Abraham Gutierrez Sanchez, Dipl.-Ing.

Probabilistic Aspects in Automata and Digraphs

DOCTORAL THESIS

to achieve the university degree of

Doktor der technischen Wissenschaften

submitted to

Graz University of Technology

Supervisor

Dr. Daniele D'Angeli

Institute of Discrete Mathematics

Graz, December 2020

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date

Signature

Abstract

In the first chapter of this thesis, we prove that a uniformly distributed random circular automaton $\mathcal{A}_n$ of order n synchronizes with high probability (w.h.p.). More precisely, we prove that

$$\mathbb{P}\left[\mathcal{A}_n \text{ synchronizes}\right] = 1 - O\left(\frac{1}{n}\right).$$

The main idea of the proof is to translate the synchronization problem into a problem concerning properties of a random matrix; these properties are then established with high probability by a careful analysis of the stochastic dependence structure among the random entries of the matrix. Additionally, we provide an upper bound for the probability of synchronization of circular automata in terms of chromatic polynomials of circulant graphs.

Multivariate regression models and ANOVA are probably the most frequently applied methods of all statistical analyses. In the second chapter of this thesis, we propose an alternative to the classic approaches that do not assume homoscedasticity or normality of the error term but assumes that a Markov chain can describe the covariates' correlations. This approach transforms the dependent covariate using a change of measure to independent covariates. The transformed estimates allow a pairwise comparison of the mean and variance of the contribution of different values of the covariates. We show that under standard moment conditions, the estimators are asymptotically normally distributed. Additionally, we test our method with simulated data and apply it to several classic data sets.

Acknowledgements

I've been thinking for some time now, how to write this section. How to express my gratitude to the people that made this journey possible, successful, instructive, and enjoyable. First, I'd like to say that my gratitude goes beyond the words written on this single-page section. In my humble opinion, gratitude is expressed best in the natural flow of daily routine—nothing fancy; simple, kind words and actions; readiness to act when needed.

The first time I encounter my advisor Daniele D'Angeli, was in the school's corridors, where the German courses took place. We weren't even acquainted with each other at that time. I couldn't have any idea about the significant role he'd play years ahead in my life when he offered me the chance to work with him in the Ph.D. Some months before I finished my Master's studies under Oliver Cooley's diligent supervision, Daniele and I discussed possible Ph.D. topics. Even though I was not fully familiarized with these topics, I had a good feeling about undertaking a Ph.D. under his supervision. He has always been extremely flexible and kind to me. These points were crucial to fuel me throughout the studies' challenging times.

I'm deeply in debt to my co-advisors, Christoph Aistleitner and Emanuele Rodaro. Two outstanding sources of inspiration, encouragement, and great humour.

I'm very grateful to all my collaborators. They allowed me to grow as a scientist through rich interactions and the exchange of ideas. Here is a list of my current collaborators in alphabetic order by last name: Christoph Aistleitner, Mickaël Buchet, Sylvain Clair, Oliver Cooley, Daniele D'Angeli, Manfred Liebmann, José Luis Martínez Morales, Sebastian Müller, Emanuele Rodaro, Amnon Rosenmann and Stjepan Šebek.

I wish to sincerely thank the reviewers of this work, Cyril Nicaud and Marc Peigné. Thanks for kindly accepting to do the review, especially during Covid-19 times. I am pretty sure you have an extra tight schedule in this period. Thanks for your kind and useful feedback.

I wish to thank all my friends. They significantly contributed to make my journey a delightful one. Two special mentions to Daniel El-Baz and Marc Alexandre Munsch, whose great stories, remarks, sense of humour, and storytelling skills always kept me "on the edge of the seat". Thanks, Reinhard Lüftenegger, for your neat company and for teaching me how to ski.

There are many more people I wish to thank, but space restrictions compel me to stop. To those of you that I haven't mentioned, I want you to know the following: this section might stop here, but my gratitude does not.

Contents

Introduction

The central point of this thesis is to use probabilistic tools to study discrete structures: circular automata and DAGs (directed acyclic digraphs). In the first chapter, we pose a question of the following type: "how probable is that one discrete structure (a circular automaton) will have a given property (synchronizing) if we choose one of them at random?" In the second chapter, we pose a reverse engineering question: "given an indirect information about a discrete random structure (DAG with random variables in each node), what can we say about the nature of its internal states (moments of random variables associated to each node)?"

The main motivation of Chapter I are the exiting results made in recent years by Berlinkov and Nicaud on the probabilistic Černý conjecture ([Berlinkov, 2016], [Nicaud, 2019]). The main tools used in Chapter I are related to graph theory, automata theory and probabilistic tools. In more specific terms, one of the hardest bottlenecks in the first chapter was a tailor-made treatment of a non-Lipschitz random variable (see Lemma 7). The lack of general results in the field of concentration inequalities for non-Lipschitz random variables forced us to create a tailor-made solution; this proof strongly contrasts with the relative simplicity of Lemma's 6 proof, for which we were able to apply a ready-to-use tool (MacDiarmid's Inequality, Azuma's inequality) for Lipschitz random variables. I also must add that there might be an alternative way to prove Lemma 7 via chromatic-polynomials. We were unfortunately not able to realize this alternative proof because the necessary results are in some cases generalizations of well-known open problems in the chromatic-polynomial's community. Nevertheless, we consider that these connections can be of interest to the graph-theoretical community. This was the main motivation to add Section 1.6 in Chapter I.

The problem in Chapter II was proposed to me by Alessandro Chiancone, who at the time was working in Know-Center in Graz. As a whole, I find the motivation of this chapter's problem very natural and easy to explain. This was one of the main reasons I felt so motivated to work on this problem. This allowed to create a suitable mathematical model and solve the first moment case in the uniform and non-uniform cases. In addition, this natural motivation was big help to easily enlist the invaluable expertise of Sebastian Müller. Joining forces, we lifted the solution to the second moment and figured out the asymptotic distributions of the constructed estimators. The main tools used in the second chapter are related to Markov chains, change of measures and stopped random walks. One of my favorite tools, used in this chapter, is the Rényi's version of Ascombe's theorem which is a central limit theorem in which the deterministic sample size n is replaced by a random variable. This result was essential to calculate the asymptotic distribution of the estimators we constructed. I also find quite interesting the change of measure we did in order to retrieve information about the moments in the Markov chain case.

The contents of Chapter I are based on the paper [Aistleitner et al., 2021] which was written in collaboration with Christoph Aistleitner, Daniele D'Angeli, Emanuele Rodaro and Amnon Rosenmann. The contents of Chapter II are based on the papers [Gutierrez and Müller, 2019] and [Gutierrez and Müller, 2020] which were written in collaboration with Sebastian Müller.

The following subsections are meant to be an informal explanation of the thesis' chapters.

Chapter I

This subsection is an informal explanation of Chapter I. Let us consider the following puzzle (see Figure 1). The rules are the following: given a *state* s (0,1,2 or 3, see Figure 1) and word w formed by the letters a and b (e.g aba), we position ourselves into the state s and then follow the arrows according to the letters (from left to right) in the word w. For example, if we begin in the state 0 and the given word is aba, first we move from state 0 to state 1 after reading a and then we stay in state 1 after reading the b and finally we move from state 1 into state 2 after reading the last letter a in the word. So, we say that the *ending state* in this case is 2; we abbreviate this in the following way

$$0 \cdot aba = 2.$$

Now that we understand the rules of the game, the puzzle is to find out if there is a word w such that

$$0 \cdot w = 0; \quad 1 \cdot w = 0;$$
$$2 \cdot w = 0; \quad 3 \cdot w = 0;$$

i.e. no matter where you begin, the word w will send you to the state 0 at the end. Such a word exits indeed, one example of a word that accomplishes this is $baaabaaabaaab$ or in short notation $ba^3ba^3ba^3b$; actually, this is just one of many words that accomplish this task.

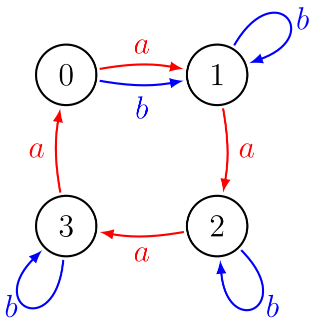


Figure 1: The Černý automaton of order four.

We now ask a more general question: is there a word such w such that

$$0 \cdot w = s; \quad 1 \cdot w = s;$$
$$2 \cdot w = s; \quad 3 \cdot w = s;$$

for some state s? i.e. now we are only interested in a word w with a unique ending state; we say that such a word is *synchronizing* (or *reset*) word for the automaton. Of course, our previous word $ba^3ba^3ba^3b$ is synchronizing but now there are new words like ab^3ab^3a which have ending state 0 and are *shorter*. In particular, one can ask: which is the shortest synchronizing word? One can prove that the *minimal* word for the automaton in Figure 1 is ab^3ab^3a.

A *deterministic finite automaton* $\mathcal{A}$ is a tuple of states Q and input alphabet Σ which are mappings from Q to Q, they basically tell you "if you are in state s and you read the letter l, then you must go to state $\hat{s}$". If an automaton has synchronizing words, we say that the automaton is *synchronizing*. An automaton is not necessarily synchronizing (see Figure 2); it is relatively easy to construct a non-synchronizing automaton: just take each letter in the alphabet Σ to be a permutation of Q. Let us explain this, if l induces a permutation, we have that $|Q \cdot l| = |Q|$ i.e. the cardinality of the state-set stays the same after applying l to each state, therefore if every letter in alphabet is a permutation, we have that $|Q \cdot w| = |Q|$ for every word w; on the other hand, a synchronizing word $\hat{w}$ satisfies that $|Q \cdot \hat{w}| = 1$ because every state will be sent to a unique state. Therefore it is impossible for an automaton with at least two states to be synchronizing if every word in its alphabet induces a permutation. In addition, there are non-synchronizing automata whose letters are not all permutations; consider the automaton on Figure 2b.

Figure 2: Examples of non-synchronizing automata. The letters in automaton (a) are all permutations. The letter b of automaton (b) is not a permutation.

We are interested in automata of the form $\mathcal{A} = \langle \{0, 1, \ldots, n-1\}, \{a, b\} \rangle$ where a is the cycle

$$0 \xrightarrow{a} 1 \xrightarrow{a} 2 \xrightarrow{a} \ldots \xrightarrow{a} n-1 \xrightarrow{a} 0,$$

and b is a mapping from $\{0, 1, \ldots, n-1\}$ to itself. This class of automata is called *circular*

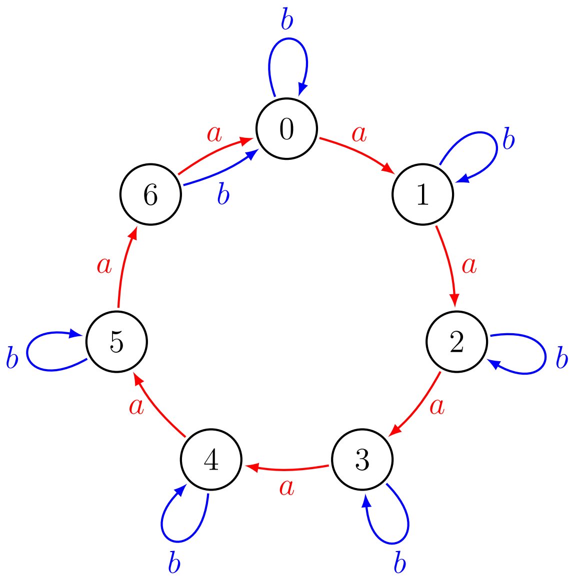


Figure 3: A circular automaton with six states.

automata with n states and two letters; in particular Figures 1, 2a, 3 are examples of circular automata with two letters and four, two and six states respectively. We are interested in the following question: if we choose one circular automaton with n states and two letters uniformly at random, how probable is that the chosen automaton is synchronizing? The main result of Chapter I answer this question: if n is "big enough," it is very probable that a circular automaton with n states will be synchronizing. In more precise terms, we prove

Theorem. *There exist a constant $c > 0$ such that*

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n : \ \mathcal{A}_n(\mathbf{b}) \textit{ synchronizes}\}\right] \geq 1 - \frac{c}{n}.$$

In words, an uniformly randomly chosen circular automaton $\mathcal{A}_n(\mathbf{b})$ synchronizes with high probability as $n \to \infty$.

Chapter II

This subsection is an informal explanation of Chapter II. Imagine that we have a network that is an arrangement of machines into c columns and at most r rows in each column (see Figure 4). The raw material is delivered from the source node s on the left-hand side of the network into a random machine of the first column, then it moves into a random machine of the second column and so on until it ends up in the terminal node t as a finished product p. To the product p, we associate $\vec{p}$, which is the path that p followed in the network; we call $\vec{p}$ the production-path of p. A finished product might have flaws at the end of its production, therefore, each finished product p gets assigned a quality-score $b(p)$. Given the data $\{(\vec{p_1}, b(p_1)), (\vec{p_2}, b(p_2)), \ldots, (\vec{p_k}, b(p_k))\}$, what can we say about the "broken" machines? Can we identify them? Let us begin with a simple

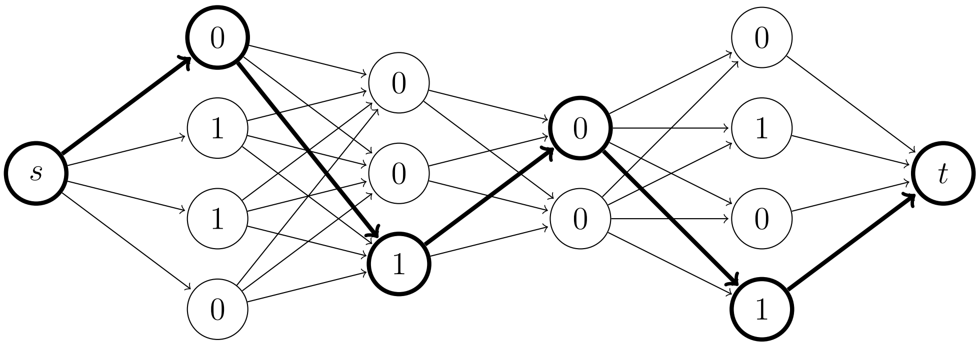


Figure 4: An illustration of a network with 4 columns. The highlighted path is $\vec{p} = (1, 3, 1, 4)$ and its quality-score is $b(p) = 0 + 1 + 0 + 1 = 2$.

example: assume that the machine in row i and column j is assigned the value $S(i, j) \in \mathbb{R}$. As an example, we can think on the binary case $S(i, j) \in \{0, 1\}$ where a broken machine is assigned a 1 and a non-broken machine a 0. Furthermore, the product's quality in a specific product p is defined as

$$b(p) = \sum_{j=1}^{c} S(\vec{p}[j], j)$$

i.e. the addition of the assigned values of the machines in its production-path $\vec{p}$. For example, in the binary case, $b(p)$ is the number of broken machines lying in its production-path $\vec{p}$ (see Figure 4 for an example). Can we find the values $S(i, j)$? The answer to this question depends on the network and the assumptions we can make about the values $S(i, j)$. For example, we will soon see that in the binary setting $S(i, j) \in \{0, 1\}$, if we have enough data and under mild assumptions (at least one non-broken machine per column) we can recover the values $S(i, j)$. In the general setting $S(i, j) \in \mathbb{R}$, the problem can be transformed into a neat linear algebra problem, where $\{S(i, j)\}_{i,j}$ is the set of variables. Then

$$b(p) = \sum_{j=1}^{c} S(\vec{p}[j], j)$$

is a linear equation on the variables $S(i, j)$. Furthermore, if there are w machines in the network, let us numerate the machines in a fixed order from 1 to w, then we can associate to each product

p a vector $e(p) \in \mathbb{R}^w$ were

$$e(p)[k] = \begin{cases} 1, & \text{if the production-path of } p \text{ passes through machine } k; \\ 0, & \text{otherwise;} \end{cases}$$

see Figure 5 for an example. So the question is: given the vectors $e(p)$, can we find the values $S(i,j)$? The answer depends on the network at hand (see Figures 5 and 7 for examples). For large enough networks (enough columns and enough machines per column) one can prove that the vectors

$$\vec{v}_k = (0, 0, \ldots, 0, \underbrace{1}_{k}, 0, \ldots, 0)$$

are linearly independent from the vectors $e(p)$. Therefore, if we only know that $S(i,j) \in \mathbb{R}$, the values $S(i,j)$ can not be calculated given the pairs $(\vec{p}, b(p))$ (see Figure 5 for an example). Even if we assume that $S(i,j) \in \{0,1\}$, there are cases where it is not possible to calculate these values (see Figures 7 and 8 for examples). On the other hand, vectors of the form

$$\vec{v}_{k,k'} = (0, 0, \ldots, 0, \underbrace{1}_{k}, 0, \ldots, \underbrace{-1}_{k'}, 0, \ldots, 0)$$

where the machines associated to k, k' belong to the same column, can be generated as linear combinations of the vectors $e(p)$ which means that differences of the type $S(i,j) - S(i',j)$ can be calculated. For example, in the binary case $S(i,j) \in \{0,1\}$, the value $S(i,j) - S(i',j)$ equals $1, -1$ or 0. If $S(i,j) - S(i',j) = 1$ then $S(i,j) = 1$ meaning that the machine located in row i and column j is broken; if $S(i,j) - S(i',j) = -1$ then $S(i',j) = 1$ and if $S(i,j) - S(i',j) = 0$ then $S(i,j) = S(i',j)$ which means that both machines are either broken or non-broken. So, in the binary case, if we assume that there is at least one non-broken machine in each column, the problem of recognizing the broken machines in the network can be indeed solved given that the amount of data (enough equations) is sufficient.

In the general case $S(i,j) \in \mathbb{R}$, given the data $(\vec{p}_1, b(p_1)), (\vec{p}_2, b(p_2)), \ldots, (\vec{p}_x, b(p_x))$ for x products, how can we know if the value $S(i,j) - S(i',j)$ can be retrieved from this data? Let k, k' be the numbers associated to the machines with locations (i,j) and (i',j) respectively and let r be the rank of the matrix with rows $e(p_1), \ldots, e(p_x)$ and let r' the rank of the matrix with rows $e(p_1), \ldots, e(p_x), \vec{v}_{k,k'}$ then: if $r = r'$ this means that $\vec{v}_{k,k'}$ can be expressed as a linear combination in terms of the vectors $e(p_1), \ldots, e(p_t)$ which means that the value of $S(i,j) - S(i',j)$ can indeed be calculated out of the data $(\vec{p}_1, b(p_1)), (\vec{p}_2, b(p_2)), \ldots, (\vec{p}_x, b(p_x))$. On the other hand, if $r < r'$ then the vector $\vec{v}_{k,k'}$ is independent from the vectors $e(p_1), \ldots, e(p_t)$, which means that the value of $S(i,j) - S(i',j)$ can not be retrieved from the data $(\vec{p}_1, b(p_1)), (\vec{p}_2, b(p_2)), \ldots, (\vec{p}_x, b(p_x))$. Therefore, if $r < r'$, we need to wait until $r = r'$ in order to calculate the value $S(i,j) - S(i',j)$.

The previous example shows that the problem is relatively simple when the variables $S(i,j)$ are constants. If we assume that the variables $S(i,j)$ are random variables, we can no longer use the previous approach. Considering random variables instead of constants is a more realistic scenario: the "amount" of mistakes behaves randomly every time a product passes through a broken machine. In Chapter II we study this setting and create estimators for the values $\mathbb{E}[S(i,j)] - \mathbb{E}[S(i',j)]$ and $\mathbb{V}[S(i,j)] - \mathbb{V}[S(i',j)]$ which allows to compare the expected values and variances of the random variables of machines that are located in the same column. We treat two main cases: when the transitions from column to column are uniform and the case when a Markov chain describes the transitions.

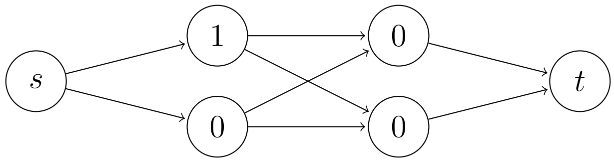


Figure 5: In this network the equations associated to production paths are $S(1,1)+S(1,2)=1$, $S(1,1)+S(2,2)=1$, $S(2,1)+S(1,2)=0$, $S(2,1)+S(2,2)=0$ which in the binary case $S(i,j)\in\{0,1\}$ is enough to deduce that $S(1,1)=1$ and $S(2,1)=S(1,2)=S(2,2)=0$. Let us order the machines in the following way $S(1,1), S(2,1), S(1,2), S(2,2)$ then the corresponding $e(p)$ vectors are $(1,0,1,0), (1,0,0,1), (0,1,1,0), (0,1,0,1)$ which are linearly independent from $(1,0,0,0)$ and by symmetry also linearly independent from $(0,1,0,0)$ and $(0,0,1,0)$ and $(0,0,0,1)$. This means that it is not possible to calcualte the values $S(i,j)$ in the general case $S(i,j)\in\mathbb{R}$.

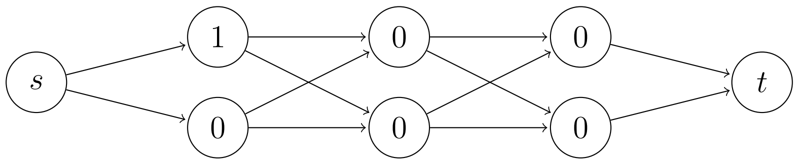


Figure 6: In this network, it is possible to calculate the values $S(i,j)$ in the binary case $S(i,j)\in\{0,1\}$ but it is not possible to do it in the general case $S(i,j)\in\mathbb{R}$.

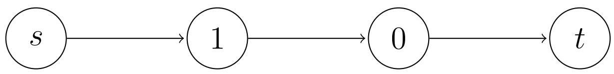


Figure 7: In this network it is not possible to know which machine $(S(1,1), S(1,2))$ is broken since the only equation at hand is $S(1,1)+S(1,2)=1$ which is not enough to deduce if $S(1,1)=1$ or $S(1,2)=1$.

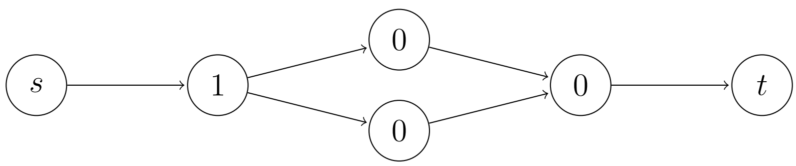


Figure 8: In this network it is not possible to know if $S(1,1)=1$ or $S(1,3)=1$ given the pairs $(\vec{p}, b(p))$.

Chapter 1

Circular automata synchronize with high probability

In this chapter we prove that a uniformly distributed random circular automaton $\mathcal{A}_n$ of order n synchronizes with high probability (w.h.p.). More precisely, we prove that

$$\mathbb{P}\left[\mathcal{A}_n \text{ synchronizes}\right] = 1 - O\left(\frac{1}{n}\right).$$

The main idea of the proof is to translate the synchronization problem into a problem concerning properties of a random matrix; these properties are then established with high probability by a careful analysis of the stochastic dependence structure among the random entries of the matrix. Additionally, we provide an upper bound for the probability of synchronization of circular automata in terms of chromatic polynomials of circulant graphs.

The contents of this chapter are based on the paper [Aistleitner et al., 2021] which was written in collaboration with Christoph Aistleitner, Daniele D'Angeli, Emanuele Rodaro and Amnon Rosenmann.

1.1 Introduction

A *complete deterministic finite automaton* (DFA) is a tuple $\mathcal{A} = (Q, L)$, where $Q := \{q_1, q_2, \ldots, q_n\}$ is a finite set of *states* and $L := \{a_1, a_2, \ldots, a_k\}$ is a finite set of mappings $a_i : Q \to Q$, where $a(q) = q'$ is also written as $qa = q'$, $q, q' \in Q$, $a \in L$. The number of states n is the *order* of $\mathcal{A}$. Each a_i is called a *letter* and a sequence $w = a_{i_1} a_{i_2} \ldots a_{i_r} \in L^*$ is a *word* of *length* r. The action of L on Q naturally extends to an action of L^* on Q, defined recursively by $q(aw) = (qa)w$, $q \in Q$, $a \in L$, $w \in L^*$. This action further extends to an action of L^* on subsets of Q by $\{q_{i_1}, q_{i_2}, \ldots, q_{i_k}\}w = \{q_{i_1}w, q_{i_2}w, \ldots, q_{i_k}w\}$. We say that the subset $S = \{q_{i_1}, q_{i_2}, \ldots, q_{i_k}\} \subseteq Q$ *synchronizes* if there exists a word $w \in L^*$ such that $q_{i_1}w = q_{i_2}w = \ldots = q_{i_k}w$ (equivalently, we say that w synchronizes S). If the set Q synchronizes then we say that $\mathcal{A}(Q, L)$ *synchronizes* (or that it is a synchronizing automaton). A word $w \in L^*$ that synchronizes Q is called a *synchronizing* (or *reset*) word of $\mathcal{A}$.

The following simple criterion for synchronization is well-known and plays a crucial role throughout the chapter:

Claim 1. *$\mathcal{A} = (Q, L)$ synchronizes $\iff$ every pair of states $q, q' \in Q$ synchronizes.*

Proof. It is clear that if Q is synchronized by a reset word w then w synchronizes every pair of states of Q. Conversely, a reset word for Q can be formed by concatenating words w_i that synchronize pairs of states until we end up with a single state. □

The synchronization property may be described in terms of the graph representation of $\mathcal{A}$. The set Q of states comprises the vertices of the graph and for each pair of states q, q' and a letter $a \in L$ such that $qa = q'$ there is an arrow $(q, q')_a$ labeled with $a \in L$ and connecting q to q'. Each $q \in Q$ and $w = a_{i_1} a_{i_2} \ldots a_{i_k} \in L^*$ defines a directed path

$$\gamma(qw) := q, qa_{i_1}, qa_{i_1}a_{i_2}, \ldots, qa_{i_1}a_{i_2} \cdots a_{i_k}$$

that begins in q and ends in qw. $\mathcal{A}$ then synchronizes if and only if there is a word w, such that the paths $\{\gamma(qw) : q \in Q\}$ have a common endpoint.

Synchronizing automata have been intensely studied by theoretical computer scientists as well as pure mathematicians since the 1960's; see [Volkov, 2008] for a detailed introduction on synchronization of automata. A driving force in this research field is the Černỳ conjecture.

Conjecture 2 (The Černỳ conjecture). A synchronizing automaton $\mathcal{A}$ of order n has a shortest synchronizing word of length at most $(n-1)^2$.

The bound in the Černỳ conjecture is tight: In [Cerny, 1964] Černỳ provided a series of synchronizing circular automata $C_2, C_3, \ldots$, such that C_n has order n and its shortest synchronizing word is of size exactly $(n-1)^2$ (see Fig. 1.1). Furthermore, the Černỳ series of circular automata $C_2, C_3, \ldots$ is the only known infinite series of automata whose shortest synchronizing words are of length $(n-1)^2$ [Ananichev et al., 2010]. The best known general upper bounds for the size of shortest synchronizing words of an automaton with n states are of order $O(n^3)$ [Frankl, 1982][Pin, 1983][Szykuła, 2017][Shitov, 2019]. Nevertheless, there are many classes of automata for which the Černỳ conjecture has been established (see [Volkov, 2008] for examples).

During the last decade, probabilistic approaches to the synchronization problem have been developed. Typical questions in this setting are: let $\mathcal{A}(\{0, 1, \ldots, n-1\}, L)$ be a uniformly chosen DFA with k letters on a certain probability space, is it true that with high probability the automaton $\mathcal{A}(\{0, 1, \ldots, n-1\}, L)$ is synchronizing? Does the Černỳ conjecture hold with high probability? Here we give a (non-comprehensive) list of recent achievements in this probabilistic setting:

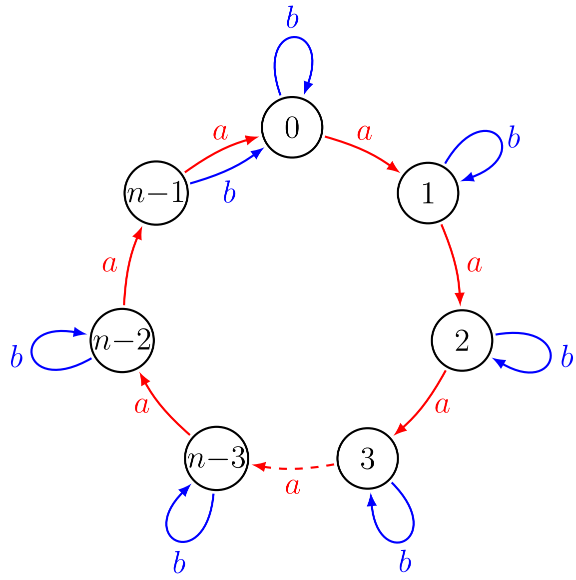


Figure 1.1: The automaton C_n

- In [Skvortsov and Zaks, 2010] the authors study random automata $\mathcal{A}$ where the number of letters k grow together with n. In particular, they prove that $\mathcal{A}$ synchronizes w.h.p. when $k(n)$ grows fast enough;
- In [Berlinkov, 2016] the author proves that $\mathbb{P}[\mathcal{A} \text{ synchronizes}] = 1 - O(n^{-k/2})$, for arbitrary $k \geq 2$, it is also proved that $\mathbb{P}[\mathcal{A} \text{ synchronizes}] = 1 - \Theta(1/n)$ for $k = 2$;
- In [Nicaud, 2019] the author proves that $\mathcal{A}$ admits w.h.p. a synchronizing word of length $O(n \log^3 n)$ for arbitrary $k \geq 2$;
- In [Berlinkov and Nicaud, 2018] the authors prove that if $\mathcal{A}$ is uniformly chosen among the strongly-connected almost-group automata then $\mathcal{A}$ synchronizes with probability $1 - \Theta((2^{k-1} - 1)n^{-2(k-1)})$ for arbitrary $k \geq 2$.

