$

On the other hand, let us now deal with $LCbB_n^1$. In that case:

1. The longest common subsequence starts with $b-1$ blocks both containing $\tilde{k}_1$ ones, $\tilde{k}_2$ zeros, $\tilde{k}_3$ ones, ..., and $\tilde{k}_{b-1}$ ones, while $\tilde{r}_i = 1/2 + (-1)^{i+1}/2$.

2. $\tilde{N}_r^*$ is the number of ones for r odd, or zeros for r even, occurring before the r-th block.

3. $\tilde{S}_{0,1}^{(n)}$ is the number of zeros after the occurrence of the last one.

4. $\tilde{\mathcal{C}}_n = \left\{ (\tilde{k}_1, ..., \tilde{k}_{b-1}) : \tilde{k}_1 \in \tilde{\mathcal{C}}_{n,1}, \tilde{k}_2 \in \tilde{\mathcal{C}}_{n,2}(\tilde{k}_1), ..., \tilde{k}_{b-1} \in \tilde{\mathcal{C}}_{n,b-1}(\tilde{k}_1, ..., \tilde{k}_{b-2}) \right\}$
   $where\ \tilde{\mathcal{C}}_{n,1} = \left\{ 0 \le \tilde{k}_1 \le (N_1(X)) \wedge (N_1(Y)) \right\}, and\ for\ i = 2, ..., b-1,$
   $\tilde{\mathcal{C}}_{n,i}(\tilde{k}_1, ..., \tilde{k}_i) = \left\{ 0 \le \tilde{k}_i \le (N_{\tilde{r}_i}(X) - N_i^*(X)) \wedge (N_{\tilde{r}_i}(Y) - N_i^*(Y)) \right\}.$

With these modifications, our final combinatorial expression for $LCbB_n$ is:

$$
\begin{aligned}
LCbB_n = \max_{\mathcal{C}_n} \min_{k=1,2} \Bigg(& \frac{1-2p_k}{1-p_k}(k_1+k_3+...+k_{b-1}) + \frac{p_k}{1-p_k}\frac{b}{2} + (1-p_k)S_{0,1}^{(n)} + p_k n \\
& + (1-p_k)\sqrt{\frac{p_k n}{(1-p_k)^2}} \sum_{j=1}^{N_0} \frac{N_1^{T_0^{j-1},T_0^j} - \frac{p_k}{1-p_k}}{\sqrt{\frac{p_k n}{(1-p_k)^2}}} - (T_0^1 - 1)\mathbb{1}_{\{k_1>0\}} \\
& - \sum_{1\le i\le b-1, i \text{ odd}} \left(\sum_{j=N_i^*+2}^{N_i^*+k_i} \left(\frac{N_1^{T_0^{j-1},T_0^j} - \frac{p_k}{1-p_k}}{\sqrt{\frac{p_k n}{(1-p_k)^2}}} \right) \sqrt{\frac{p_k n}{(1-p_k)^2}} \right) \Bigg) \\
\bigvee \max_{\tilde{\mathcal{C}}_n} \min_{k=1,2} \Bigg(& \frac{2p_k-1}{p_k}(\tilde{k}_1+\tilde{k}_3+...+\tilde{k}_{b-1}) + \frac{1-p_k}{p_k}\frac{b}{2} + p_k\tilde{S}_{0,1}^{(n)} + (1-p_k)n \\
& + p_k\sqrt{\frac{(1-p_k)n}{p_k^2}} \sum_{j=1}^{N_1} \frac{N_0^{T_1^{j-1},T_1^j} - \frac{1-p_k}{p_k}}{\sqrt{\frac{(1-p_k)n}{p_k^2}}} - (T_1^1 - 1)\mathbb{1}_{\{\tilde{k}_1>0\}} \\
& - \sum_{1\le i\le b-1, i \text{ odd}} \left(\sum_{j=\tilde{N}_i^*+2}^{\tilde{N}_i^*+\tilde{k}_i} \left(\frac{N_0^{T_1^{j-1},T_1^j} - \frac{1-p_k}{p_k}}{\sqrt{\frac{(1-p_k)n}{p_k^2}}} \right) \sqrt{\frac{(1-p_k)n}{p_k^2}} \right) \Bigg). \qquad (2.10)
\end{aligned}
$$

To finish this chapter, let us now study the asymptotic mean of $LCbB_n^0$ and $LCbB_n^1$, which will help guide us on the centerings in the following sections. Take $LCbB_n^0$ for illustration. Note that (2.9) is in the form for the convenience of introducing Brownian approximation in Chapter 3, let us simplify it for the moment, and rewrite $LCbB_n^0/n$ as:

$$
\begin{aligned}
\frac{LCbB_n^0}{n} = \max_{\mathcal{C}_n} \min_{k=1,2} \Bigg(& \frac{\frac{1-2p_k}{1-p_k}(k_1+k_3+...+k_{b-1}) + p_k n}{n} + \frac{(1-p_k)\big(N_1^{0,T_0^{N_0}} - \frac{p_k}{1-p_k}N_0\big)}{n} \\
& - \sum_{1\le i\le b-1, i \text{ odd}} \frac{N_1^{T_0^{N_i^*+1},T_0^{N_i^*+k_i}} - \frac{p_k}{1-p_k}(k_i-1)}{n} \Bigg) + o_{\mathbb{P}}(1), \quad (2.11)
\end{aligned}
$$

where $o_{\mathbb{P}}(1)$ denotes a term converging to zero in probability as n tends to infinity.

Note that in (2.11), all the $k_i's$ are of order $o(n)$, we now proceed to show that the first terms with $k_i's$ in the max-min functional will determine the centering. It can be done by the following estimation on the order of last two terms in (2.11). Let $\alpha \in (0, 1/4)$ small,

and let the event

$$E_n^\alpha := \left\{ \left| \frac{N_1^{T_0^{j_1}, T_0^{j_2}} - p_k(T_0^{j_2} - T_0^{j_1})}{\sqrt{n}} \right| \leq n^{\alpha - 1/2} \sqrt{T_0^{j_2} - T_0^{j_1}}, \textit{for} \, \forall \, 0 \leq j_1 \leq j_2 \leq N_0 \right\}.$$

Note that there are at most C_{n+1}^2 combinations of possible pairs of $\{j_1, j_2\}$, as an application of Hoeffding's inequality, we have $\mathbb{P}((E_n^\alpha)^c) \leq G_n \exp\left(-J_n n^{2\alpha}\right)$, where G_n and J_n are quadratic and linear in n, respectively. Then according to Borel-Cantelli lemma, E_n^α will happen for all n large enough. In other words, we could now instead use n^α as an upper bound of fluctuation of $N_1^{T_0^{j_1}, T_0^{j_2}}$ which is normalized by $\sqrt{n}$.

Then, recall the following elementary inequality, used throughout: Let a_k, b_k, c_k, d_k, $1 \leq k \leq K$, be reals, then see [20],

$$\begin{aligned} &\left| \max_{k=1,\ldots,K} (a_k \wedge b_k) - \max_{k=1,\ldots,K} \left((a_k + c_k) \wedge (b_k + d_k) \right) \right| \\ &\qquad \leq \max_{k=1,2,\ldots,K} (|c_k| \vee |d_k|) \,. \end{aligned} \tag{2.12}$$

Therefore, when E_n^α occurs,

$$\begin{aligned} &\left| \frac{LCbB_n^0}{n} - \max_{\mathcal{C}_n} \min_{k=1,2} \left(\frac{\frac{1-2p_k}{1-p_k}(k_1 + k_3 + \ldots + k_{b-1}) + p_k n}{n} \right) \right| \\ \leq \quad & \max_{\mathcal{C}_n} \max_{X,Y} \left| \frac{(1-p_k)\left(N_1^{0, T_0^{N_0}} - \frac{p_k}{1-p_k} N_0\right)}{n} - \sum_{1 \leq i \leq b-1, \, i \text{ odd}} \frac{N_1^{T_0^{N_i^*+1}, T_0^{N_i^*+k_i}} - \frac{p_k}{1-p_k}(k_i - 1)}{n} \right| \\ \leq \quad & L(p_k, b) n^{\alpha - 1/2}, \end{aligned}$$

where $L(p_k, b)$ is a functional independent of n. Therefore as $n \to \infty$, $LCbB_n^0/n$ almost surely tends to

$$\lim_{n \to \infty} \max_{\mathcal{C}_n} \min_{k=1,2} \left(\frac{\frac{1-2p_k}{1-p_k}(k_1 + k_3 + \ldots + k_{b-1}) + p_k n}{n} \right) := \mu(p_1, p_2).$$

Moreover, since $\mu(p_1, p_2)$ is integrable in finite measure space, dominated convergence theorem then gives $\mathbb{E}(LCbB_n^0)/n \to \mu(p_1, p_2)$. It is then straightforward to show that, under different scenarios of p_1 and p_2 (as indicated in Theorem 1.1), $\mu(p_1, p_2)$, which is also the asymptotic mean of $LCbB_n^1$, is equal to $\max(p_1 \wedge p_2, (1-p_1) \wedge (1-p_2), 2p_1p_2 - p_1 - p_2 + 1) = \max(p_1, 1-p_2, 2p_1p_2 - p_1 - p_2 + 1)$. Given these proper centerings, we are now ready to dive deep into more details of the limiting behavior in the following chapters.

CHAPTER 3

THE UNIFORM CASE

In this chapter, the limiting law of $LCbB_n$ is derived when $p_1 = p_2 = 1/2$. Again start with $LCbB_n^0$. For uniform draw, (2.10) simplifies to

$$\frac{LCbB_n^0 - n/2}{\sqrt{2n}} = \max_{\mathcal{C}_n} \min\Bigg(M_{b,n}(X) + \frac{1}{2\sqrt{2n}} \left(S_{0,1}^{(n)}(X) + b + 2(T_0^1(X) - 1)\mathbb{1}_{\{k_1>0\}} \right), \\ M_{b,n}(Y) + \frac{1}{2\sqrt{2n}} \left(S_{0,1}^{(n)}(Y) + b + 2(T_0^1(Y) - 1)\mathbb{1}_{\{k_1>0\}} \right) \Bigg), \tag{3.1}$$

where

$$M_{b,n} = \frac{1}{2} \sum_{j=1}^{N_0} \frac{N_1^{T_0^{j-1}, T_0^j} - 1}{\sqrt{2n}} - \sum_{1 \le i \le b-1,\ i\ odd} \left(\sum_{j=N_i^*+2}^{N_i^*+k_i} \frac{N_1^{T_0^{j-1}, T_0^j} - 1}{\sqrt{2n}} \right), \tag{3.2}$$

and with the inner summation not present if $k_i < 2$.

Applying (2.12) to (3.1) again, leads to

$$\Bigg| \max_{\mathcal{C}_n} \min\Bigg(M_{b,n}(X) + \frac{1}{2\sqrt{2n}} \left(S_{0,1}^{(n)}(X) + b + 2(T_0^1(X) - 1)\mathbb{1}_{\{k_1>0\}} \right), \\ M_{b,n}(Y) + \frac{1}{2\sqrt{2n}} \left(S_{0,1}^{(n)}(Y) + b + 2(T_0^1(Y) - 1)\mathbb{1}_{\{k_1>0\}} \right) \Bigg) \\ - \left(\max_{\mathcal{C}_n} \min \left(M_{b,n}(X), M_{b,n}(Y) \right) \right) \Bigg| \\ \le \frac{1}{2\sqrt{2n}} \max_{\mathcal{C}_n} \Bigg(\left| S_{0,1}^{(n)}(X) + b + 2(T_0^1(X) - 1)\mathbb{1}_{\{k_1>0\}} \right| \vee \\ \left| S_{0,1}^{(n)}(Y) + b + 2(T_0^1(Y) - 1)\mathbb{1}_{\{k_1>0\}} \right| \Bigg).$$

Now since $|S_{0,1}^{(n)} + b + 2(T_0^1 - 1)\mathbb{1}_{\{k_1>0\}}|/(2\sqrt{2n})$ converges in probability, to 0, as $n \to \infty$, the limiting behavior of the right-hand side of (3.1) is then the same as the one of

$$\max_{\mathcal{C}_n} \min \Big(M_{b,n}(X), M_{b,n}(Y) \Big). \tag{3.3}$$

Before proceeding to the details of $M_{b,n}$, let us take another look at the random variables N_i^*. First, $N_1^* = 0$. Now N_2^* is the number of ones in the first block which contains k_1 zeros, and so

$$N_2^* = \max \left(T_0^{k_1+N_1^*} - T_0^{1+N_1^*} + 1 - k_1, 0 \right).$$

Similarly, N_3^* is the sum of k_1 and the number of zeros in the second block. Formally, for $1 \leq i \leq b-1$, let s_i be the number of ones for odd i, and zeros for even i, in the i-th block, so

$$s_i = \max \left(T_{r_i}^{k_i+N_i^*} - T_{r_i}^{1+N_i^*} - k_i + 1, 0 \right),$$

and clearly

$$\mathbb{E}(s_i) = \max(k_i - 1, 0).$$

Therefore,

$$N_i^* = s_1 + k_2 + s_3 + k_4 + ... + k_{i-2} + s_{i-1},$$

for i even; and

$$N_i^* = k_1 + s_2 + k_3 + s_4 + ... + k_{i-2} + s_{i-1},$$

for i odd.

Next, set $k_i' = \max(k_i - 1, 0)$ for i odd; and $k_i' = k_i$ for i even. Correspondingly, let

the constraints

$$\mathcal{C}_n^{'} = \left\{ (k_1', ..., k_{b-1}') : k_1' \in \mathcal{C}_{n,1}^{'}, k_2' \in \mathcal{C}_{n,2}^{'}(k_1'), ..., k_{b-1}' \in \mathcal{C}_{n,b-1}^{'}(k_1', ..., k_{b-2}') \right\}, \tag{3.4}$$

where for $i = 2, ..., b-1$, $\mathcal{C}_{n,1}^{'} = \{0 \leq k_1' \leq (N_0(X)) \wedge (N_0(Y))\}$,

and $\mathcal{C}_{n,i}^{'}(k_1', ..., k_i') = \{0 \leq k_i' \leq (N_{r_i}(X) - N_i^*(X)) \wedge (N_{r_i}(Y) - N_i^*(Y))\}$

By induction, $\mathbb{E}(N_i^*) = \sum_{j=1}^{i-1} k_j'$. For a closer look at the $s_i's$, since Lemma 2.2 continues to hold for $\left(N_0^{T_1^{j-1}, T_1^j}\right)_{j \geq 1}$, when $k_i \geq 1$ s_i is equal to the sum of $k_i - 1$ geometric variables on $\mathbb{N}$ with parameter $1/2$. Also, by definition

$$s_i = N_{r_i}^{T_{1-r_i}^{N_i^*+1}, T_{1-r_i}^{N_i^*+2}} + ... + N_{r_i}^{T_{1-r_i}^{N_i^*+k_i-1}, T_{1-r_i}^{N_i^*+k_i}},$$

and by [10, Lemma 3.1], the $s_i's, i = 1, 2, ..., b-1$ are independent with mean $k_i - 1$ and probability generating function

$$\mathbb{E}(x^{s_i}) = \left(\frac{1}{2-x}\right)^{k_i - 1}.$$

Now let us bring the Brownian approximation into play. Define B_n' to be the continuous polygonal process on $[0, 1]$,

$$B_n'(t) = \frac{N_1^{0, T_0^{\lfloor tn \rfloor}} - (tn+1)/2 + (tn - \lfloor tn \rfloor \mathbb{1}_{X_{\lfloor tn \rfloor + 1} = 1})}{\sqrt{2n}},$$

for $t \in [0, 1]$, and $\lfloor . \rfloor$ is the floor function. In other words, B_n' is the polygonal process on $[0, 1]$, linearly interpolating between the values

$$B_n'\left(\frac{k}{n}\right) = \sum_{j=1}^{k+1} \frac{Z_j}{\sqrt{n}}, \ \textit{for } k = 0, 1, ..., n, \tag{3.5}$$

where

$$Z_j = \frac{N_1^{T_0^{j-1}, T_0^j} - 1}{\sqrt{2}}, \tag{3.6}$$

where, by Lemma 2.2 the $N_1^{T_0^{j-1}, T_0^j}, j = 1, ..., k+1$ are i.i.d.with mean 1 and variance 2 for uniform case. Also, notice that for the sake of indexing convenience our polygonal process contains one more term, $Z_{j+1}/\sqrt{n}$, than

$$B_n\left(\frac{k}{n}\right) = \sum_{j=1}^{k} \frac{Z_j}{\sqrt{n}}, \ \textit{for } k = 0, 1, ..., n.$$

Let us now invoke the Donsker's Theorem (also known as Donsker's invariance principle, see [5]), and the Continuous mapping theorem, which will play a significant role in this thesis.

