Estimating the Generator Matrix from Discret Data Set

Romayssa Touaimia (Algeria)
Department Probability and Statistics, Mathematical Institute
Eötvös Loránd University, Budapest, Hungary

A thesis submitted for the degree of
MSc in Mathematics

2019. May

Declaration

I herewith declare that I have produced this thesis without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other Hungarian or foreign examination board.

The thesis work was conducted under the supervision of Mr. Pröhle Tamás at Eötvös Loránd University.

Budapest, May, 2019

Acknowledgements

First and foremost, I would like to thank my supervision and advisor Mr. Prőhle Tamás for introducing me to this topic, for his help through the difficulties that I faced, his patience with me, his helpful suggestions and his assistance in every step through the process. I am grateful for the many discussions that we had.

I would like to thank my parents whom were always there for me, supported me and believed in me.

Abstract

Continuous Markov chain is determined by its transition rates which are the entries of the generator matrix. This generator matrix is used to fully describe the process.

In this work, we will first start with an introduction about the importance of the application of the Markov Chains in real life, then we will give some definitions that help us to understand the discrete and continuous time Markov chains, then we will see the relation between transition matrix and the generator matrix.

We will discuss the embedding problem in several cases, i.e. we will study under what conditions a true generator does and does not exist for a transition matrix and its uniqueness and how can wee choose the correct one if its not unique.

We will study the estimation of generator matrix of a continuous Markov chain form discrete data observation and see the relation between it and the embedding problem. If the empirical transition matrix is embeddable we will see how can we obtain the estimation of the generator matrix. If the empirical transition matrix is not embeddable we will give some methods to deal with this problem, then we will have an application on the financial time series using the R project.

Contents

1

Introduction

Markov chain is a stochastic process that was named after Andrey Andrei Markov. This process is based on the memory-less property which is named as Markov property, meaning that the past and future are independent when present in known. Markov chains can be applied in different disciplines, such as queues, games inventory, speech levels, biology, physics, biochemistry, and finance, etc.
We use the Markov chain in modeling different processes since it makes the analysis of the models much easier.
There are two types of Markov chains; discrete and continuous time.

The continuous-time Markov chain with a finite set of states has many applications in real life. In finance analyzing the financial time series using continuous Markov chain can help us to understand the change of the prices and predict the future of the financial market trends as well as the risks associated with them.

The estimation of the continuous Markov process parameters when this process has been observed continuously in some interval, It is simple and well known. On the other hand, the estimation of these parameters from discrete time observations which is closely related to the embedding problem is not an easy task. This later estimation problem was considered as an important statistical problem in a wide range of applications, such as the analysis of gene sequence data; describing the dynamics of open quantum systems in physics; causal interference in epidemiology; or in rating based credit risk modeling.

The main difficulties faced by when solving this problem are the existence and uniqueness of these estimators.

2

Discrete and Continuous time Markov Chains

In mathematics stochastic process is knowing as the study how random variables evolving over time.

Definition 1 (Conditional probability)
Let A, B be two events from the sample space S, we denote the conditional probability of the event A given B by $P(A|B)$ where:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{2.1}$$

Using the equation (2.1) and for A, B and C being an events from the sample space S we get that:

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{P(A \cap B \cap C)}{P(B \cap C)} \cdot \frac{P(B \cap C)}{P(C)},$$

this implies:

$$P(A \cap B|C) = P(A|B \cap C) \cdot P(B|C). \tag{2.2}$$

Definition 2 (partition of the sample space)
We say that the set of events $A_1, A_2, \ldots, A_n$ is a partition of the sample space S if:
1 - $A_i \cap A_j = \emptyset, \quad i \neq j, \quad i, j = 1, 2, \ldots, n$
2 - $A_1 \cup A_2 \cup \ldots \cup A_n = S$
3 - $P(A_i) > 0$ *for any i=1,2,...,n.*

Theorem 1 (Law of Total Probability)
If $A_1, A_2, \ldots, A_n$ is a partition of the sample space S then for any event B we have that

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i). \tag{2.3}$$

Definition 3 (stochastic process)
The collection of random variables $\{X(t), t \geq 0\}$ indexed by time t is stochastic process and we say that:

- $\{X(t), t \geq 0\}$ *is discrete time stochastic process if* $t \in \{0, 1, 2, \ldots\} \equiv \mathbb{N}_0$.
- $\{X(t), t \geq 0\}$ *is continuous time stochastic process if* $t \in [0, \infty)$.

Definition 4
A state of space is the set of the real values that a random variable can take, we denote it by S.

- *If S consists of a countable and finite number of states then S is a discrete finite states pace.*

2.1 Discrete-time Markov Chains

Definition 5 (the discrete time Markov chain)
The discrete time stochastic process $\{X(n), n \in N_0\}$ over the state space $S = \{1, 2, \ldots, \ell\}$ is a discrete time Markov chain if the conditional probability distribution of $X(n+1)$ given $X(n), \ldots, X(0)$ depends only on $X(n)$:

$$\begin{gathered} P(X(n+1) = j | X(n) = i, X(n-1) = i_{n-1}, \ldots, X(0) = i_0) = \\ = P(X(n+1) = j | X(n) = i), \\ \textit{for any } n \in \{0, 1, 2, \ldots\} \textit{ and } j, i, i_{n-1}, \ldots, i_0 \in S \end{gathered}$$

And this is what we call the Markov property or memory-less property.

- *We say that the chain is in state i at time n if $X(n)=i$.*

Time homogenous discrete Markov chain:
We say that a discrete Markov chain is time homogenous if $P(X(n+1) = j | X(n) = i)$ is independent from the time n.

In this section, we consider that the discrete Markov chain $\{X(n), n \in N_0\}$ over the discrete finite state space $S = \{1, 2, \ldots, \ell\}$ is time-homogenous chain.

2.1.1 Transition matrix

Definition 6 (transition probability)
The process moves successively after one time unit from a state to another state with

probability called by the transition probability and each move is called a step.
The chain moves with one step from state i to state j with transition probability denoted by p_{ij}, where $p_{ij} = P(X(n+1) = j|X(n) = i)$.
Since our chain is time homogenous in particular, we have:

$$\begin{aligned} p_{ij} &= P(X(1) = j|X(0) = i) \\ &= P(X(2) = j|X(1) = i) \\ &= P(X(3) = j|X(2) = i) \\ &= \ldots \end{aligned}$$

- *p_{ij} for $j, i \in S$ are called the one-step transition probabilities.*

If the chain is on state i the probability that the process will be on state j after m steps are denoted by $p_{ij}^{(m)}$, where $p_{ij}^{(m)} = P(X(n+m) = j|X(n) = i)$.
And again since it is a time homogenous chain in particular, we have:

$$\begin{aligned} p_{ij}^{m} &= P(X(n+m) = j|X(n) = i) \\ &= P(X(m) = j|X(0) = i) \\ &= P(1+m) = j|X(1) = i) \\ &= P(X(2+m) = j|X(2) = i) \\ &= \ldots \end{aligned}$$

- *$p_{ij}^{(m)}$ for $i, j \in S$ are called the m-step transition probabilities.*

Definition 7 (m-step transition matrix)
The m-step transition matrix for $m \in \{1, 2, \ldots\}$ is the $\ell \times \ell$ stochastic matrix such that its entries are the m-step transition probabilities, i.e. it satisfies:
1 - $0 \leq p_{ij}^{(m)} \leq 1$, $i \neq j$, $\forall i, j \in S$
2 - $\sum_{j=1}^{\ell} p_{ij}^{(m)} = 1$, $\forall i \in S$.
and we denote it by $P^{(m)}$, hence we have that $P^{(m)} = \left[p_{ij}^{(m)}\right]_{i,j \in S}$

$$P^{(m)} = \begin{bmatrix} p_{11}^{(m)} & p_{12}^{(m)} & \cdots & p_{1\ell}^{(m)} \\ p_{21}^{(m)} & p_{22}^{(m)} & \cdots & p_{2\ell}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\ell 1}^{(m)} & p_{\ell 2}^{(m)} & \cdots & p_{\ell\ell}^{(m)} \end{bmatrix}$$

Definition 8 (The transition matrix)
The transition matrix is one-step transition matrix (m=1), such that its entries are the

transition probabilities and we denote it by P*, hence we have that* $P = P^{(1)} = \left[p_{ij}\right]_{i,j\in S}$

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1\ell} \\ p_{21} & p_{2,2} & \cdots & p_{2\ell} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\ell 1} & p_{\ell 2} & \cdots & p_{\ell\ell} \end{bmatrix}$$

Remark 1 $P^0 = I$ *here* I *is the* $\ell \times \ell$ *identity matrix.*

The Chapman-Kolmogorov equation
We have the (n+m)-probability distribution

$$p_{ij}^{(n+m)} = P(X(n+m) = j|X(0) = i).$$

Using the equations (2.2) and (2.3), we get

$$\begin{aligned} p_{i,j}^{(n+m)} &= \sum_{k=1}^{\ell} P(X(n+m) = j, X(n) = k|X(0) = i) \\ &= \sum_{k=1}^{\ell} P(X(n+m) = j|X(n) = k, X(0) = i)P(X(n) = k|X(0) = i) \end{aligned}$$

And now using the the Markov property implies that:

$$p_{i,j}^{(n+m)} = \sum_{k=1}^{\ell} P(X(n+m) = j|X(n) = k)P(X(n) = k|X(0) = i),$$

finally we obtain:

$$p_{i,j}^{(n+m)} = \sum_{k=1}^{\ell} p_{i,k}^{(m)} p_{i,j}^{(n)} = \sum_{k=1}^{\ell} p_{k,j}^{(n)} p_{i,j}^{(m)}, \tag{2.4}$$

the equation (2.4) is the the Chapman-Kolmogorov equation.
Now from the equation (2.4) we can see that $P^{(m)} = P^{(1)}P^{(1)} \ldots P^{(1)} = P{\cdot}P \ldots P = P^m$ which means that the m-step transition matrix $P^{(m)}$ is equal to P multiplied by it self m times P^m .

2.1.2 The probability distribution

Definition 9
The state probability distribution $\pi_i(n)$ *is the probability of being in state* i *at time* n*.*

$$\pi_i(n) = P(X(n) = i).$$

• *The row vector* $\pi(n) = (\pi_1(n), \pi_2(n), \ldots, \pi_\ell(n))$ *is the distribution over the state space at time* n.
We have that:
  1 - $\pi_i(n) \geq 0$ *for any* $i \in S$.
  2 - $\sum_{i=1}^{\ell} \pi_i(n) = 1$.

Definition 10 (The initial probability distribution)
The initial probability distribution over the state space S *is defined by:*

$$\pi(0) = (\pi_1(0), \pi_2(0), \ldots, \pi_\ell(0)).$$

Property 1 *The probability state distribution at any time* n *depends on the transition matrix* P *and the initial distribution* $\pi(0)$.
We have that

$$\pi_i(n+1) = P(X(n+1) = i) = \sum_{k=1}^{\ell} P(X(n+1) = i | X(n) = k) P(X(n) = k),$$

hence

$$\pi_i(n+1) = \sum_{j=1}^{\ell} p_{ij} \pi_j(n), \tag{2.5}$$

writing (2.5) in the form of matrix multiplication:

$$\pi(n+1) = \pi(n)P, \tag{2.6}$$

equivalency

$$\pi_i(n) = \pi(0)P^{(n)}. \tag{2.7}$$

Definition 11 (The stationary distribution)
The distribution $\pi^* = (\pi_1^*, \pi_2^*, ..., \pi_\ell^*)$ *is said to be a stationary distribution if*

$$\pi^* = \pi^* P, \qquad i.e \qquad \pi_j^* = \sum_{j=1}^{\ell} \pi_i^* p_{ij}, \qquad \forall j \in S. \tag{2.8}$$

• If the stationary distribution exists then it is unique.

