

The Pennsylvania State University  
The Graduate School

APPROXIMATE REALIZATION OF HIDDEN MARKOV MODELS

A Thesis in  
Electrical Engineering  
by  
Ibrahim Ekrem Bardakci

© 2014 Ibrahim Ekrem Bardakci

Submitted in Partial Fulfillment  
of the Requirements  
for the Degree of

Master of Science

December 2014

The thesis of Ibrahim Ekrem Bardakci was reviewed and approved\* by the following:

Constantino Lagoa  
Professor of Electrical Engineering  
Thesis Advisor

Kenneth Jenkins  
Professor of Electrical Engineering

Kultegin Aydin  
Professor of Electrical Engineering  
Head of the Department of Electrical Engineering

\*Signatures are on file in the Graduate School.

# Abstract

The realization problem for hidden Markov models (HMMs) has been studied over years. However, it does not always have solution and requires the exact string probabilities of all string of finite length. In real world applications, it is not feasible and only finite number of approximate string probabilities of strings up to certain length are usually available. Since the string probabilities are approximate, the realization algorithms will produce an outcome of HMM that has high order. Accordingly, it may be more useful to make a good low order approximate realization of string probabilities rather than try to match them exactly.

In this thesis, we aim to give a heuristic as a solution to approximate realization problem for HMM by presenting a low-order approximation of the HMM given a sequence of observations. We propose an algorithm to find an HMM that approximately generates the given string probabilities which are computed from the observation sequence. The algorithm first minimizes the rank of Hankel matrix constructed from an HMM using nuclear norm generalization, subsequently given the rank information finds an HMM using non-negative matrix factorization that approximately realizes the given string probabilities.

# Table of Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>Acknowledgments</b>	<b>ix</b>
<b>Chapter 1</b>	
<b>Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Historical Review . . . . .	2
1.3 Motivation and Objectives . . . . .	3
<b>Chapter 2</b>	
<b>Preliminaries</b>	<b>5</b>
2.1 Markov Process . . . . .	5
2.2 Hidden Markov Models . . . . .	6
2.2.1 Moore Hidden Markov Model . . . . .	7
2.2.2 Mealy Hidden Markov Model . . . . .	9
2.3 Generalized Block Hankel Matrix . . . . .	10
2.3.1 Brief Introduction . . . . .	10
2.3.2 Construction and Properties of Generalized Block Hankel Matrix . . . . .	12
2.4 Matrix Factorizations . . . . .	13
2.4.1 Singular Value Decomposition . . . . .	13
2.4.2 Non-negative Matrix Factorization . . . . .	13
2.5 Other Definitions and Notations . . . . .	14

<b>Chapter 3</b>	
<b>Problem Statement</b>	<b>16</b>
3.1 The Problem . . . . .	16
3.2 The Approximate Realization Algorithm . . . . .	18
<b>Chapter 4</b>	
<b>Simulation and Results</b>	<b>24</b>
4.1 Simulation Results . . . . .	24
4.1.1 Case 1: $t = 2, N = 2$ . . . . .	24
4.1.2 Case 2: $t = 2, N = 3$ . . . . .	44
4.1.3 Result Discussion . . . . .	64
<b>Chapter 5</b>	
<b>Conclusions</b>	<b>69</b>
<b>Appendix A</b>	
<b>Simulation Results for Case 3</b>	<b>70</b>
A.1 Case 3: $t = 3, N = 2$ . . . . .	70
<b>Appendix B</b>	
<b>Simulation Results for Case 4</b>	<b>87</b>
B.1 Case 4: $t = 3, N = 3$ . . . . .	87
<b>Bibliography</b>	<b>106</b>

# List of Figures

2.1	Markov process with state set $\{0,1,2\}$ . . . . .	6
2.2	Moore hidden Markov model . . . . .	7
2.3	Mealy hidden Markov model . . . . .	9
4.1	Kullback-Leibler divergence between the $\tilde{\mathcal{H}}$ and $\mathcal{H}_n$ for system order 1,2,...,7 for Case 1 . . . . .	26
4.2	Frobenius norm value between the $\tilde{\mathcal{H}}$ and $\mathcal{H}_n$ for system order 1,2,...,7 for Case 1 . . . . .	27
4.3	Singular values of $\mathcal{H}$ for $t = 2, N = 2, L_{seq} = 10^6$ . . . . .	29
4.4	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 2, L_{seq} = 10^6$ . . . . .	30
4.5	Singular values of $\mathcal{H}_n$ for $t = 2, N = 2, L_{seq} = 10^6$ . . . . .	31
4.6	Singular values of $\mathcal{H}$ for $t = 2, N = 2, L_{seq} = 10^5$ . . . . .	33
4.7	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 2, L_{seq} = 10^5$ . . . . .	34
4.8	Singular values of $\mathcal{H}_n$ for $t = 2, N = 2, L_{seq} = 10^5$ . . . . .	35
4.9	Singular values of $\mathcal{H}$ for $t = 2, N = 2, L_{seq} = 10^4$ . . . . .	37
4.10	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 2, L_{seq} = 10^4$ . . . . .	38
4.11	Singular values of $\mathcal{H}_n$ for $t = 2, N = 2, L_{seq} = 10^4$ . . . . .	39
4.12	Singular values of $\mathcal{H}$ for $t = 2, N = 2, L_{seq} = 10^3$ . . . . .	41
4.13	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 2, L_{seq} = 10^3$ . . . . .	42
4.14	Singular values of $\mathcal{H}_n$ for $t = 2, N = 2, L_{seq} = 10^3$ . . . . .	43
4.15	Kullback-Leibler divergence between the $\tilde{\mathcal{H}}$ and $\mathcal{H}_n$ for system order 1,2,...,7 for Case 2 . . . . .	45
4.16	Frobenius norm value between the $\tilde{\mathcal{H}}$ and $\mathcal{H}_n$ for system order 1,2,...,7 for Case 2 . . . . .	46
4.17	Singular values of $\mathcal{H}$ for $t = 2, N = 3, L_{seq} = 10^6$ . . . . .	48
4.18	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 3, L_{seq} = 10^6$ . . . . .	49
4.19	Singular values of $\mathcal{H}_n$ for $t = 2, N = 3, L_{seq} = 10^4$ . . . . .	50
4.20	Singular values of $\mathcal{H}$ for $t = 2, N = 3, L_{seq} = 10^5$ . . . . .	52
4.21	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 3, L_{seq} = 10^5$ . . . . .	53
4.22	Singular values of $\mathcal{H}_n$ for $t = 2, N = 3, L_{seq} = 10^5$ . . . . .	54
4.23	Singular values of $\mathcal{H}$ for $t = 2, N = 3, L_{seq} = 10^4$ . . . . .	57

4.24	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 3, L_{seq} = 10^4$ . . . .	57
4.25	Singular values of $\mathcal{H}_n$ for $t = 2, N = 3, L_{seq} = 10^4$ . . . . .	59
4.26	Singular values of $\mathcal{H}$ for $t = 2, N = 3, L_{seq} = 10^3$ . . . . .	61
4.27	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 2, N = 3, L_{seq} = 10^3$ . . . .	62
4.28	Singular values of $\mathcal{H}_n$ for $t = 2, N = 3, L_{seq} = 10^3$ . . . . .	63
A.1	Singular values of $\mathcal{H}$ for $t = 3, N = 2, L_{seq} = 10^6$ . . . . .	74
A.2	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 2, L_{seq} = 10^6$ . . . .	74
A.3	Singular values of $\mathcal{H}$ for $t = 3, N = 2, L_{seq} = 10^5$ . . . . .	77
A.4	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 2, L_{seq} = 10^5$ . . . .	78
A.5	Singular values of $\mathcal{H}$ for $t = 3, N = 2, L_{seq} = 10^4$ . . . . .	81
A.6	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 2, L_{seq} = 10^4$ . . . .	81
A.7	Singular values of $\mathcal{H}$ for $t = 3, N = 2, L_{seq} = 10^3$ . . . . .	84
A.8	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 2, L_{seq} = 10^3$ . . . .	85
B.1	Singular values of $\mathcal{H}$ for $t = 3, N = 3, L_{seq} = 10^6$ . . . . .	91
B.2	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 3, L_{seq} = 10^6$ . . . .	91
B.3	Singular values of $\mathcal{H}$ for $t = 3, N = 3, L_{seq} = 10^5$ . . . . .	95
B.4	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 3, L_{seq} = 10^5$ . . . .	95
B.5	Singular values of $\mathcal{H}$ for $t = 3, N = 3, L_{seq} = 10^4$ . . . . .	99
B.6	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 3, L_{seq} = 10^4$ . . . .	100
B.7	Singular values of $\mathcal{H}$ for $t = 3, N = 3, L_{seq} = 10^3$ . . . . .	103
B.8	The difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ for $t = 3, N = 3, L_{seq} = 10^3$ . . . .	104

# List of Tables

4.1	Change of $\lambda$ value as a function of $L_{seq}$ for Case 1 . . . . .	64
4.2	Change of $\lambda$ value as a function of $L_{seq}$ for Case 2 . . . . .	64
4.3	$\sigma(\tilde{\mathcal{H}})$ , $\sigma(\mathcal{H})$ , and $\sigma(\mathcal{H}_n)$ values as a function of $L_{seq}$ for Case 1 . . . . .	65
4.4	$\sigma(\tilde{\mathcal{H}})$ , $\sigma(\mathcal{H})$ , and $\sigma(\mathcal{H}_n)$ values as a function of $L_{seq}$ for Case 2 . . . . .	65
4.5	String probabilities of strings of length up to 3 for Case 1 . . . . .	66
4.6	String probabilities of strings of length up to 3 for Case 2 . . . . .	67
4.7	Change of $\ \tilde{\mathcal{H}} - \mathcal{H}\ _F$ and $\ \tilde{\mathcal{H}} - \mathcal{H}_n\ _F$ as a function of $L_{seq}$ for Case 1 . . . . .	68
4.8	Change of $\ \tilde{\mathcal{H}} - \mathcal{H}\ _F$ and $\ \tilde{\mathcal{H}} - \mathcal{H}_n\ _F$ as a function of $L_{seq}$ for Case 2 . . . . .	68



# Acknowledgments

First and foremost, I want to deeply appreciate my advisor Professor Constantino Lagoa for his support, guidance and encouragement along the way to fulfilling my degree and his patience in advising during this research.

I would also like to give thanks to my committee member Professor Kenneth Jenkins for extending his valuable time to review my thesis.

Special thanks to Korkut Bekiroglu for his help and guidance at the every step of my research.

I would like to thank my all friends for their friendships, supports and helps when I needed it.

Special thanks to my friend Sevket Fatih Catpinar for his encouragement and support for leading me to United States for my graduate study.

Lastly, but not the least, I would like to express special gratitude to my family for all their love and support.

# Introduction

## 1.1 Background

Hidden Markov models (HMMs) were first introduced to the literature by Blackwell and Koopmans in 1957 [1] and the use of HMMs in engineering applications dates back to late 1970s. Hidden Markov models now have various practical applications such as speech processing, source coding, communication, control systems, computational biology, computer vision and computer and network security. However, many theoretical questions concerning identification, realization and estimation of HMMs are still unanswered.

The identification problem for HMMs is given a sequence of observations or an output string finding an HMM that models the output string. Traditional techniques for identifying hidden Markov model parameters (emission and transition probabilities) use maximum likelihood estimation technique based either on specialized Expectation-Maximization techniques, such as the Baum-Welch algorithm [2], or on numerical optimization techniques, such as the gradient descent algorithm [3]. In either case, HMM parameters are estimated over several training iterations.

The subspace inspired identification method is also considered as a solution for identification problem of HMMs in [4]. Unlike the Baum-Welch algorithm, in subspace inspired identification method, state sequence is estimated from the given output sequence, then system matrices are calculated given the output sequence

and estimated state sequence.

The realization problem for HMMs is given string probabilities finding a HMM that corresponds to the given string probabilities and it has been studied over years in literature [5, 6, 7, 8, 9]. However, the realization problem for HMMs does not always have solution and it requires the exact string probabilities of all string of finite length. In practical applications it is not feasible and only finite number of string probabilities are available.

In order to apply realization algorithms to HMM with finite number of string probabilities, the partial realization problem is introduced. Unlike the realization problem, the partial realization problem always has a solution [10]. In practice, approximate string probabilities of strings up to certain length are usually given. Since the string probabilities differ from the true probabilities, the partial realization algorithms will produce an outcome of HMM that has high order [11]. Accordingly, it may be more useful to make a good low order approximate realization of string probabilities rather than try to match them exactly. In this thesis, we aim to give a heuristic as a solution to approximate realization problem.

## 1.2 Historical Review

The realization problem for HMMs first introduced in literature by Gilbert [5]. Gilbert stated that a process has to have finite Hankel rank in order to be realizable by HMM. A counterexample that a process having finite Hankel rank does not have HMM realization was given by Fox and Rubin in [6]. Sufficient condition for realization of an HMM was provided by Dharmakadhari [7] and later on Heller [8] extended the argument by giving the probabilistic characterization of HMMs and he derived the necessary and sufficient condition for the existence of an HMM. For the detailed description of these arguments developed over years, see [12]. Forty years later, Anderson [9], stated that solutions for HMM realization given in [8, 7] actually was not valid. He then presented a new stochastic realization technique for HMMs with the assumption that given string probabilities have an HMM realization. The proposed technique in [9] demonstrates that a positive factorization

of a certain Hankel matrix can be used to realize the HMMs probability matrices. In [13], the complete realization problem was streamlined.

Since a solution to realization problem is not always guaranteed, the partial realization problem considered to model an HMM to realize string probabilities of strings up to length  $t$ . In [10], Vidyasagar provided a solution to partial realization problem by modeling the string probabilities with a  $(t - 1)$ -step Markov process and it was proven in [10] that this solution satisfy the consistency and nonnegativity conditions of string probabilities of strings up to certain length.

A more relaxed version of the partial realization problem, approximate partial realization, was first considered in [11]. Later on, in [14, 4] non-negative matrix factorization (NMF) algorithms were proposed based on Anderson's HMM stochastic realization approach [9] to recover the HMMs parameters.

### 1.3 Motivation and Objectives

It is well known that problems involving minimizing the rank of a Hankel matrix is often studied in linear system theory, e.g., in designing a low-order, linear time-invariant (LTI) system directly from convex specifications on its impulse response. In conventional linear system theory, the Hankel matrix can be factorized using the impulse response of a finite dimensional linear system with rational transfer function, and the rank of the Hankel matrix gives the order information of the system. The generalized Hankel matrix formed from an HMM has some superficial similarities to Hankel matrix [13]. The generalized Hankel matrix constructed from stationary HMM can also be factorized and it has rank at most  $N$  i.e., number of states or the order of the model [9].

Following the intuition of stationary Hankel matrix rank condition, in this thesis, we present an algorithm to find an HMM that realizes the given string probabilities. Given the finite output sequence, the algorithm first minimizes the rank of Hankel matrix constructed from an HMM using nuclear norm generalization, subsequently given the rank information finds an HMM that approximately generates

the given string probabilities.

The thesis is organized as follows. In chapter 2, we introduce the notation for HMMs, and define some concepts that will be used in the rest of the thesis. In chapter 3, we state the problem and present the algorithm as a solution to the problem. In chapter 4, we apply our algorithm on simulation examples for various parameters and discuss the results.

## Preliminaries

### 2.1 Markov Process

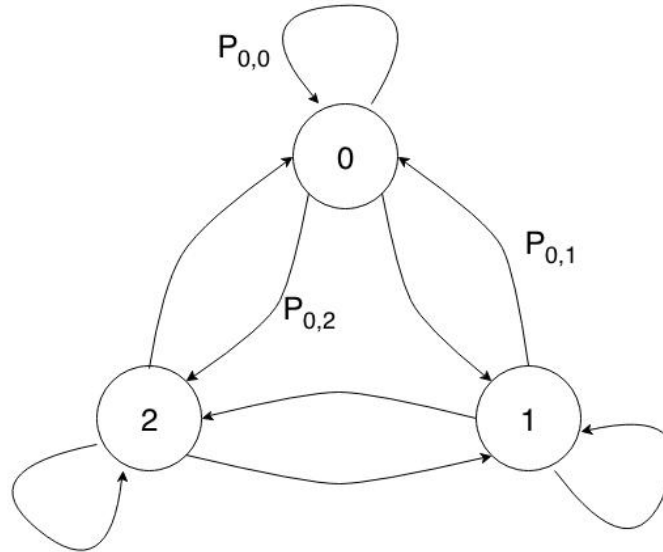
Consider a stochastic process  $\{X_{\mathbb{X}}, \mathbb{X} = 0, 1, 2, \dots\}$  that takes values on a finite or countable number of possible values. Being in state  $i$  at time  $n$  is represented by  $X_n = i$  and  $P_{ij}$  is defined as probability of transition from state  $j$  whenever the process is in state  $i$ , i.e.

$$P\{x_{n+1} = j | x_n = i, \dots, x_0 = i_0\} = P\{x_{n+1} = j | x_n = i\} = P_{ij} \quad (2.1)$$

for all states  $i_0, \dots, i_{n-1}, i, j$ . This property is called Markov property and the stochastic process having the Markov property is called as Markov process [15]. Equation 2.1. can be stated as follows; the conditional distribution of any future state  $x_{n+1}$  given the past states  $x_0, \dots, x_{n-1}$  and the present state  $x_n$ , is independent of the past state and depends only on the present state. Because of the aforementioned reason, Markov property is also called as memoryless property of a stochastic process. In Markov process, at each time step, the process moves from present state to next state according to transition probabilities. The transition probabilities  $P_{ij}$  fulfill the following conditions;

$$\begin{aligned} P_{ij} &\geq 0, \quad i, j \geq 0; \\ \sum_{j=0}^{\infty} P_{ij} &= 1, \quad i = 0, 1, \dots \end{aligned}$$

A graphical demonstration of a Markov model with three states is given in Figure 2.1.



**Figure 2.1.** Markov process with state set  $\{0,1,2\}$ .

## 2.2 Hidden Markov Models

Markov models are used to model situations where the state and transition structure are known. However, in practice, the model that is generating output sequences is "hidden" and only outputs produced by the model are known. Hidden Markov models (HMMs) are extension to Markov models that allow us to represent these situations by estimating the model and transition probabilities from observation data sequences [16]. The HMM is termed discrete if the output alphabet is finite and in this thesis discrete time hidden Markov model will be our interest. A discrete HMM is determined by two elements: an unobservable "hidden" Markov chain having a finite number of states, and a set of observations. At each discrete time instant, the process is assumed to be in a state, and an observation is generated by the probability distribution associated with the current state (in Moore HMM) or the state transition (in Mealy HMM).

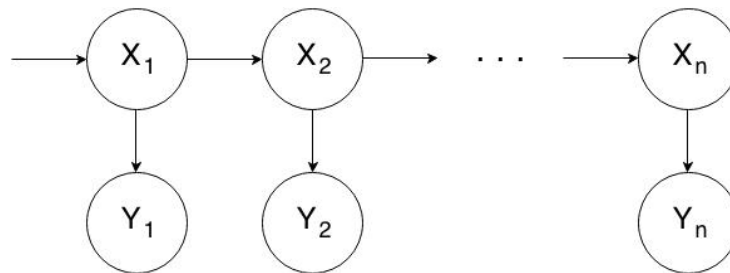
In this thesis, we consider discrete time HMMs with values in a finite set and will use the following terminology and definitions for representing the parameters and structure of HMMs:

- 1) The set  $\mathbb{X}$  is the state set and  $\mathbb{Y}$  is the output set of the HMM.
- 2)  $|\mathbb{X}| = N$  is the number of states of the Markov chain underlying the HMM in the finite-state space  $\mathbb{X} = 0, 1, 2, \dots, N - 1$  of the Markov process.  $x_t$  is defined as the state of the system at time  $t$ .
- 3)  $|\mathbb{Y}| = M$  is the number of distinct symbols or outputs generated by the HMM in the finite-output space  $\mathbb{Y} = 0, 1, 2, \dots, M - 1$ .  $y_t$  is defined as the output of the system at time  $t$ .

There are two types of HMMs named Moore and Mealy HMMs which were introduced to the literature in [17] and in [12] respectively. In Mealy HMM the generation of the output at time  $t$  given the present state  $x_t$  and the event of going to the next state  $x_{t+1}$  given the present state  $x_t$  are dependent. However, in Moore HMM, these events are independent [4]. In other words, Mealy HMM produces output at each state transition and Moore HMM produces outputs at each state. In this section, these two types of HMMs are described.

### 2.2.1 Moore Hidden Markov Model

The Moore HMM is represented by  $(\mathbb{X}, \mathbb{Y}, A, B, \pi(1))$  where  $\mathbb{X}$  and  $\mathbb{Y}$  are the state and the output sets of the process.



**Figure 2.2.** Moore hidden Markov model

In this model, let  $x$  be the state sequence and  $y$  be the output sequence of the HMM.



$A \in \mathbb{R}_+^{|\mathbb{X}| \times |\mathbb{X}|}$  with  $Ae = e$  is the state transition matrix and has the form of

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots \\ a_{21} & a_{22} & \cdots & \cdots \\ \vdots & \vdots & a_{ij} & \cdots \\ \vdots & \vdots & \vdots & a_{|\mathbb{X}| \times |\mathbb{X}|} \end{bmatrix} \quad (2.2)$$

. The element  $a_{ij}$  is defined as

$$a_{ij} = P(x_{t+1} = j | x_t = i). \quad (2.3)$$

$B \in \mathbb{R}_+^{|\mathbb{X}| \times |\mathbb{Y}|}$  with  $Be = e$  is the output matrix and has the form of

$$B = \begin{bmatrix} b(y_1) & b(y_2) & \cdots & b(y_{|\mathbb{Y}|}) \end{bmatrix}$$

$$B = \begin{bmatrix} b_1(y_1) & b_1(y_2) & \cdots & \cdots \\ b_2(y_1) & b_2(y_2) & \cdots & \cdots \\ \vdots & \vdots & b_i(y) & \cdots \\ \vdots & \vdots & \vdots & b_{|\mathbb{X}|}(y_{|\mathbb{Y}|}) \end{bmatrix} \quad (2.4)$$

The element  $b$  is a mapping from  $\mathbb{Y}$  to  $\mathbb{R}_+^{|\mathbb{X}|}$  and defined as  $b_i(y) = P(y_t = y | x_t = i)$ .  $\pi(1) \in \mathbb{R}_+^{1 \times |\mathbb{X}|}$  is a vector defined as  $\pi_i(1) = P(x_1 = i)$  with  $\pi(1)e = 1$  is the initial state distribution.

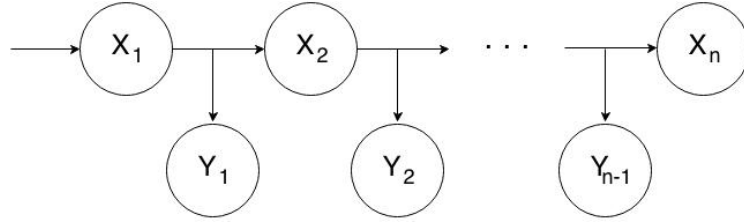
The conditions  $Ae = e$ ,  $Be = e$  and  $\pi(1)e = 1$  are called as consistency conditions of the HMM. Moore HMM is called stationary if  $\pi(1)A = \pi(1)$ . String probabilities generated by a Moore HMM are given by

$$\mathcal{P}(y) = \pi(1) \text{diag}(b(y_1)) A \cdots \text{diag}(b(y_{|y|})) Ae \quad (2.5)$$

where  $y = y_1 y_2 \dots y_{|y|} \in \mathbb{Y}^*$ .

### 2.2.2 Mealy Hidden Markov Model

The Mealy HMM is represented by  $(\mathbb{X}, \mathbb{Y}, A, \pi(1))$  where  $\mathbb{X}$  and  $\mathbb{Y}$  are the state and the output sets of the process.



**Figure 2.3.** Mealy hidden Markov model

Let again  $x$  be the state sequence and  $y$  be the output sequence of the HMM. Then, define

$$A_{\mathbb{X}} = \sum_{u \in \mathbb{Y}} A(u), \quad (2.6)$$

where  $A$  is a mapping from  $\mathbb{Y}$  to  $\mathbb{R}_+^{|\mathbb{X}| \times |\mathbb{X}|}$  and the matrix  $A_{\mathbb{X}}$  is called the state transition matrix.

$A(u) \in \mathbb{R}_+^{|\mathbb{X}| \times |\mathbb{X}|}$  has the form of

$$A(u) = \begin{bmatrix} a_{11}(u) & a_{12}(u) & \cdots & \cdots \\ a_{21}(u) & a_{22}(u) & \cdots & \cdots \\ \vdots & \vdots & a_{ij}(u) & \cdots \\ \vdots & \vdots & \vdots & a_{|\mathbb{X}| \times |\mathbb{X}|}(u) \end{bmatrix} \quad (2.7)$$

The element  $a_{ij}(u)$  is

$$a_{ij}(u) = P(x_{t+1} = j, y_{t+1} = u | x_t = i), \quad (2.8)$$

i.e. the probability of going from state  $i$  to state  $j$  while generating output symbol  $u$ , and  $A(u)$  satisfies,

$$A(uv) = A(u)A(v). \quad (2.9)$$

Then, the string probabilities generated by a Mealy HMM are given by

$$\mathcal{P}(v) = \pi^\top A(v)e. \quad (2.10)$$

Here, if  $v$  is a sequence  $y_1, y_2, \dots, y_m$ , then

$$\mathcal{P}(v) = \pi^\top A(y_1)A(y_2) \cdots A(y_m)e. \quad (2.11)$$

Where  $e = [1 \ 1 \ \cdots \ 1]^\top$ , and  $\pi$  is initial state distribution and is obtained by

$$\pi A_{\mathcal{X}} = \pi. \quad (2.12)$$

## 2.3 Generalized Block Hankel Matrix

### 2.3.1 Brief Introduction

Following [9], let  $\mathbb{Y}$  denotes the output set, and  $\mathbb{Y}^*$  denotes the set of all possible finite strings including the empty string  $\emptyset$  over  $\mathbb{Y} = \{0, 1, 2, \dots, m\}$  i.e.  $\mathbb{Y}^* = [\emptyset \ \mathbb{Y}^1 \ \mathbb{Y}^2 \ \cdots \ \mathbb{Y}^m]$ . Here,  $|u|$  is the number of symbols in  $u$  and  $\mathbb{Y}^m \subset \mathbb{Y}^*$  is the set of strings with length  $m$ . The strings  $u_i$  of  $\mathbb{Y}^*$  are listed lexicographically as last lexical order (*llo*) i.e. from left to right and first lexical order (*flo*) i.e. from right to left such that the length  $|u_i|$  increases monotonically with  $i$ . The empty sequence is included as the first element of the listing. For  $\mathbb{Y} = \{0, 1\}$ , the two enumerations are

$$\mathbb{Y}_{llo}^* = \{\emptyset, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, \dots\}$$

and

$$\mathbb{Y}_{flo}^* = \{\emptyset, 0, 1, 00, 10, 01, 11, 000, 100, 010, 110, 001, 101, 011, 111, \dots\}$$

The generalized Hankel matrix  $\mathcal{H}$  is infinite matrix with elements  $\mathcal{P}(\mathbf{u}_i \mathbf{v}_j)$  where  $u_i, v_j$  are  $i_{th}$  and  $j_{th}$  elements of two orderings respectively.

Thus, with  $\mathcal{P}(\emptyset)$  denoting the frequency of the empty string, the structure of the

generalized block Hankel matrix with two output symbols will be

$$\mathcal{H} = \begin{bmatrix} \mathcal{P}(\emptyset) & \mathcal{P}(0) & \mathcal{P}(1) & \mathcal{P}(00) & \mathcal{P}(01) & \mathcal{P}(10) & \mathcal{P}(11) & \cdots \\ \mathcal{P}(0) & \mathcal{P}(00) & \mathcal{P}(01) & \mathcal{P}(000) & \mathcal{P}(001) & \mathcal{P}(010) & \mathcal{P}(011) & \cdots \\ \mathcal{P}(1) & \mathcal{P}(10) & \mathcal{P}(11) & \mathcal{P}(100) & \mathcal{P}(101) & \mathcal{P}(110) & \mathcal{P}(111) & \cdots \\ \mathcal{P}(00) & \mathcal{P}(000) & \mathcal{P}(001) & \mathcal{P}(0000) & \mathcal{P}(0001) & \mathcal{P}(0010) & \mathcal{P}(0011) & \cdots \\ \mathcal{P}(10) & \mathcal{P}(100) & \mathcal{P}(101) & \mathcal{P}(1000) & \mathcal{P}(1001) & \mathcal{P}(1010) & \mathcal{P}(1011) & \cdots \\ \mathcal{P}(01) & \mathcal{P}(010) & \mathcal{P}(011) & \mathcal{P}(0100) & \mathcal{P}(0101) & \mathcal{P}(0110) & \mathcal{P}(0111) & \cdots \\ \mathcal{P}(11) & \mathcal{P}(110) & \mathcal{P}(111) & \mathcal{P}(1100) & \mathcal{P}(1101) & \mathcal{P}(1110) & \mathcal{P}(1111) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (2.13)$$

The term "generalized" block Hankel matrix is originated from the block Hankel matrix that is used in linear control systems because it has some similarities. In a standard block Hankel matrix, the blocks are constant along antidiagonals. However, in the generalized Hankel matrix the antidiagonal blocks contain the same probabilities. For example, the blocks  $\mathcal{H}_{(3,1)}$ ,  $\mathcal{H}_{(2,2)}$  and  $\mathcal{H}_{(1,3)}$  consist of the same string probabilities. In addition, the blocks along a certain anti-diagonal can be constructed from each other. The finite generalized block Hankel matrix  $\mathcal{H}_{(1:t_1, 1:t_2)}$  is defined as the subblock of  $\mathcal{H}$  and it is given by

$$\mathcal{H}_{(1:t_1, 1:t_2)} := [\mathcal{P}(\mathbf{u}_i \mathbf{v}_j)], \quad (2.14)$$

with  $|\mathbf{u}_i| \leq t_1 - 1$ ,  $|\mathbf{v}_j| \leq t_2 - 1$ .

$$\mathcal{H} = \begin{bmatrix} \mathcal{H}_{(1,1)} & \mathcal{H}_{(1,2)} & \cdots & \mathcal{H}_{(1,t_2)} \\ \mathcal{H}_{(2,1)} & \mathcal{H}_{(2,2)} & \cdots & \mathcal{H}_{(2,t_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{H}_{(t_1,1)} & \mathcal{H}_{(t_1,2)} & \cdots & \mathcal{H}_{(t_1,t_2)} \end{bmatrix} \quad (2.15)$$

Suppose  $\mathcal{M} = (m_1, m_2, \dots, m_m)$  is an ordered subset of strings from  $\mathbb{Y}^*$  in first lexicographical ordering and  $\mathcal{N} = (n_1, n_2, \dots, n_n)$  in last lexicographical ordering. Then the matrix  $\mathcal{H}_{(\mathcal{M}, \mathcal{N})}$  is defined as the submatrix of  $\mathcal{H}$  given by

$$(\mathcal{H}_{(\mathcal{M}, \mathcal{N})})_{ij} = \mathcal{P}(\mathbf{m}_i \mathbf{n}_j) \quad (2.16)$$

Also the string probabilities in block Hankel matrix satisfy the following consistency conditions [9] :

$$\mathcal{P}(\emptyset) = 1,$$

$$\sum_{y \in \mathbb{Y}} \mathcal{P}(\mathbf{u}y) = \mathcal{P}(\mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{Y}^*,$$

$$\sum_{\mathbf{u} \in \mathbb{Y}^k} \mathcal{P}(\mathbf{u}) = 1, \quad \forall k \in \mathbb{N}.$$

For a given process to be stationary, in addition,

$$\sum_{y \in \mathbb{Y}} \mathcal{P}(y\mathbf{u}) = \mathcal{P}(\mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{Y}^*,$$

so that we have a stationary processes that,

$$\sum_{y \in \mathbb{Y}} \mathcal{P}(\mathbf{u}y) = \sum_{y \in \mathbb{Y}} \mathcal{P}(y\mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{Y}^*.$$

With abuse of language  $\mathcal{H}$  will be called Hankel matrix in the rest of the thesis.

### 2.3.2 Construction and Properties of Generalized Block Hankel Matrix

Data analysis problems of stochastic processes typically involve a sequence  $y^{(T)}$  of length  $T$  with  $T \leq \infty$  instead of an infinite sample process  $y$  [4]. When the sample sequence  $y^{(T)}$  has finite number of elements, the generalized block Hankel matrix consists of approximate string probabilities. Let  $T$  be the length of sample sequence. The calculation of approximate string probabilities is defined as follows:

$$\tilde{\mathcal{P}}^{(t)}(\mathbf{u}) = \frac{\text{No of occurrences of a certain string } \mathbf{u}}{T - |\mathbf{u}| + 1} \tag{2.17}$$

$$\tilde{\mathcal{P}}^{(t)}(\mathbf{u}) = \sum_{y \in \mathbb{Y}} \tilde{\mathcal{P}}^{(t)}(\mathbf{u}y)$$

Approximate string probabilities  $\tilde{\mathcal{P}}^{(t)}(\mathbf{u})$  of a sample process  $\mathbf{y}$  satisfy the following consistency conditions:

$$\tilde{\mathcal{P}}(\emptyset) = 1,$$

$$\sum_{y \in \mathbb{Y}} \tilde{\mathcal{P}}(\mathbf{u}y) = \mathcal{P}(\mathbf{u}), \forall \mathbf{u} \in \mathbb{Y}^*,$$

$$\sum_{u \in \mathbb{Y}^k} \tilde{\mathcal{P}}(\mathbf{u}) = 1, \forall k \in \mathbb{N}.$$

For a given process to be stationary, in addition,

$$\sum_{y \in \mathbb{Y}} \tilde{\mathcal{P}}(y\mathbf{u}) = \tilde{\mathcal{P}}(\mathbf{u}), \forall \mathbf{u} \in \mathbb{Y}^*,$$

so that we have a stationary processes that,

$$\sum_{y \in \mathbb{Y}} \tilde{\mathcal{P}}(\mathbf{u}y) = \sum_{y \in \mathbb{Y}} \tilde{\mathcal{P}}(y\mathbf{u}), \forall \mathbf{u} \in \mathbb{Y}^*.$$

## 2.4 Matrix Factorizations

### 2.4.1 Singular Value Decomposition

The singular value decomposition SVD of a matrix  $A \in \mathbb{R}^{m \times n}$  of rank  $r$  is the factorization of  $A$  into the product of three matrices  $A = U\Sigma V^\top$  where  $U \in \mathbb{R}^{m \times m}$  orthogonal matrix whose columns are the eigenvectors of  $AA^\top$  and  $V \in \mathbb{R}^{n \times n}$  orthogonal matrix whose columns are the eigenvectors of  $A^\top A$  and the matrix  $\Sigma$  is diagonal with positive real entries. The matrix  $\Sigma$  has the form

$$\Sigma = \begin{bmatrix} \sigma_{(1)} & 0 & \cdots & 0 \\ 0 & \sigma_{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_{(r)} \end{bmatrix}$$

with  $\sigma_{(1)} \geq \sigma_{(2)} \geq \cdots \geq \sigma_{(r)} > 0$  and  $r = \text{rank}(A)$ . In the above,  $\sigma_{(1)}, \sigma_{(2)}, \cdots, \sigma_{(r)}$  are the square root of the eigenvalues of  $A^\top A$  and they are called the singular values of  $A$ .