Theorem 3.1. *Let $\xi_1, \xi_2, ...$ be a sequence of i.i.d. random variables having mean 0 and finite positive variance σ^2. Let $S_n = \xi_1 + ... + \xi_n$ $(S_0 = 0)$, and*

$$B_n(t) = \frac{1}{\sigma\sqrt{n}} S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) \frac{1}{\sigma\sqrt{n}} \xi_{\lfloor nt \rfloor + 1} \tag{3.7}$$

be the function defined by linear interpolation between its values $B_n(i/n) = S_i/(\sigma\sqrt{n})$, $for\ i = 1, ..., n$, then $B_n(t) \Longrightarrow B(t)$, where $B(t)$ is standard Brownian motion.

Theorem 3.2. *Let $(X_n)_{n\geq 1}$ and X be random elements defined on a metric space S. Denote S' as another metric space, and let a function $g : S \to S'$, which has discontinuity sets of probability 0. Then the convergence of $g(X_n)$ to $g(X)$ is preserved from the convergence of X_n to X, for convergence in distribution, convergence in probability and almost surely convergence.*

According to Donsker's Theorem, B_n is the Brownian approximation of the sum of i.i.d. normalized random variables Z_j, where specifically,

$$\mathbb{P}(Z_j = k) = (1/2)^{\sqrt{2}k+2},$$

where $k = -1/\sqrt{2}, 0, 1/\sqrt{2}, ...$ In addition, $B_n \Longrightarrow B$ in the space $C[0,1]$ equipped with the supremum norm. Throughout the thesis, the underlying probability space is assumed to be rich enough that all random variables and Brownian motions we study can be well-defined on it.

Now (3.3) rewrites as:

$$\begin{aligned}
&\max_{\mathcal{C}'_n}\left(\left(\frac{1}{2}B'_n\left(\frac{N_0(X)-1}{n}\right)-\sum_{\substack{1\le i\le b-1\\ i\ odd}}\left(B'_n\left(\frac{N_i^*(X)+k'_i}{n}\right)-B'_n\left(\frac{N_i^*(X)}{n}\right)\right)\right)\right.\\
&\qquad\left.\bigwedge\left(\frac{1}{2}B'_n\left(\frac{N_0(Y)-1}{n}\right)-\sum_{\substack{1\le i\le b-1\\ i\ odd}}\left(B'_n\left(\frac{N_i^*(Y)+k'_i}{n}\right)-B'_n\left(\frac{N_i^*(Y)}{n}\right)\right)\right)\right)\\
=&\max_{\mathcal{C}'_n}\left(\left(\frac{1}{2}B_n\left(\frac{N_0(X)-1}{n}\right)-\sum_{\substack{1\le i\le b-1\\ i\ odd}}\left(B_n\left(\frac{N_i^*(X)+k'_i}{n}\right)-B_n\left(\frac{N_i^*(X)}{n}\right)\right)\right)\right.\\
&\qquad\left.\bigwedge\left(\frac{1}{2}B_n\left(\frac{N_0(Y)-1}{n}\right)-\sum_{\substack{1\le i\le b-1\\ i\ odd}}\left(B_n\left(\frac{N_i^*(Y)+k'_i}{n}\right)-B_n\left(\frac{N_i^*(Y)}{n}\right)\right)\right)\right)+o_{\mathbb{P}}(1).
\end{aligned} \tag{3.8}$$

In the last equality we notice that the functionals with B'_n only differ from the functionals with B_n by the sum of $b+1$ standard normal random variables divided by $\sqrt{n}$, which converges to 0 in probability.

In similarity to [10], and from (3.8), three limiting behaviors need to be considered. At first, $B_n\left((N_0-1)/n\right)$ can be approximated by $B_n\left(\mathbb{E}\left((N_0-1)/n\right)\right)$. Next, the increments $B_n\left(\frac{N_i^*+k'_i}{n}\right)-B_n\left(\frac{N_i^*}{n}\right)$ are also concentrated around the Brownian approximation at the mean:

$$B_n\left(\mathbb{E}\left(\frac{N_i^*+k'_i}{n}\right)\right)-B_n\left(\mathbb{E}\left(\frac{N_i^*}{n}\right)\right)=B_n\left(\sum_{j=1}^{i}\frac{k'_j}{n}\right)-B_n\left(\sum_{j=1}^{i-1}\frac{k'_j}{n}\right).$$

Finally, the constraints need to be de-randomized. As in [10], it follows that

- $B_n\left((N_0-1)/n\right)$

$$\frac{LCbB_n^0-n/2}{\sqrt{2n}}=\max_{\mathcal{C}_n'}\min\left(\frac{1}{2}B_n^1\left(\frac{1}{2}\right)-Q_n(X),\frac{1}{2}B_n^2\left(\frac{1}{2}\right)-Q_n(Y)\right)$$
$$+o_{\mathbb{P}}(1), \qquad (3.9)$$

where

$$Q_n=\sum_{1\leq i\leq b-1, i\ odd}\left(B_n\left(\frac{N_i^*+k_i'}{n}\right)-B_n\left(\frac{N_i^*}{n}\right)\right).$$

- $B_n\left(\frac{N_i^*+k_i'}{n}\right)-B_n\left(\frac{N_i^*}{n}\right)$

$$\frac{LCbB_n^0-n/2}{\sqrt{2n}}=\max_{\mathcal{C}_n'}\min_{k=1,2}\left(\frac{1}{2}B_n^k\left(\frac{1}{2}\right)-\sum_{\substack{1\leq i\leq b-1\\ i\ odd}}\left(B_n^k\left(\sum_{j=1}^{i}\frac{k_j'}{n}\right)-B_n^k\left(\sum_{j=1}^{i-1}\frac{k_j'}{n}\right)\right)\right)+o_{\mathbb{P}}(1). \qquad (3.10)$$

- $\mathcal{C}_n'$

$$\frac{LCbB_n^0-n/2}{\sqrt{2n}}=\max_{\mathcal{C}_n'}\min_{k=1,2}\left(\frac{1}{2}B_n^k\left(\frac{1}{2}\right)-\sum_{\substack{1\leq i\leq b-1\\ i\ odd}}\left(B_n^k\left(\sum_{j=1}^{i}\frac{k_j'}{n}\right)-B_n^k\left(\sum_{j=1}^{i-1}\frac{k_j'}{n}\right)\right)\right)+o_{\mathbb{P}}(1)$$
$$=\max_{\mathcal{V}(1/2,...,1/2)}\min_{k=1,2}\left(\frac{1}{2}B_n^k\left(\frac{1}{2}\right)-\sum_{\substack{1\leq i\leq b-1\\ i\ odd}}\left(B_n^k\left(\sum_{j=1}^{i}v_j\right)-B_n^k\left(\sum_{j=1}^{i-1}v_j\right)\right)\right)+o_{\mathbb{P}}(1), \qquad (3.11)$$

where the maximum in (3.11) is taken over

$$\mathcal{V}(p_1,...,p_{b-1})=\left\{v=(v_1,...,v_{b-1})\in[0,1]^{b-1}:\forall i=1,...,b-1,\ \sum_{j=1}^{i}v_j\leq p_i\right\}, \qquad (3.12)$$

and where B^1, B^2 are two standard Brownian motions.

Next, notice that

$$f(g^1, g^2) := \max_{\mathcal{V}(1/2,...,1/2)} \min_{k=1,2} \left(\frac{1}{2} g^k\left(\frac{1}{2}\right) - \sum_{\substack{1 \le i \le b-1 \\ i \ odd}} \left(g^k\left(\sum_{j=1}^{i} v_j\right) - g^k\left(\sum_{j=1}^{i-1} v_j\right)\right)\right),$$

is continuous in its domain, where $f : C[0,1] \times C[0,1] \to R$. Therefore according to the continuous mapping theorem,

$$\begin{aligned} \frac{LCbB_n^0 - n/2}{\sqrt{2n}} &= f(B_n^1, B_n^2) + o_{\mathbb{P}}(1) \\ &\Longrightarrow f(B^1, B^2) \\ &= \max_{\mathcal{V}(1/2,...,1/2)} \min_{k=1,2} \left(\frac{1}{2} B^k\left(\frac{1}{2}\right) - \sum_{\substack{1 \le i \le b-1 \\ i \ odd}} \left(B^k\left(\sum_{j=1}^{i} v_j\right) - B^k\left(\sum_{j=1}^{i-1} v_j\right)\right)\right). \end{aligned} \tag{3.13}$$

To proceed, define the substitution $t_i = 2\sum_{j=1}^{i} v_j$, by Brownian scaling, (3.13) follows that:

$$\begin{aligned} \frac{LCbB_n^0 - n/2}{\sqrt{n}} \Longrightarrow & \max_{\substack{0 = t_0 \le t_1 \le ... \\ \le t_{b-1} \le t_b = 1}} \min_{k=1,2} \left(\frac{1}{2} B^k(1) - \sum_{\substack{1 \le i \le b-1 \\ i \ odd}} \left(B^k(t_i) - B^k(t_{i-1})\right)\right) \\ = & \max_{\substack{0 = t_0 \le t_1 \le ... \\ \le t_{b-1} \le t_b = 1}} \min_{k=1,2} \left(\frac{1}{2} B^k(1) + \sum_{i=1}^{b-1} (-1)^i B^k(t_i)\right). \end{aligned} \tag{3.14}$$

Note that for any given $0 = t_0 \le t_1 \le ... \le t_{b-1} \le t_b = 1$, the vector $\Big(B(t_1), ..., B(t_{b-1}), B(t_b)\Big)$ is Gaussian, indicating that, above, the linear combination, is univariate normal and hence symmetric. A re-arrangement yields

$$\frac{LCbB_n^0 - n/2}{\sqrt{n}/2} \Longrightarrow \max_{\substack{0 = t_0 \le t_1 \le ... \\ \le t_{b-1} \le t_b = 1}} \min_{k=1,2} \left(\sum_{i=1}^{b} (-1)^{i+1} \Big(B^k(t_i) - B^k(t_{i-1})\Big)\right). \tag{3.15}$$

Finally, for b odd, most of combinatorial analysis still holds except we are now counting for zeros in the last block. After minor modification, to deal in particular with this difference, the following asymptotic behavior, which is similar with (3.14) follows:

$$\frac{LCbB_n^0 - n/2}{\sqrt{n}} \Longrightarrow \max_{\substack{0 = t_0 \le t_1 \le \dots \\ \le t_{b-1} \le t_b = 1}} \min_{k=1,2} \left(\frac{1}{2} B^k(1) - \sum_{\substack{1 \le i \le b-1 \\ i \ even}} \left(B^k(t_i) - B^k(t_{i-1}) \right) \right)$$

$$= \max_{\substack{0 = t_0 \le t_1 \le \dots \\ \le t_{b-1} \le t_b = 1}} \min_{k=1,2} \left(\frac{1}{2} B^k(1) + \sum_{j=1}^{b-1} (-1)^{i+1} B^k(t_i) \right),$$

which leads to (3.15) as well. The proof for the first part of Theorem 1.1 is thereby finished by following the same path for $LCbB_n^1$, which is identically distributed with $LCbB_n^0$ in the uniform case, and finally taking the maximum.

Remark 3.3. First, for uniform draw in the case of three or more independent sequences, the rationale before is also valid and limiting results become the max-min of three or more independent Brownian functionals. Moreover, let us now show that the limiting behavior of $LCbB_n^0$ above agrees with the one of the binary LCI_n in the uniform case. Indeed, for $b = 2$, part (i) of Theorem 1.1 rewrites:

$$\frac{LC2B_n^0 - n/2}{\sqrt{n}} \Longrightarrow \max_{0 \le t \le 1} \min \left(-\frac{1}{2} B^1(1) + B^1(t), -\frac{1}{2} B^2(1) + B^2(t) \right). \tag{3.16}$$

On the other hand, in [10] it is shown that for binary uniform case,

$$\frac{LCI_n - n/2}{\sqrt{n/2}} \Longrightarrow \max_{0 \le t \le 1} \min \Bigg(-\frac{1}{2} B_1^{(1)}(1) + \frac{1}{2} B_1^{(2)}(1) - B_1^{(2)}(t) + B_1^{(1)}(t)$$

$$- \frac{1}{2} B_2^{(1)}(1) + \frac{1}{2} B_2^{(2)}(1) - B_2^{(2)}(t) + B_2^{(1)}(t) \Bigg), \tag{3.17}$$

where B_1 and B_2 are two standard 2-dimensional Brownian motion on $[0, 1]$ with independent components.

Now for all $t \in [0,1]$, set the following two pointwise transformations:

$$B^1(t) = (B_1^{(2)}(t) - B_1^{(1)}(t))/\sqrt{2},$$

and

$$B^2(t) = (B_2^{(2)}(t) - B_2^{(1)}(t))/\sqrt{2}.$$

Then B^1, B^2 clearly are two 1-dimensional standard Brownian motions, which shows (3.17) agrees with (3.16).

CHAPTER 4

THE NON-UNIFORM CASE

This section analyzes the non-uniform setting. As indicated in Theorem 1.1, four cases need to be considered:

1. $p_1 = p_2 \neq 1/2$.
2. $p_1 < p_2 < 1/2$, or $1/2 < p_1 < p_2$.
3. $p_1 < p_2 = 1/2$, or $1/2 = p_1 < p_2$.
4. $p_1 < 1/2 < p_2$.

As before mentioned in Chapter 1, one dominant letter exists in case 1 and case 2 and therefore the starting order is expected to be inconsequential. On the other hand, $LCbB_n$ must be determined by both sequences in case 1 and case 4. Let us start with the first case, $p_1 = p_2 = p \neq 1/2$.