Definition 12 (The limiting distribution)
The distribution $\pi^* = (\pi_1, \pi_2, ..., \pi_\ell)$ *is said to be a limiting distribution if*

$$\lim_{n\to\infty} \pi(n) = \pi, \qquad i.e \qquad \lim_{n\to\infty} \pi_j(n) = \pi_j^*, \qquad \forall j \in S. \tag{2.9}$$

• The limiting distribution is a stationary distribution (from (2.8) and (2.9))
• If the limiting distribution exist then it is unique and it is independent from the initial distribution.

Definition 13 *We say that a state i accessible from state j if there is $n \in \mathbb{N}$ such that $p^n_{ij} > 0$ and we denote this by $i \to j$.*
If both $i \to j$ and $j \to i$ hold true, then we say that the states i and j communicate.
The Markov chain is said to be **irreducible** *if each two states in S communicate.*
In this case we say that the transition matrix P is irreducible.

Definition 14 *We say that the irreducible transition matrix $P = \left[p_{ij}\right]_{i,j \in S}$ of the finite homogeneous-time discrete Markov chain is reversible if for the unique stationary distribution $\pi = (\pi_1, \pi_2, ..., \pi_\ell)$ we have that:*

$$\pi_i p_{ij} = \pi_j p_{ji}, \forall i, j \in S, \tag{2.10}$$

and if this holds we say that π is reversible distribution of P.

2.2 Continuous-time Markov Chains

Definition 15 (Continuous time Markov chain)
The continuous time stochastic process $\{X(t), t \geq 0\}$ over the discrete state space $S = \{1, 2, \ldots, \ell\}$ is a continuous time Markov chain if the conditional probability distribution of $X(t+s)$ given $X(s), X(u)$ depends only on $X(s)$, i.e :

$$P(X(t) = j|X(s) = i, X(u) = k) = P(X(t) = j|X(s) = i), \quad \text{for all } u < s < t \text{ and } j, i, k, \in S$$

Time homogenous continuous Markov chain:
We say that a continuous Markov chain is time homogenous if $P(X(t+s) = j|X(s) = i)$ is independent from the time s.

In this section we consider that the continuous Markov chain $\{X(t), t \geq 0\}$ over the discrete finite state space $S = \{1, 2, \ldots, \ell\}$ is time-homogenous chain.

2.2.1 Basic definitions

• The chain moves from state i to state j at time t with transition probability denoted by $p(t)_{ij}$, where $p(t)_{ij} = P(X(t) = j|X(s) = i)$ and since our chain is time homogenous we have that:

$$p(t)_{ij} = P(X(t+s) = j|X(s) = i) = P(X(t) = j|X(0) = i), \forall s, t \geq 0.$$

Similarly to the case of discrete Markov chain we have:
• **The transition matrix** is the $(\ell \times \ell)$ stochastic matrix which its entries are the transition probabilities $p(t)_{ij}$ i.e:
1 - $0 \leq p(t)_{ij} \leq 1$, $i \neq j$, $\forall i, j \in S$
2 - $\sum_{j=1}^{\ell} p(t)_{ij} = 1 \qquad \forall i \in S$.
and we denote it by $P(t)$ with $P(0) = I$ so we have that $P(t) = \left[p(t)_{ij}\right]_{i,j \in S}$

$$P(t) = \begin{bmatrix} p(t)_{11} & p(t)_{12} & \cdots & p(t)_{1\ell} \\ p(t)_{21} & p(t)_{2,2} & \ldots & p(t)_{2\ell} \\ \vdots & \vdots & \ddots & \vdots \\ p(t)_{\ell 1} & p(t)_{\ell 2} & \cdots & p(t)_{\ell\ell} \end{bmatrix}$$

• **Chapman-Kolmogorov equation** :

$$P(t+s) = P(t)P(s), \forall t, s \geq 0.$$

• the state probability distribution $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_\ell(t))$ is the probability of being in state i at time t $\pi_i(t) = P(X(t) = i), \forall i \in S$ with:
1 - $\pi_i(t) \geq 0, \qquad \forall i \in S,\ \forall t \geq 0.$
2 - $\sum_{i=1}^{\ell} \pi_i(t) = 1, \quad \forall t \geq 0.$
• the initial probability distribution $\pi(0) = (\pi_1(0), \pi_2(0), \ldots, \pi_\ell(0))$ is the probability of being in state i at time 0 which is the starting time $\pi_i(0) = P(X(0) = i), \forall i \in S$.
• We have that:

$$\pi(t) = \pi(0)P(t). \tag{2.11}$$

• the probability distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_\ell)$ is said to stationary distribution if:

$$\pi = \pi P.$$

• The probability distribution $\pi = (\pi_1, \pi_2, ..., \pi_\ell)$ is said to be **limiting distribution** if:

$$\pi_j = \lim_{t \to 0} P(X(t) = j | X(0) = i) = \lim_{t \to 0} \pi_j(t) \qquad \forall j, i \in S.$$

2.2.2 The generator matrix

Unlike the discrete time Markov chain the process in this case moves from a state to another state, and the time spent in each state is a continuous random variable. Let us assume that the chain start at time 0 in the state i ($X(0) = i$), and let T_i be time that we leave the state i, since we start at time 0, T_i is also define the time that we spent in state i. Since we our continuous Markov chain is homogenous and satisfies the Markov property (memoryless property) we have that:

$$P(T_i \geq t + s | T_i \geq t) = P(T_i \geq s) \qquad \forall s, t\ \forall i.$$

We know that the only continuous probability distribution having this property is the exponential distribution, and let us denote the exponential distribution's parameters of T_i by λ_i i.e $T_i \sim Exponential(\lambda_i)$, so:
• The expectation of the time that the chain leaves the state i is :

$$E(T_i) = \frac{1}{\lambda_i}.$$

• for a very small Δt we have:

$$P(T_i \leq \Delta t) = 1 - \exp(-\lambda_i \Delta t) \approx 1 - (1 - \lambda_i \Delta t) = \lambda_i \Delta t.$$

Which means that the probability that the chain leaves the state i after a very small time Δt is approximately $\lambda_i \Delta t$, this is why λ_i is called the exit rate transition from state i, and we can write:

$$\lambda_i = \lim_{\Delta t \to 0} \frac{P(X(\Delta t) \neq i | X(0) = i)}{\Delta t}. \tag{2.12}$$

Now let's us define a new matrix and denote by $Q = \left[q_{ij}\right]_{i,j=1 \in S}$ such that :

$$Q = \frac{d}{dt} P(t)|_{t=0} = \lim_{\Delta t \to 0} \frac{P(\Delta t) - P(0)}{\Delta t}. \tag{2.13}$$

In particular we have:

$$\begin{aligned} q_{ij} &= \frac{d}{dt} p_{ij}(t)|_{t=0} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t) - p_{ij}(0)}{\Delta t}, \qquad i \neq j, \\ q_{ii} &= \frac{d}{dt} p_{ii}(t)|_{t=0} = \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t) - p_{ii}(0))}{\Delta t} \end{aligned}$$

we use that: $p_{ij}(0) = 0, i \neq j, p_{ii}(0) = 1$, we get:

$$\begin{aligned} p_{ij}(\Delta t) &= q_{ij} \Delta t + o(\Delta t) \qquad i \neq j \\ p_{ii}(\Delta t) &= 1 + q_{ii} \Delta t + o(\Delta t). \end{aligned}$$

Since $\sum_{j=0}^{\ell} p_{ij}(t) = 1$, we get:

$$\left. \begin{array}{lrcll} 1. & \sum_{j=1}^{\ell} q_{ij} & = & 0, & \forall i \in S \\ 2. & q_{ij} & \geq & 0, & i \neq j \quad . \end{array} \right\} \tag{2.14}$$

Hence,

$$q_{ii} = - \sum_{j=1}^{\ell} q_{ij}. \tag{2.15}$$

From (2.12) we get that :

$$\begin{aligned}\lambda_i &= \lim_{\Delta t \to 0} \frac{\sum_{j=1, j\neq i}^{\ell} (P(X(\Delta t) = j | X(0) = i)}{\Delta t} \\ &= \sum_{j=1, j\neq i}^{\ell} \lim_{\Delta t \to 0} \frac{(P(X(\Delta t) = j | X(0) = i)}{\Delta t},\end{aligned}$$

which gives us :

$$\lambda_i = \sum_{j=1, j\neq i}^{\ell} q_{ij}. \tag{2.16}$$

Combining (2.15) and (2.16) together we get that:

$$q_{ii} = -\lambda_i \qquad \forall i \in S. \tag{2.17}$$

Definition 16
A matrix $Q = \left[q_{ij}\right]_{i,j=1\in S}$ *which satisfying 2.14, is what we called* **the generator matrix** *or also known as* **the transition rate matrix** .

- q_{ij} is the transition rate from state i to state j.

The Kolmogorov forward and backward equations:
We have seen before the Kolmogorov equation, which implies that

$$p_{ij}(t+s) = \sum_{k\in S} p_{ik}(t) \cdot p_{kj}(s) \tag{2.18}$$

$$p_{ij}(s+t) = \sum_{k\in S} p_{ik}(s) \cdot p_{kj}(t), \tag{2.19}$$

differentiate the equation (2.18) with respect to s,

$$p'_{ij}(t+s) = \sum_{k\in S} p_{ik}(t) \cdot \frac{\partial}{\partial s} p_{kj}(s),$$

put $s = 0$,

$$p'_{ij}(t) = \sum_{k\in S} p_{ik}(t) \cdot q_{kj}.$$

In the matrix form we get:

$$P'(t) = P(t) \cdot Q, \tag{2.20}$$

the equation (2.20) is what we called the Kolmogorov forward equation.
Now doing same thing with the equation (2.19),

$$p'_{ij}(t) = \sum_{k \in S} q_{ik}(t) \cdot p_{kj}.$$

$$P'(t) = Q \cdot P(t), \tag{2.21}$$

the equation(2.21) is what we called the Kolmogorov backward equation.
As resulat:

$$P(t) = e^{Qt}. \tag{2.22}$$

Combining (2.11) and (2.22) yields to:

$$\pi(t) = \pi(0)\, e^{Q}. \tag{2.23}$$

Equivalency

$$\pi'(t) = \pi(t) \cdot Q\ . \tag{2.24}$$

Definition 17 *We say that a generator matrix* $Q = \left[q_{ij}\right]_{i,j=1 \in S}$ *is reversible if there exists a distribution* $\mu = (\mu_1, \mu_2, ..., \mu_\ell)$ *such that :*

$$\mu_i q_{ij} = q_{ji} \mu_j, \qquad \forall i, j \in S, \tag{2.25}$$

and if this holds we say that μ *is reversible distribution of* Q*.*

3

Embeddable Markov Matrices

A continuous-time Markov chain with a finite number ℓ of states is determined by its transition matrices $\{P(t), t > O\}$ and since we have the relation (2.22), hence it's determined by its generator matrix. If we observe this process at discrete time points $t_1 = 0, t_2 = 1, \ldots, t_n = T$ where the difference between each two consecutive discrete time points is a time unit 1, we obtain a discrete time Markov chain and let us consider it here homogenous [1] with the same time unit 1 whose transition matrix P, such that $P(1) = P$ which can be estimated easily let's say that the estimator of P is $\hat{P}$.