Since the sequence of circular automata C_n depicted in Fig. 1.1 is the only known infinite series of synchronizing automata reaching Černý 's bound $(n-1)^2$, one might suspect that the class of circular automata is somehow difficult to synchronize. However, as we show in the present chapter, it turns out that a random circular automaton is synchronizing with high probability.

The rest of the chapter is organized as follows: in Section 1.2 we present the main result together with its proof and the statement of the two key lemmas for the proof. In Section 1.3 we study the dependence structure of the random matrix used in the proof of the main result; the result obtained in this section is crucial for the proof of the key lemmas. In Section 1.4 we prove the first lemma while in Section 1.5 we prove the second one. In Section 1.6 we present some interesting connections between synchronization of circular automata and chromatic polynomials of circulant graphs. Finally, in Section 1.7 we present some possible directions towards generalizing and improving the results presented in this chapter.

1.2 Main result

Let n be a positive integer. An automaton $\mathcal{A}(\mathbb{Z}_n, L)$, where $\mathbb{Z}_n := \{0, 1, \ldots, n-1\}$ is the set of states, is called a *circular automaton* if L contains a permutation that decomposes in exactly one cycle. Let $(i)_n := i \mod n$. Let $\mathcal{M}_n$ denote the set of all mappings from $\mathbb{Z}_n$ to itself, and let $\mathbb{P}$ denote the uniform probability measure on $\mathcal{M}_n$. We will write the elements of $\mathcal{M}_n$ as vectors by identifying the mapping $\mathbf{b}(i) = b_i,\ i = 0, \dots, n-1$ with the vector $\mathbf{b} = (b_0, \dots, b_{n-1})$.

In what follows, we denote by $\mathcal{A}_n(\mathbf{b}) := (\mathbb{Z}_n, \{\mathbf{a}, \mathbf{b}\})$ a circular automaton of order $n \in \mathbb{N}$, with $\mathbf{a} : \mathbb{Z}_n \to \mathbb{Z}_n$ being the circular right shift permutation $a(i) = (i+1)_n$ and $\mathbf{b} := (b_0, ..., b_{n-1})$ being an element of $\mathcal{M}_n$. We will understand that $\mathbf{b}$ is "randomly" chosen from $\mathcal{M}_n$ according to the uniform probability measure $\mathbb{P}$, making $\mathcal{A}_n(\mathbf{b})$ a random circular automaton.

In 1977 as a consequence of a work of [Perrin, 1977] it follows that a circular automaton $\mathcal{A}(Q, L)$ of prime order synchronizes if and only if L contains a non-permutation. In 1978 [Pin, 1978] proved with combinatorial methods that a circular automaton $\mathcal{A}(Q, L)$ of prime order which has a letter of rank $\frac{n-1}{2} \leq k \leq n$ has a minimal word of size at most $(n-k)^2$. We restate Perrin's implicit theorem in a probabilistic way.

Theorem 3 ([Perrin, 1977][Pin, 1978])**.** *Let p be a prime. Then*

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_p : \ \mathcal{A}_p(\mathbf{b}) \ \textit{synchronizes}\}\right] = 1 - \frac{p!}{p^p} = 1 - \Theta\left(\frac{\sqrt{p}}{e^p}\right).$$

Thus, a uniformly distributed random circular automaton of prime order p with $k \geq 2$ letters synchronizes with high probability (w.h.p.).

Remark. Theorem 3 is not explicitly stated in [Perrin, 1977], but it's observed by [Pin, 1978] that Perrin's work implies the theorem.

The Černý conjecture holds true for the class of circular automata [Dubuc, 1998]. In a closely related work, Béal, Berlinkov and Perrin [Béal et al., 2011] gave an $O\left(n^2\right)$ upper bound for the shortest words of synchronizing automata with a single cluster.

A natural question arises: do random circular automata of order n (not necessarily prime) synchronize with high probability? We give a positive answer to this question in the following:

Theorem 4 (**Main result**)**.** *We have*

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n : \ \mathcal{A}_n(\mathbf{b}) \ \textit{synchronizes}\}\right] = 1 - O\left(\frac{1}{n}\right)$$

as $n \to \infty$. In words, a randomly chosen $\mathcal{A}_n(\mathbf{b})$ synchronizes w.h.p. as $n \to \infty$.

Remark 5. Theorem 4 does not follow from the results of Berlinkov or Nicaud. In their models, they use a random automaton $\mathcal{A}(Q, L)$ of order n where L is a collection of k mappings from Q to Q i.i.d. uniformly chosen. For a fixed k, the probability of randomly chosen k mappings to contain a permutation with exactly one cycle is bounded from above by $k \cdot \frac{n!}{n^n} \xrightarrow{n\to\infty} 0$.

Given $n \in \mathbb{N}$ and $r \in \mathbb{Z}$, we define the *n-cyclic absolute value* of r to be

$$\left|r\right|_n := \min\left\{(r)_n, (-r)_n)\right\} \in \left\{0, 1, \dots, \left\lfloor\frac{n}{2}\right\rfloor\right\}.$$

When $r, s \in \mathbb{Z}$ then $\left|r - s\right|_n$ is the *n-cyclic distance* between r and s. When the numbers $0, 1, \dots, n-1$ are identified with the vertices of a cycle of length n, the n-cyclic distance between two such numbers is the length of the shortest path between them in the cycle. We now introduce the main tool for the proof of the main theorem.

Definition. Let $\mathcal{A}_n(\mathbf{b}) := (\mathbb{Z}_n, \{\mathbf{a}, \mathbf{b}\})$ be a circular automaton with $\mathbf{b} = (b_0, b_1, \ldots, b_{n-1})$.

$$T_{\mathbf{b}} := \begin{bmatrix} \left|b_0 - b_1\right|_n & \left|b_1 - b_2\right|_n & \cdots & \left|b_k - b_{k+1}\right|_n & \cdots & \left|b_{n-1} - b_0\right|_n \\ \left|b_0 - b_2\right|_n & \left|b_1 - b_3\right|_n & \cdots & \left|b_k - b_{(k+2)_n}\right|_n & \cdots & \left|b_{n-1} - b_1\right|_n \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \left|b_0 - b_i\right|_n & \left|b_1 - b_{1+i}\right|_n & \cdots & \left|b_k - b_{(k+i)_n}\right|_n & \cdots & \left|b_{n-1} - b_{(n-1+i)_n}\right|_n \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \left|b_0 - b_{\lfloor \frac{n}{2} \rfloor}\right|_n & \left|b_1 - b_{1+\lfloor \frac{n}{2} \rfloor}\right|_n & \cdots & \left|b_k - b_{(k+\lfloor \frac{n}{2} \rfloor)_n}\right|_n & \cdots & \left|b_{n-1} - b_{(n-1+\lfloor \frac{n}{2} \rfloor)_n}\right|_n \end{bmatrix}, \tag{1.1}$$

shortly written as

$$T_{\mathbf{b}}(i,j) = \left|b_j - b_{(j+i)_n}\right|_n \text{ for } 1 \le i \le \left\lfloor \frac{n}{2} \right\rfloor \text{ and } 0 \le j \le n-1.$$

As before, $b_i = \mathbf{b}(i)$, i.e., the image of state i under $\mathbf{b}$. To be clear, note that the first row of $T_{\mathbf{b}}$ is formed of the cyclic distances of the images of states r, s such that $\left|r - s\right|_n = 1$; in general, the i-th row of $T_{\mathbf{b}}$ is formed of the cyclic distances of the images of pairs of states r, s of cyclic distance i. Notice that the columns are counted from 0 to $n-1$.

For $\mathbf{b} \in \mathcal{M}_n$ and $i = 1, \ldots, \lfloor \frac{n}{2} \rfloor$, let $R_i(\mathbf{b})$ denote the number of different entries in row i of $T_{\mathbf{b}}$:

$$R_i(\mathbf{b}) := \# \left\{ \left|b_0 - b_{(0+i)_n}\right|_n, \left|b_1 - b_{(1+i)_n}\right|_n, \ldots, \left|b_{n-1} - b_{i-1}\right|_n \right\}. \tag{1.2}$$

Set

$$\mathcal{E}_{\text{row}}(\alpha) := \bigcap_{i=1}^{\lfloor \frac{n}{2} \rfloor} \left\{ \mathbf{b} \in \mathcal{M}_n : \ R_i(\mathbf{b}) \ge \alpha \left\lfloor \frac{n}{2} \right\rfloor \right\}, \tag{1.3}$$

i.e., $\mathcal{E}_{\text{row}}(\alpha)$ contains those $\mathbf{b}$ for which every row of $T_{\mathbf{b}}$ has at least $\alpha \lfloor \frac{n}{2} \rfloor$ different elements. Its complement is

$$\mathcal{E}^c_{\text{row}}(\alpha) := \bigcup_{i=1}^{\lfloor \frac{n}{2} \rfloor} \left\{ \mathbf{b} \in \mathcal{M}_n : \ R_i(\mathbf{b}) < \alpha \left\lfloor \frac{n}{2} \right\rfloor \right\}. \tag{1.4}$$

We also define

$$\mathcal{E}_{\text{zero}}(\beta) := \left\{ \mathbf{b} \in \mathcal{M}_n : \ D(\mathbf{b}) \ge \beta \left\lfloor \frac{n}{2} \right\rfloor \right\}, \tag{1.5}$$

and its complement

$$\mathcal{E}^c_{\text{zero}}(\beta) := \left\{ \mathbf{b} \in \mathcal{M}_n : \ D(\mathbf{b}) < \beta \left\lfloor \frac{n}{2} \right\rfloor \right\}, \tag{1.6}$$

where

$$D_i(\mathbf{b}) := \begin{cases} 1, & \text{if there exist } k, l \in \mathbb{Z}_n \text{ such that } \left|k - l\right|_n = i \text{ and } \left|b_k - b_l\right|_n = 0; \\ 0, & \text{otherwise,} \end{cases}$$

and

$$D(\mathbf{b}) := \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} D_i(\mathbf{b}).$$

That is, $\mathcal{E}_{\text{zero}}(\beta)$ is the set of those $\mathbf{b}$ for which the matrix $T_{\mathbf{b}}$ has at least $\beta \lfloor \frac{n}{2} \rfloor$ rows containing the entry zero.

The proof of Theorem 4 relies on the following two lemmas.

Lemma 6. *Let $\varepsilon > 0$ and let $\alpha = 1 - e^{-1} - \varepsilon$. Then*

$$\mathbb{P}\left[\mathcal{E}^c_{\text{row}}(\alpha)\right] = O\left(\frac{1}{n}\right)$$

as $n \to \infty$.

Lemma 7. *Let $\varepsilon \in (0,1)$ and let $\beta = 1/2 - \varepsilon$. Then*

$$\mathbb{P}\left[\mathcal{E}^c_{\text{zero}}(\beta)\right] = O\left(\frac{1}{n}\right).$$

as $n \to \infty$.

Proof of Theorem 4. The main idea of the proof is to transform the question of synchronization of $\mathcal{A}_n(\mathbf{b})$ into a question concerning properties of the matrix $T_{\mathbf{b}}$. The entries $T_{\mathbf{b}}(i,j)$ of the matrix are random variables over $\mathcal{M}_n$, and to obtain our desired probability estimates we will need to understand the joint stochastic dependence structure of these random variables.

Let $\mathbf{b} \in \mathcal{M}_n$, and consider the associated Matrix $T_{\mathbf{b}}$. The first observation is that a zero in row i of $T_{\mathbf{b}}$ means that two states r, s with cyclic distance i synchronize under $\mathbf{b}$ (i.e., $\mathbf{b}_r = \mathbf{b}_s$), which implies that any pair r', s' with cyclic distance i can be synchronized with a word of the form $\mathbf{a}^l\mathbf{b}$ because $\{r', s'\}\mathbf{a}^l = \{r, s\}$ for some l. The second observation is that if the i-th row of $T_{\mathbf{b}}$ contains a number $j = |b_k - b_{(k+i)_n}|_n$ and the j-th row contains a zero, then every pair of states (r,s) with cyclic distance i can be synchronized with a word of the form $\mathbf{a}^{l_1}\mathbf{b}\mathbf{a}^{l_2}\mathbf{b}$. Indeed, we can proceed as follows: $\{r,s\} \xrightarrow{\mathbf{a}^{l_1}} \{k, (k+i)_n\} \xrightarrow{\mathbf{b}} \{b_k, b_{(k+i)_n}\}$, where this last pair has n-cyclic distance j; then $\{b_k, b_{(k+i)_n}\}$ synchronizes with a word of the form $\mathbf{a}^{l_2}\mathbf{b}$, for some l_2 because the j-th row contains a zero, thus we can synchronize $\{r,s\}$ with a word of the form $\mathbf{a}^{l_1}\mathbf{b}\mathbf{a}^{l_2}\mathbf{b}$. With these two observations, we establish sufficient conditions on $T_{\mathbf{b}}$ for the synchronization of $\mathcal{A}_n(\mathbf{b})$. The sets $\mathcal{E}_{\text{row}}(\alpha)$ and $\mathcal{E}_{\text{zero}}(\beta)$ which we defined in (1.3) and (1.5) play a crucial role.

Let $\mathbf{b} \in \mathcal{M}_n$. If $\mathbf{b}$ is contained in both $\mathcal{E}_{\text{row}}(\alpha)$ and $\mathcal{E}_{\text{zero}}(\beta)$ for some $\alpha, \beta > 0$ such that $\alpha + \beta > 1$, then $\mathcal{A}_n(\mathbf{b})$ synchronizes. This follows from the two previous observations together with the union bound. In fact, let (r,s) be any pair of different states. Set $i = |r - s|_n$. If row i contains a zero, we can synchronize $\{r,s\}$ with a word of the form $\mathbf{a}^l\mathbf{b}$; otherwise, row i contains an entry $j \neq 0$ such that row j contains a zero (because $\alpha + \beta > 1$), which implies that $\{r,s\}$ can be synchronized with a word of the form $\mathbf{a}^{l_1}\mathbf{b}\mathbf{a}^{l_2}\mathbf{b}$. Therefore, every pair of different states synchronizes, thus $\mathcal{A}_n(\mathbf{b})$ synchronizes by Claim 1. Therefore, for any $\alpha, \beta > 0$ satisfying $\alpha + \beta > 1$, we have the following bound:

$$\begin{aligned}\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n :\ \mathcal{A}_n(\mathbf{b}) \text{ synchronizes}\}\right] \geq \mathbb{P}\left[\mathcal{E}_{\text{row}}(\alpha) \cap \mathcal{E}_{\text{zero}}(\beta)\right] &= 1 - \mathbb{P}\left[\mathcal{E}^c_{\text{row}}(\alpha) \cup \mathcal{E}^c_{\text{zero}}(\beta)\right] \\ &\geq 1 - \mathbb{P}\left[\mathcal{E}^c_{\text{row}}(\alpha)\right] - \mathbb{P}\left[\mathcal{E}^c_{\text{zero}}(\beta)\right].\end{aligned} \tag{1.7}$$

Now, by this last inequality and by Lemmas 6 and 7 we obtain the bound stated in the main theorem. We choose $\varepsilon' = \frac{1}{20}$, $\alpha^\star = 1 - e^{-1} - \varepsilon' \approx 0.582$ and $\beta^\star = 1/2 - \varepsilon' = 0.45$, so that $\alpha^\star > 0$, $\beta^\star > 0$ and $\alpha^\star + \beta^\star > 1$. Then we have

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n :\ \mathcal{A}_n(\mathbf{b}) \text{ synchronizes}\}\right] \geq 1 - \underbrace{\mathbb{P}\left[\mathcal{E}^c_{\text{row}}(\alpha^\star)\right]}_{=O\left(\frac{1}{n}\right)} - \underbrace{\mathbb{P}\left[\mathcal{E}^c_{\text{zero}}(\beta^\star)\right]}_{=O\left(\frac{1}{n}\right)} = 1 - O\left(\frac{1}{n}\right)$$

as $n \to \infty$. ∎

1.3 Independence among the entries of $T_{\mathbf{b}}$

For every i and j in the range from 0 to $n-1$, the matrix entry $T_{\mathbf{b}}(i,j)$ assigns an integer value to every $\mathbf{b} \in \mathcal{M}_n$. In other words, for every such i and j, the function $T_{\mathbf{b}}(i,j): \mathcal{M}_n \mapsto \mathbb{Z}$ is an (integer-valued) random variable on the space $\mathcal{M}_n$, equipped with the uniform probability measure $\mathbb{P}$ (and with the power set of $\mathcal{M}_n$ as the natural sigma-field). It is crucial for our proof to give a criterion on pairs of indices $(i_1, j_1), \ldots, (i_k, j_k)$ which guarantees that the random variables $T_{\mathbf{b}}(i_1, j_1), \ldots, T_{\mathbf{b}}(i_k, j_k)$ are independent. First, notice that not every subset of entries of $T_{\mathbf{b}}$ is independent. For example,

$$T_{\mathbf{b}}(1,0) = \left|b_0 - b_1\right|_n, \quad T_{\mathbf{b}}(1,1) = \left|b_1 - b_2\right|_n, \quad T_{\mathbf{b}}(2,0) = \left|b_0 - b_2\right|_n$$

are clearly dependent: if the first two random variables $T_{\mathbf{b}}(1,0)$ and $T_{\mathbf{b}}(1,1)$ are zero, then $b_0 = b_1 = b_2$, which implies that $\left|b_0 - b_2\right|_n = 0$ and so $T_{\mathbf{b}}(2,0)$ necessarily also is zero. This dependence comes from the fact that there is a "cycle" of the form $b_0 \to b_1 \to b_2 \to b_0$ generated by the indices of these three random variables. Generally, it will turn out that a set of entries of $T_{\mathbf{b}}$ is independent if and only if the corresponding indices are "acyclic". We formalize this in the following

Definition. We call $\{j, (j+i)_n\}$ the *associated edge* of the matrix entry $T_{\mathbf{b}}(i,j)$. Let

$$S = \{(i_1, j_1), (i_2, j_2), \ldots, (i_k, j_k)\}$$

be a multi-set. The *associated (multi-)graph* $G(S)$ is the (multi-)graph with vertex set $\mathbb{Z}_n$ and edge (multi-)set

$$\Big\{\{j_1, (j_1+i_1)_n\},\ \{j_2, (j_2+i_2)_n\}, \ldots,\ \{(j_k, (j_k+i_k)_n\}\Big\}.$$

We call S acyclic if its associated multi-graph $G(S)$ is acyclic.

The relation between acyclic index sets and independent variables is stated in the following

Proposition 8. *The variables $T_{\mathbf{b}}(i_1, j_1), T_{\mathbf{b}}(i_2, j_2), \ldots, T_{\mathbf{b}}(i_k, j_k)$ are i.i.d. $\iff$ the (multi-)set $S = \{(i_1, j_1), (i_2, j_2), \ldots, (i_k, j_k)\}$ is acyclic. Furthermore, if the variables are independent, then*

$$\mathbb{P}\left[\bigcap_{w=1}^{k} \{\mathbf{b} \in \mathcal{M}_n : \ T_{\mathbf{b}}(i_w, j_w) = s_w\}\right] = \frac{\prod_{w=1}^{k} m_{s_w}}{n^k}, \quad \forall k \geq 1, \tag{1.8}$$

where $s_1, s_2, \ldots, s_w$ are arbitrary integers and

$$m_s = \#\{d \in \mathbb{Z}_n : \left|1-d\right|_n = s\} = \begin{cases} 2, & \textit{if } 0 < s < \frac{n}{2}; \\ 1, & \textit{if } s = 0; \\ 1, & \textit{if } s = \frac{n}{2} \textit{ and } \frac{n}{2} \in \mathbb{N}; \\ 0, & \textit{otherwise.} \end{cases}$$

Henceforth in this chapter we use the concepts "acyclic" and "independent" interchangeably when we refer to a multi-set of independent entries of $T_{\mathbf{b}}$, resp. to entries whose associated multi-graph is acyclic.

Remark 9. Note that different entries $T_{\mathbf{b}}(i,j)$, $T_{\mathbf{b}}(i',j')$ may be associated with the same edge; this only happens when n is even and $i = i' = n/2$ and $j \equiv j' \mod \frac{n}{2}$. Thus, for n odd, a pair of different entries of $T_{\mathbf{b}}$ is always acyclic/independent.

Remark 10. For a vector $\mathbf{b} \in \mathcal{M}_n$, we can write its entries $b_0, \dots, b_{n-1}$ as functions of $\mathbf{b}$. In other words, $b_0 = b_0(\mathbf{b}), \dots, b_{n-1} = b_{n-1}(\mathbf{b})$ are random variables on $\mathcal{M}_n$, equipped with the uniform measure $\mathbb{P}$. The random variables $b_0, \dots, b_{n-1}$ are independent and identically distributed over this space; this follows immediately from the fact that the uniform measure on $\mathcal{M}_n$ is a product of n one-dimensional uniform measures.

Proof of Proposition 8. First note that any two random variables $T_{\mathbf{b}}(i,j) = \left|b_j - b_{(j+i)_n}\right|_n$ and $T_{\mathbf{b}}(i',j') = \left|b_{j'} - b_{(j'+i')_n}\right|_n$ are always identically distributed. This follows from the fact that $b_0, b_1, \dots, b_{n-1}$ are i.i.d. (see Remark 10). Note also that for all s

$$\mathbb{P}\left[\left\{\mathbf{b} \in \mathcal{M}_n : \left|b_p - b_{p+q}\right|_n = s\right\}\right] = \frac{n \cdot m_s}{n^2} = \frac{m_s}{n},$$

which can seen by an easy counting argument: There are n different possible choices of b_p, and then there are m_s independent different choices of $b_{(p+q)_n}$ such that $\left|b_p - b_{p+q}\right|_n = s$. Thus equation (1.8) is just a rephrasing of the fact that the random variables are independent. Therefore, what we need to prove is that independence holds if and only if the associated (multi-)graph acyclic.

$\Rightarrow$) (by contraposition) Let $S = \{(i_1,j_1), (i_2,j_2), \dots, (i_k,j_k)\}$ be a (multi-)set which is not acyclic. Thus its associated multi-graph $G(S)$ has a cycle C of length $l \geq 2$. Let this cycle be w.l.o.g.

$$j_1 \to (j_1+i_1)_n = j_2 \to (j_2+i_2)_n = j_3 \to \dots \to (j_{l-1}+i_{l-1}) = j_l \to (j_l+i_l)_n = j_1.$$

Recall that $T_{\mathbf{b}}(i,j) = 0 \iff b_j = b_{(j+i)_n}$. Thus if for some $\mathbf{b} \in \mathcal{M}_n$ we have

$$T_{\mathbf{b}}(i_1,j_1) = T_{\mathbf{b}}(i_2,j_2) = \dots = T_{\mathbf{b}}(i_{l-1},j_{l-1}) = 0,$$

then $b_{j_1} = b_{j_2} = \dots = b_{j_l}$, and so we automatically also have $T_{\mathbf{b}}(i_l,j_l) = \left|b_{j_l} - b_{(j_l+i_l)_n}\right|_n = \left|b_{j_l} - b_{j_1}\right|_n = 0$. Thus the variables $T_{\mathbf{b}}(i_1,j_1), \dots, T_{\mathbf{b}}(i_\ell,j_\ell)$ are not independent. We conclude that an independent multi-set must be acyclic.

$\Leftarrow$) (by induction on k) Let $k \geq 2$. Assume that the multi-set $S_k = \{(i_1,j_1), (i_2,j_2), \dots, (i_k,j_k)\}$ is acyclic. We want to show that $T_{\mathbf{b}}(i_k,j_k)$ is independent of $T_{\mathbf{b}}(i_1,j_1), \dots, T_{\mathbf{b}}(i_{k-1},j_{k-1})$. This will allow us to factor out the k-th factor on the left-hand side of (1.8), leading (by induction) to the formula on the right-hand side of (1.8), which is equivalent to independence.

We distinguish two cases: The first case is when the edge $\{j_k, (j_k+i_k)_n\}$ is a connected component by itself in $G(S)$. This means that the sets

$$S_1 := \{j_1, (j_1+i_1)_n, j_2, (j_2+i_2)_n, \dots, j_k, (j_{k-1}+i_{k-1})_n\}$$

and $S_2 := \{j_k, (j_k+i_k)_n\}$ are disjoint. By construction the random variables

$$T_{\mathbf{b}}(i_1,j_1), \dots, T_{\mathbf{b}}(i_{k-1},j_{k-1})$$

depend only on b_s with $s \in S_1$, while $T_{\mathbf{b}}(i_k,j_k)$ depends only on b_s with $s \in S_2$. Since $b_0, \dots, b_{n-1}$ are independent by Remark 10, this implies that $T_{\mathbf{b}}(i_k,j_k)$ is independent of $T_{\mathbf{b}}(i_1,j_1), \dots, T_{\mathbf{b}}(i_{k-1},j_{k-1})$, as desired.

For the second case, the edge $\{j_k, (j_k+i_k)_n\}$ is not a connected component by itself in $G(S)$. Since it is also not part of a cycle by assumption,we can assume that $(j_k+i_k)_n$ is a leaf vertex in $G(S)$. In principle, $T_{\mathbf{b}}(i_k,j_k)$ depends on b_{j_k} as well as on $b_{(j_k+i_k)_n}$. However, since $T_{\mathbf{b}}(i_k,j_k)$ is defined as a cyclic distance, the conditional distribution of $T_{\mathbf{b}}(i_k,j_k)$ given b_{j_k} is always the same. In formulas, for every s_k we have

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n : T_{\mathbf{b}}(i_k,j_k) = s_k\}\right] = \mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n : T_{\mathbf{b}}(i_k,j_k) = s_k \text{ and } b_{j_k} = r\}\right] \tag{1.9}$$

for every $r \in \{0,\dots,n-1\}$. This fact can be simply established by counting the possible configurations of b_{j_k} and $b_{(j_k+i_k)_n}$. By definition, $T_{\mathbf{b}}(i_k,j_k)$ is independent of all b_ℓ with $\ell \neq j_k, (j_k+i_k)_n$. Thus for every numbers $s_1,\dots,s_k$ we have, using the independence of $b_0,\dots,b_{n-1}$ and (1.9), that

$$
\begin{aligned}
&\mathbb{P}\left[\bigcap_{w=1}^{k}\{\mathbf{b}:\ T_{\mathbf{b}}(i_w,j_w)=s_w\}\right]\\
&=\sum_{r=0}^{n-1}\mathbb{P}\left[\bigcap_{w=1}^{k}\{\mathbf{b}:\ T_{\mathbf{b}}(i_w,j_w)=s_w \text{ and } b_{j_k}=r\}\right]\\
&=\sum_{r=0}^{n-1}\mathbb{P}\left[\underbrace{\left(\bigcap_{w=1}^{k-1}\{\mathbf{b}:\ T_{\mathbf{b}}(i_w,j_w)=s_w \text{ and } b_{j_k}=r\}\right)}_{\text{depends only on } b_\ell \text{ with } \ell\neq j_k,(j_k+i_k)_n \text{ when } b_{j_k} \text{ is fixed}}\cap\underbrace{\{\mathbf{b}:\ T_{\mathbf{b}}(i_k,j_k)=s_k \text{ and } b_{j_k}=r\}}_{\text{depends only on } b_{(j_k+i_k)_n} \text{ when } b_{j_k} \text{ is fixed}}\right]\\
&=\sum_{r=0}^{n-1}\left(\mathbb{P}\left[\left(\bigcap_{w=1}^{k-1}\{\mathbf{b}:\ T_{\mathbf{b}}(i_w,j_w)=s_w \text{ and } b_{j_k}=r\}\right)\right]\mathbb{P}\left[\{\mathbf{b}:\ T_{\mathbf{b}}(i_k,j_k)=s_k \text{ and } b_{j_k}=r\}\right]\right)\\
&=\sum_{r=0}^{n-1}\left(\mathbb{P}\left[\left(\bigcap_{w=1}^{k-1}\{\mathbf{b}:\ T_{\mathbf{b}}(i_w,j_w)=s_w \text{ and } b_{j_k}=r\}\right)\right]\mathbb{P}\left[\{\mathbf{b}:\ T_{\mathbf{b}}(i_k,j_k)=s_k\}\right]\right)\\
&=\mathbb{P}\left[\left(\bigcap_{w=1}^{k-1}\{\mathbf{b}:\ T_{\mathbf{b}}(i_w,j_w)=s_w\}\right)\right]\mathbb{P}\left[\{\mathbf{b}:\ T_{\mathbf{b}}(i_k,j_k)=s_k\}\right].
\end{aligned}
$$

This is exactly the independence property that we wanted to establish. □

1.4 Proof of Lemma 6

The overview of the proof is as follows. Recall that we understand the entries of the matrix $T_{\mathbf{b}}$ as random variables. We will prove that every row of $T_{\mathbf{b}}$ contains a "large" number of independent random variables. Then we give a lower bound for the expected value of the number of different elements in each row. Then we apply McDiarmid's inequality to each row and finally we use the union bound together with the exponential decay delivered by McDiarmid's inequality to guarantee that w.h.p. every row of $T_{\mathbf{b}}$ has at least $\sim (1-e^{-1})\left\lfloor\frac{n}{2}\right\rfloor$ different elements. We denote by $C_n(i)$ the *circulant graph* on n vertices, i.e., the graph with vertex set $\mathbb{Z}_n$ where two vertices r,s are adjacent if $\left|r-s\right|_n=i$.