4.1 Proof for $p_1 = p_2 = p \neq 1/2$

Proof. Without loss of generality, let us first investigate when $p_1 = p_2 = p > 1/2$. Again let us begin with $LCbB_n^0$ and assume b even. In the constraints, let $k_i' = \max(k_i - 1, 0)$ for i odd; and $k_i' = (1-p)(k_i - 1)/p + 1$ for i even. Then the original constraint $\mathcal{C}_n$ becomes:

$$\mathcal{C}_n^{'} = \left\{ (k_1', ..., k_{b-1}') : k_1' \in \mathcal{C}_{n,1}^{'}, k_2' \in \mathcal{C}_{n,2}^{'}(k_1'), ..., k_{b-1}' \in \mathcal{C}_{n,b-1}^{'}(k_1', ..., k_{b-2}') \right\}, \tag{4.1}$$

$$where\ \mathcal{C}_{n,1}^{'} = \left\{ 0 \leq k_1' \leq (N_0(X)) \wedge (N_0(Y)) \right\},\ and\ for\ i = 2, ..., b-1,$$

$$\mathcal{C}_{n,i}^{'}(k_1', ..., k_i') = \Big\{ 0 \leq k_i' \leq \big(N_0(X) - N_i^*(X) - 1 \big) \tag{4.2}$$

$$\wedge \big(N_0(Y) - N_i^*(Y) - 1 \big) \Big\},\ for\ i\ odd;$$

$$= \Big\{ 0 \leq k_i' \leq \frac{1-p}{p} \Big(\big(N_1(X) - N_i^*(X) \big) \tag{4.3}$$

$$\wedge \big(N_1(Y) - N_i^*(Y) \big) \Big) + \frac{2p-1}{p} \Big\},\ for\ i\ even.$$

The polygonal process B_n' is now the linear interpolation between the

$$B_n' \left(\frac{k}{n} \right) = \sum_{j=1}^{k+1} \frac{Z_j}{\sqrt{n}},\ for\ k = 0, 1, ..., n, \tag{4.4}$$

where

$$Z_j = (N_1^{T_0^{j-1}, T_0^j} - p/(1-p)) / \sqrt{p/(1-p)^2}.$$

Again B_n' has one more term than B_n, the Brownian approximation of the sum of i.i.d. normalized random variables Z_j's. Let B_n^1, B_n^2 be two copies of B_n, when $p > 1/2$, from (2.9),

$$\frac{LCbB_n^0 - np}{\sqrt{\frac{pn}{(1-p)^2}}} = \max_{\mathcal{C}_n} \min \left(\left(\frac{\frac{1-2p}{1-p}(k_1+k_3+...+k_{b-1}) + \frac{p}{1-p}\frac{b}{2} + (1-p)S_{0,1}^{(n)}(X) - (T_0^1(X)-1)\mathbb{1}_{\{k_1>0\}}}{\sqrt{\frac{pn}{(1-p)^2}}} \right.\right.$$
$$+(1-p)\sum_{j=1}^{N_0(X)} \left(\frac{N_1^{T_0^{j-1},T_0^j}(X) - \frac{p}{1-p}}{\sqrt{\frac{pn}{(1-p)^2}}} \right) - \sum_{\substack{1\le i\le b-1 \\ i\ odd}} \sum_{j=N_i^*(X)+2}^{N_i^*(X)+k_i} \left(\frac{N_1^{T_0^{j-1},T_0^j}(X) - \frac{p}{1-p}}{\sqrt{\frac{pn}{(1-p)^2}}} \right) \Bigg),$$
$$\left(\frac{\frac{1-2p}{1-p}(k_1+k_3+...+k_{b-1}) + \frac{p}{1-p}\frac{b}{2} + (1-p)S_{0,1}^{(n)}(Y) - (T_0^1(Y)-1)\mathbb{1}_{\{k_1>0\}}}{\sqrt{\frac{pn}{(1-p)^2}}} \right.$$
$$\left.\left. +(1-p)\sum_{j=1}^{N_0(Y)} \left(\frac{N_1^{T_0^{j-1},T_0^j}(Y) - \frac{p}{1-p}}{\sqrt{\frac{pn}{(1-p)^2}}} \right) - \sum_{\substack{1\le i\le b-1 \\ i\ odd}} \sum_{j=N_i^*(Y)+2}^{N_i^*(Y)+k_i} \left(\frac{N_1^{T_0^{j-1},T_0^j}(Y) - \frac{p}{1-p}}{\sqrt{\frac{pn}{(1-p)^2}}} \right) \right) \right)$$
$$= \max_{\mathcal{C}_n'} \min_{k=1,2} \left(\frac{\frac{1-2p}{1-p}(k_1' + k_3' + ... + k_{b-1}')}{\sqrt{\frac{pn}{(1-p)^2}}} + (1-p)B_n^k\left(\frac{N_0-1}{n}\right) \right.$$
$$\left. - \sum_{\substack{1\le i\le b-1 \\ i\ odd}} \left(B_n^k\left(\frac{N_i^* + k_i'}{n}\right) - B_n^k\left(\frac{N_i^*}{n}\right) \right) \right) + o_{\mathbb{P}}(1). \qquad (4.5)$$

Based on the analysis of the uniform case, particularly combining (2.12) and the following results on the linear and increments terms, we have

$$\frac{LCbB_n^0 - np}{\sqrt{\frac{pn}{(1-p)^2}}} = \max_{\mathcal{C}_n'} \min_{k=1,2} \left(\frac{\frac{1-2p}{1-p}(k_1' + k_3' + ... + k_{b-1}')}{\sqrt{\frac{pn}{(1-p)^2}}} + (1-p)B_n^k(1-p) \right.$$
$$\left. - \sum_{\substack{1\le i\le b-1 \\ i\ odd}} \left(B_n^k\left(\sum_{j=1}^{i} \frac{k_j'}{n}\right) - B_n^k\left(\sum_{j=1}^{i-1} \frac{k_j'}{n}\right) \right) \right) + o_{\mathbb{P}}(1). \qquad (4.6)$$

Since $\mathbb{E}(N_i^*) = \sum_{j=1}^{i-1} k_j'$, for i odd; and $\mathbb{E}(N_i^*) = (k_1' + k_3' + ... + k_{i-1}')p/(1-p) + (i/2 - 1)(1-2p)/(1-p)$, for i even, the limit for the constraint $\mathcal{C}_{n,i}^{'}(k_1', ..., k_{i-1}')$ becomes:

$$\left\{ (k_1', ..., k_{b-1}') : 0 \le k_i' \le (1-p)n - \sum_{j=1}^{i-1} k_j' \right\},$$

or equivalently,

$$\left\{(k_1',...,k_{b-1}') : \frac{1}{n}\sum_{j=1}^{i-1} k_j' \le \frac{1}{n}\sum_{j=1}^{i} k_j' \le (1-p),\ i=1,...,b-1\right\}.$$

Denote k_i'/n by v_i, for $i = 1, ..., b-1$, then $\mathcal{C}_n'$ is replaced by $\mathcal{V}(1-p, ..., 1-p)$, where $\mathcal{V}(\bullet, ..., \bullet)$ is defined in (3.12). The proof of this heuristic claim can be done by arguments as in the uniform case. Summarizing the results so far gives the following lemma:

Lemma 4.1.

$$\frac{LCbB_n^0 - np}{\sqrt{\frac{pn}{(1-p)^2}}} = \max_{\mathcal{V}(1-p,...,1-p)} \min_{k=1,2}\left(\frac{\frac{1-2p}{1-p}\sqrt{n}(v_1+v_3+...+v_{b-1})}{\sqrt{\frac{p}{(1-p)^2}}} + (1-p)B_n^k(1-p) - \sum_{\substack{1\le i\le b-1\\ i\ odd}}\left(B_n^k\left(\sum_{j=1}^{i} v_j\right) - B_n^k\left(\sum_{j=1}^{i-1} v_j\right)\right)\right) + o_{\mathbb{P}}(1). \tag{4.7}$$

After another linear transformation of the constraints, $t_i = \sum_{j=1}^{i} v_j/(1-p)$, (4.7) becomes

$$\frac{LCbB_n^0 - np}{\sqrt{\frac{pn}{(1-p)^2}}} = \max_{\substack{0=t_0\le t_1\le ...\\ \le t_{b-1}\le t_b=1}} \min_{k=1,2}\left(\frac{\frac{1-2p}{1-p}\sqrt{n}\sum_{i=1}^{b-1}(-1)^{i-1}(1-p)t_i}{\sqrt{\frac{p}{(1-p)^2}}} - \left(\sum_{i=1}^{b-1}(-1)^{i-1}B_n^k\left((1-p)t_i\right)\right) + (1-p)B_n^k(1-p)\right) + o_{\mathbb{P}}(1). \tag{4.8}$$

Our next goal is to show that the right-hand side of the first two lines in (4.8) converge to 0 in probability. Let

$$c_n = -(1-2p)\sqrt{n/p} > 0,$$

and for $c > 0$, let

$$M_c = \max_{\substack{0 = t_0 \le t_1 \le ... \\ \le t_{b-1} \le t_b = 1}} \left(\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) - c(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \right).$$

For n sufficiently large,

$$\begin{aligned} &\sum_{i=1}^{b-1} (-1)^i B_n\left((1-p)t_i\right) - c(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \\ &\quad \ge \sum_{i=1}^{b-1} (-1)^i B_n\left((1-p)t_i\right) - c_n(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right). \end{aligned}$$

Thus for sufficiently large n and any $z > 0$,

$$\begin{aligned} &\mathbb{P}\left(\max_{\substack{0=t_0\le t_1\le ... \\ \le t_{b-1}\le t_b=1}} \left(\sum_{i=1}^{b-1} (-1)^i B_n\left((1-p)t_i\right) - c_n(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \right) > z \right) \\ &\quad \le \mathbb{P}\left(\max_{\substack{0=t_0\le t_1\le ... \\ \le t_{b-1}\le t_b=1}} \left(\sum_{i=1}^{b-1} (-1)^i B_n\left((1-p)t_i\right) - c(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \right) > z \right). \end{aligned}$$

Hence,

$$\begin{aligned} &\limsup_{n\to\infty} \mathbb{P}\left(\max_{\substack{0=t_0\le t_1\le ... \\ \le t_{b-1}\le t_b=1}} \left(\sum_{i=1}^{b-1} (-1)^i B_n\left((1-p)t_i\right) - c_n(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \right) > z \right) \\ &\le \lim_{n\to\infty} \mathbb{P}\left(\max_{\substack{0=t_0\le t_1\le ... \\ \le t_{b-1}\le t_b=1}} \left(\sum_{i=1}^{b-1} (-1)^i B_n\left((1-p)t_i\right) - c(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \right) > z \right) \\ &= \mathbb{P}\left(M_c > z\right), \end{aligned}$$

using the Continuous Mapping Theorem and the Invariance Principle to obtain the last equality.

It remains to show that $\mathbb{P}(M_c > z) \to 0$, as $c \to \infty$. Since $0 \le t_1 \le t_2 \le ... \le t_{b-1} \le 1$ implies $t_i \ge 0$ for $i = 1, ..., b-1$ and $t_i \ge t_{i-1}$ for $1 \le i \le b-1$, i odd and

$0 \leq t_1 - t_2 + ... + t_{b-1} \leq 1$, it follows that (to reduce the burden of notation, for terms on the right-hand side of each "$\leq$" below we ignore the words "all $t_i \geq 0$ and $t_i \geq t_{i-1}$ for i odd" in the subscripts),

$$\max_{\substack{0=t_0 \leq t_1 \leq ... \\ \leq t_{b-1} \leq t_b = 1}} \left(\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) - c(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \right) \leq$$

$$\max_{0 \leq t_1 - t_2 + ... + t_{b-1} \leq 1} \left(\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) - c(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right) \right).$$

So for $z > 0$, $c > 0$, and $0 < \varepsilon \leq 1$,

$$\begin{aligned}
\mathbb{P}\left(M_c > z\right) &= \mathbb{P}\left(\max_{\substack{0=t_0 \leq t_1 \leq ... \\ \leq t_{b-1} \leq t_b = 1}} \left(\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) - c(1-p)(t_1 - t_2 + ... + t_{b-1}) \right) > z \right) \\
&\leq \mathbb{P}\left(\max_{0 \leq t_1 - t_2 + ... + t_{b-1} \leq 1} \left(\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) - c(1-p)(t_1 - t_2 + ... + t_{b-1}) \right) > z \right) \\
&\leq \mathbb{P}\left(\max_{0 \leq t_1 - t_2 + ... + t_{b-1} \leq \varepsilon} \left(\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) - c(1-p)(t_1 - t_2 + ... + t_{b-1}) \right) > z \right) \\
&+ \mathbb{P}\left(\max_{\varepsilon \leq t_1 - t_2 + ... + t_{b-1} \leq 1} \left(\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) - c(1-p)(t_1 - t_2 + ... + t_{b-1}) \right) > z \right) \\
&\leq \mathbb{P}\left(\max_{0 \leq t_1 - t_2 + ... + t_{b-1} \leq \varepsilon} \sum_{i=1}^{b-1} (-1)^i B((1-p)t_i) > z \right) \\
&+ \mathbb{P}\left(\max_{0 \leq t_1 - t_2 + ... + t_{b-1} \leq 1} \sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right) > c(1-p)\varepsilon + z \right).
\end{aligned}$$

Note the independent increments ensure that for any given $(t_1, ..., t_{b-1})$, $\sum_{i=1}^{b-1} (-1)^i B\left((1-p)t_i\right)$ being normal with zero mean and variance $(1-p)(t_1 - t_2 + ... + t_{b-1})$. Also, since

$$\mathbb{P}\left(\max_{0 \leq s \leq t} B(s) > x \right) = 2(1 - \Phi(x/\sqrt{t})).$$

Therefore,

$$\begin{aligned}\mathbb{P}(M_c > z) \leq\ & 2\left(1 - \Phi(z/\sqrt{(1-p)\varepsilon})\right) \\ & + 2\left(1 - \Phi((c(1-p)\varepsilon + z)/(\sqrt{1-p}))\right).\end{aligned}$$

Now since c and ε are arbitrary, taking first $c \to +\infty$ and then $\varepsilon \to 0$, shows that

$$\mathbb{P}(M_c > z) \to 0.$$

Therefore,

$$\max_{\substack{0=t_0\leq t_1\leq ... \\ \leq t_{b-1}\leq t_b=1}} \left(\sum_{i=1}^{b-1}(-1)^i B_n\left((1-p)t_i\right) - c_n(1-p)\left(t_1 - t_2 + ... + t_{b-1}\right)\right) \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$. Finally after Brownian scaling, and for $p > 1/2$,

$$\frac{LCbB_n^0 - pn}{\sqrt{p(1-p)n}} \Longrightarrow \min(Z_1, Z_2),$$

as $n \to \infty$, and $Z_1, Z_2 \sim N(0,1)$.

It is noteworthy that from the proof above (as well as the progress in the end of Chapter 2), we conclude that everything in (2.9) has order at most $\sqrt{n}$, except for $(1-2p_k)(k_1 + k_3 + ... + k_{b-1})/(1-p_k) + p_k n$. In other words, when $p_1 \leq p_2 < 1/2$ or $1/2 < p_1 \leq p_2$, the terms $p_1 n$ and $(1-p_2)n$ will be the dominating ones in the general combinatorial expression (2.10), as n tends to infinity. Therefore, when $p_1 = p_2 = p > 1/2$, $LCbB_n^0$ is asymptotically greater than $LCbB_n^1$ and will be chosen for representing $LCbB_n$, thus completing the proof of (i).

Remark 4.2. In this section, we have shown that

$$(1-2p)\sqrt{n/p}\sum_{i=1}^{b-1}(-1)^{i-1}(1-p)t_i + \sum_{i=1}^{b-1}(-1)^i B_n((1-p)t_i) \xrightarrow{\mathbb{P}} 0,$$

when $p > 1/2$. Actually, this property also holds for $p < 1/2$. Indeed, for $p < 1/2$ we just need to simply show that instead

$$-(1-2p)\sqrt{n/p}\sum_{i=1}^{b-1}(-1)^{i-1}(1-p)t_i + \sum_{i=1}^{b-1}(-1)^{i+1}B_n((1-p)t_i) \xrightarrow{\mathbb{P}} 0.$$

Let again $c_n = (1-2p)\sqrt{n/p} > 0$, and also taking advantage of the symmetry of Brownian motion, the conclusion follows. However, this will no longer finish the proof for Theorem 1.1 (ii) since $LCbB_n$ will be centered at $(1-p)n$ when $p < 1/2$, and in this case we will switch to the second max-min functional. Details are given below.