### 2.4.2 Non-negative Matrix Factorization

The non-negative matrix factorization (NMF) problem can be stated as follows: given a non-negative matrix  $V \in \mathbb{R}_+^{n \times m}$ , find a factorization such that  $V = WQ$

with  $W \in \mathbb{R}_+^{n \times r}$  and  $Q \in \mathbb{R}_+^{r \times m}$ .  $r$  is chosen as small as possible and it is usually smaller than  $n$ , or  $m$ , [18]. In [18], the algorithms are proposed in order to solve this problem based on iterative update formulas and it is proven that the accuracy of the approximation improves monotonically with these iterative update formulas. The concept structured non-negative factorization (SNMF) problem is introduced in [19]. The importance of the SNMF is that it is used as solution to approximate positive realization problem for HMMs. The structured non-negative factorization can be stated as:

Given  $P \in \mathbb{R}_+^{p \times p}$  and  $a \in \mathbb{N}$ , find a decomposition such that  $P = VUV^T$  with  $V \in \mathbb{R}_+^{p \times a}$  and  $U \in \mathbb{R}_+^{a \times a}$ , and with  $a$  is as small as possible.

## 2.5 Other Definitions and Notations

In this section, we give the definition of some concepts that we will use in the rest of the thesis.

### L1 Norm

The  $l_1$  norm of a vector  $\vec{x}$  is denoted  $\|\vec{x}\|_1$  and is defined as the sum of absolute values of its components:

$$\|\vec{x}\|_1 = \sum_{i=1}^n |x_i| \quad (2.18)$$

### L2 Norm

The  $l_2$  norm of a vector  $\vec{x}$  is denoted  $\|\vec{x}\|_2$  and is defined as the square root of the sum of squares of absolute values of its components:

$$\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2} \quad (2.19)$$

## L-Infinity Norm

The infinity norm  $l_\infty$  of a vector  $\vec{x}$  is denoted  $\|\vec{x}\|_\infty$  and is defined as the maximum of the absolute values of its components:

$$\|\vec{x}\|_\infty = \max_i |x_i| \quad (2.20)$$

## Frobenius Norm

The Frobenius norm  $\|A\|_F$  is defined as:

$$\|A\|_F \triangleq \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{i,j}|^2} \quad (2.21)$$

## Nuclear Norm

The nuclear norm of  $\|A\|_*$  is defined as:

$$\|A\|_* \triangleq \sum_{i=1}^n \sigma_i(A) \quad (2.22)$$

## Kullback-Liebler Divergence

The Kullback-Liebler divergence between two nonnegative matrices of the same size is defined as:

$$D_{KL}(A \parallel B) = \sum_{\mu\nu} (A_{\mu\nu} \log \frac{A_{\mu\nu}}{B_{\mu\nu}} - A_{\mu\nu} + B_{\mu\nu}) \quad (2.23)$$



## Problem Statement

### 3.1 The Problem

The problem is essentially finding a low-order approximation of the HMM given a sequence of observations and can be stated in two parts. First, given an available data can one estimate the rank of the Hankel matrix (order of the model) by minimizing the nuclear norm of the Hankel matrix. Second, given the rank estimate of the Hankel matrix (order estimate of the HMM), can one estimate the system matrices (state transition and output matrix) of the model that approximately realizes given string probabilities by using non-negative matrix factorization technique.

In this problem, the string probabilities of strings up to certain length are computed from the finite observation sequence. It is given in [9] that the rank of a Hankel matrix consist of true (stationary) probabilities is less than or equal to number of states of HMM i.e.,  $rank(\mathcal{H}) \leq N$ . However, the Hankel matrix consist of approximate string probabilities usually have high rank, in turn, it produces an outcome of HMM that its order higher than  $N$ . This requires a low-order approximation of the HMM. In order to do that, the order of the HMM is assumed equal to the rank of the Hankel matrix consist of approximate string probabilities. Hence, the problem becomes rank minimization of the Hankel matrix.

However, this problem is highly non-convex and NP-hard to solve [20, 18]. A recent heuristic introduced in [21] minimizes the nuclear norm, or the sum of the

singular values of the matrix. Since the nuclear norm is a convex function, it can be optimized efficiently, and it is one of the best convex approximation of the rank function[21]. In practice, the nuclear norm generalization have been observed to produce very low-rank solutions. Because of the aforementioned reasons, in this thesis, we minimize the nuclear norm of the Hankel matrix instead of performing rank minimization.

### 3.2 The Approximate Realization Algorithm

The identification algorithm is given by

---

**Algorithm 1** The Identification Algorithm

---

- 1: Input the data  $\mathcal{O}_{1:T}$
  - 2: Compute  $\tilde{\mathcal{P}}$  from the input observation data  $\mathcal{O}_{1:T}$  and construct  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$
  - 3: Create optimization matrix  $\mathcal{H}$  and optimization vector  $\mathcal{P}$  consist of optimization variable  $x_i$
  - 4: Set  $\gamma$
  - 5: Set  $\lambda^{(0)}$  value
  - 6: **for**  $N = 1, 2, \dots, \#$  of rows of  $\mathcal{H}$  **do**
  - 7:   **for**  $t = 0, 1, 2, \dots$  **do**
  - 8:     Solve
 
$$\begin{aligned} & \min_{x_1, x_2, \dots, x_i} \|\mathcal{H}\|_* \\ & \text{subject to } 0 \leq x_i \leq 1 \\ & \quad 1/2 \|\tilde{\mathcal{P}} - \mathcal{P}\|_2^2 \leq \lambda \\ & \quad \sum_{i,j} \mathcal{H}_{uv(ij)} = 1, \quad \forall u, v. \end{aligned}$$

$$\lambda^{(t+1)} = \lambda^{(t)} + 0.1 \times 10^{-(L_{seq}+2)}$$
  - 9:     **if**  $\sigma_{N+1}(\mathcal{H})/\sigma_N(\mathcal{H}) \leq \gamma$  **then**
  - 10:       Go to step #6
  - 11:     **else**
  - 12:       Go to step #7
  - 13:     **end if**
  - 14:   **end for**
  - 15:   Compute the  $\sigma(\mathcal{H})$  and  $\|\mathcal{P} - \tilde{\mathcal{P}}\|_2$
  - 16:   Estimate the number of states  $N$  by analyzing singular values of  $\mathcal{H}$  and  $\|\mathcal{P} - \tilde{\mathcal{P}}\|_2$ .
  - 17:   **Run Algorithm 2.**
  - 18:   Reconstruct the Hankel matrix  $\mathcal{H}_n$  from  $A_n$  and  $B_n$
  - 19: **end for**
- 

In the present algorithm, given the state transition matrix  $A$  and the output

matrix  $B$ , a sample sequence of observations  $\mathcal{O}_{1:T}$  is generated by matlab's built-in function `hmmgenerate`. The observation sequence of length  $T$  of the HMM is the only input to our algorithm and it is given by,

$$\mathcal{O}_{1:T} = \mathcal{O}_1 \mathcal{O}_2 \cdots \mathcal{O}_T \quad (3.1)$$

Then, the approximate string probabilities  $\tilde{\mathcal{P}}$  are calculated by the equation 2.16 from the observation sequence. The Hankel matrix  $\tilde{\mathcal{H}}$ , consists of approximate string probabilities  $\tilde{\mathcal{P}}$ , is constructed as in section 2.3.1. As a part of algorithm, we solve an optimization problem. The optimization problem is given by,

$$\begin{aligned} \min_{x_1, x_2, \dots, x_i} \quad & \|\mathcal{H}\|_* \\ \text{subject to} \quad & 0 \leq x_i \leq 1 \\ & 1/2 \|\tilde{\mathcal{P}} - \mathcal{P}\|_2^2 \leq \lambda \\ & \sum_{i,j} \mathcal{H}_{uv(ij)} = 1, \quad \forall u, v. \end{aligned} \quad (3.2)$$

where  $x_i \in [0, 1]$  is optimization variables. The optimization matrix  $\mathcal{H}$  and the optimization vector  $\mathcal{P}$  consist of optimization variables. The constraint  $0 \leq x_i \leq 1$  ensures that we obtain valid probabilistic values. The constraint  $1/2 \|\tilde{\mathcal{P}} - \mathcal{P}\|_2^2 \leq \lambda$  means that the error, i.e., the difference between  $\mathcal{P}$  and the approximate probabilities  $\tilde{\mathcal{P}}$  in  $L_2$  norm, must be less than a given tolerance  $\lambda$ . The constraint  $\sum_{i,j} \mathcal{H}_{uv(ij)} = 1, \quad \forall u, v$  is added because of the Hankel matrix structure i.e., the sum of elements in every sub-block of a Hankel matrix is equal to 1. On the other hand, the optimization matrix  $\mathcal{H}$  is constructed by the variables  $x_i$  as in 2.3.1. in the same way as  $\tilde{\mathcal{H}}$  constructed and it fulfills all properties of generalized block

Hankel matrix. The demonstration of  $\mathcal{H}$  for  $\mathcal{H}^{7 \times 7}$  is given by,

$$\mathcal{H} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ x_2 & x_4 & x_5 & x_8 & x_9 & x_{10} & x_{11} \\ x_3 & x_6 & x_7 & x_{12} & x_{13} & x_{14} & x_{15} \\ x_4 & x_8 & x_9 & x_{16} & x_{17} & x_{18} & x_{19} \\ x_6 & x_{12} & x_{13} & x_{24} & x_{25} & x_{26} & x_{27} \\ x_5 & x_{10} & x_{11} & x_{20} & x_{21} & x_{22} & x_{23} \\ x_7 & x_{14} & x_{15} & x_{28} & x_{29} & x_{30} & x_{31} \end{bmatrix} \quad (3.3)$$

The optimization vector  $\mathcal{P}$  consists of the same variables  $x_i$  of the Hankel matrix and it is given by

$$\mathcal{P} = \left[ x_1 \quad x_2 \quad \cdots \quad x_{30} \quad x_{31} \right]^\top \quad (3.4)$$

Here, the dimension of  $\mathcal{H} \in \mathbb{R}^{m \times m}$  and  $m$  is arbitrarily chosen. The length of the vectors  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is equal to number of distinct variables in  $\mathcal{H}$ . The optimization problem given in 3.2. minimizes the nuclear norm of the Hankel matrix  $\mathcal{H}$  associated with the variables  $x_i \in [0, 1]$  subject to the probabilities  $\mathcal{P}$  are close to the approximate string probabilities  $\tilde{\mathcal{P}}$  computed from observation sequence.

In the first part of the algorithm, we aim to obtain the correct rank estimation of the Hankel matrix  $\mathcal{H}$ . One common approach to get the rank estimation is singular value decomposition (SVD). In this approach, the singular values are computed and the rate of decrease on singular values are observed. For sufficiently large observation data  $\mathcal{O}_{1:T}$ , the estimated rank of the matrix is decided by observing the significant gap between  $N_{th}$  and  $(N + 1)_{th}$  largest singular value. Following this method, we first make different rank assumptions of the Hankel matrix and we run the algorithm for each rank assumption  $N = 1, 2, \dots$ . The rank assumption is performed by setting a threshold value  $\gamma$  between  $N_{th}$  and  $(N + 1)_{th}$  singular values.

$$\sigma_{N+1}(\mathcal{H})/\sigma_N(\mathcal{H}) \leq \gamma \quad (3.5)$$

We set  $\gamma = 10^{-3}$  to ensure the gap between  $N_{th}$  and  $(N + 1)_{th}$  largest singular

value. Then, we run the algorithm for each assumption, and  $\lambda$  value is updated until the threshold value  $\gamma$  is achieved, in other words assumed rank of the Hankel matrix is obtained. The update of  $\lambda$  value is performed as

$$\lambda^{(t+1)} = \lambda^{(t)} + 0.1 \times 10^{-(L_{seq}+2)} \quad (3.6)$$

Then, we measure the distance between reconstructed Hankel matrix  $\mathcal{H}_n$  and the Hankel matrix consist of approximate string probabilities  $\tilde{\mathcal{H}}$ . The distance between two matrices is measured with Frobenius norm and Kullback-Leibler divergence. Hence, the rank assumption that gives the minimum distance is fixed as rank of the Hankel matrix (order of the HMM).

In the second part of the algorithm, we perform the non-negative matrix factorization algorithm in order to reconstruct the system matrices of the model. The non-negative matrix factorization algorithm for Moore HMMs is developed in [4]. Since we use this algorithm as a part of our algorithm, it is just given here for completeness.

---

**Algorithm 2** The Non-negative Matrix Factorization Algorithm for Moore HMM
 

---

- 1: Give the tensor  $P \in \mathbb{R}_+^{|\mathbb{Y}| \times |\mathbb{Y}|}$
- 2: Set *tol* value
- 3: Given  $N$  and  $M$ , Generate random  $U^{(0)} \in \mathbb{R}_+^{N \times N}$  and  $V^{(0)} \in \mathbb{R}_+^{M \times N}$  such that  $\sum_{ij} U_{ij}^{(0)} = \sum_{ij} P_{kl}$  and  $\sum_k V_{ki}^{(0)} = 1$ .
- 4: **for**  $t = 1, 2, \dots$  **do**
- 5:   Calculate
  1.  $U_{ij}^{(t+1)} = U_{ij}^{(t)} \sum_{\mu\nu} V_{\mu i}^{(t)} V_{\nu j}^{(t)} \frac{P_{\mu\nu}}{(V^{(t)}U^{(t)}(V^{(t)})^\top)_{\mu\nu}}$
  2. 
$$V_{ki}^{(t+1)} = V_{ki}^{(t)} \sum_{\lambda\nu} \frac{P_{k\nu}}{(V^{(t)}U^{(t+1)}(V^{(t)})^\top)_{k\nu}} U_{i\lambda}^{(t+1)} U_{v\lambda}^{(t)} + \frac{P_{\nu k}}{(V^{(t)}U^{(t+1)}(V^{(t)})^\top)_{\nu k}} U_{\lambda i}^{(t+1)} V_{v\lambda}^{(t)}$$
  3. Normalize  $V^{(t+1)}$  such that  $e^\top V^{(t+1)} = e^\top$
- 6:   **if**  $D_{KL}(P \parallel VUV^\top) = \sum_{\mu\nu} P_{\mu\nu} \log \frac{P_{\mu\nu}}{(VUV^\top)_{\mu\nu}} - P_{\mu\nu} + (VUV^\top)_{\mu\nu} \leq \textit{tol}$  **then**
- 7:     **exit**
- 8:   **else**
- 9:     Go to step #4
- 10:   **end if**
- 11: **end for**
- 12: Reconstruct the system matrices  $A_n$  and  $B_n$

$$A_n = (\textit{diag}(Ue))^{-1}U, \quad B_n = V^\top$$


---

In this algorithm, the non-negative matrix factorization is applied to the tensor  $P \in \mathbb{R}_+^{|\mathbb{Y}| \times |\mathbb{Y}| \times \dots \times |\mathbb{Y}|}$  of string probabilities of strings of length  $t$ . This tensor becomes  $P \in \mathbb{R}_+^{|\mathbb{Y}| \times |\mathbb{Y}|}$  for  $t = 2$  and given by

$$P = \begin{bmatrix} \mathcal{P}^2(x_1x_1) & \mathcal{P}^2(x_1x_2) & \cdots & \mathcal{P}^2(x_1x_{|\mathbb{Y}|}) \\ \mathcal{P}^2(x_2x_1) & \mathcal{P}^2(x_2x_2) & \cdots & \mathcal{P}^2(x_2x_{|\mathbb{Y}|}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{P}^2(x_{|\mathbb{Y}|}x_1) & \mathcal{P}^2(x_{|\mathbb{Y}|}x_2) & \cdots & \mathcal{P}^2(x_{|\mathbb{Y}|}x_{|\mathbb{Y}|}) \end{bmatrix}$$

Then, this matrix  $P$  can be (approximately) decomposed into a product  $VUV^\top$ ,

with  $V \in \mathbb{R}_+^{M \times N}$  and  $U \in \mathbb{R}_+^{N \times N}$ ,  $V$  column stochastic and the element sum of  $U$  is equal to the element sum of  $P$ . Thus, from an approximate decomposition  $P \simeq VUV^\top$ , the matrices  $A, B$  can be reconstructed as follows

$$A = (\text{diag}(Ue))^{-1}U,$$

$$B = V^\top$$



# Simulation and Results

## 4.1 Simulation Results

In this section, we present some numerical examples to demonstrate the behavior of the proposed algorithm. We performed simulations for different case by changing the number of states  $N$ , strings of length  $t$  and sequence lengths  $L_{seq}$  in order to show the effects of aforementioned parameters on the algorithm. As it is illustrated in section 2.3, number of outputs of the given model specify the size of the Hankel matrix along with the (approximate) string probabilities of strings of length  $t$ . In these simulations, the number of outputs  $M$  is fixed as 2. In the figures, the log scale is used to ease the visibility of probabilistic values.

We performed simulations for four different cases. Here, we present first two cases and we leave Case 3 and Case 4 to the appendix. In Case 1 and Case 2, we computed string probabilities of strings of length  $t = 2$ . In these simulations, the distance between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is given by  $\|\tilde{\mathcal{H}} - \mathcal{H}\|_F$  instead of  $\|\tilde{\mathcal{P}} - \mathcal{P}\|_2$  since the matrices  $\tilde{\mathcal{H}}$  and  $\mathcal{H}$  is constructed by  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  respectively.

### 4.1.1 Case 1: $t = 2, N = 2$

Here, given the state transition matrix  $A$ , the output matrix  $B$  and (approximate) string probabilities of strings of length  $t = 2$ , we obtain the  $\lambda$  value that makes the order of the model is closest the real order of the system for different sequence

lengths  $L_{seq}$ . Here,  $A$  and  $B$  matrices are arbitrarily chosen.

Consider the process described by  $(A, B)$ ;

$$A = \begin{bmatrix} 0.314 & 0.686 \\ 0.651 & 0.349 \end{bmatrix}, \quad B = \begin{bmatrix} 0.240 & 0.760 \\ 0.591 & 0.409 \end{bmatrix}. \quad (4.1)$$

The Hankel matrix consist of real string probabilities  $\mathcal{H}_r$  corresponding to  $(A, B)$  is given by 4.1. Then, since the rank of the Hankel matrix consist of real string probabilities is equal to the order of the model, the singular values of  $\mathcal{H}_r$  i.e.  $\sigma(\mathcal{H}_r)$  is obtained as follows;

$$\sigma(H_r) = \begin{bmatrix} 1.776 \\ 3.222 \times 10^{-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.2)$$

and the  $\mathcal{H}_r$  is;

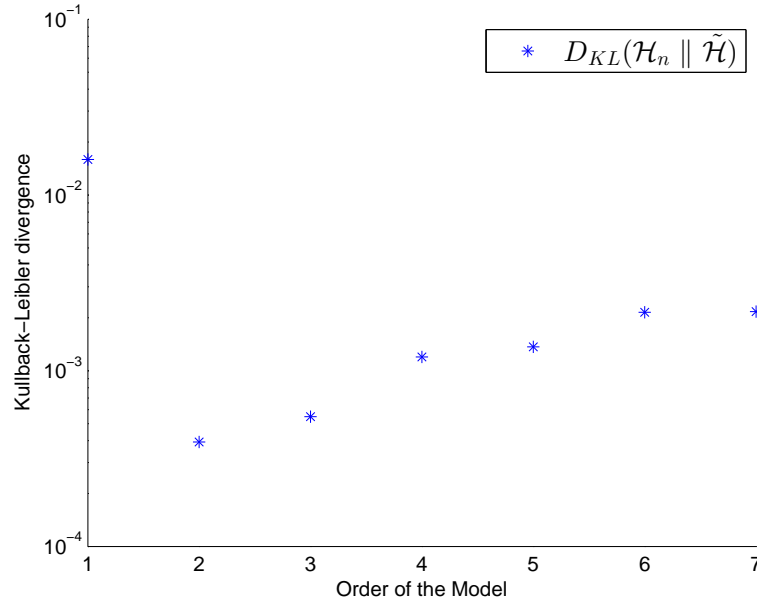
$$\mathcal{H}_r = \begin{bmatrix} 1 & 0.420 & 0.580 & 0.166 & 0.254 & 0.254 & 0.326 \\ 0.420 & 0.166 & 0.254 & 0.068 & 0.099 & 0.113 & 0.141 \\ 0.580 & 0.254 & 0.326 & 0.099 & 0.155 & 0.141 & 0.185 \\ 0.166 & 0.068 & 0.099 & 0.027 & 0.040 & 0.044 & 0.055 \\ 0.254 & 0.099 & 0.155 & 0.040 & 0.059 & 0.069 & 0.086 \\ 0.254 & 0.113 & 0.141 & 0.044 & 0.069 & 0.060 & 0.080 \\ 0.326 & 0.141 & 0.185 & 0.055 & 0.086 & 0.080 & 0.105 \end{bmatrix} \quad (4.3)$$

It is clear from the construction of  $\mathcal{H}_r$  that the probabilities in  $\mathcal{H}_r$  will be the same for all sequence lengths  $L_{seq}$ . In the simulation example, this model and the Hankel matrix corresponding to real probabilities  $\mathcal{H}_r$  are unknown; however, we provide them here to evaluate the performance of the algorithms.

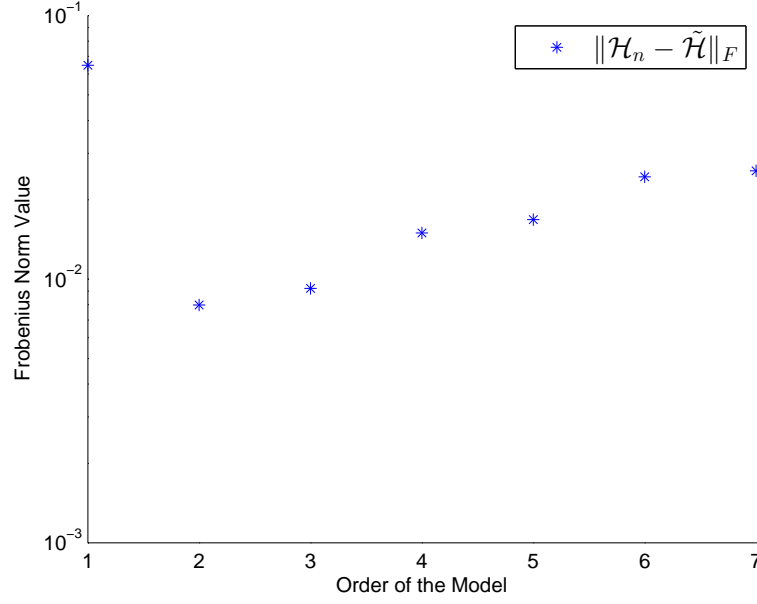
It is stated in Section 3.2. that in order to obtain best rank assumption of the Hankel matrix  $\mathcal{H}$ , the rank assumptions are made. It is performed by setting a

threshold value between  $N_{th}$  and  $(N + 1)_{th}$  singular values of  $\mathcal{H}$ . Here, we provide results of different rank assumptions for the sequence length  $T = 10^4$ . For the simulations of sequence lengths other than  $T = 10^4$ , it is assumed that the obtained rank (order) is the same. Because, as the length of sequence increases, the accuracy of the algorithm increases.

We simulated the process described by 4.1, and produced a sequence of  $T = 10^4$  observations. Then, we run the algorithm for the system order 1,2,...,7 (for each assumption). Figures 4.1 and 4.2 show the Kullback-Leibler divergence and Frobenius norm between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  as a function of the system order respectively.



**Figure 4.1.** Kullback-Leibler divergence between the  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  for system order 1,2,...,7 for Case 1



**Figure 4.2.** Frobenius norm value between the  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  for system order 1,2,...,7 for Case 1

Figure 4.1 and 4.2 show that the rank assumption that gives the minimum distance between  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  is 2. Then we fix the rank of Hankel matrix (order of HMM) to 2.

### Simulation 1.

1. We simulated this process and produced a sequence of  $T = 10^6$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.419 & 0.581 & 0.166 & 0.254 & 0.254 & 0.327 \\ 0.419 & 0.166 & 0.254 & 0.066 & 0.099 & 0.113 & 0.141 \\ 0.581 & 0.254 & 0.327 & 0.099 & 0.154 & 0.141 & 0.186 \\ 0.166 & 0.066 & 0.099 & 0.026 & 0.040 & 0.044 & 0.055 \\ 0.254 & 0.099 & 0.154 & 0.040 & 0.059 & 0.069 & 0.086 \\ 0.254 & 0.113 & 0.141 & 0.044 & 0.069 & 0.060 & 0.081 \\ 0.327 & 0.141 & 0.186 & 0.055 & 0.085 & 0.081 & 0.105 \end{bmatrix}$$

These values demonstrate that the approximate string probabilities differ from the original ones. The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.776 \\ 3.234 \times 10^{-2} \\ 5.118 \times 10^{-4} \\ 3.780 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

As it is seen from  $\sigma(\tilde{\mathcal{H}})$ , the gap between  $2_{nd}$  and  $3_{rd}$  singular value is not significant. So, it cannot be assumed that the rank of  $\tilde{\mathcal{H}}$  is 2.

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.419 & 0.581 & 0.166 & 0.254 & 0.254 & 0.327 \\ 0.419 & 0.166 & 0.254 & 0.066 & 0.099 & 0.113 & 0.141 \\ 0.581 & 0.254 & 0.327 & 0.099 & 0.154 & 0.141 & 0.186 \\ 0.166 & 0.066 & 0.099 & 0.026 & 0.040 & 0.044 & 0.055 \\ 0.254 & 0.099 & 0.154 & 0.040 & 0.059 & 0.069 & 0.086 \\ 0.254 & 0.113 & 0.141 & 0.044 & 0.069 & 0.060 & 0.081 \\ 0.327 & 0.141 & 0.186 & 0.055 & 0.085 & 0.081 & 0.105 \end{bmatrix}$$

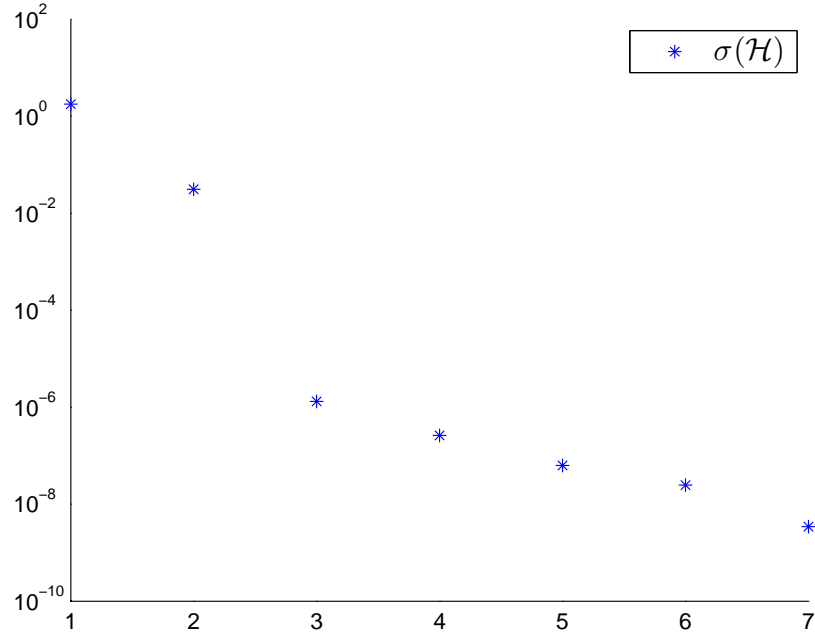
The  $\lambda$  value obtained as  $4 \times 10^{-7}$ .

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

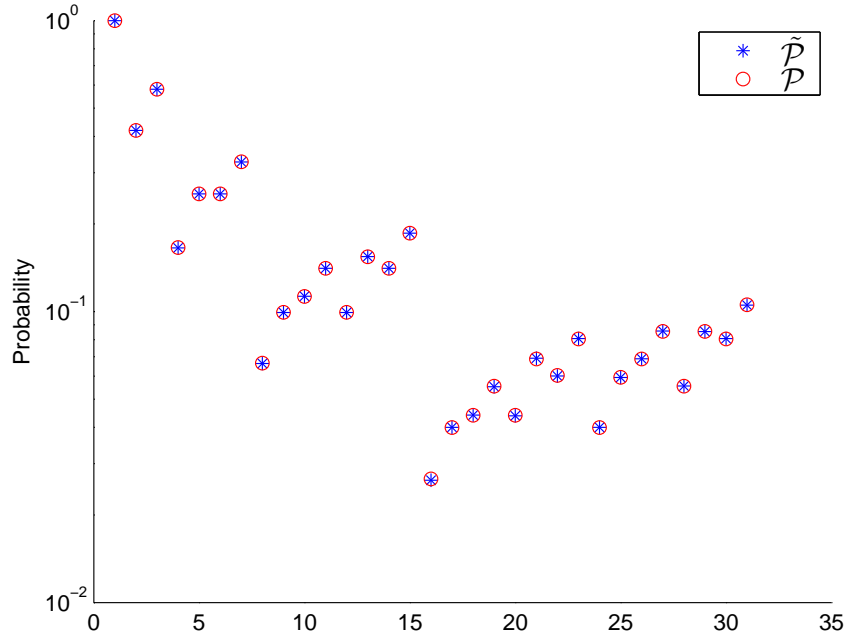
$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.776 \\ 3.180 \times 10^{-2} \\ 1.522 \times 10^{-6} \\ 3.754 \times 10^{-7} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}\|_F = 1.054 \times 10^{-3}$$



**Figure 4.3.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^6$



**Figure 4.4.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^6$

This suggests that estimated rank of  $\mathcal{H}$  i.e.,  $rank_{est}(\mathcal{H}) = 2$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

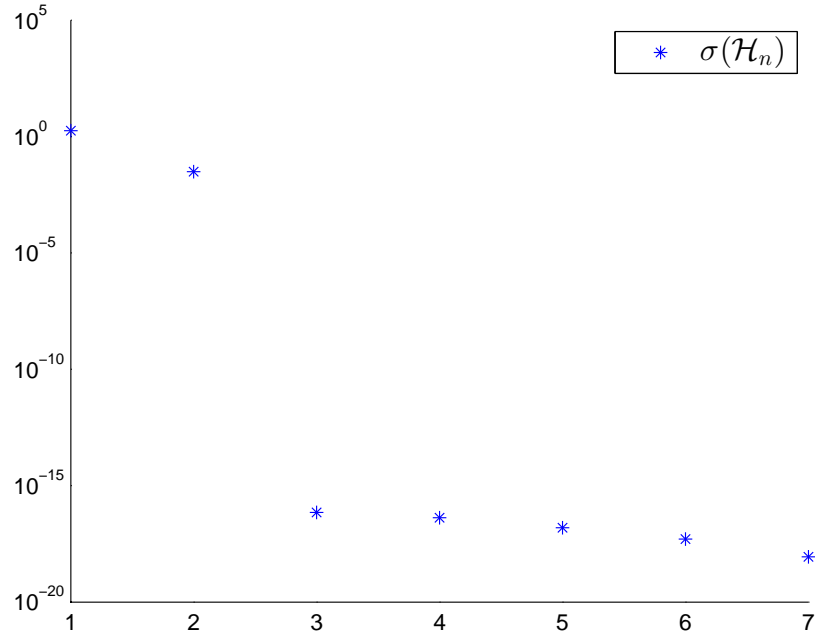
$$A_n = \begin{bmatrix} 0.321 & 0.679 \\ 0.656 & 0.344 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.243 & 0.757 \\ 0.59 & 0.41 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.419 & 0.581 & 0.166 & 0.254 & 0.254 & 0.327 \\ 0.419 & 0.166 & 0.254 & 0.067 & 0.099 & 0.113 & 0.141 \\ 0.581 & 0.254 & 0.327 & 0.099 & 0.155 & 0.141 & 0.186 \\ 0.166 & 0.067 & 0.099 & 0.027 & 0.040 & 0.044 & 0.055 \\ 0.254 & 0.099 & 0.155 & 0.040 & 0.059 & 0.069 & 0.086 \\ 0.254 & 0.113 & 0.141 & 0.044 & 0.069 & 0.061 & 0.080 \\ 0.326 & 0.141 & 0.186 & 0.055 & 0.086 & 0.080 & 0.105 \end{bmatrix}$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.776 \\ 3.165 \times 10^{-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure 4.5.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^6$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 2.544 \times 10^{-3}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.



## Simulation 2.

1. We simulated the same process and produced a sequence of  $T = 10^5$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.417 & 0.583 & 0.165 & 0.253 & 0.253 & 0.33 \\ 0.417 & 0.165 & 0.253 & 0.066 & 0.099 & 0.111 & 0.142 \\ 0.583 & 0.253 & 0.33 & 0.099 & 0.153 & 0.142 & 0.188 \\ 0.165 & 0.066 & 0.099 & 0.026 & 0.040 & 0.044 & 0.056 \\ 0.253 & 0.099 & 0.153 & 0.040 & 0.059 & 0.067 & 0.086 \\ 0.253 & 0.111 & 0.142 & 0.043 & 0.068 & 0.060 & 0.082 \\ 0.33 & 0.142 & 0.188 & 0.056 & 0.085 & 0.082 & 0.106 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.776 \\ 3.089 \times 10^{-2} \\ 2.099 \times 10^{-3} \\ 9.786 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.418 & 0.582 & 0.166 & 0.252 & 0.252 & 0.33 \\ 0.418 & 0.166 & 0.252 & 0.066 & 0.099 & 0.11 & 0.142 \\ 0.582 & 0.252 & 0.33 & 0.099 & 0.153 & 0.142 & 0.188 \\ 0.166 & 0.066 & 0.099 & 0.027 & 0.04 & 0.043 & 0.056 \\ 0.252 & 0.099 & 0.153 & 0.04 & 0.059 & 0.067 & 0.086 \\ 0.252 & 0.11 & 0.142 & 0.043 & 0.067 & 0.061 & 0.081 \\ 0.33 & 0.142 & 0.188 & 0.056 & 0.086 & 0.081 & 0.107 \end{bmatrix}$$

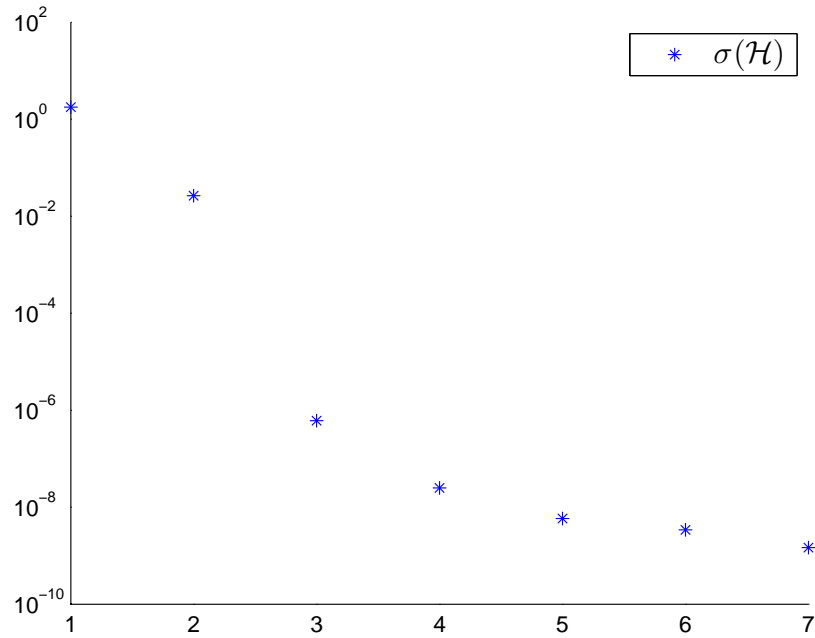
The  $\lambda$  value obtained as  $6 \times 10^{-6}$ .