We need the following property.

Claim 11. *For every i, the i-th row of $T_{\mathbf{b}}$ contains a set of at least $n-\gcd(n,i)$ random variables which are i.i.d.*

Proof. The variables in row i are given by the multi-set

$$E_i(\mathbf{b}) := \{\left|b_0-b_{(0+i)_n}\right|_n,\dots,\left|b_s-b_{(s+i)_n}\right|_n,\dots,\left|b_{n-1}-b_{i-1}\right|_n\}. \tag{1.10}$$

Let $i\neq\frac{n}{2}$. By Remark 9 the multi-set $E_i(\mathbf{b})$ does not have repeated elements and the associated multi-graph $G(E_i(\mathbf{b}))$ is isomorphic to the circulant graph $C_n(i)$. It is well know and easy to show that $C_n(i)$ is a disjoint union of $\gcd(n,i)$ cycles of length $\frac{n}{\gcd(n,i)}$ [Boesch and Tindell, 1984]. We obtain an acyclic set of variables by removing variables that correspond to exactly one edge of each cycle in $G(E_i(\mathbf{b}))$. The resulting set of variables is i.i.d. by Proposition 8. In the case $i=\frac{n}{2}$, the first $\frac{n}{2}$ variables in row $\frac{n}{2}$

$$E_{\frac{n}{2}}(\mathbf{b}) = \{\left|b_0-b_{\frac{n}{2}}\right|_n,\dots,\left|b_s-b_{(s+\frac{n}{2})_n}\right|_n,\dots,\left|b_{\frac{n}{2}-1}-b_{n-1}\right|_n\}$$

have an associated multi-graph that is isomorphic to the circulant graph $C_n(\frac{n}{2})$, which is a disjoint union of $\frac{n}{2} = \gcd(n, \frac{n}{2})$ edges. This last graph is acyclic, thus the variables are i.i.d. by Proposition 8. $\square$

We prove the following lower bound

Claim 12. *We have* $\mathbb{E}[R_i] \geq \lfloor \frac{n}{2} \rfloor (1 - e^{-1}) - 1$, *where for all* $\mathbf{b} \in \mathcal{M}_n$

$$R_i(\mathbf{b}) = \#\{ |b_0 - b_{(0+i)_n}|_n, \ldots, |b_s - b_{(s+i)_n}|_n, \ldots, |b_{n-1} - b_{i-1}|_n \}$$

(see (1.2)) is the cardinality of different elements in row i of $T_{\mathbf{b}}$.

Proof. First, for every $d \in \{0, \ldots, \lfloor \frac{n}{2} \rfloor\}$, we define the random variables

$$\delta_j^{(i)}(\mathbf{b}, d) := 1 - \mathbb{1}\{|b_j - b_{(j+i)_n}|_n = d\} = \begin{cases} 0, & \text{if } |b_j - b_{(j+i)_n}|_n = d; \\ 1, & \text{otherwise.} \end{cases}$$

and

$$r_d^{(i)}(\mathbf{b}) := \prod_{j \in \mathbb{Z}_n} \delta_j^{(i)}(\mathbf{b}, d) = \begin{cases} 0 & \text{if } \exists\, p, q \in \mathbb{Z}_n \text{ such that } d_n(p,q) = i \text{ and } d(\mathbf{b}_p, \mathbf{b}_q) = d; \\ 1, & \text{otherwise.} \end{cases}$$

Note that $r_d^{(i)}(\mathbf{b})$ is zero if the number d is included in the i-th row of $T_{\mathbf{b}}$, and that it is one otherwise. Recalling that the entries of $T_{\mathbf{b}}$ can only have values in $\{0, 1, \ldots, \lfloor \frac{n}{2} \rfloor\}$, we write the number of distinct elements in row i as

$$R_i(\mathbf{b}) = \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) - \sum_{d=0}^{\lfloor \frac{n}{2} \rfloor} r_d^{(i)}. \tag{1.11}$$

By Claim 11 there is a subset I of $\mathbb{Z}_n$ of cardinality $n - \gcd(n, i)$ such that the variables $\{\delta_w^{(i)} : w \in I\}$ are i.i.d., and thus

$$\mathbb{E}\left[r_d^{(i)} \right] = \mathbb{E}\left[\prod_{j \in \mathbb{Z}_n} \delta_j^{(i)}(\mathbf{b}, d) \right] \leq \mathbb{E}\left[\prod_{w \in I} \delta_w^{(i)}(\mathbf{b}, d) \right] = \mathbb{E}\left[\delta_0^{(i)}(\mathbf{b}, d) \right]^{n - \gcd(n,i)}.$$

Furthermore, by Proposition 8 we have $\mathbb{E}\left[\delta_0^{(i)}(\mathbf{b}, d) \right] = 1 - \frac{m_d}{n}$, and thus

$$\mathbb{E}\left[r_d^{(i)} \right] \leq \left(1 - \frac{m_d}{n} \right)^{n - \gcd(n,i)} \leq \left(1 - \frac{m_d}{n} \right)^{\frac{n}{2}} = \begin{cases} \left(1 - \frac{2}{n}\right)^{\frac{n}{2}}, & \text{if } d \neq 0, \frac{n}{2}; \\ \left(1 - \frac{1}{n}\right)^{\frac{n}{2}}, & \text{otherwise}, \end{cases}$$

for $d \in \{0, 1, \ldots, \lfloor \frac{n}{2} \rfloor\}$. Using the inequality $1 - x \leq e^{-x}$, which is valid for any real number x, we obtain

$$\mathbb{E}\left[\sum_{d=0}^{\lfloor \frac{n}{2} \rfloor} r_d^{(i)} \right] \leq \left\lfloor \frac{n}{2} \right\rfloor \underbrace{\left(1 - \frac{2}{n} \right)^{\frac{n}{2}}}_{\leq e^{-1}} + 2 \underbrace{\left(1 - \frac{1}{n} \right)^{\frac{n}{2}}}_{\leq e^{-1/2}} \leq \left\lfloor \frac{n}{2} \right\rfloor e^{-1} + 2.$$

Plugging this inequality into (1.11) yields

$$\mathbb{E}[R_i] = \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) - \mathbb{E}\left[\sum_{d=0}^{\lfloor \frac{n}{2} \rfloor} r_d^{(i)} \right] \geq \left\lfloor \frac{n}{2} \right\rfloor (1 - e^{-1}) - 1.$$

This proves Claim 12. $\square$

We introduce McDiarmid's inequality to prove Claim 14.

Definition. Let $L : (\mathbb{Z}_n)^n \to \mathbb{R}$ be a function. We say that L has *Lipschitz coefficient* $r \in \mathbb{R}^+$ if

$$|L(\overrightarrow{v}) - L(\overrightarrow{w})| \leq r$$

for every $\overrightarrow{v}, \overrightarrow{w} \in (\mathbb{Z}_n)^n$ such that $\overrightarrow{v}(j) = \overrightarrow{w}(j)$ for all j except for at most one index.

Proposition 13 (**McDiarmid's Inequality** [McDiarmid, 1989]). *Let $\bar{X} := (X_1, X_2, \ldots, X_n) \in (\mathbb{Z}_n)^n$ be a random vector where the variables $X_1, X_2, \ldots, X_n$ are independent and let $L : (\mathbb{Z}_n)^n \to \mathbb{R}$ be a function with bounded Lipschitz coefficient r. Then*

$$\text{(lower tail)} \qquad \mathbb{P}\left[L(\bar{X}) \leq \mathbb{E}\left[L(\bar{X})\right] - r\sqrt{\lambda n}\right] \leq e^{-2\lambda},$$

for all $\lambda \geq 0$.

Remark. This is just a special case of the general form of McDiarmid's inequality. The general inequality also bounds the upper tail, and allows different Lipschitz coefficients in the respective components.

In the following claim we use Proposition 13 to estimate the probability that row i of $T_{\mathbf{b}}$ has less than $\sim (1 - e^{-1}) \left\lfloor \frac{n}{2} \right\rfloor$ different elements.

Claim 14. *Let $\varepsilon > 0$. Then*

$$\mathbb{P}\left[\mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) < \left\lfloor \frac{n}{2} \right\rfloor (1 - e^{-1} - \varepsilon)\right] \leq e^{-\Theta(n)},$$

for $i = 1, 2, \ldots, \left\lfloor \frac{n}{2} \right\rfloor$.

Proof. Let $\mathbf{b} = (b_0, b_1, \ldots, b_{n-1})$. Let $E_i(\mathbf{b})$ be defined as in (1.10). The function $R_i(\mathbf{b}) := \# E_i(\mathbf{b})$ has Lipschitz coefficient 2: changing one b_j affects at most two entries, namely $\left|b_j - b_{(j+i)_n}\right|_n$ and $\left|b_{(j-i)_n} - b_j\right|_n$. Using McDiarmid's inequality, we deduce that

$$\mathbb{P}\left[\mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) \leq \mathbb{E}\left[R_i\right] - 2\sqrt{\lambda n}\right] \leq e^{-2\lambda}, \quad \forall \lambda \geq 0.$$

Using the lower bound $\mathbb{E}\left[R_i\right] \geq \left\lfloor \frac{n}{2} \right\rfloor (1 - e^{-1}) - 1$ of Claim 12 we obtain

$$\begin{aligned}
& \mathbb{P}\left[\mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) < \left(\left\lfloor \frac{n}{2} \right\rfloor (1 - e^{-1}) - 1\right) - 2\sqrt{\lambda n}\right] \\
\leq \; & \mathbb{P}\left[\mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) \leq \mathbb{E}\left[R_i\right] - 2\sqrt{\lambda n}\right] \\
\leq \; & e^{-2\lambda}, \quad \forall \lambda \geq 0.
\end{aligned}$$

Let $\varepsilon > 0$ be arbitrary but fixed, and let

$$\lambda_\varepsilon(n) := \frac{1}{4n}\left(\varepsilon \left\lfloor \frac{n}{2} \right\rfloor - 1\right)^2 = \Theta(n); \tag{1.12}$$

we observe that $\lambda_\varepsilon(n)$ is independent of i. Let $n > \frac{2}{\varepsilon}$, then plugging $\lambda = \lambda_\varepsilon(n)$ into the previous inequality yields

$$\mathbb{P}\left[\mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) < \left\lfloor \frac{n}{2} \right\rfloor (1 - e^{-1} - \varepsilon)\right] \leq e^{-2\lambda_\varepsilon(n)} = e^{-\Theta(n)}. \tag{1.13}$$

□

Recall that $\mathcal{E}_{\text{row}}(\alpha)$ contains those $\mathbf{b} \in \mathcal{M}_n$ for which every row of $T_{\mathbf{b}}$ has at least $\alpha \left\lfloor \frac{n}{2} \right\rfloor$ different elements, so that

$$\mathcal{E}^c_{\text{row}}(\alpha) = \bigcup_{i=0}^{\lfloor \frac{n}{2} \rfloor} \left\{ \mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) < \alpha \left\lfloor \frac{n}{2} \right\rfloor \right\}.$$

Let $\varepsilon > 0$ be arbitrary and let $\alpha^* = 1 - e^{-1} - \varepsilon$. Then

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}^c_{\text{row}}(\alpha^\star)\right] &= \mathbb{P}\left[\bigcup_{i=1}^{\lfloor \frac{n}{2} \rfloor} \left\{ \mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) < \alpha^\star \left\lfloor \frac{n}{2} \right\rfloor \right\} \right] \\
&\leq \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \mathbb{P}\left[\mathbf{b} \in \mathcal{M}_n : \; R_i(\mathbf{b}) < \alpha^\star \left\lfloor \frac{n}{2} \right\rfloor \right] \\
&\leq n e^{-\Theta(n)}, \qquad (1.14)
\end{aligned}$$

where we use Claim 14 for the second inequality. This proof of Lemma 6 follows by noticing that

$$n e^{-\Theta(n)} = O\left(\frac{1}{n}\right).$$

1.5 Proof of Lemma 7

The overview of the proof is as follows. We will define two random variables $\mathcal{Z}_0(\mathbf{b})$ and $\mathcal{Z}_1(\mathbf{b})$ such that

- $D(\mathbf{b}) \geq \mathcal{Z}_0(\mathbf{b}) - \mathcal{Z}_1(\mathbf{b}), \quad \forall\, \mathbf{b} : \mathbb{Z}_n \to \mathbb{Z}_n;$
- $\mathbb{E}\left[\mathcal{Z}_0 - \mathcal{Z}_1\right] \sim \frac{n}{2}.$

Then we will show that $\mathcal{Z}_0$ and $\mathcal{Z}_1$ concentrate around their respective means, and use this fact to give an upper bound on the probability that D is small. For this purpose, we note the following property.

Claim 15. *Let $\mathcal{Z}_0, \mathcal{Z}_1$ and D be random variables which take non-negative values, such that $D \geq \mathcal{Z}_0 - \mathcal{Z}_1$. Let $\nu > 0$ and let $\delta \leq \mathbb{E}\left[\mathcal{Z}_0 - \mathcal{Z}_1\right] - 2\nu$. Then*

$$\mathbb{P}\left[D < \delta\right] \leq \mathbb{P}\left[\mathcal{Z}_0 < \mathbb{E}\left[\mathcal{Z}_0\right] - \nu\right] + \mathbb{P}\left[\mathcal{Z}_1 > \mathbb{E}\left[\mathcal{Z}_1\right] + \nu\right].$$

Proof. This follows easily from the assumption that $\mathcal{Z}_0 - \mathcal{Z}_1 \leq D$ and the union bound. □

To prove concentration of $\mathcal{Z}_0$ and $\mathcal{Z}_1$ around their respective means, we use Chebyschev's inequality. Notice that $D : \mathbb{Z}_n^n \to \mathbb{Z}_n$ does not have a bounded Lipschitz coefficient, so we cannot use McDiarmid's inequality to guarantee its concentration.

1.5.1 Lower bound for $D(b)$

Recall that $D(\mathbf{b})$ counts the number of rows of $T_{\mathbf{b}}$ that contain at least one zero. Let

$$z_i = z_i(\mathbf{b}) := \#(\text{Zeros in row } i \text{ of } T_{\mathbf{b}})$$

and

$$\mathcal{Z}_0(\mathbf{b}) := \#(\text{Zeros in } T_{\mathbf{b}}) = \sum_{(i,j)\in[1,\lfloor\frac{n}{2}\rfloor]\times[0,n-1]} \mathbb{1}\left\{T_{\mathbf{b}}(i,j) = 0\right\}.$$

Then

$$D(\mathbf{b}) = \mathcal{Z}_0(\mathbf{b}) - \sum_{i=1}^{\lfloor\frac{n}{2}\rfloor} \max(z_i - 1, 0). \tag{1.15}$$

It is easy to verify that the number of non-ordered pairs of entries in the i-th row with zero value is

$$\sum_{0\le j<j'\le n-1} \mathbb{1}\left\{T_{\mathbf{b}}(i,j) = 0\right\} \mathbb{1}\left\{T_{\mathbf{b}}(i,j') = 0\right\} = \frac{z_i(z_i-1)}{2} \ge \max(z_i - 1, 0), \quad \forall i,$$

therefore

$$\mathcal{Z}_1(\mathbf{b}) := \sum_{i=1}^{\lfloor\frac{n}{2}\rfloor} \sum_{0\le j<j'\le n-1} \mathbb{1}\left\{T_{\mathbf{b}}(i,j) = 0\right\} \mathbb{1}\left\{T_{\mathbf{b}}(i,j') = 0\right\} \ge \sum_{i=1}^{\lfloor\frac{n}{2}\rfloor} \max(z_i - 1, 0).$$

From this and (1.15), we conclude that

Claim 16. *$D(\mathbf{b}) \ge \mathcal{Z}_0(\mathbf{b}) - \mathcal{Z}_1(\mathbf{b}), \quad \forall\, \mathbf{b} : \mathbb{Z}_n \to \mathbb{Z}_n.$*

1.5.2 Estimates for $\mathbb{E}[\mathcal{Z}_0]$, $\mathbb{E}[\mathcal{Z}_1]$, $\mathbb{E}[\mathcal{Z}_0 - \mathcal{Z}_1]$, $\mathbb{V}[\mathcal{Z}_0]$, $\mathbb{V}[\mathcal{Z}_1]$

In this subsection we prove that

- $\mathbb{E}[\mathcal{Z}_0 - \mathcal{Z}_1] \sim \frac{n}{2}$,
- $\mathbb{E}[\mathcal{Z}_0] = \Theta(n)$,
- $\mathbb{E}[\mathcal{Z}_1] = \Theta(n)$,
- $\mathbb{V}[\mathcal{Z}_0] = O(n)$, and
- $\mathbb{V}[\mathcal{Z}_1] = O(n)$.

For the rest of this subsection, we use the notation

$$y_{i,j} = y_{i,j}(\mathbf{b}) := \mathbb{1}\left\{T_{\mathbf{b}}(i,j) = 0\right\},$$

for $1 \le i \le \lfloor\frac{n}{2}\rfloor$ and $0 \le j \le n-1$.

Definition. The variables $y_{i_1,j_1}, y_{i_2,j_2} \dots, y_{i_k,j_k}$ are called *acyclic* if the multi-set $\bigcup_{w=1}^{k}\{(i_w, j_w)\}$ is acyclic. Let

$$G\left(\{y_{i_1,j_1}, y_{i_2,j_2} \dots, y_{i_k,j_k}\}\right) = G\left(\bigcup_{w=1}^{k}\{(i_w, j_w)\}\right)$$

be the *associated multi-graph* of the multi-set $\{y_{i_1,j_1}, y_{i_2,j_2} \dots, y_{i_k,j_k}\}$ and let $e(y_{i,j}) := \{j, (j+i)_n\}$ be the *associated edge* to $y_{i,j}$. The *length* of $e(y_{i,j})$ is $\left|j - (j+i)_n\right|_n = i$.

Remark 17. If the variables $y_{i_1,j_1}, y_{i_2,j_2} \dots, y_{i_k,j_k}$ are acyclic then they are i.i.d.; this is an immediate consequence of Proposition 8.

We begin with the easy part: the bounds for the expected values.

Claim 18. *Let $n \in \mathbb{N}$. We have* $\mathbb{E}[\mathcal{Z}_0] = \Theta(n)$, $\mathbb{E}[\mathcal{Z}_1] = \Theta(n)$, *and* $\mathbb{E}[\mathcal{Z}_0 - \mathcal{Z}_1] \geq 0.5 \left\lfloor \frac{n}{2} \right\rfloor - 1$.

Proof. Using the linearity of the expectation, we get that

$$\mathbb{E}[\mathcal{Z}_0] = \sum_{(i,j)\in[1,\left\lfloor \frac{n}{2} \right\rfloor]\times[0,n-1]} \mathbb{E}[y_{i,j}] = \left\lfloor \frac{n}{2} \right\rfloor n \frac{1}{n} = \left\lfloor \frac{n}{2} \right\rfloor = \Theta(n), \tag{1.16}$$

where for the second equality we use that

$$\mathbb{E}[y_{i,j}] = \mathbb{P}[\{\mathbf{b} : \ T_{\mathbf{b}}(i,j) = 0\}] = \mathbb{P}\left[\{\mathbf{b} : \ b_j = b_{(j+i)}\}\right] = \frac{1}{n}. \tag{1.17}$$

Now we calculate an upper bound for $\mathbb{E}[\mathcal{Z}_1]$, depending on the parity of n.

Case 1: n odd. Every product $y_{i,j}y_{i,j'}$ in the sum

$$\mathcal{Z}_1 = \sum_{i=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{0\leq j<j'\leq n-1} y_{i,j}y_{i,j'}$$

is formed of independent random variables $y_{i,j}$, $y_{i,j'}$ by Remarks 9,17. Thus

$$\begin{aligned}\mathbb{E}[\mathcal{Z}_1] = \sum_{i=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{0\leq j<j'\leq n-1} \mathbb{E}[y_{i,j}y_{i,j'}] &= \sum_{i=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{0\leq j<j'\leq n-1} \mathbb{E}[y_{i,j}]\,\mathbb{E}[y_{i,j'}] \\ &\overset{(1.17)}{=} \left\lfloor \frac{n}{2} \right\rfloor \binom{n}{2} \frac{1}{n^2} \\ &= \underbrace{\frac{1}{2}\left\lfloor \frac{n}{2} \right\rfloor \left(1 - \frac{1}{n}\right)}_{\leq \frac{1}{2}\left\lfloor \frac{n}{2} \right\rfloor} = \Theta(n).\end{aligned}$$

Case 2: n even. Using Remark 9, we write $\mathcal{Z}_1$ as

$$\mathcal{Z}_1 = \sum_{\substack{1\leq i<\frac{n}{2} \\ 0\leq j<j'\leq n-1}} y_{i,j}y_{i,j'} + \sum_{\substack{0\leq r<r'\leq n-1 \\ r\not\equiv r' \pmod{n/2}}} y_{n/2,r}y_{n/2,r'} + \sum_{s=0}^{\frac{n}{2}-1} y_{n/2,s}.$$

Every product $y_{i,j}y_{i,j'}$ in the first sum is formed of independent variables $y_{i,j}$, $y_{i,j'}$ by Remark 9 and the same is valid for the products $y_{\frac{n}{2},r}y_{\frac{n}{2},r'}$ in the second sum, therefore

$$\begin{aligned}\mathbb{E}[\mathcal{Z}_1] &= \sum_{i=1}^{\frac{n}{2}-1} \sum_{0\leq j<j'\leq n-1} \mathbb{E}[y_{i,j}]\,\mathbb{E}[y_{i,j'}] + \sum_{\substack{0\leq r<r'\leq n-1 \\ r\not\equiv r' \pmod{n/2}}} \mathbb{E}\left[y_{n/2,r}\right]\mathbb{E}\left[y_{n/2,r'}\right] + \sum_{s=0}^{\frac{n}{2}-1} \mathbb{E}\left[y_{n/2,s}\right] \\ &= \sum_{i=1}^{\frac{n}{2}-1} \sum_{0\leq j<j'\leq n-1} \frac{1}{n^2} + \sum_{\substack{0\leq r<r'\leq n-1 \\ r\not\equiv r' \pmod{n/2}}} \frac{1}{n^2} + \sum_{s=0}^{\frac{n}{2}-1} \frac{1}{n} \\ &= \left(\frac{n}{2} - 1\right) \cdot \binom{n}{2} \cdot \frac{1}{n^2} + \left(\binom{n}{2} - \frac{n}{2}\right) \cdot \frac{1}{n^2} + \frac{n}{2} \cdot \frac{1}{n} \\ &= \underbrace{\frac{1}{2} \cdot \frac{n}{2} \cdot \left(1 - \frac{1}{n} + \left(\frac{2}{n} - \frac{2}{n^2}\right)\right)}_{\leq \frac{1}{2}\cdot\frac{n}{2}+1} = \Theta(n).\end{aligned}$$

We deduce from the previous cases that $\mathbb{E}[\mathcal{Z}_1] = \Theta(n)$ and $\mathbb{E}[\mathcal{Z}_1] \leq 0.5 \left\lfloor \frac{n}{2} \right\rfloor + 1$ for all n. Using this last inequality and (1.16), we conclude that

$$\mathbb{E}[\mathcal{Z}_0] - \mathbb{E}[\mathcal{Z}_1] = \left\lfloor \frac{n}{2} \right\rfloor - \mathbb{E}[\mathcal{Z}_1] \geq 0.5 \left\lfloor \frac{n}{2} \right\rfloor - 1.$$

This concludes the proof of Claim 18. □

Now we estimate the variance of $\mathcal{Z}_0$ and $\mathcal{Z}_1$.

Claim 19. *Let $n \in \mathbb{N}$, then $\mathbb{V}[\mathcal{Z}_0] = O(n)$ and $\mathbb{V}[\mathcal{Z}_1] = O(n)$.*

Proof. Here we also divide the calculations according to the parity of n.

Case 1: n odd. We expand the variance of $\mathcal{Z}_0$ to get that

$$\mathbb{V}[\mathcal{Z}_0] = \sum_{\substack{1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor \\ 0 \leq j \leq n-1}} \mathbb{V}[y_{i,j}] + \sum_{\substack{1 \leq i,i' \leq \left\lfloor \frac{n}{2} \right\rfloor \\ 0 \leq j,j' \leq n-1 \\ (i,j) \neq (i',j')}} \operatorname{Cov}[y_{i,j}, y_{i',j'}],$$

where the covariances are calculated among pairs of independent variables $y_{i,j}, y_{i',j'}$ due to Remark 9. Thus

$$\mathbb{V}[\mathcal{Z}_0] = \sum_{\substack{1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor \\ 0 \leq j \leq n-1}} \mathbb{V}[y_{i,j}].$$

We notice that $y_{i,j}^2 = y_{i,j}$ because $y_{i,j} \in \{0,1\}$, therefore

$$\mathbb{V}[y_{i,j}] = \mathbb{E}\left[y_{i,j}^2\right] - \mathbb{E}[y_{i,j}]^2 = \frac{1}{n} - \frac{1}{n^2}, \quad \forall n \in \mathbb{N}, \tag{1.18}$$

where we use (1.17) in the last equality. Then, for all n odd, we get that

$$\mathbb{V}[\mathcal{Z}_0] = \left\lfloor \frac{n}{2} \right\rfloor n \left(\frac{1}{n} - \frac{1}{n^2} \right) = \left\lfloor \frac{n}{2} \right\rfloor \left(1 - \frac{1}{n} \right) = O(n). \tag{1.19}$$

Now we calculate

$$\mathbb{V}[\mathcal{Z}_1] = \sum_{\substack{1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor \\ 0 \leq j < j' \leq n-1}} \mathbb{V}[y_{i,j} y_{i,j'}] + \sum_{\substack{1 \leq i,r \leq \left\lfloor \frac{n}{2} \right\rfloor \\ 0 \leq j,j',s,s' \leq n-1 \\ j<j';\, s<s' \\ (i,j,j') \neq (r,s,s')}} \operatorname{Cov}[y_{i,j} y_{i,j'}, y_{r,s} y_{r,s'}]; \tag{1.20}$$

We first note that

$$\mathbb{V}[y_{i,j} y_{i,j'}] = \mathbb{E}\left[y_{i,j}^2 y_{i,j'}^2\right] - \mathbb{E}[y_{i,j} y_{i,j'}]^2 = \frac{1}{n^2} - \frac{1}{n^4}, \quad \text{for } n \text{ odd and } \forall i \text{ and } j \neq j'; \tag{1.21}$$

this follows since the variables $y_{i,j}$ and $y_{i,j'}$ are different and therefore independent (see Remark 9). Thus

$$\sum_{\substack{1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor \\ 0 \leq j < j' \leq n-1}} \mathbb{V}[y_{i,j} y_{i,j'}] = \left\lfloor \frac{n}{2} \right\rfloor \binom{n}{2} \frac{1}{n^2} \left(1 - \frac{1}{n^2} \right) = O(n). \tag{1.22}$$

For the sum of the covariances, we proceed as follows: if the variables $y_{i,j}, y_{i,j'}, y_{r,s}, y_{r,s'}$ are acyclic then they are independent (see Proposition 8), therefore

$$\operatorname{Cov}[y_{i,j} y_{i,j'}, y_{r,s} y_{r,s'}] = 0.$$

On the other hand, if the variables $y_{i,j}, y_{i,j'}, y_{r,s}, y_{r,s'}$ are not acyclic, let

$$\mathcal{Y} := \{\{y_{i,j}, y_{i,j'}, y_{r,s}, y_{r,s'}\} : (i,j,j') \neq (r,s,s'),\ j < j',\ s < s'\},$$

and let

$$Y = \{y_{i,j}, y_{i,j'}, y_{r,s}, y_{r,s'}\} \in \mathcal{Y}.$$

Then $G(Y)$ is a multi-graph with four edges $e(y_{i,j}), e(y_{i,j'}), e(y_{r,s}), e(y_{r,s'})$ such that $e(y_{i,j}) \neq e(y_{i,j'})$ and $e(y_{r,s}) \neq e(y_{r,s'})$ (see Remark 9). In particular, there cannot be 3 equal edges. If $G(Y)$ has at least one cycle, it is isomorphic to one of the multi-graphs in Figure 1.2 below.

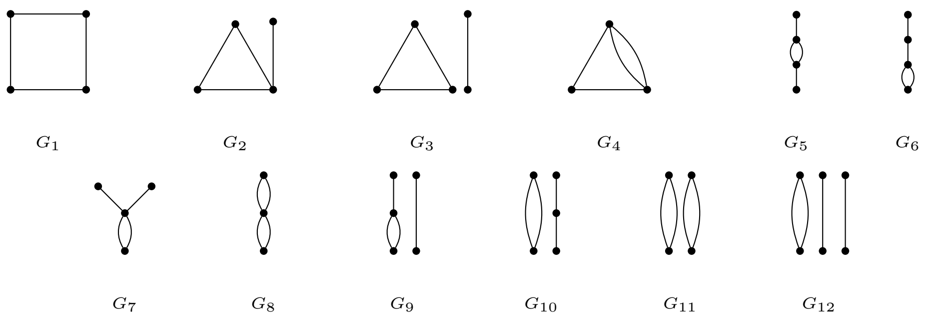


Figure 1.2: Possible non-acyclic multi-graphs for n odd.

We will now estimate the contribution of each of these possible non-acyclic multi-graphs.

