UNIVERSITY OF CALGARY

Measure Change and Filtering

by

Jia Deng

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

APRIL, 2011

© Jia Deng 2011

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Measure Change and Filtering" submitted by Jia Deng in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

Supervisor, Dr. Robert J. Elliott
Haskayne School of Business

Dr. Alfred Lehar
Haskayne School of Business

Co-supervisor,
Dr. Alexandru Badescu
Department of Mathematics and Statistics

Dr. Jean-Francois Wen
Department of Economics

Dr. Anatoliy Swishchuk
Department of Mathematics and Statistics

Dr. Alexander Melnikov
University of Alberta

Dr. Antony Ware
Department of Mathematics and Statistics

27 April 2011
Date

Abstract

State-space models are widely used in engineering, biology, finance and many other fields. Often, there are two stochastic processes in a state-space model. One, the signal, is not observed directly and is said to be hidden, but is often observed through a second observation process. Usually, there is noise present in both the unobserved and observed processes. Hidden Markov models are one of the most popular state-space models. In hidden Markov models, the hidden process is a Markov process.

In a state-space model, the problem is how to estimate the hidden states and the parameters of the model, given the observations. In order to estimate the state and parameters simultaneously, I adopt the EM algorithm and a method called "change of measure". Some of these methods were introduced in one of Robert Elliott's papers in 1994. Later, in 2006, R. J. Elliott and W. P. Malcolm gave some improvements to the related smoother. In this thesis, I apply the "measure change" and the EM algorithm to the filtering problem for the Autoregressive hidden Markov model (ARHMM). Using extensions of the Viterbi algorithm, estimates of the hidden states and parameters are obtained for a Hidden Markov model where the observed process takes values in finite discrete state space.

Sometimes, the hidden signal and the observation process have nonlinear dynamics. I apply a measure change to obtain an estimate of the joint density of the hidden signal and the parameters of such a model where the hidden signal and observation process are scalar processes and have nonlinear dynamics.

In many practical cases, the noise in the observations is correlated and has some "memory". In this thesis, I also consider the several state-space models, where the signal is observed through a real valued process which is corrupted by fractional Gaussian noise. I derive the exact estimates and approximate recursive estimates for the hidden signal and the parameters, using the change of measure method.

Simulations and applications to some practical problems are carried out to demonstrate the performance of the algorithms.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Dr. Robert Elliott, and my co-supervisor, Dr. Alexandru Badescu, for their valuable advices and guidance during my studies at the University of Calgary. I also wish to thank Leung Lung Chan for his help with my study and life in Calgary.

Finally, I would like to thank my parents for their love and encouragement over the years.

Table of Contents

List of Figures

Chapter 1

INTRODUCTION TO HIDDEN MARKOV MODELS

1.1 Markov Processes

A stochastic process which satisfies Markov property is called a Markov process, i.e., a stochastic process is a Markov process if its past and future are conditionally independent given the present. Markov processes are named after the Russian mathematician Andrey Markov.

Definition 1.1. *Suppose $\{X_t, t \in [0,\infty)\}$ is a real valued stochastic process on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{\mathcal{F}_t\}$, $t \in [0,\infty)$ be the filtration generated by $\{X_s, 0 \leq s \leq t\}$ and write $\sigma(X_s)$ to be the σ-field generated by X_s. Denote $\mathcal{B}(\Re)$ to be the Borel field on $\Re$. Then, $\{X_t, t \in [0,\infty)\}$ is said to be a Markov process if*

1. *The stochastic process $\{X_t, t \in [0,\infty)\}$ is adapted to the filtration $\{\mathcal{F}_t\}$, i.e., X_t is $\mathcal{F}_t$-measurable for all t.*
2. *For any $0 \leq s \leq t < \infty$ and any $A \in \mathcal{B}(\Re)$, we have*

$$P(X_t \in A|\mathcal{F}_s) = P(X_t \in A|\sigma(X_s)).[44][45]$$

1.2 Markov Chain

Usually, a Markov chain means a discrete-time Markov process.

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. On $(\Omega, \mathcal{F}, \mathcal{P})$ we consider a sequence of random variables $X = \{X_n, n = 0, 1, 2, ...\}$ with a finite state space $S_X := (x_1, x_2, \ldots, x_N)$. Without loss of generality, the state space of X can be identified with the set $S =$

$\{e_1, e_2, \ldots, e_N\}$, where e_i is the unit vector with unity in the ith position and zero elsewhere. At each instant the process X may change its state from the current state to another state, or remain in the same state, according to a certain probability distribution. The changes of state are called transitions. The conditional probabilities $P(X_{n+1} = e_j | X_n = e_i) \triangleq p_{ji}(n)$, $e_i, e_j \in S$ are called transition probabilities, and the probabilities $P(X_n = e_i)$, $1 \leq i \leq N$, are called marginal probabilities. X is a Markov chain, if it satisfies the Markov property, that is

$$\begin{aligned} P(X_{n+1} = e_j | X_n = e_i, X_k = e_{i_k}, k = 1, 2, ..., n-1) &= P(X_{n+1} = e_j | X_n = e_i) \\ &= p_{ji}(n), \end{aligned}$$

$$\forall \quad 1 \leq i, j, i_k \leq N, \quad and \quad n \geq 0. \tag{1.2.1}$$

Write $\Pi_n = (p_{ji}(n))$, $1 \leq i, j \leq N$. Then,

$$X_{n+1} = \Pi_n X_n + M_{n+1}. \tag{1.2.2}$$

Note that

$$\begin{aligned} E[X_{n+1} | X_n] &= \sum_{i=1}^{N} P(X_{n+1} = e_i | X_n) e_i \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} < X_n, e_j > p_{ij}(n) e_i \\ &= \sum_{j=1}^{N} < X_n, e_j > \sum_{i=1}^{N} p_{ij}(n) e_i \\ &= \sum_{j=1}^{N} < X_n, e_j > \Pi_n e_j \\ &= \Pi_n X_n, \end{aligned}$$

where $< \cdot, \cdot >$ means the scalar product.

Then $E[M_{n+1} | X_n] = E[X_{n+1} - \Pi_n X_n | X_n] = 0$. So, $M = \{M_n, n = 1, 2, ...\} \in \mathcal{R}^N$ is a sequence of martingale increments.

In addition, if the Markov process is (time-) homogeneous, then

$$P(X_{n+1} = e_j | X_n = e_i, X_k = e_{i_k}, k = 1, 2, ..., n-1) = p_{ji},$$
$$\forall \quad 1 \leq i, j, i_k \leq N, \quad and \quad n \geq 0.$$

Write $\Pi = (p_{ji})$, $1 \leq i, j \leq N$, then

$$X_{n+1} = \Pi X_n + M_{n+1}. \tag{1.2.3}$$

Otherwise, it is nonhomogeneous. Homogeneous Markov processes form the most widely used class of Markov processes.

1.3 Continuous-time Markov Chain [36]

Suppose $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space. We consider a continuous-time Markov process $\{X_t, t \in [0, \infty)\}$ on $(\Omega, \mathcal{F}, \mathcal{P})$ with a finite state space $S = \{e_1, e_2, \ldots, e_N\} \in R^N$. Write $p_t^i = P(X_t = e_i)$, $1 \leq i \leq N$, and $p_t = (p_t^1, p_t^2, ..., p_t^N)'$. Suppose for some matrix $A_t = (a_{ji}(t))$, $t \geq 0$, p_t satisfies the forward Kolmogorov equation

$$\frac{dp_t}{dt} = A_t p_t. \tag{1.3.1}$$

The transition matrix $\Phi(t, s)$ associated with A_t is defined by

$$\frac{d\Phi(t,s)}{dt} = A_t \Phi(t,s), \quad \Phi(s,s) = I, \tag{1.3.2}$$
$$\frac{d\Phi(t,s)}{ds} = -A_t \Phi(t,s), \quad \Phi(t,t) = I, \tag{1.3.3}$$

where I is the identity matrix.

Then, if X is a Markov process,

$$\begin{aligned} E[X_t|\mathcal{F}_s] &= E[X_t|X_s] \\ &= \sum_{i=1}^{N} P(X_t = e_i|X_s)e_i \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} < X_s, e_j > P(X_t = e_i|X_s = e_j)e_i \\ &= \sum_{j=1}^{N} < X_s, e_j > \sum_{i=1}^{N} \Phi(t,s)e_i \\ &= \sum_{j=1}^{N} < X_s, e_j > \Phi(t,s)e_j \\ &= \Phi(t,s)X_s. \end{aligned}$$

Write $M_t := X_t - X_0 - \int_0^t A_r X_r dr$. For $0 \leq s \leq t$, we have

$$\begin{aligned} E[M_t - M_s|\mathcal{F}_s] &= E[X_t - X_s - \int_s^t A_r X_r dr|X_s] \\ &= \Phi(t,s)X_s - X_s - \int_s^t A_r \Phi(r,s) X_s dr \\ &= 0. \end{aligned}$$

We see that $\{M_t\}$, $t \geq 0$, is an $\mathcal{R}^N$-valued martingale process with respect to the filtration generated by $\{X_t\}$, $t \geq 0$.

Therefore, $\{X_t\}$, $t \geq 0$ has the following dynamics:

$$X_t = X_0 + \int_0^t A_r X_r dr + M_t \in \mathcal{R}^N. \tag{1.3.4}$$

1.4 Hidden Markov Models

Hidden Markov Models are one of the most widely used stochastic models in engineering, biology, finance and many other fields. Having introduced Markov processes, we shall give some descriptions of hidden Markov models.

According to L. R. Rabiner[22], a hidden Markov model (HMM) is doubly stochastic process, with one Markov process that is not observed directly (hidden), but could be observed through another noisy process.

Suppose the signal, or state, process $\{X_t\}$, is a Markov process which cannot be observed directly. X_t is often called a hidden state at time t. Information concerning $\{X_t\}$ is obtained from the observation process $\{Y_t\}$, which is influenced by the hidden state, and so gives some information about the hidden states. The state space of $\{X_t\}$, S_X, is taken as the set $S = \{e_1, e_2, \ldots, e_N\}$ of unit vectors. Denote the state space of $\{Y_t\}$ as S_Y. If S_Y is finite, it can be identified with the set $S_Y = \{f_1, f_2, ..., f_M\}$, where f_j is the unit vector with unity in the jth position and zero elsewhere. Write $c_{ji} = P(Y_t = f_j | X_t = i)$, $C = (c_{ji})$. Then the dynamics of X and Y can be written as:

$$X_{t+1} = AX_t + V_{t+1}, \tag{1.4.1}$$

$$Y_{t+1} = CX_{t+1} + W_{t+1}. \tag{1.4.2}$$

Here, A is the transition probability matrix of the Markov chain. V_t and W_t represent the driving noise and measurement noise respectively. Details of this model are given in Chapter 4.

Suppose a hidden Markov model is discrete in time, in the state, and continuous in the measurement space. For example, suppose the observation process is a scalar process $\{y_t\}$. Then the dynamics of the model can be expressed as:

$$X_{t+1} = AX_t + V_{t+1}, \tag{1.4.3}$$

$$y_{t+1} = < g, X_{t+1} > + < \sigma, X_{t+1} > w_{t+1}, \tag{1.4.4}$$

where A is the transition probability matrix of the Markov chain; g and σ are N-dimensional vectors; V_t and w_t are the driving noise and scalar measurement noise respectively. The filtering and estimation problems regarding this model are discussed in Chapter 2.

There are three basic problems associated with HMMs: [22]

(1) Given the parameters of the model and the output sequence, compute the probability that a particular output sequence is produced by the model.

(2) Given the parameters of the model, find the most likely sequence of hidden states that could have generated a given output sequence.

(3) Optimize the model parameters, so as to best describe how the observations have been produced. In other words, given an output sequence, find the most likely state transition probability matrix and output probabilities, that is, find the parameters of the HMM.

L. R. Rabiner gave some classical solutions to the above three problems. For problem (1), the forward-backward algorithm is often used; for problem (2), the Viterbi algorithm is often used; for problem (3), the Baum-Welch algorithm is often used. The details of these algorithms can be found in [22] and [23]. In this thesis, we mainly discuss the following problem:

Given an output sequence, find the parameters of the HMM and the most likely sequence of hidden states which could have generated the output.

We also discuss more general discrete-time models where the state space of the scalar hidden states $\{x_t\}$ and the scalar observations $\{y_t\}$ are both continuous. That is, we consider the dynamics:

$$x_{t+1} = f(x_t) + v_{t+1}, \tag{1.4.5}$$

$$y_{t+1} = g(x_{t+1}) + w_{t+1}, \tag{1.4.6}$$

where the function $f(\bullet)$ may be either linear or nonlinear, and x_t and y_t have a linear relationship, i.e., $g(\bullet)$ is a linear function.

1.5 History and Applications of a Hidden Markov Model

Hidden Markov Models, HMMs, were first described by Leonard E. Baum and other authors in 1960s. HMMs initially appeared in some statistical papers. Later, in the second half of the 1970s, L. R. Rabiner induced HMMs into speech recognition. This was an important application of HMMs.[23]

In the late 1980s, HMMs began to be used in computational biology and bioinfomatics, to analyze biological sequences, especially the DNA sequence. Since then, many biological models based on HMMs have been introduced.

In the last 20 years, HMMs have become important models in temporal pattern recognition, such as speech, handwriting, gesture and image recognition, classification, navigation, musical score following, partial discharges and bioinformatics. They are also useful tools in other fields, such as finance and social science.

In this thesis, we shall discuss two applications of HMMs: tracking and classification.

Chapter 2

A FILTER FOR A SIMPLE LINEAR HIDDEN MARKOV MODEL

2.1 Introduction

In this chapter, we consider a discrete time, discrete finite state Markov chain, observed through a real valued function whose values are corrupted by Gaussian noise, and the relation between the hidden states and the observations is linear. The state space of the observations is continuous. This is the simplest, but most widely used hidden Markov model. Many problems in bioinformatics, finance and engineering, such as tracking, navigation, pattern recognition, and so on, can be modeled as such a model.

In order to estimate the hidden states and the parameters of the HMM, we derive the recursive estimates based on a "change of measure" and the Expectation Maximization(EM) algorithm. This method was introduced in one of Robert J. Elliott's papers in 1994 [28]. Later in 2006, R. J. Elliott and W. P. Malcolm gave some improvement to the smoother [37].

In the following sections, I first give a description of this model. Then, a new measure is constructed, under which the observations are $N(0,1)$ i.i.d random variables. Working under the new measure, recursive estimates are obtained for the states of the Markov chain, for the number of jumps from one state to another, for the occupation time in any state, and for processes related to the observations. Using the EM algorithm, estimates of all the parameters of the model are obtained, including the variance of the Gaussian noise in the observations. In the last part of the chapter, simulations are given which demonstrate the effectiveness of the method. I also apply this method to the problem of classification of DNA copy numbers.

2.2 A Simple Linear Hidden Markov Model

Assume a finite state time-homogeneous Markov chain $X = \{X_t, t = 0, 1, ...\}$ is defined on probability space $(\Omega, \mathcal{F}, P)$. Without loss of generality, the state process of X can be identified with the set of unit vectors

$$S = \{e_1, e_2, ...e_N\},$$

where $e_i = (0, 0, ..., 0, 1, 0, ..., 0)' \in R^N$.

We assume the chain is time-homogeneous and write

$$\begin{aligned} a_{j,i} &\triangleq P(X_{t+1} = e_j | X_t = e_i) \\ &= P(X_1 = e_j | X_0 = e_i). \end{aligned} \tag{2.2.1}$$

Then $A = (a_{j,i})$, $1 \le i \le N$, $1 \le j \le N$, is the matrix of transition probabilities.

Lemma 2.1. *Write $\mathcal{F}_t = \sigma\{X_0, X_1, ..., X_t\}$, and $\mathbb{F}$ for the filtration $\{\mathcal{F}_t\}$, then*

$$X_t = AX_{t-1} + M_t, \tag{2.2.2}$$

where M_t is a $(P, \mathcal{F})$ martingale increment.

Proof.

$$
\begin{aligned}
E[X_t|X_{t-1}] &= \sum_{i=1}^{N} P(X_t = e_i|X_{t-1})e_i \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} < X_{t-1}, e_j > a_{ij}e_i \\
&= \sum_{j=1}^{N} < X_{t-1}, e_j > \sum_{i=1}^{N} a_{ij}e_i \\
&= \sum_{j=1}^{N} < X_{t-1}, e_j > Ae_j \\
&= AX_{t-1}.
\end{aligned}
$$

Then

$$
\begin{aligned}
E[M_t|\mathcal{F}_{t-1}] &= E[X_t - AX_{t-1}|\mathcal{F}_{t-1}] \\
&= AX_{t-1} - AX_{t-1} \\
&= 0.
\end{aligned}
$$

□

We suppose the process X is not observed directly; rather, it is observed through another function, whose values are corrupted by Gaussian noise. All functions of X are linear. We suppose the observations $\{y_t\}$ have the form

$$y_t =< g, X_t > + < d, X_t > w_t. \tag{2.2.3}$$

Here g and d are both N dimensional vectors, and $w = \{w_t, t = 0, 1, ...\}$ is a sequence of $N(0, 1)$ independent, identically distributed (i.i.d.) random variables.

Consider a probability measure $\bar{P}$ on the measurable space $(\Omega, \mathcal{F})$ such that, under $\bar{P}$,

(1) The process X is a finite state Markov chain with transition matrix A,

(2) The observation $\{y_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables.

We call the measure $\bar{P}$ a "reference" probability.

We now construct a probability P, such that, under P, the process X is still a finite state Markov chain with transition matrix A, and $\{w_t : w_t = \frac{y_t - <g,X_t>}{<d,X_t>}\}$ is a sequence of $N(0,1)$ i.i.d. random variables.

Write $\mathcal{Y}_t = \sigma\{y_0, y_1, ..., y_t\}$, and $\mathcal{G}_t = \sigma\{X_0, y_0, X_1, y_1, ..., X_t, y_t\}$. Then the "histories", or filtrations, of the X, y and (X, y) processes are $\{\mathcal{F}_t\}$, $\{\mathcal{Y}_t\}$, $\{\mathcal{G}_t\}$.

Write

$$\begin{aligned} \bar{\lambda}_l &= \frac{\phi(w_l)}{< d, X_l > \phi(y_l)} \\ &= \frac{\phi(\frac{y_l - <g,X_l>}{<d,X_l>})}{< d, X_l > \phi(y_l)}, &\quad (2.2.4) \\ \bar{\Lambda}_0 &= 1, &\quad (2.2.5) \\ \bar{\Lambda}_t &= \prod_{l=1}^{t} \bar{\lambda}_l, \quad t = 1, 2, 3, ... \quad , &\quad (2.2.6) \end{aligned}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})$.

Definition 2.1. *Define P by putting*

$$\frac{dP}{d\bar{P}}|_{\mathcal{G}_t} = \bar{\Lambda}_t. \quad (2.2.7)$$

Theorem 2.1. *Under P, $\{w_t\}$ is a sequence of $N(0,1)$ random variables.*

Proof.

$$\begin{aligned} P(w_t \le a|\mathcal{G}_{t-1}) &= E[I(w_t \le a)|\mathcal{G}_{t-1}] \\ &= \frac{\bar{E}[\bar{\Lambda}_t I(w_t \le a)|\mathcal{G}_{t-1}]}{\bar{E}[\bar{\Lambda}_t|\mathcal{G}_{t-1}]} \\ &= \frac{\bar{E}[\bar{\lambda}_t I(w_t \le a)|\mathcal{G}_{t-1}]}{\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}]}. \end{aligned}$$

$$\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}] = \bar{E}[\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1} \vee X_t]|\mathcal{G}_{t-1}].$$

The inner expectation

$$\begin{aligned}\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1} \vee X_t] &= \int_{-\infty}^{\infty} \frac{\phi(\frac{y_t - <g, X_t>}{<d, X_t>})}{< d, X_t > \phi(y_t)} \phi(y_t) dy_t \\ &= \int_{-\infty}^{\infty} \phi(w_t) dw_t \\ &= 1.\end{aligned}$$

Then

$$\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}] = 1.$$

Similarly,

$$\begin{aligned}\bar{E}[\bar{\lambda}_t I(w_t \le a)|\mathcal{G}_{t-1} \vee X_t] &= \int_{-\infty}^{\infty} \frac{\phi(\frac{y_t - <g, X_t>}{<d, X_t>})}{< d, X_t > \phi(y_t)} I(w_t \le a) \phi(y_t) dy_t \\ &= \int_{-\infty}^{\infty} \phi(w_t) I(w_t \le a) dw_t.\end{aligned}$$

$$\begin{aligned}\bar{E}[\bar{\lambda}_t I(w_t \le a)|\mathcal{G}_{t-1}] &= \bar{E}[\bar{E}[\bar{\lambda}_t I(w_t \le a)|\mathcal{G}_{t-1} \vee X_t]|\mathcal{G}_{t-1}] \\ &= \int_{-\infty}^{\infty} \phi(w_t) I(w_t \le a) dw_t.\end{aligned}$$

So,$P(w_t \le a|G_{t-1}) = P(w_t \le a) = \int_{-\infty}^{\infty} \phi(w_t) I(w_t \le a) dw_t$, and the result follows.

□

Remark: Under P, the process X is still a finite state Markov chain with transition matrix A.

Proof.

$$\begin{aligned}E[< X_t, e_m > |\mathcal{G}_{t-1}] &= \frac{\bar{E}[\bar{\Lambda}_t < X_t, e_m > |\mathcal{G}_{t-1}]}{\bar{E}[\bar{\Lambda}_t|\mathcal{G}_{t-1}]} \\ &= \frac{\bar{E}[\bar{\lambda}_t < X_t, e_m > |\mathcal{G}_{t-1}]}{\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}]}.\end{aligned}$$

$$\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}] = \bar{E}[\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1} \vee X_t]|\mathcal{G}_{t-1}].$$

The inner expectation

$$\begin{aligned}
\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1} \vee X_t] &= \bar{E}[\frac{\phi(\frac{y_t-<g,X_t>}{<d,X_t>})}{<d,X_t>\phi(y_t)}|\mathcal{G}_{t-1} \vee X_t] \\
&= \int_{-\infty}^{\infty} \frac{\phi(\frac{y_t-<g,X_t>}{<d,X_t>})}{<d,X_t>\phi(y_t)}\phi(y_t)dy_t \\
&= \int_{-\infty}^{\infty} \phi(w_t)dw_t \\
&= 1.
\end{aligned}$$

Then

$$\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}] = 1.$$

Again, using double conditioning, we have

$$\begin{aligned}
\bar{E}[\bar{\lambda}_t <X_t,e_m> |\mathcal{G}_{t-1} \vee X_t] &= <X_t,e_m> \bar{E}[\frac{\phi(\frac{y_t-<g,X_t>}{<d,X_t>})}{<d,X_t>\phi(y_t)}|\mathcal{G}_{t-1} \vee X_t] \\
&= <X_t,e_m> \int_{-\infty}^{\infty} \frac{\phi(\frac{y_t-<g,X_t>}{<d,X_t>})}{<d,X_t>\phi(y_t)}\phi(y_t)dy_t \\
&= <X_t,e_m> \int_{-\infty}^{\infty} \phi(w_t)dw_t \\
&= <X_t,e_m>.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\bar{E}[\bar{\lambda}_t <X_t,e_m> |\mathcal{G}_{t-1}] &= \bar{E}[\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1} \vee X_t]|\mathcal{G}_{t-1}] \\
&= \bar{E}[<X_t,e_m> |\mathcal{G}_{t-1}] \\
&= \bar{E}[<X_t,e_m> |X_{t-1}] \\
&= <AX_{t-1},e_m>.
\end{aligned}$$

So,

$$E[<X_t,e_m> |\mathcal{G}_{t-1}] = <AX_{t-1},e_m>. \tag{2.2.8}$$

Therefore, under P, X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{j,i})$. □

2.3 Estimation of States and Parameters

The results were provided in [28] and [37].

2.3.1 Recursive Estimation

First of all, we shall describe how to estimate the hidden states, given the observations $\{y_t, t = 0, 1, 2, ...\}$.

Given $\mathcal{Y}_t$, write

$$\hat{X}_t = E[X_t|\mathcal{Y}_t]. \tag{2.3.1}$$

Using a version of Bayes' rule [36], we have

$$E[X_t|\mathcal{Y}_t] = \frac{\bar{E}[\bar{\Lambda}_t X_t|\mathcal{Y}_t]}{\bar{E}[\bar{\Lambda}_t|\mathcal{Y}_t]}. \tag{2.3.2}$$

Write

$$q_t = \bar{E}[\bar{\Lambda}_t X_t|\mathcal{Y}_t] \tag{2.3.3}$$

and

$$D(y_t) = \begin{pmatrix} \frac{\phi(\frac{y_t - g_1}{d_1})}{d_1\phi(y_t)} & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{\phi(\frac{y_t - g_N}{d_N})}{d_N\phi(y_t)} \end{pmatrix}. \tag{2.3.4}$$

Theorem 2.2. *The probability vector q_t is computed by the recursion*

$$q_t = D(y_t)Aq_{t-1}. \tag{2.3.5}$$

Proof.

$$\begin{aligned}
q_t &= \bar{E}[\bar{\Lambda}_t X_t|\mathcal{Y}_t] \\
&= \bar{E}[\bar{\Lambda}_{t-1}\frac{\phi(\frac{y_t-<g,X_t>}{<d,X_t>})}{<d,X_t>\phi(y_t)}X_t|\mathcal{Y}_t] \\
&= \bar{E}[\bar{\Lambda}_{t-1}\frac{\phi(\frac{y_t-<g,X_t>}{<d,X_t>})}{<d,X_t>\phi(y_t)}\sum_{i=1}^{N}<X_t,e_i>X_t|\mathcal{Y}_t] \\
&= \sum_{i=1}^{N}\bar{E}[\bar{\Lambda}_{t-1}<AX_{t-1}+M_t,e_i>|\mathcal{Y}_t]\frac{\phi(\frac{y_t-g_i}{d_i})}{d_i\phi(y_t)}e_i \\
&= \sum_{i=1}^{N}<\bar{E}[\bar{\Lambda}_{t-1}AX_{t-1}|\mathcal{Y}_{t-1}],e_i>\frac{\phi(\frac{y_t-g_i}{d_i})}{d_i\phi(y_t)}e_i \\
&= D(y_t)Aq_{t-1}.
\end{aligned}$$

□

Note that

$$\begin{aligned}
<q_t,\mathbf{1}> &= <\bar{E}[\bar{\Lambda}_t X_t|\mathcal{Y}_t],\mathbf{1}> \\
&= \bar{E}[\bar{\Lambda}_t<X_t,\mathbf{1}>|\mathcal{Y}_t] \\
&= \bar{E}[\bar{\Lambda}_t|\mathcal{Y}_t].
\end{aligned}$$

So,

$$\hat{X}_t = \frac{q_t}{<q_t,\mathbf{1}>}. \tag{2.3.6}$$

In order to estimate the parameters in this model, we must estimate several random processes. In the remainder of this section, we shall introduce these processes and derive recursive estimates based on these processes.

The first process is the counting process for the number of state transitions e_i to e_j. Denote this process by $N_t^{(j,i)}$.

Then

$$N_t^{(j,i)} \triangleq \sum_{l=1}^{t}<X_{l-1},e_i><X_l,e_j>. \tag{2.3.7}$$

Write

$$\sigma(N_t^{(j,i)}X_t) \triangleq \bar{E}[\bar{\Lambda}_t N_t^{(j,i)} X_t|\mathcal{Y}_t]. \tag{2.3.8}$$

Theorem 2.3. *The vector $\sigma(N_t^{(j,i)}X_t)$ is computed by the recursion*

$$\sigma(N_t^{(j,i)}X_t) = D(y_t)A\sigma(N_{t-1}^{(j,i)}X_{t-1}) + < q_{t-1}, e_i > D(y_t)_{j,j}a_{j,i}e_j. \tag{2.3.9}$$

Proof.

$$\begin{aligned}
& \sigma(N_t^{(j,i)}X_t) \\
= \; & \bar{E}[\bar{\Lambda}_t N_t^{(j,i)} X_t|\mathcal{Y}_t] \\
= \; & \bar{E}[\bar{\Lambda}_{t-1}\bar{\lambda}_t(N_{t-1}^{(j,i)} + < X_{t-1}, e_i >< X_t, e_j >)X_t|\mathcal{Y}_t] \\
= \; & \bar{E}[\bar{\Lambda}_{t-1}\bar{\lambda}_t(N_{t-1}^{(j,i)} + < X_{t-1}, e_i >< X_t, e_j >)\sum_{l=1}^{N} < X_t, e_l > X_t|\mathcal{Y}_t] \\
= \; & \sum_{l=1}^{N} < \bar{E}[\bar{\Lambda}_{t-1}\bar{\lambda}_{t,l}N_{t-1}^{(j,i)}(AX_{t-1} + M_t)|\mathcal{Y}_t], e_l > e_l + \\
& \bar{E}[\bar{\Lambda}_{t-1}\bar{\lambda}_{t,j} < X_{t-1}, e_i >< AX_{t-1} + M_t, e_j > |\mathcal{Y}_t]e_j \\
= \; & \sum_{l=1}^{N} < \bar{E}[\bar{\Lambda}_{t-1}N_{t-1}^{(j,i)}AX_{t-1}|\mathcal{Y}_{t-1}], e_l > \bar{\lambda}_{t,l}e_l + < \bar{E}[\bar{\Lambda}_{t-1}X_{t-1}|\mathcal{Y}_t], e_i > \bar{\lambda}_{t,j}a_{j,i}e_j \\
= \; & D(y_t)A\sigma(N_{t-1}^{(j,i)}X_{t-1}) + < q_{t-1}, e_i > D(y_t)_{j,j}a_{j,i}e_j.
\end{aligned}$$

□

Now $< X_t, \mathbf{1} >= 1$, so

$$\begin{aligned}
< \sigma(N_t^{(j,i)}, X_t), \mathbf{1} > &= < \bar{E}[\bar{\Lambda}_t N_t^{(j,i)} X_t|\mathcal{Y}_t], \mathbf{1} > \\
&= \bar{E}[< \bar{\Lambda}_t N_t^{(j,i)} X_t, \mathbf{1} > |\mathcal{Y}_t] \\
&= \bar{E}[\bar{\Lambda}_t N_t^{(j,i)} < X_t, \mathbf{1} > |\mathcal{Y}_t] \\
&= \bar{E}[\bar{\Lambda}_t N_t^{(j,i)}|\mathcal{Y}_t], \qquad (2.3.10)
\end{aligned}$$

and $\bar{E}[\bar{\Lambda}_t N_t^{(j,i)}|\mathcal{Y}_t] = \sigma(N_t^{(j,i)})$ by definition.

The second process J_t^i is defined to be the cumulative sojourn time spent by the process X in state e_i.