Let us continue to investigate $p_1 = p_2 = p < 1/2$, and in this case $(1-p)n$ will dominate in (2.10). Therefore $LCbB_n$ will have the same limiting functional as $LCbB_n^1$. In similarity to (2.9),

$$\begin{aligned}
\frac{LCbB_n^1 - (1-p)n}{\sqrt{\frac{(1-p)n}{p^2}}} &= \max_{\tilde{\mathcal{C}}_n}\min_{X,Y}\left(\frac{\frac{2p-1}{p}(\tilde{k}_1+\tilde{k}_3+...+\tilde{k}_{b-1})+\frac{1-p}{p}\frac{b}{2}+p\tilde{S}_{0,1}^{(n)}-(T_1^1-1)\mathbb{1}_{\{k_1>0\}}}{\sqrt{\frac{(1-p)n}{p^2}}}\right.\\
&\qquad \left. +p\sum_{j=1}^{N_1}\left(\frac{N_0^{T_1^{j-1},T_1^j}-\frac{1-p}{p}}{\sqrt{\frac{(1-p)n}{p^2}}}\right)-\sum_{\substack{1\le i\le b-1\\ i\ odd}}\sum_{j=N_i^*+2}^{N_i^*+k_i}\left(\frac{N_0^{T_1^{j-1},T_1^j}-\frac{1-p}{p}}{\sqrt{\frac{(1-p)n}{p^2}}}\right)\right)\\
&= \max_{\tilde{\mathcal{C}}_n'}\min_{k=1,2}\left(\frac{\frac{2p-1}{p}(\tilde{k}_1'+\tilde{k}_3'+...+\tilde{k}_{b-1}')}{\sqrt{\frac{(1-p)n}{p^2}}}+pB_n^k\left(\frac{N_0-1}{n}\right)\right.\\
&\qquad \left.-\sum_{\substack{1\le i\le b-1\\ i\ odd}}\left(B_n^k\left(\frac{N_i^*+\tilde{k}_i'}{n}\right)-B_n^k\left(\frac{N_i^*}{n}\right)\right)\right)+o_{\mathbb{P}}(1),
\end{aligned}$$

where $\tilde{k}_i^{'} = \max(\tilde{k}_i - 1, 0)$ for i odd; $p(\tilde{k}_i - 1)/(1-p) + 1$ for i even, and where

$$\tilde{\mathcal{C}}_n^{'} = \left\{(\tilde{k}_1^{'}, ..., \tilde{k}_{b-1}^{'}) : \tilde{k}_1^{'} \in \tilde{\mathcal{C}}_{n,1}^{'}, \tilde{k}_2^{'} \in \tilde{\mathcal{C}}_{n,2}^{'}(\tilde{k}_1^{'}), ..., \tilde{k}_{b-1}^{'} \in \tilde{\mathcal{C}}_{n,b-1}^{'}(\tilde{k}_1^{'}, ..., \tilde{k}_{b-2}^{'})\right\},$$

$$\textit{with } \tilde{\mathcal{C}}_{n,1}^{'} = \left\{0 \leq \tilde{k}_1^{'} \leq (N_1(X)) \wedge (N_1(Y))\right\}, \textit{ and for } i = 2, ..., b-1,$$

$$\tilde{\mathcal{C}}_{n,i}^{'}(\tilde{k}_1^{'}, ..., \tilde{k}_i^{'}) = \left\{0 \leq \tilde{k}_i^{'} \leq (N_1(X) - \tilde{N}_i^*(X) - 1) \right.$$

$$\left. \wedge (N_1(Y) - \tilde{N}_i^*(Y) - 1)\right\}, \textit{ for i odd};$$

$$\textit{and } \tilde{\mathcal{C}}_{n,i}^{'}(\tilde{k}_1^{'}, ..., \tilde{k}_i^{'}) = \left\{0 \leq \tilde{k}_i^{'} \leq \left(\frac{p}{1-p}(N_0(X) - \tilde{N}_i^*(X)) + \frac{1-2p}{1-p}\right)\right.$$

$$\left. \wedge \left(\frac{p}{1-p}(N_0(Y) - \tilde{N}_i^*(Y)) + \frac{1-2p}{1-p}\right)\right\}, \textit{ for i even}.$$

Finally, let $\tilde{v}_i = \tilde{k}_i^{'}/n$, it follows

$$\frac{LCbB_n^1 - (1-p)n}{\sqrt{\frac{(1-p)n}{p^2}}} = \max_{\mathcal{V}(p,...,p)} \min_{k=1,2} \left(\frac{\frac{1-2p}{1-p}\sqrt{n}(\tilde{v}_1 + \tilde{v}_3 + ... + \tilde{v}_{b-1})}{\sqrt{\frac{p}{(1-p)^2}}} + pB_n^k(p) \right.$$

$$\left. - \sum_{\substack{1 \leq i \leq b-1 \\ i \text{ odd}}} \left(B_n^k\left(\sum_{j=1}^{i} \tilde{v}_j\right) - B_n^k\left(\sum_{j=1}^{i-1} \tilde{v}_j\right)\right)\right) + o_{\mathbb{P}}(1)$$

$$\Longrightarrow \min\big(pB_1(p), pB_2(p)\big), \tag{4.9}$$

where to prove this last convergence result one proceeds as done for $p > 1/2$. Hence for $p < 1/2$,

$$\frac{LCbB_n^1 - (1-p)n}{\sqrt{p(1-p)n}} \Longrightarrow \min(Z_1, Z_2).$$

Therefore, when $p \neq 1/2$,

$$\frac{LCbB_n^1 - n\max(p, 1-p)}{\sqrt{p(1-p)n}} \Longrightarrow \min(Z_1, Z_2),$$

as $n \to \infty$, where $Z_1, Z_2 \sim N(0,1)$. Since, in this case and as already indicated, $LCbB_n$

has the same limiting functional as $LCbB_n^1$, the proof of Theorem 1.1 (ii) is complete. $\square$

Remark 4.3. It was conjectured in [10] that the LCI_n of two sequences with arbitrary distributions have the following limiting behavior: Let $X = (X_i)_{i\geq 1}$ and $Y = (Y_i)_{i\geq 1}$ be two sequences of i.i.d. random variables and have the same distribution. They both take values in $\mathcal{A}_m = \{\alpha_1 < ... < \alpha_m\}$, and let $p_{max} = \max_{i\in\{1,...,m\}} \mathbb{P}(X_1 = \alpha_i)$ and k be its multiplicity. Then,

$$\frac{LCI_n - np_{max}}{\sqrt{np_{max}}} \Longrightarrow \max_{\substack{0=t_0\leq t_1\leq ... \\ \leq t_{b-1}\leq t_k=1}} \min\left(\frac{\sqrt{1-kp_{max}}-1}{k}\sum_{i=1}^{k} B_1^{(i)}(1) + \sum_{i=1}^{k}\left(B_1^{(i)}(t_i) - B_1^{(i)}(t_{i-1})\right),\right.$$
$$\left.\frac{\sqrt{1-kp_{max}}-1}{k}\sum_{i=1}^{k} B_2^{(i)}(1) + \sum_{i=1}^{k}\left(B_2^{(i)}(t_i) - B_2^{(i)}(t_{i-1})\right)\right),$$

where B_1 and B_2 are two k-dimensional standard Brownian motions on $[0, 1]$, and it is not hard to verify that Theorem 1.1 (ii) also agrees with this conjecture for binary case.

4.2 Proof for $p_1 < p_2 < 1/2$, or $1/2 < p_1 < p_2$

Proof. Let us start with the proof of Theorem 1.1 (iii). First, from the previous sections, we know that $LCbB_n$ is represented either by $LCbB_n^0$, if $1/2 < p_1 < p_2$, or by $LCbB_n^1$, if $p_1 < p_2 < 1/2$. Also, given $p_1 < p_2$, in the constraints (4.7) and (4.9) p is respectively replaced by $\min(1 - p_1, 1 - p_2) = 1 - p_2$, and $\min(p_1, p_2) = p_1$. Thus, from (4.8),

$$\frac{LCbB_n^0}{\sqrt{n}} = \max_{\substack{0=t_0\leq t_1\leq ... \\ \leq t_{b-1}\leq t_b=1}} \min_{k=1,2}\left(\frac{1-2p_k}{1-p_k}\sqrt{n}\sum_{i=1}^{b-1}(-1)^{i-1}(1-p_2)t_i + p_k\sqrt{n} + \sqrt{p_k}B_n^k(1-p_k)\right.$$
$$\left. - \frac{\sqrt{p_k}}{1-p_k}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^k\left((1-p_2)t_i\right)\right) + o_{\mathbb{P}}(1), \qquad (4.10)$$

and,

$$\frac{LCbB_n^1}{\sqrt{n}} = \max_{\substack{0=t_0\le t_1\le \dots \\ \le t_{b-1}\le t_b=1}} \min_{k=1,2} \left(\frac{2p_k-1}{p_k}\sqrt{n}\sum_{i=1}^{b-1}(-1)^{i-1}p_1 t_i + (1-p_k)\sqrt{n} + \sqrt{1-p_k}B_n^k(p_k) \right.$$
$$\left. - \frac{\sqrt{1-p_k}}{p_k}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^k\left(p_1 t_i\right) \right) + o_{\mathbb{P}}(1). \quad (4.11)$$

Next, we claim that the two max-min in (4.10) and (4.11) are respectively attained for $k=1$ and $\sum_{i=1}^{b-1}(-1)^{i-1}t_i = 0$, and for $k=2$ and $\sum_{i=1}^{b-1}(-1)^{i-1}t_i = 0$. Indeed, first note that the functionals of B_n^k are of order less than $\sqrt{n}$. Thus, the inner minima in (4.10) (resp. (4.11)), is attained for $k=1$ (resp. $k=1$), when $\sum_{i=1}^{b-1}(-1)^{i-1}t_i < 1-p_1$ (resp. $\sum_{i=1}^{b-1}(-1)^{i-1}t_i > p_2$), and is attained for $k=2$ (resp. $k=2$), when $\sum_{i=1}^{b-1}(-1)^{i-1}t_i \ge 1-p_1$ (resp. $\sum_{i=1}^{b-1}(-1)^{i-1}t_i \le p_2$).

In addition, under the assumption $1/2 < p_1 < p_2$, depending on the sign of $1-2p_k$, the max-min functional on the right-hand-side of (4.10) is of order $p_1\sqrt{n}$, when $\sum_{i=1}^{b-1}(-1)^{i-1}t_i = 0$; and is of order $(2p_1p_2 - p_1 - p_2 + 1)\sqrt{n}$, when $\sum_{i=1}^{b-1}(-1)^{i-1}t_i \to 1-p_1$. However, $2p_1p_2 - p_1 - p_2 + 1 < p_1$, when $p_1 > 1/2$. Therefore the max-min in (4.10) is of order $p_1\sqrt{n}$, attained for $k=1$ and $\sum_{i=1}^{b-1}(-1)^{i-1}t_i = 0$. The second part of the claim, which indicates that the max-min in (4.11) is of order $(1-p_2)\sqrt{n}$, can be proved similarly.

Thus, in similarity of (4.8), for $1/2 < p_1 < p_2$,

$$\frac{LCbB_n^0 - p_1 n}{\sqrt{n}} = \max_{\substack{0=t_0\le t_1\le \dots \\ \le t_{b-1}\le t_b=1}} \left(\frac{1-2p_1}{1-p_1}\sqrt{n}\sum_{i=1}^{b-1}(-1)^{i-1}(1-p_2)t_i + \sqrt{p_1}B_n^1(1-p_1) \right.$$
$$\left. - \frac{\sqrt{p_1}}{1-p_1}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^1\left((1-p_2)t_i\right) \right) + o_{\mathbb{P}}(1)$$
$$\Longrightarrow \sqrt{p_1}B(1-p_1), \quad (4.12)$$

as $n\to\infty$, and $t_1 - t_2 + \dots + t_{b-1} \to 0$. Then, $(LCbB_n^0 - p_1 n)/\sqrt{np_1(1-p_1)} \Longrightarrow N(0,1)$,

simply noting that $B(1-p_1)$ is normal.

For $p_1 < p_2 < 1/2$, $LCbB_n$ will then be represented by $LCbB_n^1$, and based on proofs from the previous subsections, it is then trivial to show that

$$\frac{LCbB_n^1 - (1-p_2)n}{\sqrt{np_2(1-p_2)}} \Longrightarrow N(0,1),$$

which completes the proof of Theorem 1.1 (iii). □

4.3 Proof for $p_1 < p_2 = 1/2$, or $1/2 = p_1 < p_2$

Proof. Let us continue to Theorem 1.1 (iv). First, without loss of generality, let us first assume $1/2 = p_1 < p_2$ and b even. Now (4.10) and (4.11) respectively become:

$$\frac{LCbB_n^0}{\sqrt{n}} = \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=1}} \left(\left(\frac{\sqrt{n}}{2} + \sqrt{\frac{1}{2}} B_n^1\left(\frac{1}{2}\right) - \sqrt{2}\sum_{i=1}^{b-1}(-1)^{i-1} B_n^1\left((1-p_2)t_i\right) \right) \right.$$
$$\left. \wedge \left(\left((1-2p_2)\sum_{i=1}^{b-1}(-1)^{i-1}t_i + p_2\right)\sqrt{n} + \sqrt{p_2}B_n^2(1-p_2) - \frac{\sqrt{p_2}}{1-p_2}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^2\left((1-p_2)t_i\right) \right) \right)$$
$$+o_{\mathbb{P}}(1), \qquad (4.13)$$

and,

$$\frac{LCbB_n^1}{\sqrt{n}} = \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=1}} \left(\left(\frac{\sqrt{n}}{2} + \sqrt{\frac{1}{2}} B_n^1\left(\frac{1}{2}\right) - \sqrt{2}\sum_{i=1}^{b-1}(-1)^{i-1} B_n^1\left(\frac{t_i}{2}\right) \right) \right.$$
$$\left. \wedge \left(\left(\frac{2p_2-1}{2p_2}\sum_{i=1}^{b-1}(-1)^{i-1}t_i + 1 - p_2\right)\sqrt{n} + \sqrt{1-p_2}B_n^2(p_2) - \frac{\sqrt{1-p_2}}{p_2}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^2\left(\frac{t_i}{2}\right) \right) \right)$$
$$+o_{\mathbb{P}}(1). \qquad (4.14)$$

Clearly, in (4.13) the inner minimum is attained at the second term when $\sum_{i=1}^{b-1}(-1)^{i-1}t_i \le 1/2$; and is attained at the first term, when $\sum_{i=1}^{b-1}(-1)^{i-1}t_i > 1/2$. Note that under the

assumption $1/2 = p_1 < p_2$, $(1-2p_2)\sum_{i=1}^{b-1}(-1)^{i-1}t_i + p_2$ decreases with respect to $\sum_{i=1}^{b-1}(-1)^{i-1}t_i$. Therefore, to maximize the order of the attained minimum at the second line, we need to force $\sum_{i=1}^{b-1}(-1)^{i-1}t_i$ to be $1/2$, in other words, (4.13) becomes

$$\frac{LCbB_n^0}{\sqrt{n}} = \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=1,\\ 0\le t_1-t_2+\dots+t_{b-1}\le\frac{1}{2}}} \left(\frac{\sqrt{n}}{2}+\sqrt{\frac{1}{2}}B_n^1\left(\frac{1}{2}\right)-\sqrt{2}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^1\big((1-p_2)t_i\big)\right)+o_{\mathbb{P}}(1),$$

then, by Donsker's Theorem, followed by another Brownian scaling,

$$\begin{aligned}\frac{LCbB_n^0-n/2}{\sqrt{n}} &= \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=1,\\ 0\le t_1-t_2+\dots+t_{b-1}\le\frac{1}{2}}} \left(\sqrt{\frac{1}{2}}B_n^1\left(\frac{1}{2}\right)-\sqrt{2}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^1\big((1-p_2)t_i\big)\right)+o_{\mathbb{P}}(1)\\ &\Longrightarrow \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=1,\\ 0\le t_1-t_2+\dots+t_{b-1}\le\frac{1}{2}}} \left(\frac{1}{2}B^1(1)-\sum_{i=1}^{b-1}(-1)^{i-1}B^1\big(2(1-p_2)t_i\big)\right)\\ &\overset{\mathcal{L}}{=} \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=2,\\ 0\le t_1-t_2+\dots+t_{b-1}\le 1}} \left(-\frac{1}{2}B^1(1)+\sum_{i=1}^{b-1}(-1)^{i-1}B^1\big((1-p_2)t_i\big)\right). \qquad (4.15)\end{aligned}$$

On the other hand, for (4.14), the max-min is attained when $t_1 - t_2 + \dots + t_{b-1} \in [p_2, 1]$, and is of the order $\sqrt{n}/2$.