Claim 20. *Let $n \in \mathbb{N}$, then*

$$\#\{Y \in \mathcal{Y} : G(Y) \cong G_c\} = \begin{cases} O(n^4), & \textit{if } c = 1,2,3,5,6,7,12; \\ O(n^3), & \textit{if } c = 4,8,9,10,11. \end{cases}$$

Proof. The cases $c = 1,2,5,6,7$ can be bounded by the trivial bound $O(n^4)$, and the same for the cases $c = 4,8$ with the bound $O(n^3)$. The remaining cases $c = 3,9,10,11,12$ require better estimates than their respective trivial bounds.

First, notice that for all cases, the four edges of the multi-graph $G(\{y_{i,j}, y_{i,j'}, y_{r,s}, y_{r,s'}\})$ are divided into two pairs: $e(y_{i,j}), e(y_{i,j'})$ of length i and $e(y_{r,s}), e(y_{r,s'})$ of length r. The case G_3 is bounded by $\binom{n}{3} * 2n = O(n^4)$ because three vertices can be chosen freely to form a triangle whose edges have at most two different lengths i, r, then we choose a vertex v for the free edge and finally we choose v' such that $|v - v'|_n = i$ or $|v - v'|_n = r$ depending on the lengths of the edges in the triangle, therefore v' has only two choices.

The case G_{12} is also bounded by $O(n^4)$. To show this, we distinguish between two subcases. In the first subcase, the multi-edge is formed of the associated edges of the same pair, w.l.o.g. $e(y_{i,j}) = e(y_{i,j'})$ (this can only happen in the case n even). Then the free edges are formed of the edges $e(y_{r,s}), e(y_{r,s'})$, which have length r; we choose two vertices for the multi-edge and two more vertices v_1, v_2 (one for each of the free edges), but then the two missing vertices v_1', v_2' have at most two options each, because $|v - v_1|_n = |v_2 - v_2'|_n = r$. Thus this subcase is bounded by $O(n^4)$. The second subcase is when $e(y_{i,j}) \neq e(y_{i,j'})$ and $e(y_{r,s}) \neq e(y_{r,s'})$. Then w.l.o.g. the multi-edge is formed of the $e(y_{i,j}) = e(y_{r,s})$ then $i = r$, thus all edges have the same length; we choose two vertices v, v' for the multi-edge and two more vertices v_1, v_2 (one for each of the free edges). The missing vertices v_1', v_2' have at most two choices each because $|v_1 - v_1'|_n = |v_2 - v_2'|_n = |v - v'|_n$, which gives again a $O(n^4)$ bound.

For G_9, if we are in the case n odd, then the multi-edge is formed of edges of different groups, w.l.o.g. $e(y_{i,j}) = e(y_{r,s})$ and $i = r$. Therefore the edge attached to the multi-edge is uniquely defined because its length is determined, and the isolated edge is almost uniquely defined once one of the end points is chosen, because the other end has at most two choices. Overall, this gives the $O(n^3)$ bound. In the case n even, it can happen that w.l.o.g. $e(y_{i,j}) = e(y_{i,j'})$ but this can only happen when $i = n/2$. Then the multi-edge is uniquely defined by choosing one end, the isolated edge is defined by choosing two end points, and the last edge has at most four options since its length is already determined by the length of the isolated edge. This gives again a $O(n^3)$ bound.

For G_{10}, in the case n odd we can assume as before $e(y_{i,j}) = e(y_{r,s})$. Then $i = r$, and the multi-edge is determined by choosing two vertices and the remaining two edges are uniquely defined by the central vertex. This yields the bound $O(n^3)$. In the other case, w.l.o.g. $e(y_{i,j}) = e(y_{i,j'})$, and $i = n/2$. The multi-edge can be defined by choosing only one vertex, and the isolated path can be defined by choosing two vertices for one edge, while the remaining edge will have at most two options. This yields again a $O(n^3)$ bound.

For G_{11}, if $e(y_{i,j}) = e(y_{r,s})$, then all edges have the same length $i = r$, we can choose two vertices for the first multi-edge and one vertices for the second multi-edge, while the remaining vertex has at most two options. This yields a $O(n^3)$ bound. In the case when $e(y_{i,j}) = e(y_{i,j'})$ then $e(y_{r,s}) = e(y_{r,s'})$ and $i = r = n/2$. In this case we can choose two vertices (one for each multi-edge), and the remaining two vertices are automatically determined. This yields a $O(n^2) = O(n^3)$ bound. Thus we have established Claim 20. $\square$

We continue with the proof of Claim 19 in the case when n is odd. We observe that

$$\mathbb{E}\left[y_{i,j}y_{i,j'}y_{r,s}y_{r,s'}\right] = \mathbb{P}\left[y_{i,j}y_{i,j'}y_{r,s}y_{r,s'} = 1\right] = \mathbb{P}\left[\{\mathbf{b} : \ T_{\mathbf{b}}(i,j) = T_{\mathbf{b}}(i,j') = T_{\mathbf{b}}(r,s) = T_{\mathbf{b}}(r,s') = 0\}\right],$$

and thus for $Y = \{y_{i,j}, y_{i,j'}, y_{r,s}, y_{r,s'}\} \in \mathcal{Y}$, we have that

$$\mathbb{E}\left[y_{i,j}y_{i,j'}y_{r,s}y_{r,s'}\right] = \begin{cases} \frac{1}{n^3}, & \text{if } G(Y) \cong G_{1,2,3,5,6,7,9,10,12}; \\ \frac{1}{n^2}, & \text{if } G(Y) \cong G_{4,8,11}. \end{cases} \tag{1.23}$$

The last equation, combined with Claim 20, implies that

$$\begin{aligned} \sum_{\substack{1\le i,r\le \left\lfloor \frac{n}{2}\right\rfloor \\ 0\le j,j',s,s'\le n-1 \\ j<j';\, s<s' \\ (i,j,j')\neq(r,s,s')}} \mathrm{Cov}\left[y_{i,j}y_{i,j'}, y_{r,s}y_{r,s'}\right] &\le \sum_{\substack{1\le i,r\le \left\lfloor \frac{n}{2}\right\rfloor \\ 0\le j,j',s,s'\le n-1 \\ j<j';\, s<s' \\ (i,j,j')\neq(r,s,s')}} \mathbb{E}\left[y_{i,j}y_{i,j'}y_{r,s}y_{r,s'}\right] \\ &\le 7\cdot O(n^4)\frac{1}{n^3} + 3\cdot O(n^3)\frac{1}{n^2} + 2\cdot O(n^3)\frac{1}{n^3} \\ &= O(n). \end{aligned}$$

Using the previous inequality and (1.22) we get that

$$\mathbb{V}\left[\mathcal{Z}_1\right] = \sum_{\substack{1\le i\le \left\lfloor \frac{n}{2}\right\rfloor \\ 0\le j<j'\le n-1}} \mathbb{V}\left[y_{i,j}y_{i,j'}\right] + \sum_{\substack{1\le i,r\le \left\lfloor \frac{n}{2}\right\rfloor \\ 0\le j,j',s,s'\le n-1 \\ j<j';\, s<s' \\ (i,j,j')\neq(r,s,s')}} \mathrm{Cov}\left[y_{i,j}y_{i,j'}, y_{r,s}y_{r,s'}\right] = O(n) + O(n) = O(n). \tag{1.24}$$

This completes the proof of Claim 19 in the case when n is odd.

Case 2: n even. We estimate the variances of $\mathcal{Z}_0$ and $\mathcal{Z}_1$. For n even, we can write $\mathcal{Z}_0$ as

$$\mathcal{Z}_0 = \sum_{\substack{1\le i<\frac{n}{2}\\ 0<j\le n-1}} y_{i,j} + 2\sum_{j=0}^{\frac{n}{2}-1} y_{\frac{n}{2},j},$$

where all variables involved in the sums are mutually independent (see Remark 9). Thus

$$\mathbb{V}[\mathcal{Z}_0] = \sum_{\substack{1\le i<\frac{n}{2}\\ 0\le j\le n-1}} \mathbb{V}[y_{i,j}] + 4\sum_{j=0}^{\frac{n}{2}-1} \mathbb{V}\left[y_{\frac{n}{2},j}\right].$$

Using (1.17), we deduce that

$$\mathbb{V}[\mathcal{Z}_0] = \left(\frac{n}{2}-1\right) n\left(\frac{1}{n}-\frac{1}{n^2}\right) + 4\frac{n}{2}\left(\frac{1}{n}-\frac{1}{n^2}\right) = O(n), \tag{1.25}$$

for all n even. By Remark 9, we can write $\mathcal{Z}_1$ as

$$\mathcal{Z}_1 = \sum_{\substack{1\le i\le\frac{n}{2}\\ 0\le j<j'\le n-1\\ j\not\equiv j' \pmod{n/2}}} y_{i,j}y_{i,j'} + \sum_{s=0}^{\frac{n}{2}-1} y_{n/2,s}.$$

Therefore

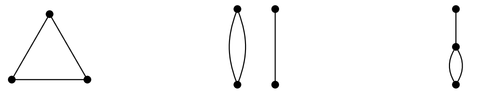
$$\begin{aligned}\mathbb{V}[\mathcal{Z}_1] = &\sum_{\substack{1\le i\le\frac{n}{2}\\ 0\le j<j'\le n-1\\ j\not\equiv j' \pmod{n/2}}} \mathbb{V}[y_{i,j}y_{i,j'}] + \sum_{s=1}^{\frac{n}{2}-1}\mathbb{V}\left[y_{n/2,s}\right] + \sum_{\substack{1\le i,r\le\frac{n}{2}\\ 0\le j,j',s,s'\le n-1\\ j<j';s<s'\\ j\not\equiv_{\frac{n}{2}} j';\, s\not\equiv_{\frac{n}{2}} s'\\ (i,j,j')\neq(r,s,s')}} \mathrm{Cov}\left[y_{i,j}y_{i,j'}, y_{r,s}y_{r,s'}\right] \\ &+ 2\sum_{\substack{1\le u\le\frac{n}{2}\\ 0\le v<v'\le n-1\\ v\not\equiv_{\frac{n}{2}} v'\\ 0\le w\le\frac{n}{2}-1}} \mathrm{Cov}\left[y_{u,v}y_{u,v'}, y_{\frac{n}{2},w}\right] + \underbrace{\sum_{\substack{0\le w,w'\le\frac{n}{2}-1\\ w\neq w'}} \mathrm{Cov}\left[y_{\frac{n}{2},w}, y_{\frac{n}{2},w'}\right]}_{=0 \text{ (by Remark 9)}}.\end{aligned} \tag{1.26}$$

We divide the analysis into three parts: the first two sums, the third sum, and the fourth sum. Using Remark 9, we write the first two sums in (1.26) as

$$\sum_{\substack{1\le i\le\frac{n}{2}\\ 0\le j<j'\le n-1\\ j\not\equiv j' \pmod{n/2}}} \mathbb{V}[y_{i,j}]\,\mathbb{V}[y_{i,j'}] + \sum_{s=1}^{\frac{n}{2}-1}\mathbb{V}\left[y_{n/2,s}\right] \overset{(1.17)}{\le} n\cdot n^2\left(\frac{1}{n}-\frac{1}{n^2}\right)^2 + n\left(\frac{1}{n}-\frac{1}{n^2}\right) = O(n). \tag{1.27}$$

The third sum in (1.26) can be bounded above in the same way as in the odd case: the associated graphs of variables $y_{i,j}, y_{i,j'}, y_{r,s}, y_{r,s'}$ with non-zero covariance in the third sum, are isomorphic to one of the graphs in Figure 1.2. Thus we can use Claim 20 and (1.23) to obtain

$$\sum_{\substack{1\le i,r\le\frac{n}{2}\\ 0\le j,j',s,s'\le n-1\\ j<j';s<s'\\ j\not\equiv_{\frac{n}{2}} j';\, s\not\equiv_{\frac{n}{2}} s'\\ (i,j,j')\neq(r,s,s')}} \mathrm{Cov}\left[y_{i,j}y_{i,j'}, y_{r,s}y_{r,s'}\right] = O(n). \tag{1.28}$$



G_{13} G_{14} G_{15}

In the fourth sum in (1.26), the variables with non-zero covariance have an associated multi-graph which is isomorphic to one of the following multi-graphs.
Let $\mathcal{X} := \left\{\{y_{u,v}, y_{u,v'}, y_{\frac{n}{2},w}\} : 1 \le u \le \frac{n}{2}; 0 \le v < v' \le n-1; v \not\equiv_{\frac{n}{2}} v'; 0 \le w \le \frac{n}{2}-1\right\}$. In the same way as Claim 20, we can prove that

$$\#\{X \in \mathcal{X} : G(X) \cong G_c\} = O(n^3), \quad c = 13, 14, 15.$$

As in (1.23), we can prove that $\mathbb{E}\left[y_{u,v}y_{u,v'}y_{\frac{n}{2},w}\right] = \frac{1}{n^2}$ for all $X = \{y_{u,v}, y_{u,v'}, y_{\frac{n}{2},w}\} \in \mathcal{X}$. Thus

$$\sum_{\substack{1\le u\le \frac{n}{2}\\ 0\le v<v'\le n-1\\ 0\le w\le \frac{n}{2}-1}} \operatorname{Cov}\left[y_{u,v}y_{u,v'}, y_{\frac{n}{2},w}\right] \le 3\cdot O(n^3)\frac{1}{n^2} = O(n). \tag{1.29}$$

Plugging (1.27),(1.28),(1.29) into (1.26) finally yields

$$\mathbb{V}\left[\mathcal{Z}_1\right] = O(n) + O(n) + 2\cdot O(n) = O(n), \tag{1.30}$$

for all n even. Equations (1.19),(1.24),(1.25) and (1.30) together yield Claim 19 in the case when n is even. Thus we have fully established Claim 19. □

1.5.3 $\mathcal{E}_{\text{zero}}(1/2-\varepsilon)$ has high probability

Using Chebyshev's inequality, we obtain that

$$\mathbb{P}\left[|\mathcal{Z}_0 - \mathbb{E}\left[\mathcal{Z}_0\right]| \ge \lambda_0\right] \le \frac{\mathbb{V}\left[\mathcal{Z}_0\right]}{\lambda_0^2}; \qquad \mathbb{P}\left[|\mathcal{Z}_1 - \mathbb{E}\left[\mathcal{Z}_1\right]| \ge \lambda_1\right] \le \frac{\mathbb{V}\left[\mathcal{Z}_1\right]}{\lambda_1^2},$$

for every $\lambda_0, \lambda_1 > 0$. In particular, this implies that

$$\mathbb{P}\left[\mathcal{Z}_0 < \mathbb{E}\left[\mathcal{Z}_0\right] - \lambda_0\right] \le \frac{\mathbb{V}\left[\mathcal{Z}_0\right]}{\lambda_0^2}; \qquad \mathbb{P}\left[\mathcal{Z}_1 > \mathbb{E}\left[\mathcal{Z}_1\right] + \lambda_1\right] \le \frac{\mathbb{V}\left[\mathcal{Z}_1\right]}{\lambda_1^2}.$$

Let $\varepsilon \in (0,1)$ be the constant from the statement of Lemma 7, and set $\nu = \varepsilon n/8$. Choosing $\lambda_0 = \lambda_1 = \nu$ and using Claims 18 and 19 we get that

$$\mathbb{P}\left[\mathcal{Z}_0 < \mathbb{E}\left[\mathcal{Z}_0\right] - \nu\right] \le \frac{\mathbb{V}\left[\mathcal{Z}_0\right]}{\nu^2} = \frac{O(n)}{n^2} = O\left(\frac{1}{n}\right);$$
$$\mathbb{P}\left[\mathcal{Z}_1 > \mathbb{E}\left[\mathcal{Z}_1\right] + \nu\right] \le \frac{\mathbb{V}\left[\mathcal{Z}_1\right]}{\nu^2} = \frac{O(n)}{n^2} = O\left(\frac{1}{n}\right).$$

By Claim 18 we have

$$\delta := (1/2-\varepsilon)\left\lfloor\frac{n}{2}\right\rfloor \le \mathbb{E}\left[\mathcal{Z}_0 - \mathcal{Z}_1\right] - 2\nu$$

for all sufficiently large n. Thus using Claim 15 we can conclude that

$$\begin{aligned}\mathbb{P}\left[\mathcal{E}^c_{\text{zero}}(1/2-\varepsilon)\right] &= \mathbb{P}\left[\left\{\mathbf{b} \in \mathcal{M}_n : \ D(\mathbf{b}) < (1/2-\varepsilon)\left\lfloor\frac{n}{2}\right\rfloor\right\}\right] \\ &\le \underbrace{\mathbb{P}\Big[\mathcal{Z}_0 < \mathbb{E}\left[\mathcal{Z}_0\right] - \nu\Big]}_{=O\left(\frac{1}{n}\right)} + \underbrace{\mathbb{P}\Big[\mathcal{Z}_1 > \mathbb{E}\left[\mathcal{Z}_1\right] + \nu\Big]}_{=O\left(\frac{1}{n}\right)} \\ &= O\left(\frac{1}{n}\right).\end{aligned}$$

This concludes the proof of Lemma 7.

1.6 Connections with chromatic polynomials of circulant graphs

As we have already seen in the proof of Claim 11, the multi-graph associated with the variables in row $i \neq \frac{n}{2}$ of $T_{\mathbf{b}}$ is the circulant graph $C_n(i)$, and the same holds for the variables in row $n/2$ if we consider the associated graph and not the associated multi-graph. Furthermore, we can express the probability of synchronization of circular automata in terms of chromatic polynomials of circulant graphs: this is a consequence of the close connection of the moments of $D(\mathbf{b})$ to chromatic polynomials of circulant graphs. We formalize this in the following results.

Definition. The *circulant graph* $C_n(i_1, i_2, \ldots, i_k)$ is a graph with vertex set $\mathbb{Z}_n$ where two vertices r, s are adjacent if $\left|r-s\right|_n \in \{i_1, i_2, \ldots, i_k\}$.

Definition. Let G be a graph with vertex set $\{0, 1, \ldots, n-1\}$. The *chromatic polynomial* $P(G; x) : \mathbb{N} \to \mathbb{N}$ of G is defined by

$$P(G;x) := \#\{\mathbf{b} \in \{0, \ldots, x-1\}^n : \ \mathbf{b} \text{ is a proper coloring of } G\}.$$

Remark 21. Let G be of order n. Then $P(G;x) = \sum_{j=1}^{n} \lambda_j x^j$, where $\lambda_j \in \mathbb{Z}$ (see, for instance, [Fengming et al., 2005]).

Claim 22. *Let $D(\mathbf{b})$ and $\mathbf{b} = (b_0, b_1, \ldots, b_{n-1}) \in \mathcal{M}_n$ be as in Lemma 7. Then*

$$\mathbb{E}[D] = \left\lfloor \frac{n}{2} \right\rfloor - \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \frac{P_i(n)}{n^n}$$

and

$$\mathbb{V}[D] = \sum_{i=1}^{n} \left(\frac{P_i(n)}{n^n} - \frac{P_i^2(n)}{n^{2n}} \right) + 2 \sum_{1 \le i < j \le \lfloor \frac{n}{2} \rfloor} \left(\frac{P_{i,j}(n)}{n^n} - \frac{P_i(n) P_j(n)}{n^{2n}} \right),$$

where P_i is the chromatic polynomial of the circulant graph $C_n(i)$ and $P_{i,j}$ is the chromatic polynomial of the circulant graph $C_n(i,j)$.

Remark 23. • It is easy to derive that $P_i(x) = \left((x-1)^{l_i} + (-1)^{l_i}(x-1)\right)^{\frac{n}{l_i}}$ where $l_i = \frac{n}{\gcd(n,i)}$, because $C_n(i)$ is a collection of $\gcd(n,i)$ disjoint cycles of length $\frac{n}{\gcd(n,i)}$ [Boesch and Tindell, 1984]. With this explicit expression, an easy corollary of Claim 22 is the estimate $\mathbb{E}[D] \sim (1-e^{-1}) \left\lfloor \frac{n}{2} \right\rfloor$.
• We could not find an explicit expression for $P_{i,j}$. The calculation of the chromatic number of circulant graphs with an arbitrary number of parameters is an NP-Hard problem [Codenotti et al., 1998]. This implies that the calculation of chromatic polynomials of circulant graphs is also NP-Hard since $\chi(G) = \operatorname{argmin}_{w \in \mathbb{N}} P(G; w) > 0$ – we believe that our unfruitful attempts to estimate $\mathbb{V}[D]$ are connected to this. To circumvent these issues, the variables $\mathcal{Z}_0$ and $\mathcal{Z}_1$ in Section 1.5 were introduced.

Proof of Claim 22. Let us recall that $D(\mathbf{b}) = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} D_i(\mathbf{b})$, where

$$D_i(\mathbf{b}) := \begin{cases} 1, & \text{if there exist } k, l \in \mathbb{Z}_n \text{ such that } \left|k-l\right|_n = i \text{ and } \left|b_k - b_l\right|_n = 0. \\ 0, & \text{otherwise,} \end{cases}$$

Then $D_i(\mathbf{b}) = 1 - x_i(\mathbf{b})$, where

$$x_i(\mathbf{b}) := \prod_{j=0}^{n} \left(1 - \mathbb{1}\{\left|b_j - b_{(j+i)_n}\right|_n = 0\}\right).$$

We observe that $x_i(\mathbf{b}) = 1$ if and only if every two numbers $r, s \in \mathbb{Z}_n$ at cyclic distance i have different images under $\mathbf{b}$ and $x_i(\mathbf{b}) = 0$ otherwise. If we consider $\mathbf{b}$ as a random coloring of $C_n(i)$, then $x_i(\mathbf{b}) = 1$ if and only if $C_n(i)$ is properly colored by $\mathbf{b}$. Thus

$$\mathbb{E}[x_i] = \mathbb{P}[\{\mathbf{b} : \; x_i(\mathbf{b}) = 1\}] = \frac{P_i(n)}{n^n}.$$

In a similar way

$$\mathbb{E}[x_i x_j] = \mathbb{P}[\{\mathbf{b} : \; x_i(\mathbf{b}) x_j(\mathbf{b}) = 1\}] = \frac{P_{i,j}(n)}{n^n}.$$

Therefore

$$\mathbb{E}[D] = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \mathbb{E}[D_i] = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} (1 - \mathbb{E}[x_i]) = \left\lfloor \frac{n}{2} \right\rfloor - \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \frac{P_i(n)}{n^n},$$

as well as

$$\mathbb{V}[D_i] = \mathbb{E}\left[D_i^2\right] - \mathbb{E}[D_i]^2 = \left(1 - \frac{P_i(n)}{n^n}\right) - \left(1 - \frac{P_i(n)}{n^n}\right)^2 = \frac{P_i(n)}{n^n} - \frac{P_i^2(n)}{n^{2n}}$$

and

$$\begin{aligned} \mathrm{Cov}[D_i, D_j] = \mathbb{E}[D_i D_j] - \mathbb{E}[D_i]\,\mathbb{E}[D_j] &= \mathbb{E}[(1 - x_i)(1 - x_j)] - \mathbb{E}[1 - x_i]\,\mathbb{E}[1 - x_j] \\ &= \mathbb{E}[x_i x_j] - \mathbb{E}[x_i]\,\mathbb{E}[x_j] \\ &= \frac{P_{i,j}(n)}{n^n} - \frac{P_i(n) P_j(n)}{n^{2n}}. \end{aligned}$$

Plugging the two previous equations into

$$\mathbb{V}[D] = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \mathbb{V}[D_i] + 2 \sum_{1 \le i < j \le \lfloor \frac{n}{2} \rfloor} \mathrm{Cov}[D_i, D_j]$$

yields Claim 22. □

We get the following relation between chromatic polynomials of circulant graphs and synchronization of circular automata. The number 0.13 in the statement of Theorem 24 comes from $1/2 - e^{-1} \approx 0.13$.

Theorem 24. *Let $\mathcal{A}_n(\mathbf{b})$ be a circulant graph as introduced in Section 1.2. Let $\varepsilon \in (0, 0.13]$, then there exist $n_\epsilon \in \mathbb{N}$ such that for all $n \ge n_\epsilon$ it holds that*

$$\mathbb{P}[\{\mathbf{b} \in \mathcal{M}_n : \; \mathcal{A}_n(\mathbf{b}) \text{ synchronizes}\}] \ge 1 - \left\lfloor \frac{n}{2} \right\rfloor \exp\left\{ -\frac{1}{2n} \left(\varepsilon \left\lfloor \frac{n}{2} \right\rfloor - 1 \right)^2 \right\} - \frac{\mathbb{V}[D]}{\left(\varepsilon \left\lfloor \frac{n}{2} \right\rfloor - 1 \right)^2},$$

where $\mathbb{V}[D]$ is as given in Claim 22.

Proof. By (1.12),(1.14) we know that

$$\mathbb{P}[\mathcal{E}^c_{\text{row}}(\alpha^\star)] \le \left\lfloor \frac{n}{2} \right\rfloor \exp\left\{ -\frac{1}{2n} \left(\varepsilon \left\lfloor \frac{n}{2} \right\rfloor - 1 \right)^2 \right\}, \tag{1.31}$$

for all $\varepsilon > 0$ and n large enough, where $\alpha^\star = 1 - e^{-1} - \varepsilon$. Using the expression for P_i in Remark 23, together with the well-know inequality $1 - x \leq e^{-x}$, $x \in \mathbb{R}$, we bound $P_i(n)/n^n$ from above

$$\begin{aligned}
\frac{\mathbb{P}i}{n^n} &= \left(\frac{n-1}{n}\right)^n \left(1 + \frac{(-1)^{\ell_i}}{(n-1)^{\ell_i - 1}}\right)^{\frac{n}{\ell_i}} \\
&\leq e^{-1} \left(1 + \frac{1}{(n-1)^{\ell_i - 1}}\right)^{\frac{n}{\ell_i}} \\
&\leq e^{-1} e^{\frac{n}{\ell_i} \cdot \frac{1}{(n-1)^{\ell_i - 1}}} \\
&= \exp\left\{-1 + \frac{n}{\ell_i \cdot (n-1)^{\ell_i - 1}}\right\}
\end{aligned}$$

and thus

$$\frac{\mathbb{P}i}{n^n} \leq \begin{cases} \exp\left\{-1 + \frac{1}{2} \cdot \left(\frac{n}{n-1}\right)\right\}, & \text{if } i = \frac{n}{2} \text{ i.e. } \ell_i = 2; \\[2ex] \exp\left\{-1 + \frac{n}{3(n-1)^2}\right\}, & \text{if } i \neq \frac{n}{2} \text{ i.e. } \ell_i \geq 3. \end{cases} \tag{1.32}$$

Using Equation 1.32 and the equation $\mathbb{E}[D] = \left\lfloor \frac{n}{2} \right\rfloor - \sum_{i=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \frac{P_i(n)}{n^n}$ from Claim 22 we get that

$$\mathbb{E}[D] \geq \left\lfloor \frac{n}{2} \right\rfloor \left(1 - \exp\left\{\frac{n}{3(n-1)^2} - 1\right\}\right) - 1 = \eta_\star.$$

By Chebyshev's inequality and elementary manipulations, we get that

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n : \ D(\mathbf{b}) < \eta_\star - \lambda\}\right] \leq \frac{\mathbb{V}[D]}{\lambda^2},$$

for all $\lambda > 0$. Let $\varepsilon > 0$. Setting $\lambda = \lambda'_\varepsilon(n) = \eta_\star - \left\lfloor \frac{n}{2} \right\rfloor (1 - e^{-1} - \varepsilon) + 1$ and noting that $\lambda > 0$ for n large enough, we get that

$$\mathbb{P}\left[\left(\mathcal{E}^{\tilde{\beta}}_{\text{zero}}\right)^c\right] = \mathbb{P}\left[\left\{\mathbf{b} \in \mathcal{M}_n : \ D(\mathbf{b}) < \left\lfloor \frac{n}{2} \right\rfloor \left(1 - e^{-1} - \varepsilon\right) - 1\right\}\right] \leq \frac{\mathbb{V}[D]}{(\lambda'_\varepsilon(n))^2} \leq \frac{\mathbb{V}[D]}{\left(\left\lfloor \frac{n}{2} \right\rfloor \varepsilon - 1\right)^2} \tag{1.33}$$

for n sufficiently large, where $\tilde{\beta} = 1 - e^{-1} - \varepsilon - 1/\left\lfloor \frac{n}{2} \right\rfloor$. Using the previous inequalities, we conclude that

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n : \ \mathcal{A}_n(\mathbf{b}) \text{ synchronizes}\}\right] \overset{(1.7)}{\geq} 1 - \mathbb{P}\left[\mathcal{E}^c_{\text{row}}(\alpha^\star)\right] - \mathbb{P}\left[\mathcal{E}^c_{\text{zero}}(\tilde{\beta})\right] \tag{1.34}$$

$$\geq 1 - \left\lfloor \frac{n}{2} \right\rfloor \exp\left\{-\frac{1}{2n}\left(\varepsilon \left\lfloor \frac{n}{2} \right\rfloor - 1\right)^2\right\} - \frac{\mathbb{V}[D]}{\left(\varepsilon \left\lfloor \frac{n}{2} \right\rfloor - 1\right)^2} \tag{1.35}$$

for n large enough where the relations $\alpha^\star, \tilde{\beta} > 0$ and $\alpha^\star + \tilde{\beta} > 1$ are valid when $\varepsilon \in (0, 0.13]$ and n is large enough. □

Actually, we formulate the following conjecture:

Conjecture 25. $\mathbb{V}[D] = O(n)$.