Here comes the problem of the existence of an estimator $\hat{Q}$ of the generator matrix Q of this continuous-time homogenous Markov chain such that $\hat{P} = e^{\hat{Q}}$ and this is what we called by the embedding problem.

Only partial results of the embedding problem are available up till now, In this chapter, we will see some of those results for the finite and homogenous transition matrix P.

Definition 18 *The transition matrix P is called embeddable if there exists a generator matrix Q such that:*

$$P = e^{Q}. \tag{3.1}$$

Remark 2 *Q is not necessarily unique.*

Definition 19 *A square matrix P is diagonalizable if there is an invertible matrix B and a diagonal matrix D such that:*

$$B^{-1}PB = D$$

or equivalency

$$P = BDB^{-1} \tag{3.2}$$

Remark For any matrix A

- $tr(A)$ is used to denote the sum of the diagonal entries of A.
- $det(A)$ is used to denote the determinate of A.

• If A is $\ell \times \ell$ diagonal matrix with diagonal entries $a_1, a_2, \ldots, a_\ell$ we denote it by $diag(a_1, a_2, \ldots, a_\ell)$.
• the diagonalization of a square matrix P is finding B and D satisfying .

Proposition 1 *Let the $\ell \times \ell$ matrix D be diagonal with γ_i ,$\forall i = 1, 2, \ldots, \ell$ are the entries on the diagonal then we have:*

$$e^D = diag(e^{\gamma_1}, e^{\gamma_2}, \ldots, e^{\gamma_\ell}), \tag{3.3}$$

and that

$$\log D = diag(\log^{\gamma_1}, \log^{\gamma_2}, \ldots, \log^{\gamma_\ell}), \tag{3.4}$$

Proposition 2 *Let the $\ell \times \ell$ matrix P be a diagonalizable then:*

$$e^P = Be^D B^{-1}, \tag{3.5}$$

and

$$\log P = B \log D B^{-1}. \tag{3.6}$$

3.1 The embedding problem of 2×2 transition matrices

The embedding problem of the 2×2 transition matrices was solved by giving a necessary and sufficient condition which is given in proposition below [1].

Proposition 3 *The 2×2 transition matrix P is embeddable if and only if :*

$$det(P) > 0,$$

or equivalency

$$tr(P) > 1.$$

Proof. First let P be a transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \text{ with } p_{11} = 1 - p_{12},\ p_{22} = 1 - p_{21},$$

we have

$$det(P) = p_{11}p_{22} - p_{12}p_{21} = (1 - p_{12})(1 - p_{21}) - p_{12}p_{21} = 1 - (p_{12} + p_{21}),$$

and

$$tr(P) = p_{11} + p_{22} = (1 - p_{12}) + (1 - p_{21}) = 2 - (p_{12} + p_{21}).$$

Now let us give the general form of a generator matrix Q which satisfying (2.14) :

$$Q = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}, \qquad \text{where } \alpha,\ \beta\ \geq 0.$$

First let us diagonalize the matrix Q, after calculation we found that eigenvalues of Q are 0 and $-(\alpha+\beta)$, and that the eigenvectors of it are $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -\frac{\alpha}{\beta} \\ 1 \end{pmatrix}$, hence we have:

$$B = \begin{pmatrix} 1 & -\frac{\alpha}{\beta} \\ 1 & 1 \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & 0 \\ 0 & -(\alpha+\beta) \end{pmatrix}, \qquad B^{-1} = \begin{pmatrix} \beta & \alpha \\ -\beta & \beta \end{pmatrix} (\alpha+\beta)^{-1},$$

thus $Q = BDB^{-1}$, using the equation (4.5) and after calculation we obtain:

$$e^{Q} = \begin{bmatrix} \beta + \alpha e^{-\alpha-\beta} & \alpha - \alpha e^{-\alpha-\beta} \\ \beta - \beta e^{-\alpha-\beta} & \alpha + \beta e^{-\alpha-\beta} \end{bmatrix} (\alpha+\beta)^{-1}.$$

Hence this satisfying solving :
$p_{12} = \alpha(\alpha+\beta)^{-1}(1-e^{-\alpha-\beta})$, $p_{21} = \beta(\alpha+\beta)^{-1}(1-e^{-\alpha-\beta})$,
Set $\delta = (\alpha+\beta)(1-e^{-\alpha-\beta})^{-1}$, i.e. $\alpha = \delta p_{12}$, $\beta = \delta p_{21}$ such that δ satistfies

$$p_{21} + p_{21} = 1 - e^{-\delta(p_{21}+p_{21})},$$

and this is possible only if $p_{12}+p_{21} \leq 1$ which is equivalent to the above two conditions in the proposition.

3.2 The embedding problem of 3×3 transition matrices

Let P be 3×3 transition matrix and let Q be the 3×3 generator matrix .
If $P = e^{Q}$ with γ is an eigenvalue of Q we have that if $\gamma \in \mathbb{R}$ then $\lambda = e^{\gamma} > 0$ is eigenvalue of P, and if $\gamma \in \mathbb{C}$ this implies that $\overline{\gamma}$ is also an eigenvalue of Q then $\lambda_1 = e^{\gamma}$ and $\lambda_2 = e^{\overline{\gamma}}$ are eigenvalues of P, we can see that if an eigenvalue of P is negative then it has multiplicity 2 [2].
Since any transition matrix has at least one eigenvalue equals to 1 then let the eigenvalues of P be $(1, \lambda_1, \lambda_2)$ so we can partial our problem to 3 cases :

1 - $\lambda_1 \neq \lambda_2$,with $0 < \lambda_1 < 1$,$0 < \lambda_2 < 1$ or $\lambda_1 = e^{\alpha+i\beta}$,$\lambda_2 = e^{\alpha-i\beta}$,$0 < \beta < \pi$.
2 - $\lambda_1 = \lambda_2 = \lambda$, $0 < \lambda < 1$.
3 - $\lambda_1 = \lambda_2 = \lambda$, $-1 < \lambda < 0$.

3.2.1 The Distinct Positive Eigenvalues or Complex Eigenvalues

This case was shown in the corollary [2] below

Corollary 1 *The 3×3 matrix P is a transition matrix with eigenvalues $(1, \lambda_1, \lambda_2)$ where $\lambda_1 \neq \lambda_2$.*
If $0 < \lambda_1 < 1$ and $0 < \lambda_2 < 1$ then P is embeddable if and only if

$$p_{ij}^2 \leq p_{ij} \frac{(\lambda_2^2 - 1) \log \lambda_1 - (\lambda_1^2 - 1) \log \lambda_2}{(\lambda_2 - 1) \log \lambda_1 - (\lambda_1 - 1) \log \lambda_2}, i \neq j.$$

If $\lambda_1 = e^{\alpha + i\beta}$,$\lambda_2 = e^{\alpha - i\beta}$,$0 < \beta < \pi$,then P is embeddable if and only if either

$$p_{ij}^2 (\beta (e^{\alpha} \cos \beta - 1) - \alpha e^{\alpha} \sin \beta) \geq p_{ij} (\beta (e^{2\alpha} \cos 2\beta - 1) - \alpha e^{2\alpha} \sin 2\beta), i \neq j$$

or

$$p_{ij}^2 ((\beta - 2\pi)(e^{\alpha} \cos \beta - 1) - \alpha e^{\alpha} \sin \beta) \geq p_{ij} ((\beta - 2\pi)(e^{2\alpha} \cos 2\beta - 1) - \alpha e^{2\alpha} \sin 2\beta), i \neq j.$$

3.2.2 The Coinciding Positive Eigenvalues

This case was shown in [2] as well.

Proposition 4 *Let P be 3×3 transition matrix with the eigenvalues $(1, \lambda, \lambda)$ where $0 < \lambda < 1$,if P is diagonable then we say that it is embeddable if and only if*

$$p_{ij}^2 \leq p_{ij} \frac{\lambda^2 \log \lambda^2 - \lambda^2 + 1}{\lambda \log \lambda - \lambda^2 + 1}$$

3.2.3 The Coinciding Negative Eigenvalues

Let P be 3×3 transition matrix with the eigenvalues $(1, \lambda, \lambda)$ such that $-1 < \lambda < 0$,then P is diagonalizable [3] which means that there exist an invertible 3×3 matrix B such that:

$$P = Bdiag(1, \lambda, \lambda) B^{-1},$$

which means also that $P^{\infty} = \lim_{n \to \infty} P^n$ exists and it's diagonalizable with eigenvalues equal to $(1, 0, 0)$, P^{∞} must consist of identical rows lets say (μ_1, μ_2, μ_3) [1], we have that:

$$P^{\infty} = Bdiag(1, 0, 0) B^{-1}.$$

It's easy to see that

$$P(\lambda, P^{\infty}) = P^{\infty} + \lambda(I - P^{\infty}), \tag{3.7}$$

with

$$P^{\infty} = \vec{e} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot (\mu_1, \mu_2, \mu_3)$$

Theorem 2 *Let $P(\lambda, P^{\infty})$ be as in (3.7) and with $\kappa \leq \lambda < 0, \kappa = -\frac{p_i}{p_i - 1}$ the following conditions are equivalent:*
(i) $P(\lambda, P^{\infty})$ is embeddable.
(ii) There exist $k \in \mathbb{N}$ and a 3×3 matrix $X2$ such that $X^2 = P^{\infty} - I$ and

$$Q = \log|\lambda|(I - P^{\infty}) + (2k+1)\pi, X \text{ is a generator matrix}$$

(iii)There exist a 3×3 matrix $X = [x_{ij}]_{i,j \in =1,2,3}$ such that $X^2 = P^{\infty} - I$ and

$$x_{ij} \geq \frac{1}{\pi} \log|\lambda| \mu_j, i \neq j.$$

Lemma 1 *$P(\lambda, P^{\infty})$ is embeddable if and only if there exist an embeddable stochastic matrix R such that $P(\lambda, P^{\infty}) = R^2$.*

Since we have here that the eigenvalues of $P(\lambda, P^{\infty})$ are $(1, \lambda, \lambda)$ with $\lambda < 0$, then $(1, i\sqrt{\lambda}, -i\sqrt{\lambda})$ are eigenvalues of R.
Now taking **theorem2** and apply it in this case yields to the following theorem [3]

Theorem 3 *Let $P(\lambda, P^{\infty})$ be as in(3.7) with $\kappa \leq \lambda < 0, \kappa = -\frac{p_i}{p_i - 1}$ we say that it's embeddable if and only if there exist a 3×3 stochastic matrix $R = [r_{ij}]_{i,j \in S}$ such that $P(\lambda, P^{\infty}) = R^2$ with either:*

$$r_{ij} \geq (1 + \frac{\sqrt{\lambda}}{\pi} \log|\lambda|)\mu_i, \forall i \neq j$$

or

$$r_{ij} \leq (1 - \frac{\sqrt{\lambda}}{3\pi} \log|\lambda|)\mu_i, \forall i \neq j.$$

Since it's not easy to calculate the square root of a matrix, there is a new necessary and sufficient condition for the irreducible stochastic matrix P with the unique stationary probability distribution μ which will be shown in the next theorem [4].