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

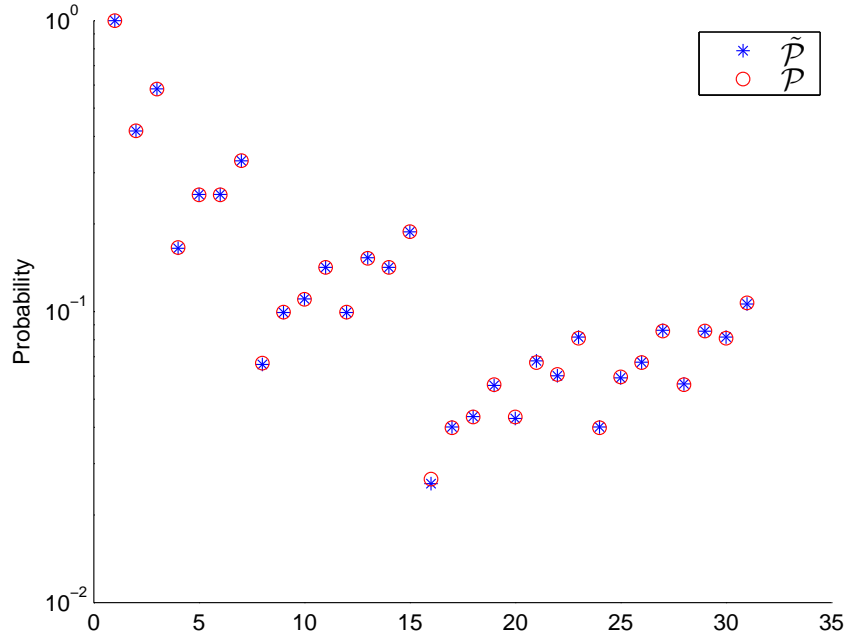
$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.777 \\ 2.855 \times 10^{-2} \\ 1.417 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}\|_F = 4.176 \times 10^{-3}$$



**Figure 4.6.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^5$



**Figure 4.7.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^5$

This again suggests that  $rank_{est}(\mathcal{H}) = 2$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

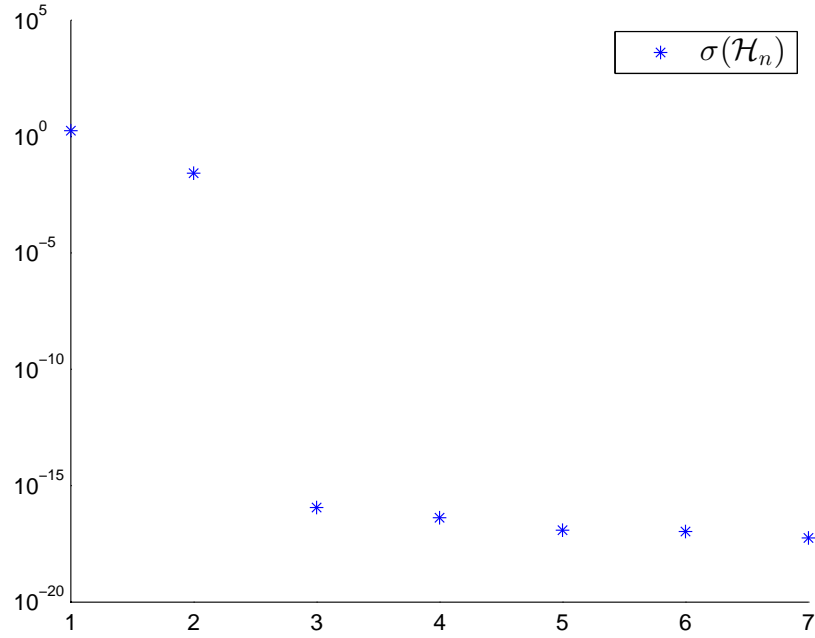
$$A_n = \begin{bmatrix} 0.311 & 0.689 \\ 0.548 & 0.452 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.632 & 0.368 \\ 0.248 & 0.752 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.418 & 0.582 & 0.166 & 0.252 & 0.252 & 0.33 \\ 0.418 & 0.166 & 0.252 & 0.066 & 0.099 & 0.11 & 0.142 \\ 0.582 & 0.252 & 0.33 & 0.099 & 0.153 & 0.142 & 0.188 \\ 0.166 & 0.066 & 0.099 & 0.027 & 0.04 & 0.044 & 0.056 \\ 0.252 & 0.099 & 0.153 & 0.04 & 0.059 & 0.067 & 0.086 \\ 0.252 & 0.11 & 0.142 & 0.043 & 0.067 & 0.061 & 0.081 \\ 0.33 & 0.142 & 0.188 & 0.056 & 0.086 & 0.081 & 0.107 \end{bmatrix}$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.776 \\ 2.846 \times 10^{-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure 4.8.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^5$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 5.297 \times 10^{-3}$$

It is seen that the values in Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.

### Simulation 3.

1. We simulated this process and produced a sequence of  $T = 10^4$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.421 & 0.579 & 0.167 & 0.254 & 0.254 & 0.325 \\ 0.421 & 0.167 & 0.254 & 0.0652 & 0.101 & 0.114 & 0.14 \\ 0.579 & 0.254 & 0.325 & 0.101 & 0.153 & 0.14 & 0.185 \\ 0.167 & 0.065 & 0.101 & 0.026 & 0.039 & 0.045 & 0.056 \\ 0.254 & 0.101 & 0.153 & 0.0391 & 0.062 & 0.068 & 0.084 \\ 0.254 & 0.114 & 0.14 & 0.0458 & 0.068 & 0.059 & 0.080 \\ 0.325 & 0.14 & 0.185 & 0.0555 & 0.085 & 0.081 & 0.104 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.775 \\ 2.701 \times 10^{-2} \\ 9.146 \times 10^{-3} \\ 3.409 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.421 & 0.579 & 0.168 & 0.253 & 0.253 & 0.326 \\ 0.421 & 0.168 & 0.253 & 0.067 & 0.102 & 0.113 & 0.141 \\ 0.579 & 0.253 & 0.326 & 0.102 & 0.151 & 0.14 & 0.185 \\ 0.168 & 0.067 & 0.102 & 0.026 & 0.040 & 0.045 & 0.056 \\ 0.253 & 0.102 & 0.151 & 0.040 & 0.0613 & 0.067 & 0.0841 \\ 0.253 & 0.113 & 0.141 & 0.045 & 0.067 & 0.060 & 0.080 \\ 0.326 & 0.14 & 0.185 & 0.056 & 0.084 & 0.080 & 0.105 \end{bmatrix}$$

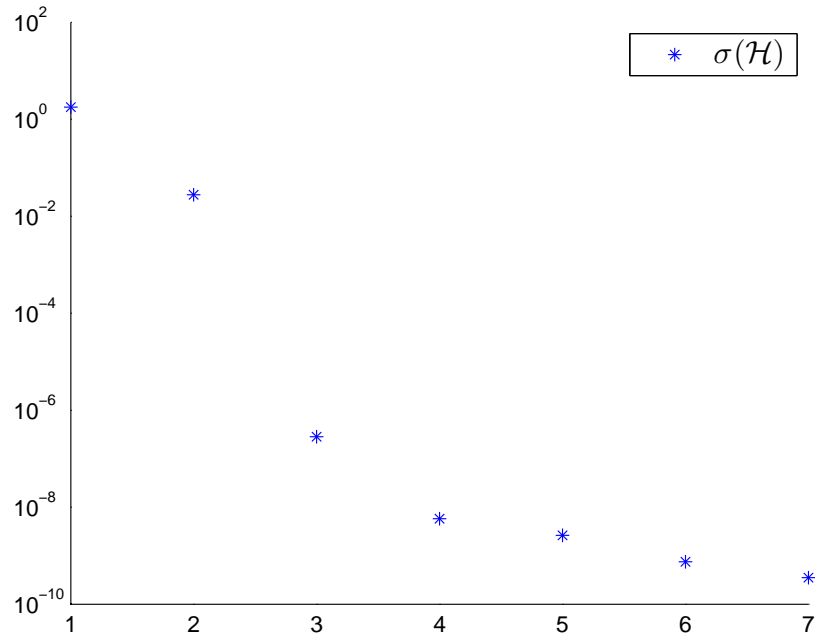
The  $\lambda$  value obtained as  $8.1 \times 10^{-5}$ .

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

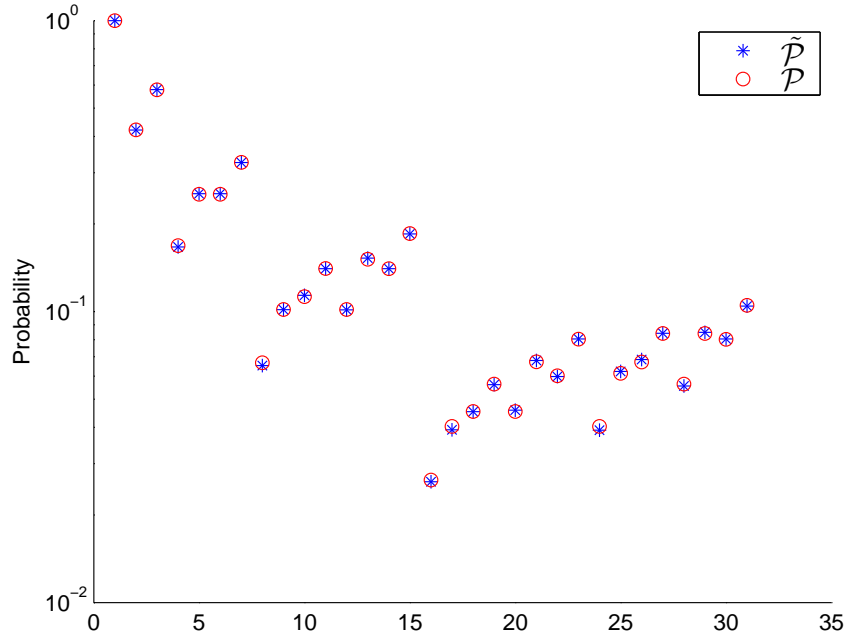
$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.774 \\ 2.632 \times 10^{-2} \\ 8.837 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}\|_F = 1.568 \times 10^{-2}$$



**Figure 4.9.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^4$



**Figure 4.10.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^4$

This again suggests that  $rank_{est}(\mathcal{H}) = 2$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

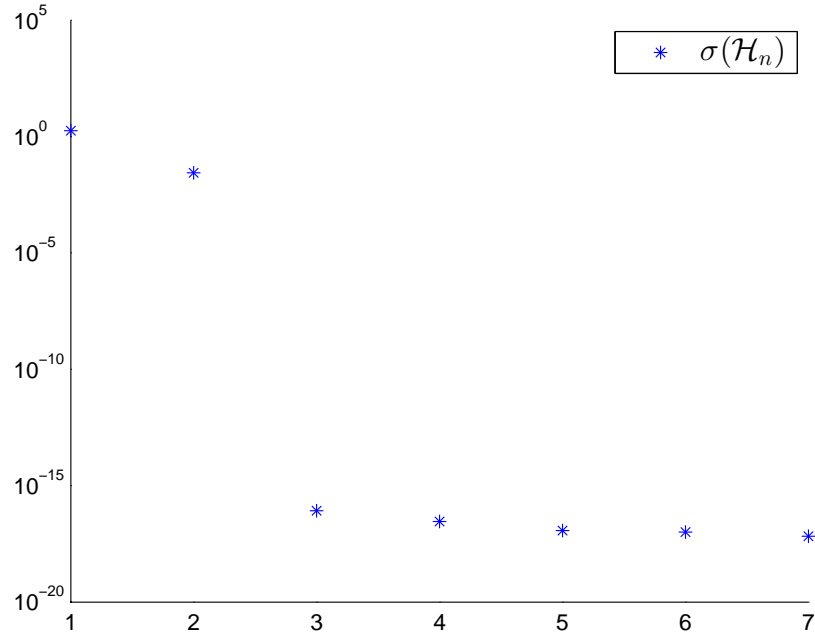
$$A_n = \begin{bmatrix} 0.233 & 0.767 \\ 0.393 & 0.607 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.087 & 0.913 \\ 0.593 & 0.407 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1.0 & 0.421 & 0.579 & 0.168 & 0.253 & 0.253 & 0.326 \\ 0.421 & 0.168 & 0.253 & 0.0683 & 0.1 & 0.112 & 0.141 \\ 0.579 & 0.253 & 0.326 & 0.1 & 0.153 & 0.141 & 0.185 \\ 0.168 & 0.0683 & 0.1 & 0.0276 & 0.0407 & 0.0442 & 0.0559 \\ 0.253 & 0.1 & 0.153 & 0.0407 & 0.0594 & 0.068 & 0.0848 \\ 0.253 & 0.112 & 0.141 & 0.0442 & 0.068 & 0.0605 & 0.0803 \\ 0.326 & 0.141 & 0.185 & 0.0558 & 0.0849 & 0.0802 & 0.105 \end{bmatrix}$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.774 \\ 2.621 \times 10^{-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure 4.11.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^4$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 1.803 \times 10^{-2}$$

It is seen that the values in Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.



### Simulation 4.

1. We simulated this process and produced a sequence of  $T = 10^3$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1.0 & 0.427 & 0.573 & 0.163 & 0.264 & 0.263 & 0.31 \\ 0.427 & 0.163 & 0.264 & 0.0662 & 0.0973 & 0.119 & 0.144 \\ 0.573 & 0.263 & 0.31 & 0.0963 & 0.166 & 0.143 & 0.166 \\ 0.163 & 0.0662 & 0.0973 & 0.0271 & 0.0391 & 0.0481 & 0.0491 \\ 0.263 & 0.0963 & 0.166 & 0.0381 & 0.0582 & 0.0712 & 0.0953 \\ 0.264 & 0.119 & 0.144 & 0.0471 & 0.0722 & 0.0592 & 0.0853 \\ 0.31 & 0.143 & 0.166 & 0.0491 & 0.0943 & 0.0843 & 0.0822 \end{bmatrix}$$

These values demonstrate that the approximate string probabilities differ from the original ones. The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.771 \\ 6.051 \times 10^{-2} \\ 1.643 \times 10^{-2} \\ 5.256 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.4)$$

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.429 & 0.571 & 0.17 & 0.26 & 0.259 & 0.312 \\ 0.429 & 0.17 & 0.26 & 0.0698 & 0.0997 & 0.115 & 0.145 \\ 0.571 & 0.259 & 0.312 & 0.0995 & 0.159 & 0.144 & 0.168 \\ 0.17 & 0.0698 & 0.0997 & 0.0281 & 0.0417 & 0.0447 & 0.055 \\ 0.259 & 0.0995 & 0.159 & 0.0415 & 0.058 & 0.0701 & 0.089 \\ 0.26 & 0.115 & 0.145 & 0.0448 & 0.0704 & 0.0661 & 0.0785 \\ 0.312 & 0.144 & 0.168 & 0.0548 & 0.0889 & 0.078 & 0.0903 \end{bmatrix}$$

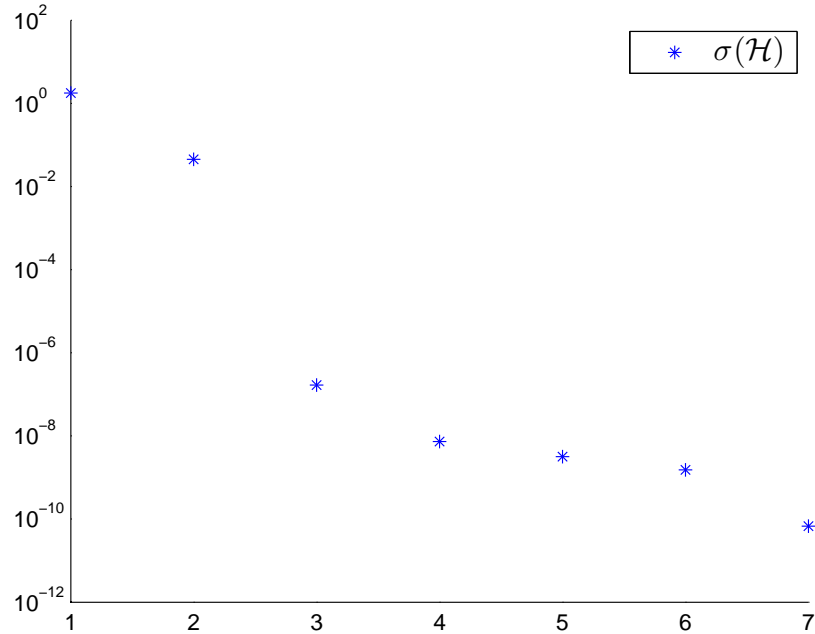
The  $\lambda$  value obtained as  $5 \times 10^{-4}$ .

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

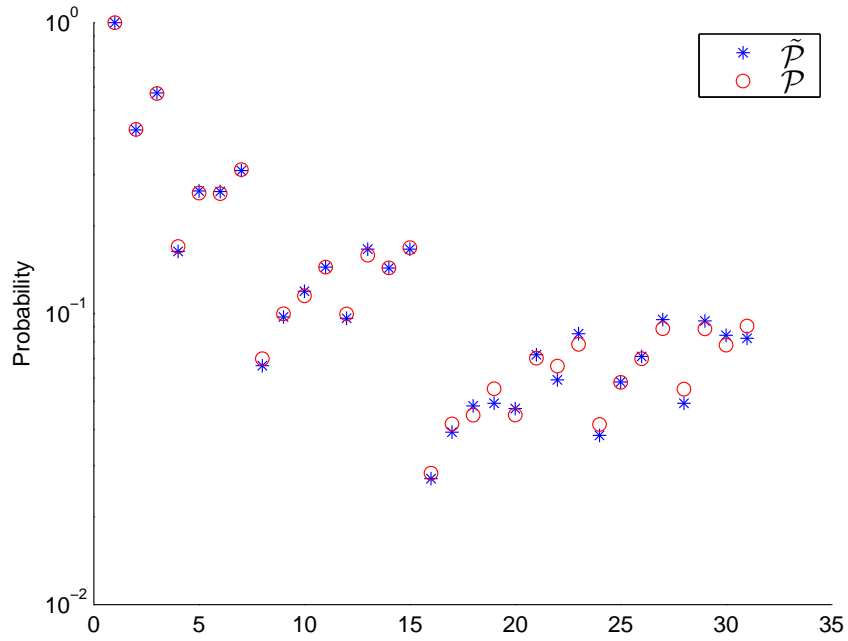
$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.770 \\ 4.513 \times 10^{-2} \\ 1.671 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}\|_F = 2.989 \times 10^{-2}$$



**Figure 4.12.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^3$



**Figure 4.13.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^3$

This again suggests that  $rank_{est}(\mathcal{H}) = 2$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

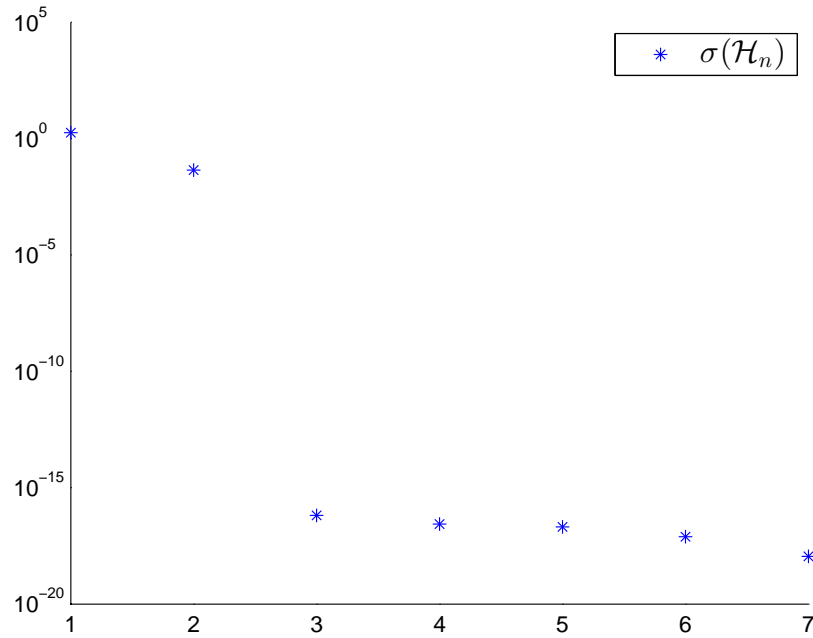
$$A_n = \begin{bmatrix} 0.319 & 0.681 \\ 0.699 & 0.301 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.618 & 0.382 \\ 0.234 & 0.766 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.429 & 0.571 & 0.17 & 0.26 & 0.259 & 0.312 \\ 0.429 & 0.17 & 0.26 & 0.069 & 0.101 & 0.12 & 0.14 \\ 0.571 & 0.259 & 0.312 & 0.1 & 0.158 & 0.139 & 0.173 \\ 0.17 & 0.069 & 0.101 & 0.028 & 0.041 & 0.046 & 0.054 \\ 0.259 & 0.1 & 0.158 & 0.041 & 0.059 & 0.074 & 0.085 \\ 0.26 & 0.12 & 0.14 & 0.046 & 0.074 & 0.062 & 0.078 \\ 0.312 & 0.139 & 0.173 & 0.054 & 0.085 & 0.077 & 0.095 \end{bmatrix}$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.770 \\ 4.375 \times 10^{-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.5)$$



**Figure 4.14.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 2$ ,  $L_{seq} = 10^3$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 2.773 \times 10^{-2}$$

It is seen that the values in Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.

### 4.1.2 Case 2: $t = 2, N = 3$

Here, given the state transition matrix  $A$ , the output matrix  $B$  and (approximate) string probabilities of strings of length  $t = 2$ , we obtain the  $\lambda$  value that makes the order of the model is closest the real order of the system for different sequence lengths  $L_{seq}$ . Here,  $A$  and  $B$  matrices are arbitrarily chosen.

Consider the process described by  $(A, B)$ ;

$$A = \begin{bmatrix} 0.522 & 0.143 & 0.345 \\ 0.116 & 0.291 & 0.593 \\ 0.387 & 0.376 & 0.236 \end{bmatrix}, \quad B = \begin{bmatrix} 0.825 & 0.175 \\ 0.288 & 0.712 \\ 0.144 & 0.856 \end{bmatrix}.$$

The Hankel matrix consist of real string probabilities  $\mathcal{H}_r$  corresponding to  $(A, B)$  is given by 4.6 and the singular values of  $\mathcal{H}_r$  i.e.  $\sigma(\mathcal{H}_r)$  is obtained as follows;

$$\sigma(H_r) = \begin{bmatrix} 1.773 \\ 4.769 \times 10^{-2} \\ 2.325 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.6)$$

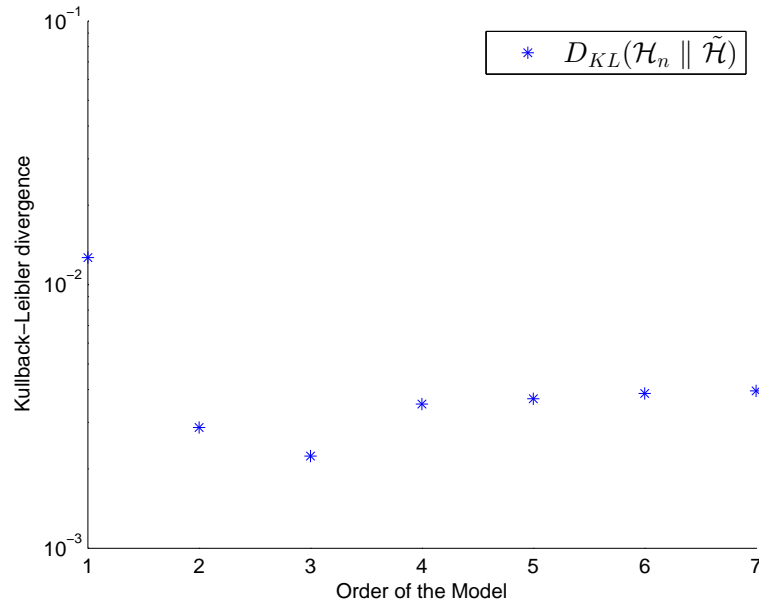
and the  $\mathcal{H}_r$  is;

$$\mathcal{H}_r = \begin{bmatrix} 1 & 0.428 & 0.572 & 0.198 & 0.23 & 0.23 & 0.342 \\ 0.428 & 0.198 & 0.23 & 0.0942 & 0.103 & 0.096 & 0.134 \\ 0.572 & 0.23 & 0.342 & 0.103 & 0.127 & 0.134 & 0.208 \\ 0.198 & 0.0942 & 0.103 & 0.0452 & 0.049 & 0.0433 & 0.0601 \\ 0.23 & 0.103 & 0.127 & 0.049 & 0.0544 & 0.0526 & 0.074 \\ 0.23 & 0.096 & 0.134 & 0.0434 & 0.0525 & 0.0523 & 0.0818 \\ 0.342 & 0.134 & 0.208 & 0.06 & 0.074 & 0.0818 & 0.127 \end{bmatrix} \quad (4.7)$$

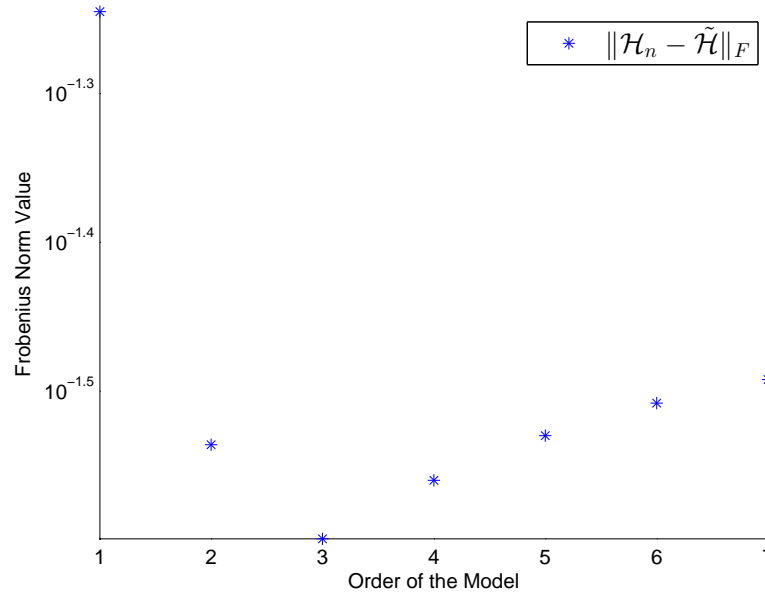
Again, we here provide results of different rank assumptions for the sequence length  $T = 10^4$ . For the simulations of sequence lengths other than  $T = 10^4$ , it is assumed

that the obtained rank (order) is the same.

We simulated the process described by 4.1, and produced a sequence of  $T = 10^4$  observations. Then, we run the algorithm for the system order  $1, 2, \dots, 7$  (for each assumption). Figures 4.15 and 4.16 show the Kullback-Leibler divergence and Frobenius norm between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  as a function of the system order respectively.



**Figure 4.15.** Kullback-Leibler divergence between the  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  for system order  $1, 2, \dots, 7$  for Case 2



**Figure 4.16.** Frobenius norm value between the  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  for system order 1,2,...,7 for Case 2

Figure 4.1 and 4.2 show that the rank assumption that gives the minimum distance between  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  is 3. Then we fix the rank of Hankel matrix (order of HMM) to 3.

### Simulation 5.

1. We simulated this process and produced a sequence of  $T = 10^6$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.428 & 0.572 & 0.198 & 0.23 & 0.23 & 0.342 \\ 0.428 & 0.198 & 0.23 & 0.0943 & 0.103 & 0.0961 & 0.134 \\ 0.572 & 0.23 & 0.342 & 0.103 & 0.127 & 0.134 & 0.208 \\ 0.198 & 0.0943 & 0.103 & 0.0452 & 0.0491 & 0.0435 & 0.0599 \\ 0.23 & 0.103 & 0.127 & 0.0491 & 0.0542 & 0.0527 & 0.074 \\ 0.23 & 0.0961 & 0.134 & 0.0435 & 0.0526 & 0.0523 & 0.0816 \\ 0.342 & 0.134 & 0.208 & 0.0599 & 0.074 & 0.0816 & 0.127 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.773 \\ 4.786 \times 10^{-2} \\ 2.331 \times 10^{-4} \\ 1.505 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.8)$$

As we can see from  $\sigma(\tilde{\mathcal{H}})$ , the gap between 3<sub>rd</sub> and 4<sub>th</sub> singular value is not significant. So, it cannot be assumed that the rank of  $\tilde{\mathcal{H}}$  is 3.

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.428 & 0.572 & 0.198 & 0.23 & 0.23 & 0.342 \\ 0.428 & 0.198 & 0.23 & 0.0942 & 0.103 & 0.0961 & 0.134 \\ 0.572 & 0.23 & 0.342 & 0.103 & 0.127 & 0.134 & 0.208 \\ 0.198 & 0.0942 & 0.103 & 0.0452 & 0.049 & 0.0435 & 0.0599 \\ 0.23 & 0.103 & 0.127 & 0.049 & 0.0544 & 0.0527 & 0.074 \\ 0.23 & 0.0961 & 0.134 & 0.0435 & 0.0526 & 0.0524 & 0.0815 \\ 0.342 & 0.134 & 0.208 & 0.0599 & 0.074 & 0.0815 & 0.127 \end{bmatrix}$$

The  $\lambda$  value obtained as  $2.1 \times 10^{-7}$ .

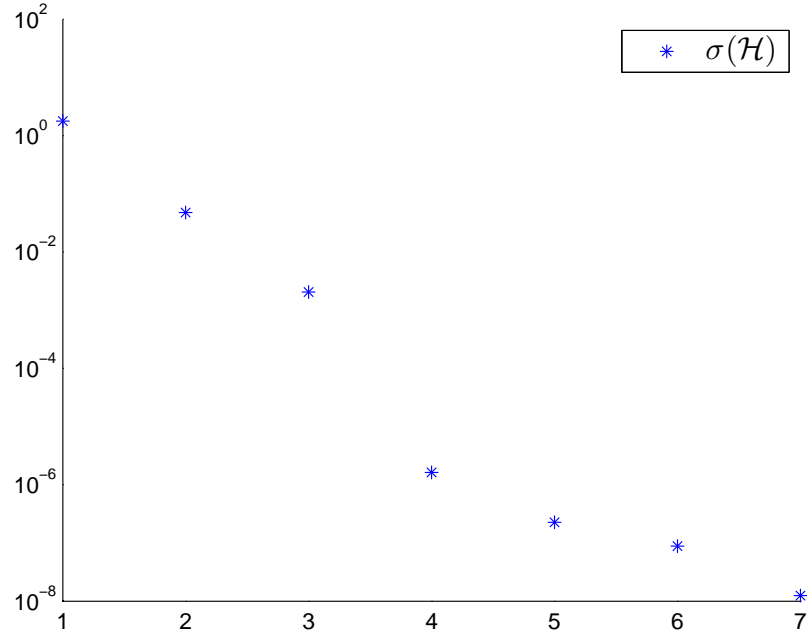
4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}_r) = \begin{bmatrix} 1.773 \\ 4.761 \times 10^{-2} \\ 2.055 \times 10^{-3} \\ 1.631 \times 10^{-6} \\ 2.260 \times 10^{-7} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

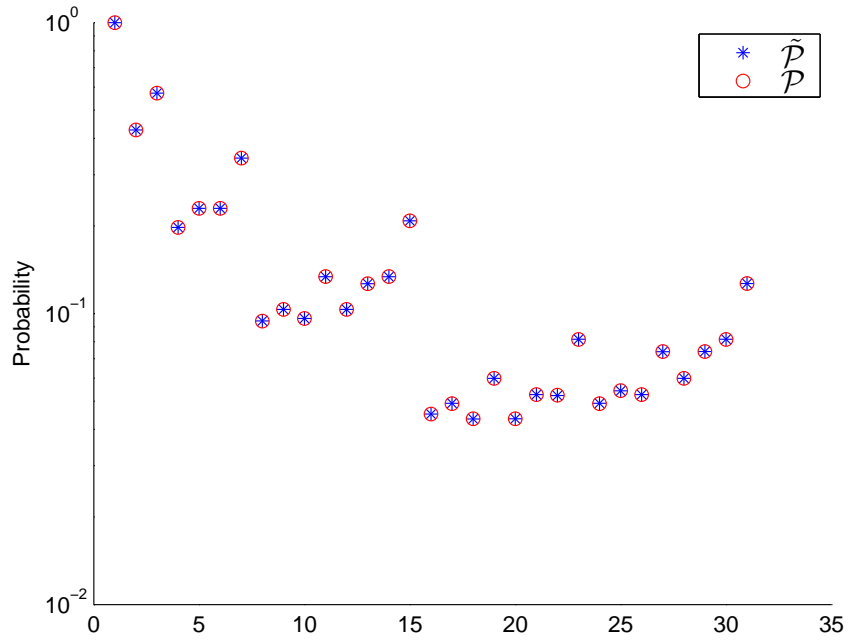


The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}\|_F = 5.895 \times 10^{-4}$$



**Figure 4.17.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^6$



**Figure 4.18.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^6$

This suggests that estimated rank of  $\mathcal{H}$  i.e.,  $rank_{est}(\mathcal{H}) = 3$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

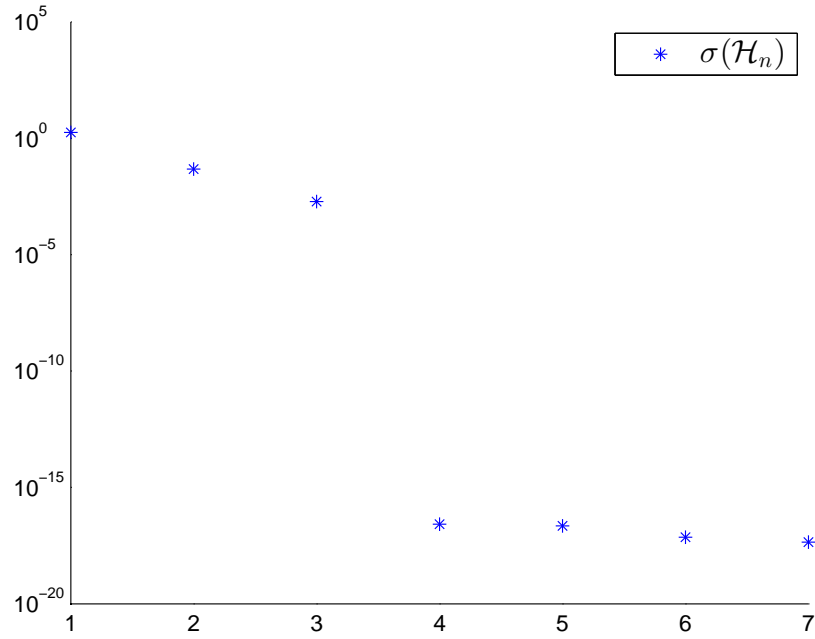
$$A_{new} = \begin{bmatrix} 0.479 & 0.143 & 0.377 \\ 0.171 & 0.567 & 0.262 \\ 0.38 & 0.308 & 0.312 \end{bmatrix}, \quad B_{new} = \begin{bmatrix} 0.212 & 0.788 \\ 0.7 & 0.3 \\ 0.365 & 0.635 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.428 & 0.572 & 0.198 & 0.23 & 0.23 & 0.342 \\ 0.428 & 0.198 & 0.23 & 0.0942 & 0.103 & 0.094 & 0.136 \\ 0.572 & 0.23 & 0.342 & 0.105 & 0.125 & 0.134 & 0.208 \\ 0.198 & 0.0942 & 0.103 & 0.0453 & 0.0489 & 0.0428 & 0.0607 \\ 0.23 & 0.105 & 0.125 & 0.0496 & 0.055 & 0.051 & 0.0744 \\ 0.23 & 0.094 & 0.136 & 0.043 & 0.051 & 0.0535 & 0.0825 \\ 0.342 & 0.134 & 0.208 & 0.0604 & 0.0736 & 0.081 & 0.127 \end{bmatrix}.$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.773 \\ 4.775 \times 10^{-2} \\ 2.177 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure 4.19.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^4$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 5.022 \times 10^{-3}$$

It is seen that the values in Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.