This conjecture can be reduced to prove that there is $g : \mathbb{N} \to \mathbb{R}$ such that $\left|\frac{P_{i,j}(n)}{n^n} - \frac{P_i(n)P_j(n)}{n^{2n}}\right| \leq g(n) = O(1/n)$ for all i, j, indeed: from Equation 1.32 we see that $0 \leq P_i(n)/n^n \leq f(n) = O(1)$ for all i, therefore the first part of the sum of $\mathbb{V}[D]$ given in Claim 22 is $\left|\sum_{i=1}^n \left(\frac{P_i(n)}{n^n} - \frac{P_i^2(n)}{n^{2n}}\right)\right| \leq$

$nf(n) = O(n)$, the second part of the sum $\sum_{1 \leq i < j \leq \lfloor \frac{n}{2} \rfloor} \left(\frac{P_{i,j}(n)}{n^n} - \frac{P_i(n)P_j(n)}{n^{2n}} \right)$ has a quadratic number of elements of the form $\frac{P_{i,j}(n)}{n^n} - \frac{P_i(n)P_j(n)}{n^{2n}}$ and it can be bounded by $O(n^2)g(n) = O(n)$ if the assumption $|\frac{P_{i,j}(n)}{n^n} - \frac{P_i(n)P_j(n)}{n^{2n}}| \leq g(n) = O(1/n)$ for all i, j is true, making $\mathbb{V}[D] = O(n) + O(n) = O(n)$. In particular, a positive answer to this chromatic-polynomial question would give an alternative proof of Theorem 4.

1.7 Future work

Let $\mathcal{A}_n(\mathbf{a}, \mathbf{b})$ be an automaton where $\mathbf{a} : \mathbb{Z}_n \to \mathbb{Z}_n$ is fixed and $\mathbf{b} \in \mathcal{M}_n$. These are natural lines of research to extend/improve the results in this chapter:

- We want to explore in more detail the strengths and limitations in the ideas presented in this chapter. For example, we think that these ideas can extend Theorem 4 to the case where $\mathbf{a} : \mathbb{Z}_n \to \mathbb{Z}_n$ is in the form of a finite number of pairwise disjoint cycles of almost-equal length. We also think that (probabilistic) upper bounds for the length of the synchronizing minimal words can be given with our techniques, in the spirit of the results of [Nicaud, 2019].
- Theorem 3 has a decay rate in $\Theta\left(\frac{\sqrt{p}}{e^p}\right)$. Based on Theorem 3 and computational simulations, we believe that this can be extended:

Conjecture 26. There is some $\varepsilon > 0$ such that

$$\mathbb{P}\left[\{\mathbf{b} \in \mathcal{M}_n : \ \mathcal{A}_n(\mathbf{b}) \text{ synchronizes}\}\right] = 1 - O(e^{-\varepsilon n})$$

as $n \to \infty$.

Chapter 2

Estimations of means and variances in a Markov linear model

Multivariate regression models and ANOVA are probably the most frequently applied methods of all statistical analyses. We study the case where the predictors are qualitative variables, and the response variable is quantitative. In this case, we propose an alternative to the classic approaches that do not assume homoscedasticity or normality of the error term but assumes that a Markov chain can describe the covariates' correlations.

This approach transforms the dependent covariate using a change of measure to independent covariates. The transformed estimates allow a pairwise comparison of the mean and variance of the contribution of different values of the covariates. We show that under standard moment conditions, the estimators are asymptotically normally distributed. We test our method with simulated data and apply it to several classic data sets.

The contents of this chapter are based on the papers [Gutierrez and Müller, 2019] and [Gutierrez and Müller, 2020] which were written in collaboration with Sebastian Müller.

2.1 Introduction

We propose a linear model for qualitative predictors, $X^{(1)}, \ldots, X^{(m)}$ and a categorical response variable Y. In contrast to classic linear approaches as linear regression and ANOVA, our approach allows us to model different variances for each category and the error term to be arbitrary. It weakens the assumptions of independent predictors of a previous work [Gutierrez and Müller, 2019], in the following way: the value of $X^{(k+1)}$ conditioned on $X^{(k)}$ is independent of $X^{(k-1)}$, for all reasonable choices of k.

Our approach is in a "probabilistic spirit" since we use Markov chains and a change of measure similar to a Girsanov transform to model the correlation between the predictors and use classic results on random walks to study the asymptotic behavior of the resulting estimators.

The approach is probably best understood with a comparison to standard linear models. Let us assume that there are two categorical predictors or groups $X^{(1)}$ and $X^{(2)}$ and a quantitative response Y. We assume that each predictor can take values in $\{1, \ldots, k_i\}, i \in \{1, 2\}$ and encode the values of these variables using the vectors[1]

$$\widehat{X}^{(i)} = (\mathbb{1}\{X^{(i)} = 1\}, \ldots, \mathbb{1}\{X^{(i)} = k_i\}). \tag{2.1}$$

Let $\alpha^{(i)} \in \mathbb{R}^{k_i}$ be real vectors and $\varepsilon^{(i)}$ be random vectors taking values in $\mathbb{R}^{k_i}$; we assume that $\mathbb{E}[\varepsilon^{(i)}] = 0$, but allow the variances of each component to be different and denote $\sigma_j^{(i)} = \mathbb{V}[\varepsilon_j^{(i)}]$. We assume the following linear relationship between the two categorical predictors and the response variable:

$$y = (\alpha^{(1)} + \varepsilon^{(1)})\widehat{x}_1^t + (\alpha^{(2)} + \varepsilon^{(2)})\widehat{x}_2^t. \tag{2.2}$$

In a standard linear regression model or ANOVA the linear relationship between predictors and the response variable looks like

$$y = \alpha^{(1)}\widehat{x}_1^t + \alpha^{(2)}\widehat{x}_2^t + \varepsilon. \tag{2.3}$$

Note that here ε is a random error that it supposed to be centered and gaussian. In particular, the classic models assume that the contribution to the random error is the same for each category. In our proposed model, we generalize this to possible different variances and also provide estimators for these differences for different categories.

We propose a modeling of dependencies between different predictors using Markov chains and a pathwise approach. Since the predictors are categorical variables, every individual of the sample corresponds to a path. For instance, in Figure 2.1 the bold path corresponds to an individual with $X^{(1)} = X^{(2)} = 1$. A Markov chain now gives the dependencies between the different categories. The probability of choosing category i in predictor one is $p_{s,1i}$; corresponding to the initial measure of a Markov chain. Now, the probability of choosing j in predictor two conditioned on having chosen i in predictor one is given by $p_{1i,2j}$. Again, each path corresponds to the categories of an individual. If the transition probability is those of a uniform distribution, the categories of the individuals are independent.

In the case of non-uniform transition, estimations for the contribution of specific factors can be biased. For instance, in the situation of Figure 2.1 assume that factor 1 has no contribution, but the category in factor 2 has a much higher contribution than the other categories of this factor. Now, if the probability that category 1 of factor 1 is chosen together with category 1 of factor 2 with much higher probability than the other categories in factor 1, the contribution of category 1 of factor 1 is generally overestimated if one works under the assumption of independence of the factors. This situation corresponds to multicollinearity in linear models. The correlation between predictor variables makes estimation and interpretation of the models

[1] In contrast to the standard encoding of categorical variables in linear models with $k_i - 1$ dummy variables, we choose k_i dummy variables.

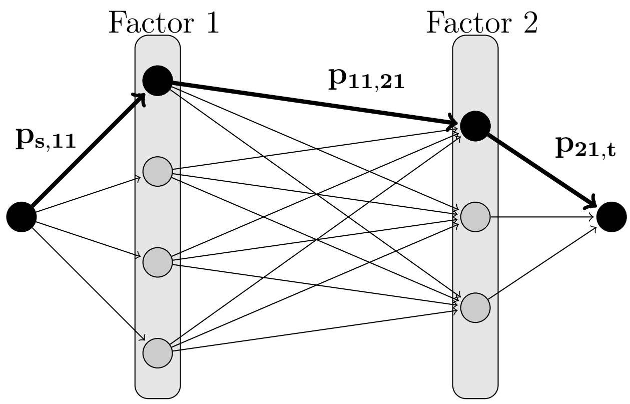


Figure 2.1: An illustration of a path passing through different workstations in a production network.

more complicated than in the independent case, since it is difficult to disentangle the impacts of different predictor variables on the response.

We solve this problem in transferring the non-uniform setting to a uniform setting using a discrete Girsanov transform, see Lemma 35. This transforms quantifies the biases introduced by the variables' dependence. Here, the martingale measure in the standard Girsanov transform corresponds to the setting of uniform transition or independence of the predictors. In other words, the Girsanov transform allows us to quantify the dependence of the predictors and change the measure to a uniform or independent setting.

As presented in Section 2.6, the estimators proposed show a good performance if the number of covariates (number of workstations in the network) is small or if the correlation structure is known (how the transitions behave). However, we have to note that the estimators may have a large variance in the case of a large number of covariates. This is because the variance of the estimator of given paths increases with the number of possible states. These effects amplify if the probabilities of the given paths are close to zero since these quantities are in the denominator of the final estimators (see Definition 1).

2.1.1 Motivation

The proposed model applies in a very general setting. However, it is motivated by real-world applications in the quality control of parallel production networks, see also the previous work [Gutierrez and Müller, 2019], and we think it might be useful to have this particular use case in mind.

In parallel production networks, the quality of the end product depends on various intermediate production steps. Therefore, a recent challenge is to achieve a level of visibility into the production flows that allow us to optimize throughput by guaranteeing at the same time given quality standards.

We consider a production network consisting of several workstations; see Figure 2.1. Each workstation is a parallel configuration of machines performing the same kind of tasks on a given part. Parts move from one workstation to another, and at each workstation, a part is assigned randomly to a machine. We assume that the production network is *acyclic*, i.e., a part does not return to a workstation where it previously received service. Furthermore, we assume that the quality of the end product is *additive*, that is, the final quality is the sum of the machines' quality contributions along the production path. Separate latent random variables model the contribution of each machine. Note that the product's quality can be replaced by the time it takes to perform specific tasks. In this language, our main result is constructing estimators that allow pairwise and multiple comparisons of the means and variances of different machines in the same workstation. These comparisons then may lead to the identification of unreliable machines.

In Section 2.6, we treat examples of different kinds to illustrate that the method applies to various kinds of use cases.

2.1.2 Related work

As mentioned above, estimations for the mean can also be conducted using multivariate regression with categorical covariates or ANOVA, e.g., [Rutherfors, 2011]. However, since in linear models, homoscedasticity is necessary, this approach naturally does not allow to compare differences of the variances. Linear regression models are often plagued by different variabilities or heteroskedasticity. We refer to [Kleiber and Zeileis, 2008, Chapter 4] for an overview of how to detect and control heteroskedasticity. In contrast to these problems, our approach naturally allows different variabilities between variables and can detect differences in the variability of values of a given variable.

Multifactor experimental designs, [Draper and Pukelsheim, 1996, Selvamuthu and Das, 2018] are also alternatives to estimate the mean differences but again rely on homoscedasticity. They are mostly used in the context of statistically planned experiments, which consists of a few experimental runs to obtain data on the characteristics under consideration. If the number of observations for each setting (or path in our notation) is sufficiently high and under further conditions described in [Draper and Pukelsheim, 1996, Section 4], these methods allow a comparison of the variances, too. While this may offer a feasible, however not direct, way to identify differences in variability if the number of covariates is small, it seems not practical in more complex situations.

There is also a connection to critical paths analysis, e.g., [Bohme et al., 2019, Schulz, 2005]. While these methods allow us to find critical paths in acyclic networks, they are not suited to compare nor estimate differences in mean and variances of given tasks.

2.1.3 Outline

In Section 2.2, we define the model using Markov chains and directed acyclic graphs. In Section 2.3 we summarize the results on the case of independent covariates from [Gutierrez and Müller, 2019]. Section 2.4 contains the main results. Lemma 35 describes the Girsanov transform . In Theorems 36, we prove asymptotic consistency of the estimators for mean and variance. In Section 2.5, their asymptotic distribution is given for the case where the correlations are known in Theorem 39, and where the correlations are unknown in Theorem 42. Last but not least, we present several examples in Section 2.6.

2.2 The model

We use a directed acyclic graph (DAG) to describe the dependencies of the covariates. A DAG is a finite directed graph with no directed cycles. It consists of a finite vertex set V and a finite set of directed edges $E = \{(v, w) : v, w \in V, v \neq w\}$. In our setting the DAG contains two special vertices: a source s and a sink t. We are interested in the paths from the source to the sink in this graph. We denote a path $\vec{p}$ in the DAG as $\vec{p} = (p_0, p_1, \dots, p_c, p_{c+1})$ where $p_0 = s$ and $p_{c+1} = t$ and $(p_i, p_{i+1}) \in E$. We define $\vec{p}[j] := p_j$. We refer to Figure 2.2 for an illustration and to [Bang-Jensen and Gutin, 2009] for more details on directed graphs.

We assume that at each step $1 \leq i \leq c$ the path $\vec{p}$ has $r_i \leq r$ different choices and the nodes in each column are always numerated starting with 1. The possible choices of a path can

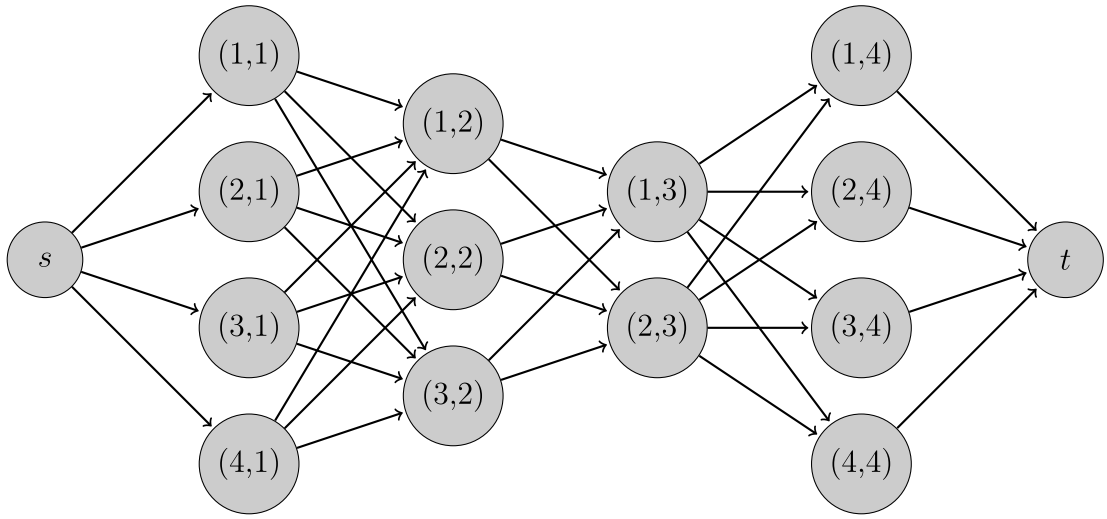


Figure 2.2: An illustration of a DAG with $c = 4$ and $r_1 = 4, r_2 = 3, r_3 = 2$ and $r_4 = 4$. Every node in column i has outgoing edges to every node in column $i + 1$, $i = 1, \dots, c - 1$.

therefore be modeled through an $r \times c$ matrix. More precisely, given a path $\vec{p}$, we associate an $r \times c$ binary matrix $V_{\vec{p}}$ that has 1's only in the nodes visited by the path:

$$V_{\vec{p}} := (V_{\vec{p}}(i, j))_{i \in [r],\, j \in [c]},$$

where we denote $[k] := \{1, 2, \dots, k\}$ for an integer k. We call $V_{\vec{p}}$ the indicator matrix of the path $\vec{p}$.

Each path contains exactly one node of each column. This chapter aims to study differences among nodes of the same columns. We think of nodes in the same columns as different possibilities for a given task, as different persons performing the same job, as different machines in the same workstation, or as variations of the same kind of treatment.

The given data consists of the list of paths $\{\vec{p}_i\}_{i=1,\dots,n}$ in the DAG and the list of outputs $\{b(\vec{p}_i)\}_{i=1,\dots,n}$ for each path. We consider the quality matrix S, which is a random matrix of size $r \times c$ with real entries:

$$S := (s(i, j))_{i \in [r],\, j \in [c]}, \qquad s(i, j) \in \mathbb{R}.$$

We model the paths with a random vector $\vec{\mathcal{P}} := (P_1, \dots, P_c)$ where the components P_i are random variables over the set $[r]$.

Throughout the chapter we work under the following standing assumptions:

Assumption 27. We assume that:

1. the paths $\vec{\mathcal{P}}_1, \vec{\mathcal{P}}_2, \dots$ are chosen independently and according to a Markov chain, see Section 2.4 for more details.

2. all entries of S have finite second moments;

3. the random variables $S(i, j), i \in [r],\ j \in [c]$ are (jointly) independent and the are independent from the paths $\vec{\mathcal{P}}_1, \vec{\mathcal{P}}_2, \dots$

Note that, we do not assume the entries of S to be identically distributed or having the same variance.

Let $\vec{p} = (p_0, p_1, \dots, p_c, p_{c+1})$ be a realization of $\vec{\mathcal{P}}$, where $p_0 = s$ and $p_{c+1} = t$ almost surely. Then, the quality of the construction path $\vec{p}$ is defined as

$$b(\vec{p}, S) := \sum_{j=1}^{c} S(p_j, j).$$

In some situation we will abbreviate $b(\vec{p}, S)$ and write only $b(\vec{p})$. We also can think of $b(\vec{p})$ as the quality (or error) cumulated along the path $\vec{p}$ or as the response variable Y in the language of linear models. We want to stress out that the model fits the description in (2.2) and that we will continue to use the "graph-based" description of the model since it is best suited for our approach and used methods.

Let us make precise how the randomness enters in our model. We choose a random path $\vec{\mathcal{P}}$ and a random matrix S. The corresponding probability measure is denoted by $\mathbb{P}$. The random choice of $\vec{\mathcal{P}}$ and S induces a random variable $b(\vec{\mathcal{P}}) = b(\vec{\mathcal{P}}, S)$ and allows to generate a sequence of i.i.d. random variables $(\vec{\mathcal{P}}_1, b(\vec{\mathcal{P}}_1)), (\vec{\mathcal{P}}_2, b(\vec{\mathcal{P}}_2)), \ldots$.

Our goal is to give estimates on the law of S by observing the paths $\vec{\mathcal{P}}$ and its cumulated qualities $b(\vec{\mathcal{P}})$. Note that, $(\vec{\mathcal{P}}_n, b(\vec{\mathcal{P}}_n))_{n\in\mathbb{N}}$ is in general not a sufficient statistic for S, i.e., we can not recover the distribution of S by only observing realizations of $(\vec{\mathcal{P}}, b(\vec{\mathcal{P}}))$, as we see in the following remark.

Remark 28. Let us consider the case $r = 1$ and $c = 2$. Let $S(1,1) \sim \mathcal{N}(0,1)$ and $S(1,2) \sim \mathcal{N}(1,1)$ and define $\tilde{S}(1,1) := S(1,1) + 1$ and $\tilde{S}(1,2) := S(1,2) - 1$. Then for any given path $\vec{p}$ we have that $\sum_{j=1}^{2} S(p_j, j) = \sum_{j=1}^{2} \tilde{S}(p_j, j)$. Hence, the statistic $(\vec{p}_n, b(\vec{p}_n))_{n\in\mathbb{N}}$ does not allow us to distinguish between S and $\tilde{S}$.

Example 29 (Binary errors). The matrix S consists of independent Bernoulli random variables $S(i,j)$. The value 1 of this Bernoulli may encode a defect and hence $b(\vec{p})$ counts the number of defects of the end product.

Example 30 (Gaussian error). The matrix S consists of independent Gaussian random variables $S(i,j)$. The quality or response $b(\vec{p})$ is then distributed as a (random) mixture of Gaussian random variables.

Given a sequence of realizations $(\vec{p}_k)_{k\in[n]}$ of $\vec{\mathcal{P}}$, we define the following matrices that are at the core of our analysis:

$$B^{(n)} := \sum_{k=1}^{n} b(\vec{p}_k) V_{\vec{p}_k}, \;\; V^{(n)} := \sum_{k=1}^{n} V_{\vec{p}_k}, n \geq 1.$$

The value $B^{(n)}(i,j)$ is the sum of all cumulated qualities of paths containing node (i,j), whereas $V^{(n)}(i,j)$ just counts the number of times node (i,j) was used. We define the sample mean matrix as the sample mean quality matrix :

$$T^{(n)} := (T^{(n)}(i,j))_{i\in[r], j\in[c]}, \;\; \text{where } T^{(n)}(i,j) := \begin{cases} \frac{B^{(n)}(i,j)}{V^{(n)}(i,j)}, & \text{if } V^{(n)}(i,j) \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The corresponding sample variance matrix $\Sigma^{(n)}$ is defined by

$$\Sigma^{(n)}(i,j) := \begin{cases} \frac{1}{V^{(n)}(i,j)} \sum_{k=1}^{n} \left(b(\vec{p}_k) V_{\vec{p}_k}(i,j) - T^{(n)}(i,j) \right)^2, & \text{if } V^{(n)}(i,j) \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Remark 31. In this chapter, we assume every path has one vertex per column. However, it is relatively straightforward to generalize this model to the case where columns can be "skipped." , we can construct a new DAG by adding a dummy node in each column. For instance, this new dummy variable may describe the effect if a certain production step was skipped. Note also that in theory, the model can be generalized to general DAGs; however, we do not expand here on this since we did not find suitable applications of the more general model.

2.3 Uniform transition model.

In this section we review some results from [Gutierrez and Müller, 2019] in the special case where the paths $\vec{\mathcal{P}}_1, \vec{\mathcal{P}}_2, \ldots$ are chosen independently and uniformly.

Theorem 32. *Let the paths $\vec{\mathcal{P}}_1, \vec{\mathcal{P}}_2, \ldots$ be chosen independently and uniformly. Let $(i,j),(i',j) \in [r]\times[c]$, then*

$$T^{(n)}(i,j) - T^{(n)}(i',j) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[S(i,j)] - \mathbb{E}[S(i',j)]$$

and

$$\Sigma^{(n)}(i,j) - \Sigma^{(n)}(i',j) \xrightarrow[n\to\infty]{a.s.} \mathbb{V}[S(i,j)] - \mathbb{V}[S(i',j)].$$

Proof. We give the proof from [Gutierrez and Müller, 2019] since it illustrates the idea behind the construction of the estimators. Let $D^{(n)} = (\vec{\mathcal{P}}_1, \ldots, \vec{\mathcal{P}}_n)$ be the the (multi-)set of the first n paths and $D^{(n)}_{(i,j)}$ be the the multi-set containing only those paths passing through (i,j). Using twice the law of large numbers and the continuous mapping theorem, we obtain

$$\begin{aligned}
T^{(n)}(i,j) &= \frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b(\vec{p}) \\
&= \frac{n}{|D^{(n)}_{(i,j)}|}\frac{1}{n}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b(\vec{p}) \\
&\xrightarrow[n\to\infty]{a.s.} \frac{1}{\mathbb{P}(\vec{\mathcal{P}}[j]=i)}\mathbb{E}[b(\vec{\mathcal{P}});\vec{\mathcal{P}}[j]=i] = \mathbb{E}\left[b(\vec{\mathcal{P}})\,|\,\vec{\mathcal{P}}[j]=i\right].
\end{aligned}$$

In the same way

$$T^{(n)}(i',j) = \frac{1}{|D^{(n)}_{(i',j)}|}\sum_{\vec{p}\in D^{(n)}_{(i',j)}} b(\vec{p}) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}\left[b(\vec{\mathcal{P}})\,|\,\vec{\mathcal{P}}[j]=i'\right].$$

Using the assumption that the paths are chosen uniformly and the definition of $b(\vec{\mathcal{P}})$, we obtain the first part of the theorem from

$$\begin{aligned}
\mathbb{E}\left[b(\vec{\mathcal{P}})|\,\vec{\mathcal{P}}[j]=i\right] - \mathbb{E}\left[b(\vec{\mathcal{P}})\,|\,\vec{\mathcal{P}}[j]=i'\right] &= \frac{\mathbb{E}\left[b(\vec{\mathcal{P}});\vec{\mathcal{P}}[j]=i\right] - \mathbb{E}\left[b(\vec{\mathcal{P}});\vec{\mathcal{P}}[j]=i'\right]}{\mathbb{P}(\vec{\mathcal{P}}[j]=i)} \\
&= \frac{\mathbb{E}\left[S(i,j);\vec{\mathcal{P}}[j]=i\right] - \mathbb{E}\left[S(i',j);\vec{\mathcal{P}}[j]=i'\right]}{\mathbb{P}(\vec{\mathcal{P}}[j]=i)} \\
&= \mathbb{E}[S(i,j)] - \mathbb{E}[S(i',j)].
\end{aligned}$$

For the second part of the theorem, we use the law of large numbers and the continuous mapping theorem to get that

$$\begin{aligned}
\Sigma^{(n)}(i,j) &= \left(\frac{1}{|D^{(n)}_{ij}|}\sum_{\vec{p}\in D^{(n)}_{ij}} b^2(\vec{p})\right) - \left(T^{(n)}(i,j)\right)^2 \\
&\xrightarrow[n\to\infty]{a.s.} \mathbb{E}\left[b(\vec{\mathcal{P}})^2|\,\vec{\mathcal{P}}[j]=i\right] - \mathbb{E}\left[b(\vec{\mathcal{P}})|\,\vec{\mathcal{P}}[j]=i\right]^2.
\end{aligned}$$

Using the definition of $b(\vec{p})$ and the jointly independence of the entries of S (see Assumption 27) we deduce with elementary calculations that

$$\mathbb{E}\left[b(\vec{\mathcal{P}})^2|\,\vec{\mathcal{P}}[j]=i\right] - \mathbb{E}\left[b(\vec{\mathcal{P}})|\,\vec{\mathcal{P}}[j]=i\right]^2 = A_j + \mathbb{V}\left[S(i,j)\right]$$

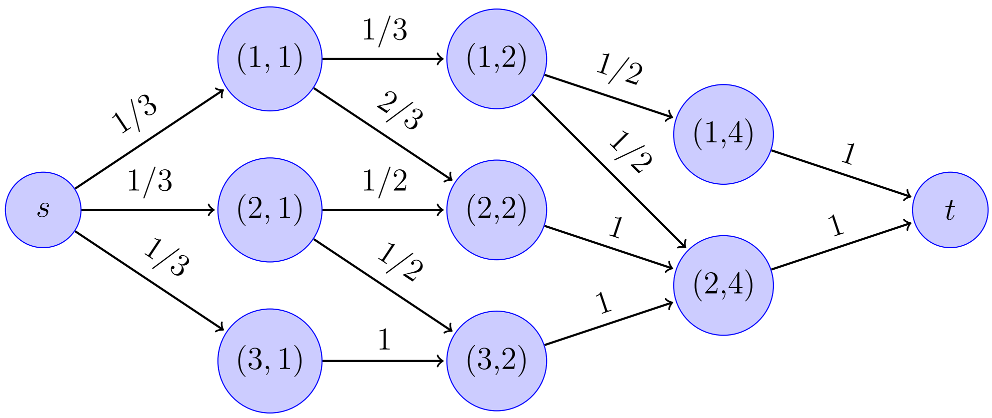


Figure 2.3: An example of the Markov chain transition model.

where

$$A_j = \sum_{\substack{l=1 \\ l \neq j}}^{c} \mathbb{V}\left[S(\vec{\mathcal{P}}[l], l) | \, \vec{\mathcal{P}}[j] = i\right]$$

is a quantity that only depend on the column j. Applying this identify for (i, j) and (i', j) we obtain that

$$\Sigma^{(n)}(i,j) - \Sigma^{(n)}(i',j) \xrightarrow[n\to\infty]{a.s.} \mathbb{V}\left[S(i,j)\right] - \mathbb{V}\left[S(i',j)\right].$$

□

2.4 Markov chain transition model.

In this Section, we will treat the more general case where the random path $\vec{\mathcal{P}}$ is the path of a Markov chain. This case corresponds to the situation where the covariates may be dependent, and their correlation follows a Markov structure. For instance, in the example of a production network, this means that the effect of machines of different production steps (or columns) are correlated.

More formally, we consider the time-inhomogeneous Markov chain on $\{1, \ldots, r\}$ with initial state s and absorbing state t. At the time 0, we start the Markov chain in the source s. The transition kernel for the first step is given by

$$Q^{(1)}(s,i) = \begin{cases} q^{(1)}_{s,i}, & \text{if } \ i \leq r_1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

The next $c - 1$ steps are defined as follows. For $2 \leq k \leq c$:

$$Q^{(k)}(i,j) = \begin{cases} q^{(k)}_{(i,j)}, & \text{if } \ i \in [r_{k-1}], j \in [r_k]); \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

Finally, the last step is determined by

$$Q^{(c+1)}(i,t) = \begin{cases} 1, & \text{if } i \in [r_c]; \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

The matrices $Q^{(k)}, k \in [c+1]$, are all supposed to be stochastic matrices. Equation (2.4) forces the initial state s to jump into a state of the first column. Equation (2.5) forces a node in column j to jump into a node of column $j + 1$. In the case that the node is in the last column c, Equation (2.6) forces it to jump into the final absorbing state t.

Remark 33. The uniform transition model is a special case of the Markov chain transition model. In fact, setting $q^{(1)}_{s,i} := 1/r_i,\ i \in [r_1]$ and $q^{(k)}_{i,j} := 1/r_k,\ i \in [r_{k-1}], j \in [r_k]$ yields the uniform transition model.