Theorem 4 *The irreducible transition matrix $P(\lambda, P^{\infty})$ with the unique stationary probability distribution μ with $P(\lambda, P^{\infty})$ be as in (3.7) and is embeddable if and only if*

$$\sqrt{\frac{\mu_1\mu_2\mu_3}{(\mu_1\mu_2+\mu_1\mu_3+\mu_2\mu_3)^2}-1} \geq \frac{\pi}{-\log\lambda}$$

when $\frac{1}{\mu_1}, \frac{1}{\mu_2}, \frac{1}{\mu_3}$ satisfy the triangle inequality,

$$\sqrt{\frac{m}{m-1}} \geq \frac{\pi}{-\log\lambda}, m = \min(\mu_1, \mu_2, \mu_3)$$

,otherwise.

3.3 The reversible embedding problem

In this subsection, we will see the the embedding problem for the finite $\ell \times \ell$ reversible transition matrix P with the reversible distribution μ which was shown in [5].
For all what we will have next in this subsection we assume that P is $\ell \times \ell$ irreducible stochastic matrix with the unique stationary distribution μ.

Definition 20 *The reversible stochastic matrix P is embeddable if there exist a reversible generator matrix Q such that:*

$$P = e^{Q}. \tag{3.8}$$

and we say that Q reversible embedding of P

Lemma 2 *If P is reversible embeddable and Q is the reversible embedding of P then Q is irreducible and the reversible distribution of it is μ which is the reversible distribution of P.*

Lemma 3 *If P is reversible, then P is diagonalizable and the eigenvalues of P are all real numbers.*

We assume that P is reversible with the positive eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_\ell$ so according to Lemma 2 there exists a diagonal matrix $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ and an invertible matrix B such that: $P = BDB^{-1}$. Let $G = B\ln DB^{-1}$,it is easy to see that $e^{G} = P$.

Let $\alpha_1, \alpha_2, \ldots, \alpha_m$ be the different eigenvalues of P with ℓ_i be the multiplicity of the eigenvalue $\alpha_i, \forall i = 1, 2, \ldots, m$, $(\ell_1 + \ell_2 + \ldots + \ell_m = \ell)$, this implies that that $D = diag(\alpha_1, \ldots, \alpha_1, \alpha_2, \ldots, \alpha_2, \alpha_m, \ldots, \alpha_m)$.
Now let $\kappa_0, \kappa_1, \ldots, \kappa_{m-1} \in \mathbb{R}$ be the solution to the following system of linear equations:

$$\begin{cases} \kappa_0 + \kappa_1\alpha_1 + \ldots + \kappa_{m-1}\alpha_1^{m-1} = \log(\alpha_1) \\ \kappa_0 + \kappa_1\alpha_2 + \ldots + \kappa_{m-1}\alpha_2^{m-1} = \log(\alpha_2) \\ \vdots \\ \kappa_0 + \kappa_1\alpha_m + \ldots + \kappa_{m-1}\alpha_m^{m-1} = \log(\alpha_m) \end{cases} \tag{3.9}$$

and since we have that $\alpha_i \neq \alpha_j, \forall i, j = 1, 2, \ldots, m$, so the solution of this linear system exists and it's unique.
We can see that (3.9) is equivalent to:

$$logD = \kappa_0 I + \kappa_1 D + \ldots + \kappa_{m-1} D^{m-1},$$

which implies that

$$G = \kappa_0 I + \kappa_1 P + \ldots + \kappa_{m-1} P^{m-1} \tag{3.10}$$

Lemma 4 *Let P be reversible with positive eigenvalues, then P is reversibly embeddable if and only if G is a reversible generator matrix.*

Lemma 5 *Let P be reversible with positive eigenvalues, then G is reversible matrix.*

Lemma 6 *Let P be reversible with positive eigenvalues, then for $G = \left[g_{ij}\right]_{i,j=1 \in S}$ we have that $\sum_{j=1}^{\ell} g_{ij} = 0$*

According to Lemma4 we only need the off-diagonal entries of G to be all positive, and form (3.10) this is equivalent to say that we need the off-diagonal entries of $\kappa_0 I + \kappa_1 P + \ldots + \kappa_{m-1} P^{m-1}$ to be all positive.
And we have that,

Lemma 7 *If P is reversibly embeddable, then P is reversible and the eigenvalues of P are all positive.*

Remark 3 *For the proofs of these lemmas and more details readers can check.*

All this will lead us to the following result which we will as a theorem [5]:

Theorem 5 *For $P^n = (p_{ij}^n), n \geq 1$,P is reversibly embeddable if and only if the following two condition hold:*
1 - P is reversible and the eigenvalues of P are all positive;
2 - $\kappa_1 p_{ij} + \kappa_2 p_{ij}^2 + \ldots + \kappa_{m-1} p_{ij}^{m-1}, i \neq j$.

The proves of all lemmas and theorems in this subsection can be found in [5].

3.4 Further results

In this subsection, we will summarize some of the results about the general case of the embedding problems, the existence and the uniqueness of the generator matrix of the $\ell \times \ell$ transition matrix P .

3.4.1 Existence

In the theorem below [6], we will see some necessary condition of the embeddability of an $\ell \times \ell$ transition matrix P i.e the existence of its generator matrix.

Theorem 6 *If the $\ell \times \ell$ transition matrix P has a valid generator matrix Q then*

1 - $det(P) > 0$;

2 - $det(P) \leq \prod_{i=1}^{\ell} p_{ii}$;

3 - no eigenvalue of P other than 1 can satisfy $|\lambda| = 1$ and any negative eigenvalue must have even (algebraic) multiplicity;

4 - for every pair of states i and j such that j is accessible from i, $p_{ij} > 0$;

5 - whenever $p_{ij} = 0$, then $p_{ij}^m = 0$, $m = 2, 3, \ldots$

6 - if P has distinct eigenvalues, then each eigenvalue γ of Q satisfies

$$|\gamma| \leq |\log(det(P))|;$$

7 - there exist $i, j \in S$ with $i \neq j$, such that $\forall k$,

$$p_{ik} = 0 \Rightarrow p_{jk} = 0,$$

and likewise there exist $i', j' \in S$ with $i' \neq j'$, such that $\forall k$,

$$p_{ki'} = 0 \Rightarrow p_{kj'} = 0;$$

8 - the entries of P must satisfy

$$p_{ik} \geq m^m r^r (m+r)^{-m-r} \sum_{j=1}^{\ell} (p_{ij} - b_m)(p_{jk} - b_r) \mathbf{1}_{p_{ij} > b_m, p_{jk} > b_r},$$

for any positive integers m and r. Here $\mathbf{1}_B$ is the indicator function of the Boolean event B and $b_m = \sum_{k=m+1}^{\infty} e^{-\alpha}\alpha^k/k!$ is the probability that $N' > m$, where N' is a Poisson random variable with mean $\alpha \equiv \max_i(-q_{ii})$.

In other word, if the $\ell \times \ell$ transition matrix P doesn't satisfies those above conditions it will not be embeddable.

3.4.2 Uniqueness

As we mentioned before if $\ell \times \ell$ transition matrix P is embeddable that doesn't mean that it has only one generator matrix, and this raised two questions here. Under what condition the embeddable transition matrix P has only one generator matrix, and if there is more than one how can we choose the good one since different generator matrices leads to different transition matrices $P(t)$.

For the first question, the answer is that if we assume that the transition matrix has several generator matrices $Q_1, Q_2, \dots$ then the one which has the smallest value of K such that:

$$K = \sum_{i,j \in S} |j - i||q_{ij}|$$

is the one that will represent the empirical transition matrices the best [7].

Now coming to the second question the next theorem gives some condition that leads to the uniqueness of the generator Q [7]:

Theorem 7 *Let $P = \left[p_{ij}\right]_{i,j \in S}$ be an $\ell \times \ell$ transition matrix*
(i) If $det(P) > 1/2$, then P has at most one generator.
(ii) If $det(P) > 1/2$ and $||P - I|| < 1/2$ (using any operator norm), then the only possible generator for P is $\log_p(P)$.
(iii) If P has distinct eigenvalues and $det(P) > e^{-\pi}$, then the only possible generator for P is $\log_p(P)$.
such that $\log_p(P)$ denote the logarithm of the matrix P that has eigenvalues all lying in the strip $\{z \in \mathbb{C}| - \pi < im(z) < \pi\}$.

Theorem 8 *If P is reversibly embeddable, then it has one unique the reversible embedding matrix Q such that $e^Q = P$ (see [5]).*

4

Estimating a Continuous Markov chain from Discrete Data

We have seen in the previous chapter that if the estimated transition matrix $\hat{P}$ is embeddable than we can find the estimator $\hat{Q}$ of generator matrix Q easily, but what if $\hat{P}$ is not embeddable. In this chapter, we will see how to calculate $\hat{P}$ and then how to find $\hat{Q}$ if $\hat{P}$ is embeddable, and if it is not we will see some methods to deal with this problem, then we will see other methods to estimate $\hat{Q}$ from the given discrete-time data without using $\hat{P}$.

4.1 The maximum likelihood estimator

For a given observation, maximum likelihood estimation is a method of estimating the parameters of a statistical model. The method obtains the parameter estimates by finding the parameter values that maximize the likelihood function. The estimates are called maximum likelihood estimates (MLE). And since the log function is an increasing function, the value which maximizes a function is the same that maximizes its log function that's why we often take the logarithm of the likelihood function tome to make it easier to find the MLE since to maximize it we just take it's derivative and find the values that make it equals to 0.

Let $Y = \{Y(t)|t \geq 0\}$ be a continuous time Markov process over the discrete finite state space $S = \{1, 2, \ldots, \ell\}$. If we observe this process continuously in the time interval $[0, T]$, hence we have the complete continuous data set $Y = \{Y(t)|0 \leq t \leq T\}$, the maximum likelihood estimation of Q estimator was solved by [8]. The likelihood function of the parameter $Q = \left[q_{ij}\right]_{i,j \in S}$ which is the generator matrix, it is given in by:

$$L_{(c)}(Q, Y) = \prod_{i=1}^{\ell} \prod_{j \neq i} q_{ij}^{N_{ij}(T)} e^{-q_{ij} R_i(T)}. \tag{4.1}$$

here (c) is to indicates the continuous time observations and $N_{ij}(T)$ denote the number of transitions from state i to state j in the time interval $[0, T]$ and

$$R_i(t) = \int_0^t \mathbf{1}_{\{Y(s)=i\}} ds$$

is the time that the process spent in state i before time t.

Taking the logarithm of (4.1) we get :

$$\log L_{(c)}(Q, Y) = \sum_{i=1}^{\ell} \sum_{j \neq i} [N_{ij}(T) \log(q_{ij}) - q_{ij} R_i(T)]. \tag{4.2}$$

and then take its partial derivative with respect to q_{ij} and make it equal to 0 we get

$$\frac{\partial log L_{(c)}(Q, Y)}{\partial q_{ij}} = 0 \Leftrightarrow \sum_{i=1}^{m} \sum_{j \neq i} [\frac{N_{ij}(T)}{q_{ij}} - R_i(T)].$$

Now we can see that the maximum likelihood estimator of Q is $\hat{Q} = \left[\hat{q}_{ij}\right]_{i,j \in S}$such that

$$\hat{q}_{ij} = \frac{N_{ij}(T)}{R_i(T)}.$$

Remark 4 *$\hat{Q}$ is a generator matrix i.e satisfies (2.14)*

We will use the above likelihood function later in chapter.