### Simulation 6.

1. We simulated the same process and produced a sequence of  $T = 10^5$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.428 & 0.572 & 0.198 & 0.23 & 0.23 & 0.342 \\ 0.428 & 0.198 & 0.23 & 0.0947 & 0.103 & 0.096 & 0.134 \\ 0.572 & 0.23 & 0.342 & 0.103 & 0.127 & 0.134 & 0.207 \\ 0.198 & 0.0947 & 0.103 & 0.045 & 0.0497 & 0.0432 & 0.0599 \\ 0.23 & 0.103 & 0.127 & 0.0497 & 0.0533 & 0.0529 & 0.0743 \\ 0.23 & 0.096 & 0.134 & 0.0431 & 0.0529 & 0.0533 & 0.0809 \\ 0.342 & 0.134 & 0.207 & 0.06 & 0.0743 & 0.0809 & 0.127 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.773 \\ 4.708 \times 10^{-2} \\ 2.709 \times 10^{-3} \\ 3.527 \times 10^{-4} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (4.9)$$

As we can see from  $\sigma(\tilde{\mathcal{H}})$ , the gap between 3<sub>rd</sub> and 4<sub>th</sub> singular value is not significant. So, it cannot be assumed that the rank of  $\tilde{\mathcal{H}}$  is 3.

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.428 & 0.572 & 0.198 & 0.23 & 0.23 & 0.341 \\ 0.428 & 0.198 & 0.23 & 0.0943 & 0.103 & 0.0961 & 0.134 \\ 0.572 & 0.23 & 0.341 & 0.103 & 0.127 & 0.134 & 0.207 \\ 0.198 & 0.0943 & 0.103 & 0.045 & 0.0495 & 0.0434 & 0.0598 \\ 0.23 & 0.103 & 0.127 & 0.0495 & 0.0537 & 0.0527 & 0.0745 \\ 0.23 & 0.0961 & 0.134 & 0.0434 & 0.0528 & 0.0533 & 0.081 \\ 0.341 & 0.134 & 0.207 & 0.0598 & 0.0745 & 0.081 & 0.126 \end{bmatrix}$$

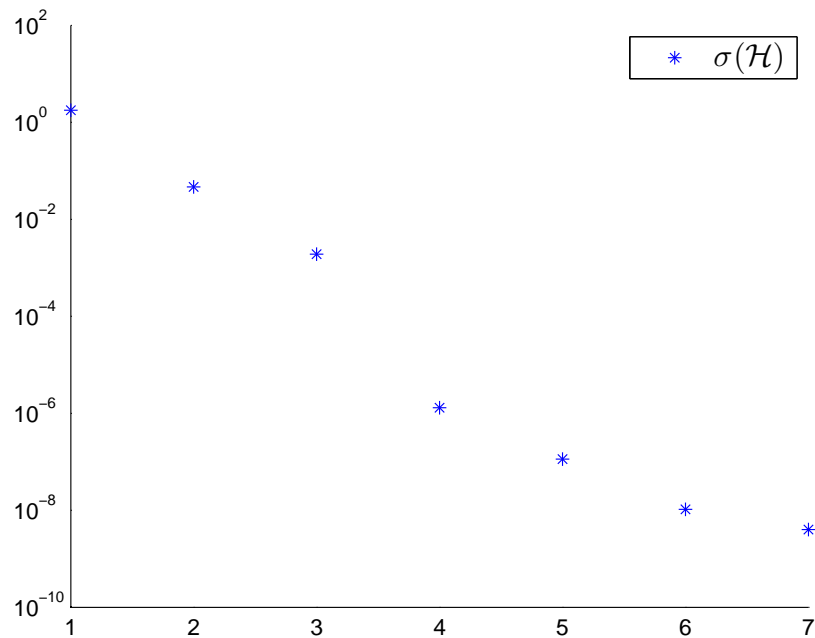
The  $\lambda$  value obtained as  $1.5 \times 10^{-6}$ .

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

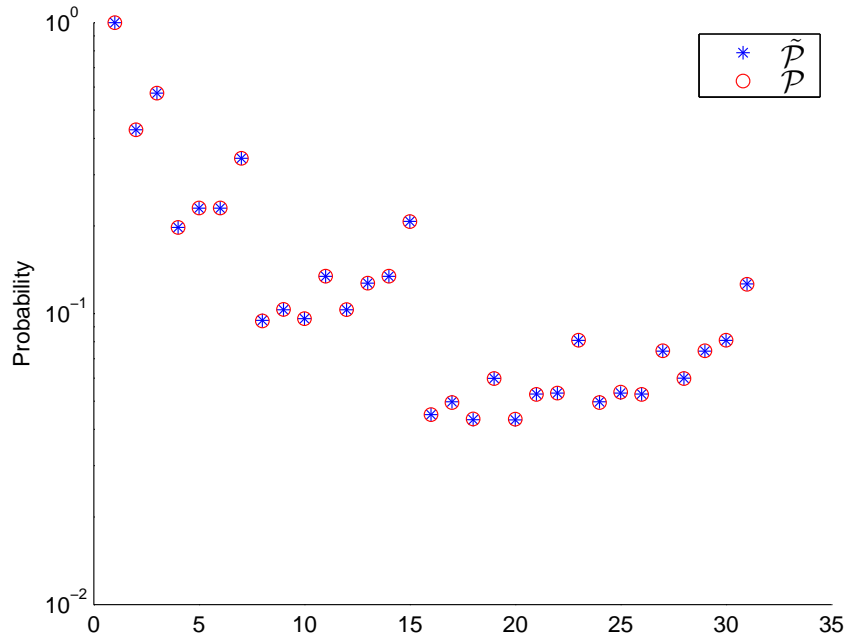
$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.773 \\ 4.638 \times 10^{-2} \\ 1.916 \times 10^{-3} \\ 1.302 \times 10^{-6} \\ 1.133 \times 10^{-7} \\ 0 \\ 0 \end{bmatrix}$$

The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}\|_F = 1.742 \times 10^{-3}$$



**Figure 4.20.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^5$



**Figure 4.21.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^5$

This suggests that estimated rank of  $\mathcal{H}$  i.e.,  $rank_{est}(\mathcal{H}) = 3$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

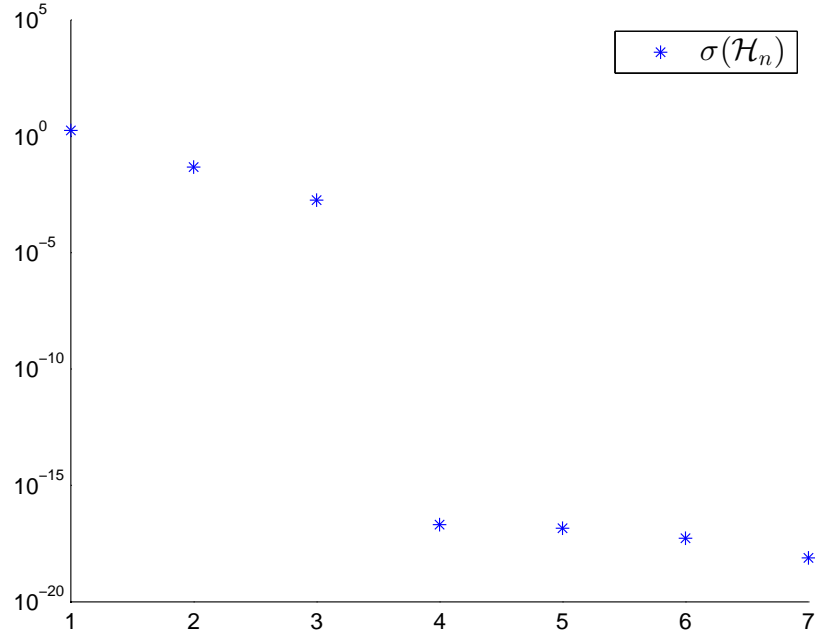
$$A_{new} = \begin{bmatrix} 0.396 & 0.322 & 0.281 \\ 0.02 & 0.363 & 0.617 \\ 0.495 & 0.282 & 0.223 \end{bmatrix}, \quad B_{new} = \begin{bmatrix} 0.084 & 0.916 \\ 0.723 & 0.277 \\ 0.455 & 0.545 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.428 & 0.572 & 0.198 & 0.23 & 0.23 & 0.341 \\ 0.428 & 0.198 & 0.23 & 0.089 & 0.108 & 0.092 & 0.139 \\ 0.572 & 0.23 & 0.341 & 0.107 & 0.124 & 0.138 & 0.204 \\ 0.198 & 0.089 & 0.108 & 0.040 & 0.049 & 0.043 & 0.065 \\ 0.23 & 0.107 & 0.124 & 0.048 & 0.058 & 0.049 & 0.074 \\ 0.23 & 0.092 & 0.139 & 0.042 & 0.049 & 0.056 & 0.083 \\ 0.341 & 0.138 & 0.204 & 0.064 & 0.074 & 0.082 & 0.122 \end{bmatrix}.$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.773 \\ 4.287 \times 10^{-2} \\ 1.898 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure 4.22.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^5$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 6.514 \times 10^{-3}$$

It is seen that the values in Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.

### Simulation 7.

1. We simulated the same process and produced a sequence of  $T = 10^4$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.429 & 0.571 & 0.194 & 0.234 & 0.234 & 0.337 \\ 0.429 & 0.194 & 0.234 & 0.0935 & 0.101 & 0.101 & 0.134 \\ 0.571 & 0.234 & 0.337 & 0.101 & 0.134 & 0.134 & 0.203 \\ 0.194 & 0.0935 & 0.101 & 0.0458 & 0.0477 & 0.0414 & 0.0595 \\ 0.234 & 0.101 & 0.134 & 0.0477 & 0.0532 & 0.0592 & 0.0743 \\ 0.234 & 0.101 & 0.134 & 0.0428 & 0.0578 & 0.0541 & 0.0797 \\ 0.337 & 0.134 & 0.203 & 0.0581 & 0.0757 & 0.0797 & 0.123 \end{bmatrix}$$

These values demonstrate that the approximate string probabilities differ from the original ones. The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.772 \\ 3.895 \times 10^{-2} \\ 8.254 \times 10^{-3} \\ 4.060 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.10)$$

As we can see from  $\sigma(\tilde{\mathcal{H}})$ , the gap between 3<sub>rd</sub> and 4<sub>th</sub> singular value is not significant. So, it cannot be assumed that the rank of  $\tilde{\mathcal{H}}$  is 3.



3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1.0 & 0.429 & 0.571 & 0.194 & 0.235 & 0.235 & 0.336 \\ 0.429 & 0.194 & 0.235 & 0.0911 & 0.102 & 0.1 & 0.135 \\ 0.571 & 0.235 & 0.336 & 0.102 & 0.133 & 0.135 & 0.201 \\ 0.194 & 0.0911 & 0.102 & 0.0439 & 0.0476 & 0.0436 & 0.0586 \\ 0.235 & 0.102 & 0.133 & 0.0473 & 0.0547 & 0.0569 & 0.0762 \\ 0.235 & 0.1 & 0.135 & 0.0448 & 0.0556 & 0.0549 & 0.0802 \\ 0.336 & 0.135 & 0.201 & 0.0576 & 0.0775 & 0.08 & 0.121 \end{bmatrix}$$

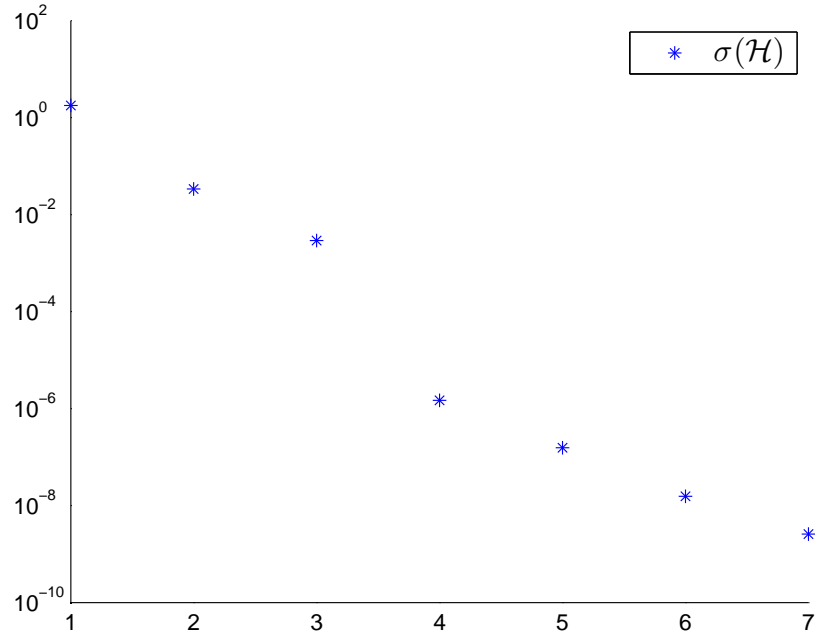
The  $\lambda$  value obtained as  $3.5 \times 10^{-5}$ .

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

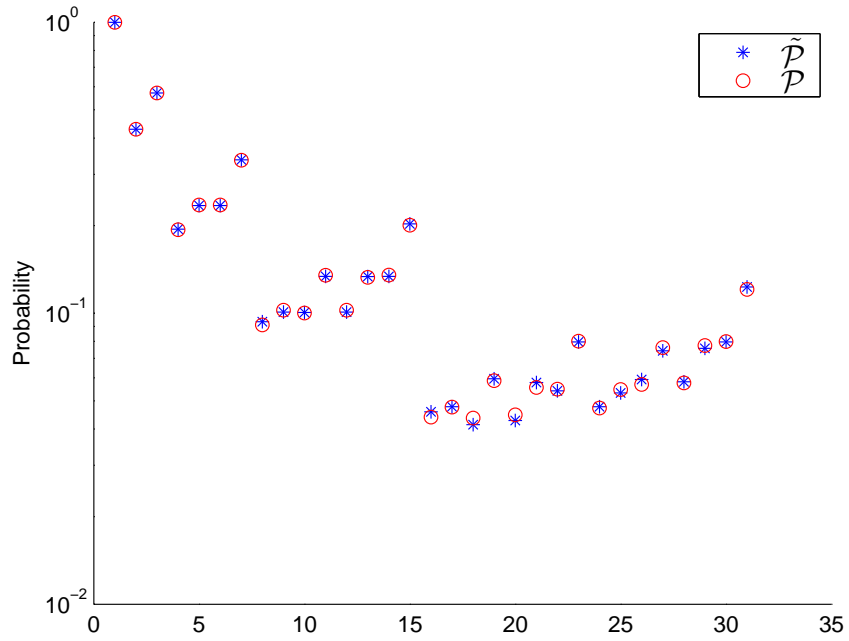
$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.772 \\ 3.383 \times 10^{-2} \\ 2.917 \times 10^{-3} \\ 1.475 \times 10^{-6} \\ 1.549 \times 10^{-7} \\ 0 \\ 0 \end{bmatrix}$$

The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}\|_F = 9.208 \times 10^{-3}$$



**Figure 4.23.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^4$



**Figure 4.24.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^4$

This suggests that estimated rank of  $\mathcal{H}$  i.e.,  $rank_{est}(\mathcal{H}) = 3$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

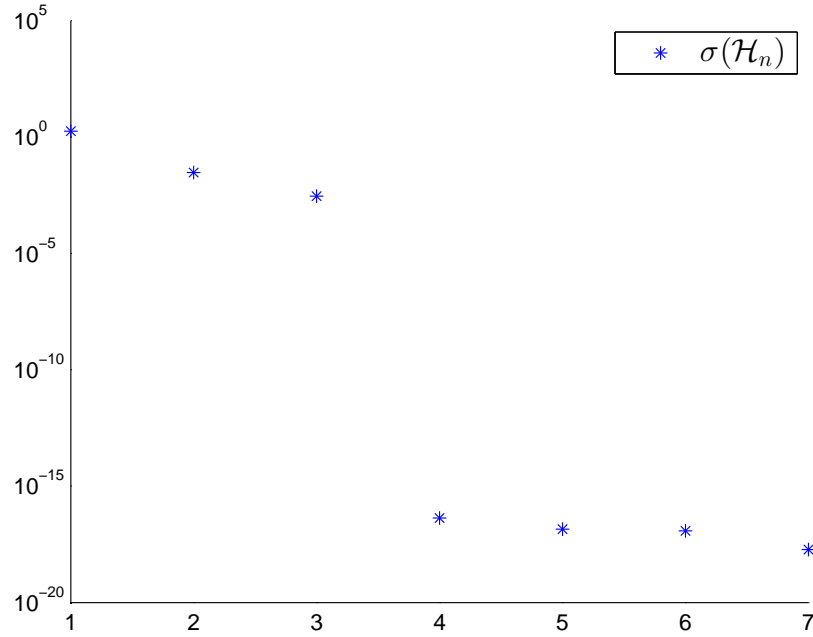
$$A_{new} = \begin{bmatrix} 0.41 & 0.407 & 0.184 \\ 0.537 & 0.173 & 0.289 \\ 0.04 & 0.451 & 0.509 \end{bmatrix}, \quad B_{new} = \begin{bmatrix} 0.532 & 0.468 \\ 0.668 & 0.332 \\ 0.084 & 0.916 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.429 & 0.571 & 0.194 & 0.235 & 0.235 & 0.336 \\ 0.429 & 0.194 & 0.235 & 0.0901 & 0.103 & 0.101 & 0.134 \\ 0.571 & 0.235 & 0.336 & 0.105 & 0.131 & 0.136 & 0.2 \\ 0.194 & 0.0901 & 0.103 & 0.0418 & 0.0484 & 0.0446 & 0.0589 \\ 0.235 & 0.105 & 0.131 & 0.0489 & 0.0559 & 0.056 & 0.0746 \\ 0.235 & 0.101 & 0.134 & 0.0453 & 0.0558 & 0.0549 & 0.0793 \\ 0.336 & 0.136 & 0.2 & 0.0602 & 0.0755 & 0.0803 & 0.12 \end{bmatrix}$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.771 \\ 3.258 \times 10^{-2} \\ 2.655 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure 4.25.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^4$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 2.531 \times 10^{-2}$$

It is seen that the values in Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.

### Simulation 8.

1. We simulated the same process and produced a sequence of  $T = 10^3$  observations  $\mathcal{O}_{1:T}$ .

2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.435 & 0.565 & 0.183 & 0.253 & 0.253 & 0.312 \\ 0.435 & 0.183 & 0.253 & 0.0802 & 0.102 & 0.119 & 0.133 \\ 0.565 & 0.253 & 0.312 & 0.102 & 0.15 & 0.134 & 0.178 \\ 0.183 & 0.0802 & 0.102 & 0.0341 & 0.0461 & 0.0451 & 0.0572 \\ 0.253 & 0.102 & 0.15 & 0.0461 & 0.0562 & 0.0742 & 0.0762 \\ 0.253 & 0.119 & 0.133 & 0.0542 & 0.0652 & 0.0592 & 0.0742 \\ 0.312 & 0.134 & 0.178 & 0.0481 & 0.0863 & 0.0752 & 0.102 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.767 \\ 3.039 \times 10^{-2} \\ 1.722 \times 10^{-2} \\ 1.882 \times 10^{-3} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (4.11)$$

As we can see from  $\sigma(\tilde{\mathcal{H}})$ , the gap between 3<sub>rd</sub> and 4<sub>th</sub> singular value is not significant. So, it cannot be assumed that the rank of  $\tilde{\mathcal{H}}$  is 3.

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.436 & 0.564 & 0.184 & 0.251 & 0.252 & 0.313 \\ 0.436 & 0.184 & 0.251 & 0.0804 & 0.104 & 0.116 & 0.136 \\ 0.564 & 0.252 & 0.313 & 0.104 & 0.147 & 0.136 & 0.177 \\ 0.184 & 0.0804 & 0.104 & 0.034 & 0.0464 & 0.0464 & 0.0577 \\ 0.252 & 0.104 & 0.147 & 0.0469 & 0.0573 & 0.0696 & 0.0779 \\ 0.251 & 0.116 & 0.136 & 0.051 & 0.065 & 0.0595 & 0.0764 \\ 0.313 & 0.136 & 0.177 & 0.0523 & 0.083 & 0.076 & 0.101 \end{bmatrix}$$

The  $\lambda$  value obtained as  $1.1 \times 10^{-4}$ .

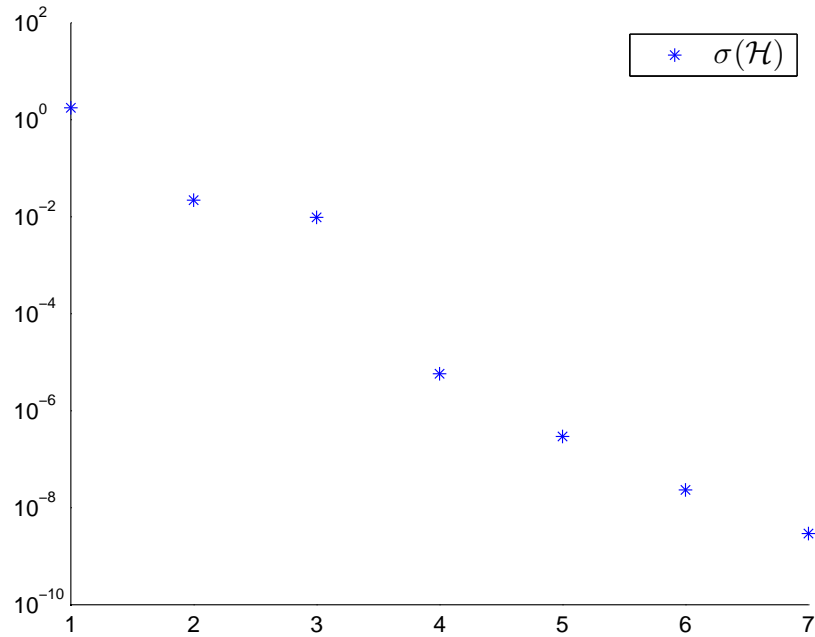
4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approxi-

mate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

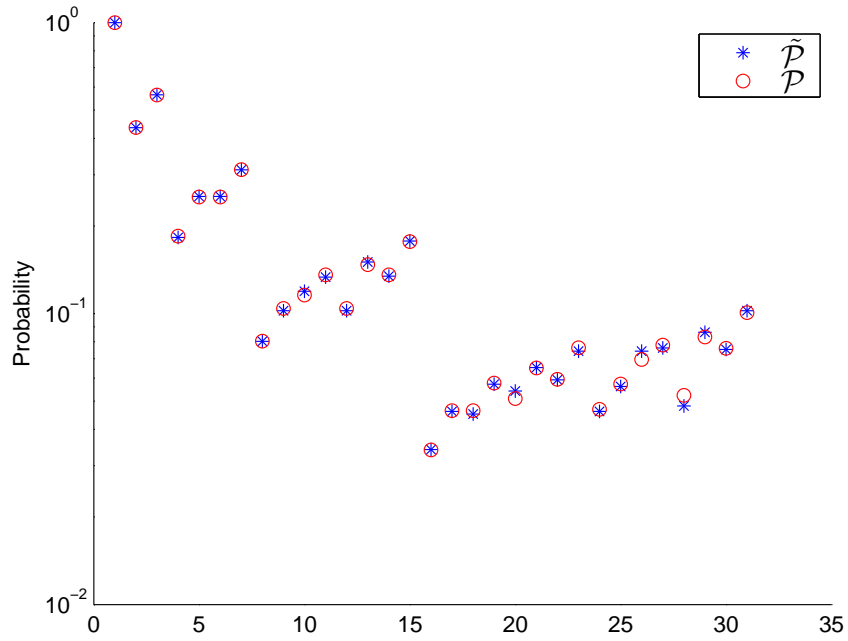
$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.766 \\ 2.196 \times 10^{-2} \\ 9.665 \times 10^{-3} \\ 5.778 \times 10^{-6} \\ 2.936 \times 10^{-7} \\ 0 \\ 0 \end{bmatrix}$$

The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  is

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 1.706 \times 10^{-2}$$



**Figure 4.26.** Singular values of  $\mathcal{H}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^3$



**Figure 4.27.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^3$

This suggests that estimated rank of  $\mathcal{H}$  i.e.,  $rank_{est}(\mathcal{H}) = 3$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

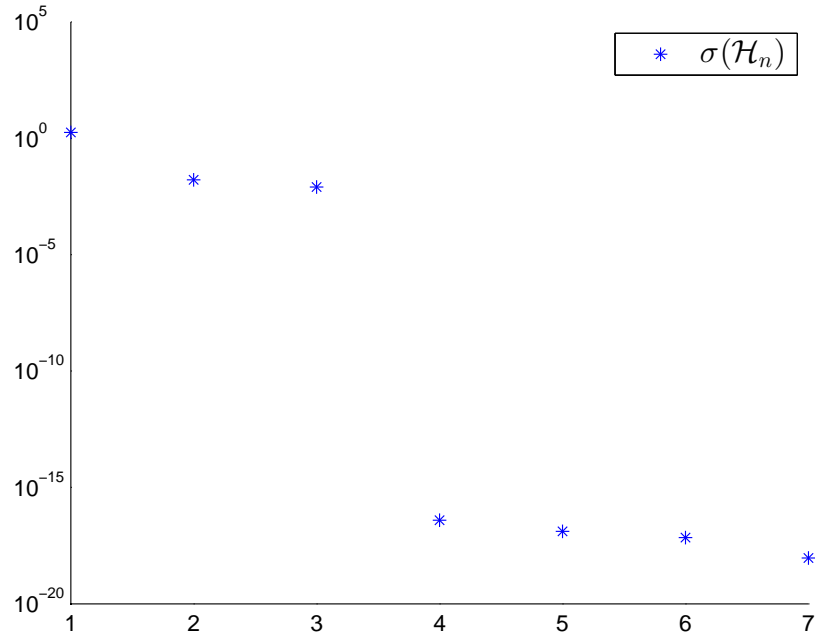
$$A_{new} = \begin{bmatrix} 0.423 & 0.398 & 0.179 \\ 0.295 & 0.234 & 0.471 \\ 0.308 & 0.525 & 0.167 \end{bmatrix}, \quad B_{new} = \begin{bmatrix} 0.188 & 0.812 \\ 0.7 & 0.3 \\ 0.352 & 0.648 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.436 & 0.564 & 0.184 & 0.251 & 0.252 & 0.313 \\ 0.436 & 0.184 & 0.251 & 0.077 & 0.107 & 0.11 & 0.141 \\ 0.564 & 0.252 & 0.313 & 0.103 & 0.149 & 0.134 & 0.178 \\ 0.184 & 0.077 & 0.107 & 0.032 & 0.045 & 0.047 & 0.060 \\ 0.252 & 0.103 & 0.149 & 0.043 & 0.059 & 0.066 & 0.083 \\ 0.251 & 0.11 & 0.141 & 0.045 & 0.065 & 0.061 & 0.080 \\ 0.313 & 0.134 & 0.178 & 0.055 & 0.079 & 0.077 & 0.101 \end{bmatrix}.$$

and the singular values of  $\mathcal{H}_n$  is given by

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.767 \\ 1.967 \times 10^{-2} \\ 8.475 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure 4.28.** Singular values of  $\mathcal{H}_n$  for  $t = 2$ ,  $N = 3$ ,  $L_{seq} = 10^3$

The distance between  $\tilde{\mathcal{H}}$  and  $\mathcal{H}_n$  in Frobenius norm is given by

$$\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F = 3.514 \times 10^{-2}$$

It is seen that the values in Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\tilde{\mathcal{H}}$  are really close.



### 4.1.3 Result Discussion

The first note about the simulation results is change of  $\lambda$  value as a function of sequence length  $L_{seq}$ . Tables 4.1 and 4.2 show the relation between  $\lambda$  and  $L_{seq}$ .

$L_{seq}$	$\lambda$	$L_{seq}$	$\lambda$
$10^6$	$4 \times 10^{-7}$	$10^6$	$2.1 \times 10^{-7}$
$10^5$	$6 \times 10^{-6}$	$10^5$	$1.6 \times 10^{-6}$
$10^4$	$8.1 \times 10^{-5}$	$10^4$	$3.5 \times 10^{-5}$
$10^3$	$5 \times 10^{-4}$	$10^3$	$1.1 \times 10^{-4}$

**Table 4.1.** Change of  $\lambda$  value as a function of  $L_{seq}$  for Case 1

**Table 4.2.** Change of  $\lambda$  value as a function of  $L_{seq}$  for Case 2

These results demonstrate that  $\lambda$  value changes depending on the size of the sample sequence  $L_{seq}$ . Tables 4.1 and 4.2 show that as the length of the sample sequence increases the  $\lambda$  value is decreases. This decrease is reasonable, because, as the length of the sample sequence goes to infinity the string probabilities converge to true probabilities, in turn, the  $\lambda$  value converges to zero and the rank of the Hankel matrix gives directly the order of the HMM.

Note that the priority of the algorithm is estimating the true order of the HMM. As we stated in Section 3.1, the estimated rank of the matrix (order of the model) is decided by observing the significant gap between  $N_{th}$  and  $(N + 1)_{th}$  largest singular value. In table 4.3 and 4.4 we provide the  $\sigma(\mathcal{H}_r)$  to check the accuracy of the algorithm. Table 4.3 and 4.4 show that we cannot have an accurate estimate of the order from the singular values of  $\tilde{\mathcal{H}}$ . Since the input data is finite, it produces an outcome of HMM that has high order and we do not observe a significant gap between  $N_{th}$  and  $(N + 1)_{th}$  largest singular value. However, this gap is obtained in optimized Hankel matrix  $\mathcal{H}$  and subsequently in reconstructed Hankel matrix  $\mathcal{H}_n$ . Although the algorithm's accuracy for estimating the singular values decreases proportional to  $L_{seq}$ , the gap condition is fulfilled and the correct order of the HMM is obtained for all data length.

$L_{seq}$	$\sigma(\mathcal{H}_r)$	$\sigma(\tilde{\mathcal{H}})$	$\sigma(\mathcal{H})$	$\sigma(\mathcal{H}_n)$
$10^6$		1.776	1.776	1.776
		$3.234 \times 10^{-2}$	$3.180 \times 10^{-2}$	$3.165 \times 10^{-2}$
		$5.118 \times 10^{-4}$	$1.522 \times 10^{-6}$	0
		$3.780 \times 10^{-4}$	$3.754 \times 10^{-7}$	0
$10^5$	1.776 $3.222 \times 10^{-2}$ 0	1.777	1.777	1.776
		$3.089 \times 10^{-2}$	$2.855 \times 10^{-2}$	$2.846 \times 10^{-2}$
		$2.099 \times 10^{-3}$	$1.417 \times 10^{-7}$	0
		$9.786 \times 10^{-4}$	0	0
$10^4$	0	1.775	1.774	1.774
		$2.701 \times 10^{-2}$	$2.632 \times 10^{-2}$	$2.621 \times 10^{-2}$
		$9.146 \times 10^{-3}$	$8.837 \times 10^{-7}$	0
		$4.409 \times 10^{-4}$	0	0
$10^3$		1.771	1.770	1.770
		$6.051 \times 10^{-2}$	$4.513 \times 10^{-2}$	$4.375 \times 10^{-2}$
		$1.643 \times 10^{-2}$	$1.671 \times 10^{-7}$	0
		$5.256 \times 10^{-3}$	0	0

**Table 4.3.**  $\sigma(\tilde{\mathcal{H}})$ ,  $\sigma(\mathcal{H})$ , and  $\sigma(\mathcal{H}_n)$  values as a function of  $L_{seq}$  for Case 1

$L_{seq}$	$\sigma(\mathcal{H}_r)$	$\sigma(\tilde{\mathcal{H}})$	$\sigma(\mathcal{H})$	$\sigma(\mathcal{H}_n)$
$10^6$		1.773	1.773	1.773
		$4.786 \times 10^{-2}$	$4.761 \times 10^{-2}$	$4.775 \times 10^{-2}$
		$2.331 \times 10^{-4}$	$2.055 \times 10^{-3}$	$2.177 \times 10^{-3}$
		$1.505 \times 10^{-4}$	$1.631 \times 10^{-6}$	0
$10^5$	1.773 $4.769 \times 10^{-2}$ $2.325 \times 10^{-3}$	1.773	1.773	1.773
		$4.708 \times 10^{-2}$	$4.638 \times 10^{-2}$	$4.287 \times 10^{-2}$
		$2.709 \times 10^{-3}$	$1.916 \times 10^{-3}$	$1.898 \times 10^{-3}$
		$3.527 \times 10^{-4}$	$1.302 \times 10^{-6}$	0
$10^4$	0 0	1.772	1.772	1.771
		$3.895 \times 10^{-2}$	$3.383 \times 10^{-2}$	$3.258 \times 10^{-2}$
		$8.254 \times 10^{-3}$	$2.917 \times 10^{-3}$	$2.655 \times 10^{-3}$
		$4.060 \times 10^{-4}$	$1.475 \times 10^{-6}$	0
$10^3$		1.767	1.766	1.767
		$3.039 \times 10^{-2}$	$2.196 \times 10^{-2}$	$1.967 \times 10^{-2}$
		$1.722 \times 10^{-2}$	$9.665 \times 10^{-3}$	$8.475 \times 10^{-3}$
		$1.882 \times 10^{-3}$	$5.778 \times 10^{-6}$	0
		0	$2.936 \times 10^{-7}$	0

**Table 4.4.**  $\sigma(\tilde{\mathcal{H}})$ ,  $\sigma(\mathcal{H})$ , and  $\sigma(\mathcal{H}_n)$  values as a function of  $L_{seq}$  for Case 2

In Tables 4.5 and 4.6, we give the true probabilities and compare them with reconstructed string probabilities of strings of length up to 3 for  $L_{seq} = 10^6, 10^5, 10^4, 10^3$  in order to show the performance of the algorithm.