A Markov chain, as given above, defines a natural probability measure on the paths in the DAG. The measure that chooses a path $\vec{p}$ according to $Q := (Q^{(1)}, Q^{(2)}, \dots Q^{(c+1)})$, and the values along $\vec{p}$ according to S is denoted by $\mathbb{P}^Q$, where

$$\mathbb{P}^Q\left[\vec{\mathcal{P}} = \vec{p}\right] = Q^{(1)}(s, \vec{p}[1]) \cdot \prod_{k=2}^{c} Q^{(k)}(\vec{p}[k-1], \vec{p}[k]).$$

We define the support of the measure Q as

$$\mathrm{Supp}^Q := \left\{\vec{p} : \mathbb{P}^Q(\vec{\mathcal{P}} = \vec{p}) > 0\right\}.$$

A node is reachable if there exists a path with positive probability through this node. We enumerate the reachable nodes in each column starting from 1. The set

$$\mathrm{Supp}^Q_{i,j} := \{\vec{p} \in \mathrm{Supp}^Q : \vec{p}[j] = i\}$$

is the set of all paths of positive probability passing through (i, j). We consider the (multi-)set $D^{(n)} = (\vec{\mathcal{P}}_1, \dots, \vec{\mathcal{P}}_n)$ of the first n paths and the multi-set $D^{(n)}_{(i,j)}$ containing only those paths passing through (i, j).

Remark 34. Let (i, k) and (i', k) be two Q-reachable nodes in the same column. A consequence of the proof of Theorem 32 is that if for all $\ell \in [r]$

$$\begin{aligned} q^{(k)}(\ell, i) &= q^{(k)}(\ell, i'), \\ q^{(k+1)}(i, \ell) &= q^{(k+1)}(i', \ell), \end{aligned}$$

then

$$T^{(n)}(i, j) - T^{(n)}(i', j) \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s.} \mathbb{E}[S(i, j)] - \mathbb{E}[S(i', j)].$$

and

$$\Sigma^{(n)}(i, j) - \Sigma^{(n)}(i', j) \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s.} \mathbb{V}[S(i, j)] - \mathbb{V}[S(i', j)].$$

Let $Q := (Q^{(1)}, Q^{(2)}, \dots, Q^{(c+1)})$ et $\widetilde{Q} := (\widetilde{Q}^{(1)}, \widetilde{Q}^{(2)}, \dots \widetilde{Q}^{(c+1)})$ be two sequences of transition matrices. We say that Q and $\widetilde{Q}$ are equivalent (as measures) if

$$Q^{(k)}(i, j) > 0 \Longleftrightarrow \widetilde{Q}^{(k)}(i, j) > 0,$$

for all meaningful choices of i, j and k. In particular, Q and $\widetilde{Q}$ are equivalent iff $\mathrm{Supp}^Q = \mathrm{Supp}^{\widetilde{Q}}$. Moreover, the induced measures on the set of paths are equivalent and there exists a discrete Radon-Nykodym derivative that allows a change of measures.

Lemma 35. *Let (i, j) be a node and let $Q, \widetilde{Q}$ be two equivalent sequences of transition matrices and let $f : P^Q \times \mathbb{R}^{r\times c} \to \mathbb{R}$ be a measurable function, then*

$$\mathbb{E}^Q\left[f(\vec{\mathcal{P}}, S) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} | \vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} | \vec{\mathcal{P}}'[j] = i\right)} \middle| \vec{\mathcal{P}}[j] = i \right] = \mathbb{E}^{\widetilde{Q}}[f(\vec{\mathcal{P}}, S) | \vec{\mathcal{P}}[j] = i]$$

where $\vec{\mathcal{P}}'$ is an independent copy of $\vec{\mathcal{P}}$.

Proof. Given a fixed path $\vec{p}$, we observe that

$$\mathbb{E}^{\widetilde{Q}}\left[f(\vec{p},S)\right]=\mathbb{E}^{Q}\left[f(\vec{p},S)\right],$$

because S is independent of the random paths $\vec{\mathcal{P}}$ by point (3) in Assumption 27. Using this, the theorem of total probability and a change of measure, we get that

$$\begin{aligned}
\mathbb{E}^{\widetilde{Q}}[f(\vec{\mathcal{P}},S)|\vec{\mathcal{P}}[j]=i] &= \sum_{\vec{p}:\vec{p}[j]=i}\mathbb{E}^{\widetilde{Q}}\left[f(\vec{p},S)\right]\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)\\
&= \sum_{\vec{p}\in\ \mathrm{Supp}^{Q}_{i,j}}\mathbb{E}^{Q}\left[f(\vec{p},S)\right]\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)}\\
&= \sum_{\vec{p}\in\ \mathrm{Supp}^{Q}_{i,j}}\mathbb{E}^{Q}\left[f(\vec{p},S)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)}\right]\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)\\
&= \mathbb{E}^{Q}\left[f(\vec{\mathcal{P}},S)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}'=\vec{\mathcal{P}}|\vec{\mathcal{P}}'[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}'=\vec{\mathcal{P}}|\vec{\mathcal{P}}'[j]=i\right)}\middle|\vec{\mathcal{P}}[j]=i\right].
\end{aligned}$$

□

We can now define the estimators for the means and variances.

Definition 1. We define

$$T^{(n)}_{Q,\widetilde{Q}}(i,j):=\frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}}\left(b(\vec{p})\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}\right)$$

and

$$\Sigma^{(n)}_{Q,\widetilde{Q}}(i,j)\ :=\ \left(\frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}}b^2(\vec{p})\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}\right)-\left(T^{(n)}_{Q,\widetilde{Q}}(i,j).\right)^2$$

Theorem 36. *Let (i,j) be a node and let $Q,\widetilde{Q}$ be two equivalent sequences of transition matrices. Then,*

$$\begin{aligned}
T^{(n)}_{Q,\widetilde{Q}}(i,j) &\xrightarrow[n\to\infty]{\mathbb{P}^{Q}-a.s.}\mathbb{E}^{\widetilde{Q}}\left[b(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j]=i\right],\\
\Sigma^{(n)}_{Q,\widetilde{Q}}(i,j) &\xrightarrow[n\to\infty]{\mathbb{P}^{Q}-a.s.}\mathbb{V}^{\widetilde{Q}}[b^2(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i].
\end{aligned}$$

Proof. The first observation is that

$$T^{(n)}_{Q,\widetilde{Q}}(i,j)=\frac{n}{|D^{(n)}_{(i,j)}|}\cdot\frac{1}{n}\sum_{k=1}^{n}\left(\mathbb{1}\{\vec{p}_k[j]=i\}b(\vec{p}_k)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}_k\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}_k\middle|\vec{\mathcal{P}}[j]=i\right)}\right).$$

Using twice the law of large numbers and the continuous mapping theorem we obtain

$$T^{(n)}_{Q,\widetilde{Q}}(i,j)\xrightarrow[n\to\infty]{\mathbb{P}^{Q}-a.s.}\mathbb{P}^{Q}(\vec{\mathcal{P}}[j]=i)^{-1}\mathbb{E}^{Q}\left[\mathbb{1}\{\vec{\mathcal{P}}[j]=i\}b(\vec{\mathcal{P}})\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}'=\vec{\mathcal{P}}\middle|\vec{\mathcal{P}}'[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}'=\vec{\mathcal{P}}\middle|\vec{\mathcal{P}}'[j]=i\right)}\right]$$

$$= \mathbb{E}^{Q}\left[b(\vec{\mathcal{P}}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)} \middle| \vec{\mathcal{P}}[j] = i \right],$$

where $\vec{\mathcal{P}}'$ is an independent copy of $\vec{\mathcal{P}}$. Now, using Lemma 35 with $f = b$ and recalling that $b(\vec{\mathcal{P}}) = b(\vec{\mathcal{P}}, S)$, we get that:

$$\mathbb{E}^{Q}\left[b(\vec{\mathcal{P}}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} | \vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} | \vec{\mathcal{P}}'[j] = i\right)} \middle| \vec{\mathcal{P}}[j] = i \right] = \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}}) | \vec{\mathcal{P}}[j] = i].$$

For the second part of the theorem, we note that

$$\begin{aligned} \Sigma^{(n)}_{Q,\widetilde{Q}}(i,j) &= \left(\frac{1}{|D^{(n)}_{(i,j)}|} \sum_{\vec{p} \in D^{(n)}_{(i,j)}} b^2(\vec{p}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}} = \vec{p} \middle| \vec{\mathcal{P}}[j] = i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}} = \vec{p} \middle| \vec{\mathcal{P}}[j] = i\right)} \right) - \left(T^{(n)}_{Q,\widetilde{Q}}(i,j) \right)^2 \\ &\xrightarrow[n\to\infty]{\mathbb{P}^Q a.s.} \mathbb{E}^{Q}\left[b^2(\vec{\mathcal{P}}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)} \middle| \vec{\mathcal{P}}[j] = i \right] - \left(\mathbb{E}^{\widetilde{Q}}\left[b(\vec{\mathcal{P}}) \middle| \vec{\mathcal{P}}[j] = i \right] \right)^2. \end{aligned}$$

Using again Lemma 35 but this time with $f = b^2$, we get that

$$\mathbb{E}^{Q}\left[b^2(\vec{\mathcal{P}}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)} \middle| \vec{\mathcal{P}}[j] = i \right] = \mathbb{E}^{\widetilde{Q}}\left[b^2(\vec{\mathcal{P}}) \middle| \vec{\mathcal{P}}[j] = i \right],$$

therefore

$$\Sigma_{Q,\widetilde{Q}}(i,j) \xrightarrow[n\to\infty]{\mathbb{P}^Q a.s.} \mathbb{E}^{\widetilde{Q}}\left[b^2(\vec{\mathcal{P}}) \middle| \vec{\mathcal{P}}[j] = i \right] - \left(\mathbb{E}^{\widetilde{Q}}\left[b(\vec{\mathcal{P}}) \middle| \vec{\mathcal{P}}[j] = i \right] \right)^2 = \mathbb{V}^{\widetilde{Q}}[b^2(\vec{\mathcal{P}}) | \vec{\mathcal{P}}[j] = i].$$

□

We obtain the following consequence of the Theorems 32 and 36.

Corollary 37. *Let $\widetilde{Q} = U$ be the transition matrix corresponding to the uniform transitions. If Q and $\widetilde{Q}$ are equivalent transition matrices then*

$$T^{(n)}_{Q,U}(i,j) - T^{(n)}_{Q,U}(i',j) \xrightarrow[n\to\infty]{\mathbb{P}^Q - a.s.} \mathbb{E}[S(i,j)] - \mathbb{E}[S(i',j)],$$

and

$$\Sigma^{(n)}_{Q,U}(i,j) - \Sigma^{(n)}_{Q,U}(i',j) \xrightarrow[n\to\infty]{\mathbb{P}^Q - a.s.} \mathbb{V}[S(i,j)] - \mathbb{V}[S(i',j)].$$

Remark 38. In the case $\widetilde{Q} = U$ the estimators can be written in a more explicit form. For example,

$$T^{(n)}_{Q,U}(i,j) = \frac{\prod_{\ell \neq j} (r_\ell)^{-1}}{|D^{(n)}_{(i,j)}|} \sum_{\vec{p} \in D^{(n)}_{(i,j)}} \left(b(\vec{p}) \frac{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}[j] = i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}} = \vec{p}\right)} \right).$$

2.5 Asymptotic distribution of the estimators

2.5.1 Known Q and $\widetilde{Q}$

In the case where both distributions, Q and $\widetilde{Q}$, are known, we deduce the asymptotic distribution of the estimators $T^{(n)}_{Q,\widetilde{Q}}(i,j)$ and $\Sigma^{(n)}_{Q,\widetilde{Q}}(i,j)$.

Theorem 39. *We have that*

$$\sqrt{|D^{(n)}_{(i,j)}|}(T^{(n)}_{Q,\widetilde{Q}}(i,j) - \mu_{i,j}) \xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}(0, \sigma^2_{i,j})$$

where

$$\mu_{i,j} = \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j] = i] \quad \text{and} \quad \sigma^2_{i,j} = \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})^2|\vec{\mathcal{P}}[j] = i] - \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j] = i]^2.$$

If we assume that the entries of S have finite forth moments, we have that

$$\sqrt{|D^{(n)}_{(i,j)}|}\left(\Sigma^{(n)}_{Q,\widetilde{Q}}(i,j) - \sigma^2_{i,j}\right) \xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}(0, (1,-1)\Sigma_{i,j}(1,-1)^T),$$

where $\Sigma_{i,j}$ equals to

$$\begin{bmatrix} \mathbb{V}^Q\left[b^2(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j]=i\right] & 2\mu_{i,j}\mathrm{Cov}^Q\left[b^2(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}}), b(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j]=i\right] \\ 2\mu_{i,j}\mathrm{Cov}^Q\left[b^2(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}}), b(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j]=i\right] & 4\mu^2_{i,j}\mathbb{V}^Q\left[b(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j]=i\right] \end{bmatrix}$$

and $C_{i,j}(\vec{p}) = \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}$.

Remark 40. The assumption of existence of second moments for S is natural because this guarantees the existence of a CLT for the first moments of its entries. In the same way, the existence of fourth moments guarantees the existence of a CLT for the variances of the entries of S.

Proof. Let us observe that

$$\begin{aligned} T^{(n)}_{Q,\widetilde{Q}}(i,j) &= \frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b(\vec{p})\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)} \\ &= \frac{1}{|D^{(n)}_{(i,j)}|}\sum_{k=1}^{n} b(\vec{\mathcal{P}}_k)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_k\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_k\middle|\vec{\mathcal{P}}[j]=i\right)}\mathbb{1}\left\{\vec{\mathcal{P}}[j]=i\right\}. \end{aligned}$$

The last sum can be interpreted as a sum of i.i.d. random variables appearing in an acceptance-rejection sampling. More precisely, we start with $\left(\vec{\mathcal{P}}_1, b(\vec{\mathcal{P}}_1)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_1\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_1\middle|\vec{\mathcal{P}}[j]=i\right)}\right)$ and $k = 1$; if $\vec{\mathcal{P}}_k[j] = i$ then we set

$$Y_{i,j} := b(\vec{\mathcal{P}}_k)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_k\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_k\middle|\vec{\mathcal{P}}[j]=i\right)} \tag{2.7}$$

and stop, otherwise we increase k and repeat until $\vec{\mathcal{P}}_K[j] = i$ for the first $K := \inf\{k|\vec{\mathcal{P}}_k[j] = i\}$. Now, letting

$$f(\vec{\mathcal{P}}_k) := b(\vec{\mathcal{P}}_k)\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_k\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^{Q}\left(\vec{\mathcal{P}}=\vec{\mathcal{P}}_k\middle|\vec{\mathcal{P}}[j]=i\right)} \tag{2.8}$$

we obtain for $y \in \mathbb{R}$ that

$$\mathbb{P}^Q[Y_{i,j} \leq y] = \sum_{k=1}^{\infty} \mathbb{P}^Q[f(\vec{\mathcal{P}}_k) \leq y | K = k]\mathbb{P}^Q[K = k] = \mathbb{P}^Q[f(\vec{\mathcal{P}}_1)|\vec{\mathcal{P}}_1[j] = i];$$

this means that the distribution of $Y_{i,j}$ equals the distribution of $f(\vec{\mathcal{P}}_1)$ conditioned on $\vec{\mathcal{P}}_1[j] = i$. Iterating this acceptance-rejection method, we see that $|D_{i,j}^{(n)}|$ describes the number of acceptances among $(\vec{\mathcal{P}}_k, f(\vec{\mathcal{P}}_k))$ for $1 \leq k \leq n$. Therefore, the estimator $T_{Q,\widetilde{Q}}^{(n)}$ has the same distribution as

$$\frac{1}{|D_{i,j}^{(n)}|} \sum_{k=1}^{|D_{i,j}^{(n)}|} Y_{i,j}^{(k)},$$

where $Y_{i,j}^{(k)}$, $k \in \mathbb{N}$ is a sequence of i.i.d. random variables distributed as $f(\vec{\mathcal{P}}_1)$ conditioned on $\vec{\mathcal{P}}_1[j] = i$. Finally, Anscombe's Theorem [Gut, 2009] implies that

$$\sqrt{|D_{(i,j)}^{(n)}|}(T_{Q,\widetilde{Q}}^{(n)}(i,j) - \mu_{i,j}) \xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}(0, \sigma_{i,j}) \tag{2.9}$$

where

$$\mu_{i,j} = \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}}_1)|\vec{\mathcal{P}}_1[j] = i] \quad \text{and} \quad \sigma_{i,j}^2 = \mathbb{E}^{\widetilde{Q}}[f(\vec{\mathcal{P}})^2|\vec{\mathcal{P}}[j] = i] - \mathbb{E}^{\widetilde{Q}}[f(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j] = i]^2.$$

For the variance, we assume that the entries of S have finite fourth moments. The estimator for the variance has the following form:

$$\Sigma_{Q,\widetilde{Q}}^{(n)}(i,j) = \left(\frac{1}{|D_{(i,j)}^{(n)}|} \sum_{\vec{p} \in D_{(i,j)}^{(n)}} b^2(\vec{p}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}} = \vec{p} \middle| \vec{\mathcal{P}}[j] = i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}} = \vec{p} \middle| \vec{\mathcal{P}}[j] = i\right)} \right) - \left(T_{Q,\widetilde{Q}}^{(n)}(i,j)\right)^2; \tag{2.10}$$

We consider the following i.i.d. vectors $\vec{Y}_{i,j}^{(k)} = (Y_{1,k}, Y_{2,k})$ where $\vec{Y}_{i,j}^{(k)}$ is distributed as

$$\left(b^2(\vec{\mathcal{P}}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)}, b(\vec{\mathcal{P}}) \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}' = \vec{\mathcal{P}} \middle| \vec{\mathcal{P}}'[j] = i\right)} \right)$$

under $\mathbb{P}^Q$ conditioned on $\vec{\mathcal{P}}'[j] = i$. We write $C_{i,j}(\vec{p}) := \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}$ and, due to Lemma 35, for $k \in \{1, 2\}$

$$\mu_{i,j}^{(k)} := \mathbb{E}^Q[b(\vec{\mathcal{P}})^k C_{i,j}(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j] = i] = \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})^k|\vec{\mathcal{P}}[j] = i], \tag{2.11}$$

and covariance matrix

$$\Sigma'_{i,j} = \begin{bmatrix} \mathbb{V}^Q\left[b^2(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j] = i\right] & \mathrm{Cov}^Q\left[b^2(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}}), b(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j] = i\right] \\ \mathrm{Cov}^Q\left[b^2(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}}), b(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j] = i\right] & \mathbb{V}^Q\left[b(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})\middle|\vec{\mathcal{P}}[j] = i\right] \end{bmatrix}. \tag{2.12}$$

We conclude from Theorem 48 that

$$\sqrt{|D_{(i,j)}^{(n)}|}\left(\tfrac{1}{|D_{(i,j)}^{(n)}|} \sum_{\vec{p} \in D_{(i,j)}^{(n)}} \left(b^2(\vec{p})C_{i,j}(\vec{p}) - \mu_{i,j}^{(2)}\right), \tfrac{1}{|D_{(i,j)}^{(n)}|} \sum_{\vec{p} \in D_{(i,j)}^{(n)}} \left(b(\vec{p})C_{i,j}(\vec{p}) - \mu_{i,j}^{(1)}\right) \right) \xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}(0, \Sigma'_{i,j}). \tag{2.13}$$

Plugging in the definition of $T^{(n)}_{Q,\widetilde{Q}}(i,j)$ and using Anscombe's delta method (see Theorem 49) with the function $h(x,y) = \left(x, \left(y+\mu^{(1)}_{i,j}\right)^2 - \mu^{(1)}_{i,j}\right)$ we obtain

$$\sqrt{|D^{(n)}_{(i,j)}|}\left(\frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}}\left(b^2(\vec{p})C_{i,j}(\vec{p}) - \mu^{(2)}_{i,j}\right), \left(T^{(n)}_{Q,\widetilde{Q}}(i,j)\right)^2 - {\mu^{(1)}_{i,j}}^2\right)$$
$$\xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}\left(0, \left(\nabla h(\vec{0})\right)^T \cdot \Sigma' \cdot \nabla h(\vec{0})\right), \tag{2.14}$$

where

$$\nabla h(\vec{0}) = J_h(\vec{0}) = \begin{bmatrix} 1 & 0 \\ 0 & 2\mu^{(1)}_{i,j} \end{bmatrix}. \tag{2.15}$$

Noting that $\mu^{(1)}_{i,j} = \mu_{i,j}$, by Equation (2.11), the latter results in

$$\Sigma_{i,j} := \left(\nabla h(\vec{0})\right)^T \cdot \Sigma' \cdot \nabla h(\vec{0}) = \begin{bmatrix} \Sigma'_{i,j}(1,1) & 2\mu_{i,j}\Sigma'_{i,j}(1,2) \\ 2\mu_{i,j}\Sigma'_{i,j}(2,1) & 4\mu^2_{i,j}\Sigma'_{i,j}(2,2) \end{bmatrix}. \tag{2.16}$$

Hence, by the continuous mapping theorem with the function $f(x,y) = x - y$ we get that

$$\sqrt{|D^{(n)}_{(i,j)}|}\left(\Sigma_{Q,\widetilde{Q}}(i,j) - \sigma^2_{i,j}\right) \xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}(0, (1,-1)\cdot\Sigma'_{i,j}\cdot(1,-1)^T),$$

which concludes the proof. □

2.5.2 Unknown Q and known $\widetilde{Q}$

If Q is unknown, we don't know a priori if it is equivalent to a matrix with a desirable form. We can deduce this by looking at the generated data: if all possible transitions in the network appear at least once, the unknown matrix Q is equivalent to a uniform transition matrix. If Q is not equivalent to a uniform matrix, we can still apply the method described in Remark 34, which is less restrictive than a uniform transition matrix.

When Q is unknown, we have to replace the probabilities $\mathbb{P}^Q\left(\vec{\mathcal{P}} = \vec{p} \middle| \vec{\mathcal{P}}[j] = i\right)$ by estimators. Define

$$\widehat{C}_{i,j}(\vec{p},n) := \mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}} = \vec{p} \middle| \vec{\mathcal{P}}[j] = i\right)\left(\frac{\sum_{k=1}^n \mathbb{1}\{\vec{p}_k = \vec{p}\}}{\sum_{k=1}^n \mathbb{1}\{\vec{p}_k[j] = i\}}\right)^{-1} \tag{2.17}$$

which is a random variable that converges $\mathbb{P}^Q$-a.s. to $C_{i,j}(\vec{p}) = \frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}=\vec{p}\middle|\vec{\mathcal{P}}[j]=i\right)}$. Therefore, we propose the following variation of $T^{(n)}_{Q,\widetilde{Q}}(i,j)$:

$$\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j) := \frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b(\vec{p})\widehat{C}_{i,j}(\vec{p},n). \tag{2.18}$$

The estimator $\Sigma^{(n)}_{Q,\widetilde{Q}}(i,j)$ for the variance becomes:

$$\widehat{\Sigma}_{Q,\widetilde{Q}}(i,j) = \left(\frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b^2(\vec{p})\widehat{C}(\vec{p},n)\right) - \left(\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j)\right)^2.$$

Lemma 41. *Let $Q, \widehat{Q}$ be two equivalent transition matrices then for all i, j, we have that*

$$\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j) - \mu_{i,j} \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} 0;$$
$$\widehat{\Sigma}_{Q,\widetilde{Q}}(i,j) - \sigma^2_{i,j} \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} 0,$$

where $\mu_{i,j} = \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}}_1)|\vec{\mathcal{P}}_1[j] = i]$ *and* $\sigma^2_{i,j} = \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})^2|\vec{\mathcal{P}}[j] = i] - \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j] = i]^2$.

Proof. Let $q_1, q_2, \dots, q_m$ be the elements of all pairwise different paths such that $q[j] = i$ and let $D^{(n)}_{(i,j)} = \{p_1, p_2, \dots, p_n\}$, then

$$\begin{aligned}\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j) &= \frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b(\vec{p})\widehat{C}_{i,j}(\vec{p},n) \\ &= \frac{n}{D^{(n)}_{(i,j)}}\sum_{l=1}^{m}\widehat{C}_{i,j}(q_l,n)\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}\{p_k = q_l\}b(p_k,S_k)\right).\end{aligned} \tag{2.19}$$

By the strong law of large numbers, we get that

$$\begin{aligned}&\frac{n}{D^{(n)}_{(i,j)}} \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} \left(\mathbb{P}^Q[\vec{\mathcal{P}}[j] = i]\right)^{-1};\\ &\widehat{C}_{i,j}(q_l,n) \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} C_{i,j}(q_l);\\ &\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}\{p_k = q_l\}b(p_k,S_k) \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} \mathbb{E}^Q[\mathbb{1}\{\vec{\mathcal{P}} = q_l\}b(\vec{\mathcal{P}},S)].\end{aligned} \tag{2.20}$$

Since m is a finite constant, we obtain, using the continuous mapping theorem and Lemma 35 in the last equality, that

$$\begin{aligned}\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j) &\xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} \left(\mathbb{P}^Q[\vec{\mathcal{P}}[j] = i]\right)^{-1}\sum_{l=1}^{m}C_{i,j}(q_l)\mathbb{E}^Q[\mathbb{1}\{\vec{\mathcal{P}} = q_l\}b(\vec{\mathcal{P}},S)]\\ &= \left(\mathbb{P}^Q[\vec{\mathcal{P}}[j] = i]\right)^{-1}\sum_{l=1}^{m}\mathbb{E}^Q[\mathbb{1}\{\vec{\mathcal{P}} = q_l\}b(\vec{\mathcal{P}},S)C_{i,j}(\vec{\mathcal{P}})]\\ &= \left(\mathbb{P}^Q[\vec{\mathcal{P}}[j] = i]\right)^{-1}\mathbb{E}^Q[\mathbb{1}\{\vec{\mathcal{P}}[j] = i\}b(\vec{\mathcal{P}},S)C_{i,j}(\vec{\mathcal{P}})]\\ &= \mathbb{E}^Q[b(\vec{\mathcal{P}},S)C_{i,j}(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j] = i]\\ &= \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}},S)|\vec{\mathcal{P}}[j] = i] = \mu_{i,j}.\end{aligned} \tag{2.21}$$

For the second part of the statement, we obtain similarly, that

$$\left(\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j)\right)^2 \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} \mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}},S)|\vec{\mathcal{P}}[j] = i]^2. \tag{2.22}$$

Furthermore,

$$\frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b^2(\vec{p})\widehat{C}(\vec{p},n) = \frac{n}{D^{(n)}_{(i,j)}}\sum_{l=1}^{m}\widehat{C}(q_l,n)\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}\{p_k = q_l\}b^2(p_k,S_k)\right) \tag{2.23}$$

In the same fashion as in (2.20) and (2.21) we get that

$$\frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b^2(\vec{p})\widehat{C}(\vec{p},n) \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} \mathbb{E}^{\widetilde{Q}}[b^2(\vec{\mathcal{P}},S)|\vec{\mathcal{P}}[j]=i]. \tag{2.24}$$

Hence, using (2.22) and (2.24) we conclude that

$$\widehat{\Sigma}_{Q,\widetilde{Q}}(i,j) = \left(\frac{1}{|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}} b^2(\vec{p})\widehat{C}(\vec{p},n)\right) - \left(\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j)\right)^2 \xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s} \sigma^2_{i,j}. \tag{2.25}$$

□

Theorem 42. *Let $Q,\widehat{Q}$ be two equivalent transition matrices then for all i,j, we have that*

$$\sqrt{|D^{(n)}_{(i,j)}|}(\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j)-\mu_{i,j}) \xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}(0,\vec{C}\Sigma_{i,j}\vec{C}^T)$$

where $\mu_{i,j}=\mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i]$ and

$$\Sigma_{i,j}(\ell,\ell') = \begin{cases} \mathbb{P}^Q\left[\vec{\mathcal{P}}=\vec{q}_\ell|\vec{\mathcal{P}}[j]=i\right]\mathbb{V}^Q\left[b\left(\vec{q}_\ell,S_{\vec{\mathcal{P}}}\right)|\vec{\mathcal{P}}[j]=i\right], & \ell=\ell'; \\ 0, & otherwise, \end{cases}$$

for $\ell,\ell'\in\{1,\dots,m\}$ and $\vec{C}=(C_{i,j}(\vec{q}_1),\dots,C_{i,j}(\vec{q}_m))$ and $C_{i,j}(\vec{p})=\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}=\vec{p}|\vec{\mathcal{P}}[j]=i\right)}$. If we assume that the entries of S have finite forth moments, we have that

$$\sqrt{|D^{(n)}_{(i,j)}|}\left(\widehat{\Sigma}^{(n)}_{Q,\widetilde{Q}}(i,j)-\sigma^2_{i,j}\right) \xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.} \mathcal{N}(0,\vec{C}'\Sigma'_{i,j}\vec{C}'^T),$$

where $\sigma^2_{i,j}=\mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})^2|\vec{\mathcal{P}}[j]=i]-\mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i]^2$ and

$$\Sigma'_{i,j}(\ell,\ell') = \begin{cases} \mathbb{P}^Q\left[\vec{\mathcal{P}}=\vec{q}_\ell|\vec{\mathcal{P}}[j]=i\right]\mathbb{V}^Q\left[b(\vec{q}_\ell)|\vec{\mathcal{P}}[j]=i\right], & \ell=\ell'\in\{1,\dots,m\}; \\ \mathbb{P}^Q\left[\vec{\mathcal{P}}=\vec{q}_\ell|\vec{\mathcal{P}}[j]=i\right]\mathbb{V}^Q\left[b^2(\vec{q}_l)|\vec{\mathcal{P}}[j]=i\right], & \ell=\ell'\in\{m+1,\dots,2m\}; \\ \mathbb{P}^Q\left[\vec{\mathcal{P}}=\vec{q}_\ell|\vec{\mathcal{P}}[j]=i\right]\mathrm{Cov}^Q\left[b^2(\vec{q}_\ell),b(\vec{q}_\ell)|\vec{\mathcal{P}}[j]=i\right], & \ell\in\{1,\dots m\},\ell`=\ell+m; \\ \mathbb{P}^Q\left[\vec{\mathcal{P}}=\vec{q}'_\ell|\vec{\mathcal{P}}[j]=i\right]\mathrm{Cov}^Q\left[b^2(\vec{q}'_\ell),b(\vec{q}'_\ell)|\vec{\mathcal{P}}[j]=i\right], & \ell'\in\{1,\dots m\},\ell=\ell'+m; \\ 0, & otherwise, \end{cases} \tag{2.26}$$

for $\ell,\ell'\in\{1,\dots,2m\}$ and $\vec{C}'=(2\mu_{i,j}C_{i,j}(\vec{q}_1),\dots,2\mu_{i,j}C_{i,j}(\vec{q}_m),C_{i,j}(\vec{q}_1),\dots,C_{i,j}(\vec{q}_m))$.