Let $Y = \{Y(t)|t \geq 0\}$ be a continuous time Markov process over the discrete finite state space $S = \{1, 2, \ldots, \ell\}$. If we observe this process at discrete time points $t_1 = 0, t_2 = 1, \ldots, tn = T$ such that difference between each two consecutive discrete time points is a time unit 1, i.e. that we have the discrete observed data set $\{Y(t_1), \ldots, Y(t_n)\}$ which consists on n observations.
The process $X = \{Y(t_1), \ldots, Y(t_n)\}$ is a time-homogenous discrete Markov chain, the likelihood function of the parameter $P = \left[p_{ij}\right]_{i,j \in S}$ which is the transition probabilities matrix it is given by

$$L_n(P) = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} p_{ij}^{K_{ij}(n)} \tag{4.3}$$

Here $K_{ij}(n)$ is the number of transitions from state i to state j during the period of n observations,taking the logarithm of (3.3) we get :

$$\log L_n(P) = \prod_{i=1}^{\ell} \prod_{j=1}^{\ell} p_{ij}^{K_{ij}(n)} \tag{4.4}$$

This likelihood function is identical to the likelihood function for n independent multinoumial distributions, so the maximum likelihood estimator of the parameter P is $\hat{P} = \left[\hat{p}_{ij}\right]_{i,j \in S}$ such that

$$\hat{p}_{ij} = \frac{K_{ij}(n)}{K_{i.}(n)} \tag{4.5}$$

here

$$K_{i.}(n) = \sum_{j=1}^{\ell} K_{ij}(n).$$

If $\hat{P}$ given by equation (4.5) is embeddable then the maximum likelihood estimator $\hat{Q}$ of the generator matrix exist and it is obtained by $\hat{Q} = \log \hat{P}$, if there is more than one in it's shown how to choose the best one . If $\hat{P}$ is not embeddable then either $\hat{Q}$ exists and there are methods to obtain $\hat{Q}$ using the transition matrix estimator $\hat{P}$ or the maximum likelihood estimator $\hat{Q}$ of the generator matrix doesn't exist.

4.2 The diagonal adjustment and weighted adjustment

We assume that the $\ell \times \ell$ transition matrix estimator $\hat{P}$ given by the equation (4.5) is not embeddable . Now we want to find a matrix $\widetilde{Q}$ such that $\hat{P} = e^{\widetilde{Q}}$ equivalency satisfying $\widetilde{Q} = \log \hat{P}$. First, we want to be sure that the logarithm function of $\hat{P}$ exist. In [7] it was shown that there is a necessary condition that guarantees the existence of the real matrix logarithm of $\hat{P}$ which will be shown in the theorem below.

Theorem 9 *Let $\hat{P}$ be $\ell \times \ell$ transition matrix, and let $F = max\{(a-1)^2 + b^2, a + b \cdot i \text{ is eigenvalue of } P, a, b \in \mathbb{R}\}$, assume that $F < 1$ then the series*

$$\widetilde{Q} = log\hat{P} = (\hat{P} - I) - \frac{(\hat{P} - I)^2}{2} - \frac{(\hat{P} - I)^3}{3} - \frac{(\hat{P} - I)^4}{4} - \dots . \tag{4.6}$$

converges geometrically quickly, and gives rise to an ℓ matrix $\widetilde{Q}$ with sum rows equal to 0 such that $e^{\widetilde{Q}} = \hat{P}$.

In the other hand if the series converges absolutely this will imply that has row sums equal to 0, so the condition $F < 1$ is not needed anymore.

For the transition matrix estimator $\hat{P}$ We can see that $\widetilde{Q}$ satisfies $e^{\widetilde{Q}} = \hat{P}$ and that its row sums are equal to 0 but due to the embedding problem is not sure that we will have the condition that its off-diagonal entries are positive which allow $\widetilde{Q}$ to be generator matrix, to solve this problem we can use the two adjustment methods which are the Diagonal Adjustment and Weighted Adjustment [7].

In all what we will have next in this section let $\hat{P}$ be the $\ell \times \ell$ transition matrix estimator given by equation (4.5) and set $\widetilde{Q} = \left[\widetilde{q}_{ij}\right]_{i,j \in S}$ be the solution of $\widetilde{Q} = \log \hat{P}$

4.2.1 The Diagonal Adjustment

Let $q_{ij}^{\mathtt{DA}}$ be the elements of the generator matrix $Q^{\mathtt{DA}}$ obtained by the Diagonal Adjustment method.
In this method, we will replace the negative values of $\widetilde{Q}$ by 0 since usually they will be very small and then we add the difference to the diagonal entries to ensure that the row sums will be 0.

$$q_{ij}^{\mathtt{DA}} = max(\widetilde{q}_{ij}, 0), i \neq j, q_{ii}^{\mathtt{DA}} = -\sum_{j=1, j\neq i}^{\ell} q_{ij}^{\mathtt{DA}}. \tag{4.7}$$

However the generator matrix $Q^{\mathtt{DA}}$ will not satisfies that $P = e^{Q^{\mathtt{DA}}}$ exactly.

4.2.2 The Weighted Adjustment

Let $q_{ij}^{\mathtt{WA}}$ be the elements of the generator matrix $Q^{\mathtt{WA}}$ obtained by the Weighted Wdjustment method. Let

$$G_i = |\widetilde{q}_{ii}| + \sum_{j\neq i} max(\widetilde{q}_{ij}, 0), B_i = \sum_{j\neq i} max(-\widetilde{q}_{ij}, 0)$$

$$q_{ij}^{\mathtt{WA}} = \begin{cases} 0 & \text{if } i \neq j \text{and } \widetilde{q}_{ij} \leq 0, \\ \widetilde{q}_{ij} - B_i|\widetilde{q}_{ij}|/G_i & \text{otherwise if } G_i > 0, \\ \widetilde{q}_{ij} & \text{otherwise if } G_i = 0. \end{cases} \tag{4.8}$$

we know that $\sum_{j=1}^{\ell} \widetilde{q}_{ij} = 0$, then $G_i \geq B_i$; hence that $q_{ij}^{\mathtt{WA}} \geq 0 \forall i \neq j$.
However the generator matrix $Q^{\mathtt{WA}}$ will not satisfies that $P = e^{Q^{\mathtt{WA}}}$ exactly.
Unfortunately, both of this methods don't give us an optimal generator matrix.

4.3 Quasi-Optimization of the Generator

This method was shown first in [9], Again let $\hat{P}$ be the $\ell \times \ell$ transition matrix estimator given by equation (4.5) and set $\widetilde{Q} = [\widetilde{q}_{ij}]_{i,j\in S}$ be the solution of $\widetilde{Q} = \log \hat{P}$. The Quasi-Optimization of the Generator method which is to find an approximation of the generator matrix is the same as finding a solution of the minimization problem

$$\min_{Q\in\mathcal{Q}} \left\| Q - \widetilde{Q} \right\| \tag{4.9}$$

such that $\mathcal{Q}$ is the set of all generator matrices and is the $\|\cdot\|$ is the Euclidean norm.
We can solve this problem row by row since the conditions() of Q are closed on each row, then solving() is equivalent to solving the ℓ independent minimization problems

$$\min_{q\in\mathcal{C}(\ell)} \sum_{i=1}^{\ell} (p_i - q_i)^2 \tag{4.10}$$

such that $p = (p_1, \ldots, p_\ell)$ is a row vector of $\widetilde{Q}$ and $q = (q_1, \ldots, q_\ell)$ is a row vector of a generator matrix $Q \in \mathcal{Q}$, which is permuted i.e $q_i \leq q_{i+1}$ this means that q_1 is diagonal element in this row, and

$$\mathcal{C}(\ell) = \{q \in \mathbb{R} | \sum q_i = 0, q_1 \leq 0, q_i \geq 0 \text{ for } i \geq 2\}.$$

The algorithm to solve (system) i.e. finding the optimal is resumed in the steps below [10]:

step1: Let $b = (b_1, \ldots, b_\ell)$ such that $b_i = p_i + \alpha$ where

$$\alpha = -\frac{1}{\ell} \sum_{i=1}^{\ell} p_i, \text{ for } i = 1, 2, \ldots, \ell.$$

step2: Compute $\tilde{p} = \pi(b)$, here π is a permutation that

orders b such, that $b_i \leq b_{i+1}$.

step3: Find the smallest m^*, for $2 \leq m \leq \ell - 1$ which satisfy

$$(\ell - m + 1)\tilde{p}_{m+1} - \Big(\tilde{p}_1 + \sum_{i=0}^{\ell-m-1} \tilde{p}_{\ell-i}\Big) \geq 0.$$

step4: Construct $q^* = (q_1^*, \ldots, q_\ell^*) \in \mathcal{C}(\ell)$, where

$$q_i^* = \begin{cases} 0 & \text{if } 2 \leq i \leq m^* \\ \tilde{p}_i - \frac{1}{k-m^*+1}\Big(\tilde{p}_i + \sum_{j=m^*+1}^{\ell} \tilde{p}_j\Big) & \text{otherwise.} \end{cases} \tag{4.11}$$

step5: Compute the inverse permutation $\pi^{-1}(q^*)$,
which is the optimal solution to this problem.

4.4 The EM algorithm

let $Y = (X, Z)$ be the complete set of data, the Expectation maximization which is known as the EM algorithm is a method to obtain the MLE with only X is observed and given.to get the MLE using this method we need two steps :
The E-step: the calculation of conditional expectation of $\log(L(\theta, Y))$ given X and the current MLE θ_0. i.e. the calculation of $\mathbb{E}[\log L(\theta, Y)|X, \theta_0]$.
The M-step: is to obtain new MLE $\hat{\theta}$ by maximization of

$$E[L(\theta, Y)|X, \theta_0],$$

then we put $\theta_0 = \hat{\theta}$, and we repeat those two step until the sequence converges.

The application of this method in in our case was shown in [11], it's obvious that

$Y = \{Y(t)|0 \leq t \leq T\}$ is the complete data and that $X = \{Y(t_1), \ldots, Y(t_n)\}$ is the discrete observed, here $t_1 = 0$ and $t_n = T$. The maximum likelihood function of a given continuous observed data set Y is given by the equation (4.1):

The E-step: From the equation (4.1) and for an initial generator matrix Q_0 we see that:

$$\mathbb{E}_{Q_0}[\log L_{(c)}(Q,Y)|X,Q_0] = \sum_{i=1}^{\ell}\sum_{j\neq i}\log(q_{ij})\mathbb{E}_{Q_0}[N_{ij}(T)|X] - \sum_{i=1}^{\ell}\sum_{j\neq i} q_{ij}\mathbb{E}_{Q_0}[R_i(T)|X]. \tag{4.12}$$