Then

$$J_t^i \triangleq \sum_{l=1}^{t} < X_l, e_i > . \tag{2.3.11}$$

Write

$$\sigma(J_t^i X_t) \triangleq \bar{E}[\bar{\Lambda}_t J_t^i X_t | \mathcal{Y}_t]. \tag{2.3.12}$$

Theorem 2.4. *The vector $\sigma(J_t^i X_t)$ is computed by the recursion*

$$\sigma(J_t^i X_t) = D(y_t) A \sigma(J_{t-1}^i X_{t-1}) + < A q_{t-1}, e_i > D(y_t)_{i,i} e_i. \tag{2.3.13}$$

Proof.

$$\begin{aligned}
& \sigma(J_t^i X_t) \\
= \ & \bar{E}[\bar{\Lambda}_t J_t^i X_t | \mathcal{Y}_t] \\
= \ & \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_t (J_{t-1}^i + < X_t, e_i >) X_t | \mathcal{Y}_t] \\
= \ & \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_t (J_{t-1}^i + < X_t, e_i >) \sum_{l=1}^{N} < X_t, e_l > X_t | \mathcal{Y}_t] \\
= \ & \sum_{l=1}^{N} < \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_{t,l} J_{t-1}^i (A X_{t-1} + M_t) | \mathcal{Y}_t], e_l > e_l + \\
& \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_{t,i} < A X_{t-1} + M_t, e_i > | \mathcal{Y}_t] e_i \\
= \ & \sum_{l=1}^{N} < \bar{E}[\bar{\Lambda}_{t-1} J_{t-1}^i A X_{t-1} | \mathcal{Y}_{t-1}], e_l > \bar{\lambda}_{t,l} e_l + < \bar{E}[\bar{\Lambda}_{t-1} A X_{t-1} | \mathcal{Y}_t], e_i > \bar{\lambda}_{t,i} e_i \\
= \ & D(y_t) A \sigma(J_{t-1}^i X_{t-1}) + < A q_{t-1}, e_i > D(y_t)_{i,i} e_i.
\end{aligned}$$

□

As before, $\sigma(J_t^i) =< \sigma(J_t^i, X_t), \mathbf{1} >$.

Finally, we define G_t^i to be the observation variance and drift.

$$G_t^i \triangleq \sum_{l=1}^{t} f(y_l) < X_l, e_i >, \tag{2.3.14}$$

where $f(y_l)$ is any function of y_l.

Write

$$\sigma(G_t^i X_t) \triangleq \bar{E}[\bar{\Lambda}_t G_t^i X_t | \mathcal{Y}_t]. \tag{2.3.15}$$

Theorem 2.5. *The probability vector $\sigma(G_t^i X_t)$ is computed by the recursion*

$$\sigma(G_t^i X_t) = D(y_t) A \sigma(G_{t-1}^i X_{t-1}) + < A q_{t-1}, e_i > f(y_t) D(y_t)_{i,i} e_i. \tag{2.3.16}$$

Proof.

$$\begin{aligned}
& \sigma(G_t^i X_t) \\
= \; & \bar{E}[\bar{\Lambda}_t G_t^i X_t | \mathcal{Y}_t] \\
= \; & \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_t (G_{t-1}^i + f(y_t) < X_t, e_i >) X_t | \mathcal{Y}_t] \\
= \; & \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_t (G_{t-1}^i + f(y_t) < X_t, e_i >) \sum_{l=1}^{N} < X_t, e_l > X_t | \mathcal{Y}_t] \\
= \; & \sum_{l=1}^{N} \{ \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_{t,l} G_{t-1}^i < X_t, e_l > e_l | \mathcal{Y}_t] + \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_{t,l} f(y_t) < X_t, e_i >< X_t, e_l > e_l | \mathcal{Y}_t] \} \\
= \; & \sum_{l=1}^{N} < \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_{t,l} G_{t-1}^i (A X_{t-1} + M_t) | \mathcal{Y}_t], e_l > e_l + \\
& \bar{E}[\bar{\Lambda}_{t-1} \bar{\lambda}_{t,i} f(y_t) < A X_{t-1} + M_t, e_i > | \mathcal{Y}_t] e_i \\
= \; & \sum_{l=1}^{N} < \bar{E}[\bar{\Lambda}_{t-1} G_{t-1}^i A X_{t-1} | \mathcal{Y}_{t-1}], e_l > \bar{\lambda}_{t,l} e_l + < \bar{E}[\bar{\Lambda}_{t-1} A X_{t-1} | \mathcal{Y}_t], e_i > f(y_t) \bar{\lambda}_{t,i} e_i \\
= \; & \sum_{l=1}^{N} < A \sigma(G_{t-1}^i X_{t-1}), e_l > \bar{\lambda}_{t,l} e_l + < A q_{t-1}, e_i > f(y_t) \bar{\lambda}_{t,i} e_i \\
= \; & D(y_t) A \sigma(G_{t-1}^i X_{t-1}) + < A q_{t-1}, e_i > f(y_t) D(y_t)_{i,i} e_i.
\end{aligned}$$

□

Again, $\sigma(G_t^i) =< \sigma(G_t^i, X_t), \mathbf{1} >$.

2.3.2 Estimation of Parameters

Having introduced the three processes in the above section, we use Expectation Maximization (EM) algorithm to re-estimate the parameters in this HMM.

The EM algorithm has two main steps.

(1) E-step: Take the expectation of the Log-likelihood function, given the observations up to time t.

$$Q_t(\theta,\theta^*) = E_{\theta^*}[L_{\theta^*}(\theta)|\mathcal{Y}_t], \tag{2.3.17}$$

where θ is the true value of the parameters, and θ^* is an estimate of the parameters at time $t-1$.

(2) M-step: Maximize the above expectation in equation (2.3.17) with respect to θ.

Here, we take the Log-likelihood function to be the logarithm of the Radon-Nikodym derivative of the new probability measure with respect to the old, see [36]:

$$L_{\theta^*}(\theta) = \log \frac{dP_\theta}{dP_{\theta^*}}. \tag{2.3.18}$$

Then,

$$Q_t(\theta,\theta^*) = E_{\theta^*}[\log \frac{dP_\theta}{dP_{\theta^*}}|\mathcal{Y}_t]. \tag{2.3.19}$$

Recall that under $P_{\theta^*} = P$, X is a Markov chain with transition matrix $A = (a_{j,i})$. Now we shall introduce another probability measure $P_{\hat{\theta}}$, so that under $P_{\hat{\theta}}$, X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{j,i})$.

Theorem 2.6. *Define* $\Lambda_0 = 1$, $\Lambda_t = \prod_{l=1}^{t}(\sum_{i,j=1}^{N}(\frac{\hat{a}_{j,i}}{a_{j,i}}) < X_l, e_j >< X_{l-1}, e_i >)$ *for* $t \geq 1$. *Define* $P_{\hat{\theta}}$ *by putting* $\frac{dP_{\hat{\theta}}}{dP_{\theta^*}}|_{\mathcal{F}_t} = \Lambda_t$. *Then under* $P_{\hat{\theta}}$, X *is a Markov chain with transition matrix* $\hat{A} = (\hat{a}_{j,i})$. *[36][30]*

Proof.

$$\begin{aligned} E_{\hat{\theta}}[< X_t, e_m > |\mathcal{F}_{t-1}] &= \frac{E[\Lambda_t < X_t, e_m > |\mathcal{F}_{t-1}]}{E[\Lambda_t|\mathcal{F}_{t-1}]} \\ &= \frac{\Lambda_{t-1}E[\lambda_t < X_t, e_m > |\mathcal{F}_{t-1}]}{\Lambda_{t-1}E[\lambda_t|\mathcal{F}_{t-1}]} \\ &= \frac{E[\sum_{i,j=1}^N (\frac{\hat{a}_{j,i}}{a_{j,i}}) < X_t, e_j >< X_{t-1}, e_i > < X_t, e_m > |\mathcal{F}_{t-1}]}{E[\sum_{i,j=1}^N (\frac{\hat{a}_{j,i}}{a_{j,i}}) < X_t, e_j >< X_{t-1}, e_i >|\mathcal{F}_{t-1}]}. \end{aligned}$$

$$\begin{aligned} & E[\sum_{i,j=1}^N (\frac{\hat{a}_{j,i}}{a_{j,i}}) < X_t, e_j >< X_{t-1}, e_i >|\mathcal{F}_{t-1}] \\ = & \sum_{i=1}^N E[(\sum_{j=1}^N (\frac{\hat{a}_{j,i}}{a_{j,i}})a_{j,i}) < X_{t-1}, e_i > |\mathcal{F}_{t-1}] \\ = & E[\sum_{i=1}^N < X_{t-1}, e_i >] \\ = & 1. \end{aligned}$$

$$\begin{aligned} & E[(\sum_{i,j=1}^N (\frac{\hat{a}_{j,i}}{a_{j,i}}) < X_t, e_j >< X_{t-1}, e_i >) < X_t, e_m > |\mathcal{F}_{t-1}] \\ = & E[\sum_{i=1}^N (\frac{\hat{a}_{m,i}}{a_{m,i}}) < X_t, e_m >< X_{t-1}, e_i >|\mathcal{F}_{t-1}] \\ = & E[\sum_{i=1}^N \hat{a}_{m,i} < X_{t-1}, e_i >]. \end{aligned}$$

So,

$$P_{\hat{\theta}}(X_t = e_m|X_{t-1} = e_i) = E_{\hat{\theta}}[< X_t, e_m > |X_{t-1} = e_i] = \hat{a}_{m,i}.$$

Therefore, X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{j,i})$. ☐

Theorem 2.7. *Given the observations up to time t, the EM estimation of $a_{j,i}$ is*

$$\hat{a}_{j,i} = \frac{\sigma(N_t^{(j,i)})}{\sigma(J_{t-1}^i)}. \tag{2.3.20}$$

Proof.

$$\frac{dP_{\hat{\theta}}}{dP_{\theta^*}} = \Lambda_t = \prod_{l=1}^{t} (\sum_{i,j=1}^{N} (\frac{\hat{a}_{j,i}}{a_{j,i}}) < X_l, e_j >< X_{l-1}, e_i >).$$

Then,

$$\begin{aligned}
L_{\theta^*}(\hat{\theta}) &= \log \frac{dP_{\hat{\theta}}}{dP_{\theta^*}} \\
&= \sum_{l=1}^{t} \sum_{i,j=1}^{N} < X_l, e_j >< X_{l-1}, e_i > (\log \hat{a}_{j,i} - \log a_{j,i}) \\
&= \sum_{i,j=1}^{N} N_t^{(j,i)} \log \hat{a}_{j,i} + R(a).
\end{aligned}$$

$$Q_t(\hat{\theta}, \theta^*) = E[L_{\theta^*}(\hat{\theta})|\mathcal{Y}_t] = \sum_{i,j=1}^{N} E[N_t^{(j,i)}|\mathcal{Y}_t] \log \hat{a}_{j,i} + \breve{R}(a).$$

Notice that $\sum_{j=1}^{N} \hat{a}_{j,i} = 1$. Write γ for the Lagrange multiplier and put

$$l(\hat{a}, \gamma) = \sum_{i,j=1}^{N} E[N_t^{(j,i)}|\mathcal{Y}_t] \log \hat{a}_{j,i} + \breve{R}(a) + \gamma(\sum_{j=1}^{N} \hat{a}_{j,i} - 1). \tag{2.3.21}$$

Setting the derivative of $l(\hat{a}, \gamma)$, in $\hat{a}_{j,i}$ and γ, to be 0, we have

$$\frac{1}{\hat{a}_{j,i}} E[N_t^{(j,i)}|\mathcal{Y}_t] + \gamma = 0,$$
$$\sum_{j=1}^{N} \hat{a}_{j,i} = 1.$$

Solving the above two equations, we get

$$\begin{aligned}
\gamma &= -\sum_{j=1}^{N} E[N_t^{(j,i)}|\mathcal{Y}_t] = -E[J_{t-1}^i|\mathcal{Y}_t], \\
\hat{a}_{j,i} &= \frac{E[N_t^{(j,i)}|\mathcal{Y}_t]}{E[J_{t-1}^i|\mathcal{Y}_t]} = \frac{\sigma(N_t^{(j,i)})}{\sigma(J_{t-1}^i)}.
\end{aligned}$$

$\square$

Theorem 2.8. *Given the observations up to time t, the EM estimation of g_i is*

$$\hat{g}_i = \frac{\sigma(G_t^i(y))}{\sigma(J_t^i)}, \tag{2.3.22}$$

where $f(y_l) = y_l$ for $G_t^i(y)$.

Proof. Define

$$\begin{aligned} \Lambda_{g0} &= 1, \\ \Lambda_{gt} &= \frac{\prod_{l=1}^{t} \frac{\phi(\frac{y_l - <\hat{g}, X_l>}{<d, X_l>})}{<d, X_l>\phi(y_l)}}{\prod_{l=1}^{t} \frac{\phi(\frac{y_l - <g, X_l>}{<d, X_l>})}{<d, X_l>\phi(y_l)}}. \end{aligned}$$

Set $P_{g^*} = P$. Define $P_{\hat{g}}$ by putting $\frac{dP_{\hat{g}}}{dP_{g^*}}|_{\mathcal{G}_t} = \Lambda_{gt}$. Then, as in the proof of Theorem 2.1, we can prove that, under $P_{\hat{g}}$, $(y_l - <\hat{g}, X_l>)/<d, X_l>$ is a sequence of $N(0,1)$ i.i.d. random variables.

Now,

$$\begin{aligned} Q_t(\hat{g}, g^*) &= E[L_{g^*}(\hat{g})|\mathcal{Y}_t] \\ &= E[\log \frac{dP_{\hat{g}}}{dP_{g^*}}|\mathcal{Y}_t] \\ &= E[\sum_{l=1}^{t} -\frac{1}{2}(\frac{y_l - <\hat{g}, X_l>}{<\theta, X_l>})^2 - R(g,d)|\mathcal{Y}_t] \\ &= E[\sum_{l=1}^{t}\sum_{i=1}^{N} <X_l, e_i>(-\frac{1}{2}(\frac{y_l - \hat{g}_i}{d_i})^2)|\mathcal{Y}_t] - \breve{R}(g,d). \end{aligned}$$

Setting the derivative of $Q_t(\hat{g}, g^*)$, in $\hat{g}_i$, to be 0, we have

$$E[\sum_{l=1}^{t} <X_l, e_i> (y_l - \hat{g}_i)|\mathcal{Y}_t] = 0.$$

That is,

$$E[J_t^i \hat{g}_i] = E[G_t^i(y)].$$

So,

$$\hat{g}_i = \frac{\sigma(G_t^i(y))}{\sigma(J_t^i)},$$

where $f(y_l) = y_l$ for $G_t^i(y)$. $\square$

Theorem 2.9. *Given the observations up to time t, the EM estimation of d_i is*

$$\hat{d}_i = \sqrt{\frac{\sigma(G_t^i(y^2)) - 2\hat{g}_i\sigma(G_t^i(y)) + \hat{g}_i^2\sigma(J_t^i)}{\sigma(J_t^i)}}, \tag{2.3.23}$$

where $f(y_l) = y_l$ for $G_t^i(y)$, $f(y_l) = y_l^2$ for $G_t^i(y^2)$.

Proof. Define

$$\begin{aligned} \Lambda_{d0} &= 1, \\ \Lambda_{dt} &= \frac{\prod_{l=1}^t \frac{\phi(\frac{y_l - <g, X_l>}{<\hat{d}, X_l>})}{<\hat{d}, X_l>\phi(y_l)}}{\prod_{l=1}^t \frac{\phi(\frac{y_l - <g, X_l>}{<d, X_l>})}{<d, X_l>\phi(y_l)}}. \end{aligned}$$

Set $P_{d^*} = P$. Define $P_{\hat{d}}$ by putting $\frac{dP_{\hat{d}}}{dP_{d^*}}|_{\mathcal{G}_t} = \Lambda_{dt}$. Then, as in the proof of Theorem 2.1, we can prove that, under $P_{\hat{d}}$, $(y_l - < g, X_l >)/ < \hat{d}, X_l >$ is a sequence of $N(0,1)$ i.i.d. random variables.

Now,

$$\begin{aligned} & Q_t(\hat{d}, d^*) \\ = \; & E[L_{d^*}(\hat{d})|\mathcal{Y}_t] \\ = \; & E[\log \frac{dP_{\hat{d}}}{dP_{d^*}}|\mathcal{Y}_t] \\ = \; & E[\sum_{l=1}^t (-\frac{1}{2}(\frac{y_l - < g, X_l >}{< \hat{d}, X_l >})^2 - \log < \hat{d}, X_l >) - R(g, d)|\mathcal{Y}_t] \\ = \; & E[\sum_{l=1}^t \sum_{i=1}^N < X_l, e_i > (-\frac{1}{2}(\frac{y_l - g_i}{\hat{d}_i})^2 - \log \hat{d}_i)|\mathcal{Y}_t] - \breve{R}(g, d). \end{aligned}$$

Setting the derivative of $Q_t(\hat{d}, d^*)$, in $\hat{d}_i$, to be 0, we have

$$E[\sum_{l=1}^{t} < X_l, e_i > (\frac{(y_l - g_i)^2}{\hat{d}_i}\frac{1}{\hat{d}_i^2} - \frac{1}{\hat{d}_i})|\mathcal{Y}_t] = 0.$$

That is,

$$E[J_t^i \hat{d}_i^2] = E[G_t^i(y^2) - 2g_i G_t^i(y) + g_i^2 J_t^i].$$

So,

$$\hat{d}_i = \sqrt{\frac{\sigma(G_t^i(y^2)) - 2\hat{g}_i \sigma(G_t^i(y)) + \hat{g}_i^2 \sigma(J_t^i)}{\sigma(J_t^i)}},$$

where $f(y_l) = y_l$ for G_t^i, $f(y_l) = y_l^2$ for $G_t^i(y^2)$. $\square$

2.4 Simulation and Results

In this section, we shall give an example to show the performance of this method. Assume the parameters in equation (2.2.2) and (2.2.3) have the following values:

$$A = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix},$$

$$g = [0.7, 0.3]',$$

$$d = [0.2, 0.2]'.$$

600 data points are generated for the simulation. The estimated results for the parameters are shown in figure 2.1 and 2.2.

The accuracy of estimating the hidden states is always higher than 70%.

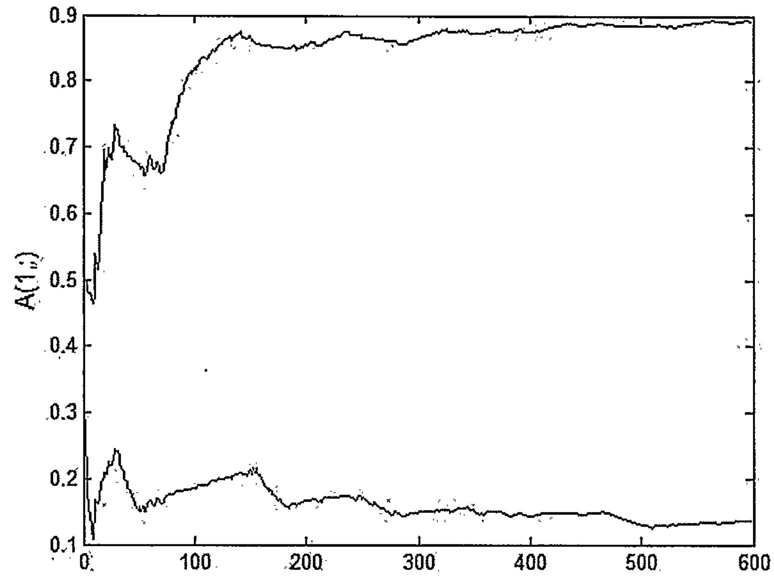
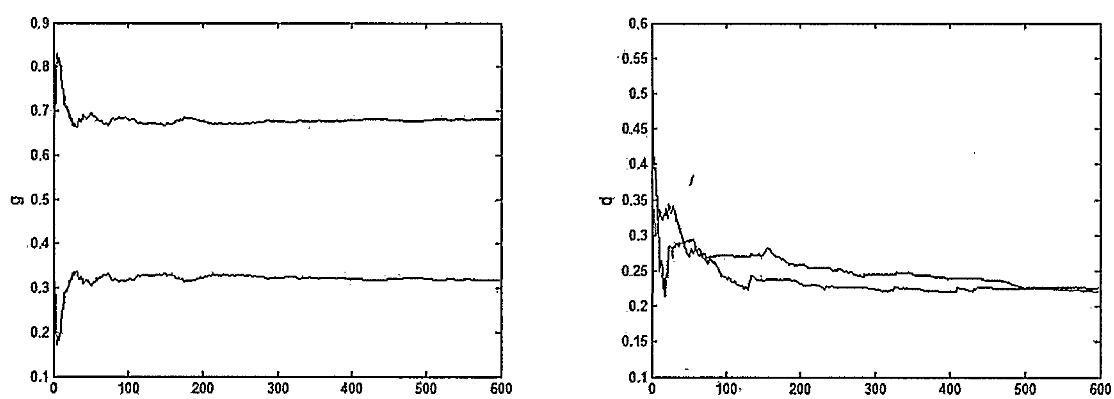


Figure 2.1: Estimated transition probabilities



(a)

(b)

Figure 2.2: Estimated parameters g and d

2.5 Application in Classification of DNA Copy Numbers

The application of comparative genomic hybridization(CGH)s enabled genome-wide analysis of gross DNA copy number imbalance. These are key genetic events in the development and progression of human cancers. The purpose of array-based Comparative Genomic Hybridization, (array CGH), is to detect and map chromosomal aberrations, on a genomic scale, in a single experiment. Since chromosomal copy numbers can not be measured directly, two samples of genomic DNA, (referred to as the reference and test DNAs), are differentially labelled with fluorescent dyes and competitively hybridized to known mapped sequences, (referred to as BACs), that are immobilized on a slide. Subsequently, the ratio of the intensities of the two fluorochromes is computed and a CGH profile is constituted for each chromosome when the log_2 of fluorescence ratios are ranked and plotted according to the physical position of their corresponding BACs on the genome. Each profile can be viewed as a succession of "segments" that represent homogeneous regions in the genome whose BACs share the same relative copy number on average. [13]

Copy number variants are regions of the genome that can occur at a variable copy number in the population. In diploid organisms, such as humans, somatic cells normally contain two copies of each gene, one inherited from each parent. However, abnormalities during the process of DNA replication and synthesis can lead to the loss or gain of DNA fragments, leading to variable gene copy numbers which may initiate or promote disease conditions. For example, the loss or gain of a number of tumor suppressor genes and oncogenes are known to promote the initiation and growth of cancers. [9] So, detection of the changes of the DNA copy numbers is very important in cancer research.

Array CGH data are normalized with a median set to be $log_2(ratio) = 0$ for regions of no change; segments with positive means represent duplicated regions in the test sample genome, and segments with negative means represent deleted regions. [13] An example is shown in figure 2.3.

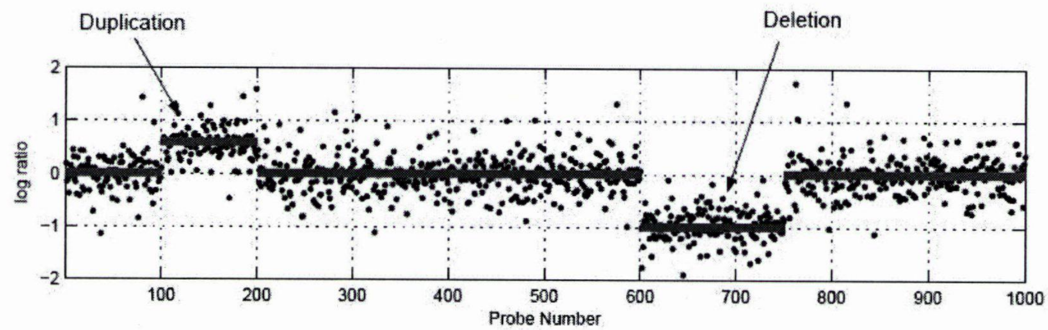


Figure 2.3: An example of array-CGH data[9]

If we suppose that different copy numbers belong to different classes, then the detection of changes of copy numbers becomes a problem of classification of different copy numbers. The different classes are not observable. What can be observed are the log_2 of fluorescence ratios corrupted by noise, which could be considered as a function of copy number with Gaussian noise. Therefore, this problem could be modeled as a hidden Markov model. The hidden states X_t are the classes of different copy numbers, and the observations are the CGH data y_t. Recall that

$$y_t =< g, X_t > + < d, X_t > w_t,$$

Here, g is the log_2 of fluorescence ratios. Usually, there are four states for the copy numbers: [41]

(1) a copy number loss (that could be either a single copy loss or a deletion), $log_2(ratio) = log_2(1/2) = -1$;

(2) copy-neutral state, $log_2(ratio) = log_2(2/2) = 0$;

(3)a single copy gain, $log_2(ratio) = log_2(3/2) = 0.585$;

(4)an amplification (i.e. multiple copy gain), $log_2(ratio) = log_2(k/2)$, $k > 3$.

For the state (4), we take $log_2(ratio) = 1.5$. So, in this problem, $g = [-1, 0, 0.585, 1.5]$. Then the algorithm described in the previous sections could be used to detect the changes of copy numbers.

The data used in this section is the array CGH profiles of 24 pancreatic adenocarcinoma cell lines and 13 primary tumor specimens from [2]. Labeled DNA fragments were hybridized to Agilent human cDNA microarrays containing 14160 cDNA clones.[2] Here I use four sample data sets to show the performance of the above method. Figure 2.4, Figure 2.5, Figure 2.6 and Figure 2.7 show the estimated states. The x-coordinate is the position of the corresponding chromosome. The y-coordinate of the top figure is the log_2 of the fluorescence ratios, and the y-coordinate of the bottom figure is the estimated states, 1, 2, 3, 4 refer to state (1), (2), (3), (4) correspondingly.

The estimated parameters for sample 6, chromosome 12 are as follows:

$$\hat{A} = \begin{pmatrix} 0.1223 & 0.1116 & 0.2845 & 0.2980 \\ 0.6151 & 0.6985 & 0.3418 & 0.4264 \\ 0.0788 & 0.0941 & 0.2586 & 0.2230 \\ 0.1838 & 0.0958 & 0.1151 & 0.0526 \end{pmatrix},$$

$$\hat{d} = [1.5794, 0.3139, 0.0650, 1.5064]',$$

The estimated parameters for sample 8, chromosome 12 are as follows:

$$\hat{A} = \begin{pmatrix} 0.2012 & 0.1052 & 0.1274 & 0.2518 \\ 0.3957 & 0.7317 & 0.3111 & 0.2892 \\ 0.1970 & 0.0516 & 0.2566 & 0.3074 \\ 0.2061 & 0.1115 & 0.3049 & 0.1516 \end{pmatrix},$$

$$\hat{d} = [1.6287, 0.2234, 0.2409, 0.4015]',$$

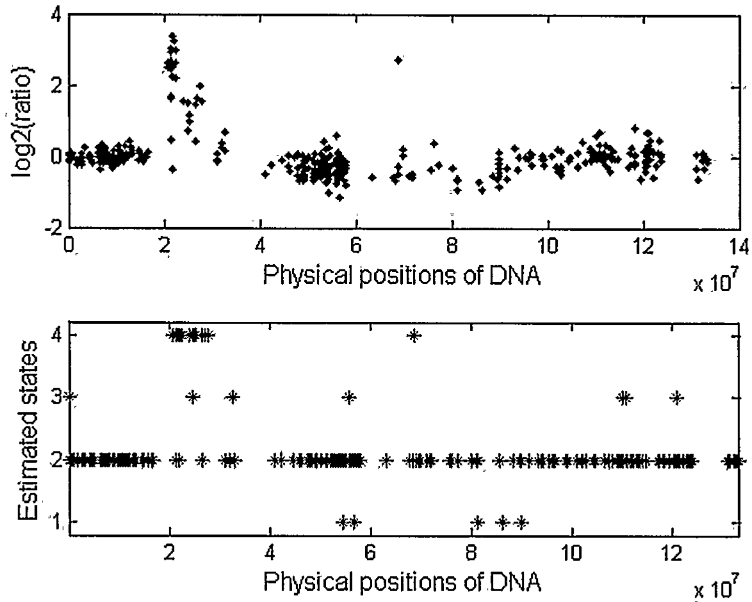


Figure 2.4: Estimated result for array-CGH data (sample 6, chromosome 12)

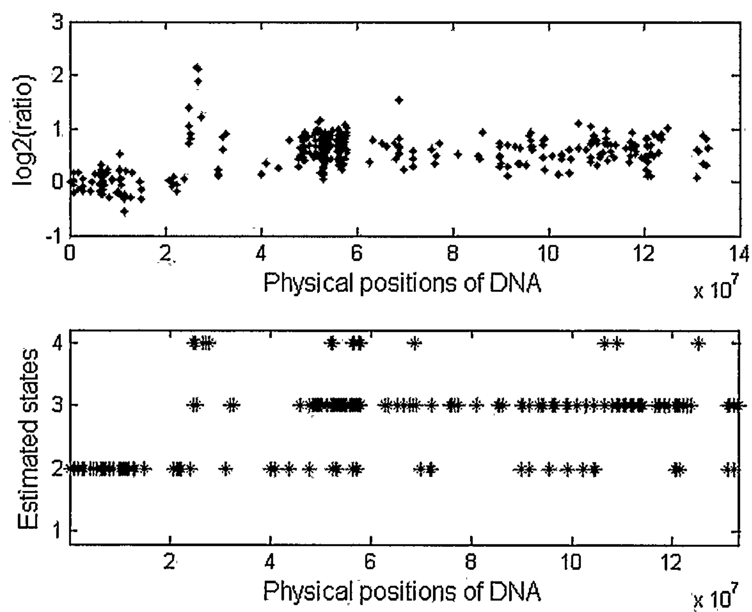


Figure 2.5: Estimated result for array-CGH data (sample 8, chromosome 12)

The estimated parameters for sample 9, chromosome 20 are as follows:

$$\hat{A} = \begin{pmatrix} 0.2365 & 0.1799 & 0.2506 & 0.1633 \\ 0.2985 & 0.2151 & 0.3340 & 0.0971 \\ 0.2547 & 0.2730 & 0.2487 & 0.1971 \\ 0.2102 & 0.3320 & 0.1666 & 0.5424 \end{pmatrix},$$

$$\hat{d} = [1.7497, 0.6543, 0.1745, 0.8063]',$$

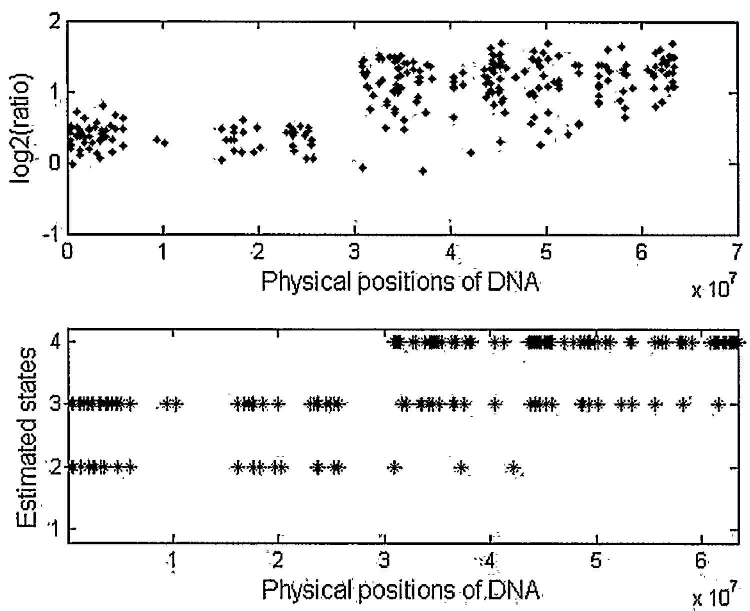


Figure 2.6: Estimated result for array-CGH data (sample 9, chromosome 20)

The estimated parameters for sample 22, chromosome 10 are as follows:

$$\hat{A} = \begin{pmatrix} 0.1726 & 0.2498 & 0.2410 & 0.1377 \\ 0.6434 & 0.6441 & 0.4063 & 0.5254 \\ 0.1270 & 0.0646 & 0.1496 & 0.1638 \\ 0.0569 & 0.0415 & 0.2030 & 0.1731 \end{pmatrix},$$

$$\hat{d} = [0.8857, 0.3553, 0.0950, 0.4474]',$$

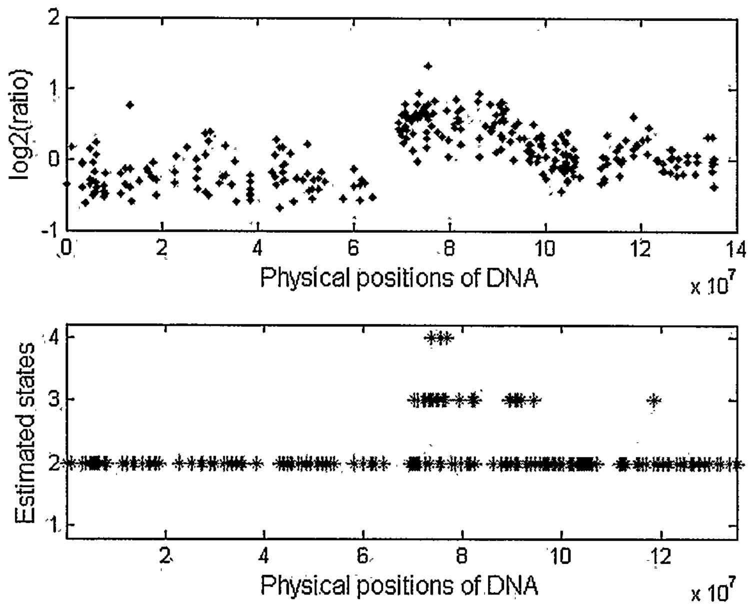


Figure 2.7: Estimated result for array-CGH data (sample 22, chromosome 10)

We see from the above estimates and figures that, although a few states were misclassified, most of the states could be estimated correctly using the method descibed in this chapter. We also see that the CGH data contains large noise. In this case, however, the results obtained by using the methods described in this chapter are still reasonable.