Therefore,

$$\begin{aligned}\frac{LCbB_n^1-n/2}{\sqrt{n}} &= \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=1,\\ p_2\le t_1-t_2+\dots+t_{b-1}\le 1}} \left(\sqrt{\frac{1}{2}}B_n^1\left(\frac{1}{2}\right)-\sqrt{2}\sum_{i=1}^{b-1}(-1)^{i-1}B_n^1\left(\frac{t_i}{2}\right)\right)+o_{\mathbb{P}}(1)\\ &\Longrightarrow \max_{\substack{0=t_0\le t_1\le\dots\\ \le t_{b-1}\le t_b=1,\\ p_2\le t_1-t_2+\dots+t_{b-1}\le 1}} \left(\frac{1}{2}B^1(1)-\sum_{i=1}^{b-1}(-1)^{i-1}B^1\big(t_i\big)\right). \qquad (4.16)\end{aligned}$$

Next, via the transformation $t_i' = 1 - t_i$ and $t_i'' = t_i'/(2(1-p_2))$, and recalling the time reversibility of Brownian motion, i.e. $\{B_s : 0 \le s \le t\}$ is equivalent in distribution to

$\{B_{t-s} - B_t : 0 \le s \le t\}$, for any given $t > 0$. Therefore (4.16) becomes

$$\begin{aligned}
\frac{LCbB_n^1 - n/2}{\sqrt{n}} \Longrightarrow & \max_{\substack{0=t_b'\le t_{b-1}'\le\dots \\ \le t_1'\le t_0'=1, \\ 0\le t_1'-t_2'+\dots+t_{b-1}'\le 1-p_2}} \left(\frac{1}{2}B^1(1) - \sum_{i=1}^{b-1}(-1)^{i-1}B^1\big(1-t_i'\big) \right) \\
\overset{\mathcal{L}}{=\!=} & \max_{\substack{0=t_b''\le t_{b-1}''\le\dots \\ \le t_1''\le t_0''=1, \\ 0\le t_1''-t_2''+\dots+t_{b-1}''\le \frac{1}{2}}} \left(\frac{1}{2}B^1(1) - \sum_{i=1}^{b-1}(-1)^{i-1}B^1\big(1-2(1-p_2)t_i''\big) \right) \\
\overset{\mathcal{L}}{=\!=} & \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=1, \\ \frac{1}{2}\le t_1-t_2+\dots+t_{b-1}\le 1}} \left(\frac{1}{2}B^1(1) - \sum_{i=1}^{b-1}(-1)^{i-1}B^1\big(2(1-p_2)t_i\big) \right) \\
\overset{\mathcal{L}}{=\!=} & \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=2, \\ 1\le t_1-t_2+\dots+t_{b-1}\le 2}} \left(-\frac{1}{2}B^1(1) + \sum_{i=1}^{b-1}(-1)^{i-1}B^1\big((1-p_2)t_i\big) \right). \qquad (4.17)
\end{aligned}$$

Combining (4.15) and (4.17), gives

$$\begin{aligned}
\frac{LCbB_n - n/2}{\sqrt{n}} \Longrightarrow & \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=2}} \left(-\frac{1}{2}B(1) + \sum_{i=1}^{b-1}(-1)^{i-1}B\big((1-p_2)t_i\big) \right) \qquad (4.18) \\
\overset{\mathcal{L}}{=\!=} & \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=1}} \left(-\frac{1}{2}B(1) + \sum_{i=1}^{b-1}(-1)^{i-1}B\big(2\min(p_1,1-p_2)t_i\big) \right),
\end{aligned}$$

when $1/2 = p_1 < p_2$. The proof for the case when $p_1 < p_2 = 1/2$ or b odd is similar. $\square$

Let us finish this section by obtaining the density of the functional in (iv) in the binary case. Note that when $b = 2$, (iv) reduces to

$$\frac{LC2B_n - n/2}{\sqrt{n}} \Longrightarrow \max_{0\le t_1\le 1}\Big(-\frac{1}{2}B(1) + B\big(2\min(p_1, 1-p_2)t_1\big)\Big)$$
$$\stackrel{\mathcal{L}}{=} \max_{0\le t_1\le 2\min(p_1,1-p_2)}\Big(-\frac{1}{2}B(1) + B\big(t_1\big)\Big)$$
$$= \max_{0\le t_1\le 2\min(p_1,1-p_2)}\Big(B\big(t_1\big) - \frac{1}{2}B\big(2\min(p_1, 1-p_2)\big)$$
$$+\frac{1}{2}B\big(2\min(p_1, 1-p_2)\big) - \frac{1}{2}B(1)\Big). \qquad (4.19)$$

Let us now invoke the following facts. First, the process $\big(-\frac{1}{2}B(s)+\max_{0\le t_1\le s} B(t_1)\big)_{s\ge 0}$ is equal in law to $\big(\sqrt{(B^1(s))^2 + (B^2(s))^2 + (B^3(s))^2}/2\big)_{s\ge 0}$, which is essentially $(\sqrt{s}\chi(3)/2)_{s\ge 0}$, where $(B^1(s), B^2(s), B^3(s))_{s\ge 0}$ is a standard three dimensional Brownian motion (see [37]). Next, according to the Markov property of Brownian motion, or simply let $\mathcal{F}_t^B = \sigma(B_s, s \le t)$ be the natural filtration, then the independence of increments of B implies that $B_{t+h}-B_t$ is independent of $\mathcal{F}_t^B$, for any $t, h \ge 0$. Now setting $t = 2\min(p_1, 1-p_2)$ and $h = 1-2\min(p_1, 1-p_2)$, the right hand of (4.17) becomes the sum of $\sqrt{2\min(p_1, 1-p_2)}\chi(3)/2$, and $B\big(2\min(p_1, 1-p_2)\big)/2-B(1)/2$, where the latter has distribution $N(0, (1-2\min(p_1, 1-p_2))/4)$. Finally, recall that a chi-distribution r.v. $\chi(k)$, with degrees of freedom $k > 0$, is supported on $[0, +\infty)$, with density $x^{k-1}e^{-x^2/2}/(2^{k/2-1}\Gamma(k/2))$. Therefore the density of the right hand side of (4.19) is given by the following convolution:

$$f(z) = \frac{2\sqrt{2}}{\Gamma(3/2)(\min(p_1, 1-p_2))^{3/2}\sqrt{(1-2(\min(p_1, 1-p_2)))\pi}}$$
$$\int_0^{+\infty} x^2 e^{-\frac{x^2}{\min(p_1,1-p_2)} - \frac{(z-x)^2}{1/2-\min(p_1,1-p_2)}}dx, \textit{ for } z \in \mathbb{R}.$$

4.4 Proof for $p_1 < 1/2 < p_2$

Proof. Finally, as for the case $p_1 < 1/2 < p_2$, note that the max-min in (4.10) (resp. (4.11)) is attained at $k = 1$ (resp. $k = 2$) and is of order $(2p_1p_2 - p_1 - p_2 + 1)\sqrt{n}$, when $\sum_{i=1}^{b-1}(-1)^{i-1}t_i = 1 - p_1$ (resp. $\sum_{i=1}^{b-1}(-1)^{i-1}t_i = p_2$).

Note that,

$$\begin{aligned} & \sqrt{p_1}B(1-p_1) - \frac{p_1}{1-p_1}B\big((1-p_2)(1-p_1)\big) \\ = \; & \frac{\sqrt{p_1}}{1-p_1}\left((1-p_1)B(1-p_1) - B\big((1-p_2)(1-p_1)\big)\right), \end{aligned}$$

and from the independent increments property of Brownian motion,

$$\begin{aligned} & (1-p_1)B(1-p_1) - B\big((1-p_2)(1-p_1)\big) \\ = \; & (1-p_1)\Big(B(1-p_1) - B\big((1-p_1)(1-p_2)\big)\Big) - p_1B\big((1-p_2)(1-p_1)\big) \\ \sim \; & N\left(0, (1-p_1)^3p_2 + p_1^2(1-p_1)(1-p_2)\right), \end{aligned}$$

thus,

$$\sqrt{p_1}B(1-p_1) - \frac{\sqrt{p_1}}{1-p_1}B\big((1-p_2)(1-p_1)\big) \sim N\left(0, \frac{p_1(p_1^2 - 2p_1p_2 + p_2)}{1-p_1}\right).$$

And similarly,

$$\begin{aligned} \sqrt{1-p_2}B(p_2) - \frac{\sqrt{1-p_2}}{p_2}B\big(p_1p_2\big) &= \frac{\sqrt{1-p_2}}{p_2}\Big(p_2B\big(p_2\big) - B\big(p_1p_2\big)\Big) \\ &= \frac{\sqrt{1-p_2}}{p_2}\Big(p_2\big(B\big(p_2\big) - B\big(p_1p_2\big)\big) - (1-p_2)B\big(p_1p_2\big)\Big) \\ &\sim N\left(0, \frac{(1-p_2)(p_2^2 - 2p_1p_2 + p_1)}{p_2}\right). \end{aligned}$$

Therefore,

$$\frac{LCbB_n^0 - n(2p_1p_2 - p_1 - p_2 + 1)}{\sqrt{n}} \Longrightarrow N\left(0, \frac{p_1(p_1^2 - 2p_1p_2 + p_2)}{1 - p_1}\right),$$
$$\frac{LCbB_n^1 - n(2p_1p_2 - p_1 - p_2 + 1)}{\sqrt{n}} \Longrightarrow N\left(0, \frac{(1 - p_2)(p_2^2 - 2p_1p_2 + p_1)}{p_2}\right).$$

Finally, by the previous analysis, $LCbB_n$ will be represented by $LCbB_n^0$ when $p_1 + p_2 > 1$, and by $LCbB_n^1$ when $p_1 + p_2 \leq 1$. Then, taking the maximum completes the proof of Theorem 1.1 (v).

□

CHAPTER 5

EXTENSIONS AND CONNECTIONS

5.1 Extensions of the $LCbB_n$ Problem

Let us start this chapter by presenting several further connections and extensions of the block problem.

- First, the limiting functionals obtained above needs to be better understood. For a single sequence, say, starting with zeros, and for $b \geq 2$, set

$$V(b) := \max_{\substack{0 = t_0 \leq t_1 \leq \ldots \\ \leq t_{b-1} \leq t_b = 1}} \sum_{j=1}^{b} (-1)^{j+1} \Big(B^i(t_j) - B^i(t_{j-1}) \Big).$$

  In particular, for $b = 2$ under uniform draw,

$$\frac{L2B_n^0 - n/2}{\sqrt{n}/2} \Longrightarrow -B(1) + 2 \max_{0=t_0 \leq t_1 \leq t_2 = 1} B(t_1) := V(2).$$

  As before mentioned, a well-known result of Pitman [37] asserts that $V(2)$ is identical in law to the radial part of a three-dimensional standard Brownian motion at time $t = 1$, i.e., a Bessel process of order 3 at $t = 1$, which is also a $\chi(3)$ random variable. In Chapter 4 we also derived the pdf of $-B(1) + 2 \max_{0 \leq t_1 \leq 2 \min(p_1, 1-p_2)} B\big(t_1\big)$ when $p_1 < p_2 = 1/2$, or $1/2 = p_1 < p_2$. However, for $b \geq 3$, the distribution of limiting functional $V(b)$ is yet to be studied.

- Another open problem related to above is to understand the limiting behavior and the growth rate of $V(b)$, as b or the mesh goes to infinity. Clearly, $V(b)$, considered as a functional of b, is non-negative and is also monotonically increasing in b.

Therefore, $\lim_{b\to\infty} V(b)$ is highly expected to be related (approximately identical) to the total variation of the standard Brownian motion in $[0,1]$. Let $TV_{[0,1]}(B) := \lim_{||\Pi_b||\to 0} \sum_{j=1}^{b} |B(t_j) - B(t_{j-1})|$, where $\Pi = \{t_0, t_1, ..., t_b\}$ is a partition of $[0,1]$, and where $||\Pi_b|| = \max_{1\le j\le b} |t_j - t_{j-1}|$ denotes the mesh of the partition. In fact,

Claim 5.1. *Let* $||\Pi_b|| = 1/b$*, then as* $b \to \infty$*,* $TV_{[0,1]}/\sqrt{b} \to \sqrt{2/\pi},\ a.s.$

Proof. Indeed, simply note that $X_i = B(t_i) - B(t_{i-1})$ are independent Gaussian random variables with mean 0 and variance $t_i - t_{i-1}$, $i = 1, ..., b$. Hence the $|X_i|$'s are independent and half-normal distributed with mean $\sqrt{t_i - t_{i-1}}\sqrt{\frac{2}{\pi}}$, and variance $(t_i - t_{i-1})\left(1 - \frac{2}{\pi}\right)$. So under a uniform partition assumption, the $|X_i|$'s are i.i.d. and

$$\sum_{i=1}^{b} |X_i|/\sqrt{b} \to \sqrt{2/\pi},\ a.s.$$

by the law of large numbers. ($Var(\sum_{i=1}^{b} |X_i|) = 1 - 2/\pi$, regardless of b or of the partition.) □

In general, the growth rate of $TV_{[0,1]}(B)$ with respect to b or to the mesh is yet to be studied, and nor does $V(b)$. The work in [34] might also be inspiring here, where the concept of 'truncated' variation were introduced. In which the truncated variation of Brownian motion, B, in the interval $[0,1]$, denoted by $TV^c[0,1]$, as

$$TV^c[0,1] = \sup_b \max_{\substack{0=t_0\le t_1\le ... \\ \le t_{b-1}\le t_b=1}} \sum_{j=1}^{b} \max(|B(t_j) - B(t_{j-1})| - c, 0),$$

for some $c > 0$. Roughly speaking, in TV^c we ignore the little changes of B which are below some threshold c, and it is shown that unlike $TV[0,1]$, $TV^c[0,1]$ is finite almost surely.

Also, of interest is the study of the asymptotic behavior of $LCbB_n$ when the word length n and block size b simultaneously grow to infinity, e.g., $b = n^\alpha$, $\alpha > 0$, small.

- A most natural question is to go beyond the binary case, to finite or even infinite countable alphabets. For finite alphabets with size m, it is expected that LbB_n (resp. $LCbB_n$) should be the maximum over $m(m-1)^{b-1}$ maximal (resp. max-min) functionals. Moreover, based on the sandwich argument developed in [22], it can also be shown that the asymptotic behavior of $LCbB_n$ for countably infinite alphabets should perform the same as for finite alphabets, Also, the comparison over more than two sequences in that case deserves to be studied as well.

- As clearly indicated by its name, $LCbB_n$ is also pertinent to LC_n, the longest common subsequence problem. Let $(X_i)_{1\leq i\leq n}$ and $(Y_i)_{1\leq i\leq n}$ be two independent sequences of i.i.d. random variables with alphabet $\mathcal{A} = \{\alpha_1 < \alpha_2 < ... < \alpha_m\}$. One central problem in LC_n is to study the limit

$$\gamma_m^* = \lim_{n\to\infty} \frac{\mathbb{E}LC_n}{n}.$$

  γ_m^* has been shown to exist (see [12]), but the exact value still remains unknown so far. Several theoretical and numerical works on bounds of γ_2^* are presented, giving a range of lower bound 0.788 and upper bound 0.826. Note that Theorem 1.1 (i) shows with the number of blocks fixed, LC_n/n tends to $1/2$ almost surely, and thus γ_2^* shrinks to $1/2$ by the dominated convergence. However, since LC_n is essentially the maximum of all possible $LCbB_n$ for $b = 1, ..., n$, it is of interest to study if the present work could be helpful there. As for variance of LC_n, it was shown that $Var(LC_n) \leq n\left(1 - \sum_{i=1}^m p_i^2\right)$, where $p_i = \mathbb{P}(X_1 = \alpha_i)$, see [41]. However, the linear order result for lower bound has only been done in some biased situations (see [23], [27]...), therefore the present work also serves as an instance of linear order variance, when the number of blocks is imposed.

- It is straightforward to extend, beyond the independent case, the solution of the problem studied in the present paper, to Markovian words. In the following, let us take

$LCbB_n^0$ for the sake of illustration.