Sequence	Exact	$L_{seq} = 10^6$	$L_{seq} = 10^5$	$L_{seq} = 10^4$	$L_{seq} = 10^3$
0	0.420	0.419	0.418	0.421	0.429
1	0.580	0.581	0.582	0.579	0.571
00	0.166	0.166	0.166	0.168	0.170
01	0.254	0.254	0.252	0.253	0.260
10	0.254	0.254	0.252	0.253	0.259
11	0.326	0.326	0.330	0.326	0.312
000	0.068	0.067	0.066	0.068	0.069
001	0.099	0.099	0.099	0.100	0.101
010	0.113	0.113	0.110	0.112	0.120
011	0.141	0.141	0.142	0.141	0.140
100	0.099	0.099	0.099	0.100	0.100
101	0.155	0.155	0.153	0.153	0.158
110	0.141	0.141	0.142	0.141	0.139
111	0.185	0.186	0.188	0.185	0.173

**Table 4.5.** String probabilities of strings of length up to 3 for Case 1

Sequence	Exact	$L_{seq} = 10^6$	$L_{seq} = 10^5$	$L_{seq} = 10^4$	$L_{seq} = 10^3$
0	0.428	0.428	0.428	0.429	0.436
1	0.572	0.572	0.572	0.571	0.564
00	0.198	0.198	0.198	0.194	0.184
01	0.230	0.230	0.230	0.235	0.251
10	0.230	0.230	0.230	0.235	0.252
11	0.342	0.342	0.342	0.336	0.313
000	0.094	0.094	0.089	0.090	0.077
001	0.103	0.103	0.108	0.103	0.107
010	0.096	0.094	0.092	0.101	0.110
011	0.134	0.136	0.139	0.134	0.141
100	0.103	0.105	0.107	0.105	0.103
101	0.127	0.125	0.124	0.131	0.149
110	0.134	0.134	0.138	0.136	0.134
111	0.208	0.208	0.204	0.200	0.178

**Table 4.6.** String probabilities of strings of length up to 3 for Case 2

It is evident that as the length of the sequence  $L_{seq}$  decreases, the noise level of the data increases. As a result, the divergence of the approximate string probabilities  $\tilde{\mathcal{P}}$  from true probabilities  $\mathcal{P}_r$  increases. Since approximate string probabilities  $\tilde{\mathcal{P}}$  are the only input to the algorithm, the difference between reconstructed probabilities  $\mathcal{P}_n$  and true probabilities  $\mathcal{P}_r$  also increases as the  $L_{seq}$  decreases. Tables 4.5 and 4.6 show that for  $L_{seq} = 10^6$ ,  $\mathcal{P}_n$  values are almost the same and for smaller sequences the statistics of the string probabilities are very close to true probabilities. Thus, we can conclude that the algorithm works well.

We stated in Chapter 3 that the algorithm consists of two steps. In the first step, the input was  $\tilde{\mathcal{H}}$  and the output was  $\mathcal{H}$ . In the second step,  $\mathcal{H}$  was used as an input and  $\mathcal{H}_n$  was obtained as an output. We give Tables 4.7 and 4.8 in order to illustrate the divergence of reconstructed probabilities  $\mathcal{P}_n$  and optimized probabilities  $\mathcal{P}$  from approximate probabilities  $\tilde{\mathcal{P}}$  as a function of  $L_{seq}$ .

$L_{seq}$	$\ \tilde{\mathcal{H}} - \mathcal{H}\ _F$	$\ \tilde{\mathcal{H}} - \mathcal{H}_n\ _F$
$10^6$	$1.054 \times 10^{-4}$	$2.544 \times 10^{-3}$
$10^5$	$4.176 \times 10^{-3}$	$5.297 \times 10^{-3}$
$10^4$	$1.568 \times 10^{-3}$	$1.803 \times 10^{-2}$
$10^3$	$2.989 \times 10^{-2}$	$2.773 \times 10^{-2}$

**Table 4.7.** Change of  $\|\tilde{\mathcal{H}} - \mathcal{H}\|_F$  and  $\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F$  as a function of  $L_{seq}$  for Case 1

$L_{seq}$	$\ \tilde{\mathcal{H}} - \mathcal{H}\ _F$	$\ \tilde{\mathcal{H}} - \mathcal{H}_n\ _F$
$10^6$	$5.895 \times 10^{-4}$	$5.022 \times 10^{-3}$
$10^5$	$1.742 \times 10^{-3}$	$6.514 \times 10^{-3}$
$10^4$	$9.208 \times 10^{-3}$	$2.531 \times 10^{-2}$
$10^3$	$1.706 \times 10^{-2}$	$3.514 \times 10^{-2}$

**Table 4.8.** Change of  $\|\tilde{\mathcal{H}} - \mathcal{H}\|_F$  and  $\|\tilde{\mathcal{H}} - \mathcal{H}_n\|_F$  as a function of  $L_{seq}$  for Case 2

Again, Tables 4.7 and 4.8 show that the divergence increases as the sequence length decreases.

## Conclusions

In this thesis, we presented an algorithm to find an HMM that generates the string probabilities that are given to the algorithm. The only input to the algorithm is finite output sequence  $\mathcal{O}_{1:T}$ . Due to finite input data, the Hankel matrix  $\tilde{\mathcal{H}}$  has high rank, in turn, the HMM constructed from  $\tilde{\mathcal{H}}$  has high order. In the first step of algorithm, the true order of the HMM was obtained by minimizing the nuclear norm of the Hankel matrix constructed from an HMM  $\tilde{\mathcal{H}}$  instead of performing rank minimization.

Subsequently in the second step of the algorithm, given the rank information, an HMM that generates the given string probabilities was reconstructed using non-negative matrix factorization technique. It is worth to mention that the priority of the algorithm was to obtain the true order of the model and keep the divergence minimum for given order between approximate  $\tilde{\mathcal{P}}$  and reconstructed probabilities  $\mathcal{P}_n$ . It was demonstrated that the algorithm efficiently estimates the order of the HMM (number of states of HMM). Moreover, given the order information, the HMM parameters (state transition and output matrix) are estimated by non-negative matrix factorization technique. It was shown that these estimates differ from the original parameters. However, it was also illustrated that the string probabilities generated from this model was close to the approximate string probabilities which were used as an input to the algorithm.

## Simulation Results for Case 3

### A.1 Case 3: $t = 3, N = 2$

In these simulations, we check the performance of the algorithm for larger size of Hankel matrix. The difference of Case 3 and 4 from the first two cases is that we now compute the string probabilities of strings of length  $t = 3$  instead of  $t = 2$ . Given the state transition matrix  $A$ , the output matrix  $B$  and (approximate) string probabilities of strings of length  $t = 3$ , we obtain the  $\lambda$  value that makes the order of the model is closest the real order of the system for different sequence lengths  $L_{seq}$ . Here,  $A$  and  $B$  matrices are again arbitrarily chosen.

Consider the process described by  $(A, B)$ ;

$$A = \begin{bmatrix} 0.389 & 0.611 \\ 0.544 & 0.456 \end{bmatrix}, \quad B = \begin{bmatrix} 0.332 & 0.668 \\ 0.855 & 0.145 \end{bmatrix}. \quad (\text{A.1})$$

The Hankel matrix consist of real string probabilities  $\mathcal{H}_r$  corresponding to  $(A, B)$  is given below. Then, since the rank of the Hankel matrix consist of real string probabilities is equal to the order of the model, the singular values of  $\mathcal{H}_r$  i.e.  $\sigma(\mathcal{H}_r)$

is obtained as follows;

$$\sigma(\mathcal{H}_r) = \begin{bmatrix} 1.94 \\ 3.763 \times 10^{-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and the  $\mathcal{H}_r$  is;

$$\mathcal{H}_r = \begin{bmatrix} 1.0 & 0.609 & 0.391 & 0.36 & 0.249 & 0.249 & 0.142 & 0.214 & 0.146 & 0.159 & 0.09 & 0.146 & 0.102 & 0.09 & 0.052 \\ 0.609 & 0.36 & 0.249 & 0.214 & 0.146 & 0.159 & 0.09 & 0.127 & 0.087 & 0.093 & 0.053 & 0.093 & 0.065 & 0.057 & 0.033 \\ 0.391 & 0.249 & 0.142 & 0.146 & 0.102 & 0.09 & 0.052 & 0.087 & 0.059 & 0.065 & 0.037 & 0.053 & 0.037 & 0.033 & 0.019 \\ 0.36 & 0.214 & 0.146 & 0.127 & 0.087 & 0.093 & 0.053 & 0.075 & 0.052 & 0.055 & 0.032 & 0.055 & 0.038 & 0.034 & 0.019 \\ 0.249 & 0.146 & 0.102 & 0.087 & 0.059 & 0.065 & 0.037 & 0.052 & 0.035 & 0.038 & 0.022 & 0.038 & 0.027 & 0.023 & 0.014 \\ 0.249 & 0.159 & 0.09 & 0.093 & 0.065 & 0.057 & 0.033 & 0.055 & 0.038 & 0.042 & 0.024 & 0.034 & 0.023 & 0.021 & 0.012 \\ 0.142 & 0.09 & 0.052 & 0.053 & 0.037 & 0.033 & 0.019 & 0.032 & 0.022 & 0.024 & 0.013 & 0.019 & 0.014 & 0.012 & 0.007 \\ 0.214 & 0.127 & 0.087 & 0.075 & 0.052 & 0.055 & 0.032 & 0.045 & 0.031 & 0.033 & 0.019 & 0.033 & 0.023 & 0.02 & 0.012 \\ 0.146 & 0.087 & 0.059 & 0.052 & 0.035 & 0.038 & 0.022 & 0.031 & 0.021 & 0.023 & 0.013 & 0.022 & 0.016 & 0.014 & 0.008 \\ 0.159 & 0.093 & 0.065 & 0.055 & 0.038 & 0.042 & 0.024 & 0.033 & 0.023 & 0.024 & 0.014 & 0.025 & 0.017 & 0.015 & 0.009 \\ 0.09 & 0.053 & 0.037 & 0.032 & 0.022 & 0.024 & 0.013 & 0.019 & 0.013 & 0.014 & 0.008 & 0.014 & 0.01 & 0.008 & 0.005 \\ 0.146 & 0.093 & 0.053 & 0.055 & 0.038 & 0.034 & 0.019 & 0.033 & 0.022 & 0.025 & 0.014 & 0.02 & 0.014 & 0.012 & 0.007 \\ 0.102 & 0.065 & 0.037 & 0.038 & 0.027 & 0.023 & 0.014 & 0.023 & 0.016 & 0.017 & 0.01 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.09 & 0.057 & 0.033 & 0.034 & 0.023 & 0.021 & 0.012 & 0.02 & 0.014 & 0.015 & 0.008 & 0.012 & 0.009 & 0.008 & 0.004 \\ 0.052 & 0.033 & 0.019 & 0.019 & 0.014 & 0.012 & 0.007 & 0.012 & 0.008 & 0.009 & 0.005 & 0.007 & 0.005 & 0.004 & 0.003 \end{bmatrix}$$

It is clear from the construction of  $\mathcal{H}_r$  that the probabilities in  $\mathcal{H}_r$  will be the same for all sequence lengths  $L_{seq}$ . In the simulation example, this model and the Hankel matrix corresponding to real probabilities  $\mathcal{H}_r$  are unknown; however, we give them here to evaluate the performance of the algorithms.

### Simulation 9.

1. We simulated this process and produced a sequence of  $T = 10^6$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$



$$\tilde{\mathcal{H}} = \begin{bmatrix} 1.0 & 0.609 & 0.391 & 0.36 & 0.249 & 0.249 & 0.142 & 0.213 & 0.146 & 0.159 & 0.09 & 0.146 & 0.103 & 0.09 & 0.052 \\ 0.609 & 0.36 & 0.249 & 0.213 & 0.146 & 0.159 & 0.09 & 0.127 & 0.087 & 0.093 & 0.053 & 0.093 & 0.066 & 0.057 & 0.033 \\ 0.391 & 0.249 & 0.142 & 0.146 & 0.103 & 0.09 & 0.052 & 0.087 & 0.06 & 0.066 & 0.037 & 0.053 & 0.037 & 0.033 & 0.019 \\ 0.36 & 0.213 & 0.146 & 0.127 & 0.087 & 0.093 & 0.053 & 0.075 & 0.051 & 0.055 & 0.031 & 0.055 & 0.039 & 0.034 & 0.019 \\ 0.249 & 0.146 & 0.103 & 0.087 & 0.06 & 0.066 & 0.037 & 0.051 & 0.035 & 0.038 & 0.022 & 0.039 & 0.027 & 0.023 & 0.014 \\ 0.249 & 0.159 & 0.09 & 0.093 & 0.066 & 0.057 & 0.033 & 0.055 & 0.038 & 0.042 & 0.024 & 0.034 & 0.024 & 0.021 & 0.012 \\ 0.142 & 0.09 & 0.052 & 0.053 & 0.037 & 0.033 & 0.019 & 0.031 & 0.022 & 0.024 & 0.013 & 0.019 & 0.014 & 0.012 & 0.007 \\ 0.213 & 0.127 & 0.087 & 0.075 & 0.051 & 0.055 & 0.031 & 0.045 & 0.031 & 0.033 & 0.019 & 0.032 & 0.023 & 0.02 & 0.011 \\ 0.146 & 0.087 & 0.06 & 0.051 & 0.035 & 0.038 & 0.022 & 0.031 & 0.021 & 0.022 & 0.013 & 0.022 & 0.016 & 0.014 & 0.008 \\ 0.159 & 0.093 & 0.066 & 0.055 & 0.038 & 0.042 & 0.024 & 0.033 & 0.022 & 0.024 & 0.014 & 0.025 & 0.017 & 0.015 & 0.009 \\ 0.09 & 0.053 & 0.037 & 0.031 & 0.022 & 0.024 & 0.013 & 0.019 & 0.013 & 0.014 & 0.008 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.146 & 0.093 & 0.053 & 0.055 & 0.039 & 0.034 & 0.019 & 0.032 & 0.022 & 0.025 & 0.014 & 0.02 & 0.014 & 0.012 & 0.007 \\ 0.103 & 0.066 & 0.037 & 0.039 & 0.027 & 0.023 & 0.014 & 0.023 & 0.016 & 0.017 & 0.01 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.09 & 0.057 & 0.033 & 0.034 & 0.024 & 0.021 & 0.012 & 0.02 & 0.014 & 0.015 & 0.009 & 0.012 & 0.009 & 0.008 & 0.004 \\ 0.052 & 0.033 & 0.019 & 0.019 & 0.014 & 0.012 & 0.007 & 0.011 & 0.008 & 0.009 & 0.005 & 0.007 & 0.005 & 0.004 & 0.003 \end{bmatrix}$$

These values demonstrate that the approximate string probabilities differ from the original ones. The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.939 \\ 3.890 \times 10^{-2} \\ 7.729 \times 10^{-4} \\ 5.729 \times 10^{-4} \\ 3.255 \times 10^{-4} \\ 1.750 \times 10^{-4} \\ 1.514 \times 10^{-4} \\ 7.216 \times 10^{-5} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{A.2})$$

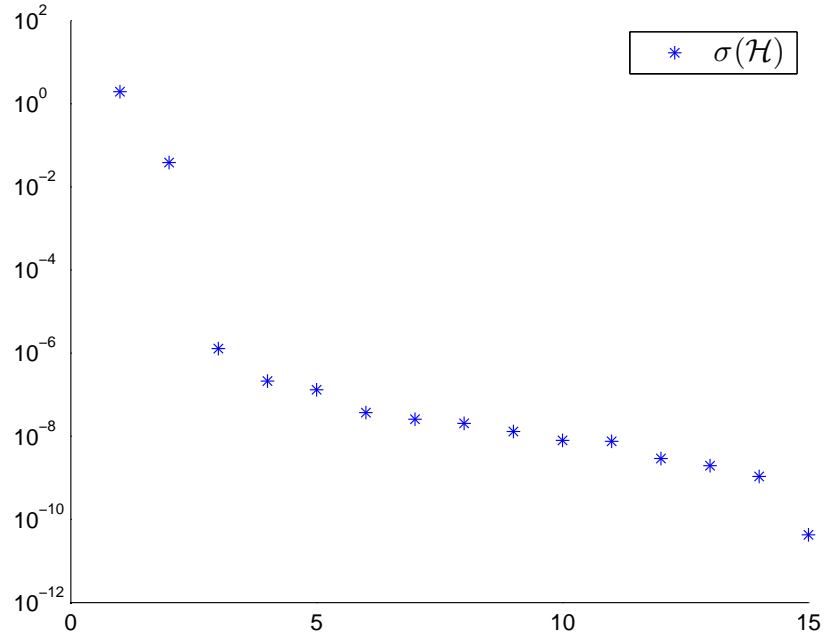
As it is seen from A.2, it cannot be assumed that the rank of  $\tilde{\mathcal{H}}$  is 2 since the difference between  $2_{nd}$  and  $3_{rd}$  singular value is not significant. Then we run the algorithm.

3. Solve the optimization problem 3.1.

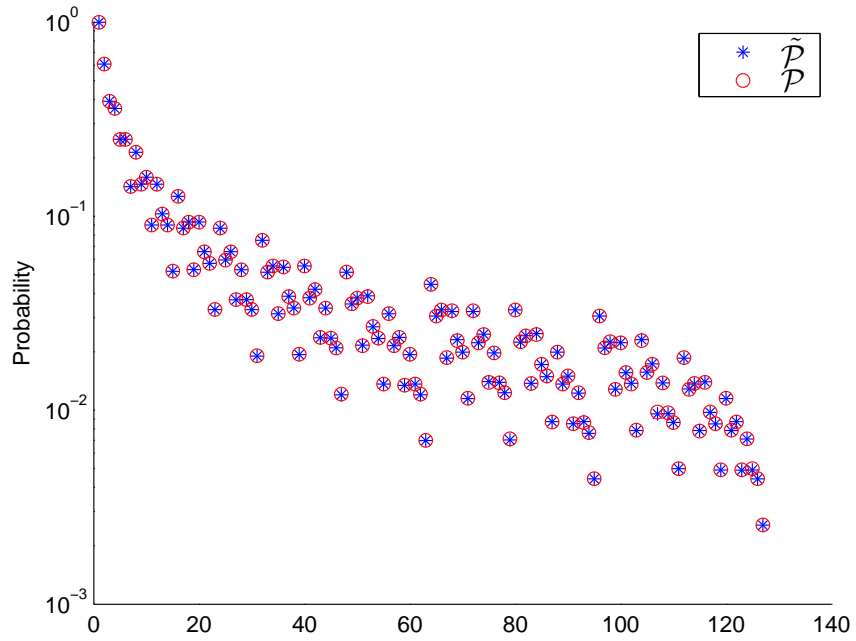
$$\mathcal{H} = \begin{bmatrix} 1.0 & 0.609 & 0.391 & 0.36 & 0.249 & 0.249 & 0.142 & 0.213 & 0.146 & 0.159 & 0.09 & 0.146 & 0.103 & 0.09 & 0.052 \\ 0.609 & 0.36 & 0.249 & 0.213 & 0.146 & 0.159 & 0.09 & 0.127 & 0.087 & 0.093 & 0.053 & 0.093 & 0.066 & 0.057 & 0.033 \\ 0.391 & 0.249 & 0.142 & 0.146 & 0.103 & 0.09 & 0.052 & 0.087 & 0.059 & 0.066 & 0.037 & 0.053 & 0.037 & 0.033 & 0.019 \\ 0.36 & 0.213 & 0.146 & 0.127 & 0.087 & 0.093 & 0.053 & 0.075 & 0.052 & 0.055 & 0.031 & 0.055 & 0.038 & 0.034 & 0.019 \\ 0.249 & 0.146 & 0.103 & 0.087 & 0.059 & 0.066 & 0.037 & 0.052 & 0.035 & 0.038 & 0.022 & 0.038 & 0.027 & 0.024 & 0.014 \\ 0.249 & 0.159 & 0.09 & 0.093 & 0.066 & 0.057 & 0.033 & 0.055 & 0.038 & 0.042 & 0.024 & 0.034 & 0.024 & 0.021 & 0.012 \\ 0.142 & 0.09 & 0.052 & 0.053 & 0.037 & 0.033 & 0.019 & 0.031 & 0.022 & 0.024 & 0.013 & 0.019 & 0.014 & 0.012 & 0.007 \\ 0.213 & 0.127 & 0.087 & 0.075 & 0.052 & 0.055 & 0.031 & 0.045 & 0.031 & 0.033 & 0.019 & 0.033 & 0.023 & 0.02 & 0.012 \\ 0.146 & 0.087 & 0.059 & 0.052 & 0.035 & 0.038 & 0.022 & 0.031 & 0.021 & 0.023 & 0.013 & 0.022 & 0.016 & 0.014 & 0.008 \\ 0.159 & 0.093 & 0.066 & 0.055 & 0.038 & 0.042 & 0.024 & 0.033 & 0.023 & 0.024 & 0.014 & 0.025 & 0.017 & 0.015 & 0.009 \\ 0.09 & 0.053 & 0.037 & 0.031 & 0.022 & 0.024 & 0.013 & 0.019 & 0.013 & 0.014 & 0.008 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.146 & 0.093 & 0.053 & 0.055 & 0.038 & 0.034 & 0.019 & 0.033 & 0.022 & 0.025 & 0.014 & 0.02 & 0.014 & 0.012 & 0.007 \\ 0.103 & 0.066 & 0.037 & 0.038 & 0.027 & 0.024 & 0.014 & 0.023 & 0.016 & 0.017 & 0.01 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.09 & 0.057 & 0.033 & 0.034 & 0.024 & 0.021 & 0.012 & 0.02 & 0.014 & 0.015 & 0.009 & 0.012 & 0.009 & 0.008 & 0.004 \\ 0.052 & 0.033 & 0.019 & 0.019 & 0.014 & 0.012 & 0.007 & 0.012 & 0.008 & 0.009 & 0.005 & 0.007 & 0.005 & 0.004 & 0.003 \end{bmatrix}$$

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.9393.831 \times 10^{-2} \\ 1.281 \times 10^{-6} \\ 2.128 \times 10^{-7} \\ 1.318 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure A.1.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^6$



**Figure A.2.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^6$

This suggests that  $rank_{est}(\mathcal{H}) = 2$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

$$A_n = \begin{bmatrix} 0.432 & 0.568 \\ 0.585 & 0.415 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.871 & 0.129 \\ 0.339 & 0.661 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1.0 & 0.609 & 0.391 & 0.36 & 0.249 & 0.249 & 0.142 & 0.213 & 0.146 & 0.159 & 0.09 & 0.146 & 0.103 & 0.09 & 0.052 \\ 0.609 & 0.36 & 0.249 & 0.213 & 0.146 & 0.159 & 0.09 & 0.127 & 0.087 & 0.093 & 0.053 & 0.093 & 0.065 & 0.057 & 0.033 \\ 0.391 & 0.249 & 0.142 & 0.146 & 0.103 & 0.09 & 0.052 & 0.087 & 0.059 & 0.065 & 0.037 & 0.053 & 0.037 & 0.033 & 0.019 \\ 0.36 & 0.213 & 0.146 & 0.127 & 0.087 & 0.093 & 0.053 & 0.075 & 0.051 & 0.055 & 0.031 & 0.055 & 0.038 & 0.034 & 0.019 \\ 0.249 & 0.146 & 0.103 & 0.087 & 0.059 & 0.065 & 0.037 & 0.051 & 0.035 & 0.038 & 0.022 & 0.038 & 0.027 & 0.024 & 0.014 \\ 0.249 & 0.159 & 0.09 & 0.093 & 0.065 & 0.057 & 0.033 & 0.055 & 0.038 & 0.042 & 0.024 & 0.034 & 0.024 & 0.021 & 0.012 \\ 0.142 & 0.09 & 0.052 & 0.053 & 0.037 & 0.033 & 0.019 & 0.031 & 0.022 & 0.024 & 0.013 & 0.019 & 0.014 & 0.012 & 0.007 \\ 0.213 & 0.127 & 0.087 & 0.075 & 0.051 & 0.055 & 0.031 & 0.044 & 0.031 & 0.033 & 0.019 & 0.033 & 0.023 & 0.02 & 0.012 \\ 0.146 & 0.087 & 0.059 & 0.051 & 0.035 & 0.038 & 0.022 & 0.031 & 0.021 & 0.023 & 0.013 & 0.022 & 0.016 & 0.014 & 0.008 \\ 0.159 & 0.093 & 0.065 & 0.055 & 0.038 & 0.042 & 0.024 & 0.033 & 0.023 & 0.024 & 0.014 & 0.025 & 0.017 & 0.015 & 0.009 \\ 0.09 & 0.053 & 0.037 & 0.031 & 0.022 & 0.024 & 0.013 & 0.019 & 0.013 & 0.014 & 0.008 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.146 & 0.093 & 0.053 & 0.055 & 0.038 & 0.034 & 0.019 & 0.033 & 0.022 & 0.025 & 0.014 & 0.02 & 0.014 & 0.012 & 0.007 \\ 0.103 & 0.065 & 0.037 & 0.038 & 0.027 & 0.024 & 0.014 & 0.023 & 0.016 & 0.017 & 0.01 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.09 & 0.057 & 0.033 & 0.034 & 0.024 & 0.021 & 0.012 & 0.02 & 0.014 & 0.015 & 0.009 & 0.012 & 0.009 & 0.008 & 0.004 \\ 0.052 & 0.033 & 0.019 & 0.019 & 0.014 & 0.012 & 0.007 & 0.012 & 0.008 & 0.009 & 0.005 & 0.007 & 0.005 & 0.004 & 0.003 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

## Simulation 10.

1. We simulated the same process and produced a sequence of  $T = 10^5$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1.0 & 0.607 & 0.393 & 0.358 & 0.249 & 0.249 & 0.144 & 0.213 & 0.145 & 0.159 & 0.091 & 0.145 & 0.104 & 0.091 & 0.053 \\ 0.607 & 0.358 & 0.249 & 0.213 & 0.145 & 0.159 & 0.091 & 0.126 & 0.087 & 0.092 & 0.053 & 0.092 & 0.066 & 0.057 & 0.033 \\ 0.393 & 0.249 & 0.144 & 0.145 & 0.104 & 0.091 & 0.053 & 0.087 & 0.058 & 0.066 & 0.038 & 0.053 & 0.038 & 0.033 & 0.02 \\ 0.358 & 0.213 & 0.145 & 0.126 & 0.087 & 0.092 & 0.053 & 0.075 & 0.051 & 0.056 & 0.031 & 0.053 & 0.039 & 0.033 & 0.019 \\ 0.249 & 0.145 & 0.104 & 0.087 & 0.058 & 0.066 & 0.038 & 0.051 & 0.036 & 0.036 & 0.022 & 0.039 & 0.027 & 0.024 & 0.014 \\ 0.249 & 0.159 & 0.091 & 0.092 & 0.066 & 0.057 & 0.033 & 0.055 & 0.037 & 0.042 & 0.024 & 0.033 & 0.024 & 0.021 & 0.012 \\ 0.144 & 0.091 & 0.053 & 0.053 & 0.038 & 0.033 & 0.02 & 0.032 & 0.021 & 0.025 & 0.013 & 0.02 & 0.014 & 0.012 & 0.007 \\ 0.213 & 0.126 & 0.087 & 0.075 & 0.051 & 0.056 & 0.031 & 0.044 & 0.031 & 0.033 & 0.018 & 0.032 & 0.024 & 0.02 & 0.011 \\ 0.145 & 0.087 & 0.058 & 0.051 & 0.036 & 0.036 & 0.022 & 0.031 & 0.021 & 0.023 & 0.013 & 0.021 & 0.015 & 0.014 & 0.008 \\ 0.159 & 0.092 & 0.066 & 0.055 & 0.037 & 0.042 & 0.024 & 0.032 & 0.023 & 0.023 & 0.014 & 0.024 & 0.017 & 0.015 & 0.009 \\ 0.091 & 0.053 & 0.038 & 0.032 & 0.021 & 0.025 & 0.013 & 0.019 & 0.013 & 0.013 & 0.008 & 0.015 & 0.01 & 0.009 & 0.005 \\ 0.145 & 0.092 & 0.053 & 0.053 & 0.039 & 0.033 & 0.019 & 0.032 & 0.021 & 0.025 & 0.014 & 0.019 & 0.014 & 0.012 & 0.007 \\ 0.104 & 0.066 & 0.038 & 0.039 & 0.027 & 0.024 & 0.014 & 0.023 & 0.016 & 0.017 & 0.01 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.091 & 0.057 & 0.033 & 0.033 & 0.024 & 0.021 & 0.012 & 0.02 & 0.013 & 0.016 & 0.009 & 0.012 & 0.009 & 0.008 & 0.005 \\ 0.053 & 0.033 & 0.02 & 0.02 & 0.014 & 0.012 & 0.007 & 0.012 & 0.008 & 0.009 & 0.005 & 0.007 & 0.005 & 0.005 & 0.003 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

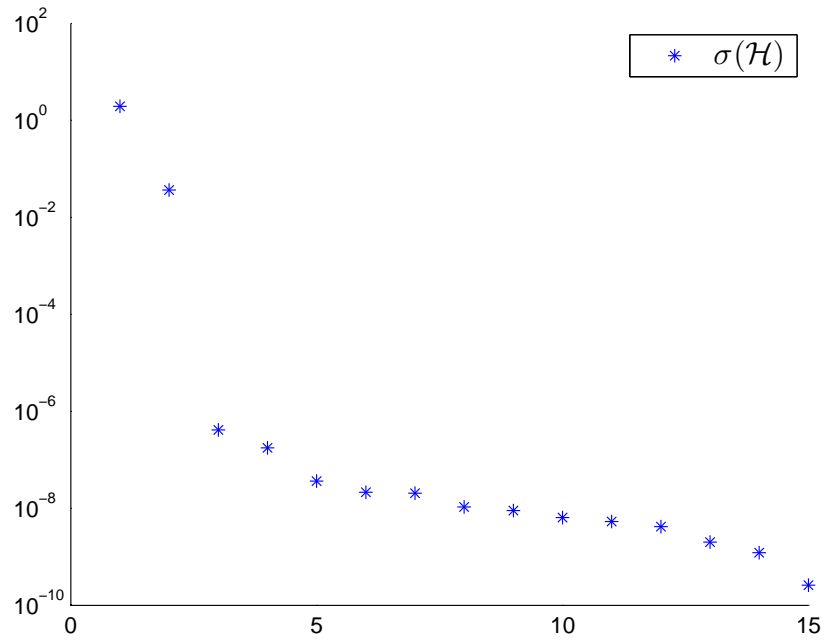
$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.938 \\ 3.881 \times 10^{-2} \\ 2.569 \times 10^{-3} \\ 1.933 \times 10^{-3} \\ 1.469 \times 10^{-3} \\ 1.024 \times 10^{-3} \\ 5.479 \times 10^{-4} \\ 1.405 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{A.3})$$