Chapter 3

A FILTER FOR AN AUTOREGRESSIVE HIDDEN MARKOV MODEL

3.1 Introduction

Autoregressive(AR) models are widely used in modeling time-varying signals. Combining a hidden Markov model and an autoregressive model, we obtain an autoregressive hidden Markov model(ARHMM), which is an extension of a hidden Markov model. There are already some methods for estimating the states and parameters of an autoregressive hidden Markov model. In this chapter, we derive new formulae, following Chapter 2, to estimate the hidden states and parameters of an ARHMM. As in Chapter 2, the formulae are recursive in the observation data, and provide on-line estimates.

In the next section, we give a brief introduction to the autoregressive hidden Markov model. We then derive the formulae for estimating hidden states and parameters based on a change of measure and the EM algorithm. In section 3.4, we describe an application of the autoregressive hidden Markov model to classification and give some simulation results. In the final section of this chapter, we apply these estimates to real data.

3.2 Autoregressive Hidden Markov Model

In this chapter, the finite state time-homogeneous Markov chain $X = \{X_t, t = 0, 1, ...\}$ is defined as in Chapter 2. The transition probabilities and dynamics of X are defined in equation (2.2.1) and (2.2.2).

The process X is not observed directly. It is observed through another function, whose values are corrupted by Gaussian noise. Here, we assume the observed process $y = \{y_t, t = 0, 1, ...\}$ is the summation of an autoregressive time series of order p and Gaussian noise. Then y can be written in the form

$$y_t = \alpha^{X_t} + \beta_1^{X_t} y_{t-1} + \beta_2^{X_t} y_{t-2} + ... + \beta_p^{X_t} y_{t-p} + \sigma^{X_t} w_t, \tag{3.2.1}$$

where $w = \{w_t, t = 0, 1, ...\}$ is a sequence of $N(0,1)$ independent, identically distributed (i.i.d.) random variables.

α^{X_t}, $\{\beta_i^{X_t}, i = 1, 2, ..., p\}$ and σ^{X_t} are parameters for the autoregressive model in state X_t. σ^{X_t} is the variance of the Gaussian noise. They all depend on the current state X_t of the chain. Since there are finitely many states, the number of the values for the parameters is finite. Then we can rewrite equation (3.2.1) as

$$y_t =< \alpha, X_t > + < \beta_1, X_t > y_{t-1} + < \beta_2, X_t > y_{t-2} + ... + < \beta_p, X_t > y_{t-p} + < \sigma, X_t > w_t, \tag{3.2.2}$$

where α, $\{\beta_i, i = 1, 2, ...p\}$, σ are N dimensional vectors.

Consider a probability measure $\bar{P}$ on the measurable space $(\Omega, \mathcal{F})$ such that, under $\bar{P}$,

1) The process X is a finite state Markov chain with transition matrix A,

2) The observation $\{y_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables.

As in Chapter 2, the measure $\bar{P}$ is called a "reference" probability.

We now construct a probability P, such that, under P, the process X is still a finite state Markov chain with transition matrix A, and $\{w_t : w_t = \frac{y_t - <\alpha, X_t> - \sum_{m=1}^{p} <\beta_m, X_t> y_{t-m}}{<\sigma, X_t>}\}$ is a sequence of $N(0,1)$ i.i.d. random variables.

Write

$$\begin{aligned}
\bar{\lambda}_l &= \frac{\phi(w_t)}{<\sigma, X_l> \phi(y_l)} \\
&= \frac{\phi(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^{p} <\beta_m, X_l> y_{l-m}}{<\sigma, X_l>})}{<\sigma, X_l> \phi(y_l)}, && (3.2.3) \\
\bar{\Lambda}_0 &= 1, && (3.2.4) \\
\bar{\Lambda}_t &= \prod_{l=1}^{t} \bar{\lambda}_l, \quad t = 1, 2, 3, \ldots \quad . && (3.2.5)
\end{aligned}$$

Definition 3.1. *Define P by putting*

$$\frac{dP}{d\bar{P}}|_{\mathcal{G}_t} = \bar{\Lambda}_t. \tag{3.2.6}$$

Theorem 3.1. *Under P, $\{w_t\}$ is a sequence of $N(0,1)$ random variables.*

Proof.

$$\begin{aligned}
P(w_t \leq a|\mathcal{G}_{t-1}) &= E[I(w_t \leq a)|\mathcal{G}_{t-1}] \\
&= \frac{\bar{E}[\bar{\Lambda}_t I(w_t \leq a)|\mathcal{G}_{t-1}]}{\bar{E}[\bar{\Lambda}_t|\mathcal{G}_{t-1}]} \\
&= \frac{\bar{E}[\bar{\lambda}_t I(w_t \leq a)|\mathcal{G}_{t-1}]}{\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}]}.
\end{aligned}$$

$$\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}] = \bar{E}[\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1} \vee X_t]|\mathcal{G}_{t-1}].$$

The inner expectation

$$\begin{aligned}
\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1} \vee X_t] &= \int_{-\infty}^{\infty} \frac{\phi(\frac{y_t - <\alpha, X_t> - \sum_{m=1}^{p} <\beta_m, X_t> y_{t-m}}{<\sigma, X_t>})}{<\sigma, X_t> \phi(y_t)} \phi(y_t) dy_t \\
&= \int_{-\infty}^{\infty} \phi(w_t) dw_t \\
&= 1.
\end{aligned}$$

Then

$$\bar{E}[\bar{\lambda}_t|\mathcal{G}_{t-1}] = 1.$$

Similarly,

$$\begin{aligned}
\bar{E}[\bar{\lambda}_t I(w_t \leq a)|\mathcal{G}_{t-1} \vee X_t] &= \int_{-\infty}^{\infty} \frac{\phi(\frac{y_t - <\alpha, X_t> - \sum_{m=1}^{p} <\beta_m, X_t> y_{t-m}}{<\sigma, X_t>})}{<\sigma, X_t> \phi(y_t)} I(w_t \leq a)\phi(y_t)dy_t \\
&= \int_{-\infty}^{\infty} \phi(w_t) I(w_t \leq a) dw_t. \qquad (3.2.7)
\end{aligned}$$

$$\begin{aligned}
\bar{E}[\bar{\lambda}_t I(w_t \leq a)|\mathcal{G}_{t-1}] &= \bar{E}[\bar{E}[\bar{\lambda}_t I(w_t \leq a)|\mathcal{G}_{t-1} \vee X_t]|\mathcal{G}_{t-1}] \\
&= \int_{-\infty}^{\infty} \phi(w_t) I(w_t \leq a) dw_t.
\end{aligned}$$

So, $P(w_t \leq a|\mathcal{G}_{t-1}) = P(w_t \leq a) = \int_{-\infty}^{\infty} \phi(w_t) I(w_t \leq a) dw_t$, and the result follows.

□

Remark: Under P, the process X is still a finite state Markov chain with transition matrix A.

This can be proved similarly as in Chapter 2.

3.3 Estimation of States and Parameters

3.3.1 Recursive Estimation

In this section we shall first describe how to estimate the hidden states, given the observations $\{y_t, t = 0, 1, 2\}$. The notations are the same as Chapter 2.

Write

$$D(y_t) = \begin{pmatrix} \frac{\phi(\frac{y_t - \alpha_1 - \sum_{m=1}^{p} \beta_{m,1} y_{t-m}}{\sigma_1})}{\sigma_1 \phi(y_t)} & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{\phi(\frac{y_t - \alpha_N - \sum_{m=1}^{p} \beta_{m,N} y_{t-m}}{\sigma_N})}{\sigma_N \phi(y_t)} \end{pmatrix}. \qquad (3.3.1)$$

Theorem 3.2. *The probability vector q_t is computed by the recursion*

$$q_t = D(y_t) A q_{t-1}. \qquad (3.3.2)$$

Proof.

$$\begin{aligned}
q_t &= \bar{E}[\bar{\Lambda}_t X_t|\mathcal{Y}_t] \\
&= \bar{E}[\bar{\Lambda}_{t-1}\frac{\phi(\frac{y_t-<\alpha,X_t>-\sum_{m=1}^{p}<\beta_m,X_t>y_{t-m}}{<\sigma,X_t>})}{<\sigma,X_t>\phi(y_t)}X_t|\mathcal{Y}_t] \\
&= \bar{E}[\bar{\Lambda}_{t-1}\frac{\phi(\frac{y_t-<\alpha,X_t>-\sum_{m=1}^{p}<\beta_m,X_t>y_{t-m}}{<\sigma,X_t>})}{<\sigma,X_t>\phi(y_t)}\sum_{i=1}^{N}<X_t,e_i>X_t|\mathcal{Y}_t] \\
&= \sum_{i=1}^{N}\bar{E}[\bar{\Lambda}_{t-1}<AX_{t-1}+M_t,e_i>|\mathcal{Y}_t]\frac{\phi(\frac{y_t-\alpha_i-\sum_{m=1}^{p}\beta_{m,i}y_{t-m}}{\sigma_i})}{\sigma_i\phi(y_t)}e_i \\
&= \sum_{i=1}^{N}<\bar{E}[\bar{\Lambda}_{t-1}AX_{t-1}|\mathcal{Y}_{t-1}],e_i>\frac{\phi(\frac{y_t-\alpha_i-\sum_{m=1}^{p}\beta_{m,i}y_{t-m}}{\sigma_i})}{\sigma_i\phi(y_t)}e_i \\
&= D(y_t)Aq_{t-1}.
\end{aligned}$$

□

Similarly as in Chapter 2,

$$\hat{X}_t = \frac{q_t}{<q_t,\mathbf{1}>}. \tag{3.3.3}$$

As in Chapter 2, in order to estimate the parameters in the ARHMM, we have to estimate the random processes $N_t^{(j,i)}$, J_t^i and G_t^i. The recursive estimates of the three processes are given in Section 2.3.1.

3.3.2 Estimation of Parameters

In this section, we use Expectation Maximization (EM) algorithm to re-estimate the parameters in the ARHMM. EM algorithm has been introduced in Chapter 2.

The EM estimation of $a_{j,i}$ is given in Theorem 2.7.

Theorem 3.3. *Given the observations up to time t, the EM estimation of α_i is*

$$\hat{\alpha}_i = \frac{\sigma(G_{t,0}^i) - \sum_{m=1}^{p}\sigma(G_{t,m}^i)\hat{\beta}_{m,i}}{\sigma(J_t^i)}, \tag{3.3.4}$$

where $f(y_l) = y_{l-m}$ for $G_{t,m}^i$, $m = 0, 1, ..., p$.

Proof. Define

$$
\begin{aligned}
\Lambda_{\alpha 0} &= 1, \\
\Lambda_{\alpha t} &= \frac{\prod_{l=1}^{t} \frac{\phi(\frac{y_l - <\hat{\alpha}, X_l> - \sum_{m=1}^{p} <\beta_m, X_l> y_{l-m}}{<\sigma, X_l>})}{<\sigma, X_l> \phi(y_l)}}{\prod_{l=1}^{t} \frac{\phi(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^{p} <\beta_m, X_l> y_{l-m}}{<\sigma, X_l>})}{<\sigma, X_l> \phi(y_l)}}.
\end{aligned}
$$

Set $P_{\alpha^*} = P$. Define $P_{\hat{\alpha}}$ by putting $\frac{dP_{\hat{\alpha}}}{dP_{\alpha^*}}|_{\mathcal{G}_t} = \Lambda_{\alpha t}$. Then, as in the proof of Theorem 2.1, we can prove that, under $P_{\hat{\alpha}}$, $(y_l - <\hat{\alpha}, X_l> - \sum_{m=1}^{p} <\beta_m, X_l> y_{l-m})/<\sigma, X_l>$ is a sequence of $N(0,1)$ i.i.d. random variables.

Now,

$$
\begin{aligned}
Q_t(\hat{\alpha}, \alpha^*) &= E[L(\hat{\alpha})|\mathcal{Y}_t] \\
&= E[\log \frac{dP_{\hat{\alpha}}}{dP_{\alpha^*}} |\mathcal{Y}_t] \\
&= E[\sum_{l=1}^{t} -\frac{1}{2} (\frac{y_l - <\hat{\alpha}, X_l> - \sum_{m=1}^{p} <\beta_m, X_l> y_{l-m}}{<\sigma, X_l>})^2 - R(\alpha, \beta, \sigma)|\mathcal{Y}_t] \\
&= E[\sum_{l=1}^{t} \sum_{i=1}^{N} <X_l, e_i> (-\frac{1}{2} (\frac{y_l - \hat{\alpha}_i - \sum_{m=1}^{p} \beta_{m,i} y_{l-m}}{\sigma_i})^2)|\mathcal{Y}_t] - \breve{R}(\alpha, \beta, \sigma).
\end{aligned}
$$

Setting the derivative of $Q_t(\hat{\alpha}, \alpha^*)$, in $\hat{\alpha}_i$, to be 0, we have

$$
E[\sum_{l=1}^{t} <X_l, e_i> (y_l - \hat{\alpha}_i - \sum_{m=1}^{p} \beta_{m,i} y_{l-m})|\mathcal{Y}_t] = 0.
$$

That is,

$$
E[J_t^i \hat{\alpha}_i] = E[G_{t,0}^i - \sum_{m=1}^{p} G_{t,m}^i \beta_{m,i}].
$$

So,

$$
\hat{\alpha}_i = \frac{\sigma(G_{t,0}^i) - \sum_{m=1}^{p} \sigma(G_{t,m}^i) \hat{\beta}_{m,i}}{\sigma(J_t^i)},
$$

where $f(y_l) = y_{l-m}$ for $G_{t,m}^i$, $m = 0, 1, ..., p$. □

Theorem 3.4. *Given the observations up to time t, the EM estimation of $\beta_{m,i}$ is*

$$\hat{\beta}_{m,i} = \frac{\sigma(G^i_{t,m0}) - \sigma(G^i_{t,m})\hat{\alpha}_i}{\sigma(G^i_{t,mm})}, \tag{3.3.5}$$

where $f(y_l) = y_{l-m}$ for $G^i_{t,m}$, $f(y_l) = y_{l-m}y_l$ for $G^i_{t,m0}$, $f(y_l) = y^2_{l-m}$ for $G^i_{t,mm}$, $m = 0, 1, ..., p$.

Proof. Define

$$\begin{aligned} \Lambda_{\beta 0} &= 1, \\ \Lambda_{\beta t} &= \frac{\prod_{l=1}^t \frac{\phi(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^p <\hat{\beta}_m, X_l> y_{l-m}}{<\sigma, X_l>})}{<\sigma, X_l>\phi(y_l)}}{\prod_{l=1}^t \frac{\phi(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^p <\beta_m, X_l> y_{l-m}}{<\sigma, X_l>})}{<\sigma, X_l>\phi(y_l)}}. \end{aligned}$$

Set $P_{\beta^*} = P$. Define $P_{\hat{\beta}}$ by putting $\frac{dP_{\hat{\beta}}}{dP_{\beta^*}}|_{\mathcal{G}_t} = \Lambda_{\beta t}$. Then, as in the proof of Theorem 2.1, we can prove that, under $P_{\hat{\beta}}$, $(y_l - <\alpha, X_l> - \sum_{m=1}^p <\hat{\beta}_m, X_l> y_{l-m})/<\sigma, X_l>$ is a sequence of $N(0,1)$ i.i.d. random variables.

Now,

$$\begin{aligned} Q_t(\hat{\beta}, \beta^*) &= E[L(\hat{\beta})|\mathcal{Y}_t] \\ &= E[\log \frac{dP_{\hat{\beta}}}{dP_{\beta^*}}|\mathcal{Y}_t] \\ &= E[\sum_{l=1}^t -\frac{1}{2}(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^p <\hat{\beta}_m, X_l> y_{l-m}}{<\sigma, X_l>})^2 - R(\alpha, \beta, \sigma)|\mathcal{Y}_t] \\ &= E[\sum_{l=1}^t \sum_{i=1}^N <X_l, e_i>(-\frac{1}{2}(\frac{y_l - \alpha_i - \sum_{m=1}^p \hat{\beta}_{m,i} y_{l-m}}{\sigma_i})^2)|\mathcal{Y}_t] - \breve{R}(\alpha, \beta, \sigma). \end{aligned}$$

Setting the derivative of $Q_t(\hat{\beta}, \beta^*)$, in $\hat{\beta}_{m,i}$, to be 0, we have

$$E[\sum_{l=1}^t <X_l, e_i> (y_l - \alpha_i - \sum_{n=1}^p \hat{\beta}_{m,i} y_{l-m}) y_{l-m}|\mathcal{Y}_t] = 0.$$

That is,

$$E[G^i_{t,mm}\hat{\beta}_{m,i}] = E[G^i_{t,m0} - G^i_{t,m}\alpha_i].$$

So,

$$\hat{\beta}_{m,i} = \frac{\sigma(G^i_{t,m0}) - \sigma(G^i_{t,m})\hat{\alpha}_i}{\sigma(G^i_{t,mm})},$$

where $f(y_l) = y_{l-m}$ for $G^i_{t,m}$, $f(y_l) = y_{l-m}y_l$ for $G^i_{t,m0}$, $f(y_l) = y^2_{l-m}$ for $G^i_{t,mm}$, $m = 0, 1, ..., p$. □

Theorem 3.5. *Given the observations up to time t, the EM estimation of σ_i is*

$$\begin{aligned}\hat{\sigma}_i &= ((\sigma(G^i_{t,00}) - 2\hat{\alpha}_i\sigma(G^i_{t,0}) - 2\sum_{m=1}^{p}\hat{\beta}_{m,i}\sigma(G^i_{t,m0}) + 2\sum_{m=1}^{p}\hat{\alpha}_i\hat{\beta}_{m,i}\sigma(G^i_{t,m}) + \\ &\quad \hat{\alpha}_i^2\sigma(J^i_t) + \sum_{m=1}^{p}\sum_{n=1}^{p}\hat{\beta}_{m,i}\hat{\beta}_{n,i}\sigma(G^i_{t,mn}))/(\sigma(J^i_t)))^{1/2}, \qquad (3.3.6)\end{aligned}$$

where $f(y_l) = y_l^2$ for $G^i_{t,00}$, $f(y_l) = y_{l-m}$ for $G^i_{t,m}$, $f(y_l) = y_{l-m}y_l$ for $G^i_{t,m0}$, $f(y_l) = y_{l-m}y_{l-n}$ for $G^i_{t,mn}$, $m, n = 0, 1, ..., p$.

Proof. Define

$$\begin{aligned}\Lambda_{\sigma 0} &= 1, \\ \Lambda_{\sigma t} &= \frac{\prod_{l=1}^{t}\frac{\phi(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^{p}<\beta_m, X_l> y_{l-m}}{<\hat{\sigma}, X_l>})}{<\hat{\sigma}, X_l>\phi(y_l)}}{\prod_{l=1}^{t}\frac{\phi(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^{p}<\beta_m, X_l> y_{l-m}}{<\sigma, X_l>})}{<\sigma, X_l>\phi(y_l)}}.\end{aligned}$$

Set $P_{\sigma^*} = P$. Define $P_{\hat{\sigma}}$ by putting $\frac{dP_{\hat{\sigma}}}{dP_{\sigma^*}}|_{\mathcal{G}_t} = \Lambda_{\sigma t}$. Then, as in the proof of Theorem 2.1, we can prove that, under $P_{\hat{\alpha}}$, $(y_l - <\alpha, X_l> - \sum_{m=1}^{p} <\beta_m, X_l> y_{l-m})/<\hat{\sigma}, X_l>$ is a sequence of $N(0, 1)$ i.i.d. random variables.

Now,

$$
\begin{aligned}
& Q_t(\hat{\sigma}, \sigma^*) \\
= \ & E[L(\hat{\sigma})|\mathcal{Y}_t] \\
= \ & E[\log \frac{dP_{\hat{\sigma}}}{dP_{\sigma^*}}|\mathcal{Y}_t] \\
= \ & E[\sum_{l=1}^{t}(-\frac{1}{2}(\frac{y_l - <\alpha, X_l> - \sum_{m=1}^{p} <\beta_m, X_l> y_{l-m}}{<\hat{\sigma}, X_l>})^2 - log <\hat{\sigma}, X_l>) - R(\alpha, \beta, \sigma)|\mathcal{Y}_t] \\
= \ & E[\sum_{l=1}^{t}\sum_{i=1}^{N} <X_l, e_i> (-\frac{1}{2}(\frac{y_l - \alpha_i - \sum_{m=1}^{p}\beta_{m,i}y_{l-m}}{\hat{\sigma}_i})^2 - log\hat{\sigma}_i)|\mathcal{Y}_t] - \breve{R}(\alpha, \beta, \sigma).
\end{aligned}
$$

Setting the derivative of $Q_t(\hat{\sigma}, \sigma^*)$, in $\hat{\sigma}_i$, to be 0, we have

$$
E[\sum_{l=1}^{t} <X_l, e_i> (\frac{(y_l - \alpha_i - \sum_{m=1}^{p}\beta_{m,i}y_{l-m})^2}{\hat{\sigma}_i}\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\hat{\sigma}_i})|\mathcal{Y}_t] = 0.
$$

That is,

$$
\begin{aligned}
& E[J_t^i \hat{\sigma}_i^2] \\
= \ & E[G_{t,00}^i - 2\alpha_i G_{t,0}^i - 2\sum_{m=1}^{p}\beta_{m,i}G_{t,m0}^i + 2\sum_{m=1}^{p}\alpha_i\beta_{m,i}G_{t,m}^i + \alpha_i^2 J_t^i + \sum_{m=1}^{p}\sum_{n=1}^{p}\beta_{m,i}\beta_{n,i}G_{t,mn}^i].
\end{aligned}
$$

So,

$$
\begin{aligned}
\hat{\sigma}_i \ = \ & ((\sigma(G_{t,00}^i) - 2\hat{\alpha}_i\sigma(G_{t,0}^i) - 2\sum_{m=1}^{p}\hat{\beta}_{m,i}\sigma(G_{t,m0}^i) + 2\sum_{m=1}^{p}\hat{\alpha}_i\hat{\beta}_{m,i}\sigma(G_{t,m}^i) + \\
& \hat{\alpha}_i^2\sigma(J_t^i) + \sum_{m=1}^{p}\sum_{n=1}^{p}\hat{\beta}_{m,i}\hat{\beta}_{n,i}\sigma(G_{t,mn}^i))/(\sigma(J_t^i)))^{1/2},
\end{aligned}
$$

where $f(y_l) = y_l^2$ for $G_{t,00}^i$, $f(y_l) = y_{l-m}$ for $G_{t,m}^i$, $f(y_l) = y_{l-m}y_l$ for $G_{t,m0}^i$, $f(y_l) = y_{l-m}y_{l-n}$ for $G_{t,mn}^i$, $m, n = 0, 1, ..., p$. □

3.4 Classification and Simulation Results

In some classification problems, the object belongs to one class at one time, and may jump to another class at the next time, according to certain probability. Then, the states of the object could be modeled as a Markov chain. However, the states (classes) may not be observed directly. What possibly can be observed is a sequence of data, generated by different AR models corresponding to different classes. In this case, we could use the ARHMM to model the problem, and apply the above method to estimate the hidden states (classes).

In order to show the effectiveness of the method, we consider an example. Assume there are two classes, and assign the following values to the parameters.

$$A = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix},$$

$$\alpha = [0.7, 0.3]',$$

$$\beta = [0.4, 0.6]',$$

$$\sigma = [0.1, 0.1]'.$$

We generate 300 data for the simulation. The estimated results for the hidden classes are shown in Figure 3.1.

The accuracy of estimating the hidden classes is always higher than 75%.

3.5 Application in Predicting EL NINO Phenomenon

One of the applications of the ARHMM for classification is prediction of the EL NINO phenomenon. What are observed are temperatures of the sea water. The values of the temperatures could be modeled by two AR models, according to two classes, "there is an EL NINO phenomenon" and "there is no EL NINO phenomenon".

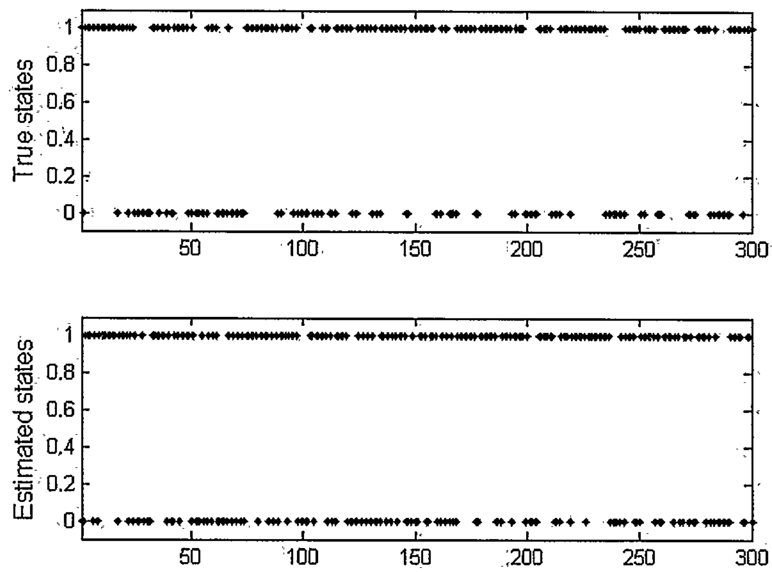


Figure 3.1: Estimated classes

The data are downloaded from the website: "http://www.pmel.noaa.gov/tao/". We use the temperature values of the sea water at 0^0N,155^0W and 0^0N,170^0W, from July 21st, 1991 to January 9th, 1997. Figure 3.2 shows the distribution of temperature of the sea water from 140^0E to 100^0W along the equator, from July 21st, 1991 to January 9th, 1997. If the warm tongue takes up most of the area from east to west, we could say there is an EL NINO phenomenon at that time.[49]

Figure 3.3 shows the estimated result. The points in the first plot of Figure 3.3 show that there are EL NINO at those times. Compare Figure 3.2 and Figure 3.3, we see that our estimation is consistent with the real situation. For example, the warm tongue takes up most of the area from east to west on the right in Figure 3.2, so there is an EL NINO phenomenon during those years. The estimated result in Figure 3.3 also shows that there is an EL NINO phenomenon during the years on the right of the figure.

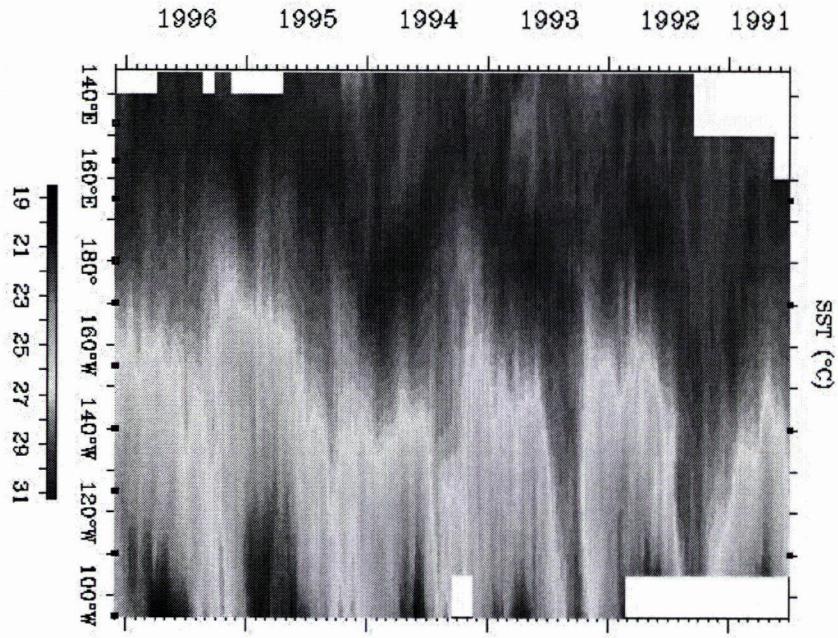


Figure 3.2: Temperature distribution

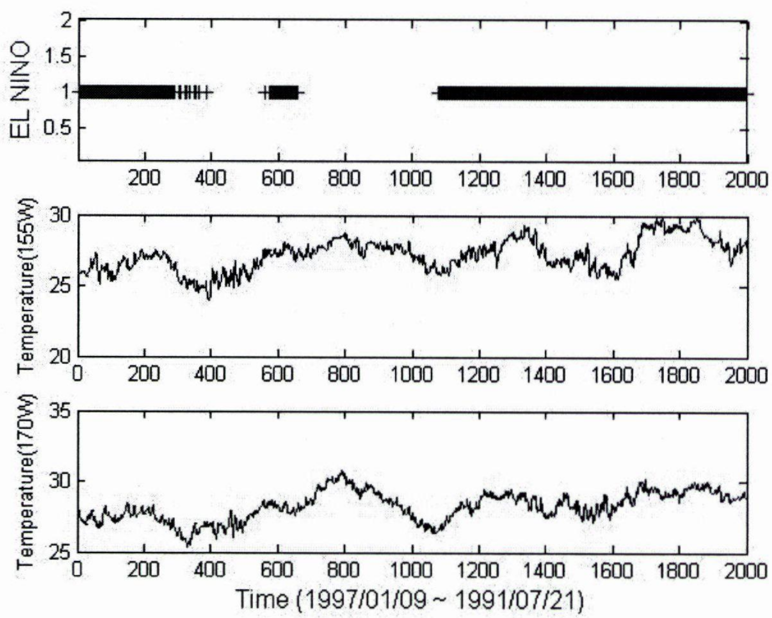


Figure 3.3: Estimated results

Chapter 4

A VITERBI SMOOTHER FOR A DISCRETE STATE SPACE MODEL

4.1 Introduction

Recursive estimates for the parameters of discrete time Markov chains observed in Gaussian noise have been discussed in Chapter 2 and Chapter 3. The recursive estimate for the hidden Markov model in which both the hidden states and the observations are discrete is derived in [36]. In this chapter, we derive new filter and smoother update formulae, based on a change of measure method and extensions of the Viterbi algorithm. The formulae are recursive in the observation data, and could possibly be used in biological sequence analysis and communications.