Let $X = (X_n)_{n\geq 0}$ and $Y = (Y_n)_{n\geq 0}$ be two binary time-homogeneous Markov chains with, again, for simplicity, the same initial distribution and the same transition matrix. Let

$$\mathbb{P}(X_{n+1} = 1|X_n = 0) = \mathbb{P}(Y_{n+1} = 1|Y_n = 0) = p_{01},$$

and

$$\mathbb{P}(X_{n+1} = 0|X_n = 1) = \mathbb{P}(Y_{n+1} = 0|Y_n = 1) = p_{10}.$$

For $0 < p_{01} + p_{10} \leq 2$, let the law of X_0 and Y_0 be the invariant distribution

$$(\pi_1, \pi_2) = (p_{10}/(p_{01} + p_{10}), p_{01}/(p_{01} + p_{10}))$$

while, for $p_{01} = p_{10} = 0$, let $(\pi_1, \pi_2) = (1, 0)$.

Then, combining techniques as above with those developed for LI_n, in [21], we make the following statement.

Corollary 5.1. *With the setting presented above, for* $0 < p_{01} = p_{10}$*,*

$$\frac{LCbB_n^0 - n/2}{\sqrt{n}/2} \Longrightarrow \max_{\substack{0 = t_0 \leq t_1 \leq \ldots \\ \leq t_{b-1} \leq t_b = 1}} \min_{k=1,2} \sqrt{\frac{1 - p_{01}}{p_{01}}} \left(\sum_{i=1}^{b} (-1)^{i+1} \Big(B^k(t_i) - B^k(t_{i-1}) \Big) \right),$$

where for $p_{01} = p_{10} = 1$*, the limiting distribution above is understood to be degenerated at the origin. For* $p_{01} \neq p_{10}$ *or* $p_{01} = p_{10} = 0$*,*

$$\frac{LCbB_n^0 - n \max(\pi_1, \pi_2)}{\sqrt{n}} \Longrightarrow Z \sim N(0, \sigma^2),$$

where $\sigma^2 = p_{01}p_{10}(2 - p_{01} - p_{10})/(p_{01} + p_{10})^3$*, for* $p_{01} \neq p_{10}$*, and* $\sigma^2 = 0$*, for* $p_{01} = p_{10} = 0$*.*

An outlined proof has been included in the end of this chapter, since some combinatorial set-up required are presented in section 5.2. Indeed, the methodology we utilized for full two-sequence comparison in the main body, can also be applied here to Markovian context, but could be rather cumbersome. Fortunately, the proof then can be much shortened thanks to the alternative approach developed in section 5.2. A further extension could be to the hidden Markov models setting.

- As well known, both LI_n and LCI_n have interpretations in directed last-passage percolation theory (e.g., see [9]): Let $T_2(n,m)$ be the last-passage time from $(0,0)$ to (n,m), let $\omega_{i,j}, i \geq 0, j \geq 1$ be the time spent by a path going through the vertex (i,j), and $\Pi_2(n,m)$ be the set of all directed unit step North-East paths from $(0,0)$ to (n,m) on the non-negative lattice $\mathbb{Z}_+^2$, that is,

$$\Pi_2(n,m) := \Big\{(u_1,...,u_{n+m}) \in (\mathbb{Z}_+^2)^{n+m} : u_1 = (0,1), u_{n+m} = (n,m) \\ u_{i+1} - u_i \in \{(0,1),(1,0)\}, i = 1,...,n+m-1\Big\}.$$

For $\mathbb{Z}_+^3$, similarly define $T_3(n,n,m)$ to be the last-passage time from $(0,0,0)$ to (n,n,m), and $\Pi_3(n,n,m)$ to be the set of all paths in $\mathbb{Z}_+^3$ from $(0,0,0)$ to (n,n,m) taking either upwards unit steps or horizontal steps of any length but neither parallel to the x-axis nor to the y-axis, i.e.

$$\Pi_3(n,n,m) := \Big\{(u_1,...,u_{n+m}) \in (\mathbb{Z}_+^3)^{n+m} : u_1 = (0,0,1), u_{n+m} = (n,n,m) \\ u_{i+1} - u_i \in \{(0,0,1),(a,b,0)|ab \neq 0\}, i = 1,...,n+m-1\Big\}.$$

In the context of random words, we interpret $\omega_{i,j}$ as $\mathbb{1}_{\{X_i=\alpha_j\}}$, and $\omega_{0,j} = 0,\ j \geq 1$, where $X = (X_i)_{1\leq i\leq n}$ is a sequence of i.i.d. random variables taking values in an equiprobable alphabet $\{\alpha_1 < ... < \alpha_m\}$; and $\omega_{i,j,k} = \mathbb{1}_{\{X_i=Y_j=\alpha_k\}}$, while $\omega_{0,0,k} = 0,\ k \geq 1$. Clearly $\omega_{i,j}$ and $\omega_{i,j,k}$ are i.d. but dependent. Indeed, $(\omega_{i,j})_i$ are inde-

pendent for a fixed letter α_j, but for each i the row sums to one (similarly for $\omega_{i,j,k}$). Then, [9] shows that

$$LI_n = T_2(n,m) := \max_{\pi \in \Pi_2(n,m)} \left(\sum_{(i,j)\in\pi} \omega_{i,j} \right),$$

$$LCI_n = T_3(n,n,m) := \max_{\pi \in \Pi_3(n,n,m)} \left(\sum_{(i,j,k)\in\pi} \omega_{i,j,k} \right).$$

In the block context, LbB_n and $LCbB_n$ also enjoy similar percolation interpretations. Let us first discuss LbB_n. Instead of the final destination point (n,m) as for LI_n, we now allow the process to stop whenever it hits the vertical boundary $x = n$, also the constraint on the size of vertical steps being necessarily 1 is replaced by requiring the number of vertical steps, upwards or downwards, to be exactly b, i.e., let

$$\begin{aligned}\Pi_2(n,m,b) := \Big\{ & (u_1,u_2,...,u_{n+b}) \in (\mathbb{Z}_+^2)^{n+b} : u_1 = (1,0), \\ & \{u_{i+1} - u_i, i = 1,...,n+b-1\} \textit{ vertically containing} \\ & (0,d_1),(0,d_2),...(0,d_b), \text{ where } 1 \le |d_j| \le m-1, \\ & 0 \le \sum_{l=1}^{j} d_l \le m-1, \textit{ for } 1 \le j \le b; \\ & \textit{and horizontally containing } n-1 \textit{ times } (1,0) \Big\}.\end{aligned}$$

Then

$$LbB_n = T_2(n,m,b) := \max_{\pi \in \Pi_2(n,m,b)} \left(\sum_{(i,j)\in\pi} \omega_{i,j} \right),$$

where $T_2(n,m,b)$ is the last-passage time from $(0,0)$ to hitting the vertical boundary $x = n$ over all horizontal or vertical paths with b vertical segments (either upwards

or downwards). Finally for $LCbB_n$, let

$$\Pi_3(n,n,m,b) := \Big\{(u_1,u_2,...,u_{n+b}) \in (\mathbb{Z}_+^3)^{n+b} : u_1 = (0,0,0),$$
$$\{u_{i+1}-u_i, i=1,...,n+b-1\}$$
$$\textit{vertically containing } (0,0,d_1),(0,0,d_2),...,$$
$$(0,0,d_b), \text{ where } 1 \le |d_j| \le m-1,$$
$$0 \le \sum_{l=1}^{j} d_l \le m-1, \textit{ for } 1 \le j \le b; \textit{ and horizontally}$$
$$\textit{the remaining } n-1 \textit{ terms not containing } (a,b,0) \textit{ for } ab=0\Big\}.$$

Then,

$$LCbB_n = T_3(n,n,m,b) := \max_{\pi \in \Pi_3(n,n,m,b)} \left(\sum_{(i,j,k)\in\pi} \omega_{i,j,k} \right),$$

where $T_3(n,n,m,b)$ is the last-passage time from $(0,0,0)$ to hitting the vertical boundary $x = n$ or the horizontal boundary $y = n$ over all paths with b vertical upwards or downwards segments.

Note that when two sequences are identical, $\Pi_3(n,n,m,b)$ is indeed equivalent to $\Pi_2(n,m,b)$ by equating the first and second coordinates. Also, note that for LCI_n, we have proceeded without assuming each step of which increases one of the coordinates by 1; and for LbB_n and $LCbB_n$, we have further loosened the directedness on the last coordinates. Therefore it would be of interest to investigate the limiting behavior of $\Pi_2(n,m,b)$ and $\Pi_3(n,n,m,b)$, with fixed m and b, and with the weights to be replaced by the ones i.i.d exponential, geometric or beyond.

5.2 An Alternative Approach To the One-Sequence Problem

We finish this chapter by presenting an alternative approach dealing with the one sequence case, LbB_n. Let us first state the result of LbB_n, which was derived from Theorem 1.1.

Corollary 5.1. *Let $(X_i)_{i\geq 1}$ be a sequences of i.i.d. Bernoulli random variables with parameter p, and without loss of generality $0 < p < 1$. Then,*

(i) For $p = 1/2$,

$$\frac{LbB_n - n/2}{\sqrt{n}/2} \Longrightarrow \max_{\substack{0=t_0\leq t_1\leq ... \\ \leq t_{b-1}\leq t_b=1}} \left(\sum_{i=1}^{b}(-1)^{i+1}\Big(B^1(t_i) - B^1(t_{i-1})\Big)\right) \bigvee \max_{\substack{0=t_0\leq t_1\leq ... \\ \leq t_{b-1}\leq t_b=1}} \left(\sum_{i=1}^{b}(-1)^{i+1}\Big(B^2(t_i) - B^2(t_{i-1})\Big)\right), \tag{5.1}$$

where B^1, B^2 are standard Brownian motions defined on $[0, 1]$.

(ii) For $p \neq 1/2$,

$$\frac{LbB_n - n\max(p, 1-p)}{\sqrt{p(1-p)n}} \Longrightarrow Z, \tag{5.2}$$

where Z is a standard normal.

Let us first consider the uniform case, and as usual, first assume b to be even. Now let us define b_i, a_i as the number of zeros and ones in $X_1, X_2, ..., X_i$, respectively. By convention set $b_0 = a_0 = 0$, and LbB_n^0 and LbB_n^1 are analogously defined.

Therefore

$$\begin{aligned} LbB_n^0 &= \max_{\substack{0=k_0\leq k_1\leq ... \\ \leq k_{b-1}\leq k_b=n}} \big((b_{k_1} - b_0) + (a_{k_2} - a_{k_1}) + ... + (b_{k_{b-1}} - b_{k_{b-2}}) + (a_n - a_{k_{b-1}})\big) \\ &= \max_{\substack{0=k_0\leq k_1\leq ... \\ \leq k_{b-1}\leq k_b=n}} \big((b_{k_1} - a_{k_1}) + (a_{k_2} - b_{k_2}) + ... + (b_{k_{b-1}} - a_{k_{b-1}}) + a_b\big). \end{aligned} \tag{5.3}$$

Now we set

$$Z_i = \begin{cases} 1 & \text{if } X_i = 0,\ w.p.\ 1/2 \\ -1 & \text{if } X_i = 1,\ w.p.\ 1/2 \end{cases} \quad \text{for } i = 1, 2, ..., n.$$

Then define $S_{k_i} = \sum_{j=1}^{k_i} Z_j$, for $i = 1, 2, ..., n$, and clearly $S_{k_i} = b_{k_i} - a_{k_i}$.

Therefore

$$LbB_n^0 = \max_{\substack{0=k_0\le k_1\le ...\\ \le k_{b-1}\le k_b=n}} \left(S_{k_1} - S_{k_2} + S_{k_3} - S_{k_4} + ... + S_{k_{b-1}} + a_n\right)$$
$$= \frac{n}{2} - \frac{1}{2}S_n + \max_{\substack{0=k_0\le k_1\le ...\\ \le k_{b-1}\le k_b=n}} \left(S_{k_1} - S_{k_2} + S_{k_3} - S_{k_4} + ... + S_{k_{b-1}}\right). \tag{5.4}$$

Note that $\mathbb{E}(Z_i) = 0, \quad \mathbb{E}(Z_i^2) = 1, \ Var(Z_i) = 1,$ and $Var(S_n) = n$. Here we define the polygonal function $\widehat{B}_n(t) = \frac{1}{\sqrt{n}}S_{[nt]} + \frac{1}{\sqrt{n}}(nt - [nt])Z_{[nt]+1}$, for $0 \le t \le 1$. Then substitute k_i by $t_i = k_i/n$, it follows by the Donsker's Theorem and the Continuous Mapping Theorem,

$$\frac{LbB_n^0 - n/2}{\sqrt{n}} = -\frac{1}{2}\widehat{B}_n(1) + \max_{\substack{0=t_0\le t_1\le ...\\ \le t_{b-1}\le t_b=1}} \left(\widehat{B}_n(t_1) - \widehat{B}_n(t_2) + ... + \widehat{B}_n(t_{b-1})\right) \tag{5.5}$$
$$\Longrightarrow -\frac{1}{2}B(1) + \max_{\substack{0=t_0\le t_1\le ...\\ \le t_{b-1}\le t_b=1}} \Big(B(t_1) - B(t_2) + ... + B(t_{b-1})\Big),$$

where B is standard Brownian motion.

Now for odd b, similar with (5.4) we have

$$LbB_n^0 = \frac{n}{2} + \frac{1}{2}S_n + \max_{\substack{0=k_0\le k_1\le ...\\ \le k_{b-1}\le k_b=n}} \left(S_{k_1} - S_{k_2} + S_{k_3} - S_{k_4} + ... - S_{k_{b-1}}\right)$$
$$= \frac{n}{2} + \frac{1}{2}\widehat{B}_n(1)\sqrt{n} + \max_{\substack{0=t_0\le t_1\le ...\\ \le t_{b-1}\le t_b=1}} \left(\widehat{B}_n(t_1) - \widehat{B}_n(t_2) + ... - \widehat{B}_n(t_{b-1})\right)\sqrt{n}.$$

It follows from the Invariance Principle and the Continuous Mapping Theorem,

$$\frac{LbB_n^0 - n/2}{\sqrt{n}} \Longrightarrow \frac{1}{2}B(1) + \max_{\substack{0=t_0\le t_1\le ...\\ \le t_{b-1}\le t_b=1}} \Big(B(t_1) - B(t_2) + ... - B(t_{b-1})\Big).$$

Combining both cases and the same rationale also applies for LbB_n^1. We conclude Corollary A.1. (i) after rearranging and taking the maximum.

For non-uniform draw, assume b even, we define accordingly

$$Z_i = \begin{cases} 1 & \text{if } X_i = 0,\ w.p.\ 1-p \\ -1 & \text{if } X_i = 1,\ w.p.\ p \end{cases}$$

Then it follows $\mu = \mathbb{E}(Z_i) = 1-2p,\ \ \sigma^2 = Var(Z_i) = 4p(1-p)$.

Let $\widehat{B}_n(t) = \frac{S_{[nt]}-\mu[nt]}{\sigma\sqrt{n}} + (nt - [nt])\frac{Z_{[nt]+1}-\mu}{\sigma\sqrt{n}}$, for $0 \le t \le 1$, and (5.3) rewrites

$$\begin{aligned} LbB_n^0 \ &= \ pn - \frac{1}{2}\sigma\sqrt{n}\widehat{B}_n(1) \\ &+ \max_{\substack{0=t_0\le t_1\le \dots \\ \le t_{b-1}\le t_b = 1}} \left(\sigma\sqrt{n}\left(\sum_{i=1}^{b-1}(-1)^{i-1}\widehat{B}_n(t_i)\right) + \mu n\left(t_1 - t_2 + \dots + t_{b-1}\right)\right), \end{aligned}$$

on the other hand, for LbB_n^1,

$$\begin{aligned} LbB_n^1 \ &= \ (1-p)n + \frac{1}{2}\sigma\sqrt{n}\widehat{B}_n(1) \\ &+ \max_{\substack{0=t_0\le t_1\le \dots \\ \le t_{b-1}\le t_b = 1}} \left(\sigma\sqrt{n}\left(\sum_{i=1}^{b}(-1)^{i-1}\widehat{B}_n(t_i)\right) + \mu n\left(t_1 - t_2 + \dots + t_{b-1} - t_b\right)\right), \end{aligned}$$

which are quite similar to (4.8) and (4.9) on two-sequence case which can follow the proof back there to show limiting behavior.