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

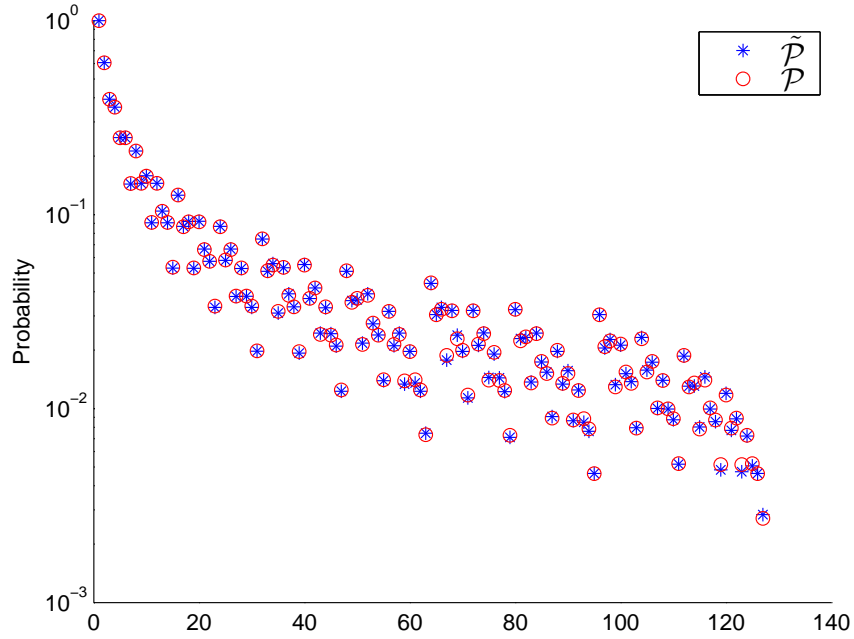
$$\mathcal{H} = \begin{bmatrix} 1.0 & 0.607 & 0.393 & 0.358 & 0.249 & 0.249 & 0.145 & 0.213 & 0.145 & 0.158 & 0.091 & 0.145 & 0.104 & 0.091 & 0.054 \\ 0.607 & 0.358 & 0.249 & 0.213 & 0.145 & 0.158 & 0.091 & 0.126 & 0.087 & 0.092 & 0.053 & 0.092 & 0.066 & 0.057 & 0.034 \\ 0.393 & 0.249 & 0.145 & 0.145 & 0.104 & 0.091 & 0.054 & 0.087 & 0.058 & 0.066 & 0.038 & 0.053 & 0.038 & 0.034 & 0.02 \\ 0.358 & 0.213 & 0.145 & 0.126 & 0.087 & 0.092 & 0.053 & 0.075 & 0.051 & 0.055 & 0.032 & 0.053 & 0.038 & 0.033 & 0.02 \\ 0.249 & 0.145 & 0.104 & 0.087 & 0.058 & 0.066 & 0.038 & 0.051 & 0.035 & 0.037 & 0.021 & 0.038 & 0.028 & 0.024 & 0.014 \\ 0.249 & 0.158 & 0.091 & 0.092 & 0.066 & 0.057 & 0.034 & 0.055 & 0.037 & 0.042 & 0.024 & 0.033 & 0.024 & 0.021 & 0.012 \\ 0.145 & 0.091 & 0.054 & 0.053 & 0.038 & 0.034 & 0.02 & 0.032 & 0.021 & 0.024 & 0.014 & 0.02 & 0.014 & 0.012 & 0.007 \\ 0.213 & 0.126 & 0.087 & 0.075 & 0.051 & 0.055 & 0.032 & 0.044 & 0.03 & 0.032 & 0.019 & 0.032 & 0.023 & 0.02 & 0.012 \\ 0.145 & 0.087 & 0.058 & 0.051 & 0.035 & 0.037 & 0.021 & 0.03 & 0.021 & 0.022 & 0.013 & 0.022 & 0.015 & 0.013 & 0.008 \\ 0.158 & 0.092 & 0.066 & 0.055 & 0.037 & 0.042 & 0.024 & 0.032 & 0.022 & 0.023 & 0.014 & 0.024 & 0.018 & 0.015 & 0.009 \\ 0.091 & 0.053 & 0.038 & 0.032 & 0.021 & 0.024 & 0.014 & 0.019 & 0.013 & 0.014 & 0.008 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.145 & 0.092 & 0.053 & 0.053 & 0.038 & 0.033 & 0.02 & 0.032 & 0.022 & 0.024 & 0.014 & 0.019 & 0.014 & 0.012 & 0.007 \\ 0.104 & 0.066 & 0.038 & 0.038 & 0.028 & 0.024 & 0.014 & 0.023 & 0.015 & 0.018 & 0.01 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.091 & 0.057 & 0.034 & 0.033 & 0.024 & 0.021 & 0.012 & 0.02 & 0.013 & 0.015 & 0.009 & 0.012 & 0.009 & 0.008 & 0.005 \\ 0.054 & 0.034 & 0.02 & 0.02 & 0.014 & 0.012 & 0.007 & 0.012 & 0.008 & 0.009 & 0.005 & 0.007 & 0.005 & 0.005 & 0.003 \end{bmatrix}$$

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.937 \\ 3.652 \times 10^{-2} \\ 4.124 \times 10^{-7} \\ 1.762 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure A.3.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^5$



**Figure A.4.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^5$

This suggests that  $rank_{est}(\mathcal{H}) = 2$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

$$A_n = \begin{bmatrix} 0.356 & 0.644 \\ 0.683 & 0.317 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.778 & 0.222 \\ 0.425 & 0.575 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1.0 & 0.607 & 0.393 & 0.358 & 0.249 & 0.249 & 0.145 & 0.213 & 0.145 & 0.158 & 0.09 & 0.145 & 0.104 & 0.09 & 0.054 \\ 0.607 & 0.358 & 0.249 & 0.213 & 0.145 & 0.158 & 0.09 & 0.126 & 0.087 & 0.092 & 0.053 & 0.092 & 0.066 & 0.056 & 0.034 \\ 0.393 & 0.249 & 0.145 & 0.145 & 0.104 & 0.09 & 0.054 & 0.087 & 0.058 & 0.066 & 0.038 & 0.053 & 0.038 & 0.034 & 0.02 \\ 0.358 & 0.213 & 0.145 & 0.126 & 0.087 & 0.092 & 0.053 & 0.075 & 0.051 & 0.055 & 0.032 & 0.054 & 0.039 & 0.033 & 0.02 \\ 0.249 & 0.145 & 0.104 & 0.087 & 0.058 & 0.066 & 0.038 & 0.051 & 0.035 & 0.037 & 0.021 & 0.039 & 0.028 & 0.023 & 0.014 \\ 0.249 & 0.158 & 0.09 & 0.092 & 0.066 & 0.056 & 0.034 & 0.055 & 0.037 & 0.042 & 0.024 & 0.033 & 0.023 & 0.021 & 0.013 \\ 0.145 & 0.09 & 0.054 & 0.053 & 0.038 & 0.034 & 0.02 & 0.032 & 0.021 & 0.024 & 0.014 & 0.02 & 0.014 & 0.013 & 0.008 \\ 0.213 & 0.126 & 0.087 & 0.075 & 0.051 & 0.055 & 0.032 & 0.044 & 0.03 & 0.033 & 0.019 & 0.032 & 0.023 & 0.02 & 0.012 \\ 0.145 & 0.087 & 0.058 & 0.051 & 0.035 & 0.037 & 0.021 & 0.03 & 0.021 & 0.022 & 0.013 & 0.022 & 0.016 & 0.013 & 0.008 \\ 0.158 & 0.092 & 0.066 & 0.055 & 0.037 & 0.042 & 0.024 & 0.033 & 0.022 & 0.024 & 0.014 & 0.025 & 0.018 & 0.015 & 0.009 \\ 0.09 & 0.053 & 0.038 & 0.032 & 0.021 & 0.024 & 0.014 & 0.019 & 0.013 & 0.014 & 0.008 & 0.014 & 0.01 & 0.008 & 0.005 \\ 0.145 & 0.092 & 0.053 & 0.054 & 0.039 & 0.033 & 0.02 & 0.032 & 0.022 & 0.025 & 0.014 & 0.019 & 0.014 & 0.012 & 0.007 \\ 0.104 & 0.066 & 0.038 & 0.039 & 0.028 & 0.023 & 0.014 & 0.023 & 0.016 & 0.018 & 0.01 & 0.014 & 0.01 & 0.009 & 0.005 \\ 0.09 & 0.056 & 0.034 & 0.033 & 0.023 & 0.021 & 0.013 & 0.02 & 0.013 & 0.015 & 0.008 & 0.012 & 0.009 & 0.008 & 0.005 \\ 0.054 & 0.034 & 0.02 & 0.02 & 0.014 & 0.013 & 0.008 & 0.012 & 0.008 & 0.009 & 0.005 & 0.007 & 0.005 & 0.005 & 0.003 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

### Simulation 11.

1. We simulated the same process and produced a sequence of  $T = 10^4$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1.0 & 0.605 & 0.395 & 0.357 & 0.248 & 0.248 & 0.148 & 0.209 & 0.147 & 0.157 & 0.091 & 0.147 & 0.101 & 0.091 & 0.056 \\ 0.605 & 0.357 & 0.248 & 0.209 & 0.147 & 0.157 & 0.091 & 0.123 & 0.086 & 0.091 & 0.056 & 0.094 & 0.063 & 0.056 & 0.036 \\ 0.395 & 0.248 & 0.148 & 0.147 & 0.101 & 0.091 & 0.056 & 0.086 & 0.061 & 0.066 & 0.035 & 0.053 & 0.038 & 0.036 & 0.021 \\ 0.357 & 0.209 & 0.147 & 0.123 & 0.086 & 0.091 & 0.056 & 0.074 & 0.049 & 0.054 & 0.033 & 0.054 & 0.036 & 0.034 & 0.022 \\ 0.248 & 0.147 & 0.101 & 0.086 & 0.061 & 0.066 & 0.035 & 0.049 & 0.037 & 0.037 & 0.024 & 0.039 & 0.027 & 0.021 & 0.013 \\ 0.248 & 0.157 & 0.091 & 0.094 & 0.063 & 0.056 & 0.036 & 0.057 & 0.037 & 0.039 & 0.024 & 0.032 & 0.023 & 0.023 & 0.013 \\ 0.148 & 0.091 & 0.056 & 0.053 & 0.038 & 0.036 & 0.021 & 0.029 & 0.024 & 0.027 & 0.011 & 0.021 & 0.015 & 0.013 & 0.008 \\ 0.209 & 0.123 & 0.086 & 0.074 & 0.049 & 0.054 & 0.033 & 0.044 & 0.03 & 0.031 & 0.018 & 0.032 & 0.021 & 0.02 & 0.012 \\ 0.147 & 0.086 & 0.061 & 0.049 & 0.037 & 0.037 & 0.024 & 0.03 & 0.02 & 0.022 & 0.015 & 0.022 & 0.015 & 0.014 & 0.01 \\ 0.157 & 0.094 & 0.063 & 0.057 & 0.037 & 0.039 & 0.024 & 0.034 & 0.023 & 0.023 & 0.013 & 0.023 & 0.016 & 0.016 & 0.009 \\ 0.091 & 0.053 & 0.038 & 0.029 & 0.024 & 0.027 & 0.011 & 0.015 & 0.014 & 0.014 & 0.01 & 0.016 & 0.011 & 0.006 & 0.005 \\ 0.147 & 0.091 & 0.056 & 0.054 & 0.036 & 0.034 & 0.022 & 0.033 & 0.021 & 0.022 & 0.015 & 0.019 & 0.015 & 0.015 & 0.008 \\ 0.101 & 0.066 & 0.035 & 0.039 & 0.027 & 0.021 & 0.013 & 0.024 & 0.015 & 0.017 & 0.01 & 0.014 & 0.008 & 0.008 & 0.005 \\ 0.091 & 0.056 & 0.036 & 0.032 & 0.023 & 0.023 & 0.013 & 0.017 & 0.016 & 0.017 & 0.006 & 0.013 & 0.009 & 0.008 & 0.005 \\ 0.056 & 0.036 & 0.021 & 0.021 & 0.015 & 0.013 & 0.008 & 0.013 & 0.009 & 0.01 & 0.004 & 0.008 & 0.005 & 0.005 & 0.003 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.935 \\ 3.441 \times 10^{-2} \\ 1.178 \times 10^{-2} \\ 8.023 \times 10^{-3} \\ 4.481 \times 10^{-3} \\ 3.141 \times 10^{-3} \\ 9.644 \times 10^{-4} \\ 6.341 \times 10^{-5} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{A.4})$$

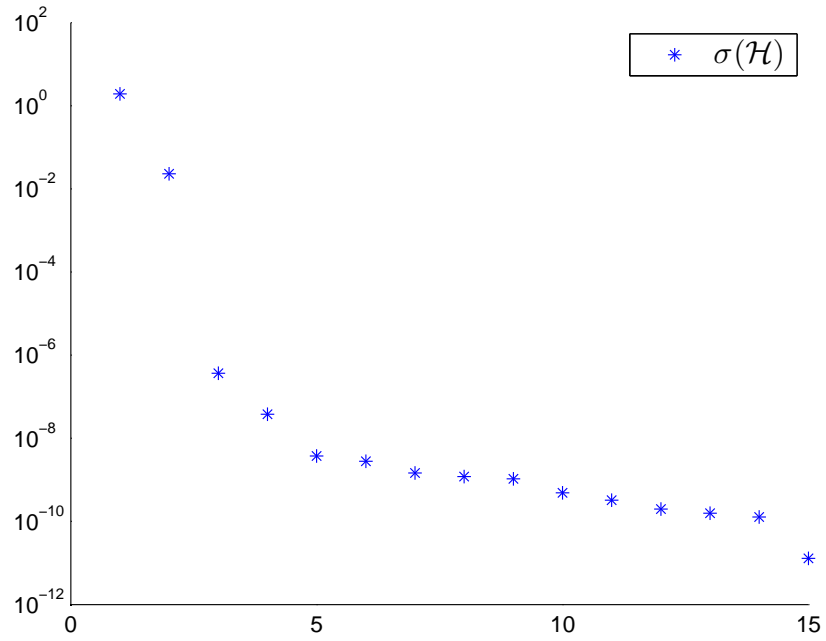
3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .



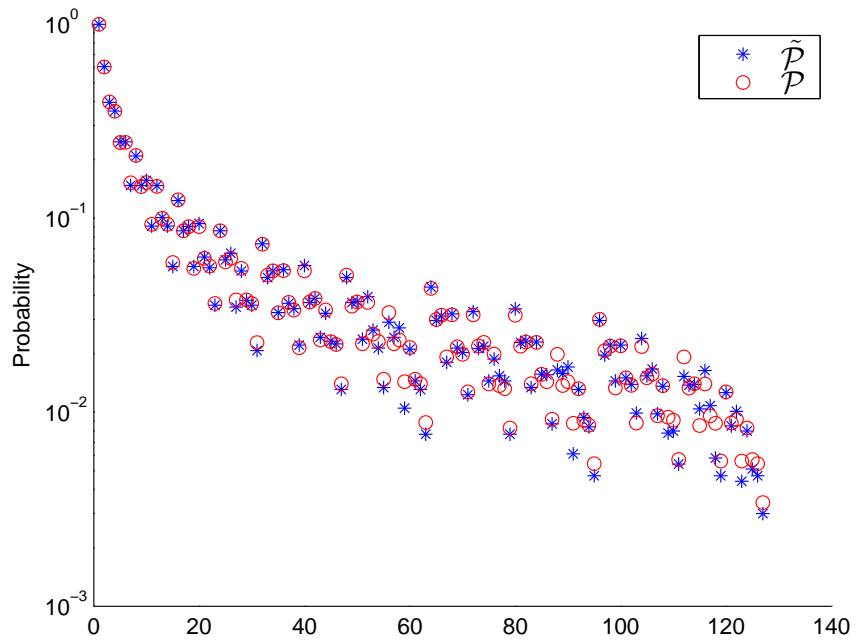
$$\mathcal{H} = \begin{bmatrix} 1.0 & 0.602 & 0.398 & 0.356 & 0.246 & 0.246 & 0.152 & 0.211 & 0.146 & 0.153 & 0.093 & 0.146 & 0.1 & 0.093 & 0.059 \\ 0.602 & 0.356 & 0.246 & 0.211 & 0.146 & 0.153 & 0.093 & 0.124 & 0.086 & 0.091 & 0.055 & 0.091 & 0.062 & 0.057 & 0.036 \\ 0.398 & 0.246 & 0.152 & 0.146 & 0.1 & 0.093 & 0.059 & 0.086 & 0.06 & 0.062 & 0.038 & 0.055 & 0.038 & 0.036 & 0.023 \\ 0.356 & 0.211 & 0.146 & 0.124 & 0.086 & 0.091 & 0.055 & 0.074 & 0.051 & 0.054 & 0.033 & 0.054 & 0.037 & 0.034 & 0.022 \\ 0.246 & 0.146 & 0.1 & 0.086 & 0.06 & 0.062 & 0.038 & 0.051 & 0.035 & 0.037 & 0.023 & 0.037 & 0.025 & 0.023 & 0.015 \\ 0.246 & 0.153 & 0.093 & 0.091 & 0.062 & 0.057 & 0.036 & 0.054 & 0.037 & 0.039 & 0.024 & 0.034 & 0.023 & 0.022 & 0.014 \\ 0.152 & 0.093 & 0.059 & 0.055 & 0.038 & 0.036 & 0.023 & 0.033 & 0.023 & 0.024 & 0.014 & 0.022 & 0.015 & 0.014 & 0.009 \\ 0.211 & 0.124 & 0.086 & 0.074 & 0.051 & 0.054 & 0.033 & 0.043 & 0.03 & 0.032 & 0.019 & 0.032 & 0.022 & 0.02 & 0.013 \\ 0.146 & 0.086 & 0.06 & 0.051 & 0.035 & 0.037 & 0.023 & 0.03 & 0.021 & 0.022 & 0.013 & 0.022 & 0.015 & 0.014 & 0.009 \\ 0.153 & 0.091 & 0.062 & 0.054 & 0.037 & 0.039 & 0.024 & 0.032 & 0.022 & 0.023 & 0.014 & 0.023 & 0.016 & 0.014 & 0.009 \\ 0.093 & 0.055 & 0.038 & 0.033 & 0.023 & 0.024 & 0.014 & 0.019 & 0.013 & 0.014 & 0.009 & 0.014 & 0.01 & 0.009 & 0.006 \\ 0.146 & 0.091 & 0.055 & 0.054 & 0.037 & 0.034 & 0.022 & 0.032 & 0.022 & 0.023 & 0.014 & 0.02 & 0.014 & 0.013 & 0.008 \\ 0.1 & 0.062 & 0.038 & 0.037 & 0.025 & 0.023 & 0.015 & 0.022 & 0.015 & 0.016 & 0.01 & 0.014 & 0.009 & 0.009 & 0.006 \\ 0.093 & 0.057 & 0.036 & 0.034 & 0.023 & 0.022 & 0.014 & 0.02 & 0.014 & 0.014 & 0.009 & 0.013 & 0.009 & 0.009 & 0.005 \\ 0.059 & 0.036 & 0.023 & 0.022 & 0.015 & 0.014 & 0.009 & 0.013 & 0.009 & 0.009 & 0.006 & 0.008 & 0.006 & 0.005 & 0.003 \end{bmatrix}$$

3. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.932 \\ 2.299 \times 10^{-2} \\ 3.679 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure A.5.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^4$



**Figure A.6.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^4$

This suggests that  $\text{rank}_{est}(\mathcal{H}) = 2$ .

4. Estimate the matrices  $A_n$  and  $B_n$ .

$$A_n = \begin{bmatrix} 0.415 & 0.585 \\ 0.469 & 0.531 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.225 & 0.775 \\ 0.904 & 0.096 \end{bmatrix}.$$

5. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1.0 & 0.602 & 0.398 & 0.356 & 0.246 & 0.246 & 0.152 & 0.211 & 0.145 & 0.152 & 0.094 & 0.145 & 0.1 & 0.094 & 0.058 \\ 0.602 & 0.356 & 0.246 & 0.211 & 0.145 & 0.152 & 0.094 & 0.125 & 0.086 & 0.09 & 0.056 & 0.09 & 0.062 & 0.058 & 0.036 \\ 0.398 & 0.246 & 0.152 & 0.145 & 0.1 & 0.094 & 0.058 & 0.086 & 0.059 & 0.062 & 0.038 & 0.056 & 0.038 & 0.036 & 0.022 \\ 0.356 & 0.211 & 0.145 & 0.125 & 0.086 & 0.09 & 0.056 & 0.074 & 0.051 & 0.053 & 0.033 & 0.053 & 0.037 & 0.034 & 0.021 \\ 0.246 & 0.145 & 0.1 & 0.086 & 0.059 & 0.062 & 0.038 & 0.051 & 0.035 & 0.037 & 0.023 & 0.037 & 0.025 & 0.024 & 0.015 \\ 0.246 & 0.152 & 0.094 & 0.09 & 0.062 & 0.058 & 0.036 & 0.053 & 0.037 & 0.038 & 0.024 & 0.034 & 0.024 & 0.022 & 0.014 \\ 0.152 & 0.094 & 0.058 & 0.056 & 0.038 & 0.036 & 0.022 & 0.033 & 0.023 & 0.024 & 0.015 & 0.021 & 0.015 & 0.014 & 0.009 \\ 0.211 & 0.125 & 0.086 & 0.074 & 0.051 & 0.053 & 0.033 & 0.044 & 0.03 & 0.031 & 0.019 & 0.031 & 0.022 & 0.02 & 0.013 \\ 0.145 & 0.086 & 0.059 & 0.051 & 0.035 & 0.037 & 0.023 & 0.03 & 0.021 & 0.022 & 0.013 & 0.022 & 0.015 & 0.014 & 0.009 \\ 0.152 & 0.09 & 0.062 & 0.053 & 0.037 & 0.038 & 0.024 & 0.031 & 0.022 & 0.023 & 0.014 & 0.023 & 0.016 & 0.015 & 0.009 \\ 0.094 & 0.056 & 0.038 & 0.033 & 0.023 & 0.024 & 0.015 & 0.019 & 0.013 & 0.014 & 0.009 & 0.014 & 0.01 & 0.009 & 0.006 \\ 0.145 & 0.09 & 0.056 & 0.053 & 0.037 & 0.034 & 0.021 & 0.031 & 0.022 & 0.023 & 0.014 & 0.02 & 0.014 & 0.013 & 0.008 \\ 0.1 & 0.062 & 0.038 & 0.037 & 0.025 & 0.024 & 0.015 & 0.022 & 0.015 & 0.016 & 0.01 & 0.014 & 0.01 & 0.009 & 0.006 \\ 0.094 & 0.058 & 0.036 & 0.034 & 0.024 & 0.022 & 0.014 & 0.02 & 0.014 & 0.015 & 0.009 & 0.013 & 0.009 & 0.008 & 0.005 \\ 0.058 & 0.036 & 0.022 & 0.021 & 0.015 & 0.014 & 0.009 & 0.013 & 0.009 & 0.009 & 0.006 & 0.008 & 0.006 & 0.005 & 0.003 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

## Simulation 12.

1. We simulated the same process and produced a sequence of  $T = 10^3$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1.0 & 0.623 & 0.377 & 0.367 & 0.256 & 0.255 & 0.122 & 0.215 & 0.152 & 0.172 & 0.084 & 0.152 & 0.104 & 0.084 & 0.037 \\ 0.623 & 0.367 & 0.256 & 0.215 & 0.152 & 0.172 & 0.084 & 0.131 & 0.084 & 0.103 & 0.049 & 0.101 & 0.071 & 0.057 & 0.027 \\ 0.377 & 0.255 & 0.122 & 0.152 & 0.104 & 0.084 & 0.037 & 0.084 & 0.067 & 0.068 & 0.035 & 0.051 & 0.033 & 0.027 & 0.01 \\ 0.367 & 0.215 & 0.152 & 0.131 & 0.084 & 0.103 & 0.049 & 0.084 & 0.046 & 0.057 & 0.027 & 0.065 & 0.037 & 0.033 & 0.016 \\ 0.255 & 0.152 & 0.104 & 0.084 & 0.067 & 0.068 & 0.035 & 0.046 & 0.038 & 0.045 & 0.022 & 0.035 & 0.033 & 0.024 & 0.011 \\ 0.256 & 0.172 & 0.084 & 0.101 & 0.071 & 0.057 & 0.027 & 0.056 & 0.044 & 0.046 & 0.025 & 0.037 & 0.02 & 0.02 & 0.007 \\ 0.122 & 0.084 & 0.037 & 0.051 & 0.033 & 0.027 & 0.01 & 0.028 & 0.023 & 0.023 & 0.01 & 0.014 & 0.013 & 0.007 & 0.003 \\ 0.215 & 0.131 & 0.084 & 0.084 & 0.046 & 0.057 & 0.027 & 0.05 & 0.034 & 0.027 & 0.019 & 0.039 & 0.018 & 0.018 & 0.009 \\ 0.152 & 0.084 & 0.067 & 0.046 & 0.038 & 0.045 & 0.022 & 0.034 & 0.012 & 0.03 & 0.008 & 0.026 & 0.019 & 0.015 & 0.007 \\ 0.172 & 0.101 & 0.071 & 0.056 & 0.044 & 0.046 & 0.025 & 0.034 & 0.022 & 0.031 & 0.013 & 0.028 & 0.018 & 0.018 & 0.007 \\ 0.084 & 0.051 & 0.033 & 0.028 & 0.023 & 0.023 & 0.01 & 0.012 & 0.016 & 0.014 & 0.009 & 0.008 & 0.015 & 0.006 & 0.004 \\ 0.152 & 0.103 & 0.049 & 0.065 & 0.037 & 0.033 & 0.016 & 0.035 & 0.03 & 0.02 & 0.017 & 0.024 & 0.009 & 0.014 & 0.002 \\ 0.104 & 0.068 & 0.035 & 0.035 & 0.033 & 0.024 & 0.011 & 0.021 & 0.014 & 0.025 & 0.008 & 0.013 & 0.011 & 0.006 & 0.005 \\ 0.084 & 0.057 & 0.027 & 0.037 & 0.02 & 0.02 & 0.007 & 0.02 & 0.017 & 0.015 & 0.005 & 0.01 & 0.01 & 0.005 & 0.002 \\ 0.037 & 0.027 & 0.01 & 0.014 & 0.013 & 0.007 & 0.003 & 0.008 & 0.006 & 0.008 & 0.005 & 0.004 & 0.003 & 0.002 & 0.001 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

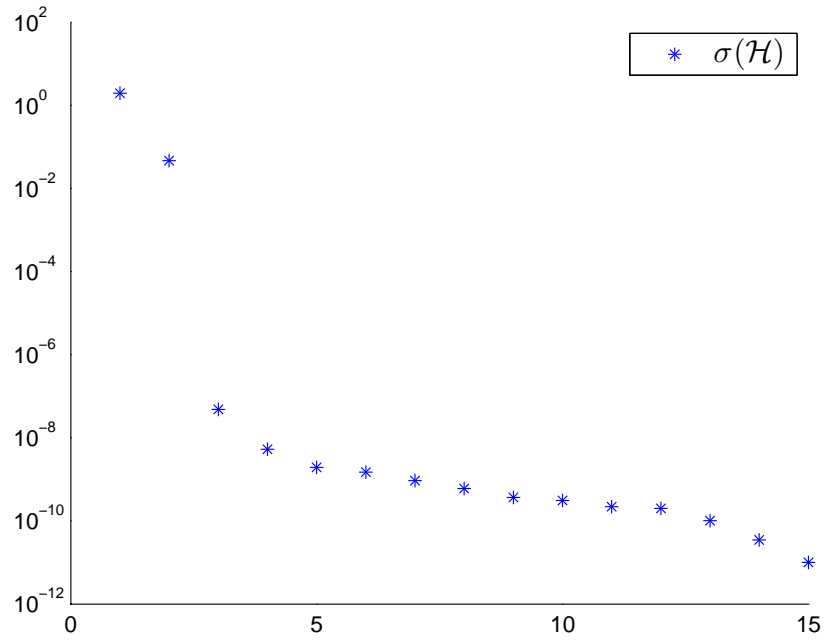
$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.9577.510 \times 10^{-2} \\ 2.941 \times 10^{-2} \\ 1.650 \times 10^{-2} \\ 1.103 \times 10^{-2} \\ 4.672 \times 10^{-3} \\ 2.927 \times 10^{-3} \\ 7.023 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{A.5})$$

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

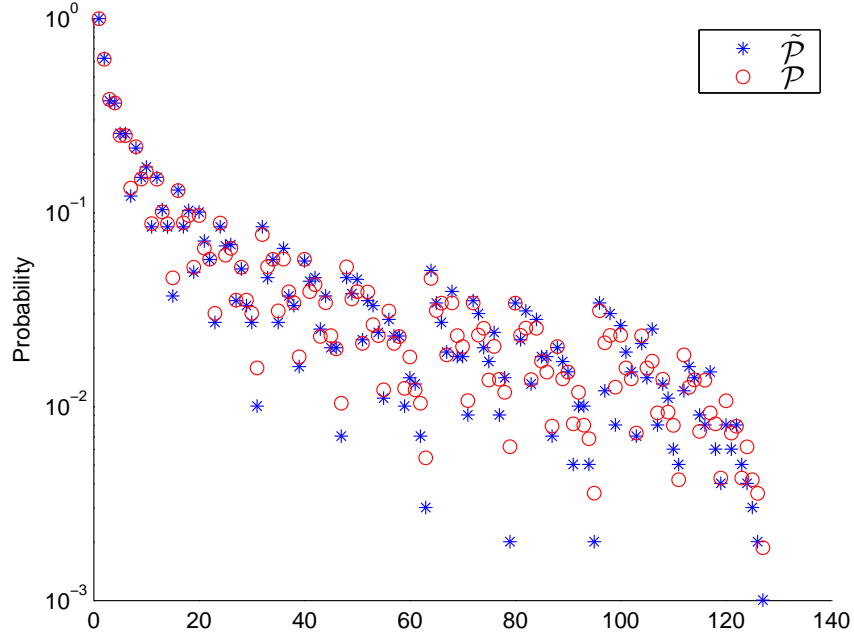
$$\mathcal{H} = \begin{bmatrix} 1.0 & 0.617 & 0.383 & 0.367 & 0.25 & 0.249 & 0.134 & 0.218 & 0.149 & 0.162 & 0.088 & 0.149 & 0.101 & 0.088 & 0.046 \\ 0.617 & 0.367 & 0.25 & 0.218 & 0.149 & 0.162 & 0.088 & 0.13 & 0.089 & 0.097 & 0.052 & 0.097 & 0.065 & 0.057 & 0.03 \\ 0.383 & 0.249 & 0.134 & 0.149 & 0.101 & 0.088 & 0.046 & 0.088 & 0.06 & 0.065 & 0.035 & 0.052 & 0.035 & 0.03 & 0.016 \\ 0.367 & 0.218 & 0.149 & 0.13 & 0.089 & 0.097 & 0.052 & 0.077 & 0.053 & 0.057 & 0.031 & 0.058 & 0.039 & 0.034 & 0.018 \\ 0.249 & 0.149 & 0.101 & 0.088 & 0.06 & 0.065 & 0.035 & 0.053 & 0.036 & 0.039 & 0.021 & 0.039 & 0.026 & 0.023 & 0.012 \\ 0.25 & 0.162 & 0.088 & 0.097 & 0.065 & 0.057 & 0.03 & 0.057 & 0.039 & 0.042 & 0.023 & 0.034 & 0.023 & 0.02 & 0.01 \\ 0.134 & 0.088 & 0.046 & 0.052 & 0.035 & 0.03 & 0.016 & 0.031 & 0.021 & 0.023 & 0.012 & 0.018 & 0.012 & 0.01 & 0.005 \\ 0.218 & 0.13 & 0.089 & 0.077 & 0.053 & 0.057 & 0.031 & 0.046 & 0.031 & 0.034 & 0.018 & 0.034 & 0.023 & 0.02 & 0.011 \\ 0.149 & 0.088 & 0.06 & 0.053 & 0.036 & 0.039 & 0.021 & 0.031 & 0.021 & 0.023 & 0.013 & 0.023 & 0.016 & 0.014 & 0.007 \\ 0.162 & 0.097 & 0.065 & 0.057 & 0.039 & 0.042 & 0.023 & 0.034 & 0.023 & 0.025 & 0.014 & 0.025 & 0.017 & 0.015 & 0.008 \\ 0.088 & 0.052 & 0.035 & 0.031 & 0.021 & 0.023 & 0.012 & 0.018 & 0.013 & 0.014 & 0.007 & 0.014 & 0.009 & 0.008 & 0.004 \\ 0.149 & 0.097 & 0.052 & 0.058 & 0.039 & 0.034 & 0.018 & 0.034 & 0.023 & 0.025 & 0.014 & 0.02 & 0.014 & 0.012 & 0.006 \\ 0.101 & 0.065 & 0.035 & 0.039 & 0.026 & 0.023 & 0.012 & 0.023 & 0.016 & 0.017 & 0.009 & 0.014 & 0.009 & 0.008 & 0.004 \\ 0.088 & 0.057 & 0.03 & 0.034 & 0.023 & 0.02 & 0.01 & 0.02 & 0.014 & 0.015 & 0.008 & 0.012 & 0.008 & 0.007 & 0.004 \\ 0.046 & 0.03 & 0.016 & 0.018 & 0.012 & 0.01 & 0.005 & 0.011 & 0.007 & 0.008 & 0.004 & 0.006 & 0.004 & 0.004 & 0.002 \end{bmatrix}$$

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.9494.663 \times 10^{-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure A.7.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^3$



**Figure A.8.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 2$ ,  $L_{seq} = 10^3$

This suggests that  $rank_{est}(\mathcal{H}) = 2$ .

5. Estimate the matrices  $A_n$  and  $B_n$  by solving the problem 3.8.

$$A_n = \begin{bmatrix} 0.427 & 0.573 \\ 0.671 & 0.329 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.832 & 0.168 \\ 0.364 & 0.636 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1.0 & 0.617 & 0.383 & 0.367 & 0.25 & 0.249 & 0.134 & 0.22 & 0.147 & 0.164 & 0.086 & 0.147 & 0.103 & 0.086 & 0.048 \\ 0.617 & 0.367 & 0.25 & 0.22 & 0.147 & 0.164 & 0.086 & 0.132 & 0.088 & 0.096 & 0.051 & 0.096 & 0.067 & 0.056 & 0.031 \\ 0.383 & 0.249 & 0.134 & 0.147 & 0.103 & 0.086 & 0.048 & 0.088 & 0.059 & 0.067 & 0.035 & 0.051 & 0.035 & 0.031 & 0.017 \\ 0.367 & 0.22 & 0.147 & 0.132 & 0.088 & 0.096 & 0.051 & 0.079 & 0.053 & 0.058 & 0.031 & 0.057 & 0.04 & 0.033 & 0.018 \\ 0.249 & 0.147 & 0.103 & 0.088 & 0.059 & 0.067 & 0.035 & 0.053 & 0.035 & 0.038 & 0.02 & 0.04 & 0.028 & 0.023 & 0.013 \\ 0.25 & 0.164 & 0.086 & 0.096 & 0.067 & 0.056 & 0.031 & 0.058 & 0.038 & 0.044 & 0.023 & 0.033 & 0.023 & 0.02 & 0.011 \\ 0.134 & 0.086 & 0.048 & 0.051 & 0.035 & 0.031 & 0.017 & 0.031 & 0.02 & 0.023 & 0.012 & 0.018 & 0.013 & 0.011 & 0.006 \\ 0.22 & 0.132 & 0.088 & 0.079 & 0.053 & 0.058 & 0.031 & 0.047 & 0.032 & 0.035 & 0.018 & 0.034 & 0.024 & 0.02 & 0.011 \\ 0.147 & 0.088 & 0.059 & 0.053 & 0.035 & 0.038 & 0.02 & 0.032 & 0.021 & 0.023 & 0.012 & 0.023 & 0.016 & 0.013 & 0.007 \\ 0.164 & 0.096 & 0.067 & 0.058 & 0.038 & 0.044 & 0.023 & 0.035 & 0.023 & 0.025 & 0.013 & 0.026 & 0.018 & 0.015 & 0.008 \\ 0.086 & 0.051 & 0.035 & 0.031 & 0.02 & 0.023 & 0.012 & 0.018 & 0.012 & 0.013 & 0.007 & 0.014 & 0.01 & 0.008 & 0.004 \\ 0.147 & 0.096 & 0.051 & 0.057 & 0.04 & 0.033 & 0.018 & 0.034 & 0.023 & 0.026 & 0.014 & 0.019 & 0.013 & 0.012 & 0.006 \\ 0.103 & 0.067 & 0.035 & 0.04 & 0.028 & 0.023 & 0.013 & 0.024 & 0.016 & 0.018 & 0.01 & 0.013 & 0.009 & 0.008 & 0.004 \\ 0.086 & 0.056 & 0.031 & 0.033 & 0.023 & 0.02 & 0.011 & 0.02 & 0.013 & 0.015 & 0.008 & 0.012 & 0.008 & 0.007 & 0.004 \\ 0.048 & 0.031 & 0.017 & 0.018 & 0.013 & 0.011 & 0.006 & 0.011 & 0.007 & 0.008 & 0.004 & 0.006 & 0.004 & 0.004 & 0.002 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

# Appendix B

## Simulation Results for Case 4

### B.1 Case 4: $t = 3, N = 3$

Here, given the state transition matrix  $A$ , the output matrix  $B$  and (approximate) string probabilities of strings of length  $t = 3$ , we obtain the  $\lambda$  value that makes the order of the model is closest the real order of the system for different sequence lengths  $L_{seq}$ . Here,  $A$  and  $B$  matrices are arbitrarily chosen.