This chapter is organized as follows. In the next section, we give a brief introduction to the discrete state hidden Markov model. In section 4.3, we derive the filter based recursive estimates of the parameters in the hidden Markov model. In section 4.4, we derive the smoother based recursive estimates of the parameters in the hidden Markov model. In section 4.5, we give the backward Viterbi filter and the simulation results. In section 4.6, we give the backward Viterbi smoother and the simulation results. In the final section, we give some conclusions.

4.2 Hidden Markov Model

In this chapter, the finite state time-homogeneous Markov chain $X = \{X_t, t = 0, 1, ...\}$ is defined as in Chapter 2. The transition probabilities and the dynamics of X are defined by equations (2.2.1) and (2.2.2).

We suppose the process X is not observed directly; rather, we observe a second finite state process $Y = \{Y_t, t = 0, 1, ...\}$, where

$$Y_{t+1} = c(X_t, W_{t+1}). \tag{4.2.1}$$

Here $W = \{W_t, t = 0, 1, ...\}$ is a sequence of independent, identically distributed (i.i.d.) random variables.

Suppose the range of $c(\cdot, \cdot)$ consists of M points; then we can again identify the range of $c(.,.)$ with the set of unit vectors

$$Q = \{f_1, f_2, ..., f_M\}, \tag{4.2.2}$$

where $f_j = (0, 0, ..., 0, 1, 0, ..., 0)' \in R^M$.

Write $\mathcal{G}_t = \sigma\{X_0, y_0, X_1, y_1, ..., X_t, y_t\}$. Suppose

$$P(Y_{t+1} = f_j | \mathcal{G}_t) = P(Y_{t+1} = f_j | X_t), \tag{4.2.3}$$

and write

$$c_{j,i} \triangleq P(Y_{t+1} = f_j | X_t = e_i), \quad C = (c_{j,i}), 1 \leq i \leq N, 1 \leq j \leq M.$$

Then, $E[Y_{t+1}|X_t] = CX_t$. If $W_{t+1} := Y_{t+1} - CX_t$, then

$$\begin{aligned} E[W_{t+1}|\mathcal{G}_t] &= E[Y_{t+1} - CX_t|\mathcal{G}_t] \\ &= CX_t - CX_t \\ &= 0. \end{aligned}$$

So, $W_{t+1} = Y_{t+1} - CX_t$ is a $(P, \mathcal{G}_{t+1})$ martingale increment. [36] Therefore,

$$Y_{t+1} = CX_t + W_{t+1}. \tag{4.2.4}$$

Consider a probability measure $\bar{P}$ on the measurable space $(\Omega, \mathcal{F})$ such that, under $\bar{P}$,

1) The process X is a finite state Markov chain with transition matrix A,

2) The observation $\{Y_t\}$ is a sequence of i.i.d. random variables and

$$\bar{P}\{Y_{t+1}^j = 1|\mathcal{G}_t\} = \bar{P}\{Y_{t+1}^j = 1\} = 1/M. \tag{4.2.5}$$

We call the measure $\bar{P}$ a "reference" probability.

We now construct a probability P, such that, under P, the process X is still a finite state Markov chain with a transition matrix A, and $P(Y_{t+1} = f_j|X_t = e_i) = c_{j,i}$.

Write

$$\bar{\lambda}_{l+1} = M \sum_{j=1}^{M} \sum_{i=1}^{N} c_{j,i} < Y_{l+1}, f_j >< X_l, e_i > . \tag{4.2.6}$$

$$\bar{\Lambda}_t = \prod_{l=1}^{t} \bar{\lambda}_l. \tag{4.2.7}$$

Definition 4.1. *Define P by putting*

$$\frac{dP}{d\bar{P}}|_{\mathcal{G}_t} = \bar{\Lambda}_t. \tag{4.2.8}$$

Theorem 4.1. *Under P, the process X is still a finite state Markov chain with transition matrix A and $\{Y_t\}$ is a sequence of random variables such that $P(Y_{t+1}^j = 1|X_t = e_i) = c_{j,i}$. [36]*

Proof.

$$\begin{aligned}
\bar{E}[\bar{\lambda}_{t+1}|\mathcal{G}_t] &= \bar{E}[M\sum_{j=1}^{M}\sum_{i=1}^{N} c_{j,i} < Y_{t+1}, f_j >< X_t, e_i > |\mathcal{G}_t] \\
&= M\bar{E}[\sum_{j=1}^{M} c_{j,i} < Y_{t+1}, f_j > |\mathcal{G}_t] \\
&= M\sum_{j=1}^{M} c_{j,i}\bar{E}[< Y_{t+1}, f_j > |\mathcal{G}_t] \\
&= M\sum_{j=1}^{M} c_{j,i}\frac{1}{M} \\
&= 1.
\end{aligned}$$

$$\begin{aligned}
P(Y_{t+1}^{j} = 1|X_t = e_i) &= E[< Y_{t+1}, f_j > |X_t = e_i] \\
&= \frac{\bar{E}[\bar{\Lambda}_{t+1} < Y_{t+1}, f_j > |X_t = e_i]}{\bar{E}[\bar{\Lambda}_{t+1}|X_t = e_i]} \\
&= \frac{\bar{E}[\bar{\lambda}_{t+1} < Y_{t+1}, f_j > |X_t = e_i]}{\bar{E}[\bar{\lambda}_{t+1}|X_t = e_i]} \\
&= \bar{E}[\bar{\lambda}_{t+1} < Y_{t+1}, f_j > |X_t = e_i] \\
&= \bar{E}[M\sum_{m=1}^{M}\sum_{l=1}^{N} c_{m,l} < Y_{t+1}, f_m >< X_t, e_l >< Y_{t+1}, f_j > |X_t = e_i] \\
&= M\bar{E}[c_{j,i} < Y_{t+1}, f_j > |X_t = e_i] \\
&= c_{j,i}.
\end{aligned}$$

$$\begin{aligned}
E[X_{t+1}|\mathcal{G}_t] &= \frac{\bar{E}[\bar{\Lambda}_{t+1}X_{t+1}|\mathcal{G}_t]}{\bar{E}[\bar{\Lambda}_{t+1}|\mathcal{G}_t]} \\
&= \frac{\bar{E}[\bar{\lambda}_{t+1}X_{t+1}|\mathcal{G}_t]}{\bar{E}[\bar{\lambda}_{t+1}|\mathcal{G}_t]} \\
&= \bar{E}[\bar{\lambda}_{t+1}X_{t+1}|\mathcal{G}_t] \\
&= \bar{E}[M\sum_{m=1}^{M}\sum_{l=1}^{N} c_{m,l} < Y_{t+1}, f_m >< X_t, e_l > X_{t+1}|\mathcal{G}_t] \\
&= \bar{E}[\sum_{l=1}^{N} < X_t, e_l > X_{t+1}\bar{E}[M\sum_{m=1}^{M} c_{m,l} < Y_{t+1}, f_m > |\mathcal{G}_t]|\mathcal{G}_t] \\
&= \bar{E}[\sum_{l=1}^{N} < X_t, e_l > X_{t+1}|\mathcal{G}_t] \\
&= \bar{E}[X_{t+1}|X_t] \\
&= AX_t.
\end{aligned}$$

So, the result follows. □

The real world dynamics take place under P. However, $\bar{P}$ is a nicer measure under which to work.

4.3 Filter Based Estimation

In this section, we shall derive filter based estimates for the parameters A and C. Following the steps in Chapter 2, we must first estimate the unobserved state process X and some processes related to X.

Write

$$B(Y_{t+1}) = M \begin{pmatrix} \sum_{m=1}^{M} c_{m,1} < Y_{t+1}, f_m > & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sum_{m=1}^{M} c_{m,N} < Y_{t+1}, f_m > \end{pmatrix}. \tag{4.2.9}$$

Theorem 4.2. *The unnormalized probability vector q_t satisfies the recursion*

$$q_{t+1} = AB(Y_{t+1})q_t. \tag{4.2.10}$$

Proof.

$$\begin{aligned} q_{t+1} &= \bar{E}[\bar{\Lambda}_{t+1}X_{t+1}|\mathcal{Y}_{t+1}] \\ &= \bar{E}[\bar{\Lambda}_t(AX_t + M_{t+1}) \cdot M \sum_{j=1}^{M}\sum_{i=1}^{N} c_{j,i} < Y_{t+1}, f_j >< X_t, e_i > |\mathcal{Y}_{t+1}] \\ &= M \sum_{j=1}^{M}\sum_{i=1}^{N} \bar{E}[\bar{\Lambda}_t AX_t c_{j,i} < Y_{t+1}, f_j >< X_t, e_i > |\mathcal{Y}_{t+1}] \\ &= M \sum_{j=1}^{M}\sum_{i=1}^{N} \bar{E}[\bar{\Lambda}_t < X_t, e_i > |\mathcal{Y}_t] A e_i c_{j,i} < Y_{t+1}, f_j > \\ &= M \sum_{j=1}^{M}\sum_{i=1}^{N} < \bar{E}[\bar{\Lambda}_t X_t|\mathcal{Y}_t], e_i > A e_i c_{j,i} < Y_{t+1}, f_j > \\ &= M \sum_{j=1}^{M}\sum_{i=1}^{N} < q_t, e_i > A e_i c_{j,i} < Y_{t+1}, f_j > \\ &= AB(Y_{t+1})q_t. \end{aligned}$$

□

As in Chapter 2,

$$E[X_t|\mathcal{Y}_t] = \frac{q_t}{\sum_{i=1}^{N} < q_t, e_i >}. \tag{4.2.11}$$

Theorem 4.3. *The vector $\sigma(N_t^{(j,i)}X_t)$ is computed by the recursion*

$$\sigma(N_{t+1}^{(j,i)}X_{t+1}) = AB(Y_{t+1})\sigma(N_t^{(j,i)}X_t) + < q_t, e_i > a_{j,i}e_j[B(Y_{t+1})]_{i,i}, \qquad (4.2.12)$$

where $N_t^{(j,i)}$ is defined in Chapter 2.

Proof.

$$\begin{aligned}
& \sigma(N_{t+1}^{(j,i)}X_{t+1}) \\
= \;& \bar{E}[\bar{\Lambda}_{t+1}N_{t+1}^{(j,i)}X_{t+1}|\mathcal{Y}_{t+1}] \\
= \;& \bar{E}[\bar{\Lambda}_t(N_t^{(j,i)} + < X_t, e_i >< X_{t+1}, e_j >)X_{t+1}M\sum_{m=1}^{M}\sum_{l=1}^{N} c_{m,l} < Y_{t+1}, f_m >< X_t, e_l > |\mathcal{Y}_{t+1}] \\
= \;& M\sum_{m=1}^{M}\sum_{l=1}^{N} < \sigma(N_t^{(j,i)}X_t), e_l > Ae_l c_{m,l} < Y_{t+1}, f_m > + \\
& M\sum_{m=1}^{M}\bar{E}[\bar{\Lambda}_t < X_t, e_i > |\mathcal{Y}_t]a_{j,i}e_j c_{m,i} < Y_{t+1}, f_m > \\
= \;& AB(Y_{t+1})\sigma(N_t^{(j,i)}X_t) + < q_t, e_i > a_{j,i}e_j[B(Y_{t+1})]_{i,i}.
\end{aligned}$$

□

Theorem 4.4. *The vector $\sigma(J_t^iX_t)$ is computed by the recursion*

$$\sigma(J_{t+1}^iX_{t+1}) = AB(Y_{t+1})\sigma(J_t^iX_t) + < q_t, e_i > Ae_i[B(Y_{t+1})]_{i,i}, \qquad (4.2.13)$$

where $J_t^i \triangleq \sum_{l=1}^{t} < X_{l-1}, e_i >$.

Proof.

$$
\begin{aligned}
& \sigma(J^i_{t+1}X_{t+1}) \\
= \; & \bar{E}[\bar{\Lambda}_{t+1}J^i_{t+1}X_{t+1}|\mathcal{Y}_{t+1}] \\
= \; & \bar{E}[\bar{\Lambda}_t(J^i_t + <X_t, e_i>)(AX_t + M_{t+1})M\sum_{m=1}^{M}\sum_{l=1}^{N} c_{m,l} <Y_{t+1}, f_m><X_t, e_l> |\mathcal{Y}_{t+1}] \\
= \; & M\sum_{m=1}^{M}\sum_{l=1}^{N} <\sigma(J^i_t X_t), e_l> Ae_l c_{m,l} <Y_{t+1}, f_m> + \\
& M\sum_{m=1}^{M} <q_t, e_i> Ae_i c_{m,i} <Y_{t+1}, f_m> \\
= \; & AB(Y_{t+1})\sigma(J^i_t X_t) + <q_t, e_i> Ae_i[B(Y_{t+1})]_{i,i}.
\end{aligned}
$$

$\square$

Theorem 4.5. *The probability vector $\sigma(G^i_t X_t)$ is computed by the recursion*

$$\sigma(G^i_{t+1}X_{t+1}) = AB(Y_{t+1})\sigma(G^i_t X_t) + <q_t, e_i> f(Y_{t+1})Ae_i[B(Y_{t+1})]_{i,i}, \qquad (4.2.14)$$

where $G^i_t \triangleq \sum_{l=1}^{t} f(Y_l) < X_{l-1}, e_i >$.

Proof.

$$
\begin{aligned}
& \sigma(G^i_{t+1}X_{t+1}) \\
= \; & \bar{E}[\bar{\Lambda}_{t+1}G^i_{t+1}X_{t+1}|\mathcal{Y}_{t+1}] \\
= \; & \bar{E}[\bar{\Lambda}_t(G^i_t + f(Y_{t+1}) <X_t, e_i>)(AX_t + M_{t+1})M\sum_{m=1}^{M}\sum_{l=1}^{N} c_{m,l} <Y_{t+1}, f_m><X_t, e_l> |\mathcal{Y}_{t+1}] \\
= \; & AB(Y_{t+1})\sigma(G^i_t X_t) + <q_t, e_i> f(Y_{t+1})Ae_i[B(Y_{t+1})]_{i,i}.
\end{aligned}
$$

$\square$

As before,

$$
\begin{aligned}
\sigma(N^{(j,i)}_t) &= <\sigma(N^{(j,i)}_t, X_t), \mathbf{1}>, \\
\sigma(J^i_t) &= <\sigma(J^i_t, X_t), \mathbf{1}>, \\
\sigma(G^i_t) &= <\sigma(G^i_t, X_t), \mathbf{1}>.
\end{aligned}
$$

Finally, using the Expectation Maximization(EM) algorithm, the parameters can be re-estimated by the formulae:

$$\hat{a}_{j,i} = \frac{\sigma(N_t^{(j,i)})}{\sigma(J_t^i)}, \tag{4.2.15}$$

$$\hat{c}_{j,i} = \frac{\sigma(G_t^i)}{\sigma(J_t^i)}, \tag{4.2.16}$$

with $f(Y_l) =< Y_l, f_j >$.

The details of this algorithm and the derivatives of the estimates are given in [1] and [36].

4.4 Smoother Based Estimation

In this section, we derive the estimate of X_t, given the information $\mathcal{Y}_T$, for $0 \leq k \leq T$.

From Bayes' Theorem,

$$E[X_t|\mathcal{Y}_T] = \frac{\bar{E}[\bar{\Lambda}_{0,T}X_t|\mathcal{Y}_T]}{\bar{E}[\bar{\Lambda}_{0,T}|\mathcal{Y}_T]}. \tag{4.4.1}$$

Here

$$\bar{E}[\bar{\Lambda}_{0,T}X_t|\mathcal{Y}_T] = \bar{E}[\bar{\Lambda}_{0,t}X_t\bar{E}[\bar{\Lambda}_{t+1,T}|\mathcal{Y}_T \vee \mathcal{F}_t]|\mathcal{Y}_T], \tag{4.4.2}$$

where $\bar{\Lambda}_{t+1,T} = \prod_{l=t+1}^{T} \bar{\lambda}_l$.

Using the Markov property, we have

$$\bar{E}[\bar{\Lambda}_{t+1,T}|\mathcal{Y}_T \vee \mathcal{F}_t] = \bar{E}[\bar{\Lambda}_{t+1,T}|\mathcal{Y}_T \vee \sigma(X_t)]. \tag{4.4.3}$$

Write

$$v_{t,T} = (< v_{t,T}, e_1 >, ..., < v_{t,T}, e_N >)', \tag{4.4.4}$$

where $< v_{t,T}, e_i > \triangleq \bar{E}[\bar{\Lambda}_{t+1,T}|\mathcal{Y}_T \vee X_t = e_i]$. We put $v_{T,T} = \mathbf{1} \in R^N$.

Lemma 4.1. *The process v is computed by the backward recursion*

$$v_{t,T} = B(Y_{t+1})A'v_{t+1,T}. \tag{4.4.5}$$

Proof.

$$\begin{aligned}
& < v_{t,T}, e_i > \\
= \;& \bar{E}[\bar{\Lambda}_{t+2,T} M \sum_{m=1}^{M}\sum_{l=1}^{N} c_{m,l} < Y_{t+1}, f_m >< X_t, e_l > |\mathcal{Y}_T \vee X_t = e_i] \\
= \;& M \sum_{m=1}^{M} \bar{E}[\bar{\Lambda}_{t+2,T}|\mathcal{Y}_T \vee X_t = e_i] c_{m,i} < Y_{t+1}, f_m > \\
= \;& M \sum_{m=1}^{M}\sum_{j=1}^{N} \bar{E}[< X_{t+1}, e_j > \bar{\Lambda}_{t+2,T}|\mathcal{Y}_T \vee X_t = e_i] c_{m,i} < Y_{t+1}, f_m > \\
= \;& M \sum_{m=1}^{M}\sum_{j=1}^{N} \bar{E}[< X_{t+1}, e_j > \bar{E}[\bar{\Lambda}_{t+2,T}|\mathcal{Y}_T \vee X_t = e_i \vee X_{t+1} = e_j]|\mathcal{Y}_T \vee X_t = e_i] c_{m,i} \cdot \\
& < Y_{t+1}, f_m > \\
= \;& M \sum_{m=1}^{M}\sum_{j=1}^{N} \bar{E}[< X_{t+1}, e_j >< v_{t+1,T}, e_j > |\mathcal{Y}_T \vee X_t = e_i] c_{m,i} < Y_{t+1}, f_m > \\
= \;& M \sum_{m=1}^{M}\sum_{j=1}^{N} a_{j,i} < v_{t+1}, e_j > c_{m,i} < Y_{t+1}, f_m > .
\end{aligned}$$

Then, $v_{t,T} = B(Y_{t+1})A'v_{t+1,T}$. □

Theorem 4.6. *The unnormalized smoothed estimate for the process X, at the time-index k, is given by*

$$\bar{E}[\bar{\Lambda}_{0,T} X_t|\mathcal{Y}_T] = diag < q_t, e_i > v_{t,T}. \tag{4.4.6}$$

Proof.

$$\begin{aligned}
\bar{E}[\bar{\Lambda}_{0,T} < X_t, e_i > |\mathcal{Y}_T] &= \bar{E}[\bar{\Lambda}_{0,t} < X_t, e_i > \bar{E}[\bar{\Lambda}_{t+1,T}|\mathcal{Y}_T \vee X_t = e_i]|\mathcal{Y}_T] \\
&= < q_t, e_i >< v_{t,T}, e_i > .
\end{aligned}$$

So,

$$\begin{aligned} \bar{E}[\bar{\Lambda}_{0,T}X_t|\mathcal{Y}_T] &= \sum_{i=1}^{N} < q_t, e_i >< v_{t,T}, e_i > e_i \\ &= diag< q_t, e_i > v_{t,T}. \end{aligned}$$

□

Write $\sigma_T(N_t^{(j,i)}X_t) \triangleq \bar{E}[\bar{\Lambda}_{0,T}N_t^{(j,i)}X_t|\mathcal{Y}_T]$.

Theorem 4.7. *The smoothed estimate for the quantity $\sigma_T(N_t^{(j,i)}X_t)$ is given by*

$$\sigma_T(N_t^{(j,i)}X_t) = diag< \sigma(N_t^{(j,i)}X_t), e_l > v_{t,T}. \tag{4.4.7}$$

Proof.

$$\begin{aligned} &\bar{E}[\bar{\Lambda}_{0,T}N_t^{(j,i)} < X_t, e_l > |\mathcal{Y}_T] \\ =\ & \bar{E}[\bar{\Lambda}_{0,t}N_t^{(j,i)} < X_t, e_l > \bar{E}[\bar{\Lambda}_{t+1,T}|\mathcal{Y}_T \vee X_t = e_l]|\mathcal{Y}_T] \\ =\ & < \sigma(N_t^{(j,i)}X_t), e_l >< v_{t,T}, e_l > . \end{aligned}$$

So,

$$\begin{aligned} &\sigma_T(N_t^{(j,i)}X_t) \\ =\ & \sum_{l=1}^{N} < \sigma(N_t^{(j,i)}X_t), e_l >< v_{t,T}, e_l > e_l \\ =\ & diag< \sigma(N_t^{(j,i)}X_t), e_l > v_{t,T}. \end{aligned}$$

□

Write $\sigma_T(G_t^iX_t) \triangleq \bar{E}[\bar{\Lambda}_{0,T}G_t^iX_t|\mathcal{Y}_T]$.

Theorem 4.8. *The smoothed estimate for the quantity $\sigma_T(G_t^iX_t)$ is given by*

$$\sigma_T(G_t^iX_t) = diag< \sigma(G_t^iX_t), e_l > v_{t,T}. \tag{4.4.8}$$

Proof.

$$
\begin{aligned}
& \bar{E}[\bar{\Lambda}_{0,T} G_t^i < X_t, e_l > |\mathcal{Y}_T] \\
= \;& \bar{E}[\bar{\Lambda}_{0,t} G_t^i < X_t, e_l > \bar{E}[\bar{\Lambda}_{t+1,T} | \mathcal{Y}_T \vee X_t = e_l] | \mathcal{Y}_T] \\
= \;& < \sigma(G_t^i X_t), e_l >< v_{t,T}, e_l > .
\end{aligned}
$$

So,

$$
\begin{aligned}
& \sigma_T(G_t^i X_t) \\
= \;& \sum_{l=1}^{N} < \sigma(G_t^i X_t), e_l >< v_{t,T}, e_l > e_l \\
= \;& diag < \sigma(G_t^i X_t), e_l > v_{t,T}.
\end{aligned}
$$

□

Write $\sigma_T(J_t^i X_t) \triangleq \bar{E}[\bar{\Lambda}_{0,T} J_t^i X_t | \mathcal{Y}_T]$.

Similarly,

$$
\sigma_T(J_t^i X_t) = diag < \sigma(J_t^i X_t), e_l > v_{t,T}. \tag{4.4.9}
$$

Similarly to section 3, estimates for the parameters are given by

$$
\hat{a}_{j,i} = \frac{\sigma_T(N_T^{(j,i)})}{\sigma_T(J_T^i)}, \tag{4.4.10}
$$

$$
\hat{c}_{j,i} = \frac{\sigma_T(G_T^i)}{\sigma_T(J_T^i)}. \tag{4.4.11}
$$

4.5 Viterbi Algorithm

The basic idea of the Viterbi algorithm is that the expected values represented by summations in the recursive estimates are replaced by maximum likelihoods. That is, the sums are replaced by maxima.

For example, from (4.2.10),

$$q_{t+1}(j) = \sum_{i=1}^{N} \{a_{j,i}[B(Y_{t+1})]_{j,j} q_t(i)\},$$

where $[B(Y_{t+1})]_{j,j} = \sum_{m=1}^{M} c_{m,j} < Y_{t+1}, f_m >$.

Instead of this summation, we recursively define new unnormalized probabilities $q_t^* = [q_t^*(1), q_t^*(2), ..., q_t^*(N)]'$ by

$$q_{t+1}^*(j) = \max_i \{a_{j,i}[B(Y_{t+1})]_{i,i} q_t^*(i)\}. \tag{4.5.1}$$

Certainly,

$$q_{t+1}^*(j) > 0. \tag{4.5.2}$$

We can then define Viterbi probabilities by

$$\rho_{t+1}(j) := \frac{q_{t+1}^*(j)}{\sum_{n=1}^{N} q_{t+1}^*(n)}. \tag{4.5.3}$$

So,

$$\sum_{j=1}^{N} \rho_{t+1}(j) = 1. \tag{4.5.4}$$

$\rho_{t+1}(j)$ is an estimate of the conditional probability, that $X_{t+1} = e_j$, given $Y_1, Y_2, ..., Y_{t+1}$.

The quantity q_t^* is an approximation of q_t. Similarly, we can define $\sigma^*(N_t^{(j,i)} X_t)$, $\sigma^*(J_t^i X_t)$ and $\sigma^*(G_t^i X_t)$ by

$$\sigma^*(N_{t+1}^{(j,i)} X_{t+1})(m) = \max_l a_{m,l}[B(Y_{t+1})]_{l,l} \sigma^*(N_t^{(j,i)} X_t)(l) + q_t^*(i) a_{j,i} \delta_{m,j} [B(Y_{t+1})]_{i,i}. \tag{4.5.5}$$

$$\sigma^*(J_{t+1}^i X_{t+1})(m) = \max_l a_{m,l}[B(Y_{t+1})]_{l,l} \sigma^*(J_t^i X_t)(l) + q_t^*(i) a_{m,i} [B(Y_{t+1})]_{i,i}. \tag{4.5.6}$$

$$\sigma^*(G_{t+1}^i X_{t+1})(m) = \max_l a_{m,l}[B(Y_{t+1})]_{l,l} \sigma^*(G_t^i X_t)(l) + q_t^*(i) f(Y_{t+1}) a_{m,i} [B(Y_{t+1})]_{i,i}. \tag{4.5.7}$$

For example, $\sigma^*(N_t^{(j,i)} < X_t, e_l >)$ is an estimate of the expected value of $N_t^{(j,i)}$ and the probability that $X_t = e_l$, given $Y_1, Y_2, ..., Y_t$.

Following the results in section 3, estimates for the parameters could then be computed by

$$\hat{a}_{j,i} = \frac{\sigma^*(N_t^{(j,i)})}{\sigma^*(J_t^i)}, \tag{4.5.8}$$

$$\hat{c}_{j,i} = \frac{\sigma^*(G_t^i)}{\sigma^*(J_t^i)}. \tag{4.5.9}$$

To demonstrate the performance of the Viterbi filter presented in this chapter, we consider an example. Assume there are two hidden states in the model, and the observations also have two states. The transition matrices A and C are

$$A = \begin{pmatrix} 0.9 & 0.7 \\ 0.1 & 0.3 \end{pmatrix},$$

$$C = \begin{pmatrix} 0.8 & 0.9 \\ 0.2 & 0.1 \end{pmatrix}.$$

The simulation results for the matrices A and C, using the original filter and the Viterbi filter, are shown in Figure 4.1 and Figure 4.2. We can see from the figures that the estimated values converge to the true values of the parameters. The differences of the estimated values and the true values are caused by the noisy system.

A method for estimating X_t is:

Set $\hat{X}_t(I) = 1$ and $\hat{X}_t(n) = 0$, $1 \le n \le N, n \ne I$, where $I = arg\max_j \{\rho_t^*(j)\}$.