Remark 5.1. Corollary 5.1 recovers the result from [22] for binary words and fixing sequence starting with a zero. Also, 5.1 still holds in the case when the vacuous blocks (of length zero) are not allowed, where the constraints in (5.3) are to be replaced by " $<$ ", due to almost surely continuity.

Remark 5.2. It is noteworthy that in this approach we do have constraints k_js in a deterministic sets, which is more straightforward than the random one used in $LCbB_n$, however,

the combinatorial method mentioned above does not hold for two-sequence case. Let us take the example in Chapter 1 again. For $X = 0010$, $Y = 0101$, and $LC3B_4 = 3$ with subsequence 010. However, this optima cannot be obtained from simultaneously picking k_1, k_2 from $1, 2, 3$ for both sequences. In other words, $LCbB_n^0$ is not exactly (less than or equal to) the maximum over constraints k_js of the minimum of two independent copies of (5.5), i.e.,

$$LCbB_n^0 \neq \max_{\substack{0=k_0\leq k_1\leq \dots \\ \leq k_{b-1}\leq k_b=n}} \min_{\kappa=1,2} \left(\frac{n}{2} - \frac{1}{2}S_n^\kappa + S_{k_1}^\kappa - S_{k_2}^\kappa + S_{k_3}^\kappa - S_{k_4}^\kappa + \dots + S_{k_{b-1}}^\kappa \right).$$

Nevertheless, in the main part we have shown that as n tends to infinity, the right hand side of above, after centered at $n/2$ and normalized by $\sqrt{n}$, is different from

$$\max_{\substack{0=t_0\leq t_1\leq \dots \\ \leq t_{b-1}\leq t_b=1}} \min_{k=1,2} \left(-\frac{1}{2}\widehat{B}_n^k(1) + \widehat{B}_n^k(t_1) - \widehat{B}_n^k(t_2) + \dots + \widehat{B}_n^k(t_{b-1}) \right)$$

by a term goes to zero in probability. Hence the extreme cases can be dismissed and do not affect the limiting behavior of $LCbB_n^0$.

Remark 5.3. Finally, let us finish this chapter by outlining the proof for Corollary 5.1. Recall that in [21], the eigenvalues of the transition matrix

$$\begin{pmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{pmatrix}$$

are denoted by $\lambda_1 = 1$ and $-1 < \lambda_2 = 1 - p_{10} - p_{10} < 1$, associated with left eigenvectors $(\pi_1, \pi_2) = (p_{10}/(p_{01} + p_{10}), p_{01}/(p_{01} + p_{10}))$, which is the stationary distribution, and $(1, -1)$. Now let us assume the sequence initiates with the stationary distribution for simplicity. With the random variables Z_k and S_k defined in (5.4), it is shown that $\mu := \mathbb{E}(Z_k) = (p_{10} - p_{01})/(p_{01} + p_{10}) = \pi_1 - \pi_2$, and $\mathbb{E}(S_k) = k\mu$. Moreover,

$\sigma^2 := Var(Z_k) = 1 - (\frac{p_{10}-p_{01}}{p_{01}+p_{10}})^2 = \frac{4p_{01}p_{10}}{(p_{01}+p_{10})^2}$, and $S_k/k \to \sigma^2(1+\lambda_2)/(1-\lambda_2)$. Let $\tilde{\sigma} = \sigma\sqrt{(1+\lambda_2)/(1-\lambda_2)}$, and define the polygonal function

$$\widehat{B}_n(t) = \frac{S_{[nt]} - \mu[nt]}{\sigma\sqrt{n(1+\lambda_2)/(1-\lambda_2)}} + (nt - [nt])\frac{Z_{[nt]+1} - \mu}{\sigma\sqrt{n(1+\lambda_2)/(1-\lambda_2)}}.$$

According to (5.4) and assuming b even, it follows

$$\begin{aligned}
LCbB_n^0 &= \pi_2 n \\
&\quad + \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=1}} \min_{k=1,2}\left(-\frac{1}{2}\tilde{\sigma}\sqrt{n}\widehat{B}_n^k(1) + \tilde{\sigma}\sqrt{n}\sum_{i=1}^{b-1}(-1)^{i-1}\widehat{B}_n^k(t_i) + (\pi_1-\pi_2)n\sum_{i=1}^{b-1}(-1)^{i-1}t_i\right) \\
&\quad + o_{\mathbb{P}}(1) \\
&= \pi_{max} n - (\pi_{max} - \pi_2)n \\
&\quad + \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=1}} \min_{k=1,2}\left(-\frac{1}{2}\tilde{\sigma}\sqrt{n}\widehat{B}_n^k(1) + \tilde{\sigma}\sqrt{n}\sum_{i=1}^{b-1}(-1)^{i-1}\widehat{B}_n^k(t_i) + (\pi_1-\pi_2)n\sum_{i=1}^{b-1}(-1)^{i-1}t_i\right) \\
&\quad + o_{\mathbb{P}}(1). \qquad (5.6)
\end{aligned}$$

First, when $\pi_{max} = \pi_1 = \pi_2$, i.e., $0 < p_{01} = p_{10} < 1$, then $\tilde{\sigma}^2 = (1-p_{01})/p_{01}$, and (5.6) becomes

$$\begin{aligned}
\frac{LCbB_n^0 - n\pi_{max}}{\sqrt{n(1-p_{01})/p_{01}}} &= \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=1}} \min_{k=1,2}\left(-\frac{1}{2}\widehat{B}_n^k(1) + \sum_{i=1}^{b-1}(-1)^{i-1}\widehat{B}_n^k(t_i)\right) + o_{\mathbb{P}}(1) \\
&\Longrightarrow \max_{\substack{0=t_0\le t_1\le\dots \\ \le t_{b-1}\le t_b=1}} \min_{k=1,2}\left(-\frac{1}{2}B^k(1) + \sum_{i=1}^{b-1}(-1)^{i-1}B^k(t_i)\right).
\end{aligned}$$

Secondly, when $\pi_{max} = \pi_2 > \pi_1$, then

$$\frac{LCbB_n^0 - n\pi_{max}}{\tilde{\sigma}\sqrt{n}} = \max_{\substack{0 = t_0 \le t_1 \le \dots \\ \le t_{b-1} \le t_b = 1}} \min_{k=1,2} \Bigg(-\frac{1}{2}\widehat{B}_n^k(1) + \sum_{i=1}^{b-1} (-1)^{i-1} \widehat{B}_n^k(t_i) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_1) \sum_{i=1}^{b-1} (-1)^{i-1} t_i \Bigg) + o_{\mathbb{P}}(1);$$

or when $\pi_{max} = \pi_1 > \pi_2$, then

$$\frac{LCbB_n^0 - n\pi_{max}}{\tilde{\sigma}\sqrt{n}} = \max_{\substack{0 = t_0 \le t_1 \le \dots \\ \le t_{b-1} \le t_b = 1}} \min_{k=1,2} \Bigg(\frac{1}{2}\widehat{B}_n^k(1) + \sum_{i=1}^{b} (-1)^{i-1} \widehat{B}_n^k(t_i) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2) \sum_{i=1}^{b} (-1)^{i-1} t_i \Bigg) + o_{\mathbb{P}}(1).$$

For the last two situations it is trivial to show that $(LCbB_n^0 - n\pi_{max})/(\tilde{\sigma}\sqrt{n})$ converges in law to $-\frac{1}{2}B(1)$, based on the same method shown in Chapter 5.

Finally, for the degenerate cases, if p_{01} or p_{10} is equal to 0, which means $LCbB_n^0$ being 0 or n. On the other hand, if $p_{01} = p_{10} = 1$, i.e. the sequence then keeps going back and forth between 0 and 1, indicating $LCbB_n \sim n/2$. The proof of Corollary 5.1 is hence finished.

CHAPTER 6

CONCLUSION

In this dissertation we have studied the asymptotic behavior of $LCbB_n$, the length of the longest common subsequences containing a fixed number, b, of blocks, for two sequences of i.i.d. Bernoulli random variables with possibly two different parameters. This is a generalization of LCI_n with many applications in the areas such as biology and linguistics. For example, two DNA-strings are assumed to share a common ancestor if they have a long common sub-strings. Each nitrogenous base in the chromosome, when considered as a block, can be arbitrarily permuted or repeated in the common sub-strings. However, the total number of blocks of nitrogenous in the chromosome may be fixed, in order to form a meaningful DNA piece. So understanding the asymptotic behavior of $LCbB_n$ is beneficial in multiple scenarios, e.g. for two independent DNA-strings who lengths are both growing large, we would like to estimate how likely they have to have a long common sub-string just by "chance".

The proof of main theorem 1.1 starts by development on combinatorics and probability, and at the end of Chapter 2, we gave an explicit combinatorial expression of $LCbB_n$, as the maximum of two max-min functionals over random constraints. Moreover, we derived the proper centerings for different scenarios of p_1 and p_2. Then following this path, in Chapter 3 we first proved that for the uniform case, $LCbB_n$, centered by $n/2$ and scaled by $\sqrt{n}/2$, tends to the maximum of two max-min 1-dimensional Brownian motion functionals. The limiting functional contains b alternating terms and is highly related to the total variation of Brownian motion. On the other hand, we showed that for uniform case, the random variables $LCbB_n^0$ and $LCbB_n^1$ share the same distribution, and that for $LCbB_n$, to achieve the maximum, needs to fully use both counting orders. Going beyond the uniform case,

we finished the proof of the main theorem for other non-uniform scenarios case by case in Chapter 4. Comparing p_1 and $1-p_2$ appears in all non-uniform cases. Intuitively, this is not surprising in view of the following facts. First, the number of blocks becomes inconsequential when a dominating term appears in either sequence. Indeed, the longest common subsequence is asymptotically a string mainly consisting of the most frequently occurring letter. In addition, the $LCbB_n$ of $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ is the same as that of $(1-X_i)_{i\geq 1}$ and $(1-Y_i)_{i\geq 1}$, and since it is assumed that $p_1 \leq p_2$ throughout the thesis, the possible smaller one of two dominating terms in $LCbB_n^0$ and $LCbB_n^1$ is either p_1 or $1-p_2$. From then on, in Chapter 5, we presented several connections and extensions of the $LCbB_n$ problem. First, we gave a succinct and alternative approach to LbB_n, but unfortunately the method cannot be generalized to the two-sequence problem. Then, a connection to the Markov context has been made and the Markov two-sequence $LCbB_n$ could be achieved accordingly in view of Theorem 1.1. We finished this chapter by showing that $LCbB_n$ also has interpretation as last-passage percolation, with proper weights on the lattices. Finally, this dissertation concludes with an appendix, containing Monte-Carlo simulations for the asymptotics of $LCbB_n$ and the growth order of the limiting functional.

Appendices

APPENDIX A

MONTE-CARLO SIMULATION FOR LC2BN

A.1 $LC2B_n$ simulation: Brute-force and linear approach

Below we first present a brute-force method to simulate $LC2B_n$. It loops through both sequences and thus has time complexity $O(n^2)$. It is also clear that for general $LCbB_n$, the time complexity of this brute-force approach becomes exponential with b, and is of the order $O(n^{2b-2})$, making it rather time-consuming. However, in the binary case much computation could be saved. Indeed, note that it was never optimal to break the consecutive blocks, we may filter both sequences and only keep the indices where they are changing from a zero to a one, or vice versa.

A more efficient method in linear time, also built for $LC2B_n$, is included. From an algorithmic perspective, Dynamic Programming (DP) has been widely used in computing LI_n, LC_n and LCI_n with time complexity $O(n^2)$. However, in those cases the resulting subsequences are required to be strictly increasing. If this "strict" condition is dropped, which is also known as the longest common and weakly increasing subsequences problem ($LCWI_n$), it has been shown that it is hard to solve in $O(n^2)$ time in theory (see [38]), and concrete efficient algorithm for small-sized alphabet remains sparse as well. One remarkable result comes from Duraj [15], who proposed $LCWI_n$ can be solved in linear time for a special case of a 3-letter alphabet. However, no subquadratic time algorithm has been found for the general case, and it will be of great interest to see more work on $LCWI_n$ since it will be highly related to $LCbB_n$. Now, for the binary case, we will be able to simulate $LC2B_n$ efficiently enough in linear time.