Hence we need to calculate $\mathbb{E}_{Q_0}[N_{ij}(T)|X]$ and $\mathbb{E}_{Q_0}[R_i(T)|X]$, and since we have that:

$$\mathbb{E}_{Q_0}[N_{ij}(T)|X] = \sum_{k=0}^{n-1}\widetilde{F}^{ij}_{X_k X_{k+1}}(t_{k+1} - t_k) \tag{4.13}$$

such that

$$\widetilde{F}^{ij}_{kl}(t) = \mathbb{E}[N_{ij}(t)|Y(t) = l, Y(0) = k], \tag{4.14}$$

and that:

$$\mathbb{E}_{Q_0}[R_i(T)|X] = \sum_{k=0}^{n-1}\widetilde{M}^{i}_{X_k X_{k+1}}(t_{k+1} - t_k), \tag{4.15}$$

such that:

$$\widetilde{M}^{i}_{kl}(t) = \mathbb{E}[R_i(t)|Y(t) = l, Y(0) = k], \tag{4.16}$$

it is enough to calculate $\widetilde{M}^{i}_{kl}(t)$ and $\widetilde{F}^{ij}_{kl}(t)$,we choose $\lambda = \max_{i\in S}(-q_{ii}, 0)$ and define $B = I + \lambda^{-1}Q_0$ and let e_j, e'_j denoting the unit vector with the j-th coordinate equal to 1 and it's transpose respectively, it's shown in that:

$$\widetilde{M}^{i}_{kl}(t) = \frac{M^{i}_{kl}(t)}{e'_k e^{Q_0 t} e_l} \tag{4.17}$$

such that

$$M^i(t) = [M^{i}_{kl}(t)]_{k,l\in S} = e^{-\lambda t}\lambda^{-1}\sum_{n=1}^{\infty}\frac{(\lambda t)^{n+1}}{(n+1)!}\sum_{s=0}^{n}B^s(e_i e'_i)B^{n-s}. \tag{4.18}$$

and that:

$$\widetilde{F}^{ij}_{kl}(t) = \frac{F^{ij}_{kl}(t)}{e'_k e^{Q_0 t} e_l}, \tag{4.19}$$

such that

$$F^{ij}(t) = [F^{ij}_{kl}(t)]_{k,l\in S} = q_0 ij e^{-\lambda t}\lambda^{-1}\sum_{n=1}^{\infty}\frac{(\lambda t)^{n+1}}{(n+1)!}\sum_{s=0}^{n}B^s(e_i e'_j)B^{n-s}. \tag{4.20}$$

M-step We can see that the value $\hat{Q} = [\hat{q}_{ij}]_{i,j \in S}$, where

$$\hat{q}_{ij} = \frac{\mathbb{E}_{Q_0}[N_{ij}(T)|X]}{\mathbb{E}_{Q_0}[R_i(T)|X]}, \quad \forall i \neq j, \tag{4.21}$$

is what maximize $\mathbb{E}_{Q_0}[\log L_{(c)}(Q, Y)|X, Q_0]$ so is the new MLE is $\hat{Q}$. We put $\hat{Q} = Q_0$ and we repeat this until the sequence converges.

The convergence of the series of the obtained generator matrices $\{Q_k\}_{k=1}^{K}$ depends on the choice of the initial generator matrix Q_0, it's good idea to choose it in such way that $det(e^{Q_k})$ is much greater than 0. We also have to choose Q_0 to be in the interior of $\mathcal{Q}$, here $\mathcal{Q}$ is set of all generator matrices.

4.5 Markov Chain Monte Carlo (Gibbs Sampler)

In this section we will see how can we use the Monte Carlo Markov Chain estimation methods which was shown in [11]. There are several Monte Carlo Markov chain methods and one of them is the Gibbs sampler which what we will employ here. By using this method we will avoid the non-existence problem of the maximum likelihood estimator besides that it is easier to implement computationally.
The Gibbs sampler is an algorithm that samples from the conditional distribution of some joint distribution. For a better understanding how this algorithm works I will start giving its steps in the the general case, suppose that we want to obtain K samples of $X = (x_1, \ldots, x_n)$ from the initial joint distribution $p(x_1, \ldots, x_n)$, let the ith sample denoted by $X^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$ the algorithm is:
1. Initialize $X^{(0)} = (x_1^{(0)}, \ldots, x_n^{(0)})$ for $t = 0$.
2. Now for i=0,...,k
To sample $X^{(i+1)}$ we sample each component of it $x_j^{(i+1)}, j = 1, \ldots, n$ by updating it according to the probability distribution specified by

$$p(x_j^{(i+1)}|x_1^{(i+1)}, \ldots, x_{(j-1)}^{(i+1)}, x_{(j+1)}^{(i)}, \ldots, x_{(n)}^{(i)}).$$

Theorem 10 (Bayes' theorem)

$$p(\phi|Y) = \frac{p(Y|\phi)p(\phi)}{p(Y)}, \tag{4.22}$$

we can ignore the normalization of the constant $p(Y)$, thus utilizing that

$$p(\phi|Y) \propto p(Y|\phi)p(\phi), \tag{4.23}$$

here $\propto$ refers to the proportion. $p(\phi)$ is the prior distribution which is the probability before observing any data, $p(\phi|Y)$ is the posterior distribution which is the conditional distribution of the parameter ϕ given data Y. It's the distribution after the new data Y observed, and $p(Y|\phi)$ is the likelihood function as in (4.1).

Now let us consider that for a complete data we only have the discrete time observations $X = \{Y(t_1), \ldots, Y(t_n)\}$ observed at time $(t_1 = 0, \ldots, t_n = T)$, in this case, we will apply the two sites Q and J Gibbs simple for a given, thus we draw J from a given (Q, X) and Q for a given (J, X), hence we will obtain $\{Q^{(k)}, J^{(k)}\}_{k=1}^{K}$, where the Qs are the generator matrices and the Js are the simulated Markov chains samples such that, and do that we first choose a good prior distribution to be the initial probability distribution.
For the prior distribution of Q in [11] the gamma distribution was proposed:

$$p(Q) = \prod_{i=1}^{\ell} \prod_{j \neq i} q_{ij}^{\alpha_{ij}-1} e^{q_{ij}-\beta_i}, \tag{4.24}$$

where $\alpha_{ij} > 0$, $i, j \in S$, and $\beta_i > 0$, $i \in S$, are know constant that we choose, we can see that q_{ij} a re gamma-distributed with shape α_{ij} and rate $1/\beta_i$ and we denote this by $q_{ij} \sim \Gamma(\alpha_{ij}, 1/\beta_i)$, hence the mean and the variance are $\frac{\alpha_{ij}}{\beta_i}$ and $\frac{\alpha_{ij}}{\beta_i^2}$ respectively.
Using and since the likelihood function for the complete data is given in the equation (4.1), the posterior distribution of Q is given by

$$p(Q|J, X) = P(Q|J) \propto p(Q)L_{(c)}(Q, Y) = \prod_{i=1}^{\ell} \prod_{j \neq i} q_{ij}^{N_{ij}(T)+\alpha_{ij}-1} e^{-q_{ij}(R_i(T)+\beta_i)}. \tag{4.25}$$

We can see that the posterior distribution follows gamma distribution in particular q_{ij} is gamma-distributed, which make it tractable to draw q_{ij} from $\Gamma(+\alpha_{ij}, 1/(+\beta_i))$.

We can summarize the Gibbs sample algorithm as follow [10]. :
(i) Construct the initial generator matrix Q_0 by drawing $q_{ji,0}$ from the prior distribution for $j \neq i$.
(ii) Iterate the following steps up to K times
1. Simulate a new sample of continuous Markov chain J with generator matrix Q such that all the observation are realized.
2. From the simulated Markov chain, we calculate $N_{ij}(T)$ and $R_i(T)$.
3. Construct a new Q by drawing q_{ij} from the posterior distribution $\Gamma(+\alpha_{ij}, 1/(+\beta_i))$.
4. Save this new Q and use it to the next simulation.

(iii) For $Q_1, \ldots, Q_k$ be the sequence of the generator matrices obtained by the Gibbs sampler algorithm we drop some proportion lets say the first N proportion (burn-in) and then take the mean of the remaining ones.
Which means that the estimator $\hat{Q}$ of the generator matrix Q is given by:

$$\hat{Q} = \frac{1}{K - N} \sum_{i=N+1}^{K} Q_i. \tag{4.26}$$

The problems of this method are the choice of the α and β and the number of iteration and the number of iteration K before we know that the sample has convergent (burn-in).

5

Adoption for Finantial Time Series

The `ctmcd` is a package in R project using to estimate the continuous time Markov chain parameters i.e. the generators matrices when only the discrete time observations are available. The implemented methods to derive these estimates are:
-diagonal adjustment (DA)
-weighted adjustment (WA)
-quasi-optimization (QO)
-an instance of the EM algorithm (EM)
-a Gibbs Sampler (GS).

We use the command "gm(tm, te, method, ...)", where
tm:is the matrix of absolute transition frequencies if the method is the The EM algorithm (EM) or the Gibbs Sample (GS) or the matrix of relative transition frequencies which is the transition matrix if the method is diagonal adjustment (DA), weighted adjustment (WA) or quasi-optimization (QO).
te: is the time path which is in our case the unit time 1.
method: the method that we want to use:"DA", "WA", "QO", "EM", "GS".

In my application, I will use a discrete data set of gold price that was observed monthly from the year 1950 till 2018, i.e the time unit here is 1 month (Gold price time series - 1950-2018). beside the ctcmd package [12], I will use the packages in [13] and[14].

5.1 Presentation of the analyzed data set

The plot of gold price time series.
The price difference Plot.
The price relative Difference Plot.