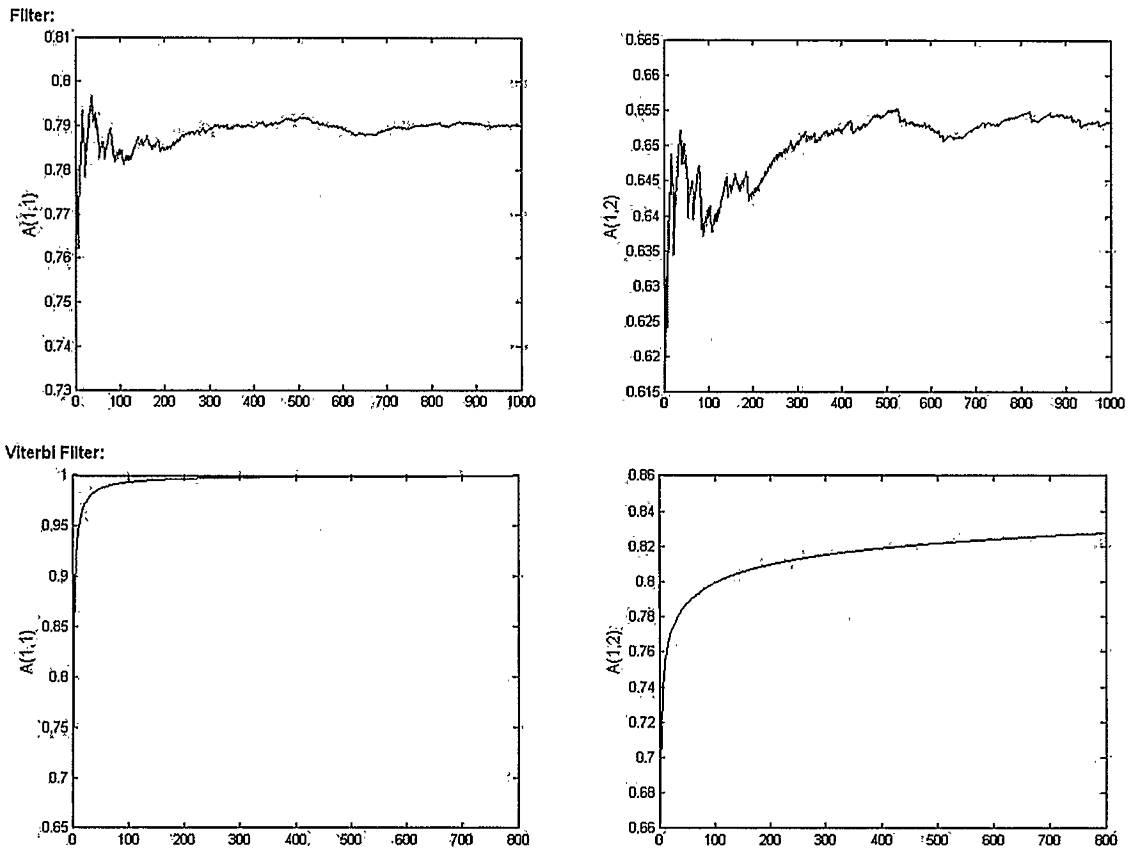


Figure 4.1: Estimated parameter A using the original filter and the Viterbi filter

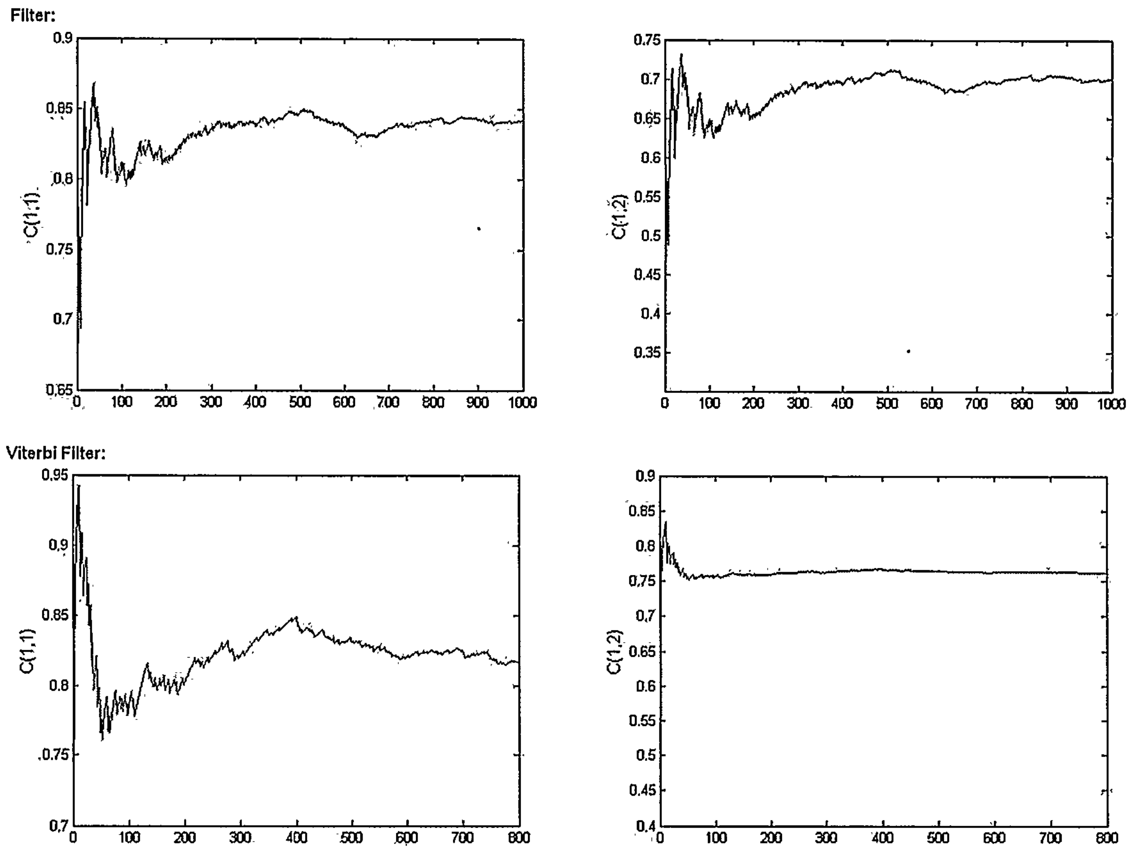


Figure 4.2: Estimated parameter C using the original filter and the Viterbi filter

4.6 Viterbi Smoother

Recall the backward process v defined by

$$v_{t,T} = \bar{E}[\bar{\Lambda}_{t+1} X_t | \mathcal{Y}_T \vee \mathcal{F}_t].$$

Then from Lemma 4.1, $v_{t,T} = B(Y_{t+1})A' v_{t+1,T}$, with $v_{T,T} = \mathbf{1}$.

We now define a Viterbi smoother, by again replacing the sum by a maximum.

Define a process $v^*_{t,T} = [v^*_{t,T}(1), v^*_{t,T}(2), ..., v^*_{t,T}(N)]'$ by

$$v^*_{t,T}(j) = [B(Y_{t+1})]_{j,j}[\max_i \{a_{i,j} v^*_{t+1,T}(i)\}], \tag{4.6.1}$$

$$v^*_{T,T} = \mathbf{1}. \tag{4.6.2}$$

Motivated by the results of section 4.5, we define Viterbi smoothed estimates in the following way.

$$\begin{aligned} \sigma_T^*(N_t^{(j,i)}X_t) &= diag(\sigma^*(N_t^{(j,i)}X_t)) \cdot v_{t,T}^*. & (4.6.3) \\ \sigma_T^*(G_t^iX_t) &= diag(\sigma^*(G_t^iX_t)) \cdot v_{t,T}^*. & (4.6.4) \\ \sigma_T^*(J_t^iX_t) &= diag(\sigma^*(J_t^iX_t)) \cdot v_{t,T}^*. & (4.6.5) \end{aligned}$$

To demonstrate the performance of the Viterbi smoother presented above, we consider the same example as in section 5. The estimated results for the matrices A and C are shown in Figure 4.3 and Figure 4.4. Again, the estimated values converge to the true values of the parameters.

Write

$$q_t'^* = diag < q_t^*, e_i > v_{t,T}^*. \tag{4.6.6}$$

$$\rho_t'^*(j) := \frac{q_t'^*(j)}{\sum_{n=1}^N q_t'^*(n)}. \tag{4.6.7}$$

The method for estimating X_t is:

Set $\hat{X}_t(I) = 1$ and $\hat{X}_t(n) = 0$, $1 \leq n \leq N, n \neq I$, where $I = arg\max_j \{\rho_t'^*(j)\}$.

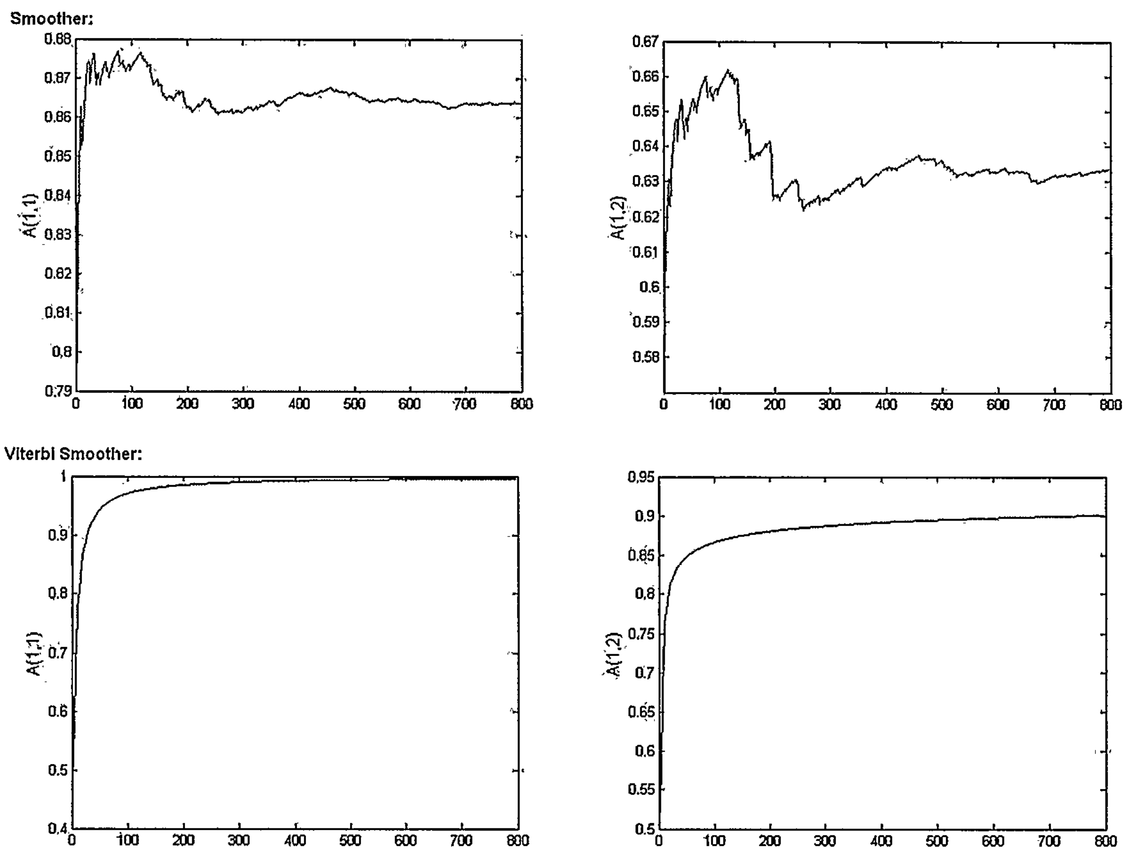


Figure 4.3: Estimated parameter A using the original smoother and the Viterbi smoother

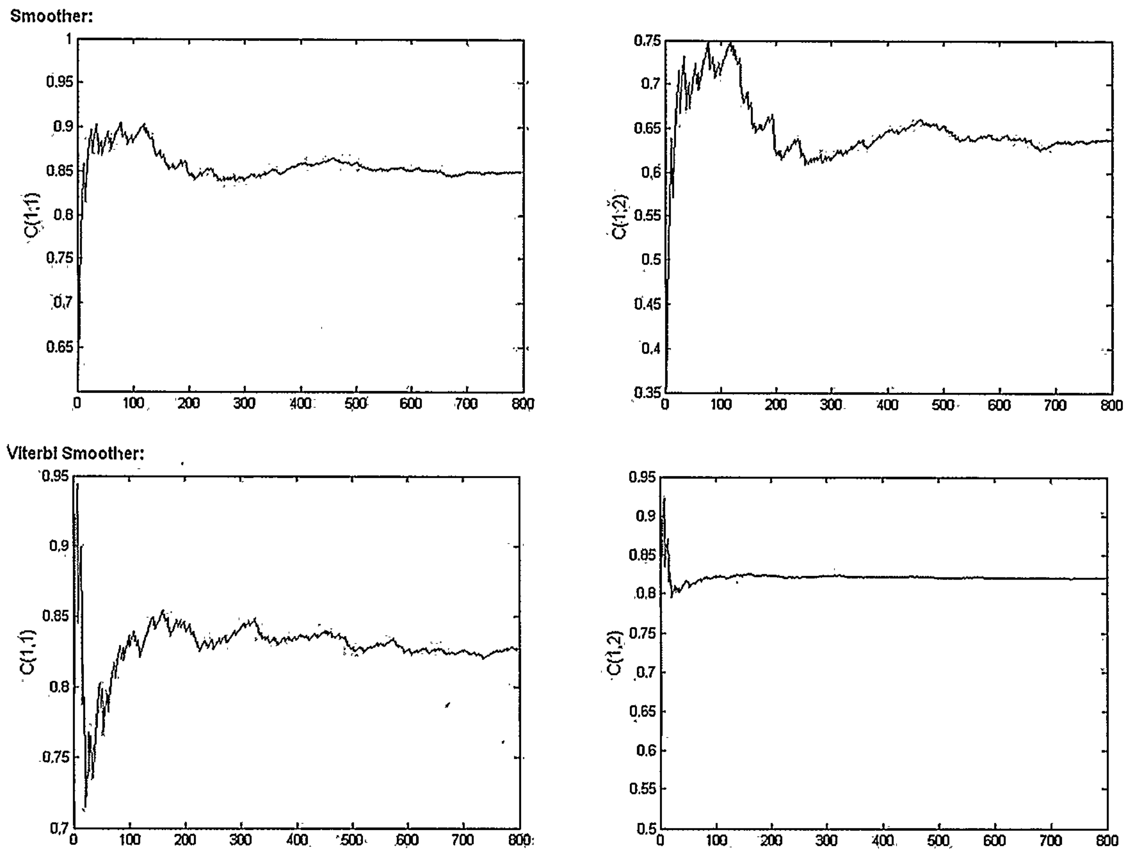


Figure 4.4: Estimated parameter C using the original smoother and the Viterbi smoother

4.7 Conclusions

The Viterbi algorithm can be considered as replacing expected values by maximum likelihoods. We have introduced new Viterbi-type algorithms related to parameter estimation and smoothing.

Chapter 5

A FILTER FOR A HIDDEN MARKOV CHAIN OBSERVED IN FRACTIONAL GAUSSIAN NOISE

5.1 Introduction

In the previous two chapters, we discussed hidden Markov models, where the noise in the observations is assumed to be Gaussian. However, in many practical cases in engineering, physics and finance, it has been observed that the noise in the observations has some long term "memory" correlation. For example, a long-range dependence structure has been noted in squared stock returns and also exchange rates, such as the Yen-Dollar rate. Consequently, a long memory stochastic volatility model has been suggested. [4] [17] Other examples of long term memory are the measurements of IP (Internet Protocol) traffic and the model of Local Area Network (LAN) Ethernet traces. [38] [48] In this chapter we consider a discrete time, finite state Markov chain, observed through a real valued process which is corrupted by fractional Gaussian noise. This is an example of noise with correlation. The relation between the hidden states and the observations is linear. We derive estimates for the parameters and hidden states, using the change of measure method and the EM algorithm. [1] [36] [28] [30] [32] [37]

The chapter is arranged as follows. In the next two sections, we give a description of fractional Gaussian noise and the model used in this chapter. In section 5.4, we describe the change of measure method. In section 5.5, we derive the formulae for estimating the parameters and hidden states. In section 5.6, we derive the formulae for approximately estimating the parameters and hidden states. In section 5.7, we give the Viterbi estimation of the parameters and states. In the final section, we give some conclusions.

5.2 Fractional Differencing

The results are quoted from the paper of Elliott and Miao [31].

Let Z denote the set of integers, $Z = \{..., -2, -1, 0, 1, 2, ...\}$, and Z^+ denote the set of non-negative integers, $Z = \{0, 1, 2, ...\}$. We define a set of functions $\mathcal{L} = \{f\}$ on Z^+ with values in R, i.e., $f : Z^+ \to R$. We suppose that if $i < 0$, then $f(i) = 0$. These functions could be considered as infinite sequences: $f(0) = f_0$, $f(1) = f_1$, $f(2) = f_2$,..., $f(i) = f_i$,....

Definition 5.1. *If $f^1 \in \mathcal{L}$ and $f^2 \in \mathcal{L}$, the convolution product $f^1 * f^2$ is defined by*

$$(f^1 * f^2)(n) = \sum_{i=0}^{\infty} f_i^1 f_{n-i}^2 = \sum_{i=0}^{n} f_i^1 f_{n-i}^2. \tag{5.2.1}$$

Considering the first few terms

$$\begin{aligned}
(f^1 * f^2)(0) &= f_0^1 f_0^2, \\
(f^1 * f^2)(1) &= f_0^1 f_1^2 + f_1^1 f_0^2, \\
(f^1 * f^2)(2) &= f_0^1 f_2^2 + f_1^1 f_1^2 + f_2^1 f_0^2,
\end{aligned}$$

and

$$(f^1 * f^2)(n) = f_0^1 f_n^2 + f_1^1 f_{n-1}^2 + ... + f_{n-1}^1 f_1^2 + f_n^1 f_0^2.$$

If $f^2 = (0, 0, 0, ..., 0, ...)$, then for all $n \in Z^+$,

$$(f^1 * f^2)(n) = \sum_{i=0}^{n} f_i^1 \cdot 0 = 0.$$

In this set of functions, consider the function u, which is defined as

$$u = (u_0, u_1, u_2, ...) = (1, 1, 1, ...). \tag{5.2.2}$$

Then, for any sequence $f = \{f_i, i = 0, 1, 2, ...\}$, and for any $n \in Z^+$,

$$(u * f)(n) = \sum_{i=0}^{n} u_i f_{n-i} = f_1 + f_2 + ... + f_n. \tag{5.2.3}$$

Therefore, convolution with u is the summation operator.

Consider the function $I \in \mathcal{L}$ given by

$$I = (1, 0, 0, ...).$$

Then for any function $f \in \mathcal{L}$,

$$(I * f)(n) = (f * I)(n) = \sum_{i=0}^{n} I_i f_{n-i} = 0 \cdot f_1 + 0 \cdot f_2 + ...0 \cdot f_{n-1} + 1 \cdot f_n = f_n.$$

So, I is the identity operator for convolution multiplication.

The convolution powers of u are

$$\begin{aligned}
u^2 &= u * u = (1, 2, 3, ...), \\
u^3 &= u * u^2 = (1, 3, 6, ...), \\
u^4 &= u * u^3 = (1, 4, 10, ...), \\
&\quad ...... \\
u^k &= (1, \frac{k}{1!}, \frac{k(k+1)}{2!}, \frac{k(k+1)(k+2)}{3!}, ...), \\
&\quad ......
\end{aligned}$$

In fact, for any $r \in R$, $\{u^r\}$ could be defined as in [31]

$$\begin{aligned}
u^r &= (1, \frac{r}{1!}, \frac{r(r+1)}{2!}, \frac{r(r+1)(r+2)}{3!}, ...). \qquad (5.2.4) \\
u^0 &= (1, 0, 0, ..., 0, ...) = I.
\end{aligned}$$

Theorem 5.1. *For any $r, s \in R$,*

$$u^r * u^s = u^{r+s}. \tag{5.2.5}$$

Proof. Write

$$\begin{aligned}
v &= (v_0, v_1, v_2, ...) \\
&= u^r * u^s \\
&= (1, \frac{r}{1!}, \frac{r(r+1)}{2!}, ...) * (1, \frac{s}{1!}, \frac{s(s+1)}{2!}, ...) \\
&= (1, r+s, \frac{r(r+1)}{2!} + rs + \frac{s(s+1)}{2!}, ...).
\end{aligned} \tag{5.2.6}$$

$$u^{r+s} = (1, \frac{r+s}{1!}, \frac{(r+s)(r+s+1)}{2!}, ...).$$

We know that for $|x| < 1$,

$$\begin{aligned}
(1-x)^{-r} &= 1 + rx + \frac{r(r+1)}{2!}x^2 + \frac{r(r+1)(r+2)}{3!}x^3 + ..., \\
(1-x)^{-s} &= 1 + sx + \frac{s(s+1)}{2!}x^2 + \frac{s(s+1)(s+2)}{3!}x^3 + ...
\end{aligned}$$

However,

$$(1-x)^{-r}(1-x)^{-s} = 1 + (r+s)x + (\frac{r(r+1)}{2!} + rs + \frac{s(s+1)}{2!})x^2 + ... \tag{5.2.7}$$

Also,

$$\begin{aligned}
(1-x)^{-r}(1-x)^{-s} &= (1-x)^{-(r+s)} \\
&= 1 + (r+s)x + \frac{(r+s)(r+s+1)}{2!}x^2 + ... \\
&= u_0^{r+s} + u_1^{r+s}x + u_2^{r+s}x + ...
\end{aligned}$$

Comparing (5.2.6) and (5.2.7), we have

$$\begin{aligned}(1-x)^{-r}(1-x)^{-s} &= v_0 + v_1 x + v_2 x^2 + ... \\ &= (1-x)^{-(r+s)} \\ &= u_0^{r+s} + u_1^{r+s} x + u_2^{r+s} x + ...\end{aligned}$$

Since x is arbitrary, we have $v_i = u_i^{r+s}$, for all $i \geq 1$. So, $u^r * u^s = u^{r+s}$. $\square$

Corollary 5.1. *For any $r \in R$,*

$$u^r * u^{-r} = u^0 = I. \tag{5.2.8}$$

5.3 Hidden Markov Model With Fractional Gaussian Noise

In this chapter we consider a finite state time-homogeneous Markov chain $X = \{X_t, t = 0, 1, ...\}$ as in Chapter 2. The transition probabilities and the dynamics of X are defined by equations (2.2.1) and (2.2.2).

Assume $w = \{w_t, t = 0, 1, 2, ...\}$ is a sequence of $N(0, 1)$ independent, identically distributed (i.i.d.) random variables. The fractional Gaussian noise $w^r = \{w_t^r, t = 0, 1, 2, ...\}$ used in this chapter is defined as

$$w_t^r \triangleq (u^r * w)(t) = \sum_{k=0}^{\infty} u_k^r w_{t-k}. \tag{5.3.1}$$

Then, w^r is a sequence of Gaussian random variables which have memory and are correlated. Also,

$$\begin{aligned} E[w_t^r] &= 0, \\ Var(w_t^r) &= \sum_{k=0}^{t} (u_k^r)^2, \\ Cov(w_t^r, w_{t-1}^r) &= \sum_{k=0}^{t-1} u_{t-k}^r u_{t-1-k}^r + 1, \\ Cor(w_t^r, w_{t-1}^r) &= \frac{Cov(w_t^r, w_{t-1}^r)}{\sqrt{Var(w_t^r)Var(w_{t-1}^r)}}. \end{aligned}$$

We suppose the process X is not observed directly; rather, it is observed through another process, whose values are corrupted by fractional Gaussian noise. All functions of X are linear. We consider the following model for the observations:

$$y_t =< g, X_t > +w_t^r, \tag{5.3.2}$$

where g is an N dimensional vector, and $w^r = \{w_t^r, t = 0, 1, ...\}$ is a sequence of fractional Gaussian random variables as described above.

From (5.2.4), u^{-r} is the series

$$(1, \frac{-r}{1!}, \frac{-r(-r+1)}{2!}, \frac{-r(-r+1)(-r+2)}{3!}, ...).$$

Then,

$$\begin{aligned} (u^{-r} * y)(t) &= (u^{-r} * < g, X >)(t) + (u^{-r} * w^r)(t) \\ &= (u^{-r} * < g, X >)(t) + (u^{-r} * u^r * w)(t). \end{aligned}$$

By Theorem 5.1,

$$(u^{-r} * y)(t) = (u^{-r} * < g, X >)(t) + w_t.$$

Write

$$\begin{aligned} z_t &= (u^{-r} * y)(t), \\ \gamma_t(X_0, X_1, ..., X_t) &= (u^{-r} * < g, X >)(t). \end{aligned}$$

Then,

$$\begin{aligned} z_0 &= y_0, \\ z_1 &= y_1 - ry_0, \\ z_2 &= y_2 - ry_1 + \frac{-r(-r+1)}{2!} y_0, \end{aligned}$$

and so on.

Also,

$$\begin{aligned} \gamma_0(X_0) &= < g, X_0 >, \\ \gamma_1(X_0, X_1) &= < g, X_1 > -r < g, X_0 >, \\ \gamma_2(X_0, X_1, X_2) &= < g, X_2 > -r < g, X_1 > + \frac{-r(-r+1)}{2!} < g, X_0 >, \end{aligned}$$

and so on.

Then (5.3.2) implies the following equation.

$$z_t = \gamma_t + w_t. \tag{5.3.3}$$

These are the dynamics of z under the 'real world' probability P.

5.4 Change of Measure

Consider a probability measure $\bar{P}$ on the measurable space $(\Omega, \mathcal{F})$ such that, under $\bar{P}$,

(1) The process X is a finite state Markov chain with transition matrix A,

(2) $\{z_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables.

We call the measure $\bar{P}$ a "reference" probability.

We now construct the 'real world' probability P from $\bar{P}$, such that, under P, the process X is still a finite state Markov chain with transition matrix A, and $\{w_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables, where $w_t = z_t - \gamma_t$

Write $\mathcal{G}_t^z = \sigma\{X_0, z_0, X_1, z_1, ..., X_t, z_t\}$, so $\{\mathcal{G}_t^z\}$ is the filtration generated by (X, z).

Write

$$\lambda_l = \frac{\phi(z_l - \gamma_l)}{\phi(z_l)}, \tag{5.4.1}$$

$$\Lambda_0 = 1, \tag{5.4.2}$$

$$\Lambda_t = \prod_{l=1}^{t} \lambda_l, \quad t = 1, 2, 3, ... \quad , \tag{5.4.3}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})$.

Definition 5.2. *Define P by putting*

$$\frac{dP}{d\bar{P}}|_{\mathcal{G}_t^z} = \Lambda_t. \tag{5.4.4}$$

Theorem 5.2. *Define $w_t = z_t - \gamma_t(X_0, X_1, ..., X_t)$ for $t \in \{0, 1, 2, ...\}$. Then, under P, $\{w_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables.*

Proof.

$$\begin{aligned} P(w_t \leq a | G_{t-1}^z) &= E[I(w_t \leq a) | G_{t-1}^z] \\ &= \frac{\bar{E}[\Lambda_t I(w_t \leq a) | G_{t-1}^z]}{\bar{E}[\Lambda_t | G_{t-1}^z]} \\ &= \frac{\bar{E}[\lambda_t I(w_t \leq a) | G_{t-1}^z]}{\bar{E}[\lambda_t | G_{t-1}^z]}. \end{aligned}$$

$$\bar{E}[\lambda_t | \mathcal{G}_{t-1}^z] = \bar{E}[\bar{E}[\lambda_t | \mathcal{G}_{t-1}^z \vee X_t] | \mathcal{G}_{t-1}^z].$$

The inner expectation

$$\begin{aligned} \bar{E}[\lambda_t | G_{t-1}^z \vee X_t] &= \int_{-\infty}^{\infty} \frac{\phi(z_t - \gamma_t)}{\phi(z_t)} \phi(z_t) dz_t \\ &= \int_{-\infty}^{\infty} \phi(w_t) dw_t \\ &= 1. \end{aligned}$$

Then

$$\bar{E}[\lambda_t | \mathcal{G}^z_{t-1}] = 1.$$

Similarly,

$$\begin{aligned}\bar{E}[\lambda_t I(w_t \le a) | G^z_{t-1} \vee X_t] &= \int_{-\infty}^{\infty} \frac{\phi(z_t - \gamma_t)}{\phi(z_t)} I(w_t \le a)\phi(z_t) dz_t \\ &= \int_{-\infty}^{\infty} \phi(w_t) I(w_t \le a) dw_t.\end{aligned}$$

$$\begin{aligned}\bar{E}[\lambda_t I(w_t \le a) | \mathcal{G}^z_{t-1}] &= \bar{E}[\bar{E}[\lambda_t I(w_t \le a) | \mathcal{G}^z_{t-1} \vee X_t] | \mathcal{G}^z_{t-1}] \\ &= \int_{-\infty}^{\infty} \phi(w_t) I(w_t \le a) dw_t.\end{aligned}$$

So, $P(w_t \le a | G_{t-1}) = P(w_t \le a) = \int_{-\infty}^{\infty} \phi(w_t) I(w_t \le a) dw_t$, and the result follows. That is, under P, $w_t = z_t - \gamma_t(X_0, X_1, ..., X_t)$ is an i.i.d. sequence of $N(0,1)$ random variables. Consequently, $z_t = \gamma_t(X_0, X_1, ..., X_t) + w_t$. □

Corollary 5.2. *Under P, $y_t := (u^r * z)(t) = (u^r * \gamma)(t) + (u^r * w)(t)$. That is,*

$$y_t =< g, X_t > + w_t^r.$$

The process X remains a finite state Markov chain with transition matrix A.

Write

$$\begin{aligned}\mathcal{Y}_t &= \sigma\{y_0, y_1, ..., y_t\}, \\ \mathcal{Z}_t &= \sigma\{z_0, z_1, ..., z_t\}, \\ \mathcal{G}^y_t &= \sigma\{X_0, y_0, X_1, y_1, ..., X_t, y_t, X_{t+1}\}.\end{aligned}$$

Then the filtrations of the X, y, z and the (X, y) processes are $\{\mathcal{F}_t\}$, $\{\mathcal{Y}_t\}$, $\{\mathcal{Z}_t\}$ and $\{\mathcal{G}^y_t\}$. Note that, $\{\mathcal{Y}_t\} = \{\mathcal{Z}_t\}$, $\{\mathcal{G}^y_t\} = \{\mathcal{G}^z_t\}$.

5.5 Exact Estimation of States and Parameters

In this section we shall derive the exact estimates for the parameters and hidden states.

5.5.1 Estimating the Hidden States

First we describe how to estimate the hidden states, given the observations $\{y_t, t = 0, 1, 2, ...\}$.

Write

$$q_t = \bar{E}[\Lambda_t X_t | \mathcal{Z}_t]. \tag{5.5.1}$$

Assume q_0 is known, then we have the following theorem.

Theorem 5.3. *The unnormalized probability vector q_t is computed by*

$$q_t = \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_t=1}^{N} \frac{\phi(z_1 - \gamma_1(X_0, e_{i_t}))...\phi(z_t - \gamma_t(X_0, e_{i_t}, ..., e_{i_1}))}{\phi(z_1)...\phi(z_t)} a_{i_{t-1}, i_t} ... a_{i_1, i_2} e_{i_1} < q_0, e_{i_t} >. \tag{5.5.2}$$

Proof.

$$
\begin{aligned}
& q_t \\
= \;& \bar{E}[\Lambda_t X_t | \mathcal{Z}_t] \\
= \;& \bar{E}[\Lambda_{t-1} \frac{\phi(z_t - \gamma_t(X_0, ..., X_t))}{\phi(z_t)} X_t | \mathcal{Z}_t] \\
= \;& \bar{E}[\Lambda_{t-1} \frac{\phi(z_t - \gamma_t(X_0, ..., X_t))}{\phi(z_t)} \sum_{i_1=1}^{N} < X_t, e_{i_1} > X_t | \mathcal{Z}_t] \\
= \;& \sum_{i_1=1}^{N} \bar{E}[\Lambda_{t-1} \phi(z_t - \gamma_t(X_0, ..., X_{t-1}, e_{i_1})) < AX_{t-1} + M_t, e_{i_1} > |\mathcal{Z}_t] \frac{e_{i_1}}{\phi(z_t)} \\
= \;& \sum_{i_1=1}^{N} \bar{E}[\Lambda_{t-1} \phi(z_t - \gamma_t(X_0, ..., X_{t-1}, e_{i_1})) < AX_{t-1}, e_{i_1} > |\mathcal{Z}_t] \frac{e_{i_1}}{\phi(z_t)} \\
= \;& \sum_{i_1=1}^{N} \bar{E}[\Lambda_{t-2} \frac{\phi(z_{t-1} - \gamma_{t-1}(X_0, ..., X_{t-1}))}{\phi(z_{t-1})} \phi(z_t - \gamma_t(X_0, ..., X_{t-1}, e_{i_1})) \cdot \\
& \sum_{i_2=1}^{N} < X_{t-1}, e_{i_2} > < AX_{t-1}, e_{i_1} > |\mathcal{Z}_t] \frac{e_{i_1}}{\phi(z_t)} \\
= \;& \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \bar{E}[\Lambda_{t-2} \phi(z_{t-1} - \gamma_{t-1}(X_0, ..., X_{t-2}, e_{i_2})) \phi(z_t - \gamma_t(X_0, ..., X_{t-2}, e_{i_2}, e_{i_1})) \cdot \\
& < AX_{t-2} + M_{t-1}, e_{i_2} > |\mathcal{Z}_t] \frac{a_{i_1,i_2} e_{i_1}}{\phi(z_{t-1})\phi(z_t)} \\
& \ldots\ldots \\
= \;& \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_t=1}^{N} \bar{E}[\Lambda_0 \phi(z_1 - \gamma_1(X_0, e_{i_t})) ... \phi(z_t - \gamma_t(X_0, e_{i_t}, ..., e_{i_1})) < AX_0, e_{i_t} > |\mathcal{Z}_t] \cdot \\
& \frac{a_{i_{t-1},i_t} ... a_{i_1,i_2} e_{i_1}}{\phi(z_1) ... \phi(z_t)} \\
= \;& \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_t=1}^{N} \frac{\phi(z_1 - \gamma_1(X_0, e_{i_t})) ... \phi(z_t - \gamma_t(X_0, e_{i_t}, ..., e_{i_1}))}{\phi(z_1) ... \phi(z_t)} a_{i_{t-1},i_t} ... a_{i_1,i_2} e_{i_1} < q_0, e_{i_t} >.
\end{aligned}
$$

□

Similarly as before,

$$
\hat{X}_t = \frac{q_t}{< q_t, \mathbf{1} >}. \tag{5.5.3}
$$

5.5.2 Estimating the Parameters

In order to estimate the parameters in this model, we have to estimate several random processes, as in the previous chapters.