Proof. The key result to get the asymptotic distribution in this case is the multi-dimensional Anscombe's theorem, Theorem 48 in Appendix 2.6. In order to apply this result we do some preparations. Using Lemma 35, we write

$$\mu_{i,j}=\mathbb{E}^{\widetilde{Q}}[b(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i]=\mathbb{E}^Q[b(\vec{\mathcal{P}})C_{i,j}(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i],$$

with $C_{i,j}(\vec{\mathcal{P}})=\frac{\mathbb{P}^{\widetilde{Q}}\left(\vec{\mathcal{P}}'=\vec{\mathcal{P}}|\vec{\mathcal{P}}'[j]=i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}'=\vec{\mathcal{P}}|\vec{\mathcal{P}}'[j]=i\right)}$. Let $\vec{q}_1,\vec{q}_2,\dots,\vec{q}_m$ be all the pairwise different paths such that $\mathrm{Supp}^Q_{i,j}$. Notice that in fact $b(\vec{\mathcal{P}})=b(\vec{\mathcal{P}},S)$ and that $\vec{\mathcal{P}}$ and S are drawn from a product measure.

Hence, it makes sense to write also $b(\vec{q}) = b(\vec{q}, S)$ for a given path $\vec{q}$. For any realization $(\vec{p}_n, S_n)$ of $\mathbb{P}^Q$ or $\mathbb{P}^{\widetilde{Q}}$ we also write $b(\vec{p}_n) = b(\vec{p}_n, S_n)$. We can decompose $\mu_{i,j}$ using all possible paths:

$$\mu_{i,j} = \sum_{\ell=1}^{m} \mathbb{E}^Q[b(\vec{q}_\ell)C_{i,j}(\vec{q}_\ell)\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\}|\vec{\mathcal{P}}[j] = i]. \tag{2.27}$$

Let us write $\tilde{\mu}_\ell = \mathbb{E}^Q[b(\vec{q}_\ell)C_{i,j}(\vec{q}_\ell)\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\}|\vec{\mathcal{P}}[j] = i]$ and $\mu_\ell = \mathbb{E}^Q[b(\vec{q}_\ell)C_{i,j}(\vec{q}_\ell)|\vec{\mathcal{P}}[j] = i]$, and consider

$$\begin{aligned}
\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j) - \mu_{i,j} &= \left(\frac{1}{|D^{(n)}_{(i,j)}|} \sum_{\vec{p} \in D^{(n)}_{(i,j)}} b(\vec{p})\widehat{C}_{i,j}(\vec{p}, n) \right) - \mu_{i,j} \\
&= \frac{1}{|D^{(n)}_{(i,j)}|} \sum_{\ell=1}^{m} \widehat{C}_{i,j}(\vec{q}_l, n) \left(\sum_{\vec{p} \in D^{(n)}_{(i,j)}} \mathbb{1}\{\vec{p} = \vec{q}_\ell\} b(\vec{p}) \right) - \sum_{\ell=1}^{m} \tilde{\mu}_\ell \\
&= \frac{1}{|D^{(n)}_{(i,j)}|} \sum_{\ell=1}^{m} \widehat{C}_{i,j}(\vec{q}_\ell, n) \left(\sum_{\vec{p} \in D^{(n)}_{(i,j)}} \mathbb{1}\{\vec{p} = \vec{q}_\ell\} \left(b(\vec{q}_\ell) - \mathbb{E}^Q[b(q_\ell)|\vec{\mathcal{P}}[j] = i] \right) \right) \\
&\quad + \sum_{\ell=1}^{m} \left(\widehat{C}_{i,j}(\vec{q}_\ell, n) \frac{|D^{(n)}_\ell|}{|D^{(n)}_{(i,j)}|} \mathbb{E}^Q[b(q_\ell)|\vec{\mathcal{P}}[j] = i] \right) - \sum_{\ell=1}^{m} \tilde{\mu}_\ell
\end{aligned}$$

with $|D^{(n)}_\ell| = \sum_{k=1}^n \mathbb{1}\{\vec{p}_k = \vec{q}_\ell\}$ for $\ell \in \{1, \dots, m\}$. By independence, we deduce for every $\ell \in \{1, \dots, m\}$ that

$$\begin{aligned}
\tilde{\mu}_\ell &= \mathbb{E}^Q[b(\vec{q}_\ell)C_{i,j}(\vec{q}_\ell)\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\}|\vec{\mathcal{P}}[j] = i] \\
&= \mathbb{E}^Q[b(\vec{q}_\ell)|\vec{\mathcal{P}}[j] = i] \cdot C_{i,j}(\vec{q}_\ell) \cdot \mathbb{P}^Q[\vec{\mathcal{P}} = \vec{q}_\ell|\vec{\mathcal{P}}[j] = i] \\
&= \mathbb{E}^Q[b(\vec{q}_\ell)|\vec{\mathcal{P}}[j] = i] \cdot \mathbb{P}^{\widetilde{Q}}[\vec{\mathcal{P}} = \vec{q}_\ell|\vec{\mathcal{P}}[j] = i].
\end{aligned} \tag{2.28}$$

Moreover, recalling that $\widehat{C}_{i,j}(\vec{q}_\ell, n) := \mathbb{P}^{\widetilde{Q}} \left(\vec{\mathcal{P}} = \vec{q}_\ell \middle| \vec{\mathcal{P}}[j] = i \right) \frac{|D^{(n)}_{(i,j)}|}{|D^{(n)}_\ell|}$, it follows from the previous equality that

$$\widehat{C}_{i,j}(\vec{q}_\ell, n) \frac{|D^{(n)}_\ell|}{|D^{(n)}_{(i,j)}|} \mathbb{E}^Q[b(q_\ell)|\vec{\mathcal{P}}[j] = i] = \mathbb{P}^{\widetilde{Q}} \left[\vec{\mathcal{P}} = \vec{q}_\ell \middle| \vec{\mathcal{P}}[j] = i \right] \mathbb{E}^Q[b(q_\ell)|\vec{\mathcal{P}}[j] = i] = \tilde{\mu}_\ell \tag{2.29}$$

for every $\ell \in \{1, \dots, m\}$, implying that

$$\widehat{T}^{(n)}_{Q,\widetilde{Q}}(i,j) - \mu_{i,j} = \sum_{\ell=1}^{m} \widehat{C}_{i,j}(\vec{q}_\ell, n) \left(\frac{1}{|D^{(n)}_{(i,j)}|} \sum_{\vec{p} \in D^{(n)}_{(i,j)}} \mathbb{1}\{\vec{p} = \vec{q}_\ell\} \left(b(q_\ell) - \mathbb{E}^Q[b(q_\ell)|\vec{\mathcal{P}}[j] = i] \right) \right). \tag{2.30}$$

Since $\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\}$ and $b(q_\ell) = b(q_\ell, S)$ are independent we also have that

$$\begin{aligned}
\mathbb{E}^Q \left[\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\} \mathbb{E}^Q \left[b(q_\ell)|\vec{\mathcal{P}}[j] = i \right] |\vec{\mathcal{P}}[j] = i \right] &= \mathbb{E}^Q \left[b(q_\ell)|\vec{\mathcal{P}}[j] = i \right] \mathbb{E}^Q \left[\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\}|\vec{\mathcal{P}}[j] = i \right] \\
&= \mathbb{E}^Q \left[b(q_\ell)|\vec{\mathcal{P}}[j] = i \right] \mathbb{E}^Q \left[\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\}|\vec{\mathcal{P}}[j] = i \right] \\
&= \mathbb{E}^Q \left[\mathbb{1}\{\vec{\mathcal{P}} = \vec{q}_\ell\} b(q_\ell)|\vec{\mathcal{P}}[j] = i \right]
\end{aligned} \tag{2.31}$$

Hence,

$$\mathbb{E}^{Q}\left[\mathbb{1}\{\vec{\mathcal{P}}=\vec{q}_\ell\}\left(b(q_\ell)-\mathbb{E}^{Q}\left[b(q_\ell)|\vec{\mathcal{P}}[j]=i\right]\right)|\vec{\mathcal{P}}[j]=i\right]=0, \quad \text{for all } \ell=1,\dots,m. \tag{2.32}$$

We now define the random variables Y in order to apply Theorem 48. In the following we assume all random variables to be distributed under $\mathbb{P}^Q$ conditioned on $\vec{\mathcal{P}}[j]=i$. Define for $k\in\{1,\dots,n\}$,

$$\Xi^{(k)}=\sum_{\ell=1}^{m}\ell\cdot\mathbb{1}\{\vec{\mathcal{P}}_k=\vec{q}_\ell\}$$

the variable indicating which path passing through (i,j) is chosen. For $\ell\in\{1,\dots,m\}$ define

$$Y_\ell^{(k)}=\mathbb{1}\{\Xi^{(k)}=\ell\}\left(b(q_\ell)-\mathbb{E}^{Q}[b(q_\ell)|\vec{\mathcal{P}}[j]=i]\right) \tag{2.33}$$

and

$$Y^{(k)}=\left(Y_1^{(k)},\dots,Y_m^{(k)}\right)$$

which by Equation 2.32 satisfies

$$\mathbb{E}^{Q}\left[Y^{(k)}|\vec{\mathcal{P}}[j]=i\right]=\vec{0}.$$

Furthermore, we define

$$N(n)=|D_{(i,j)}^{(n)}|.$$

Then we have that,

$$\frac{N(n)}{n}\xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s.}\theta>0.$$

We define for $\ell\in\{1,\dots,m\}$

$$S_{N(n)}^{(\ell)}=\sum_{k=1}^{N(n)}Y_\ell^{(k)}.$$

Then, by Theorem 48, we get that

$$S_{N(n)}:=\frac{1}{\sqrt{N(n)}}\left(S_{N(n))}^{(1)},\dots,S_{N(n))}^{(m)}\right)\xrightarrow[n\to\infty]{\mathbb{P}^Q-distr.}\mathcal{N}(0,\Sigma) \tag{2.34}$$

where

$$\Sigma(\ell,\ell')=\begin{cases}\mathbb{P}^{Q}\left[\vec{\mathcal{P}}=\vec{q}_\ell|\vec{\mathcal{P}}[j]=i\right]\mathbb{V}^{Q}\left[b\left(\vec{q}_\ell\right)|\vec{\mathcal{P}}[j]=i\right] & \ell=\ell';\\ 0 & \text{otherwise.}\end{cases} \tag{2.35}$$

We also know that $\widehat{C}_{i,j}(\vec{q}_\ell,n)$ converges almost surely to the constant $C(\vec{q}_\ell)$ for every $\ell=1,\dots,m$ and therefore,

$$\left(\widehat{C}_{i,j}(\vec{q}_1,n),\dots,\widehat{C}_{i,j}(\vec{q}_m,n)\right)\xrightarrow[n\to\infty]{\mathbb{P}^Q-a.s}\vec{C}=\left(C_{i,j}(\vec{q}_1),\dots,C_{i,j}(\vec{q}_m)\right). \tag{2.36}$$

Hence, by the multidimensional Slutsky's Theorem (Lemma 2.8 in [Van der Vaart, 2000]), the continuous mapping theorem, and the fact that a linear transformation of a gaussian vector is again gaussian we have by Equation (2.30) that

$$\frac{1}{\sqrt{|D_{(i,j)}^{(n)}|}}\left(\widehat{T}_{Q,\widetilde{Q}}^{(n)}(i,j)-\mu_{i,j}\right)\xrightarrow[n\to\infty]{\mathbb{P}-distr.}\mathcal{N}\left(0,\vec{C}\cdot\Sigma\cdot\vec{C}^T\right), \tag{2.37}$$

which concludes the first part of the proof.

For the second estimator we proceed in a similar way. As in Equation (2.27), we get that

$$\mathbb{E}^{\tilde{Q}}\left[b^2(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i\right]=\sum_{l=1}^{m}\mathbb{P}^{\tilde{Q}}\left[\vec{\mathcal{P}}=q_\ell|\vec{\mathcal{P}}[j]=i\right]\mathbb{E}^{Q}\left[b^2(q_\ell)|\vec{\mathcal{P}}[j]=i\right], \tag{2.38}$$

and we also see that

$$\begin{aligned}\frac{1}{|D_{i,j}^{(n)}|}\sum_{\vec{p}\in D_{i,j}^{(n)}}b^2(\vec{p})\hat{C}(\vec{p},n)=&\sum_{l=1}^{m}\hat{C}(q_\ell,n)\left(\frac{1}{|D_{i,j}^{(n)}|}\sum_{\vec{p}\in D_{i,j}^{(n)}}\mathbb{1}\{\vec{p}=q_\ell\}\left(b^2(q_\ell)-\mathbb{E}^{Q}\left[b^2(q_\ell)|\vec{\mathcal{P}}[j]=i\right]\right)\right)\\&+\sum_{l=1}^{m}\frac{|D_l^{(n)}|}{|D_{i,j}^{(n)}|}\hat{C}(q_\ell,n)\mathbb{E}^{Q}\left[b^2\left(q_\ell\right)|\vec{\mathcal{P}}[j]=i\right]\end{aligned} \tag{2.39}$$

where

$$\frac{|D_l^{(n)}|}{|D_{i,j}^{(n)}|}\hat{C}(q_\ell,n)\mathbb{E}^{Q}\left[b^2\left(q_\ell\right)|\vec{\mathcal{P}}[j]=i\right]=\mathbb{P}^{\tilde{Q}}\left[\vec{\mathcal{P}}=q_\ell|\vec{\mathcal{P}}[j]=i\right]\mathbb{E}^{Q}\left[b^2(q_\ell)|\vec{\mathcal{P}}[j]=i\right]. \tag{2.40}$$

Equations (2.38) and (2.40) imply that

$$\begin{aligned}&\frac{1}{|D_{i,j}^{(n)}|}\sum_{\vec{p}\in D_{i,j}^{(n)}}\left(b^2(\vec{p})\hat{C}(\vec{p},n)\right)-\mathbb{E}^{\tilde{Q}}\left[b^2(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i\right]=\\&\sum_{l=1}^{m}\hat{C}(q_\ell,n)\left(\frac{1}{|D_{i,j}^{(n)}|}\sum_{\vec{p}\in D_{i,j}^{(n)}}\mathbb{1}\{\vec{p}=q_\ell\}\left(b^2(q_\ell)-\mathbb{E}^{Q}\left[b^2(q_\ell,S_{\vec{\mathcal{P}}})|\vec{\mathcal{P}}[j]=i\right]\right)\right)\end{aligned} \tag{2.41}$$

which implies that

$$\begin{aligned}&\Sigma_{Q,\tilde{Q}}^{(n)}(i,j)-\sigma_{i,j}^2=\frac{1}{|D_{i,j}^{(n)}|}\sum_{\vec{p}\in D_{i,j}^{(n)}}\left(b^2(\vec{p})\hat{C}(\vec{p},n)\right)-\mathbb{E}^{\tilde{Q}}\left[b^2(\vec{\mathcal{P}})|\vec{\mathcal{P}}[j]=i\right]+\left(\widehat{T}_{Q,\tilde{Q}}^{(n)}(i,j)\right)^2-\mu_{i,j}^2\\&=\sum_{l=1}^{m}\hat{C}(q_\ell,n)\left(\frac{1}{|D_{i,j}^{(n)}|}\sum_{\vec{p}\in D_{i,j}^{(n)}}\mathbb{1}\{\vec{p}=q_\ell\}\left(b^2(q_\ell)-\mathbb{E}^{Q}\left[b^2(q_\ell,)|\vec{\mathcal{P}}[j]=i\right]\right)\right)\\&+\left(\sum_{\ell=1}^{m}\widehat{C}_{i,j}(\vec{q}_\ell,n)\left(\frac{1}{|D_{(i,j)}^{(n)}|}\sum_{\vec{p}\in D_{(i,j)}^{(n)}}\mathbb{1}\{\vec{p}=\vec{q}_\ell\}\left(b(q_\ell)-\mathbb{E}^{Q}[b(q_\ell)|\vec{\mathcal{P}}[j]=i]\right)\right)\right)\left(\widehat{T}_{Q,\tilde{Q}}^{(n)}(i,j)+\mu_{i,j}\right)\end{aligned} \tag{2.42}$$

where we use Equation (2.30) in the second line of the last equality. We are now almost ready to apply Theorem 48. In the following we assume all random variables to be distributed under $\mathbb{P}^Q$ conditioned on $\vec{\mathcal{P}}[j]=i$. For $\ell\in\{1,\dots,m\}$ define

$$W_\ell^{(k)}=\mathbb{1}\{\Xi^{(k)}=\ell\}\left(b(q_\ell)-\mathbb{E}^{Q}[b(q_\ell)|\vec{\mathcal{P}}[j]=i]\right) \tag{2.43}$$

where by Equation 2.32, we have for all $\ell=1,\dots,m$ that

$$\mathbb{E}^{Q}\left[W_\ell^{(k)}|\vec{\mathcal{P}}[j]=i\right]=0.$$

Furthermore, for $l = m+1, \dots, 2m$, define

$$W_\ell^{(k)} = \mathbb{1}\{\Xi^{(k)} = \ell - m\} \left(b^2(q_{\ell-m}) - \mathbb{E}^Q \left[b^2(q_{\ell-m}) | \vec{\mathcal{P}}[j] = i \right] \right)$$

where for all $\ell = m+1, \dots, 2m$ we get that $\mathbb{E}^Q \left[W_\ell^{(k)} \right]$ equals to

$$\mathbb{E}^Q \left[\mathbb{1}\{\vec{\mathcal{P}} = q_{\ell-m}\} b^2(q_{\ell-m}) | \vec{\mathcal{P}}[j] = i \right] - \mathbb{E}^Q \left[b^2(q_{\ell-m}) | \vec{\mathcal{P}}[j] = i \right] \mathbb{E}^Q \left[\mathbb{1}\{\vec{\mathcal{P}} = q_{\ell-m}\} | \vec{\mathcal{P}}[j] = i \right] = 0.$$

We then define

$$W^{(k)} = \left(W_1^{(k)}, \dots, W_{2m}^{(k)} \right)$$

which by the previous calculations satisfies for all $k = 1, \dots, n$ that

$$\mathbb{E}^Q \left[W^{(k)} | \vec{\mathcal{P}}[j] = i \right] = \vec{0}.$$

We define for $\ell \in \{1, \dots, 2m\}$

$$S_{N(n)}^{\prime(\ell)} = \sum_{k=1}^{N(n)} W_\ell^{(k)},$$

where $N(n) = |D_{(i,j)}^{(n)}|$. It follows from Theorem 48 that

$$S'_{N(n)} := \frac{1}{\sqrt{N(n)}} \left(S_{N(n))}^{(1)'}, \dots, S_{N(n))}^{(2m)'} \right) \xrightarrow[n \to \infty]{\mathbb{P}^Q - distr.} \mathcal{N}(0, \Sigma'') = (S'_1, \dots, S'_{2m}) \tag{2.44}$$

where the covariance matrix is given by

$$\Sigma'_{i,j}(\ell, \ell') = \begin{cases} \mathbb{P}^Q \left[\vec{\mathcal{P}} = \vec{q}_\ell | \vec{\mathcal{P}}[j] = i \right] \mathbb{V}^Q \left[b(\vec{q}_\ell) | \vec{\mathcal{P}}[j] = i \right], & \ell = \ell' \in \{1, \dots, m\}; \\ \mathbb{P}^Q \left[\vec{\mathcal{P}} = \vec{q}_\ell | \vec{\mathcal{P}}[j] = i \right] \mathbb{V}^Q \left[b^2(\vec{q}_l) | \vec{\mathcal{P}}[j] = i \right], & \ell = \ell' \in \{m+1, \dots, 2m\}; \\ \mathbb{P}^Q \left[\vec{\mathcal{P}} = \vec{q}_\ell | \vec{\mathcal{P}}[j] = i \right] \mathrm{Cov}^Q \left[b^2(\vec{q}_\ell), b(\vec{q}_\ell) | \vec{\mathcal{P}}[j] = i \right], & \ell \in \{1, \dots m\}, \ell‘ = \ell + m; \\ \mathbb{P}^Q \left[\vec{\mathcal{P}} = \vec{q}_{\ell'} | \vec{\mathcal{P}}[j] = i \right] \mathrm{Cov}^Q \left[b^2(\vec{q}_{\ell'}), b(\vec{q}_{\ell'}) | \vec{\mathcal{P}}[j] = i \right], & \ell' \in \{1, \dots m\}, \ell = \ell' + m; \\ 0, & \text{otherwise.} \end{cases} , \tag{2.45}$$

We also know that the vector

$$\vec{C}^{(n)'} = \left(\hat{C}_{Q,\tilde{Q}}(q_1, n), \dots, \hat{C}_{Q,\tilde{Q}}(q_m, n), \widehat{T}_{Q,\tilde{Q}}^{(n)}(i,j) + \mu_{i,j} \right) \xrightarrow[n \to \infty]{\mathbb{P}^Q - a.s} \vec{C}' := (C_{i,j}(\vec{q}_1), \dots, C_{i,j}(\vec{q}_m), 2\mu_{i,j}).$$

The remaining part of the proof is similar to the end of the first part of the proof. □

2.6 Examples

The first example illustrates how we can unbias estimates if the covariates are dependent; we use simulation data to compare the estimates with the “real” values[2] .

[2] An implementation of the estimators in the language R together with the following examples can be found at `https://github.com/NaitsabesMue/MarkovLinearModel`

Example 43 (Unbiasing). Let us consider a DAG with 2 lines and 2 columns, and a random matrix S with first moments

$$\mathbb{E}[S] = \begin{pmatrix} 0 & 1 \\ -2 & 2 \end{pmatrix}$$

The transition Q of the Markov chain are given as

$$q^{(1)}(s,1) = q^{(1)}(s,2) = 1/2,$$

$$q^{(2)}(1,1) = 3/4,\ q^{(2)}(1,2) = 1/4,\ q^{(2)}(2,1) = 1/4,\ \text{and}\ q^{(2)}(2,2) = 3/4.$$

We calculate

$$\mathbb{E}^Q[b(\vec{p})|\vec{p}[1] = 1] = \frac{3}{4}\cdot 1 + \frac{1}{4}\cdot 2 = \frac{5}{4}$$

$$\mathbb{E}^Q[b(\vec{p})|\vec{p}[1] = 2] = \frac{1}{4}\cdot(-1) + \frac{3}{4}\cdot 0 = -\frac{1}{4}$$

and

$$\mathbb{E}^Q[b(\vec{p})|\vec{p}[2] = 1] = \frac{3}{4}\cdot 1 + \frac{1}{4}\cdot(-1) = \frac{1}{2} = \frac{1}{4}\cdot 2 + \frac{3}{4}\cdot 0 = \mathbb{E}^Q[b(\vec{p})|\vec{p}[2] = 2]$$

This implies that the biased estimators $T^{(n)}(i,j)$ won't be able to detect the difference in quality in the second column and would underestimate the difference in the first column. However, Corollary 37 ensures that

$$T^{(n)}_{Q,U}(i,j) = \frac{1}{2|D^{(n)}_{(i,j)}|}\sum_{\vec{p}\in D^{(n)}_{(i,j)}}\left(b(\vec{p})\frac{\mathbb{P}^Q\left(\vec{\mathcal{P}}'[j] = i\right)}{\mathbb{P}^Q\left(\vec{\mathcal{P}}' = \vec{p}\right)}\right)$$

allows to detect the differences in the second column. If we write $\vec{p}_1 = (s,1,1),(1,2),r),\vec{p}_2 = (s,(1,1),(2,2),r),\vec{p}_3 = (s,(2,1),(1,2),r)$ and $\vec{p}_4 = (s,(1,1),(2,2),r)$ the estimators for the second column become:

$$T^{(n)}_{Q,U}(1,2) = \frac{1}{2|D^{(n)}_{(1,2)}|}\frac{1}{2}\sum_{\vec{p}\in D^{(n)}_{(1,2)}}\left(\frac{8}{3}b(\vec{p})\mathbb{1}\{\vec{p} = \vec{p}_1\} + 8b(\vec{p})\mathbb{1}\{\vec{p} = \vec{p}_3\}\right),$$

$$T^{(n)}_{Q,U}(2,2) = \frac{1}{2|D^{(n)}_{(2,2)}|}\frac{1}{2}\sum_{\vec{p}\in D^{(n)}_{(2,2)}}\left(8b(\vec{p})\mathbb{1}\{\vec{p} = \vec{p}_2\} + \frac{8}{3}b(\vec{p})\mathbb{1}\{\vec{p} = \vec{p}_4\}\right).$$

Let us consider some data from simulations. We suppose S to be a matrix with independent Gaussian entries. The means are given by $\mathbb{E}[S]$ above, and the variances are chosen as follows

$$\mathbb{V}(S) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}. \tag{2.46}$$

We perform $n = 1000$ simulation runs and compare the estimations under the assumption of uniform transition, known transitions, and estimated transitions in Tables 2.1 and 2.2. The asymptotic distributions of the estimators also allow constructing confidence intervals or tests for the difference.

Example 44 (Wafer production). Our study was motivated by a root-cause analysis in the wafer fabrication. Wafer fabrication is, in general, a procedure of many repeated sequential processes. For instance, a simplified illustration consists of 12 subsequent fabrication steps, see [contributors, 2019], where intermediate measurement of qualities is not feasible. Our concrete example treated up to 30 different steps and more than 90 machines. Unfortunately, since our industrial partner insists on fulfilling an NDA, we cannot publish any more information about the project. Probably for the same reasons, we could not find publicly available data on other industrial projects.

	$\mathbb{E}[S(1,j)] - \mathbb{E}[S(2,j)]$	$T^{(n)}(1,j) - T^{(n)}(2,j)$	$T^{(n)}_{Q,U}(1,j) - T^{(n)}_{Q,U}(2,j)$	$\widehat{T}^{(n)}_{Q,U}(1,j) - \widehat{T}^{(n)}_{Q,U}(2,j)$
$j=1$	2	1.54	2.15	2.01
$j=2$	-1	-0.13	-1.21	-1.04

Table 2.1: Different estimations for the differences in mean in Example 43; values rounded to the third decimal, $n = 1000$.

	$\mathbb{V}[S(1,j)] - \mathbb{V}[S(2,j)]$	$\Sigma^{(n)}(1,j) - \Sigma^{(n)}(2,j)$	$\Sigma^{(n)}_{Q,U}(1,j) - \Sigma^{(n)}_{Q,U}(2,j)$	$\widehat{\Sigma}^{(n)}_{Q,U}(1,j) - \widehat{\Sigma}^{(n)}_{Q,U}(2,j)$
$j=1$	1	0.99	1.08	1.05
$j=2$	0	0.68	0.13	0.21

Table 2.2: Different estimations for the differences in variance in Example 43; values rounded to the third decimal, $n = 1000$.

In the following, we treat some textbook examples to illustrate the possible application of our method.

Example 45 (Tooth growth). We consider the classical tooth growth data set, [Crampton, 1947] or [R Documentation,], that studies the effect of vitamin C on tooth growth in Guinea pigs. We also refer to [Greenwood, 2020] for a detailed analysis of this data set. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two sources of vitamin C; ascorbic acid (VC) or orange juice (OJ). This experiment is balanced and corresponds to the uniform case, under the additional assumption that every path appears 10 times. See Figure 2.4 for violin plots of the six different groups.

The figure suggests that the mean tooth growth increases with the dosage level and that OJ seems to lead to higher growth rates than VC except at a 2 mg/day dosage. The variability around the means looks to be different, however relatively small to the differences among the means. There might be some skew in the responses in some of the groups (e.g., the group OJ with dose 0.5 has a right skew) that may violate the standard normality assumption. However, a classic analysis does not reveal statistical evidence against constance variances and the normality assumption. The fitted coefficients in a multivariate linear model then give an increase of 3.7 from VC to OJ and that the lengths increase with a dose of 1mg/day (resp. 2mg/day) by 9.13 (resp. 15.49) to the baseline of 0.5 mg/day. These differences are reported with p-values smaller than 0.001. Without having to verify the above conditions, our method obtains the same value for the difference in mean contribution. Also, we can estimate the differences of the variances: the variance increases from VC to OJ by 23.87. Passing from 0.5mg/day to 1mg/day (resp. 2mg/day) increases the variance by 0.71 (resp. 5.70). As expected, these values coincide with the pairwise difference of the empirical variances in each group.

The next examples treats a situation where variances are not constant.