Consider the process described by  $(A, B)$ ;

$$A = \begin{bmatrix} 0.4 & 0.226 & 0.374 \\ 0.782 & 0.074 & 0.144 \\ 0.046 & 0.873 & 0.081 \end{bmatrix}, \quad B = \begin{bmatrix} 0.467 & 0.533 \\ 0.282 & 0.718 \\ 0.633 & 0.367 \end{bmatrix}. \quad (\text{B.1})$$

The Hankel matrix consist of real string probabilities  $\mathcal{H}_r$  corresponding to  $(A, B)$  is given by B.1. Then, since the rank of the Hankel matrix consist of real string probabilities is equal to the order of the model, the singular values of  $\mathcal{H}_r$  i.e.  $\sigma(\mathcal{H}_r)$



is obtained as follows;

$$\sigma(\mathcal{H}_r) = \begin{bmatrix} 1.891 \\ 2.387 \times 10^{-2} \\ 2.421 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and the  $\mathcal{H}_r$  is;

$$\mathcal{H}_r = \begin{bmatrix} 1 & 0.445 & 0.555 & 0.192 & 0.254 & 0.254 & 0.301 & 0.082 & 0.11 & 0.116 & 0.138 & 0.11 & 0.144 & 0.138 & 0.163 \\ 0.445 & 0.192 & 0.254 & 0.082 & 0.11 & 0.116 & 0.138 & 0.035 & 0.047 & 0.05 & 0.059 & 0.051 & 0.065 & 0.063 & 0.074 \\ 0.555 & 0.254 & 0.301 & 0.11 & 0.144 & 0.138 & 0.163 & 0.047 & 0.063 & 0.066 & 0.078 & 0.059 & 0.079 & 0.074 & 0.089 \\ 0.192 & 0.082 & 0.11 & 0.035 & 0.047 & 0.05 & 0.059 & 0.015 & 0.02 & 0.021 & 0.025 & 0.022 & 0.028 & 0.027 & 0.032 \\ 0.254 & 0.11 & 0.144 & 0.047 & 0.063 & 0.066 & 0.078 & 0.02 & 0.026 & 0.029 & 0.034 & 0.029 & 0.037 & 0.036 & 0.042 \\ 0.254 & 0.116 & 0.138 & 0.051 & 0.065 & 0.063 & 0.074 & 0.021 & 0.029 & 0.03 & 0.036 & 0.027 & 0.036 & 0.034 & 0.041 \\ 0.301 & 0.138 & 0.163 & 0.059 & 0.079 & 0.074 & 0.089 & 0.025 & 0.034 & 0.036 & 0.043 & 0.032 & 0.042 & 0.041 & 0.048 \\ 0.082 & 0.035 & 0.047 & 0.015 & 0.02 & 0.021 & 0.025 & 0.007 & 0.009 & 0.009 & 0.011 & 0.009 & 0.012 & 0.012 & 0.014 \\ 0.11 & 0.047 & 0.063 & 0.02 & 0.026 & 0.029 & 0.034 & 0.009 & 0.011 & 0.012 & 0.014 & 0.013 & 0.016 & 0.016 & 0.018 \\ 0.116 & 0.051 & 0.065 & 0.021 & 0.029 & 0.03 & 0.036 & 0.009 & 0.012 & 0.013 & 0.016 & 0.013 & 0.017 & 0.016 & 0.019 \\ 0.138 & 0.059 & 0.079 & 0.025 & 0.034 & 0.036 & 0.043 & 0.011 & 0.014 & 0.016 & 0.018 & 0.016 & 0.02 & 0.02 & 0.023 \\ 0.11 & 0.05 & 0.059 & 0.022 & 0.028 & 0.027 & 0.032 & 0.009 & 0.013 & 0.013 & 0.015 & 0.012 & 0.016 & 0.015 & 0.017 \\ 0.144 & 0.066 & 0.078 & 0.029 & 0.037 & 0.036 & 0.042 & 0.012 & 0.017 & 0.017 & 0.02 & 0.015 & 0.021 & 0.019 & 0.023 \\ 0.138 & 0.063 & 0.074 & 0.027 & 0.036 & 0.034 & 0.041 & 0.012 & 0.015 & 0.017 & 0.02 & 0.015 & 0.019 & 0.019 & 0.022 \\ 0.163 & 0.074 & 0.089 & 0.032 & 0.042 & 0.041 & 0.048 & 0.014 & 0.018 & 0.019 & 0.023 & 0.017 & 0.023 & 0.022 & 0.026 \end{bmatrix}$$

It is clear from the construction of  $\mathcal{H}_r$  that the probabilities in  $\mathcal{H}_r$  will be the same for all sequence lengths  $L_{seq}$ . In the simulation example, this model and the Hankel matrix corresponding to real probabilities  $\mathcal{H}_r$  are unknown; however, we give them here to evaluate the performance of the algorithms.

### Simulation 13.

1. We simulated this process and produced a sequence of  $T = 10^6$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.445 & 0.555 & 0.191 & 0.254 & 0.254 & 0.301 & 0.082 & 0.109 & 0.116 & 0.137 & 0.109 & 0.144 & 0.137 & 0.164 \\ 0.445 & 0.191 & 0.254 & 0.082 & 0.109 & 0.116 & 0.137 & 0.035 & 0.047 & 0.05 & 0.059 & 0.051 & 0.065 & 0.063 & 0.074 \\ 0.555 & 0.254 & 0.301 & 0.109 & 0.144 & 0.137 & 0.164 & 0.047 & 0.063 & 0.066 & 0.078 & 0.059 & 0.079 & 0.074 & 0.09 \\ 0.191 & 0.082 & 0.109 & 0.035 & 0.047 & 0.05 & 0.059 & 0.015 & 0.02 & 0.021 & 0.025 & 0.022 & 0.029 & 0.027 & 0.032 \\ 0.254 & 0.109 & 0.144 & 0.047 & 0.063 & 0.066 & 0.078 & 0.02 & 0.026 & 0.029 & 0.034 & 0.029 & 0.037 & 0.036 & 0.042 \\ 0.254 & 0.116 & 0.137 & 0.051 & 0.065 & 0.063 & 0.074 & 0.021 & 0.029 & 0.03 & 0.036 & 0.027 & 0.036 & 0.034 & 0.041 \\ 0.301 & 0.137 & 0.164 & 0.059 & 0.079 & 0.074 & 0.09 & 0.025 & 0.034 & 0.036 & 0.043 & 0.032 & 0.042 & 0.041 & 0.049 \\ 0.082 & 0.035 & 0.047 & 0.015 & 0.02 & 0.021 & 0.025 & 0.006 & 0.009 & 0.009 & 0.011 & 0.009 & 0.012 & 0.012 & 0.014 \\ 0.109 & 0.047 & 0.063 & 0.02 & 0.026 & 0.029 & 0.034 & 0.009 & 0.011 & 0.012 & 0.014 & 0.013 & 0.016 & 0.016 & 0.018 \\ 0.116 & 0.051 & 0.065 & 0.021 & 0.029 & 0.03 & 0.036 & 0.009 & 0.012 & 0.013 & 0.016 & 0.013 & 0.017 & 0.016 & 0.019 \\ 0.137 & 0.059 & 0.079 & 0.025 & 0.034 & 0.036 & 0.043 & 0.011 & 0.014 & 0.016 & 0.018 & 0.016 & 0.02 & 0.02 & 0.023 \\ 0.109 & 0.05 & 0.059 & 0.022 & 0.029 & 0.027 & 0.032 & 0.009 & 0.013 & 0.013 & 0.016 & 0.012 & 0.016 & 0.014 & 0.018 \\ 0.144 & 0.066 & 0.078 & 0.029 & 0.037 & 0.036 & 0.042 & 0.012 & 0.017 & 0.017 & 0.02 & 0.015 & 0.021 & 0.019 & 0.023 \\ 0.137 & 0.063 & 0.074 & 0.027 & 0.036 & 0.034 & 0.041 & 0.012 & 0.015 & 0.017 & 0.02 & 0.014 & 0.019 & 0.019 & 0.022 \\ 0.164 & 0.074 & 0.09 & 0.032 & 0.042 & 0.041 & 0.049 & 0.014 & 0.018 & 0.019 & 0.023 & 0.018 & 0.023 & 0.022 & 0.027 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.892 \\ 2.329 \times 10^{-2} \\ 2.660 \times 10^{-3} \\ 4.895 \times 10^{-4} \\ 3.660 \times 10^{-4} \\ 2.791 \times 10^{-4} \\ 1.836 \times 10^{-4} \\ 8.021 \times 10^{-5} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{B.2})$$

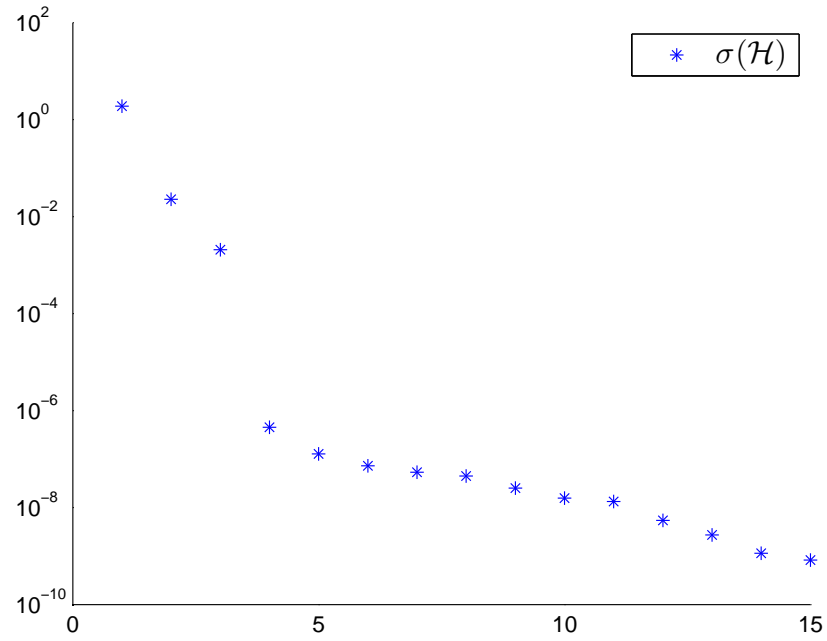
As it is seen from  $\sigma(\tilde{\mathcal{H}})$ , the gap between 3<sub>rd</sub> and 4<sub>th</sub> singular value is not significant. So, it cannot be assumed that the rank of  $\tilde{\mathcal{H}}$  is 3.

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

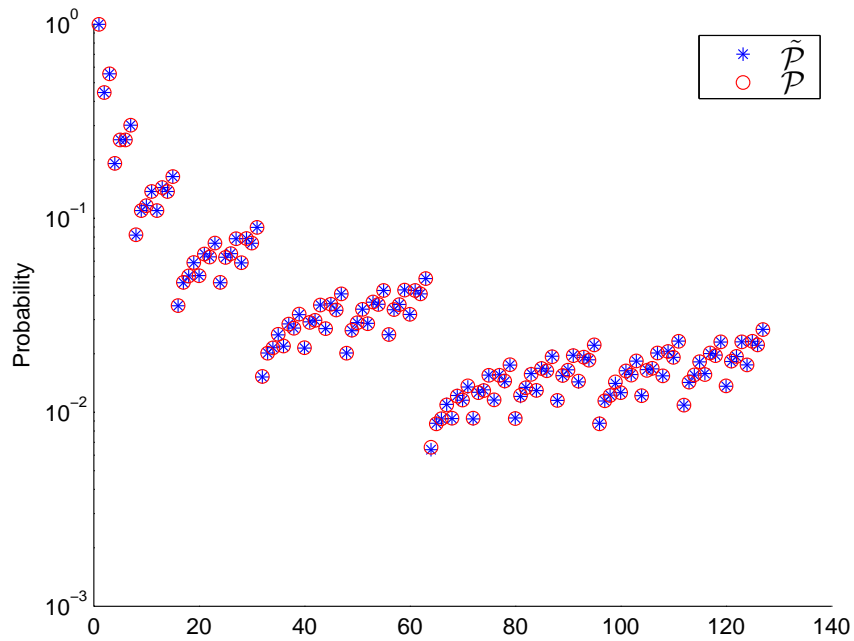
$$\mathcal{H} = \begin{bmatrix} 1 & 0.445 & 0.555 & 0.192 & 0.253 & 0.253 & 0.302 & 0.082 & 0.109 & 0.116 & 0.137 & 0.109 & 0.144 & 0.137 & 0.164 \\ 0.445 & 0.192 & 0.253 & 0.082 & 0.109 & 0.116 & 0.137 & 0.035 & 0.047 & 0.05 & 0.059 & 0.051 & 0.065 & 0.063 & 0.074 \\ 0.555 & 0.253 & 0.302 & 0.109 & 0.144 & 0.137 & 0.164 & 0.047 & 0.063 & 0.066 & 0.078 & 0.059 & 0.078 & 0.074 & 0.09 \\ 0.192 & 0.082 & 0.109 & 0.035 & 0.047 & 0.05 & 0.059 & 0.015 & 0.02 & 0.021 & 0.025 & 0.022 & 0.028 & 0.027 & 0.032 \\ 0.253 & 0.109 & 0.144 & 0.047 & 0.063 & 0.066 & 0.078 & 0.02 & 0.027 & 0.029 & 0.034 & 0.029 & 0.037 & 0.036 & 0.042 \\ 0.253 & 0.116 & 0.137 & 0.051 & 0.065 & 0.063 & 0.074 & 0.021 & 0.029 & 0.03 & 0.036 & 0.027 & 0.036 & 0.034 & 0.041 \\ 0.302 & 0.137 & 0.164 & 0.059 & 0.078 & 0.074 & 0.09 & 0.025 & 0.034 & 0.036 & 0.043 & 0.032 & 0.042 & 0.041 & 0.049 \\ 0.082 & 0.035 & 0.047 & 0.015 & 0.02 & 0.021 & 0.025 & 0.007 & 0.009 & 0.009 & 0.011 & 0.009 & 0.012 & 0.012 & 0.014 \\ 0.109 & 0.047 & 0.063 & 0.02 & 0.027 & 0.029 & 0.034 & 0.009 & 0.011 & 0.012 & 0.014 & 0.013 & 0.016 & 0.016 & 0.018 \\ 0.116 & 0.051 & 0.065 & 0.021 & 0.029 & 0.03 & 0.036 & 0.009 & 0.012 & 0.013 & 0.016 & 0.013 & 0.017 & 0.016 & 0.019 \\ 0.137 & 0.059 & 0.078 & 0.025 & 0.034 & 0.036 & 0.043 & 0.011 & 0.014 & 0.016 & 0.018 & 0.016 & 0.02 & 0.02 & 0.023 \\ 0.109 & 0.05 & 0.059 & 0.022 & 0.028 & 0.027 & 0.032 & 0.009 & 0.013 & 0.013 & 0.016 & 0.012 & 0.016 & 0.014 & 0.018 \\ 0.144 & 0.066 & 0.078 & 0.029 & 0.037 & 0.036 & 0.042 & 0.012 & 0.016 & 0.017 & 0.02 & 0.015 & 0.021 & 0.019 & 0.023 \\ 0.137 & 0.063 & 0.074 & 0.027 & 0.036 & 0.034 & 0.041 & 0.012 & 0.015 & 0.016 & 0.02 & 0.014 & 0.019 & 0.019 & 0.022 \\ 0.164 & 0.074 & 0.09 & 0.032 & 0.042 & 0.041 & 0.049 & 0.014 & 0.018 & 0.019 & 0.023 & 0.018 & 0.023 & 0.022 & 0.027 \end{bmatrix}$$

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.892 \\ 2.265 \times 10^{-2} \\ 2.064 \times 10^{-3} \\ 4.532 \times 10^{-7} \\ 1.275 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure B.1.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^6$



**Figure B.2.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^6$

This suggests that  $\text{rank}_{est}(\mathcal{H}) = 3$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

$$A_n = \begin{bmatrix} 0.212 & 0.455 & 0.333 \\ 0.323 & 0.283 & 0.393 \\ 0.584 & 0.162 & 0.254 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.222 & 0.778 \\ 0.608 & 0.392 \\ 0.538 & 0.462 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.445 & 0.555 & 0.192 & 0.253 & 0.253 & 0.302 & 0.082 & 0.11 & 0.116 & 0.138 & 0.109 & 0.145 & 0.137 & 0.165 \\ 0.445 & 0.192 & 0.253 & 0.082 & 0.11 & 0.116 & 0.138 & 0.035 & 0.047 & 0.05 & 0.059 & 0.05 & 0.066 & 0.062 & 0.075 \\ 0.555 & 0.253 & 0.302 & 0.109 & 0.145 & 0.137 & 0.165 & 0.046 & 0.062 & 0.066 & 0.079 & 0.058 & 0.078 & 0.075 & 0.09 \\ 0.192 & 0.082 & 0.11 & 0.035 & 0.047 & 0.05 & 0.059 & 0.015 & 0.02 & 0.021 & 0.025 & 0.022 & 0.029 & 0.027 & 0.033 \\ 0.253 & 0.109 & 0.145 & 0.046 & 0.062 & 0.066 & 0.079 & 0.02 & 0.027 & 0.028 & 0.034 & 0.028 & 0.038 & 0.036 & 0.043 \\ 0.253 & 0.116 & 0.138 & 0.05 & 0.066 & 0.062 & 0.075 & 0.021 & 0.028 & 0.03 & 0.036 & 0.027 & 0.036 & 0.034 & 0.041 \\ 0.302 & 0.137 & 0.165 & 0.058 & 0.078 & 0.075 & 0.09 & 0.025 & 0.033 & 0.036 & 0.042 & 0.032 & 0.043 & 0.041 & 0.049 \\ 0.082 & 0.035 & 0.047 & 0.015 & 0.02 & 0.021 & 0.025 & 0.006 & 0.009 & 0.009 & 0.011 & 0.009 & 0.012 & 0.011 & 0.014 \\ 0.109 & 0.046 & 0.062 & 0.02 & 0.027 & 0.028 & 0.034 & 0.009 & 0.011 & 0.012 & 0.014 & 0.012 & 0.016 & 0.015 & 0.018 \\ 0.116 & 0.05 & 0.066 & 0.021 & 0.028 & 0.03 & 0.036 & 0.009 & 0.012 & 0.013 & 0.015 & 0.013 & 0.017 & 0.016 & 0.02 \\ 0.137 & 0.058 & 0.078 & 0.025 & 0.033 & 0.036 & 0.042 & 0.011 & 0.014 & 0.015 & 0.018 & 0.015 & 0.02 & 0.019 & 0.023 \\ 0.11 & 0.05 & 0.059 & 0.022 & 0.029 & 0.027 & 0.033 & 0.009 & 0.012 & 0.013 & 0.016 & 0.012 & 0.015 & 0.015 & 0.018 \\ 0.145 & 0.066 & 0.079 & 0.028 & 0.038 & 0.036 & 0.043 & 0.012 & 0.016 & 0.017 & 0.021 & 0.015 & 0.02 & 0.02 & 0.023 \\ 0.138 & 0.062 & 0.075 & 0.027 & 0.036 & 0.034 & 0.041 & 0.011 & 0.015 & 0.016 & 0.019 & 0.015 & 0.02 & 0.019 & 0.022 \\ 0.165 & 0.075 & 0.09 & 0.032 & 0.043 & 0.041 & 0.049 & 0.014 & 0.018 & 0.02 & 0.023 & 0.018 & 0.023 & 0.022 & 0.027 \end{bmatrix}$$

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.892 \\ 2.293 \times 10^{-2} \\ 1.749 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

### Simulation 14.

1. We simulated the same process and produced a sequence of  $T = 10^5$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.444 & 0.556 & 0.19 & 0.253 & 0.253 & 0.303 & 0.081 & 0.11 & 0.116 & 0.137 & 0.11 & 0.143 & 0.137 & 0.166 \\ 0.444 & 0.19 & 0.253 & 0.081 & 0.11 & 0.116 & 0.137 & 0.035 & 0.046 & 0.051 & 0.059 & 0.051 & 0.065 & 0.063 & 0.074 \\ 0.556 & 0.253 & 0.303 & 0.11 & 0.143 & 0.137 & 0.166 & 0.046 & 0.064 & 0.066 & 0.078 & 0.059 & 0.079 & 0.074 & 0.092 \\ 0.19 & 0.081 & 0.11 & 0.035 & 0.046 & 0.051 & 0.059 & 0.015 & 0.02 & 0.021 & 0.025 & 0.022 & 0.028 & 0.027 & 0.032 \\ 0.253 & 0.11 & 0.143 & 0.046 & 0.064 & 0.066 & 0.078 & 0.02 & 0.026 & 0.029 & 0.035 & 0.029 & 0.037 & 0.036 & 0.042 \\ 0.253 & 0.116 & 0.137 & 0.051 & 0.065 & 0.063 & 0.074 & 0.022 & 0.03 & 0.029 & 0.036 & 0.027 & 0.036 & 0.033 & 0.042 \\ 0.303 & 0.137 & 0.166 & 0.059 & 0.079 & 0.074 & 0.092 & 0.024 & 0.034 & 0.036 & 0.042 & 0.032 & 0.043 & 0.042 & 0.051 \\ 0.081 & 0.035 & 0.046 & 0.015 & 0.02 & 0.021 & 0.025 & 0.006 & 0.009 & 0.009 & 0.011 & 0.009 & 0.012 & 0.011 & 0.013 \\ 0.11 & 0.046 & 0.064 & 0.02 & 0.026 & 0.029 & 0.035 & 0.009 & 0.011 & 0.012 & 0.014 & 0.013 & 0.016 & 0.016 & 0.019 \\ 0.116 & 0.051 & 0.065 & 0.022 & 0.03 & 0.029 & 0.036 & 0.01 & 0.012 & 0.014 & 0.016 & 0.013 & 0.016 & 0.016 & 0.019 \\ 0.137 & 0.059 & 0.079 & 0.024 & 0.034 & 0.036 & 0.042 & 0.01 & 0.014 & 0.016 & 0.019 & 0.016 & 0.02 & 0.019 & 0.023 \\ 0.11 & 0.051 & 0.059 & 0.022 & 0.028 & 0.027 & 0.032 & 0.009 & 0.013 & 0.013 & 0.016 & 0.012 & 0.015 & 0.014 & 0.018 \\ 0.143 & 0.066 & 0.078 & 0.029 & 0.037 & 0.036 & 0.042 & 0.012 & 0.017 & 0.016 & 0.02 & 0.015 & 0.021 & 0.018 & 0.024 \\ 0.137 & 0.063 & 0.074 & 0.027 & 0.036 & 0.033 & 0.042 & 0.011 & 0.016 & 0.017 & 0.019 & 0.014 & 0.019 & 0.018 & 0.023 \\ 0.166 & 0.074 & 0.092 & 0.032 & 0.043 & 0.042 & 0.051 & 0.013 & 0.018 & 0.019 & 0.023 & 0.018 & 0.024 & 0.023 & 0.027 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

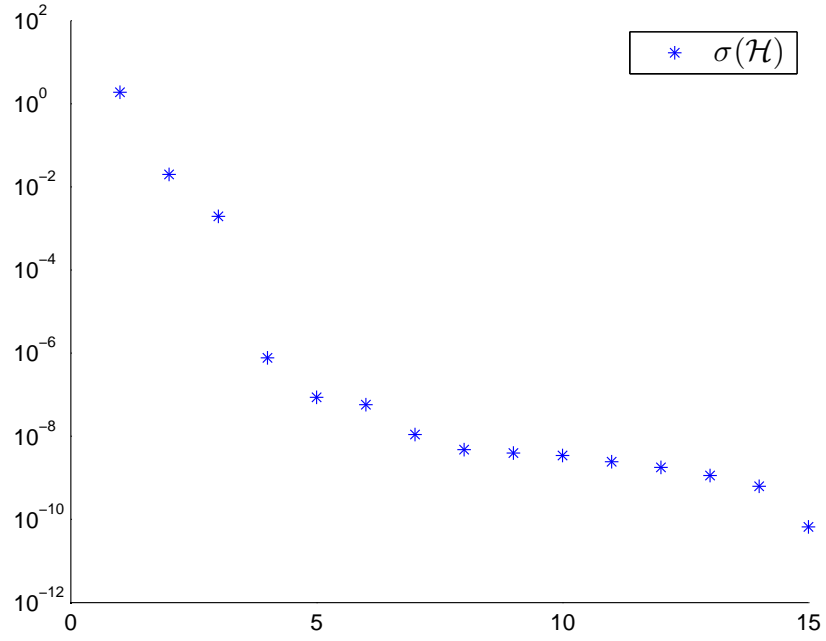
$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.892 \\ 2.319 \times 10^{-2} \\ 4.714 \times 10^{-3} \\ 2.633 \times 10^{-4} \\ 1.089 \times 10^{-4} \\ 8.642 \times 10^{-4} \\ 5.474 \times 10^{-4} \\ 3.045 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{B.3})$$

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

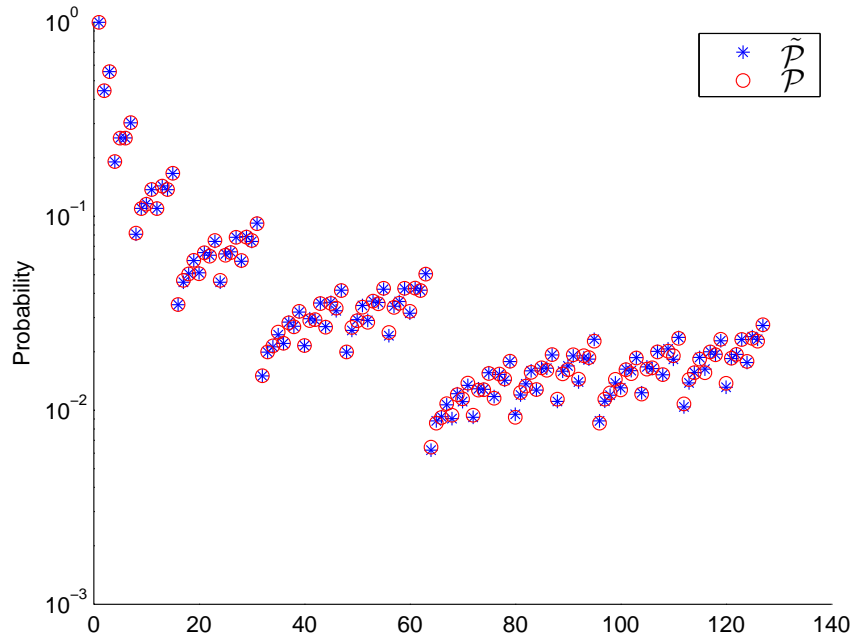
$$\mathcal{H} = \begin{bmatrix} 1.0 & 0.444 & 0.556 & 0.191 & 0.252 & 0.252 & 0.304 & 0.082 & 0.11 & 0.115 & 0.137 & 0.11 & 0.143 & 0.137 & 0.167 \\ 0.444 & 0.191 & 0.252 & 0.082 & 0.11 & 0.115 & 0.137 & 0.035 & 0.047 & 0.05 & 0.059 & 0.051 & 0.065 & 0.062 & 0.075 \\ 0.556 & 0.252 & 0.304 & 0.11 & 0.143 & 0.137 & 0.167 & 0.047 & 0.063 & 0.065 & 0.078 & 0.059 & 0.078 & 0.075 & 0.092 \\ 0.191 & 0.082 & 0.11 & 0.035 & 0.047 & 0.05 & 0.059 & 0.015 & 0.02 & 0.021 & 0.025 & 0.022 & 0.028 & 0.027 & 0.032 \\ 0.252 & 0.11 & 0.143 & 0.047 & 0.063 & 0.065 & 0.078 & 0.02 & 0.027 & 0.029 & 0.034 & 0.028 & 0.037 & 0.035 & 0.043 \\ 0.252 & 0.115 & 0.137 & 0.051 & 0.065 & 0.062 & 0.075 & 0.022 & 0.029 & 0.029 & 0.035 & 0.027 & 0.036 & 0.034 & 0.041 \\ 0.304 & 0.137 & 0.167 & 0.059 & 0.078 & 0.075 & 0.092 & 0.025 & 0.034 & 0.036 & 0.042 & 0.032 & 0.043 & 0.041 & 0.05 \\ 0.082 & 0.035 & 0.047 & 0.015 & 0.02 & 0.021 & 0.025 & 0.006 & 0.009 & 0.009 & 0.011 & 0.009 & 0.012 & 0.011 & 0.014 \\ 0.11 & 0.047 & 0.063 & 0.02 & 0.027 & 0.029 & 0.034 & 0.009 & 0.011 & 0.012 & 0.014 & 0.013 & 0.016 & 0.015 & 0.019 \\ 0.115 & 0.051 & 0.065 & 0.022 & 0.029 & 0.029 & 0.035 & 0.009 & 0.012 & 0.013 & 0.016 & 0.013 & 0.017 & 0.016 & 0.019 \\ 0.137 & 0.059 & 0.078 & 0.025 & 0.034 & 0.036 & 0.042 & 0.011 & 0.014 & 0.016 & 0.018 & 0.016 & 0.02 & 0.019 & 0.023 \\ 0.11 & 0.05 & 0.059 & 0.022 & 0.028 & 0.027 & 0.032 & 0.009 & 0.013 & 0.013 & 0.015 & 0.012 & 0.015 & 0.014 & 0.018 \\ 0.143 & 0.065 & 0.078 & 0.028 & 0.037 & 0.035 & 0.043 & 0.012 & 0.016 & 0.017 & 0.02 & 0.015 & 0.02 & 0.019 & 0.023 \\ 0.137 & 0.062 & 0.075 & 0.027 & 0.036 & 0.034 & 0.041 & 0.011 & 0.015 & 0.016 & 0.019 & 0.014 & 0.019 & 0.019 & 0.023 \\ 0.167 & 0.075 & 0.092 & 0.032 & 0.043 & 0.041 & 0.05 & 0.014 & 0.019 & 0.019 & 0.023 & 0.018 & 0.023 & 0.023 & 0.028 \end{bmatrix}$$

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.892 \\ 1.987 \times 10^{-2} \\ 1.947 \times 10^{-3} \\ 7.699 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure B.3.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^5$



**Figure B.4.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^5$



This suggests that  $\text{rank}_{est}(\mathcal{H}) = 3$ .