Theorem 5.4. *The vector $\sigma(N_t^{(j,i)}X_t)$ is computed by*

$$\begin{aligned}
&\sigma(N_t^{(j,i)}X_t) \\
&= \sum_{l_1=1}^{N} \cdots \sum_{l_{t-1}=1}^{N} \{< q_0, e_{l_{t-1}} > [\lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) \ldots \lambda_t(e_{l_{t-1}}, \ldots, e_{l_1}, e_i, e_j) a_{l_{t-2},l_{t-1}} \ldots a_{l_1,l_2} a_{i,l_1} a_{ji} e_j + \\
&\ldots + \lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) \ldots \lambda_t(e_{l_{t-1}}, \ldots, e_{l_{t-k+1}}, e_i, e_j, e_{l_{t-k}}, \ldots, e_{l_1}) \cdot \\
&a_{l_{t-2},l_{t-1}} \ldots a_{i,l_{t-k+1}} a_{ji} a_{l_{t-k},j} a_{l_{t-k-1},l_{t-k}} \ldots a_{l_1,l_2} e_{l_1} + \ldots] + \\
&< q_0, e_i > \lambda_1(e_i, e_j) \lambda_2(e_i, e_j, e_{l_{t-1}}) \ldots \lambda_t(e_i, e_j, e_{l_{t-1}}, \ldots, e_{l_1}) a_{ji} a_{l_{t-1},j} a_{l_{t-2},l_{t-1}} \ldots a_{l_1,l_2} e_{l_1}\}.
\end{aligned} \tag{5.5.4}$$

Proof.

$$\begin{aligned}
\sigma(N_t^{(j,i)}X_t) &= \bar{E}[\Lambda_t N_t^{(j,i)} X_t | \mathcal{Z}_t] \\
&= \bar{E}[\Lambda_t \sum_{l=1}^{t} < X_{l-1}, e_i >< X_l, e_j > X_t | \mathcal{Z}_t] \\
&= \sum_{l=1}^{t} \bar{E}[\Lambda_t < X_{l-1}, e_i >< X_l, e_j > X_t | \mathcal{Z}_t].
\end{aligned}$$

$$
\begin{aligned}
&\bar{E}[\Lambda_t < X_{t-1}, e_i >< X_t, e_j > X_t | \mathcal{Z}_t] \\
= \quad & \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_t) < X_{t-1}, e_i >< AX_{t-1} + M_t, e_j > |\mathcal{Z}_t] e_j \\
= \quad & \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_{t-1}, e_j) < X_{t-1}, e_i >< AX_{t-1}, e_j > |\mathcal{Z}_t] e_j \\
= \quad & \bar{E}[\Lambda_{t-2}\lambda_{t-1}(X_0, ..., X_{t-2}, e_i)\lambda_t(X_0, ..., X_{t-2}, e_i, e_j) < AX_{t-2} + M_{t-1}, e_i > |\mathcal{Z}_t] a_{ji} e_j \\
= \quad & \bar{E}[\Lambda_{t-2}\lambda_{t-1}(X_0, ..., X_{t-2}, e_i)\lambda_t(X_0, ..., X_{t-2}, e_i, e_j) \sum_{l_1=1}^{N} < X_{t-2}, e_{l_1} > < AX_{t-2}, e_i > |\mathcal{Z}_t] \cdot \\
& a_{ji} e_j \\
= \quad & \sum_{l_1=1}^{N} \bar{E}[\Lambda_{t-2}\lambda_{t-1}(X_0, ..., X_{t-3}, e_{l_1}, e_i)\lambda_t(X_0, ..., X_{t-3}, e_{l_1}, e_i, e_j) < X_{t-2}, e_{l_1} > |\mathcal{Z}_t] a_{i,l_1} a_{ji} e_j \\
\ldots\ldots & \\
= \quad & \sum_{l_1=1}^{N} \cdots \sum_{l_{t-1}=1}^{N} \bar{E}[\Lambda_0 \lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) ... \lambda_t(e_{l_{t-1}}, ..., e_{l_1}, e_i, e_j) < X_0, e_{l_{t-1}} > |\mathcal{Z}_t] \cdot \\
& a_{l_{t-2}, l_{t-1}} ... a_{l_1, l_2} a_{i, l_1} a_{ji} e_j \\
= \quad & \sum_{l_1=1}^{N} \cdots \sum_{l_{t-1}=1}^{N} < q_0, e_{l_{t-1}} > \lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) ... \lambda_t(e_{l_{t-1}}, ..., e_{l_1}, e_i, e_j) a_{l_{t-2}, l_{t-1}} ... a_{l_1, l_2} a_{i, l_1} a_{ji} e_j.
\end{aligned}
$$

For $0 < k < t$,

$$
\begin{aligned}
& \bar{E}[\Lambda_t < X_{k-1}, e_i >< X_k, e_j > X_t | \mathcal{Z}_t] \\
= \; & \bar{E}[\Lambda_t < X_{k-1}, e_i >< X_k, e_j > \sum_{l_1=1}^{N} < X_t, e_{l_1} > X_t | \mathcal{Z}_t] \\
= \; & \sum_{l_1=1}^{N} \bar{E}[\Lambda_{t-1} \lambda_t(X_0, ..., X_{t-1}, e_{l_1}) < X_{k-1}, e_i >< X_k, e_j >< AX_{t-1}, e_{l_1} > | \mathcal{Z}_t] e_{l_1} \\
= \; & \sum_{l_1=1}^{N} \bar{E}[\Lambda_{t-1} \lambda_t(X_0, ..., X_{t-1}, e_{l_1}) < X_{k-1}, e_i >< X_k, e_j > \cdot \\
& \sum_{l_2=1}^{N} < X_{t-1}, e_{l_2} > < AX_{t-1}, e_{l_1} > | \mathcal{Z}_t] e_{l_1} \\
= \; & \sum_{l_1=1}^{N} \sum_{l_2=1}^{N} \bar{E}[\Lambda_{t-2} \lambda_{t-1}(X_0, ..., X_{t-2}, e_{l_2}) \lambda_t(X_0, ..., X_{t-2}, e_{l_2}, e_{l_1}) < X_{k-1}, e_i >< X_k, e_j > \cdot \\
& < AX_{t-2}, e_{l_2} > | \mathcal{Z}_t] a_{l_1, l_2} e_{l_1}
\end{aligned}
$$

......

$$
\begin{aligned}
= \; & \sum_{l_1=1}^{N} \cdots \sum_{l_{t-k}=1}^{N} \bar{E}[\Lambda_k \lambda_{k+1}(X_0, ..., X_k, e_{l_{t-k}}) ... \lambda_t(X_0, ..., X_k, e_{l_{t-k}}, ..., e_{l_1}) < X_{k-1}, e_i > \cdot \\
& < X_k, e_j >< AX_k, e_{l_{t-k}} > | \mathcal{Z}_t] a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} e_{l_1} \\
= \; & \sum_{l_1=1}^{N} \cdots \sum_{l_{t-k}=1}^{N} \bar{E}[\Lambda_k \lambda_{k+1}(X_0, ..., X_{k-1}, e_j, e_{l_{t-k}}) ... \lambda_t(X_0, ..., X_{k-1}, e_j, e_{l_{t-k}}, ..., e_{l_1}) \cdot \\
& < X_{k-1}, e_i >< AX_{k-1}, e_j > | \mathcal{Z}_t] a_{l_{t-k}, j} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} e_{l_1} \\
= \; & \sum_{l_1=1}^{N} \cdots \sum_{l_{t-k}=1}^{N} \bar{E}[\Lambda_k \lambda_{k+1}(X_0, ..., X_{k-1}, e_j, e_{l_{t-k}}) ... \lambda_t(X_0, ..., X_{k-1}, e_j, e_{l_{t-k}}, ..., e_{l_1}) \cdot \\
& < X_{k-1}, e_i > | \mathcal{Z}_t] a_{ji} a_{l_{t-k}, j} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} e_{l_1} \\
= \; & \sum_{l_1=1}^{N} \cdots \sum_{l_{t-k}=1}^{N} \bar{E}[\Lambda_{k-2} \lambda_{k-1}(X_0, ..., X_{k-2}, e_i) ... \lambda_t(X_0, ..., X_{k-2}, e_i, e_j, e_{l_{t-k}}, ..., e_{l_1}) \cdot \\
& \sum_{l_{t-k+1}=1}^{N} < X_{k-2}, e_{l_{t-k+1}} > < AX_{k-2}, e_i > | \mathcal{Z}_t] a_{ji} a_{l_{t-k}, j} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} e_{l_1}
\end{aligned}
$$

$$= \sum_{l_1=1}^{N} \cdots \sum_{l_{t-k+1}=1}^{N} \bar{E}[\Lambda_{k-2} \cdot$$
$$\lambda_{k-1}(X_0, ..., X_{k-3}, e_{l_{t-k+1}}, e_i) ... \lambda_t(X_0, ..., X_{k-3}, e_{l_{t-k+1}}, e_i, e_j, e_{l_{t-k}}, ..., e_{l_1}) \cdot$$
$$< X_{k-2}, e_{l_{t-k+1}} > |\mathcal{Z}_t] a_{i,l_{t-k+1}} a_{ji} a_{l_{t-k},j} a_{l_{t-k-1},l_{t-k}} ... a_{l_1,l_2} e_{l_1}$$

......

$$= \sum_{l_1=1}^{N} \cdots \sum_{l_{t-1}=1}^{N} \bar{E}[\Lambda_0 \lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) ... \lambda_t(e_{l_{t-1}}, ..., e_{l_{t-k+1}}, e_i, e_j, e_{l_{t-k}}, ..., e_{l_1}) < X_0, e_{l_{t-1}} > |\mathcal{Z}_t] \cdot$$
$$a_{l_{t-2},l_{t-1}} ... a_{i,l_{t-k+1}} a_{ji} a_{l_{t-k},j} a_{l_{t-k-1},l_{t-k}} ... a_{l_1,l_2} e_{l_1}$$
$$= \sum_{l_1=1}^{N} \cdots \sum_{l_{t-1}=1}^{N} < q_0, e_{l_{t-1}} > \lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) ... \lambda_t(e_{l_{t-1}}, ..., e_{l_{t-k+1}}, e_i, e_j, e_{l_{t-k}}, ..., e_{l_1}) \cdot$$
$$a_{l_{t-2},l_{t-1}} ... a_{i,l_{t-k+1}} a_{ji} a_{l_{t-k},j} a_{l_{t-k-1},l_{t-k}} ... a_{l_1,l_2} e_{l_1}.$$

So,

$$\sigma(N_t^{(j,i)} X_t)$$
$$= \sum_{l_1=1}^{N} \cdots \sum_{l_{t-1}=1}^{N} \{< q_0, e_{l_{t-1}} > [\lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) ... \lambda_t(e_{l_{t-1}}, ..., e_{l_1}, e_i, e_j) a_{l_{t-2},l_{t-1}} ... a_{l_1,l_2} a_{i,l_1} a_{ji} e_j +$$
$$... + \lambda_1(e_{l_{t-1}}, e_{l_{t-2}}) ... \lambda_t(e_{l_{t-1}}, ..., e_{l_{t-k+1}}, e_i, e_j, e_{l_{t-k}}, ..., e_{l_1}) \cdot$$
$$a_{l_{t-2},l_{t-1}} ... a_{i,l_{t-k+1}} a_{ji} a_{l_{t-k},j} a_{l_{t-k-1},l_{t-k}} ... a_{l_1,l_2} e_{l_1} + ...] +$$
$$< q_0, e_i > \lambda_1(e_i, e_j) \lambda_2(e_i, e_j, e_{l_{t-1}}) ... \lambda_t(e_i, e_j, e_{l_{t-1}}, ..., e_{l_1}) a_{ji} a_{l_{t-1},j} a_{l_{t-2},l_{t-1}} ... a_{l_1,l_2} e_{l_1}\}.$$

□

Theorem 5.5. *The probability vector $\sigma(G_t^i X_t)$ is computed by*

$$\begin{aligned}
&\sigma(G_t^i X_t)\\
=\ & \sum_{l_1=1}^{N}\cdots\sum_{l_t=1}^{N}\{<q_0, e_{l_t}> [\lambda_1(e_{l_t}, e_{l_{t-1}})\ldots\lambda_t(e_{l_t}, e_{l_{t-1}}\ldots, e_{l_1}, e_i)a_{l_{t-1},l_t}\ldots a_{l_1,l_2}a_{i,l_1}f(z_t)e_i +\\
&+\ldots + \lambda_1(e_{l_t}, e_{l_{t-1}})\ldots\lambda_t(e_{l_t}, \ldots, e_{l_{t-k+1}}, e_i, e_{l_{t-k}}, \ldots, e_{l_1})\cdot\\
&a_{l_{t-1},l_t}\ldots a_{i,l_{t-k+1}}a_{l_{t-k},i}a_{l_{t-k-1},l_{t-k}}\ldots a_{l_1,l_2}f(z_k)e_{l_1} + \ldots] +\\
&<q_0, e_i> \lambda_1(e_i, e_{l_t})\ldots\lambda_t(e_i, e_{l_t}, \ldots, e_{l_1})a_{l_t,i}a_{l_{t-1},l_t}a_{l_{t-2},l_{t-1}}\ldots a_{l_1,l_2}f(z_0)e_{l_1}\}.
\end{aligned} \tag{5.5.5}$$

Proof.

$$\begin{aligned}
\sigma(G_t^i X_t) &= \bar{E}[\Lambda_t G_t^i X_t | \mathcal{Z}_t]\\
&= \bar{E}[\Lambda_t \sum_{l=1}^{t} f(z_l) < X_l, e_i > X_t | \mathcal{Z}_t]\\
&= \sum_{l=1}^{t} \bar{E}[\Lambda_t f(z_l) < X_l, e_i > X_t | \mathcal{Z}_t].
\end{aligned}$$

$$\begin{aligned}
&\bar{E}[\Lambda_t f(z_t) < X_t, e_i > X_t | \mathcal{Z}_t]\\
=\ & \bar{E}[\Lambda_t < X_t, e_i > | \mathcal{Z}_t] f(z_t) e_i\\
=\ & \bar{E}[\Lambda_{t-1}\lambda_t(X_0, \ldots, X_{t-1}, e_i) < AX_{t-1} + M_t, e_i > | \mathcal{Z}_t] f(z_t) e_i\\
=\ & \bar{E}[\Lambda_{t-1}\lambda_t(X_0, \ldots, X_{t-1}, e_i) \sum_{l_1=1}^{N} < X_{t-1}, e_{l_1} > < AX_{t-1}, e_i > | \mathcal{Z}_t] f(z_t) e_i\\
=\ & \sum_{l_1=1}^{N} \bar{E}[\Lambda_{t-1}\lambda_t(X_0, \ldots, X_{t-2}, e_{l_1}, e_i) < X_{t-1}, e_{l_1} > | \mathcal{Z}_t] a_{i,l_1} f(z_t) e_i\\
&\ldots\ldots\\
=\ & \sum_{l_1=1}^{N}\cdots\sum_{l_t=1}^{N} \bar{E}[\Lambda_0\lambda_1(e_{l_t}, e_{l_{t-1}})\ldots\lambda_t(e_{l_t}, e_{l_{t-1}}\ldots, e_{l_1}, e_i) < X_0, e_{l_t} > | \mathcal{Z}_t] a_{l_{t-1},l_t}\ldots a_{l_1,l_2}a_{i,l_1} f(z_t) e_i\\
=\ & \sum_{l_1=1}^{N}\cdots\sum_{l_t=1}^{N} < q_0, e_{l_t} > \lambda_1(e_{l_t}, e_{l_{t-1}})\ldots\lambda_t(e_{l_t}, e_{l_{t-1}}\ldots, e_{l_1}, e_i) a_{l_{t-1},l_t}\ldots a_{l_1,l_2}a_{i,l_1} f(z_t) e_i.
\end{aligned}$$

For $0 \leq k < t$,

$$
\begin{aligned}
& \bar{E}[\Lambda_t f(z_k) < X_k, e_i > X_t | \mathcal{Z}_t] \\
= \;& \bar{E}[\Lambda_t < X_k, e_i > \sum_{l_1=1}^{N} < X_t, e_{l_1} > X_t | \mathcal{Z}_t] f(z_k) \\
= \;& \sum_{l_1=1}^{N} \bar{E}[\Lambda_{t-1} \lambda_t(X_0, ..., X_{t-1}, e_{l_1}) < X_k, e_i >< AX_{t-1}, e_{l_1} > |\mathcal{Z}_t] f(z_k) e_{l_1} \\
= \;& \sum_{l_1=1}^{N} \bar{E}[\Lambda_{t-1} \lambda_t(X_0, ..., X_{t-1}, e_{l_1}) < X_k, e_i > \sum_{l_2=1}^{N} < X_{t-1}, e_{l_2} > < AX_{t-1}, e_{l_1} > |\mathcal{Z}_t] f(z_k) e_{l_1} \\
= \;& \sum_{l_1=1}^{N} \sum_{l_2=1}^{N} \bar{E}[\Lambda_{t-1} \lambda_t(X_0, ..., X_{t-2}, e_{l_2}, e_{l_1}) < X_k, e_i >< X_{t-1}, e_{l_2} > |\mathcal{Z}_t] a_{l_1, l_2} f(z_k) e_{l_1}
\end{aligned}
$$

......

$$
\begin{aligned}
= \;& \sum_{l_1=1}^{N} ... \sum_{l_{t-k}=1}^{N} \bar{E}[\Lambda_k \lambda_{k+1}(X_0, ..., X_{k-1}, e_i, e_{l_{t-k}}) ... \lambda_t(X_0, ..., X_{k-1}, e_i, e_{l_{t-k}}, ..., e_{l_1}) < X_k, e_i > \cdot \\
& < AX_k, e_{l_{t-k}} > |\mathcal{Z}_t] a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} f(z_k) e_{l_1} \\
= \;& \sum_{l_1=1}^{N} ... \sum_{l_{t-k}=1}^{N} \bar{E}[\Lambda_k \lambda_{k+1}(X_0, ..., X_{k-1}, e_i, e_{l_{t-k}}) ... \lambda_t(X_0, ..., X_{k-1}, e_i, e_{l_{t-k}}, ..., e_{l_1}) \cdot \\
& < X_k, e_i > |\mathcal{Z}_t] a_{l_{t-k}, i} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} f(z_k) e_{l_1} \\
= \;& \sum_{l_1=1}^{N} ... \sum_{l_{t-k}=1}^{N} \bar{E}[\Lambda_{k-1} \lambda_k(X_0, ..., X_{k-1}, e_i) ... \lambda_t(X_0, ..., X_{k-1}, e_i, e_{l_{t-k}}, ..., e_{l_1}) \cdot \\
& \sum_{l_{t-k+1}=1}^{N} < X_{k-1}, e_{l_{t-k+1}} > < AX_{k-1}, e_i > |\mathcal{Z}_t] a_{l_{t-k}, i} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} f(z_k) e_{l_1} \\
= \;& \sum_{l_1=1}^{N} ... \sum_{l_{t-k+1}=1}^{N} \bar{E}[\Lambda_{k-1} \lambda_k(X_0, ..., X_{k-2}, e_{l_{t-k+1}}, e_i) ... \lambda_t(X_0, ..., X_{k-2}, e_{l_{t-k+1}}, e_i, e_{l_{t-k}}, ..., e_{l_1}) \cdot \\
& < X_{k-1}, e_{l_{t-k+1}} > |\mathcal{Z}_t] a_{i, l_{t-k+1}} a_{l_{t-k}, i} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} f(z_k) e_{l_1}
\end{aligned}
$$

......

$$
\begin{aligned}
= \;& \sum_{l_1=1}^{N} ... \sum_{l_t=1}^{N} \bar{E}[\Lambda_0 \lambda_1(e_{l_t}, e_{l_{t-1}}) ... \lambda_t(e_{l_t}, ..., e_{l_{t-k+1}}, e_i, e_{l_{t-k}}, ..., e_{l_1}) < X_0, e_{l_t} > |\mathcal{Z}_t] \cdot \\
& a_{l_{t-1}, l_t} ... a_{i, l_{t-k+1}} a_{l_{t-k}, i} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} f(z_k) e_{l_1} \\
= \;& \sum_{l_1=1}^{N} ... \sum_{l_t=1}^{N} < q_0, e_{l_t} > \lambda_1(e_{l_t}, e_{l_{t-1}}) ... \lambda_t(e_{l_t}, ..., e_{l_{t-k+1}}, e_i, e_{l_{t-k}}, ..., e_{l_1}) \cdot \\
& a_{l_{t-1}, l_t} ... a_{i, l_{t-k+1}} a_{l_{t-k}, i} a_{l_{t-k-1}, l_{t-k}} ... a_{l_1, l_2} f(z_k) e_{l_1}.
\end{aligned}
$$

So,

$$
\begin{aligned}
&\sigma(G_t^i X_t) \\
= &\sum_{l_1=1}^{N} \cdots \sum_{l_t=1}^{N} \{ < q_0, e_{l_t} > [\lambda_1(e_{l_t}, e_{l_{t-1}}) \ldots \lambda_t(e_{l_t}, e_{l_{t-1}} \ldots, e_{l_1}, e_i) a_{l_{t-1}, l_t} \ldots a_{l_1, l_2} a_{i, l_1} f(z_t) e_i + \\
&+ \ldots + \lambda_1(e_{l_t}, e_{l_{t-1}}) \ldots \lambda_t(e_{l_t}, \ldots, e_{l_{t-k+1}}, e_i, e_{l_{t-k}}, \ldots, e_{l_1}) \cdot \\
&a_{l_{t-1}, l_t} \ldots a_{i, l_{t-k+1}} a_{l_{t-k}, i} a_{l_{t-k-1}, l_{t-k}} \ldots a_{l_1, l_2} f(z_k) e_{l_1} + \ldots] + \\
&< q_0, e_i > \lambda_1(e_i, e_{l_t}) \ldots \lambda_t(e_i, e_{l_t}, \ldots, e_{l_1}) a_{l_t, i} a_{l_{t-1}, l_t} a_{l_{t-2}, l_{t-1}} \ldots a_{l_1, l_2} f(z_0) e_{l_1} \}.
\end{aligned}
$$

□

In (5.5.5), let $f(z_l) = 1$, $0 \leq l \leq t$. Then

$$
\begin{aligned}
\sigma(J_t^i X_t) &\triangleq \bar{E}[\bar{\Lambda}_t J_t^i X_t | \mathcal{Z}_t] \\
&= \sum_{l_1=1}^{N} \cdots \sum_{l_t=1}^{N} \{ < q_0, e_{l_t} > [\lambda_1(e_{l_t}, e_{l_{t-1}}) \ldots \lambda_t(e_{l_t}, e_{l_{t-1}} \ldots, e_{l_1}, e_i) a_{l_{t-1}, l_t} \ldots a_{l_1, l_2} a_{i, l_1} e_i + \\
&\quad + \ldots + \lambda_1(e_{l_t}, e_{l_{t-1}}) \ldots \lambda_t(e_{l_t}, \ldots, e_{l_{t-k+1}}, e_i, e_{l_{t-k}}, \ldots, e_{l_1}) \cdot \\
&\quad a_{l_{t-1}, l_t} \ldots a_{i, l_{t-k+1}} a_{l_{t-k}, i} a_{l_{t-k-1}, l_{t-k}} \ldots a_{l_1, l_2} e_{l_1} + \ldots] + \\
&\quad < q_0, e_i > \lambda_1(e_i, e_{l_t}) \ldots \lambda_t(e_i, e_{l_t}, \ldots, e_{l_1}) a_{l_t, i} a_{l_{t-1}, l_t} a_{l_{t-2}, l_{t-1}} \ldots a_{l_1, l_2} e_{l_1} \}.
\end{aligned}
\tag{5.5.6}
$$

Again,

$$
\begin{aligned}
\sigma(N_t^{(j,i)}) &= < \sigma(N_t^{(j,i)}, X_t), \mathbf{1} >, \\
\sigma(G_t^i) &= < \sigma(G_t^i, X_t), \mathbf{1} >, \\
\sigma(J_t^i) &= < \sigma(J_t^i, X_t), \mathbf{1} > .
\end{aligned}
$$

The parameters are again estimated using the EM algorithm discussed in Chapter 2.

Given the observations up to time t, the EM estimates of $a_{j,i}$ and $\gamma_t(X_0, \hat{X}_1, ..., \hat{X}_{t-1}, e_i)$ are

$$\begin{aligned} \hat{a}_{j,i} &= \frac{\sigma(N_t^{(j,i)})}{\sigma(J_{t-1}^i)}, & (5.5.7) \\ \hat{\gamma}_t(X_0, \hat{X}_1, ..., \hat{X}_{t-1}, e_i) &= \frac{\sigma(G_t^i)}{\sigma(J_t^i)}, & (5.5.8) \end{aligned}$$

where in Theorem 5.5, $f(z_l) = z_l$ for G_t^i. The proof is similar to the proof of Theorem 2.7 and Theorem 2.8 of Chapter 2.

Then, the estimator for g_i up to time t is

$$\hat{g}_i(t) = (u^r * \hat{\gamma}(X_0, \hat{X}_1, ..., \hat{X}_{t-1}, e_i))(t). \tag{5.5.9}$$

5.6 Approximate Estimation of States and Parameters

In this section we give recursive approximate estimates of the parameters and hidden states.

Theorem 5.6. *The probability vector $\tilde{q}_t$ is approximately computed by the recursion*

$$\tilde{q}_t = \sum_{i=1}^{N} < A\tilde{q}_{t-1}, e_i > \frac{\phi(z_t - \gamma_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i))}{\phi(z_t)} e_i, \tag{5.6.1}$$

where, for $0 < l \le t$, $\tilde{X}_l = \frac{\tilde{q}_l}{<\tilde{q}_l, 1>}$.

Proof.

$$\begin{aligned} & q_t \\ = \ & \bar{E}[\Lambda_t X_t | \mathcal{Z}_t] \\ = \ & \bar{E}[\Lambda_{t-1} \frac{\phi(z_t - \gamma_t(X_0, ..., X_t))}{\phi(z_t)} X_t | \mathcal{Z}_t] \\ = \ & \bar{E}[\Lambda_{t-1} \frac{\phi(z_t - \gamma_t(X_0, ..., X_t))}{\phi(z_t)} \sum_{i=1}^{N} < X_t, e_i > X_t | \mathcal{Z}_t]. \qquad (5.6.2) \end{aligned}$$

Using $\tilde{X}_i$ to approximate X_i, $i = 1, 2, ..., t-1$ in (5.6.2), we define

$$\begin{aligned}
&\tilde{q}_t \\
= &\sum_{i=1}^{N} \bar{E}[\Lambda_{t-1} < AX_{t-1} + M_t, e_i > |\mathcal{Z}_t] \frac{\phi(z_t - \gamma_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i))}{\phi(z_t)} e_i \\
= &\sum_{i=1}^{N} < \bar{E}[\Lambda_{t-1} AX_{t-1}|\mathcal{Z}_t], e_i > \frac{\phi(z_t - \gamma_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i))}{\phi(z_t)} e_i \\
= &\sum_{i=1}^{N} < Aq_{t-1}, e_i > \frac{\phi(z_t - \gamma_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i))}{\phi(z_t)} e_i,
\end{aligned}$$

where, for $0 < l \leq t$, $\tilde{X}_l = \frac{\tilde{q}_l}{<\tilde{q}_l, 1>}$.