Example 46 (Biomass response). We consider the experiment in [Sokolowska K and MC., 2017] that studied the impacts of Nitrogen (N) additions on the mass of two feather moss species (Pleurozium schreberi (PS) and Hylocomium (HS)) in an experimental forest in Sweden. More details on the classic analysis of this data set can be found in [Greenwood, 2020]. The study used a randomized block design. Here, pre-specified areas were divided into three experimental units of area 0.1 hectare, and one of the three treatments was randomly applied. This procedure resulted in a balanced design with six replicates at each combination of species and treatment. The three treatments involved different levels of Nitrogen applied immediately after snowmelt: no additional Nitrogen (control), 12.5 kg N per ha and year, (N12.5), and 50 kg N per ha and year (N50). The study's primary objective was whether the treatments would have differential

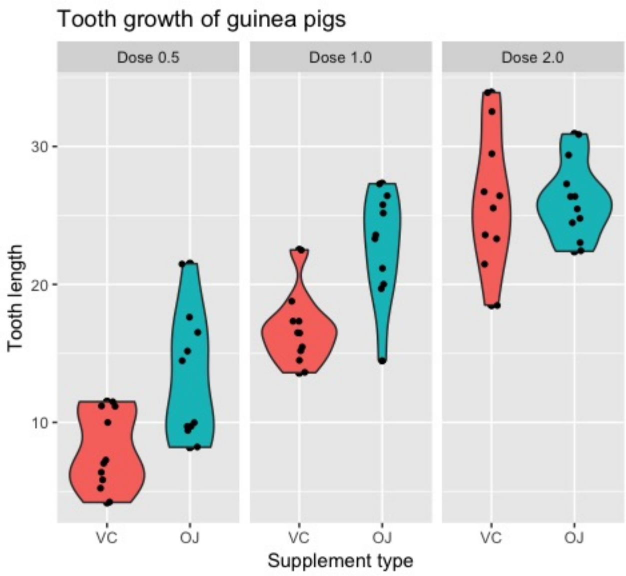


Figure 2.4: Violin plot of odontoplast growth (in microns).

impacts on the two species; while they measured additional variables, we will only observe the variables mentioned above. The violin plot in Figure 2.5 provides a first overview of the responses. Figure 2.5 indicates some differences in variability in the different groups. Classical residual versus fitted plots reveal that there is a problem with non-constant variances. The normality assumption seems, however, to be verified. A standard remedy for non-constant variance is to transform the data to a logarithmic scale. This transformation reduces the differences in variances but induces a slight violation of the normality assumption. Our method does not demand such a transformation, and we can directly perform the pairwise comparison. However, since the experiment is balanced, we obtain the same estimates as in the multivariate regression model: the treatment with N12.5 (resp. N50) decreases the growth by -488.55 (resp. -1138.76), and the species PS has in mean more 955.78 mass grow per ha and year than the species HS. Our method allows also quantifying the difference in variability. The increase in variance from HS to PS is estimated with 287118.5, and the loss of variability in N12.5 (resp. N50) is 29235.43 (resp. 252389.76) compared to the control group.

We now turn to an examples with non-constant variability and dependency between the predictors.

Example 47 (California test score data). We consider the California test score data, see [Zeiles, Achim,]. This data concerns all 420 $K-6$ and $K-8$ districts in California between 1998 and 1999. Test scores are on the Stanford 9 standardized test administered to 5th-grade students. We add a new variable, called score, that is the mean of the English and math results. This score will be our response variable.

The original data set contains many possible school characteristics or predictors for the score, as enrollment, number of teachers, number of students, number of computers per classroom, and expenditures per student. Besides, there are demographic variables for the students that are averaged across each district. These demographic variables include the percentage of English learners, that is, students for whom English is a second language. We call this variable “english”.

This dataset is an interesting textbook example since it illustrates many different aspects of multivariate regression models. In the following, we do not aim to conduct a scientific study of this data set, but only want to illustrate our method. A natural guess is that the student-teacher ratio may impact the pupils’ test scores. For this purpose, we add a variable STR that gives the ratio between students and teachers to the dataset. The correlation between STR

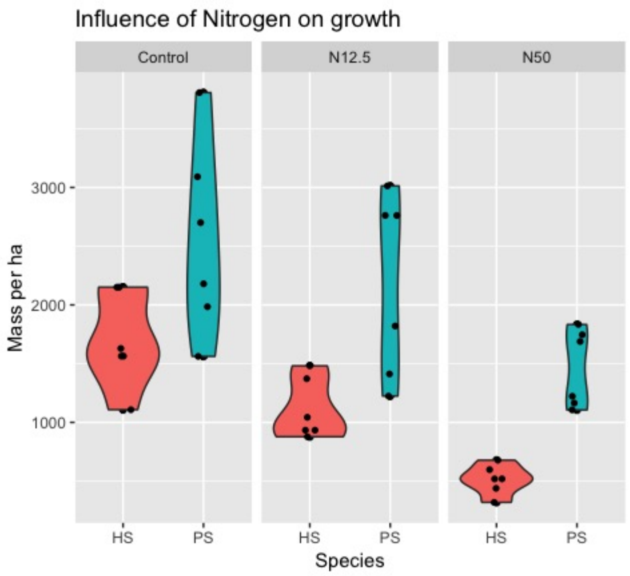


Figure 2.5: Violin plot of moss growth.

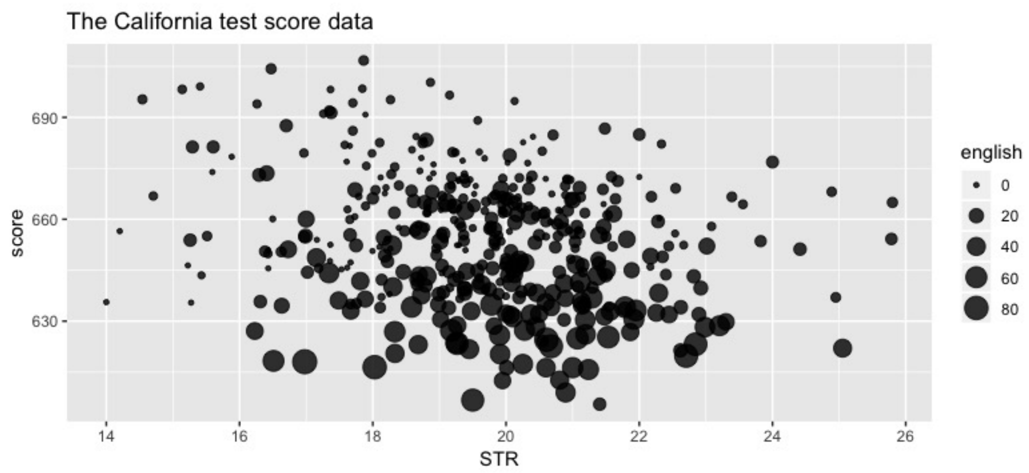


Figure 2.6: Bubble plot of the score as a function of STR. The sizes of the bubbles correspond to the percentage of students for whom English is a second language.

and scores is about -0.22 and can be considered statistically significant, e.g., the Pearson's correlation test leads to a p-value of less than 0.1%. However, the conclusion that smaller class sizes lead to better performances might be made too fast, or their impact may be overestimated. Other variables may influence the test scores, too. The next variable that we are looking at is english; its correlation with scores is -0.64, which can be considered statistically significant. Moreover, STR and english correlate 0.19. A first visual glimpse of the data can be obtained through Figure 2.6.

Our method is a priori suited to categorical predictors. We, therefore, discretize our data and create 5 groups of equal size for STR and english. The groups in STR are defined by the following breaks (rounded to two decimals): 14.00, 18.16, 19.27, 20.08, 21.08, and 25.80. The new variable is called STRCat. For instance, group 1 contains all observations with STR between 14.00 and 18.16. The five groups of english are defined by the breaks: 0.00, 1.16, 5.01, 13.14, 30.72, and 85.54. The new variable is called englishCat. Figure 2.7 shows violin plots of discretized data and gives a strong indication of non-constant variances. Our Markov model

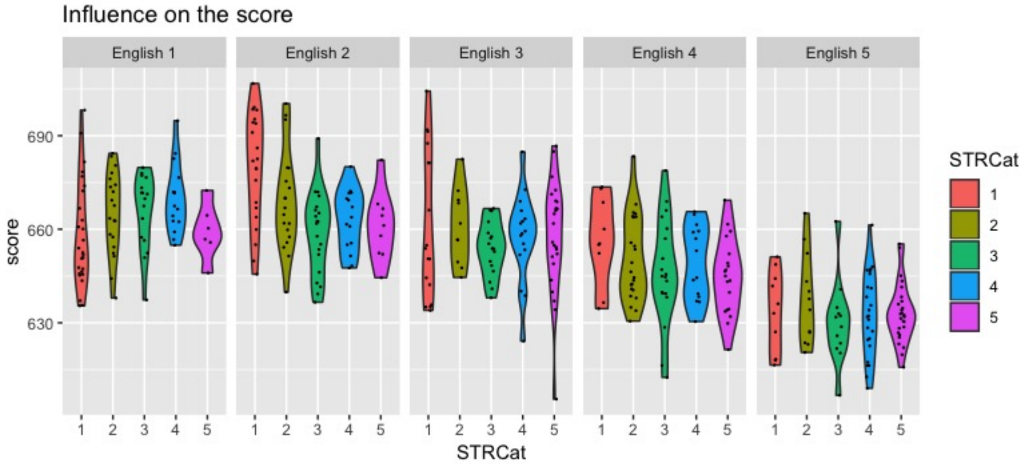


Figure 2.7: Influence of student-teacher ratio and percentage of the non-native speaker on the score.

can now be applied to the adapted data. Since we only have two predictors, the assumption of Markov dependencies is satisfied. We consider the variable english as the first column and the variable STRcat as the second.

The transition of our Markov chain describes the correlation between these two categorical. Since we choose every group in englischCat of equal size $\widehat{Q}^{(1)}$ is the uniform distribution of $\{1, \ldots, \}$. The second Markov transition kernel is estimated by:

$$\widehat{Q}^{(2)} = \begin{pmatrix} 0.32 & 0.24 & 0.19 & 0.18 & 0.07 \\ 0.25 & 0.23 & 0.24 & 0.18 & 0.11 \\ 0.20 & 0.12 & 0.20 & 0.19 & 0.29 \\ 0.11 & 0.27 & 0.23 & 0.17 & 0.23 \\ 0.12 & 0.14 & 0.14 & 0.29 & 0.31 \end{pmatrix}. \tag{2.47}$$

We compare the different estimators for the differences in means. Let $\Delta(i,j) = \mathbb{E}[S(1,j)] - \mathbb{E}[S(i,j)]$ for $i \in \{2,5\}$ and $j \in \{1, \ldots\}$. We denote $\Delta_U^{(n)}(i,j) = T_U^{(n)}(1,j) - T_U^{(n)}(i,j)$ for the estimation under the assumption of uniform transitions and obtain

$$\Delta_U^{(n)} = \begin{pmatrix} 3.91 & -3.37 \\ -5.45 & -8.66 \\ -14.71 & -8.81 \\ -30.29 & -13.41 \end{pmatrix}. \tag{2.48}$$

The unbiased estimations are given by $\widehat{\Delta}_{Q,U}^{(n)}(i,j) = \widehat{T}_{Q,U}^{(n)}(1,j) - \widehat{T}_{Q,U}(i,j)$:

$$\widehat{\Delta}_{Q,U}^{(n)} = \begin{pmatrix} 2.68 & -1.11 \\ -5.41 & -6.69 \\ -14.14 & -3.34 \\ -30.32 & -6.64 \end{pmatrix}. \tag{2.49}$$

To compare the model with a multivariate regression we give the corresponding estimates resulting from the standard linear regression:

$$\Delta_{\text{regression}}^{(n)} = \begin{pmatrix} 4.39 & -0.66 \\ -4.29 & -6.29 \\ -13.68 & -2.88 \\ -29.09 & -5.26 \end{pmatrix}. \tag{2.50}$$

We can observe that the biased estimator $T_U^{(n)}$ leads to an overestimation of the effect of STR on the scores. Finally, let us consider the estimates for the differences in variances. For $i \in \{2,5\}$ and $j \in \{1,\dots\}$ let $\Gamma(i,j) = \mathbb{V}[S(1,j)] - \mathbb{V}[S(i,j)]$ and its estimators $\Gamma_U^{(n)}(i,j) = \Sigma_U^{(n)}(1,j) - \Sigma_U^{(n)}(i,j)$ and $\widehat{\Gamma}_{Q,U}^{(n)}(i,j) = \widehat{\Sigma}_{Q,U}^{(n)}(1,j) - \widehat{\Sigma}_{Q,U}(i,j)$. We obtain

$$\Gamma_U^{(n)} = \begin{pmatrix} -64.24 & 185.54 \\ -72.11 & 198.51 \\ -0.56 & 149.28 \\ 47.12 & 198.30 \end{pmatrix} \text{ and } \widehat{\Gamma}_{Q,U}^{(n)} = \begin{pmatrix} -68.59 & 190.41 \\ -84.87 & 190.47 \\ -29.32 & 196.33 \\ 9.40 & 244.44 \end{pmatrix}. \tag{2.51}$$

For instance, this shows that the variances in the group 1 and 5 of english are similar and are smaller than in the other groups. The variance in group of STRcat is much higher than in the other four groups. We also see that the unbiasing leads to very different estimates on the variability.

Appendix

A multidimensional version of Anscombe's theorem

We present a multi-dimensional version of the classical Anscombe's Theorem. The proof follows with simple modification the argument given by Renyi in his proof of Anscombe's theorem [Gut, 2009]; it is presented here for the sake of completeness.

Theorem 48 (Multidimensional Anscombe)**.** *Let* $Y^{(i)} := (Y_1^{(i)}, Y_2^{(i)}, \dots, Y_m^{(i)})$*, for* $i \geq 1$*, be a sequence of i.i.d. real-valued random vectors with* $\mathbb{E}[Y^{(i)}] = 0 \in \mathbb{R}^m$ *and covariance matrix* Σ*. Let* $N(t)$ *be a random integer-valued random variable such that* $N(t)/t \xrightarrow[t\to\infty]{a.s.} \theta \in \mathbb{R}^+$*, then*

$$\frac{1}{\sqrt{N(t)}} \sum_{i=1}^{N(t)} Y^{(i)} \xrightarrow[t\to\infty]{distr.} \mathcal{N}(0, \Sigma).$$

Proof. Let $n(t) := \lfloor \theta t \rfloor$ and let $S_k := \sum_{i=1}^{k} Y^{(i)}$ and let $S_k^{(j)} = \sum_{i=1}^{k} Y_j^{(i)}$ then

$$\frac{S_{N(t)}}{\sqrt{N(t)}} = \left(\left(\frac{S_{n(t)}^{(1)}}{\sqrt{n(t)}} + \frac{S_{N(t)}^{(1)} - S_{n(t)}^{(1)}}{\sqrt{n(t)}} \right) \sqrt{\frac{n(t)}{N(t)}}, \dots, \left(\frac{S_{n(t)}^{(m)}}{\sqrt{n(t)}} + \frac{S_{N(t)}^{(m)} - S_{n(t)}^{(m)}}{\sqrt{n(t)}} \right) \sqrt{\frac{n(t)}{N(t)}} \right). \tag{2.52}$$

The first observation is that, since $n(t)$ is deterministic, due to the multi-dimensional central limit theorem we have that

$$\left(\frac{S_{n(t)}^{(1)}}{\sqrt{n(t)}}, \dots, \frac{S_{n(t)}^{(m)}}{\sqrt{n(t)}} \right) \xrightarrow[t\,\to\,\infty]{\text{distr.}} \mathcal{N}(0, \Sigma), \tag{2.53}$$

where Σ is the covariance matrix of the random vector $Y^{(1)}$. Next, let $\epsilon \in (0, 1/3)$ be given and $n_1(t) := \lfloor n(t)(1 - \epsilon^3) \rfloor + 1$ and $n_2(t) := \lfloor n(t)(1 + \epsilon^3) \rfloor$, then

$$\mathbb{P}\left[\bigcup_{i=1}^{m} \left\{ \left| S_{N(t)}^{(i)} - S_{n(t)}^{(i)} \right| > \epsilon\sqrt{n} \right\} \right] \leq \sum_{i=1}^{m} \mathbb{P}\left[\left| S_{N(t)}^{(i)} - S_{n(t)}^{(i)} \right| > \epsilon\sqrt{n} \right], \tag{2.54}$$

by the union bound. Let $\sigma_i^2 := \mathbb{E}\left[\left(Y_i^{(1)}\right)^2\right] < \infty$, then we also know that

$$
\begin{aligned}
\mathbb{P}\left[\left|S_{N(t)}^{(i)} - S_{n(t)}^{(i)}\right| > \epsilon\sqrt{n(t)}\right] &= \mathbb{P}\left[\left|S_{N(t)}^{(i)} - S_{n(t)}^{(i)}\right| > \epsilon\sqrt{n(t)},\, N(t) \in [n_1(t), n_2(t)]\right] \\
&\quad + \mathbb{P}\left[\left|S_{N(t)}^{(i)} - S_{n(t)}^{(i)}\right| > \epsilon\sqrt{n(t)},\, N(t) \notin [n_1(t), n_2(t)]\right] \\
&\leq \mathbb{P}\left[\max_{n_1(t)\leq n\leq n(t)} \left|S_n^{(i)} - S_{n(t)}^{(i)}\right| > \epsilon\sqrt{n(t)}\right] \\
&\quad + \mathbb{P}\left[\max_{n(t)\leq n\leq n_2(t)} \left|S_n^{(i)} - S_{n(t)}^{(i)}\right| > \epsilon\sqrt{n(t)}\right] \\
&\quad + \mathbb{P}\left[N(t) \notin [n_1(t), n_2(t)]\right] \\
&\leq \frac{(n(t) - n_1(t))\sigma_i^2}{\epsilon^2 n(t)} + \frac{(n_2(t) - n(t))\sigma_i^2}{\epsilon^2 n(t)} \qquad \text{(Kolmogorov's inequality)} \\
&\quad + \mathbb{P}\left[N(t) \notin [n_1(t), n_2(t)]\right] \\
&\leq 3\epsilon
\end{aligned}
$$

for all $i = 1, \ldots, m$ where the last inequality is valid for t sufficiently large. Plugging this last estimation in Inequality (2.54) yields for t sufficiently large

$$
\mathbb{P}\left[\bigcup_{i=i}^{m} \left\{\left|S_{N(t)}^{(i)} - S_{n(t)}^{(i)}\right| > \epsilon\sqrt{n}\right\}\right] \leq 3m\epsilon,
$$

for any $\epsilon \in (0, 1/3)$. Since ϵ can be chosen arbitrarily small we deduce that

$$
\left(\frac{S_{N(t)}^{(1)} - S_{n(t)}^{(1)}}{\sqrt{n(t)}}, \ldots, \frac{S_{N(t)}^{(m)} - S_{n(t)}^{(m)}}{\sqrt{n(t)}}\right) \xrightarrow[t\to\infty]{\text{prob.}} (0, 0, \ldots, 0).
$$

By noticing that $\sqrt{\frac{n(t)}{N(t)}} \xrightarrow[t\to\infty]{\text{prob.}} 1$ and using the multidimensional version of Slutsky's theorem (Lemma 2.8 in [Van der Vaart, 2000]), we deduce that

$$
\sqrt{\frac{n(t)}{N(t)}}\left(\frac{S_{N(t)}^{(1)} - S_{n(t)}^{(1)}}{\sqrt{n(t)}}, \ldots, \frac{S_{N(t)}^{(m)} - S_{n(t)}^{(m)}}{\sqrt{n(t)}}\right) \xrightarrow[t\to\infty]{\text{prob.}} (0, 0, \ldots, 0),
$$

where the last convergence is indeed in probability since it is a convergence in distribution to a constant. Using this last equation, Equation (2.52), and the multidimensional Slutsky theorem, we conclude that

$$
\frac{S_{N(t)}}{\sqrt{N(t)}} \xrightarrow[t\to\infty]{\text{distr.}} \mathcal{N}(0, \Sigma).
$$

□

An Anscombe version of the multivariate delta method

Here we present a modification of the multivariate delta method that also works in the case when n is replaced with a random variable. The proof is a simple modification of the argument used for the multivariate delta method and it is presented for the sake of completeness.

Theorem 49 (Anscombe's multivariate delta method). *Let $\theta \in \mathbb{R}^k$ and $\{T_n\}_{n\in\mathbb{N}}$ be a sequence of k dimensional random vectors and $\{\mathcal{X}_n\}_{n\in\mathbb{N}}$ be sequence of natural valued random variables such that*

- $\sqrt{\mathcal{X}_n}(T_{\mathcal{X}_n} - \theta) \xrightarrow[n\to\infty]{distr.} \mathcal{N}_k(0, \Sigma)$.
- $T_{\mathcal{X}_n} \xrightarrow[n\to\infty]{prob.} \theta$

Furthermore, let $h : \mathbb{R}^k \to \mathbb{R}^m$ be once differentiable at θ with the gradient matrix $\nabla h(\theta)$. Then

$$\sqrt{\mathcal{X}_n}\,(h(T_{\mathcal{X}_n}) - h(\theta)) \xrightarrow[n\to\infty]{distr.} \mathcal{N}_k(0, \nabla h(\theta)^T \Sigma \nabla h(\theta)).$$

***Proof*.** By the definition of differentiability of a vector field, we have that

$$h(x) = h(\theta) + (x - \theta) \cdot \nabla h(\theta) + |x - \theta| R_2(x) \tag{2.55}$$

where $|R_2(x)| \xrightarrow[x\to\theta]{} 0$. In particular, we have that

$$\sqrt{\mathcal{X}_n} \cdot (h(T_{\mathcal{X}_n}) - h(\theta)) = \sqrt{\mathcal{X}_n} \cdot (T_{\mathcal{X}_n} - \theta) \cdot \nabla h(\theta) + \left(\sqrt{\mathcal{X}_n} \cdot |T_{\mathcal{X}_n} - \theta|\right) R_2(T_{\mathcal{X}_n}). \tag{2.56}$$

On the other hand, it follows from the assumptions and the definition of R_2 that

- $\sqrt{\mathcal{X}_n} \cdot (T_{\mathcal{X}_n} - \theta) = (\sqrt{\mathcal{X}_n}\,(T_{\mathcal{X}_n} - \theta)) \xrightarrow[n\to\infty]{distr.} \mathcal{N}_k(0, \Sigma)$,
- $R_2(T_{\mathcal{X}_n}) \xrightarrow[n\to\infty]{prob.} 0$.

Therefore, using the multidimensional Slutsky's theorem (Lemma 2.8 in [Van der Vaart, 2000]), we get that

$$\left(\sqrt{\mathcal{X}_n} \cdot |T_{\mathcal{X}_n} - \theta|\right) R_2(T_{\mathcal{X}_n}) \xrightarrow[n\to\infty]{prob.} 0, \tag{2.57}$$

where the last convergence is in probability because it is towards a constant. Using once more the multidimensional Slutsky's theorem together with Equations (2.56),(2.57) we conclude that

$$\sqrt{\mathcal{X}_n} \cdot (h(T_{\mathcal{X}_n}) - h(\theta)) \xrightarrow[n\to\infty]{\text{distr.}} \mathcal{N}_k(0, \nabla h(\theta)^T \Sigma \nabla h(\theta)). \tag{2.58}$$

□

Bibliography

[Aistleitner et al., 2021] Aistleitner, C., D'Angeli, D., Gutierrez, A., Rodaro, E., and Rosenmann, A. (2021). Circular automata synchronize with high probability. *Journal of Combinatorial Theory, Series A*. To Appear.

[Ananichev et al., 2010] Ananichev, D. S., Gusev, V. V., and Volkov, M. V. (2010). Slowly synchronizing automata and digraphs. *CoRR*, abs/1005.0129.

[Bang-Jensen and Gutin, 2009] Bang-Jensen, J. and Gutin, G. (2009). *Digraphs*. Springer Monographs in Mathematics. Springer-Verlag London, Ltd., London, second edition. Theory, algorithms and applications.

[Béal et al., 2011] Béal, M.-P., Berlinkov, M. V., and Perrin, D. (2011). A quadratic upper bound on the size of a synchronizing word in one-cluster automata. *International Journal of Foundations of Computer Science*, 22(02):277–288.

[Berlinkov, 2016] Berlinkov, M. V. (2016). On the probability of being synchronizable. In *Algorithms and discrete applied mathematics*, volume 9602 of *Lecture Notes in Comput. Sci.*, pages 73–84. Springer, [Cham].

[Berlinkov and Nicaud, 2018] Berlinkov, M. V. and Nicaud, C. (2018). Synchronizing random almost-group automata. In *International Conference on Implementation and Application of Automata*, pages 84–96. Springer.

[Boesch and Tindell, 1984] Boesch, F. and Tindell, R. (1984). Circulants and their connectivities. *J. Graph Theory*, 8(4):487–499.

[Bohme et al., 2019] Bohme, D., Geimer, M., Wolf, F., and Arnold, L. (2019). Identifying the root causes of wait states in large-scale parallel applications. In *2010 39th International Conference on Parallel Processing*, pages 90–100.

[Cerny, 1964] Cerny, J. (1964). Poznamka k homogenym eksperimentom s konechnymi automatami, math. *Fyz. Cas*, 14:208–215.

[Codenotti et al., 1998] Codenotti, B., Gerace, I., and Vigna, S. (1998). Hardness results and spectral techniques for combinatorial problems on circulant graphs. *Linear Algebra and its Applications*, 285(1):123 – 142.

[contributors, 2019] contributors, W. (2019). Wafer fabrication.

[Crampton, 1947] Crampton, E. W. (1947). The Growth of the Odontoblasts of the Incisor Tooth as a Criterion of the Vitamin C Intake of the Guinea Pig: Five Figures. *The Journal of Nutrition*, 33(5):491–504.

[Draper and Pukelsheim, 1996] Draper, N. R. and Pukelsheim, F. (1996). An overview of design of experiments. *Statist. Papers*, 37(1):1–32.

[Dubuc, 1998] Dubuc, L. (1998). Sur les automates circulaires et la conjecture de černỳ. *RAIRO-Theoretical Informatics and Applications*, 32(1-3):21–34.

[Fengming et al., 2005] Fengming, D., Khee-meng, K., et al. (2005). *Chromatic polynomials and chromaticity of graphs*. World Scientific.

[Frankl, 1982] Frankl, P. (1982). An extremal problem for two families of sets. *European Journal of Combinatorics*, 3(2):125–127.

[Greenwood, 2020] Greenwood, M. C. (2020). *Intermediate Statistics with R*. Version 2.2.

[Gut, 2009] Gut, A. (2009). *Stopped random walks*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition. Limit theorems and applications.

[Gutierrez and Müller, 2019] Gutierrez, A. and Müller, S. (2019). Quality analysis in acyclic production networks. *Stoch. Qual. Control*, 34(2):59–66.

[Gutierrez and Müller, 2020] Gutierrez, A. and Müller, S. (2020). Estimations of means and variances in a markov linear model. *arXiv preprint arXiv:2010.06999*. Submitted.

[Kleiber and Zeileis, 2008] Kleiber, C. and Zeileis, A. (2008). *Applied Econometrics with R*. Springer New York.

[McDiarmid, 1989] McDiarmid, C. (1989). On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188.

[Nicaud, 2019] Nicaud, C. (2019). The Černý conjecture holds with high probability. *J. Autom. Lang. Comb.*, 24(2-4):343–365.

[Perrin, 1977] Perrin, D. (1977). Codes asynchrones. *Bulletin de la Société mathématique de France*, 105:385–404.

[Pin, 1978] Pin, J.-E. (1978). Sur un cas particulier de la conjecture de Cerny. In *Automata, languages and programming (Fifth Internat. Colloq., Udine, 1978)*, volume 62 of *Lecture Notes in Comput. Sci.*, pages 345–352. Springer, Berlin-New York.

[Pin, 1983] Pin, J.-E. (1983). On two combinatorial problems arising from automata theory. In *North-Holland Mathematics Studies*, volume 75, pages 535–548. Elsevier.

[R Documentation,] R Documentation. https://stat.ethz.ch/r-manual/r-patched/library/datasets/html/toothgrowth.html.

[Rutherfors, 2011] Rutherfors, A. (2011). *ANOVA and ANCOVA - A GLM Approach*. Wiley.

[Schulz, 2005] Schulz, M. (2005). Extracting critical path graphs from mpi applications. In *2005 IEEE International Conference on Cluster Computing*, pages 1–10.

[Selvamuthu and Das, 2018] Selvamuthu, D. and Das, D. (2018). *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*. Springer Singapore.

[Shitov, 2019] Shitov, Y. (2019). An improvement to a recent upper bound for synchronizing words of finite automata. *Journal of Automata, Languages and Combinatorics*, 24:367–373.

[Skvortsov and Zaks, 2010] Skvortsov, E. and Zaks, Y. (2010). Synchronizing random automata. *Discrete Mathematics and Theoretical Computer Science*, 12(4):95–108.

[Sokolowska K and MC., 2017] Sokolowska K, T. M. and MC., N. (2017). Symplasmic and apoplasmic transport inside feather moss stems of pleurozium schreberi and hylocomium splendens. *Annals of botany*, 120(5):1805–817.

[Szykuła, 2017] Szykuła, M. (2017). Improving the upper bound on the length of the shortest reset words. *arXiv preprint arXiv:1702.05455*.

[Van der Vaart, 2000] Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

[Volkov, 2008] Volkov, M. V. (2008). Synchronizing automata and the Černỳ conjecture. In *International Conference on Language and Automata Theory and Applications*, pages 11–27. Springer.

[Zeiles, Achim,] Zeiles, Achim. https://www.rdocumentation.org/packages/aer/versions/1.2-9/topics/caschoolsl.