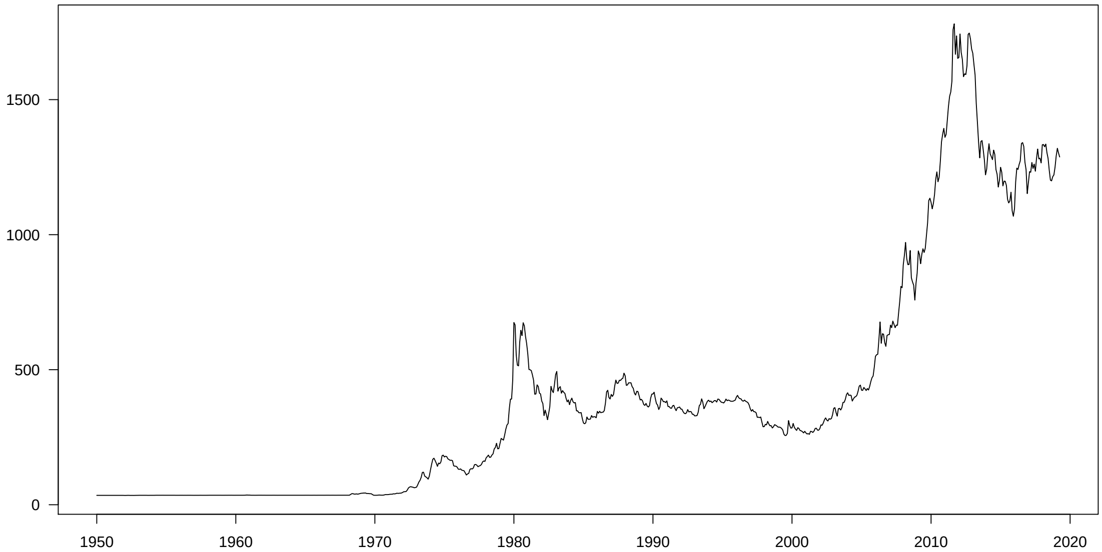

Figure 5.1: Gold price time series - 1950-2018

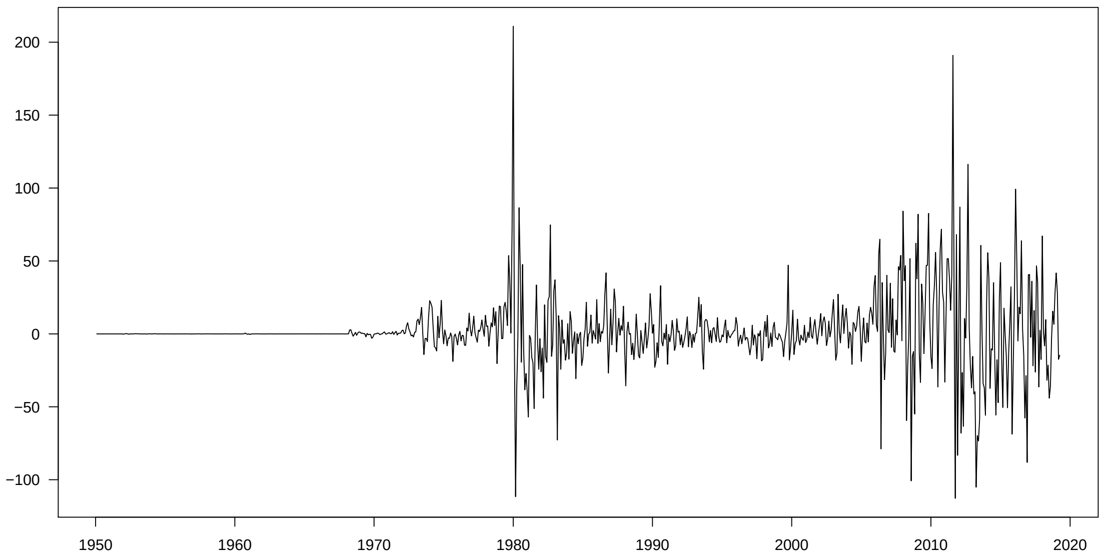

Figure 5.2: Gold price difference time series - 1950-2018

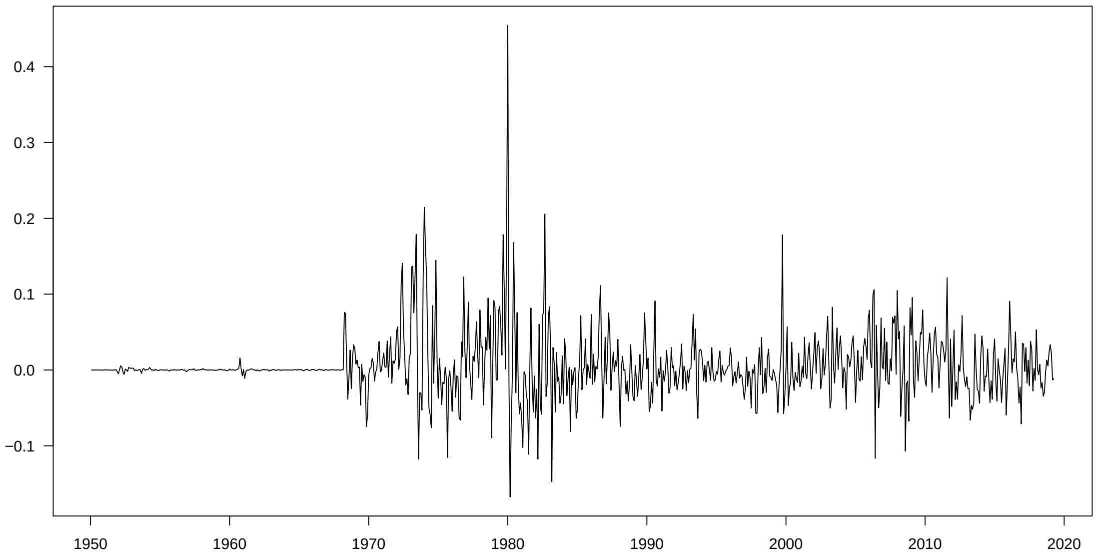

Figure 5.3: Gold price relative differences time series - 1950-2018

5.2 Discretization

Our states here will be the 5 different changes in the price let 1 be the big negative change, 2 the small negative change, 3 there in no change, 4 the small positive change and 5 be the big positive change
State statistics for relative differences.

```
tapply(D$reldiff,D$cat5,fivenum)
# the state   min      Q1       median   Q3       Max
#         1  -0.1676  -0.0554  -0.0386  -0.0250  -0.0181
#         2 -0.0177  -0.0137  -0.0097  -0.0041  -0.0010
#         3 -0.0009  -0.0003   0.0000   0.0003   0.0011
#         4   0.0011   0.0031   0.0083   0.0157   0.0234
#         5   0.0235   0.0334   0.0490   0.0776   0.4549
```

The state frequencies:

```
#  -2   -1   0   1   2
# 150  151 198 166 166
```

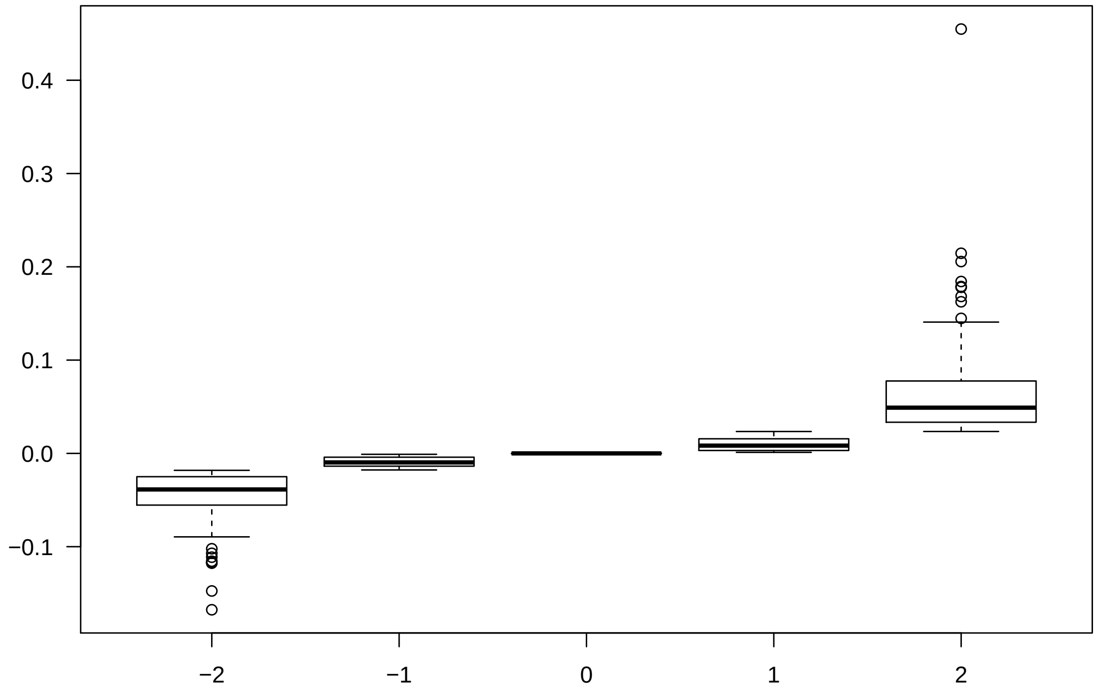

Figure 5.4: Diagnostique plot of discreditation - five states

5.3 Calculations of the transition matrices

The matrix of the absolute transition frequencies

F5

	-2	-1	0	1	2
1	63	28	0	23	35
2	22	54	9	43	23
3	0	6	182	10	0
4	26	39	6	56	39
5	39	24	0	34	69

The matrix of the relative transition frequencies, which is the transition probability matrix

```
P5
          -2    -1     0     1     2
    1  0.423 0.188 0.000 0.154 0.235
    2  0.146 0.358 0.060 0.285 0.152
    3  0.000 0.030 0.919 0.051 0.000
    4  0.157 0.235 0.036 0.337 0.235
    5  0.235 0.145 0.000 0.205 0.416
```