Taking $\tilde{q}_{t-1}$ as an approximation to q_{t-1}, we have

$$\tilde{q}_t = \sum_{i=1}^{N} < A\tilde{q}_{t-1}, e_i > \frac{\phi(z_t - \gamma_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i))}{\phi(z_t)} e_i.$$

□

Theorem 5.7. *The vector $\sigma(N_t^{(j,i)} X_t)$ is approximately computed by the recursion*

$$\begin{aligned}
\tilde{\sigma}(N_t^{(j,i)} X_t) \;=\; &\sum_{l=1}^{N} < \tilde{\sigma}(N_{t-1}^{(j,i)} X_{t-1}), e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l) e_l + \\
&< \tilde{q}_{t-1}, e_i > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_j) a_{j,i} e_j. \qquad (5.6.3)
\end{aligned}$$

Proof.

$$\begin{aligned}
&\sigma(N_t^{(j,i)} X_t) \\
= &\bar{E}[\Lambda_t N_t^{(j,i)} X_t|\mathcal{Z}_t] \\
= &\bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_t)(N_{t-1}^{(j,i)} + < X_{t-1}, e_i >< X_t, e_j >) X_t|\mathcal{Z}_t] \\
= &\bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_t)(N_{t-1}^{(j,i)} + < X_{t-1}, e_i >< X_t, e_j >) \sum_{l=1}^{N} < X_t, e_l > X_t|\mathcal{Z}_t] \\
= &\sum_{l=1}^{N} < \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_{t-1}, e_l) N_{t-1}^{(j,i)} (AX_{t-1} + M_t)|\mathcal{Z}_t], e_l > e_l + \\
&\bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_{t-1}, e_j) < X_{t-1}, e_i >< AX_{t-1} + M_t, e_j > |\mathcal{Z}_t] e_j.
\end{aligned}$$

Similarly as before, we define

$$\begin{aligned}
&\tilde{\sigma}(N_t^{(j,i)}X_t) \\
= &\sum_{l=1}^{N} < \bar{E}[\Lambda_{t-1}N_{t-1}^{(j,i)}AX_{t-1}|\mathcal{Z}_{t-1}], e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l)e_l + \\
&< \bar{E}[\Lambda_{t-1}X_{t-1}|\mathcal{Y}_t], e_i > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_j)a_{j,i}e_j \\
= &\sum_{l=1}^{N} < A\sigma(N_{t-1}^{(j,i)}X_{t-1}), e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l)e_l + \\
&< q_{t-1}, e_i > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_j)a_{j,i}e_j.
\end{aligned}$$

Taking $\tilde{\sigma}(N_{t-1}^{(j,i)}X_{t-1})$ as an approximation to $\sigma(N_{t-1}^{(j,i)}X_{t-1})$, and $\tilde{q}_{t-1}$ as an approximation to q_{t-1}, we have

$$\begin{aligned}
\tilde{\sigma}(N_t^{(j,i)}X_t) &= \sum_{l=1}^{N} < \tilde{\sigma}(N_{t-1}^{(j,i)}X_{t-1}), e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l)e_l + \\
&< \tilde{q}_{t-1}, e_i > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_j)a_{j,i}e_j.
\end{aligned}$$

□

Theorem 5.8. *The probability vector $\sigma(G_t^iX_t)$ is approximately computed by the recursion*

$$\begin{aligned}
\tilde{\sigma}(G_t^iX_t) &= \sum_{l=1}^{N} < A\tilde{\sigma}(G_{t-1}^iX_{t-1}), e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l)e_l + \\
&< A\tilde{q}_{t-1}, e_i > f(z_t)\lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i)e_i. \qquad (5.6.4)
\end{aligned}$$

Proof.

$$
\begin{aligned}
& \sigma(G_t^i X_t) \\
= \;& \bar{E}[\Lambda_t G_t^i X_t | \mathcal{Z}_t] \\
= \;& \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_t)(G_{t-1}^i + f(z_t) < X_t, e_i >)X_t | \mathcal{Z}_t] \\
= \;& \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_t)(G_{t-1}^i + f(z_t) < X_t, e_i >) \sum_{l=1}^{N} < X_t, e_l > X_t | \mathcal{Z}_t] \\
= \;& \sum_{l=1}^{N} \{ \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_{t-1}, e_l) G_{t-1}^i < X_t, e_l > e_l | \mathcal{Z}_t] + \\
& \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_{t-1}, e_l) f(z_t) < X_t, e_i >< X_t, e_l > e_l | \mathcal{Z}_t] \} \\
= \;& \sum_{l=1}^{N} < \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_{t-1}, e_l) G_{t-1}^i (AX_{t-1} + M_t) | \mathcal{Z}_t], e_l > e_l + \\
& \bar{E}[\Lambda_{t-1}\lambda_t(X_0, ..., X_{t-1}, e_i) f(z_t) < AX_{t-1} + M_t, e_i > | \mathcal{Z}_t] e_i.
\end{aligned}
$$

Similarly as before, we define

$$
\begin{aligned}
& \tilde{\sigma}(G_t^i X_t) \\
= \;& \sum_{l=1}^{N} < \bar{E}[\Lambda_{t-1} G_{t-1}^i A X_{t-1} | \mathcal{Z}_{t-1}], e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l) e_l + \\
& < \bar{E}[\Lambda_{t-1} A X_{t-1} | \mathcal{Z}_t], e_i > f(z_t) \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i) e_i \\
= \;& \sum_{l=1}^{N} < A\sigma(G_{t-1}^i X_{t-1}), e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l) e_l + \\
& < A q_{t-1}, e_i > f(z_t) \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i) e_i.
\end{aligned}
$$

Taking $\tilde{\sigma}(G_{t-1}^i X_{t-1})$ as an approximation to $\sigma(G_{t-1}^i X_{t-1})$, and $\tilde{q}_{t-1}$ as an approximation to q_{t-1}, we have

$$
\begin{aligned}
\tilde{\sigma}(G_t^i X_t) \;=\; & \sum_{l=1}^{N} < A\tilde{\sigma}(G_{t-1}^i X_{t-1}), e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l) e_l + \\
& < A\tilde{q}_{t-1}, e_i > f(z_t) \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i) e_i.
\end{aligned}
$$

□

Setting $f(y_t) = 1$ in equation (5.6.4), we obtain

$$\begin{aligned}\tilde{\sigma}(J_t^i X_t) &= \sum_{l=1}^{N} < A\tilde{\sigma}(J_{t-1}^i X_{t-1}), e_l > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_l)e_l + \\ &\quad < A\tilde{q}_{t-1}, e_i > \lambda_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_i)e_i. \qquad (5.6.5)\end{aligned}$$

The estimates of $a_{j,i}$ and g_i are then given by (5.5.7) and (5.5.9).

5.7 Viterbi Estimation of States and Parameters

Following Chapter 4, we shall give the Viterbi estimates in this section.

From (5.6.1),

$$\tilde{q}_t(j) = \sum_{i=1}^{N} a_{ji} \frac{\phi(z_t - \gamma_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_j))}{\phi(z_t)} \tilde{q}_{t-1}(i).$$

Instead of this summation, we recursively define new unnormalized probabilities $q_t^* = [q_t^*(1), q_t^*(2), ..., q_t^*(N)]'$ by

$$q_t^*(j) = \max_i a_{ji} \frac{\phi(z_t - \gamma_t(X_0, \tilde{X}_1, ..., \tilde{X}_{t-1}, e_j))}{\phi(z_t)} q_{t-1}^*(i). \qquad (5.7.1)$$

Certainly, $q_t^*(j) > 0$. The Viterbi probabilities $\{\rho_t\}$ are defined in (4.5.3). Also, $\rho_t(j)$ is an estimate of the conditional probability, that $X_t = e_j$, given $Z_1, Z_2, ..., Z_t$.

Write $X_t^* = (\rho_t(1), \rho_t(2), ..., \rho_t(N))$. In (5.7.1), instead of $\tilde{X}_i$, we use X_i^* to approximate X_i, $i = 1, 2, ..., t-1$, then

$$q_t^*(j) = \max_i a_{ji} \frac{\phi(z_t - \gamma_t(X_0, X_1^*, ..., X_{t-1}^*, e_j))}{\phi(z_t)} q_{t-1}^*(i). \qquad (5.7.2)$$

The quantity q_k^* is an approximation of q_k. Similarly, we can define $\sigma^*(N_k^{(j,i)} X_k)$, $\sigma^*(J_k^i X_k)$ and $\sigma^*(G_k^i X_k)$ by

$$\begin{aligned}\sigma^*(N_t^{(j,i)} X_t)(m) &= \max_l \{a_{ml} \sigma^*(N_{t-1}^{(j,i)} X_{t-1})(l) \lambda_t(X_0, X_1^*, ..., X_{t-1}^*, e_m)\} + \\ &\quad q_{t-1}^*(i) \lambda_t(X_0, X_1^*, ..., X_{t-1}^*, e_j) a_{j,i} \delta_{m,j}. \qquad (5.7.3)\end{aligned}$$

$$\begin{aligned}\sigma^*(G_t^iX_t)(m) &= \max_l \{a_{ml}\sigma^*(G_{t-1}^iX_{t-1})(l)\lambda_t(X_0,X_1^*,...,X_{t-1}^*,e_m)\} + \\ &\quad \max_l \{a_{il}q_{t-1}^*(l)f(z_t)\lambda_t(X_0,X_1^*,...,X_{t-1}^*,e_i)\delta_{m,i}\}. \end{aligned} \tag{5.7.4}$$

$$\begin{aligned}\sigma^*(J_t^iX_t)(m) &= \max_l \{a_{ml}\sigma^*(J_{t-1}^iX_{t-1})(l)\lambda_t(X_0,X_1^*,...,X_{t-1}^*,e_m)\} + \\ &\quad \max_l \{a_{il}q_{t-1}^*(l)\lambda_t(X_0,X_1^*,...,X_{t-1}^*,e_i)\delta_{m,i}\}. \end{aligned} \tag{5.7.5}$$

Following the results in section 5.5, estimates for the parameters could then be computed by

$$\hat{a}_{j,i} = \frac{\sigma^*(N_t^{(j,i)})}{\sigma^*(J_{t-1}^i)}, \tag{5.7.6}$$

$$\hat{\gamma}_t(X_0,X_1^*,...,X_{t-1}^*,e_i) = \frac{\sigma^*(G_t^i)}{\sigma^*(J_t^i)}, \tag{5.7.7}$$

where in (5.7.4), $f(z_l) = z_l$ for G_t^i.

The estimator for g_i up to time t is still given by

$$\hat{g}_i(t) = (u^\tau * \hat{\gamma}(X_0,X_1^*,...,X_{t-1}^*,e_i))(t).$$

5.8 Conclusions

In this chapter we have obtained exact estimates for the parameters and hidden states in the hidden Markov model, with the noise in the observations being fractional Gaussian noise. It is shown that, using change of measure method, the parameters can be estimated based on the history of the observations. We then gave recursive approximate and Viterbi estimates for these parameters and states. For the recursive approximate estimates we do not need to go back through all the time steps to the values at time 0.

Chapter 6

A FILTER FOR A STATE SPACE MODEL WITH FRACTIONAL GAUSSIAN NOISE

6.1 Introduction

State space models are widely used in finance, speech processing, image processing and control systems. A state space model is an extension of a hidden Markov model. In such a model the signal of interest is hidden but observed through another stochastic process. Both the unobservable and the observed processes are corrupted by noise. However, the hidden signal could be any process, not necessary a Markov process. Often, the noise in the signal and observations is assumed to be Gaussian. In that case, the Kalman filter, the extended Kalman filter and the Wonham filter are well known methods for estimating the hidden signal.

However, as mentioned in Chapter 5, in many practical cases the noise in the observations has some 'memory'. In this chapter, we consider a discrete time, state space model, where the signal is observed through a real valued process which is corrupted by fractional Gaussian noise. The relation between the signals at different times and the relation between the hidden signal and the observations are both linear. We derive the estimates for the hidden signal and the parameters, using the change of measure method and the EM algorithm.

An important question is to estimate the error between our approximate filter and an exact filter. However, without an exact filter, this appears to be a difficult problem.

This chapter is arranged as follows. In the next section, we give a brief description of fractional Gaussian noise and the model used in this chapter. In section 6.3, we derive the filter for estimating the hidden signal. In section 6.4, we derive the formulae for approximately estimating the parameters and the hidden signal. In the final section, we give some conclusions.

6.2 State Space Model with Fractional Gaussian Noise

Fractional Gaussian noise is defined in Chapter 5. The real-valued, (one dimensional), state and observations of the system considered in this chapter satisfy the dynamics

$$x_t = ax_{t-1} + bv_t, \tag{6.2.1}$$

$$y_t = cx_t + dw_t^r. \tag{6.2.2}$$

Here a, b, c and d are unknown parameters; $v = \{v_t, t = 0, 1, ...\}$ is the process noise, which is a sequence of $N(0,1)$ i.i.d. random variables; $w^r = \{w_t^r, t = 0, 1, ...\}$ is the measurement noise, which is a sequence of fractional Gaussian random variables as described in Chapter 5.

As in Chapter 5, write

$$\begin{aligned} z_t &= (u^{-r} * y)(t), \\ \gamma_t(x_0, x_1, ..., x_t) &= (u^{-r} * x)(t). \end{aligned}$$

Then, (6.2.2) implies the following equation.

$$z_t = c\gamma_t(x_0, ..., x_t) + dw_t. \tag{6.2.3}$$

These remain the dynamics of z under the 'real world' probability P.

6.3 Filtering

Consider a probability measure $\bar{P}$ on the measurable space $(\Omega, \mathcal{F})$ such that, under $\bar{P}$,

(1) $\{x_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables,

(2) $\{z_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables,

and $\{x_t\}$ and $\{z_t\}$ are independent of each other.

We call the measure $\bar{P}$ a "reference" probability.

We now construct the 'real world' probability P from $\bar{P}$, such that, under P, $\{v_t\}$ and $\{w_t\}$ are sequences of $N(0,1)$ i.i.d. random variables, where $v_t = b^{-1}(x_t - ax_{t-1})$, $w_t = d^{-1}(z_t - c\gamma_t(x_0, ..., x_t))$.

Write $\mathcal{G}_t^z = \sigma\{x_0, z_0, x_1, z_1, ..., x_t, z_t\}$, so $\{\mathcal{G}_t^z\}$ is the filtration generated by (x, z). (However, see the definition of $\mathcal{G}_t^y$ after Corollary 6.1.)

Write

$$\begin{aligned} \phi(x) &= \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2}), \\ \lambda_0 &= \frac{\phi(d^{-1}(z_0 - c\gamma_0(x_0)))}{d\phi(z_0)}, \end{aligned} \tag{6.3.1}$$

and for $l \geq 1$,

$$\lambda_l = \frac{\phi(b^{-1}(x_l - ax_{l-1}))\phi(d^{-1}(z_l - c\gamma_l(x_0, ..., x_l)))}{bd\phi(x_l)\phi(z_l)}, \tag{6.3.2}$$

$$\Lambda_t = \prod_{l=0}^{t} \lambda_l, \quad t = 1, 2, 3, ... \quad . \tag{6.3.3}$$

Definition 6.1. *Define P by putting*

$$\frac{dP}{d\bar{P}}|_{\mathcal{G}_t^z} = \Lambda_t. \tag{6.3.4}$$

Theorem 6.1. *Define $v_t = b^{-1}(x_t - ax_{t-1})$, $w_t = d^{-1}(z_t - c\gamma_t(x_0, x_1, ..., x_t))$ for $t \in \{0, 1, 2, ...\}$. Then, under P, $\{v_t\}$ and $\{w_t\}$ are sequences of $N(0,1)$ i.i.d. random variables.*

Proof.

$$\begin{aligned}
E(f(v_t)g(w_t)|\mathcal{G}^z_{t-1}) &= \frac{\bar{E}[\Lambda_t f(v_t)g(w_t)|\mathcal{G}^z_{t-1}]}{\bar{E}[\Lambda_t|\mathcal{G}^z_{t-1}]} \\
&= \frac{\bar{E}[\lambda_t f(v_t)g(w_t)|\mathcal{G}^z_{t-1}]}{\bar{E}[\lambda_t|\mathcal{G}^z_{t-1}]}.
\end{aligned}$$

$$\begin{aligned}
\bar{E}[\lambda_t|\mathcal{G}^z_{t-1}] &= \bar{E}[\frac{\phi(b^{-1}(x_t - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{bd\phi(x_t)\phi(z_t)}|\mathcal{G}^z_{t-1}] \\
&= \bar{E}[\frac{\phi(b^{-1}(x_t - ax_{t-1}))}{b\phi(x_t)}\bar{E}[\frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{d\phi(z_t)}|\mathcal{G}^z_{t-1}, x_t]|\mathcal{G}^z_{t-1}].
\end{aligned}$$

$$\begin{aligned}
\bar{E}[\frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{d\phi(z_t)}|\mathcal{G}^z_{t-1}, x_t] &= \int_{-\infty}^{\infty} \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{d\phi(z_t)}\phi(z_t)dz_t \\
&= 1.
\end{aligned}$$

$$\begin{aligned}
\bar{E}[\frac{\phi(b^{-1}(x_t - ax_{t-1}))}{b\phi(x_t)}|\mathcal{G}^z_{t-1}] &= \int_{-\infty}^{\infty} \frac{\phi(b^{-1}(x_t - ax_{t-1}))}{b\phi(x_t)}\phi(x_t)dx_t \\
&= 1.
\end{aligned}$$

$$\begin{aligned}
&\bar{E}[\lambda_t f(v_t)g(w_t)|\mathcal{G}^z_{t-1}] \\
=\ & \bar{E}[\frac{\phi(b^{-1}(x_t - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{bd\phi(x_t)\phi(z_t)} \cdot \\
& f(b^{-1}(x_t - ax_{t-1}))g(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))|\mathcal{G}^z_{t-1}] \\
=\ & \bar{E}[\frac{\phi(b^{-1}(x_t - ax_{t-1}))}{b\phi(x_t)} f(b^{-1}(x_t - ax_{t-1})) \cdot \\
& \bar{E}[\frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{d\phi(z_t)} g(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))|\mathcal{G}^z_{t-1}, x_t]|\mathcal{G}^z_{t-1}].
\end{aligned}$$

$$\begin{aligned}
&\bar{E}[\frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{d\phi(z_t)} g(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))|\mathcal{G}^z_{t-1}, x_t] \\
=\ & \int_{-\infty}^{\infty} \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{d\phi(z_t)} g(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))\phi(z_t)dz_t \\
=\ & \int_{-\infty}^{\infty} \phi(\xi)g(\xi)d\xi.
\end{aligned}$$

$$\bar{E}[\lambda_t f(v_t) g(w_t) | \mathcal{G}^z_{t-1}] = \int_{-\infty}^{\infty} \phi(\zeta) f(\zeta) d\zeta \int_{-\infty}^{\infty} \phi(\xi) g(\xi) d\xi.$$

The result follows. That is, under P, $v_t = b^{-1}(x_t - ax_{t-1})$, $w_t = d^{-1}(z_t - c\gamma_t(x_0, x_1, ..., x_t))$ are i.i.d. sequences of $N(0,1)$ random variables. Consequently, under P, $x_t = ax_{t-1} + bv_t$ and $z_t = c\gamma_t(x_0, x_1, ..., x_t) + dw_t$. □

Corollary 6.1. *Under P, $y_t := (u^r * z)(t) = c(u^r * \gamma)(t) + d(u^r * w)(t)$. That is,*

$$y_t = cx_t + dw^r_t.$$

Write

$$\begin{aligned} \mathcal{Y}_t &= \sigma\{y_0, y_1, ..., y_t\}, \\ \mathcal{Z}_t &= \sigma\{z_0, z_1, ..., z_t\}, \\ \mathcal{G}^y_t &= \sigma\{x_0, y_0, x_1, y_1, ..., x_t, y_t\}. \end{aligned}$$

Then the filtrations of the x, y, z and (x, y) processes are $\{\mathcal{F}_t\}$, $\{\mathcal{Y}_t\}$, $\{\mathcal{Z}_t\}$ and $\{\mathcal{G}^y_t\}$. Note that, $\{\mathcal{Y}_t\} = \{\mathcal{Z}_t\}$, so $\{\mathcal{G}^y_t\} = \{\mathcal{G}^z_t\}$.

First we shall describe how to estimate the hidden states, given the observations $\{y_t, t = 0, 1, 2, ...\}$.

Using a version of Bayes' rule [36], we have

$$E[g(x_t)|\mathcal{Z}_t] = \frac{\bar{E}[\Lambda_t g(x_t)|\mathcal{Z}_t]}{\bar{E}[\Lambda_t|\mathcal{Z}_t]}. \tag{6.3.5}$$

Assume

$$\bar{E}[\Lambda_t g(x_t)|\mathcal{Z}_t] = \int g(x)\alpha_t(x)dx. \tag{6.3.6}$$

If $p_t(\bullet)$ denotes the normalized conditional density, such that $E[g(x_t)|\mathcal{Z}_t] = \int g(x)p_t(x)dx$, then $p_t(x) = \alpha_t(x)[\int \alpha_t(u)du]^{-1}$.

Theorem 6.2. *The unnormalized density $\alpha_t(\bullet)$ is computed by*

$$\alpha_0(x) = \lambda_0 = \frac{\phi(d^{-1}(z_0 - c\gamma_0(x_0)))}{d\phi(z_0)} p_0(x),$$

where $p_0(x)$ is the initial density of x_0, and for $t \geq 1$,

$$\begin{aligned}
&\alpha_t(x) \\
= \; & \frac{1}{b^t d^t \phi(z_t)...\phi(z_1)} \int ... \int \phi(b^{-1}(x - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_{t-1}, x))) \cdot \\
& \phi(b^{-1}(x_{t-1} - ax_{t-2}))\phi(d^{-1}(z_{t-1} - c\gamma_t(x_0, ..., x_{t-1})))...\phi(b^{-1}(x_1 - ax_0)) \cdot \\
& \phi(d^{-1}(z_1 - c\gamma_t(x_0, x_1)))p_0(x_0)dx_{t-1}...dx_0.
\end{aligned} \tag{6.3.7}$$

Proof. It is easy to see $\alpha_0(x) = \lambda_0 p_0(x)$. When $t \geq 1$,

$$\begin{aligned}
& \int g(x)\alpha_t(x)dx \\
= \; & \bar{E}[\Lambda_t g(x_t)|\mathcal{Z}_t] \\
= \; & \bar{E}[\Lambda_{t-1}\lambda_t g(x_t)|\mathcal{Z}_t] \\
= \; & \bar{E}[\Lambda_{t-1}\frac{\phi(b^{-1}(x_t - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_t)))}{bd\phi(x_t)\phi(z_t)} g(x_t)|\mathcal{Z}_t] \\
= \; & \frac{1}{bd\phi(z_t)}\bar{E}[\Lambda_{t-1}\int \frac{\phi(b^{-1}(x - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_{t-1}, x)))}{\phi(x)} g(x)\phi(x)dx|\mathcal{Z}_t] \\
= \; & \frac{1}{bd\phi(z_t)}\bar{E}[\Lambda_{t-2}\frac{\phi(b^{-1}(x_{t-1} - ax_{t-2}))\phi(d^{-1}(z_{t-1} - c\gamma_{t-1}(x_0, ..., x_{t-1})))}{bd\phi(x_{t-1})\phi(z_{t-1})} \cdot \\
& \int \phi(b^{-1}(x - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_{t-1}, x)))g(x)dx|\mathcal{Z}_t] \\
= \; & ...... \\
= \; & \frac{1}{b^t d^t \phi(z_t)...\phi(z_1)} \int ... \int \phi(b^{-1}(x - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_{t-1}, x))) \cdot \\
& \phi(b^{-1}(x_{t-1} - ax_{t-2}))\phi(d^{-1}(z_{t-1} - c\gamma_t(x_0, ..., x_{t-1})))...\phi(b^{-1}(x_1 - ax_0)) \cdot \\
& \phi(d^{-1}(z_1 - c\gamma_t(x_0, x_1)))p_0(x_0)g(x)dxdx_{t-1}...dx_0.
\end{aligned}$$

So,

$$
\begin{aligned}
&\alpha_t(x) \\
= \ & \frac{1}{b^t d^t \phi(z_t)...\phi(z_1)} \int ... \int \phi(b^{-1}(x - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, ..., x_{t-1}, x))) \cdot \\
& \phi(b^{-1}(x_{t-1} - ax_{t-2}))\phi(d^{-1}(z_{t-1} - c\gamma_t(x_0, ..., x_{t-1})))...\phi(b^{-1}(x_1 - ax_0)) \cdot \\
& \phi(d^{-1}(z_1 - c\gamma_t(x_0, x_1)))p_0(x_0)dx_{t-1}...dx_0.
\end{aligned}
$$

□

From Theorem 6.2 and (6.3.5), we have the estimates of x_t

$$\hat{x}_t = E[x_t|\mathcal{Z}_t] = \frac{\bar{E}[\Lambda_t x_t|\mathcal{Z}_t]}{\bar{E}[\Lambda_t|\mathcal{Z}_t]} = \frac{\int x\alpha_t(x)dx}{\int \alpha_t(x)dx}.$$

6.4 Approximate Estimation of States and Parameters

6.4.1 Approximate Estimation of the States

In this section, we give recursive approximate estimates of the parameters and hidden states.

Theorem 6.3. *The unnormalized density $\alpha_t(\bullet)$ is approximately computed by the recursion*

$$\tilde{\alpha}_t(x) = \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)} \int \phi(b^{-1}(x - au))\tilde{\alpha}_{t-1}(u)du. \quad (6.4.1)$$

The approximate normalized densities are defined by

$$\tilde{p}_t(x) = \tilde{\alpha}_t(x) / \int \tilde{\alpha}_t(u)du.$$

The approximate means are then given as

$$\tilde{x}_t = \int_R x\tilde{p}_t(x)dx.$$

These means are the quantities used in (6.4.1).

Proof. The values $z_0, z_1, z_2, \dots$ are observed sequentially. We have assumed that the distribution x_0 is described by an a priori density $p_0(x)$.

The recursion is initialized as follows.

$$\alpha_0(x) = \tilde{\alpha}_0(x) = \frac{\phi(d^{-1}(z_0 - cx))}{d\phi(z_0)} p_0(x),$$

and $x_0 = \hat{x}_0 = \tilde{x}_0 = E_{p_0}[x]$.

For $t \geq 1$,

$$\begin{aligned}
&\int g(x)\alpha_t(x)dx \\
= \;& \bar{E}[\Lambda_t g(x_t)|\mathcal{Z}_t] \\
= \;& \bar{E}[\Lambda_{t-1}\lambda_t g(x_t)|\mathcal{Z}_t] \\
= \;& \bar{E}[\Lambda_{t-1}\frac{\phi(b^{-1}(x_t - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, \dots, x_t)))}{bd\phi(x_t)\phi(z_t)} g(x_t)|\mathcal{Z}_t]. \qquad (6.4.2)
\end{aligned}$$

Using $\tilde{x}_i$ to approximate x_i, $i = 1, 2, \dots, t-1$ in (6.4.2), we define $\tilde{\alpha}_t(\bullet)$ such that

$$\begin{aligned}
&\int g(x)\tilde{\alpha}_t(x)dx \\
= \;& \bar{E}[\Lambda_{t-1}\frac{\phi(b^{-1}(x_t - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, \dots, \tilde{x}_{t-1}, x_t)))}{bd\phi(x_t)\phi(z_t)} g(x_t)|\mathcal{Z}_t] \\
= \;& \frac{1}{bd\phi(z_t)}\bar{E}[\Lambda_{t-1}\int \frac{\phi(b^{-1}(x - ax_{t-1}))\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, \dots, \tilde{x}_{t-1}, x)))}{\phi(x)} g(x)\phi(x)dx|\mathcal{Z}_t] \\
= \;& \frac{1}{bd\phi(z_t)}\int\int \phi(b^{-1}(x - au))\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, \dots, \tilde{x}_{t-1}, x)))g(x)\alpha_{t-1}(u)dxdu,
\end{aligned}$$

where, for $0 < l \leq t$, $\tilde{x}_l = \frac{\int x\tilde{\alpha}_l(x)dx}{\int \tilde{\alpha}_l(x)dx}$.

Taking $\tilde{\alpha}_{t-1}(\bullet)$ as an approximation to $\alpha_{t-1}(\bullet)$, we have

$$\tilde{\alpha}_t(x) = \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, \dots, \tilde{x}_{t-1}, x)))}{bd\phi(z_t)} \int \phi(b^{-1}(x - au))\tilde{\alpha}_{t-1}(u)du.$$

□

Note that since the dynamics of x_t and y_t are both linear, the density $\tilde{p}_t(\bullet)$ is also normally distributed with mean $\mu_t = E[x_t|\mathcal{Y}_t]$ and variance $R_t = E[(x_t - \mu_t)^2|\mathcal{y}_t]$. Then we have the following theorem.

Theorem 6.4. *The mean μ_t and variance R_t are approximately computed by the recursion*

$$\tilde{R}_t = A_t, \tag{6.4.3}$$

$$\tilde{\mu}_t = A_t B_t, \tag{6.4.4}$$

where $A_t = (\frac{1}{b^2} - \frac{a^2 \bar{R}_{t-1}}{b^4})^{-1}$, $B_t = \frac{a\tilde{\mu}_{t-1}\bar{R}_{t-1}}{b^2 \tilde{R}_{t-1}}$, *and* $\bar{R}_t = (\frac{a^2}{b^2} + \frac{1}{\tilde{R}_t})^{-1}$.

Proof. By Theorem 6.3,

$$\begin{aligned}
&\tilde{\alpha}_t(x) \\
= \ & \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)} \int \phi(b^{-1}(x - au))\tilde{\alpha}_{t-1}(u)du \\
\propto \ & \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)} \int exp(-\frac{1}{2b^2}(x - au)^2 - \frac{1}{2R_{t-1}}(u - \mu_{t-1})^2)du \\
\propto \ & \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)} exp(-\frac{x^2}{2b^2} - \frac{\mu_{t-1}^2}{2R_{t-1}}) \cdot \\
& \int exp(-\frac{1}{2}(u^2(\frac{a^2}{b^2} + \frac{1}{R_{t-1}}) - 2u(\frac{ax}{b^2} + \frac{\mu_{t-1}}{R_{t-1}})))du.
\end{aligned}$$

Taking $\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, a\tilde{x}_{t-1})$ as an approximation to $\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)$, and taking $\tilde{\mu}_{t-1}$, $\tilde{R}_{t-1}$ to approximate μ_{t-1}, R_{t-1}, we have

$$\begin{aligned}
\tilde{\alpha}_t(x) \ \propto \ & \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, a\tilde{x}_{t-1})))}{bd\phi(z_t)} exp(-\frac{x^2}{2b^2} - \frac{\tilde{\mu}_{t-1}^2}{2\tilde{R}_{t-1}}) \cdot \\
& \int exp(-\frac{1}{2}(u^2(\frac{a^2}{b^2} + \frac{1}{\tilde{R}_{t-1}}) - 2u(\frac{ax}{b^2} + \frac{\tilde{\mu}_{t-1}}{\tilde{R}_{t-1}})))du.
\end{aligned}$$

Write

$$\begin{aligned}
K(x) &= \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, a\tilde{x}_{t-1})))}{bd\phi(z_t)} exp(-\frac{x^2}{2b^2} - \frac{\tilde{\mu}_{t-1}^2}{2\tilde{R}_{t-1}}), \\
\bar{R}_t^{-1} &= \frac{a^2}{b^2} + \frac{1}{\tilde{R}_t}, \\
\beta_t(x) &= \frac{ax}{b^2} + \frac{\tilde{\mu}_t}{\tilde{R}_t}.
\end{aligned}$$

Then,

$$
\begin{aligned}
& \tilde{\alpha}_t(x) \\
\propto \ & K(x) \int exp(-\frac{1}{2}(u^2 \bar{R}_{t-1}^{-1} - 2u\beta_{t-1}(x)))du \\
\propto \ & K(x) \int exp(-\frac{1}{2\bar{R}_{t-1}}(u^2 - 2u\beta_{t-1}(x)\bar{R}_{t-1} + (\beta_{t-1}(x)\bar{R}_{t-1})^2 - (\beta_{t-1}(x)\bar{R}_{t-1})^2))du \\
\propto \ & K(x)exp(\frac{\beta_{t-1}^2(x)\bar{R}_{t-1}}{2}) \int exp(-\frac{1}{2\bar{R}_{t-1}}(u - \beta_{t-1}(x)\bar{R}_{t-1})^2)du \\
\propto \ & K(x)exp(\frac{\beta_{t-1}^2(x)\bar{R}_{t-1}}{2})\sqrt{2\pi\bar{R}_{t-1}} \\
\propto \ & K_1 exp(-\frac{1}{2}(x^2(\frac{1}{b^2} - \frac{a^2\bar{R}_{t-1}}{b^4}) - 2x\frac{a\tilde{\mu}_{t-1}\bar{R}_{t-1}}{b^2\tilde{R}_{t-1}})) \\
\propto \ & K_2 exp(-\frac{1}{2A_t}(x - A_tB_t)^2),
\end{aligned}
$$

where K_1 and K_2 are constant in x.