4. Estimate the matrices  $A_n$  and  $B_n$ .

$$A_n = \begin{bmatrix} 0.105 & 0.738 & 0.156 \\ 0.201 & 0.064 & 0.735 \\ 0.323 & 0.3 & 0.377 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.741 & 0.259 \\ 0.328 & 0.672 \\ 0.372 & 0.628 \end{bmatrix}.$$

5. Construct the  $\mathcal{H}_{new}$  corresponding to the  $A_{new}$  and  $B_{new}$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.444 & 0.556 & 0.191 & 0.252 & 0.252 & 0.304 & 0.083 & 0.109 & 0.114 & 0.138 & 0.109 & 0.144 & 0.138 & 0.165 \\ 0.444 & 0.191 & 0.252 & 0.083 & 0.109 & 0.114 & 0.138 & 0.036 & 0.047 & 0.049 & 0.06 & 0.049 & 0.065 & 0.063 & 0.075 \\ 0.556 & 0.252 & 0.304 & 0.109 & 0.144 & 0.138 & 0.165 & 0.047 & 0.062 & 0.065 & 0.079 & 0.059 & 0.079 & 0.075 & 0.09 \\ 0.191 & 0.083 & 0.109 & 0.036 & 0.047 & 0.049 & 0.06 & 0.015 & 0.02 & 0.021 & 0.026 & 0.021 & 0.028 & 0.027 & 0.032 \\ 0.252 & 0.109 & 0.144 & 0.047 & 0.062 & 0.065 & 0.079 & 0.02 & 0.027 & 0.028 & 0.034 & 0.028 & 0.037 & 0.036 & 0.043 \\ 0.252 & 0.114 & 0.138 & 0.049 & 0.065 & 0.063 & 0.075 & 0.021 & 0.028 & 0.029 & 0.036 & 0.027 & 0.036 & 0.034 & 0.041 \\ 0.304 & 0.138 & 0.165 & 0.059 & 0.079 & 0.075 & 0.09 & 0.026 & 0.034 & 0.036 & 0.043 & 0.032 & 0.043 & 0.041 & 0.049 \\ 0.083 & 0.036 & 0.047 & 0.015 & 0.02 & 0.021 & 0.026 & 0.007 & 0.009 & 0.009 & 0.011 & 0.009 & 0.012 & 0.012 & 0.014 \\ 0.109 & 0.047 & 0.062 & 0.02 & 0.027 & 0.028 & 0.034 & 0.009 & 0.012 & 0.012 & 0.015 & 0.012 & 0.016 & 0.015 & 0.018 \\ 0.114 & 0.049 & 0.065 & 0.021 & 0.028 & 0.029 & 0.036 & 0.009 & 0.012 & 0.013 & 0.015 & 0.013 & 0.017 & 0.016 & 0.019 \\ 0.138 & 0.059 & 0.079 & 0.026 & 0.034 & 0.036 & 0.043 & 0.011 & 0.015 & 0.015 & 0.018 & 0.015 & 0.02 & 0.02 & 0.024 \\ 0.109 & 0.049 & 0.06 & 0.021 & 0.028 & 0.027 & 0.032 & 0.009 & 0.012 & 0.013 & 0.015 & 0.012 & 0.016 & 0.015 & 0.018 \\ 0.144 & 0.065 & 0.079 & 0.028 & 0.037 & 0.036 & 0.043 & 0.012 & 0.016 & 0.017 & 0.02 & 0.015 & 0.021 & 0.019 & 0.023 \\ 0.138 & 0.063 & 0.075 & 0.027 & 0.036 & 0.034 & 0.041 & 0.012 & 0.015 & 0.016 & 0.02 & 0.015 & 0.019 & 0.019 & 0.022 \\ 0.165 & 0.075 & 0.09 & 0.032 & 0.043 & 0.041 & 0.049 & 0.014 & 0.018 & 0.019 & 0.024 & 0.018 & 0.023 & 0.022 & 0.027 \end{bmatrix}$$

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.892 \\ 2.001 \times 10^{-2} \\ 1.281 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

### Simulation 15.

1. We simulated the same process and produced a sequence of  $T = 10^4$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1.0 & 0.447 & 0.553 & 0.195 & 0.252 & 0.252 & 0.301 & 0.083 & 0.112 & 0.118 & 0.134 & 0.112 & 0.14 & 0.134 & 0.167 \\ 0.447 & 0.195 & 0.252 & 0.083 & 0.112 & 0.118 & 0.134 & 0.035 & 0.047 & 0.054 & 0.059 & 0.05 & 0.067 & 0.061 & 0.073 \\ 0.553 & 0.252 & 0.301 & 0.112 & 0.14 & 0.134 & 0.167 & 0.047 & 0.065 & 0.064 & 0.076 & 0.062 & 0.073 & 0.073 & 0.094 \\ 0.195 & 0.083 & 0.112 & 0.035 & 0.047 & 0.054 & 0.059 & 0.015 & 0.02 & 0.023 & 0.025 & 0.023 & 0.031 & 0.027 & 0.032 \\ 0.252 & 0.112 & 0.14 & 0.047 & 0.065 & 0.064 & 0.076 & 0.02 & 0.027 & 0.031 & 0.034 & 0.028 & 0.037 & 0.035 & 0.041 \\ 0.252 & 0.118 & 0.134 & 0.05 & 0.067 & 0.061 & 0.073 & 0.022 & 0.028 & 0.03 & 0.038 & 0.029 & 0.033 & 0.031 & 0.042 \\ 0.301 & 0.134 & 0.167 & 0.062 & 0.073 & 0.073 & 0.094 & 0.025 & 0.037 & 0.035 & 0.038 & 0.033 & 0.04 & 0.042 & 0.052 \\ 0.083 & 0.035 & 0.047 & 0.015 & 0.02 & 0.023 & 0.025 & 0.006 & 0.009 & 0.01 & 0.01 & 0.008 & 0.015 & 0.011 & 0.013 \\ 0.112 & 0.047 & 0.065 & 0.02 & 0.027 & 0.031 & 0.034 & 0.009 & 0.011 & 0.013 & 0.014 & 0.015 & 0.016 & 0.015 & 0.019 \\ 0.118 & 0.05 & 0.067 & 0.022 & 0.028 & 0.03 & 0.038 & 0.01 & 0.013 & 0.014 & 0.014 & 0.013 & 0.017 & 0.016 & 0.021 \\ 0.134 & 0.062 & 0.073 & 0.025 & 0.037 & 0.035 & 0.038 & 0.011 & 0.014 & 0.017 & 0.02 & 0.015 & 0.02 & 0.019 & 0.02 \\ 0.112 & 0.054 & 0.059 & 0.023 & 0.031 & 0.027 & 0.032 & 0.011 & 0.012 & 0.014 & 0.017 & 0.012 & 0.014 & 0.015 & 0.017 \\ 0.14 & 0.064 & 0.076 & 0.028 & 0.037 & 0.035 & 0.041 & 0.012 & 0.016 & 0.016 & 0.021 & 0.016 & 0.019 & 0.016 & 0.025 \\ 0.134 & 0.061 & 0.073 & 0.029 & 0.033 & 0.031 & 0.042 & 0.012 & 0.017 & 0.016 & 0.017 & 0.014 & 0.017 & 0.019 & 0.023 \\ 0.167 & 0.073 & 0.094 & 0.033 & 0.04 & 0.042 & 0.052 & 0.013 & 0.02 & 0.019 & 0.021 & 0.019 & 0.023 & 0.023 & 0.028 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

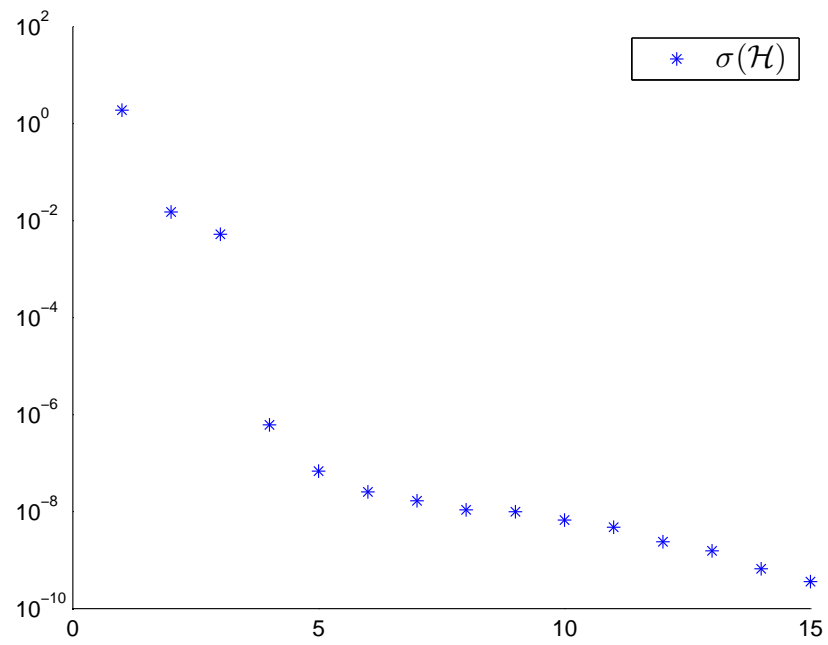
$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.891 \\ 2.313 \times 10^{-2} \\ 1.337 \times 10^{-2} \\ 5.993 \times 10^{-3} \\ 3.375 \times 10^{-3} \\ 3.154 \times 10^{-3} \\ 1.538 \times 10^{-3} \\ 5.733 \times 10^{-4} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{B.4})$$

3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

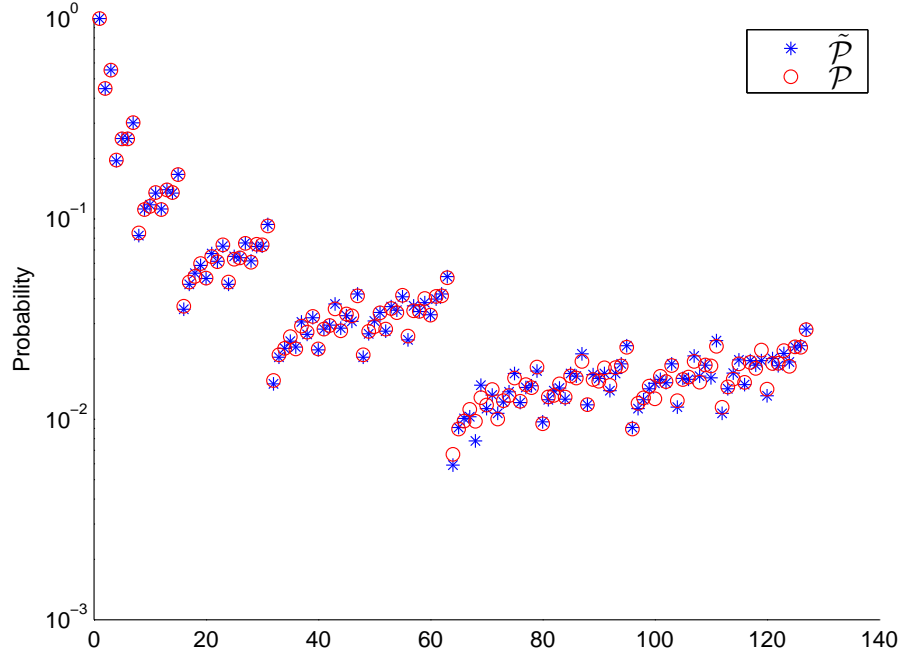
$$\mathcal{H} = \begin{bmatrix} 1 & 0.447 & 0.553 & 0.196 & 0.251 & 0.251 & 0.302 & 0.085 & 0.111 & 0.115 & 0.135 & 0.111 & 0.139 & 0.135 & 0.166 \\ 0.447 & 0.196 & 0.251 & 0.085 & 0.111 & 0.115 & 0.135 & 0.037 & 0.048 & 0.052 & 0.06 & 0.051 & 0.065 & 0.061 & 0.074 \\ 0.553 & 0.251 & 0.302 & 0.111 & 0.139 & 0.135 & 0.166 & 0.048 & 0.063 & 0.064 & 0.076 & 0.061 & 0.075 & 0.074 & 0.092 \\ 0.196 & 0.085 & 0.111 & 0.037 & 0.048 & 0.052 & 0.06 & 0.016 & 0.021 & 0.023 & 0.026 & 0.022 & 0.029 & 0.027 & 0.033 \\ 0.251 & 0.111 & 0.139 & 0.048 & 0.063 & 0.064 & 0.076 & 0.021 & 0.027 & 0.029 & 0.034 & 0.028 & 0.036 & 0.034 & 0.042 \\ 0.251 & 0.115 & 0.135 & 0.051 & 0.065 & 0.061 & 0.074 & 0.022 & 0.028 & 0.029 & 0.036 & 0.028 & 0.034 & 0.033 & 0.041 \\ 0.302 & 0.135 & 0.166 & 0.061 & 0.075 & 0.074 & 0.092 & 0.026 & 0.035 & 0.035 & 0.04 & 0.033 & 0.041 & 0.041 & 0.051 \\ 0.085 & 0.037 & 0.048 & 0.016 & 0.021 & 0.023 & 0.026 & 0.007 & 0.009 & 0.01 & 0.011 & 0.01 & 0.013 & 0.012 & 0.014 \\ 0.111 & 0.048 & 0.063 & 0.021 & 0.027 & 0.029 & 0.034 & 0.009 & 0.012 & 0.013 & 0.015 & 0.013 & 0.016 & 0.015 & 0.019 \\ 0.115 & 0.051 & 0.065 & 0.022 & 0.028 & 0.029 & 0.036 & 0.009 & 0.013 & 0.013 & 0.015 & 0.013 & 0.016 & 0.016 & 0.019 \\ 0.135 & 0.061 & 0.075 & 0.026 & 0.035 & 0.035 & 0.04 & 0.011 & 0.015 & 0.016 & 0.019 & 0.015 & 0.019 & 0.018 & 0.022 \\ 0.111 & 0.052 & 0.06 & 0.022 & 0.029 & 0.027 & 0.033 & 0.01 & 0.012 & 0.013 & 0.016 & 0.012 & 0.015 & 0.014 & 0.018 \\ 0.139 & 0.064 & 0.076 & 0.028 & 0.036 & 0.034 & 0.042 & 0.012 & 0.016 & 0.016 & 0.019 & 0.015 & 0.019 & 0.018 & 0.023 \\ 0.135 & 0.061 & 0.074 & 0.028 & 0.034 & 0.033 & 0.041 & 0.012 & 0.016 & 0.015 & 0.018 & 0.015 & 0.018 & 0.018 & 0.023 \\ 0.166 & 0.074 & 0.092 & 0.033 & 0.041 & 0.041 & 0.051 & 0.014 & 0.019 & 0.019 & 0.022 & 0.018 & 0.023 & 0.023 & 0.028 \end{bmatrix}$$

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.890 \\ 1.504 \times 10^{-2} \\ 5.199 \times 10^{-3} \\ 6.130 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure B.5.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^4$



**Figure B.6.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^4$

This suggests that  $rank_{est}(\mathcal{H}) = 3$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

$$A_n = \begin{bmatrix} 0.133 & 0.468 & 0.399 \\ 0.408 & 0.363 & 0.229 \\ 0.816 & 0.063 & 0.121 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.364 & 0.636 \\ 0.251 & 0.749 \\ 0.796 & 0.204 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.447 & 0.553 & 0.196 & 0.251 & 0.251 & 0.302 & 0.087 & 0.11 & 0.114 & 0.137 & 0.108 & 0.143 & 0.133 & 0.169 \\ 0.447 & 0.196 & 0.251 & 0.087 & 0.11 & 0.114 & 0.137 & 0.038 & 0.049 & 0.05 & 0.06 & 0.049 & 0.065 & 0.06 & 0.076 \\ 0.553 & 0.251 & 0.302 & 0.108 & 0.143 & 0.133 & 0.169 & 0.048 & 0.06 & 0.066 & 0.077 & 0.058 & 0.076 & 0.075 & 0.093 \\ 0.196 & 0.087 & 0.11 & 0.038 & 0.049 & 0.05 & 0.06 & 0.017 & 0.021 & 0.022 & 0.026 & 0.021 & 0.028 & 0.026 & 0.033 \\ 0.251 & 0.108 & 0.143 & 0.048 & 0.06 & 0.066 & 0.077 & 0.021 & 0.027 & 0.027 & 0.033 & 0.028 & 0.038 & 0.034 & 0.043 \\ 0.251 & 0.114 & 0.137 & 0.049 & 0.065 & 0.06 & 0.076 & 0.022 & 0.027 & 0.03 & 0.035 & 0.026 & 0.034 & 0.034 & 0.042 \\ 0.302 & 0.133 & 0.169 & 0.058 & 0.076 & 0.075 & 0.093 & 0.026 & 0.032 & 0.035 & 0.041 & 0.032 & 0.043 & 0.041 & 0.052 \\ 0.087 & 0.038 & 0.049 & 0.017 & 0.021 & 0.022 & 0.026 & 0.007 & 0.009 & 0.01 & 0.012 & 0.009 & 0.013 & 0.012 & 0.015 \\ 0.108 & 0.048 & 0.06 & 0.021 & 0.027 & 0.027 & 0.033 & 0.009 & 0.012 & 0.012 & 0.015 & 0.012 & 0.015 & 0.014 & 0.018 \\ 0.114 & 0.049 & 0.065 & 0.022 & 0.027 & 0.03 & 0.035 & 0.01 & 0.012 & 0.012 & 0.015 & 0.013 & 0.017 & 0.015 & 0.02 \\ 0.133 & 0.058 & 0.076 & 0.026 & 0.032 & 0.035 & 0.041 & 0.011 & 0.014 & 0.015 & 0.018 & 0.015 & 0.02 & 0.018 & 0.023 \\ 0.11 & 0.05 & 0.06 & 0.021 & 0.028 & 0.026 & 0.033 & 0.01 & 0.012 & 0.013 & 0.015 & 0.011 & 0.015 & 0.015 & 0.019 \\ 0.143 & 0.066 & 0.077 & 0.028 & 0.038 & 0.034 & 0.043 & 0.013 & 0.016 & 0.017 & 0.02 & 0.015 & 0.019 & 0.019 & 0.024 \\ 0.137 & 0.06 & 0.076 & 0.026 & 0.034 & 0.034 & 0.042 & 0.012 & 0.015 & 0.016 & 0.019 & 0.015 & 0.019 & 0.019 & 0.024 \\ 0.169 & 0.075 & 0.093 & 0.032 & 0.043 & 0.041 & 0.052 & 0.014 & 0.018 & 0.02 & 0.023 & 0.018 & 0.024 & 0.023 & 0.029 \end{bmatrix}$$

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.891 \\ 1.695 \times 10^{-2} \\ 5.708 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

### Simulation 16.

1. We simulated the same process and produced a sequence of  $T = 10^3$  observations  $\mathcal{O}_{1:T}$ .
2. Compute  $\tilde{\mathcal{P}}$  from the data  $\mathcal{O}$  and build  $\tilde{\mathcal{H}}$  from  $\tilde{\mathcal{P}}$

$$\tilde{\mathcal{H}} = \begin{bmatrix} 1 & 0.427 & 0.573 & 0.174 & 0.253 & 0.252 & 0.321 & 0.074 & 0.099 & 0.103 & 0.151 & 0.099 & 0.153 & 0.15 & 0.171 \\ 0.427 & 0.174 & 0.253 & 0.074 & 0.099 & 0.103 & 0.151 & 0.027 & 0.047 & 0.042 & 0.057 & 0.039 & 0.063 & 0.078 & 0.072 \\ 0.573 & 0.252 & 0.321 & 0.099 & 0.153 & 0.15 & 0.171 & 0.047 & 0.052 & 0.06 & 0.092 & 0.06 & 0.089 & 0.072 & 0.098 \\ 0.174 & 0.074 & 0.099 & 0.027 & 0.047 & 0.042 & 0.057 & 0.009 & 0.018 & 0.019 & 0.028 & 0.017 & 0.025 & 0.027 & 0.03 \\ 0.252 & 0.099 & 0.153 & 0.047 & 0.052 & 0.06 & 0.092 & 0.018 & 0.029 & 0.023 & 0.029 & 0.022 & 0.038 & 0.051 & 0.041 \\ 0.253 & 0.103 & 0.151 & 0.039 & 0.063 & 0.078 & 0.072 & 0.017 & 0.022 & 0.02 & 0.043 & 0.035 & 0.043 & 0.029 & 0.043 \\ 0.321 & 0.15 & 0.171 & 0.06 & 0.089 & 0.072 & 0.098 & 0.03 & 0.03 & 0.04 & 0.049 & 0.026 & 0.046 & 0.043 & 0.055 \\ 0.074 & 0.027 & 0.047 & 0.009 & 0.018 & 0.019 & 0.028 & 0.004 & 0.005 & 0.005 & 0.013 & 0.006 & 0.013 & 0.014 & 0.014 \\ 0.099 & 0.047 & 0.052 & 0.018 & 0.029 & 0.023 & 0.029 & 0.005 & 0.013 & 0.014 & 0.015 & 0.011 & 0.012 & 0.013 & 0.016 \\ 0.103 & 0.039 & 0.063 & 0.017 & 0.022 & 0.02 & 0.043 & 0.005 & 0.012 & 0.013 & 0.009 & 0.007 & 0.013 & 0.022 & 0.021 \\ 0.15 & 0.06 & 0.089 & 0.03 & 0.03 & 0.04 & 0.049 & 0.013 & 0.017 & 0.01 & 0.02 & 0.015 & 0.025 & 0.029 & 0.02 \\ 0.099 & 0.042 & 0.057 & 0.017 & 0.025 & 0.027 & 0.03 & 0.008 & 0.009 & 0.007 & 0.018 & 0.011 & 0.016 & 0.012 & 0.018 \\ 0.153 & 0.06 & 0.092 & 0.022 & 0.038 & 0.051 & 0.041 & 0.009 & 0.013 & 0.013 & 0.025 & 0.024 & 0.027 & 0.017 & 0.024 \\ 0.151 & 0.078 & 0.072 & 0.035 & 0.043 & 0.029 & 0.043 & 0.018 & 0.017 & 0.015 & 0.028 & 0.012 & 0.017 & 0.019 & 0.024 \\ 0.171 & 0.072 & 0.098 & 0.026 & 0.046 & 0.043 & 0.055 & 0.013 & 0.013 & 0.025 & 0.021 & 0.014 & 0.029 & 0.024 & 0.031 \end{bmatrix}$$

The singular values of the  $\tilde{\mathcal{H}}$  is given by;

$$\sigma(\tilde{\mathcal{H}}) = \begin{bmatrix} 1.905 \\ 5.625 \times 10^{-2} \\ 4.285 \times 10^{-2} \\ 1.926 \times 10^{-2} \\ 1.142 \times 10^{-2} \\ 9.326 \times 10^{-3} \\ 5.585 \times 10^{-3} \\ 1.562 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (\text{B.5})$$

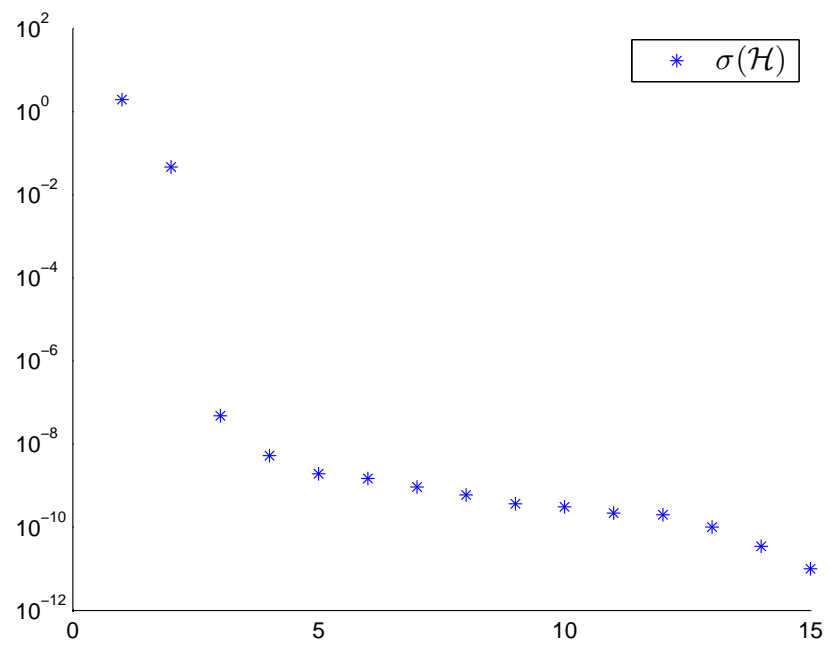
3. Solve the optimization problem 3.1. and obtain optimized Hankel matrix  $\mathcal{H}$ .

$$\mathcal{H} = \begin{bmatrix} 1 & 0.431 & 0.569 & 0.181 & 0.249 & 0.249 & 0.32 & 0.078 & 0.103 & 0.107 & 0.142 & 0.103 & 0.146 & 0.142 & 0.178 \\ 0.431 & 0.181 & 0.249 & 0.078 & 0.103 & 0.107 & 0.142 & 0.033 & 0.045 & 0.045 & 0.058 & 0.044 & 0.063 & 0.064 & 0.078 \\ 0.569 & 0.249 & 0.32 & 0.103 & 0.146 & 0.142 & 0.178 & 0.045 & 0.058 & 0.062 & 0.084 & 0.059 & 0.083 & 0.078 & 0.1 \\ 0.181 & 0.078 & 0.103 & 0.033 & 0.045 & 0.045 & 0.058 & 0.014 & 0.019 & 0.019 & 0.026 & 0.019 & 0.026 & 0.026 & 0.032 \\ 0.249 & 0.103 & 0.146 & 0.045 & 0.058 & 0.062 & 0.084 & 0.019 & 0.026 & 0.026 & 0.033 & 0.026 & 0.036 & 0.038 & 0.045 \\ 0.249 & 0.107 & 0.142 & 0.044 & 0.063 & 0.064 & 0.078 & 0.02 & 0.025 & 0.026 & 0.037 & 0.027 & 0.038 & 0.034 & 0.044 \\ 0.32 & 0.142 & 0.178 & 0.059 & 0.083 & 0.078 & 0.1 & 0.025 & 0.034 & 0.036 & 0.047 & 0.033 & 0.045 & 0.044 & 0.056 \\ 0.078 & 0.033 & 0.045 & 0.014 & 0.019 & 0.019 & 0.026 & 0.006 & 0.008 & 0.008 & 0.011 & 0.008 & 0.011 & 0.012 & 0.014 \\ 0.103 & 0.045 & 0.058 & 0.019 & 0.026 & 0.026 & 0.033 & 0.008 & 0.011 & 0.011 & 0.015 & 0.011 & 0.015 & 0.014 & 0.018 \\ 0.107 & 0.044 & 0.063 & 0.02 & 0.025 & 0.026 & 0.037 & 0.008 & 0.011 & 0.011 & 0.014 & 0.011 & 0.015 & 0.017 & 0.02 \\ 0.142 & 0.059 & 0.083 & 0.025 & 0.034 & 0.036 & 0.047 & 0.011 & 0.015 & 0.015 & 0.019 & 0.015 & 0.021 & 0.021 & 0.026 \\ 0.103 & 0.045 & 0.058 & 0.019 & 0.026 & 0.026 & 0.032 & 0.008 & 0.011 & 0.011 & 0.015 & 0.011 & 0.015 & 0.014 & 0.018 \\ 0.146 & 0.062 & 0.084 & 0.026 & 0.036 & 0.038 & 0.045 & 0.011 & 0.014 & 0.015 & 0.021 & 0.016 & 0.023 & 0.02 & 0.026 \\ 0.142 & 0.064 & 0.078 & 0.027 & 0.038 & 0.034 & 0.044 & 0.011 & 0.015 & 0.016 & 0.021 & 0.014 & 0.02 & 0.019 & 0.025 \\ 0.178 & 0.078 & 0.1 & 0.033 & 0.045 & 0.044 & 0.056 & 0.014 & 0.019 & 0.02 & 0.026 & 0.018 & 0.026 & 0.025 & 0.031 \end{bmatrix}$$

4. Estimate  $N$ . Analyze singular values of  $\mathcal{H}$  and the difference between

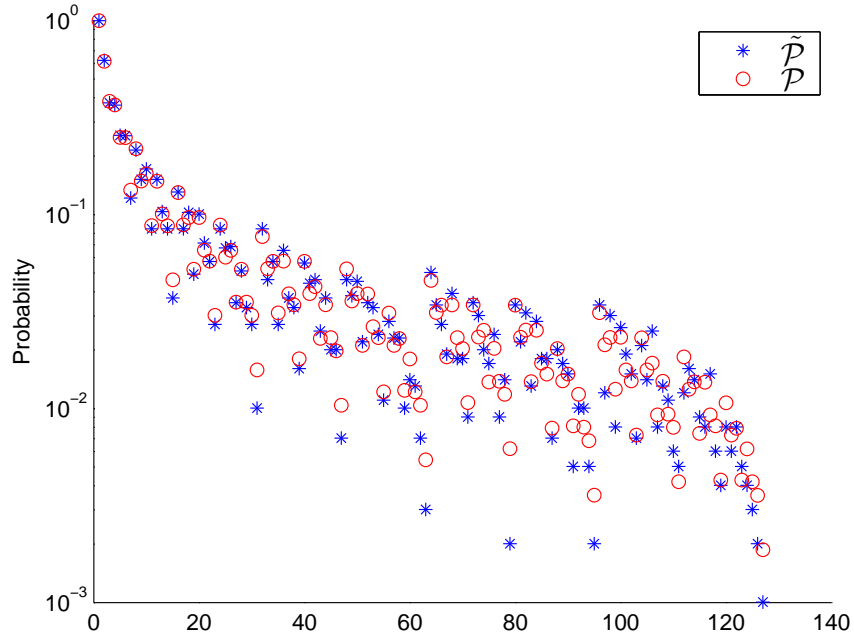
approximate  $\tilde{\mathcal{P}}$  and estimated probabilities  $\mathcal{P}$ .

$$\sigma(\mathcal{H}) = \begin{bmatrix} 1.901 \\ 1.961 \times 10^{-2} \\ 6.441 \times 10^{-3} \\ 4.498 \times 10^{-7} \\ 3.948 \times 10^{-7} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$



**Figure B.7.** Singular values of  $\mathcal{H}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^3$





**Figure B.8.** The difference between  $\tilde{\mathcal{P}}$  and  $\mathcal{P}$  for  $t = 3$ ,  $N = 3$ ,  $L_{seq} = 10^3$

This suggests that  $rank_{est}(\mathcal{H}) = 3$ .

5. Estimate the matrices  $A_n$  and  $B_n$ .

$$A_n = \begin{bmatrix} 0.334 & 0.498 & 0.169 \\ 0.442 & 0.223 & 0.335 \\ 0.309 & 0.444 & 0.247 \end{bmatrix}, \quad B_n = \begin{bmatrix} 0.364 & 0.636 \\ 0.247 & 0.753 \\ 0.786 & 0.214 \end{bmatrix}.$$

6. Construct the  $\mathcal{H}_n$  corresponding to the  $A_n$  and  $B_n$

$$\mathcal{H}_n = \begin{bmatrix} 1 & 0.431 & 0.569 & 0.181 & 0.249 & 0.249 & 0.32 & 0.076 & 0.105 & 0.108 & 0.141 & 0.103 & 0.146 & 0.137 & 0.183 \\ 0.431 & 0.181 & 0.249 & 0.076 & 0.105 & 0.108 & 0.141 & 0.032 & 0.044 & 0.046 & 0.06 & 0.045 & 0.063 & 0.06 & 0.081 \\ 0.569 & 0.249 & 0.32 & 0.103 & 0.146 & 0.137 & 0.183 & 0.043 & 0.06 & 0.063 & 0.082 & 0.057 & 0.08 & 0.079 & 0.105 \\ 0.181 & 0.076 & 0.105 & 0.032 & 0.044 & 0.046 & 0.06 & 0.013 & 0.018 & 0.019 & 0.025 & 0.019 & 0.027 & 0.026 & 0.034 \\ 0.249 & 0.103 & 0.146 & 0.043 & 0.06 & 0.063 & 0.082 & 0.018 & 0.025 & 0.026 & 0.034 & 0.026 & 0.037 & 0.035 & 0.047 \\ 0.249 & 0.108 & 0.141 & 0.045 & 0.063 & 0.06 & 0.081 & 0.019 & 0.026 & 0.028 & 0.036 & 0.025 & 0.035 & 0.035 & 0.046 \\ 0.32 & 0.137 & 0.183 & 0.057 & 0.08 & 0.079 & 0.105 & 0.024 & 0.033 & 0.035 & 0.045 & 0.033 & 0.046 & 0.045 & 0.06 \\ 0.076 & 0.032 & 0.044 & 0.013 & 0.018 & 0.019 & 0.025 & 0.006 & 0.008 & 0.008 & 0.01 & 0.008 & 0.011 & 0.011 & 0.014 \\ 0.103 & 0.043 & 0.06 & 0.018 & 0.025 & 0.026 & 0.034 & 0.008 & 0.011 & 0.011 & 0.014 & 0.011 & 0.015 & 0.015 & 0.019 \\ 0.108 & 0.045 & 0.063 & 0.019 & 0.026 & 0.028 & 0.036 & 0.008 & 0.011 & 0.011 & 0.015 & 0.011 & 0.016 & 0.015 & 0.02 \\ 0.137 & 0.057 & 0.08 & 0.024 & 0.033 & 0.035 & 0.045 & 0.01 & 0.014 & 0.014 & 0.019 & 0.014 & 0.02 & 0.019 & 0.026 \\ 0.105 & 0.046 & 0.06 & 0.019 & 0.027 & 0.026 & 0.034 & 0.008 & 0.011 & 0.012 & 0.015 & 0.011 & 0.015 & 0.015 & 0.019 \\ 0.146 & 0.063 & 0.082 & 0.026 & 0.037 & 0.035 & 0.047 & 0.011 & 0.015 & 0.016 & 0.021 & 0.015 & 0.021 & 0.02 & 0.027 \\ 0.141 & 0.06 & 0.081 & 0.025 & 0.035 & 0.035 & 0.046 & 0.011 & 0.015 & 0.015 & 0.02 & 0.014 & 0.02 & 0.02 & 0.026 \\ 0.183 & 0.079 & 0.105 & 0.033 & 0.046 & 0.045 & 0.06 & 0.014 & 0.019 & 0.02 & 0.026 & 0.019 & 0.026 & 0.026 & 0.034 \end{bmatrix}$$

$$\sigma(\mathcal{H}_n) = \begin{bmatrix} 1.902 \\ 1.493 \times 10^{-2} \\ 5.935 \times 10^{-3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Even though the system matrices of reconstructed model  $A_n, B_n$  is diverges from the original system matrices  $A$  and  $B$ , the Hankel matrix of the reconstructed model  $\mathcal{H}_n$  and  $\mathcal{H}$  are really close.

# Bibliography

- [1] BLACKWELL, D. and L. KOOPMANS (1957) “On the identifiability problem for functions of finite Markov chains,” *The Annals of Mathematical Statistics*, pp. 1011–1015.
- [2] BAUM, L. E. (1972) “An equality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes,” *Inequalities*, **3**, pp. 1–8.
- [3] BALDI, P. and Y. CHAUVIN (1994) “Smooth on-line learning algorithms for hidden Markov models,” *Neural Computation*, **6**(2), pp. 307–318.
- [4] VANLUYTEN, B. (2008) *Realization, identification and filtering for hidden Markov models using matrix factorization techniques*, Ph.D. thesis, PhD Thesis.
- [5] GILBERT, E. J. (1959) “On the identifiability problem for functions of finite Markov chains,” *The Annals of Mathematical Statistics*, pp. 688–697.
- [6] FOX, M. and H. RUBIN (1968) “Functions of processes with Markovian states,” *The Annals of Mathematical Statistics*, pp. 938–946.
- [7] DHARMADHIKARI, S. (1963) “Sufficient conditions for a stationary process to be a function of a finite Markov chain,” *The Annals of Mathematical Statistics*, pp. 1033–1041.
- [8] HELLER, A. (1965) “On stochastic processes derived from Markov chains,” *The Annals of Mathematical Statistics*, pp. 1286–1291.
- [9] ANDERSON, B. D. O. (1999) “Mathematics of Control , Signals , and Systems The Realization Problem for Hidden Markov Models \*,” **120**, pp. 80–120.
- [10] VIDYASAGAR, M. (2006) “A realization theory for hidden Markov models: the partial realization problem,” in *Proc. of the 17th International Symposium on Mathematical Theory of Networks and Systems*, pp. 2145–2150.

- [11] FINESSO, L. and P. SPREIJ (2006) “Nonnegative matrix factorization and I-divergence alternating minimization,” *Linear Algebra and its Applications*, **416**(2), pp. 270–287.
- [12] PICCI, G. (1978) “On the internal structure of finite-state stochastic processes,” in *Recent Developments in Variable Structure Systems, Economics and Biology*, Springer, pp. 288–304.
- [13] VIDYASAGAR, M. (2005) “The realization problem for hidden Markov models: The complete realization problem,” in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC’05. 44th IEEE Conference on*, IEEE, pp. 6632–6637.
- [14] FINESSO, L., A. GRASSI, and P. SPREIJ (2006) “Approximation of stationary processes by hidden Markov models,” *arXiv preprint math/0606591*.
- [15] ROSS, S. M. (2014) *Introduction to probability models*, Academic press.
- [16] ELLIOTT, R. J., L. AGGOUN, and J. B. MOORE (1994) *Hidden Markov Models*, Springer.
- [17] BAUM, L. E. and T. PETRIE (1966) “Statistical inference for probabilistic functions of finite state Markov chains,” *The annals of mathematical statistics*, pp. 1554–1563.
- [18] DAVID, J. (1994) “Algorithms for Analysis, Design of Robust Controllers,” *status: published*.
- [19] VANLUYTEN, B., J. C. WILLEMS, and B. DE MOOR (2008) “Structured non-negative matrix factorization with applications to hidden Markov realization and clustering,” *Linear Algebra and its applications*, **429**(7), pp. 1409–1424.
- [20] VANDENBERGHE, L. and S. BOYD (1996) “Semidefinite programming,” *SIAM review*, **38**(1), pp. 49–95.
- [21] RECHT, B., M. FAZEL, and P. A. PARRILO (2010) “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM review*, **52**(3), pp. 471–501.