5.4 Five different estimation of the Q matrix

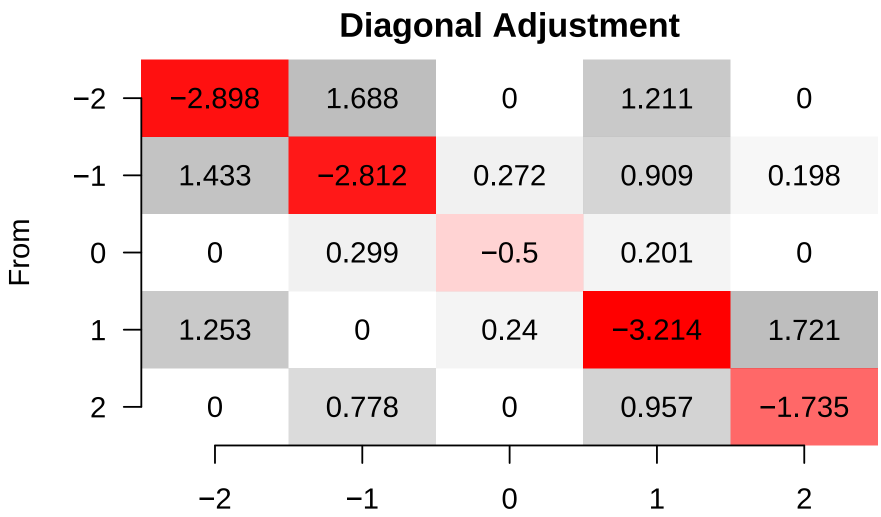

Figure 5.5: Q5_DA - BBB

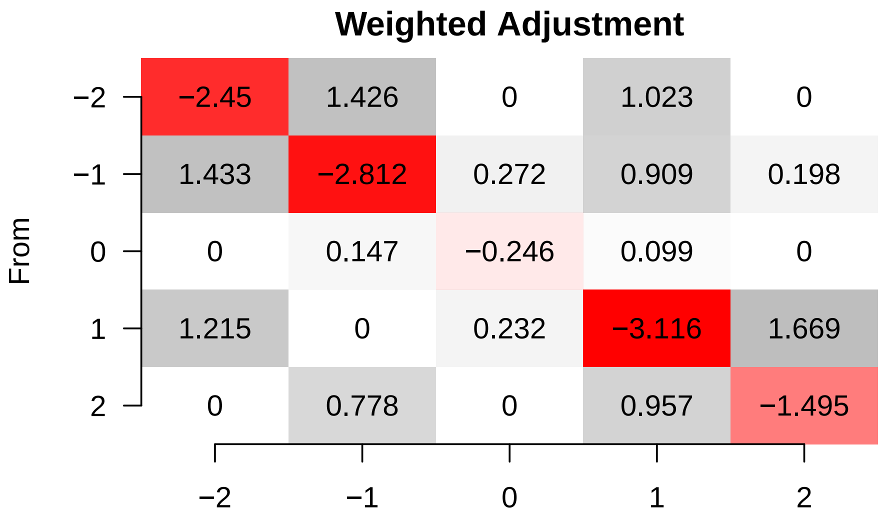

Figure 5.6: Q5_WA - BBB

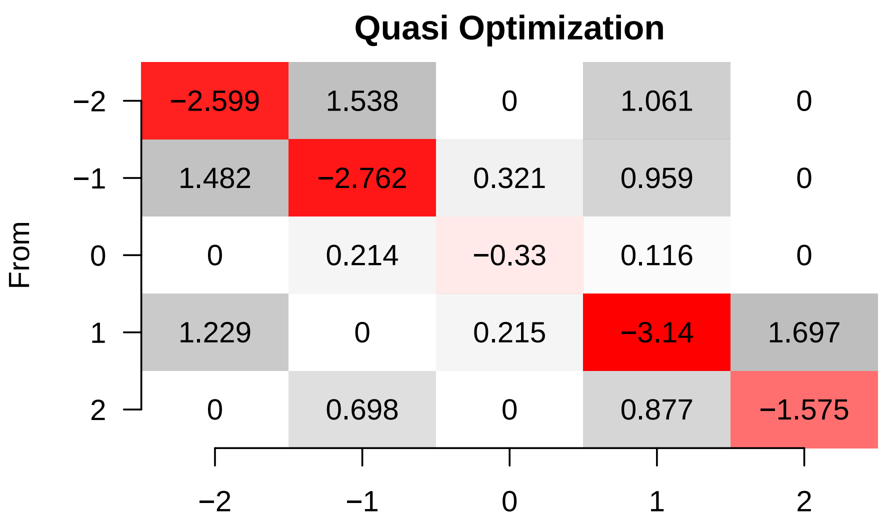

Figure 5.7: Q5_QO - BBB

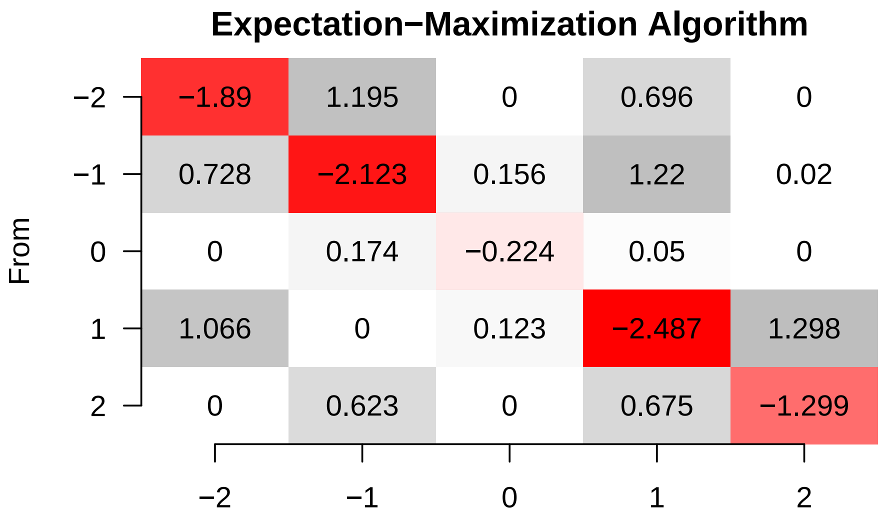

Figure 5.8: Q5_EM - BBB

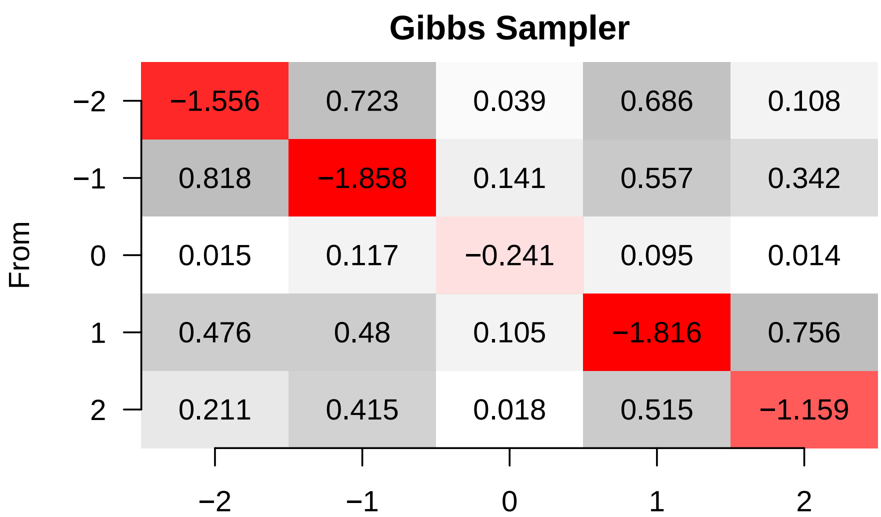

Figure 5.9: Q5_GS - BBB

The confidence 5 state Markov chain by method of Gibbs Sampling

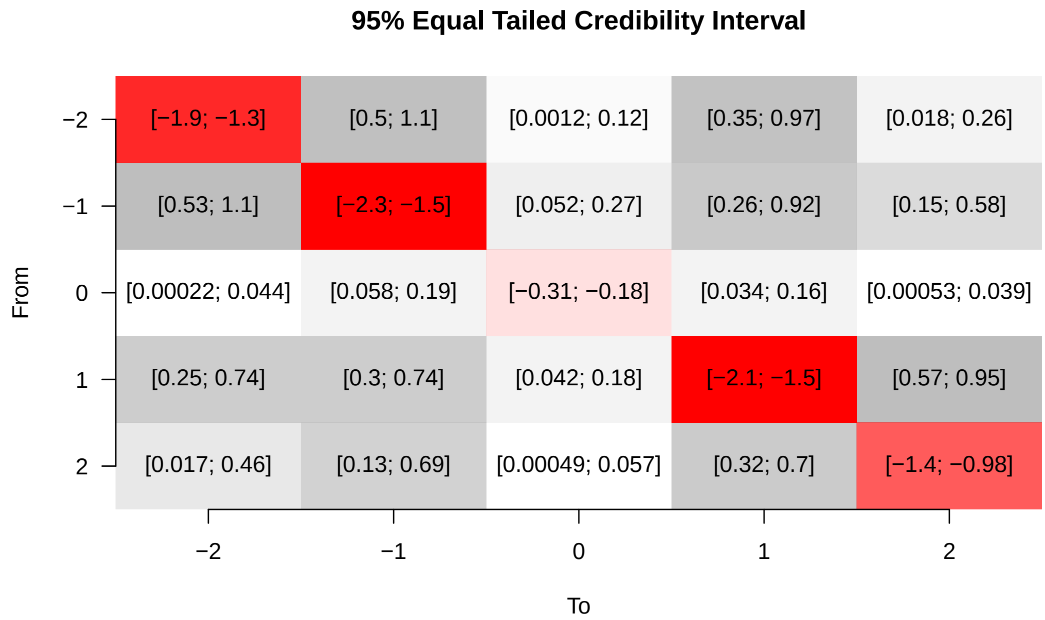

Figure 5.10: Credibility intervals for 5 states Markov chain - for "GS" based generator matrix

5.5 Examination of the estimations retention time

After doing same thing with having more states with always taking first state as the highest negative change and the last one in the highest positive change.
Now I will do the comparison between the methods, here I will give two examples of the 7 and 11 states case and in both the exterminated state is the first state, i.e. the highest negative change.

```
# summary
#    chain: 7 states
#    the exterminated state: "1"
#    the retention parameter by Q matrices estimated bey different methods
#       DA        WA        QO        EM        GS
#       -1.750890 -1.750890 -1.746555 -1.510837 -1.325919
#    the estimated parameter from the process path: 1.344
#    equal tailed 95% confidence intervall by GS method: [-1.667389, -1.063208]
```

```
# summary
#    chain: 11 states
#    the examinated state: "1"
#    the retention parameter by Q matrices estimated bey different methods
#       DA        WA        QO        EM        GS
#       -4.271130 -2.526609 -2.875514 -1.774952 -1.377196
#    the estimated parameter from the process path: 1.41176
#    equal tailed 95% confidence intervall by GS method: [-1.4495714, -0.9089183]
```

From this results we can see that, in both cases the Gibbs sampler method has the smallest error, and that the Diagonal adjustment method is the worst among all of them specially when number of states is becoming large.
Hence can rank the method starting from the best to the worst one as follow: GS method, EM method, QA method, WA method, DA method.

General Conclusion

Conclusion

From our previous work, we can derive several conclusions:

(1) There are sufficient and necessary conditions under which the transition Matrix P is embeddable in the partial cases: 2×2, 3×3 and the reversible square matrices.

(2) There are some conditions under which the transition matrix P is not embeddable.

(3) If a transition matrix P is embeddable then the generator matrix Q which satisfies $P = e^Q$ exists and is not necessarily unique.

(4) From the continuous Markov chain with transition matrices $\{P(t), t > 1\}$ we can obtain homogenous Markov chain with transition matrix P such that $P = P(1)$.

(5) We can estimate the maximum likelihood estimator $\hat{P}$ of the transition matrix P.

(6) If $\hat{P}$ is embeddable then we can find the maximum likelihood estimator $\hat{Q}$ of the generator matrix Q using the relation $\hat{P} = e^{\hat{Q}}$, and in case of the non-uniqueness, there is a method to choose the best one.

(7) If P is not embeddable then either $\hat{Q}$ doesn't exist, or it does exist and we can detriment by the methods: the diagonal adjustment, the weighted adjustment, the Quasi-Optimization, the EM algorithm and the Markov Chain Monte Carlo (Gibbs sampler).

(8) Using the MCMC(Gibbs sampler) method we can deal with the problem of the non-existence of the $\hat{Q}$.

(9) From the application in the R project, we can see that the estimator of the generator matrix obtained by the MCMC(Gibbs sampler) has the best fit.

Future work:

As we have seen that the embeddable problem was studied only at partial cases, and since obtaining the generator matrix from embeddable transition matrix is more valid and better than estimating it, it would be very interesting and worth to study the embedding problem and found the condition under which the $n \times n$ transition matrix for any n=4,5,. . . is embeddable.

List of Figures

References

[1] John Frank Charles Kingman. **The imbedding problem for finite Markov chains**. Probability Theory and Related Fields, **no. 1**:14–24, 1962. (12), (13), (15)

[2] Søren Johansen. **Some results on the imbedding problem for finite Markov chains**. Journal of the London Mathematical Society, **no. 2**:345–351, 1974. (14), (15)

[3] Philippe Carette. **Characterizations of Embeddable 3 3 Stochastic Matrices with a Negative Eigenvalue**. New York J. Math, **no. 120**, 1995. (15), (16)

[4] Chen Yong; Jianmin Chen. **On the imbedding problem for three-state time homogeneous Markov chains with coinciding negative eigenvalues**. Journal of Theoretical Probability, **24.4**:928–938, 2011. (16)

[5] Chen Jia. **A solution to the reversible embedding problem for finite Markov chains**. Statistics & Probability Letters, **116**:122–130, 2016. (17), (18), (20)

[6] Lijing Lin. **Roots of stochastic matrices and fractional matrix powers**. Diss. Manchester Institute for Mathematical Sciences, The University of Manchester, 2011. (19)

[7] Israel Robert B.; Jeffrey S. Rosenthal; and Jason Z. Wei. **Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings**. Mathematical Finance, **11.2**:245–265, 2001. (20), (23)

[8] Patrick Billingsley. Statistical inference for Markov processes, **Vol. 2**. University of Chicago Press, 1961. (21)

[9] Kreinin Alexander; Marina Sidelnikova. **Regularization algorithms for transition matrices**. Algo Research Quarterly, **4.1/2**:23–40, 2001. (24)

[10] Yasunari Inamura. **Estimating continuous time transition matrices from discretely observed data**. Bank of Japan, **06-07**, 2006. (25), (28)

[11] Bladt Mogens; and Michael Sorensen. **Statistical inference for discretely observed Markov jump processes**. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **67.3**:395–410, 2005. (25), (27), (28)

[12] Marius Pfeuffer, Greig Smith, and Goncalo dos Reis. ctmcd: Estimating the Parameters of a Continuous-Time Markov Chain from Discrete-Time Data, 2017. R package version 1.2. (30)

[13] Ephraim Hanks. ctmcmove: Modeling Animal Movement with Continuous-Time Discrete-Space Markov Chains, 2017. R package version 1.2.8. (30)

[14] Jon Fintzi. ECctmc: Simulation from Endpoint-Conditioned Continuous Time Markov Chains, 2017. R package version 0.2.4. (30)