A.2 Code Snippet and Figures

```
import numpy as np
from scipy.stats import norm, bernoulli
from math import sqrt

from datetime import datetime as dt

import matplotlib.pyplot as plt
%matplotlib inline

class simulation:

    # Function simLC2Bn() simulates the result of LC2Bn with the
    following parameters in different mode.
    # 'p1,p2', 'n' - Bernoulli r.v. parameter, length of both sequences.
    # 'iteration' - number of runs of experiment.
    # 'mode = 'rough' - linearly and simultaneously loop over both
    sequences, the outcome will be less than comprehensive search, with
    an error of the order Op(\sqrt{n}). However, it is much faster with
    time complexity of O(n).
    # 'mode = 'comp' - comprehensive search, exhaustively find the
    maximum number of length over both sequences in time complexity O(n
    ^2).
    # 'mode'= 'linear' - efficiently find LC2Bn in linear time.
    # 'timing' - if True, print the time cost for the first 10 iteration
     as well as the total estimated time cost. Suggested when the time-
    consuming comprehensive mode is used, or when n>=20000 in linear
    mode.
    # 'strict': if True, vacuous blocks of length 0 are not allowed.
```

```
def simLC2Bn(p1, p2, n=10000, iteration=500, mode='comp', strict=
False, timing=True):

    if timing:
        t0 = dt.now()
        print('Started at: ' + str(t0.strftime('%H:%M:%S')))

    results, LC2Bn_lst, LC2Bn0_lst, LC2Bn1_lst = [], [], [], []
    rLC2Bn0_lst, rLC2Bn1_lst, sLC2Bn0_lst, sLC2Bn1_lst = [], [], [],
 []

    for k in range(iteration):

        X, Y = bernoulli.rvs(p1, size=n), bernoulli.rvs(p2, size=n)

        rLC2Bn0, rLC2Bn1 = 0, 0
        sLC2Bn0, sLC2Bn1 = 0, 0

        if mode=='rough' or mode=='comp':
            for i in range(n+1):
                xlst1, xlst2 = X[:i], X[i:]
                ylst1, ylst2 = Y[:i], Y[i:]
                n00, n10 = min(i - np.sum(xlst1), i - np.sum(ylst1))
, min(np.sum(xlst2), np.sum(ylst2))
                n11, n01 = min(np.sum(xlst1), np.sum(ylst1)), min(n-
i - np.sum(xlst2), n-i - np.sum(ylst2))

                rLC2Bn0 = max(rLC2Bn0, n00+n10)
                rLC2Bn1 = max(rLC2Bn1, n11+n01)

        if mode=='comp':
            X_01_ix = [i for i in range(n-1) if X[i]==0 and X[i
+1]==1]
```

```
            X_10_ix = [i for i in range(n-1) if X[i]==1 and X[i
+1]==0]
            Y_01_ix = [i for i in range(n-1) if Y[i]==0 and Y[i
+1]==1]
            Y_10_ix = [i for i in range(n-1) if Y[i]==1 and Y[i
+1]==0]

            for i in X_01_ix:
                tmpx00, tmpx10 = i+1-np.sum(X[:(i+1)]), np.sum(X[(i
+1):])
                for j in Y_01_ix:
                    tmpy00, tmpy10 = j+1-np.sum(Y[:(j+1)]), np.sum(Y
[(j+1):])
                    n00, n10 = min(tmpx00, tmpy00), min(tmpx10,
tmpy10)
                    sLC2Bn0 = max(sLC2Bn0, n00+n10)

            for i in X_10_ix:
                tmpx11, tmpx01 = np.sum(X[:(i+1)]), n-i-1-np.sum(X[(
i+1):])
                for j in Y_10_ix:
                    tmpy11, tmpy01 = np.sum(Y[:(j+1)]), n-j-1-np.sum
(Y[(j+1):])
                    n11, n01 = min(tmpx11, tmpy11), min(tmpx01,
tmpy01)
                    sLC2Bn1 = max(sLC2Bn1, n11+n01)

        LC2Bn0, LC2Bn1 = max(rLC2Bn0, sLC2Bn0), max(rLC2Bn1, sLC2Bn1
)
        LC2Bn = max(LC2Bn0, LC2Bn1)

        if mode=='linear':
            n1X, n1Y = np.sum(X), np.sum(Y)
```

```
            n0X, n0Y = n-n1X, n-n1Y

            X0_ix = [-1]+[i for (i,value) in enumerate(X) if value
==0]
            X1_ix = [-1]+[i for (i,value) in enumerate(X) if value
==1]
            Y0_ix = [-1]+[i for (i,value) in enumerate(Y) if value
==0]
            Y1_ix = [-1]+[i for (i,value) in enumerate(Y) if value
==1]

            tmp_01_lst, tmp_10_lst = [], []

            for i in range(min(n0X, n0Y)+1):
                X_after_lst, Y_after_lst = X[(X0_ix[i]+1):], Y[(
Y0_ix[i]+1):]
                n1X_after_lst, n1Y_after_lst = np.sum(X_after_lst),
np.sum(Y_after_lst)
                if (strict and i*n1X_after_lst*n1Y_after_lst !=0) or
 (not strict):
                    tmp_01_lst.append(min(i+n1X_after_lst, i+
n1Y_after_lst))

            for i in range(min(n1X, n1Y)+1):
                X_after_lst, Y_after_lst = X[(X1_ix[i]+1):], Y[(
Y1_ix[i]+1):]
                n0X_after_lst, n0Y_after_lst = len(X_after_lst)-np.
sum(X_after_lst), len(Y_after_lst)-np.sum(Y_after_lst)
                if (strict and i*n0X_after_lst*n0Y_after_lst !=0) or
 (not strict):
                    tmp_10_lst.append(min(i+n0X_after_lst, i+
n0Y_after_lst))
```

```
                tmp_01_lst, tmp_10_lst = np.array(tmp_01_lst), np.array(
tmp_10_lst)
                LC2Bn0_tmp, LC2Bn1_tmp = np.max(tmp_01_lst), np.max(
tmp_10_lst)
                rLC2Bn0, rLC2Bn1 = min(n0X, n0Y), min(n1X, n1Y)

                LC2Bn0, LC2Bn1 = max(LC2Bn0_tmp, rLC2Bn0), max(
LC2Bn1_tmp, rLC2Bn1)
                LC2Bn = max(LC2Bn0, LC2Bn1)

            if p1<0.5<p2:
                pp = 2*p1*p2-p1-p2+1
                if p1 > 1-p2:
                    res = (LC2Bn - n * pp)/sqrt(n*(p1*(p1**2-2*p1*p2+p2)
/(1-p1)))
                else:
                    res = (LC2Bn - n * pp)/sqrt(n*(1-p2)*(p2**2-2*p1*p2+
p1)/p2)
            else:
                res = (LC2Bn - n*max(p1, 1-p2))/(sqrt(n * max(p1, 1-p2)
* min(1-p1, p2)))

            results.append(res); LC2Bn_lst.append(LC2Bn); LC2Bn0_lst.
append(LC2Bn0); LC2Bn1_lst.append(LC2Bn1)
            rLC2Bn0_lst.append(rLC2Bn0); rLC2Bn1_lst.append(rLC2Bn1);
sLC2Bn0_lst.append(sLC2Bn0); sLC2Bn1_lst.append(sLC2Bn1)

            if timing and k==10:
                t1 = dt.now()
                m, s = divmod((t1-t0).total_seconds(), 60)
                h, m = divmod(m, 60)
                print('Ten iteration cost {}h:{}m:{}s'.format(int(h),
int(m), int(s)))
```

```
            m, s = divmod((t1-t0).total_seconds() * (iteration//10),
 60)
            h, m = divmod(m, 60)
            print('Expect to finish in {}h:{}m:{}s, at {}'.format(
int(h), int(m), int(s), (t0+(t1-t0)*(iteration//10)).strftime('%H:%M
:%S')))

    results, LC2Bn_lst, LC2Bn0_lst, LC2Bn1_lst = np.array(results),
np.array(LC2Bn_lst), np.array(LC2Bn0_lst), np.array(LC2Bn1_lst)
    rLC2Bn0_lst, rLC2Bn1_lst, sLC2Bn0_lst, sLC2Bn1_lst = np.array(
rLC2Bn0_lst), np.array(rLC2Bn1_lst), np.array(sLC2Bn0_lst), np.array
(sLC2Bn1_lst)

    if timing:
        t2 = dt.now()
        m, s = divmod((t2-t0).total_seconds(), 60)
        h, m = divmod(m, 60)
        print('Finshed at: {}, total cost {}h:{}m:{}s'.format(t2.
strftime('%H:%M:%S'), int(h), int(m), int(s)))

    return results, LC2Bn_lst, LC2Bn0_lst, LC2Bn1_lst


# Function brownian() simulates Brownian motions with the following
parameters.
# N: Size of the uniform partition of time interval [0, T].
# m: Number of Brownian motions to generate simultaneously.
# T: End of time interval.
# sigma: Standard deviation of Brownian motion at time t divided by
sqrt(t).
# plot: Make a plot of simulated Brownian motions if True, suggested
 when m small.
```

```
def brownian(N=2000, m=1000, T=1, sigma=1, plot=False):

    dt = T/N

    x = np.empty((m,N+1))
    x[:, 0] = 0
    x0 = np.array(x[:, 0])
    res = x[:,1:]

    r = norm.rvs(size=x0.shape + (N,), scale=sigma*sqrt(dt))
    np.cumsum(r, axis=-1, out=res)

    res += np.expand_dims(x0, axis=-1)
    x[:,1:] = res

    if plot:
        t = np.linspace(0, T, N+1)
        for k in range(m):
            plt.plot(t, x[k])
        plt.xlabel('t', fontsize=16)
        plt.ylabel('x', fontsize=16)
        plt.grid(True)
        plt.show()

    return x
```

Finally, some figures and tables are included as record of simulation results.

p_1	p_2	μ_{LC2B_n}	std_{LC2B_n}	$\mu_{LC2B_n^0}$	$std_{LC2B_n^0}$	$\mu_{LC2B_n^1}$	$std_{LC2B_n^1}$
0.1	0.1	26971.0856	42.9462	26970.9844	42.9549	26970.9848	42.9538
0.1	0.2	24000.2276	69.1092	23999.9625	69.1109	23999.9600	69.1076
0.1	0.5	15060.4689	80.9980	15042.7270	83.6495	15042.6144	83.3110
0.5	0.5	15094.8498	37.1779	15069.1133	45.5258	15068.8760	45.6263

Table A.1: Simulation of $LC2B_n$ when $n = 30000$ and $iteration = 30000$.

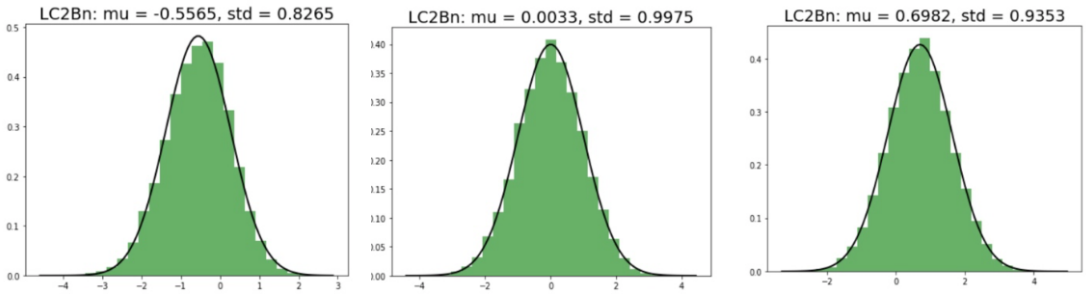

Figure A.1: Histogram of $LC2B_n$ when $p_1 = 0.1, p_2 = 0.1; p_1 = 0.1, p_2 = 0.2; p_1 = 0.1, p_2 = 0.5$.

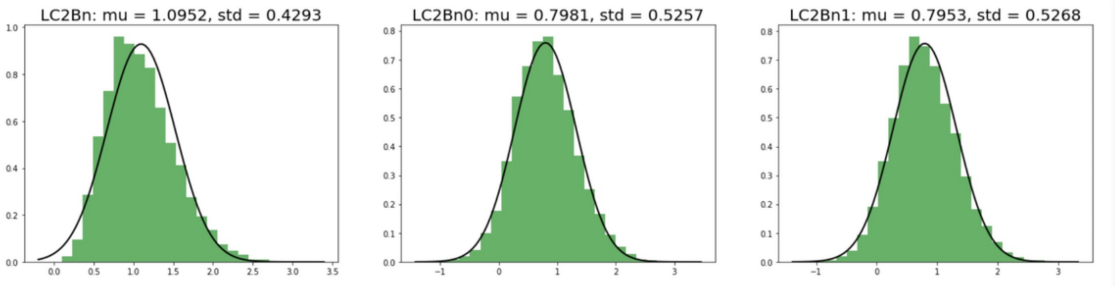

Figure A.2: Histogram of $LC2B_n$, $LC2B_n^0$, and $LC2B_n^1$ when $p_1 = 0.5, p_2 = 0.5$.

REFERENCES

[1] G. Anderson, A. Guionnet, and O. Zeitouni, "An introduction to ranom matrices," *Cambridge Studies in Advanced Mathematics*, vol. 118, 2009.

[2] J. Baik, P. Deift, and K. Johansson, "On the distribution of the length of the longest increasing subsequence of random permutations," *J. Amer. Math. Soc.*, vol. 12, 4 1999.

[3] Y. Baryshnikov, "Gues and queues," *Probab. Theory Relat. Fields*, vol. 119, pp. 256–274, 2001.

[4] F. Benaych-Georges and C. Houdré, "Gue minors, maximal brownian functionals and longest increasing subsequences in random words," *Markov Processes Relat.Fields*, vol. 21, pp. 109–126, 2015.

[5] P. Billingsley, "Convergence of probability measures," 1999.

[6] T. Bodineau and J. Martin, "A universality property for last-passage percolation paths close to the axis," *Elec. Comm. Prob.*, vol. 10, 2005.

[7] A. Borodin and P. Salminen, "Handbook of brownian motion facts and formulae," *Probability and Its Applications*, 2002.

[8] P. Bougerol and T. Jeulin, "Paths in weyl chambers and random matrices," *Probab. Theory and Relat. Fields*, vol. 124, 4 2002.

[9] J. Breton and C. Houdré, "Asymptotics for random young diagrams when the word length and alphabet size simultaneously grow to infinity," *Bernoulli*, vol. 16, pp. 471–492, 2 2010.

[10] ——, "On the limiting law of the length of the longest common and increasing subsequences in random words," *Stochastic Processes and their Applications*, vol. 127, pp. 1676–1720, 2017.

[11] G. Chistyakov and F. Götze, "Distribution of the shape of markovian random words," *Probab. Theory Related Fields*, vol. 129, pp. 18–36, 1 2004.

[12] V. Chvtal and D. Sankoff, "Longest common subsequences of two random sequences," *Journal of Applied Probability*, vol. 12, 2 1975.

[13] A. Delcher and S. Kasif, "Aligment of whole genomes," *Nucleic Acids Research*, vol. 27, 11 1999.

[14] C. Deslandes and C. Houdré, "On the limiting law of the length of the longest common and increasing subsequences in random words with arbitrary distributions," *ArXiv #Math.PR/1906.06544*,

[15] L. Duraj, "A linear algorithm for 3-letter longest common weakly increasing subsequence," *Information Processing Letters*, vol. 113, 3 Feb. 2013.

[16] W. Fulton, "Young tableaux: With applications to representation theory and geometry," *London Mathematical Society Student Texts*, 1996.

[17] P. W. Glynn and W. Whitt, "Departures from many queues in series," *Ann. Appl. Probab.*, vol. 1, 4 1991.

[18] J. Gravner, C. Tracy, and H. Widom, "Limit theorems for height fluctuations in a class of discrete space and time growth models," *J. Stat. Phys.*, vol. 102, pp. 1085–1132, 2001.

[19] C. Houdré and U. Işlak, "Asymptotic representation theory of the symmetric group and its applications in analysis," *AMS, Translations of Mathematical Monographs*, vol. 219, 2014.

[20] C. Houdré, J. Lember, and H. Maztinger, "On the longest common increasing binary subsequence," *C.R. Acad. Sci., Paris Ser. I*, vol. 343, 2006.

[21] C. Houdré and T. Litherland, "Asymptotics for the length of the longest increasing subsequence of a binary markov random word," *ArXiv #Math.PR/1110.1324v2*, 2012.

[22] ——, "On the longest increasing subsequence for finite and countable alphabets," *High Dimensional Probability V*, vol. The Luminy Volume, pp. 185–212, 2009.

[23] C. Houdré and H. Matzinger, "On the variance of the optimal alignments score for binary random words and an asymmetric scoring function," *Journal of Statistical Physics*, vol. 164, 3 Aug. 2016.

[24] C. Houdré and H. Xu, "On the limiting shape of young diagrams associated with inhomogeneous random words," *High Dimensional Probability VI. Progress in Probability*, vol. 66, 2013.

[25] A. Its, C. Tracy, and H. Widom, "Random words, toeplitz determinants, and integrable systems i," *Random matrix models and their applications*, vol. 40, pp. 245–258, 2001.

[26] A. Its, H. Widom, and C. Tracy, "Random words, toeplitz determinants, and integrable systems ii," *Phys. D, Adv. in nonlin. mathem. and sci.*, pp. 199–224, 2001.

[27] J. S. J. Lember H. Matzinger and F. Zucca, "Lower bounds for moments of global scores of pairwise markov chains," *ArXiv:1602.05560 [math]*, Feb. 2016.

[28] K. Johansson, "Discrete orthogonal polynomial ensembles and the plancherel measure," *Ann. Math.*, vol. 2, pp. 259–296, 1 2001.

[29] ——, "Transversal fluctuations for increasing subsequences on the plane," *Probab Theory Relat Fields*, vol. 116, 4 Apr. 2000.

[30] I. Karatzas and S. Shrev, "Brownian motion and stochastic calculus," 1998.

[31] S. Kerov, "Asymptotic representation theory of the symmetric group and its applications in analysis," *AMS, Translations of Mathematical Monographs*, vol. 219, 2003.

[32] G. Kuperberg, "Random words, quantum statistics, central limits, random matrices," *Methods Appl. Anal.*, vol. 9, 1 2002.

[33] T. Litherland and C. Houdré, "On the limiting shape of young diagrams associated with markov random words," *ArXiv #Math.PR/1110.4570*, 2011.

[34] R. Łochowski, "On truncated variation of brownian motion with drift," *Bulletin of The Polish Academy of Sciences Mathematics*, vol. 56, 2008.

[35] R. Łochowski and P. Miłos, "On truncated variation, upward truncated variation and downward truncated variation for diffusions," *Stochastic Processes and their Applications*, vol. 123, 2 2013.

[36] J. Martin, "Limiting shape for directed percolation models," *The Annals of Probability*, vol. 32, 4 2004.

[37] J. Pitman, "One-dimensional brownian motion and the three-dimensional bessel process," *Adv. Appl. Prob.*, vol. 7, 511–526, 1975.

[38] A. Polak, "Why is it hard to beat o(n2) for longest common weakly increasing subsequence?" *ArXiv:1703.01143 [cs.CC]*, 2017.

[39] D. Romik, "The surprising mathematics of longest increasing subsequences," *Institute of Mathematical Statistics Textbooks*, 2015.

[40] T. Seppäläinen, "A scaling limit for queues in series," *Ann. Appl. Probab.*, vol. 7, 4 1997.

[41] J. M. Steele, "An efron-stein inequality for nonsymmetric statistics," *The Annals of Statistics*, vol. 14, 2 Jan. 1986.

[42] C. Tracy and H. Widom, "On the distribution of the lengths of the longest increasing monotone subsequences in random words," *Probab. Theor. Rel. Fields.*, vol. 119, pp. 350–380, 2001.