So, the result follows. □

6.4.2 Approximate Estimation of the Parameters

Now we shall show how to estimate the parameters a, b, c and d in the model when they are unknown. The EM algorithm is used.

Similar as Chapter 2, we take

$$
Q_t(\theta, \theta^*) = E_{\theta^*}[\log \frac{dP_\theta}{dP_{\theta^*}}|\mathcal{G}_t^z]. \tag{6.4.5}
$$

In our model, we define

$$
\frac{dP_\theta}{dP_{\theta^*}}|\mathcal{G}_t^z = \prod_{l=0}^{t} \eta_l, \tag{6.4.6}
$$

where

$$
\begin{aligned}
\eta_0 &= \frac{d^*\phi(d^{-1}(z_0 - c\gamma_0))}{d\phi(d^{*-1}(z_0 - c^*\gamma_0))}, && (6.4.7) \\
\eta_l &= \frac{d^*\phi(d^{-1}(z_l - c\gamma_l(x_0, ..., x_l)))b^*\phi(b^{-1}(x_l - ax_{l-1}))}{d\phi(d^{*-1}(z_l - c^*\gamma_l(x_0, ..., x_l)))b\phi(b^{*-1}(x_l - a^*x_{l-1}))}. && (6.4.8)
\end{aligned}
$$

Then

$$
\begin{aligned}
Q(\theta, \theta^*) &= E_{\theta^*}[\log \frac{dP_\theta}{dP_{\theta^*}} | \mathcal{G}_t^z] \\
&= -t \log b - (t+1) \log d - \frac{1}{2} E_{\theta^*}[\sum_{l=1}^{t} b^{-2}(x_l - a x_{l-1})^2 | \mathcal{Z}_t] - \\
&\quad \frac{1}{2} E_{\theta^*}[\sum_{l=1}^{t} d^{-2}(z_l - c\gamma_l(x_0, ..., x_l))^2 | \mathcal{Z}_t] + R(\theta^*), \qquad (6.4.9)
\end{aligned}
$$

where $R(\theta^*)$ does not contain θ.

Set $\frac{\partial Q(\theta,\theta^*)}{\partial \theta} = 0$, we get

$$
\begin{aligned}
a &= E_{\theta^*}[\sum_{l=1}^{t} x_l x_{l-1} | \mathcal{Z}_t](E_{\theta^*}[\sum_{l=1}^{t} x_{l-1}^2 | \mathcal{Z}_t])^{-1}, \qquad (6.4.10) \\
b^2 &= \frac{1}{t} E_{\theta^*}[\sum_{l=1}^{t} (x_l - a x_{l-1})^2 | \mathcal{Z}_t], \qquad (6.4.11) \\
c &= E_{\theta^*}[\sum_{l=1}^{t} z_l \gamma_l(x_0, ..., x_l) | \mathcal{Z}_t](E_{\theta^*}[\sum_{l=1}^{t} \gamma_l^2(x_0, ..., x_l) | \mathcal{Z}_t])^{-1}, \qquad (6.4.12) \\
d^2 &= \frac{1}{t+1} E_{\theta^*}[\sum_{l=1}^{t} (z_l - c\gamma_l(x_0, ..., x_l))^2 | \mathcal{Z}_t]. \qquad (6.4.13)
\end{aligned}
$$

Write $T_t^{x(0)} = \sum_{l=0}^{t} x_l^2$, $T_t^{x(1)} = \sum_{l=1}^{t} x_l x_{l-1}$, $T_t^{x(2)} = \sum_{l=0}^{t} x_{l-1}^2$, $T_t^{\gamma(0)} = \sum_{l=0}^{t} \gamma_l^2(x_0, ..., x_l)$, $U_t = \sum_{l=1}^{t} z_l \gamma_l(x_0, ..., x_l)$, then

$$
\begin{aligned}
a &= E_{\theta^*}[T_t^{x(1)} | \mathcal{Z}_t](E_{\theta^*}[T_t^{x(2)} | \mathcal{Z}_t])^{-1}, \qquad (6.4.14) \\
b^2 &= \frac{1}{t} E_{\theta^*}[T_t^{x(0)} - 2a T_t^{x(1)} + a^2 T_t^{x(2)} | \mathcal{Z}_t], \qquad (6.4.15) \\
c &= E_{\theta^*}[U_t | \mathcal{Z}_t](E_{\theta^*}[T_t^{\gamma(0)} | \mathcal{Z}_t])^{-1}, \qquad (6.4.16) \\
d^2 &= \frac{1}{t+1} E_{\theta^*}[2cU_t + T_t^{\gamma(0)} + \sum_{l=1}^{t} z_l^2 | \mathcal{Z}_t]. \qquad (6.4.17)
\end{aligned}
$$

Next, we develop the finite - dimensional filter for $T_t^{x(M)}$, $M = 0, 1, 2$, $T_t^{\gamma(0)}$ and U_t.

Definition 6.2. *Define the processes*

$$\alpha_t(x) = \bar{E}[\Lambda_t I(x_t \in dx)|\mathcal{Z}_t], \tag{6.4.18}$$

$$\beta_t^{x(M)}(x) = \bar{E}[\Lambda_t T_t^{x(M)} I(x_t \in dx)|\mathcal{Z}_t], \quad M = 0,1,2, \tag{6.4.19}$$

$$\beta_t^{\gamma(0)}(x) = \bar{E}[\Lambda_t T_t^{\gamma(M)} I(x_t \in dx)|\mathcal{Z}_t], \tag{6.4.20}$$

$$\delta_t(x) = \bar{E}[\Lambda_t U_t I(x_t \in dx)|\mathcal{Z}_t]. \tag{6.4.21}$$

Then, for any function $g(\bullet)$,

$$\bar{E}[\Lambda_t g(x_t)|\mathcal{Z}_t] = \int \alpha_t(x) g(x) dx, \tag{6.4.22}$$

$$\bar{E}[\Lambda_t T_t^{x(M)} g(x_t)|\mathcal{Z}_t] = \int \beta_t^{x(M)}(x) g(x) dx, \quad M = 0,1,2, \tag{6.4.23}$$

$$\bar{E}[\Lambda_t T_t^{\gamma(M)} g(x_t)|\mathcal{Z}_t] = \int \beta_t^{\gamma(0)}(x) g(x) dx, \tag{6.4.24}$$

$$\bar{E}[\Lambda_t U_t g(x_t)|\mathcal{Z}_t] = \int \delta_t(x) g(x) dx. \tag{6.4.25}$$

Theorem 6.5. *The unnormalized density $\beta_t^{x(M)}(\bullet)$, $M = 0,1,2$, $\beta_t^{\gamma(0)}(\bullet)$ and $\delta_t(\bullet)$ are approximately computed by the recursions*

$$\begin{aligned} \tilde{\beta}_t^{x(0)}(x) = {} & \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)}[\int \phi(b^{-1}(x - au))\tilde{\beta}_{t-1}^{x(0)}(u)du + \\ & x^2 \int \phi(b^{-1}(x - au))\tilde{\alpha}_{t-1}(u)du], \end{aligned} \tag{6.4.26}$$

$$\begin{aligned} \tilde{\beta}_t^{x(1)}(x) = {} & \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)}[\int \phi(b^{-1}(x - au))\tilde{\beta}_{t-1}^{x(1)}(u)du + \\ & x \int u\phi(b^{-1}(x - au))\tilde{\alpha}_{t-1}(u)du], \end{aligned} \tag{6.4.27}$$

$$\begin{aligned} \tilde{\beta}_t^{x(2)}(x) = {} & \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)}[\int \phi(b^{-1}(x - au))\tilde{\beta}_{t-1}^{x(2)}(u)du + \\ & \int u^2\phi(b^{-1}(x - au))\tilde{\alpha}_{t-1}(u)du], \end{aligned} \tag{6.4.28}$$

$$\begin{aligned}
\tilde{\beta}_t^{\gamma(0)}(x) &= \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)}[\int \phi(b^{-1}(x-au))\tilde{\beta}_{t-1}^{\gamma(0)}(u)du + \\
&\quad \gamma_t^2(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x) \int \phi(b^{-1}(x-au))\tilde{\alpha}_{t-1}(u)du], \qquad (6.4.29) \\
\tilde{\delta}_t(x) &= \frac{\phi(d^{-1}(z_t - c\gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)))}{bd\phi(z_t)}[\int \phi(b^{-1}(x-au))\tilde{\delta}_{t-1}(u)du + \\
&\quad \gamma_t(x_0, \tilde{x}_1, ..., \tilde{x}_{t-1}, x)z_t \int \phi(b^{-1}(x-au))\tilde{\alpha}_{t-1}(u)du]. \qquad (6.4.30)
\end{aligned}$$

The proof of Theorem 6.5 is similar to the proof of Theorem 6.3.

Write $\sigma_{t+1} = \bar{R}_t^{-1}$, $\Sigma_t = \frac{a}{b^2}\sigma_t^{-1}$ and $S_t = \sigma_{t+1}^{-1}\tilde{R}_t^{-1}\tilde{\mu}_t$. Then we have the following theorems.

Theorem 6.6. *The densities* $\tilde{\beta}_t^{x(M)}(\bullet)$, $M = 0, 1, 2$ *and* $\tilde{\beta}_t^{\gamma(0)}(\bullet)$ *are defined by*

$$\begin{aligned}
\tilde{\beta}_t^{x(M)}(x) &= [a_t^{x(M)} + b_t^{x(M)}x + d_t^{x(M)}x^2]\tilde{\alpha}_t(x), \quad M = 0, 1, 2, \qquad (6.4.31) \\
\tilde{\beta}_t^{\gamma(0)}(x) &= [a_t^{\gamma(0)} + b_t^{\gamma(0)}x + d_t^{\gamma(0)}x^2]\tilde{\alpha}_t(x), \qquad (6.4.32)
\end{aligned}$$

where $a_t^{x(M)}$, $b_t^{x(M)}$, $d_t^{x(M)}$, $a_t^{\gamma(0)}$, $b_t^{\gamma(0)}$ and $d_t^{\gamma(0)}$ are computed by the following recursions:

$$
\begin{aligned}
a_{t+1}^{x(0)} &= a_t^{x(0)} + b_t^{x(0)} S_t + d_t^{x(0)} \sigma_{t+1}^{-1} + d_t^{x(0)} S_t^2, \\
a_0^{x(0)} &= 0, && (6.4.33) \\
b_{t+1}^{x(0)} &= \Sigma_{t+1}(b_t^{x(0)} + 2d_t^{x(0)} S_t), \quad b_0^{x(0)} = 0, && (6.4.34) \\
d_{t+1}^{x(0)} &= \Sigma_{t+1}^2 d_t^{x(0)} + 1, \quad d_0^{x(0)} = 1, && (6.4.35) \\
a_{t+1}^{x(1)} &= a_t^{x(1)} + b_t^{x(1)} S_t + d_t^{x(0)} \sigma_{t+1}^{-1} + d_t^{x(0)} S_t^2, \\
a_0^{x(1)} &= 0, && (6.4.36) \\
b_{t+1}^{x(1)} &= \Sigma_{t+1}(b_t^{x(1)} + 2d_t^{x(1)} S_t) + S_t, \quad b_0^{x(1)} = 0, && (6.4.37) \\
d_{t+1}^{x(1)} &= \Sigma_{t+1}^2 d_t^{x(1)} + \Sigma_{t+1}, \quad d_0^{x(1)} = 0, && (6.4.38) \\
a_{t+1}^{x(2)} &= a_t^{x(2)} + b_t^{x(2)} S_t + d_t^{x(2)} \sigma_{t+1}^{-1} + (d_t^{x(2)} + 1) S_t^2, \\
a_0^{x(2)} &= 0, && (6.4.39) \\
b_{t+1}^{x(2)} &= \Sigma_{t+1}(b_t^{x(2)} + 2(d_t^{x(2)} + 1) S_t), \quad b_0^{x(2)} = 0, && (6.4.40) \\
d_{t+1}^{x(2)} &= \Sigma_{t+1}^2 (d_t^{x(2)} + 1), \quad d_0^{x(2)} = 0, && (6.4.41) \\
a_{t+1}^{\gamma(0)} &= a_t^{\gamma(0)} + b_t^{\gamma(0)} S_t + d_t^{\gamma(0)} \sigma_{t+1}^{-1} + d_t^{\gamma(0)} S_t^2 + \gamma_{t+1}^2(x_0, \tilde{x}_1, \ldots, \tilde{x}_t, a\tilde{x}_t), \\
a_0^{\gamma(0)} &= x_0^2, && (6.4.42) \\
b_{t+1}^{\gamma(0)} &= \Sigma_{t+1}(b_t^{\gamma(0)} + 2d_t^{\gamma(0)} S_t), \quad b_0^{\gamma(0)} = 0, && (6.4.43) \\
d_{t+1}^{\gamma(0)} &= \Sigma_{t+1}^2 d_t^{\gamma(0)}, \quad d_0^{\gamma(0)} = 0. && (6.4.44)
\end{aligned}
$$

Proof. When $t = 0$, (6.4.19) and (6.4.26) indicate $\tilde{\beta}_t^{x(0)}(x)$ has the form (6.4.31) with $a_0^{x(0)} = 0$, $b_0^{x(0)} = 0$ and $d_0^{x(0)} = 1$. Assume (6.4.31) holds at time t, then at time $t+1$, using (6.4.26), (6.4.31) and Theorem 6.4, we have

$$\begin{aligned}\tilde{\beta}^{x(0)}_{t+1}(x) &= \frac{\phi(d^{-1}(z_{t+1}-c\gamma_t(x_0,\tilde{x}_1,...,\tilde{x}_t,x)))}{bd\phi(z_{t+1})}\cdot \\ &\quad \int [a^{x(0)}_t + b^{x(0)}_t u + d^{x(0)}_t u^2]\tilde{\alpha}_t(u)\phi(b^{-1}(x-au))du + x^2\tilde{\alpha}_{t+1}(x) \\ &= \Phi(x,z_{t+1})\int [a^{x(0)}_t + b^{x(0)}_t u + d^{x(0)}_t u^2]\tilde{\alpha}_t(u)\phi(b^{-1}(x-au))du + x^2\tilde{\alpha}_{t+1}(x),\end{aligned} \tag{6.4.45}$$

where $\Phi(x,z_{t+1}) = \frac{\phi(d^{-1}(z_{t+1}-c\gamma_t(x_0,\tilde{x}_1,...,\tilde{x}_t,x)))}{bd\phi(z_{t+1})}$.

Write $\xi_t = \frac{ax}{b^2} + \frac{\tilde{\mu}_t}{\tilde{R}_t}$. Write the first term of (6.4.45) as I_1.

$$\begin{aligned} & I_1 \\ =\ & \frac{\Phi(x,z_{t+1})}{2\pi b\sqrt{\tilde{R}_t}}\int exp\{-\frac{1}{2}[b^{-2}(x-au)^2 + \tilde{R}_t^{-1}(u-\tilde{\mu}_t)^2]\}[a^{x(0)}_t + b^{x(0)}_t u + d^{x(0)}_t u^2]du \\ =\ & \frac{\Phi(x,z_{t+1})}{2\pi b\sqrt{\tilde{R}_t}}exp[-\frac{1}{2}(\frac{x^2}{b^2}+\frac{\tilde{\mu}_t^2}{\tilde{R}_t})]\cdot \\ & \int exp\{-\frac{1}{2}[(\frac{a^2}{b^2}+\frac{1}{\tilde{R}_t})u^2 - 2(\frac{ax}{b^2}+\frac{\tilde{\mu}_t}{\tilde{R}_t})u]\}[a^{x(0)}_t + b^{x(0)}_t u + d^{x(0)}_t u^2]du \\ =\ & \frac{\Phi(x,z_{t+1})}{2\pi b\sqrt{\tilde{R}_t}}exp[-\frac{1}{2}(\frac{x^2}{b^2}+\frac{\tilde{\mu}_t^2}{\tilde{R}_t})]\int exp[-\frac{1}{2}(\sigma_{t+1}u^2 - 2\xi_t u)][a^{x(0)}_t + b^{x(0)}_t u + d^{x(0)}_t u^2]du \\ =\ & \frac{\Phi(x,z_{t+1})}{\sqrt{2\pi}b\sqrt{\tilde{R}_t\sigma_{t+1}}}exp[-\frac{1}{2}(\frac{x^2}{b^2}+\frac{\tilde{\mu}_t^2}{\tilde{R}_t} - \sigma_{t+1}^{-1}\xi_t^2)]\cdot \\ & \int exp[-\frac{1}{2}\sigma_{t+1}(u-\sigma_{t+1}^{-1}\xi_t)^2][a^{x(0)}_t + b^{x(0)}_t u + d^{x(0)}_t u^2]du \\ =\ & \frac{\Phi(x,z_{t+1})}{\sqrt{2\pi}b\sqrt{\tilde{R}_t\sigma_{t+1}}}exp[-\frac{1}{2}(\frac{x^2}{b^2}+\frac{\tilde{\mu}_t^2}{\tilde{R}_t} - \sigma_{t+1}^{-1}\xi_t^2)][a^{x(0)}_t + b^{x(0)}_t\sigma_{t+1}^{-1}\xi_t + d^{x(0)}_t(\sigma_{t+1}^{-1}+\sigma_{t+1}^{-2}\xi_t^2)] \\ =\ & \frac{\Phi(x,z_{t+1})}{\sqrt{2\pi\tilde{R}_{t+1}}}exp[-\frac{1}{2}\tilde{R}_{t+1}^{-1}(x-\tilde{\mu}_{t+1})^2][a^{x(0)}_t + b^{x(0)}_t\sigma_{t+1}^{-1}(\frac{ax}{b^2}+\frac{\tilde{\mu}_t}{\tilde{R}_t}) + \\ & d^{x(0)}_t(\sigma_{t+1}^{-1}+\sigma_{t+1}^{-2}(\frac{ax}{b^2}+\frac{\tilde{\mu}_t}{\tilde{R}_t})^2)] \\ =\ & \tilde{\alpha}_{t+1}(x)[a^{x(0)}_t + b^{x(0)}_t\sigma_{t+1}^{-1}\frac{a}{b^2}x + b^{x(0)}_t\sigma_{t+1}^{-1}\frac{\tilde{\mu}_t}{\tilde{R}_t} + d^{x(0)}_t\sigma_{t+1}^{-1} + d^{x(0)}_t\sigma_{t+1}^{-2}(\frac{a^2}{b^4}x^2 + 2\frac{a\tilde{\mu}_t}{b^2\tilde{R}_t}x + \frac{\tilde{\mu}_t^2}{\tilde{R}_t^2})].\end{aligned}$$

So,

$$
\begin{aligned}
&\tilde{\beta}_{t+1}^{x(0)}(x) \\
= \; & [a_t^{x(0)} + b_t^{x(0)}\sigma_{t+1}^{-1}\frac{a}{b^2}x + b_t^{x(0)}\sigma_{t+1}^{-1}\frac{\tilde{\mu}_t}{\tilde{R}_t} + d_t^{x(0)}\sigma_{t+1}^{-1} + d_t^{x(0)}\sigma_{t+1}^{-2}(\frac{a^2}{b^4}x^2 + 2\frac{a\tilde{\mu}_t}{b^2\tilde{R}_t}x + \frac{\tilde{\mu}_t^2}{\tilde{R}_t^2}) + \\
& x^2]\tilde{\alpha}_{t+1}(x) \\
= \; & [a_t^{x(0)} + b_t^{x(0)}\Sigma_{t+1}x + b_t^{x(0)}S_t + d_t^{x(0)}\sigma_{t+1}^{-1} + d_t^{x(0)}(\Sigma_{t+1}^2 x^2 + 2\Sigma_{t+1}S_t x + S_t^2) + x^2]\tilde{\alpha}_{t+1}(x) \\
= \; & [a_{t+1}^{x(0)} + b_{t+1}^{x(0)}x + d_{t+1}^{x(0)}x^2]\tilde{\alpha}_{t+1}(x),
\end{aligned}
$$

where $a_{t+1}^{x(0)} = a_t^{x(0)} + b_t^{x(0)}S_t + d_t^{x(0)}\sigma_{t+1}^{-1} + d_t^{x(0)}S_t^2$, $b_{t+1}^{x(0)} = \Sigma_{t+1}(b_t^{x(0)} + 2d_t^{x(0)}S_t)$, $d_{t+1}^{x(0)} = \Sigma_{t+1}^2 d_t^{x(0)} + 1$.

The result for $\tilde{\beta}_t^{x(0)}(\bullet)$ follows. The proofs for $\tilde{\beta}_t^{x(M)}(\bullet)$, $M = 1, 2$ and $\tilde{\beta}_t^{\gamma(0)}(\bullet)$ are similar. □

Theorem 6.7. *The density $\tilde{\delta}_t(\bullet)$ is defined by*

$$\tilde{\delta}_t(x) = [\bar{a}_t + \bar{b}_t x]\tilde{\alpha}_t(x), \tag{6.4.46}$$

where $\bar{a}_t$ and $\bar{b}_t$ are computed by the following recursions:

$$
\begin{aligned}
\bar{a}_{t+1} &= \bar{a}_t + \bar{b}_t S_t + \gamma_{t+1}(x_0, \tilde{x}_1, ..., \tilde{x}_t, a\tilde{x}_t)z_{t+1}, \\
\bar{a}_0 &= x_0 z_0, && (6.4.47) \\
\bar{b}_{t+1} &= \Sigma_{t+1}\bar{b}_t, \quad \bar{b}_0 = 0. && (6.4.48)
\end{aligned}
$$

The proofs of Theorem 6.7 are similar to the proof of Theorem 6.6.

Then, finite - dimensional filters for $T_t^{x(M)}$, $M = 0, 1, 2$, $T_t^{\gamma(0)}$ and U_t are approximately computed by

$$
\begin{aligned}
E[T_t^{x(M)}|\mathcal{Z}_t] &= a_t^{x(M)} + b_t^{x(M)}\tilde{\mu}_t + d_t^{x(M)}(\tilde{R}_t + \tilde{\mu}_t^2), \quad M = 0, 1, 2, && (6.4.49) \\
E[T_t^{\gamma(0)}|\mathcal{Z}_t] &= a_t^{\gamma(0)} + b_t^{\gamma(0)}\tilde{\mu}_t + d_t^{\gamma(0)}(\tilde{R}_t + \tilde{\mu}_t^2), && (6.4.50) \\
E[U_t|\mathcal{Z}_t] &= \bar{a}_t + \bar{b}_t\tilde{\mu}_t. && (6.4.51)
\end{aligned}
$$

6.5 Conclusions

In this chapter, we derived estimates for the parameters and hidden states in the state space model, with the noise in the observations being fractional Gaussian noise. It is shown that, using change of measure method, the conditional density of the hidden signal based on the history of the observations can be estimated. We also developed recursive approximate estimates of the density. The mean and variance of the approximate density were also estimated recursively. Finally, we derived maximum likelihood estimates of the parameters, using the EM algorithm.

Chapter 7

A NONLINEAR FILTER

7.1 Introduction

The Kalman filter is used for models with linear dynamics and additive, Gaussian noise. For mild non-linearities the extended Kalman filter often provides a good sub-optimal estimate. Particle filters have recently been popular for calculating Monte-Carlo estimates. However, when there is no noise in the state dynamics the particle filter does not work.

In this chapter we consider a scalar state process with non-linear dynamics $x = \{x_t, t = 0, 1, ...\}$, where $x_t = g(\theta, x_{t-1})$ for $t = 1, 2, ...$ Here θ is an unknown parameter. The process x is observed in Gaussian noise through a process $y = \{y_t, t = 0, 1, ...\}$, where $y_t = x_t + w_t$. Here $w = \{w_t, t = 0, 1, ...\}$ is a sequence of independent $N(0, \sigma)$ random variables.

Write $p_t(\theta, x)$ for the joint conditional density of (θ, x) given observations $y_0, y_1, ..., y_t$. By adapting methods from Chapter 2, we obtain a recursion for a discrete approximation to $p_t(\theta, x)$.

A particular case of our model is when x has Logistic dynamics given by

$$x_t = \theta x_{t-1}(1 - x_{t-1}). \tag{7.1.1}$$

Previous results for this problem are obtained in the paper [15] by Leung, Zhu and Ding.

7.2 Dynamics

Consider a state process $x = \{x_t, t = 0, 1, 2, ...\}$ whose dynamics depend on an unknown parameter θ, and whose evolution is described by:

$$x_t = g(\theta, x_{t-1}), \tag{7.2.1}$$

for $t = 1, 2, ...$

For simplicity we suppose the x_t and θ are scalar valued.

The state process is observed in additive white noise so we suppose we have a sequence of observations $y = \{y_t, t = 0, 1, 2, ...\}$ such that $y_t = x_t + w_t$. Here $w = \{w_t, t = 0, 1, 2, ...\}$ is a sequence of independent Gaussian random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Under P each w_t is Gaussian with mean 0 and variance σ.

However, to obtain the recursion we suppose that under another probability measure $\bar{P}$ the $y = \{y_t\}$ is a sequence of independent, Gaussian random variables, each having mean 0 and variance σ.

Considering the following $\sigma-$fields, or "histories" as Chapter 6:

$$\begin{aligned}
\mathcal{F}_t &= \sigma\{x_0, x_1, ..., x_t\}, \\
\mathcal{Y}_t &= \sigma\{y_0, y_1, ..., y_t\}, \\
\mathcal{G}_t &= \sigma\{x_0, y_0, x_1, y_1, ..., x_t, y_t\}.
\end{aligned}$$

Write $\phi_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$ for the $N(0, \sigma)$ density. Consider, for $k = 0, 1, 2, ...$, the variables

$$\Lambda_t := \prod_{l=0}^{t} \frac{\phi_\sigma(y_l - x_l)}{\phi_\sigma(y_l)}. \tag{7.2.2}$$

Then the following result can be provided as in [36].

Lemma 7.1. *Define the probability measure P in terms of $\bar{P}$ by setting*

$$\frac{dP}{d\bar{P}}|_{\mathcal{G}_{t-1}} := \Lambda_t. \tag{7.2.3}$$

Then under P the random variables $w_t := y_t - x_t$ are i.i.d. and $N(0, \sigma)$. That is, under P, as required

$$y_t = x_t + w_t. \tag{7.2.4}$$

Proof.

$$\begin{aligned} P(w_t \leq a|\mathcal{G}_{t-1}) &= E[I_{w_t \leq a}|\mathcal{G}_{t-1}] \\ &= \frac{\bar{E}[\Lambda_t I_{w_t \leq a}|\mathcal{G}_{t-1}]}{\bar{E}[\Lambda_t|\mathcal{G}_{t-1}]} \\ &= \frac{\Lambda_{t-1}\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)} I_{w_t \leq a}|\mathcal{G}_{t-1}]}{\Lambda_{t-1}\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)}|\mathcal{G}_{t-1}]} \\ &= \frac{\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)} I_{w_t \leq a}|\mathcal{G}_{t-1}]}{\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)}|\mathcal{G}_{t-1}]}. \end{aligned}$$

Write

$$w_t = y_t - x_t.$$

Then

$$\begin{aligned} \bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)}|\mathcal{G}_{t-1} \vee x_t] &= \int_{\Re} \frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)} \phi_\sigma(y_t) dy_t \\ &= \int_{\Re} \phi_\sigma(y_t - x_t) d(x_t + w_t) \\ &= \int_{\Re} \phi_\sigma(w_t) dw_t \\ &= 1. \end{aligned}$$

$$\begin{aligned} \bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)}|\mathcal{G}_{t-1}] &= \bar{E}[\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)}|\mathcal{G}_{t-1} \vee x_t]|\mathcal{G}_{t-1}] \\ &= 1. \end{aligned}$$

$$\begin{aligned}\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)} I_{w_t \le a} | \mathcal{G}_{t-1} \vee x_t] &= \int_{\Re} \frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)} I_{w_t \le a} \phi_\sigma(y_t) dy_t \\ &= \int_{\Re} \phi_\sigma(y_t - x_t) I_{w_t \le a} d(x_t + w_t) \\ &= \int_{\Re} \phi_\sigma(w_t) I_{w_t \le a} dw_t \\ &= \int_{-\infty}^{a} \phi_\sigma(w_t) dw_t.\end{aligned}$$

$$\begin{aligned}\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)} I_{w_t \le a} | \mathcal{G}_{t-1}] &= \bar{E}[\bar{E}[\frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)} I_{w_t \le a} | \mathcal{G}_{t-1} \vee x_t] | \mathcal{G}_{t-1}] \\ &= \int_{-\infty}^{a} \phi_\sigma(w_t) dw_t.\end{aligned}$$

So we have

$$P(w_t | \mathcal{G}_{t-1}) = \int_{-\infty}^{a} \phi_\sigma(w_t) dw_t,$$

and the result follows. □

7.3 Filtering

We wish to determine recursive estimates of θ and x_t given the observations $y_0, y_1, ..., y_t$. Consider arbitrary bounded, measurable functions h and f. Then we wish to estimate, (under the measure P),

$$E[h(\theta) f(x_t) | \mathcal{Y}_t] = \frac{\bar{E}[\Lambda_t h(\theta) f(x_t) | \mathcal{Y}_t]}{\bar{E}[\Lambda_t | \mathcal{Y}_t]}. \tag{7.3.1}$$

Write the numerator as

$$\sigma_t\big(h(\theta) f(x_t)\big) = \bar{E}[\Lambda_t h(\theta) f(x_t) | \mathcal{Y}_t]. \tag{7.3.2}$$

Then $\sigma_t(1) = \bar{E}[\Lambda_t|\mathcal{Y}_t]$. σ_t is a continuous linear functional on functions (h, f); suppose it is given by a density q_t so that

$$\sigma_t\big(h(\theta)f(x_t)\big) = \int_{\Re}\int_{\Re} h(\theta)f(x)q_t(\theta, x)d\theta dx. \tag{7.3.3}$$

Then q_t is an unnormalized conditional density of (θ, x_t) given $\mathcal{Y}_t$ and

$$\begin{aligned} P(\theta \in d\theta, x_t \in dx|\mathcal{Y}_t) &= E[I_{\theta\in d\theta}I_{x_t\in dx}|\mathcal{Y}_t] \\ &= \frac{q_t(\theta, x)d\theta dx}{\int_{\Re}\int_{\Re} q_t(y, z)dydz}. \end{aligned} \tag{7.3.4}$$

We shall obtain a recursion for $q_t(\theta, x)$.

Theorem 7.1.

$$q_0(\theta, x) = \frac{\phi_\sigma(y_0 - x_0)}{\phi_\sigma(y_0)}p_0(\theta, x),$$

where $p_0(\theta, x)$ *is the initial density of* (θ, x_0). *For* $t = 1, 2, ...,$

$$q_t(\theta, g(\theta, x)) = \phi_\sigma(y_t)^{-1}g_x(\theta, x)^{-1}\phi_\sigma(y_t - g(\theta, x))q_{t-1}(\theta, x). \tag{7.3.5}$$

Proof. It is easy to see $q_0(\theta, x) = \frac{\phi_\sigma(y_0 - x_0)}{\phi_\sigma(y_0)}p_0(\theta, x)$. When $t \geq 1$,

$$\begin{aligned} &\sigma_t(h(\theta)f(x_t)) \\ =\ & \bar{E}[\Lambda_t h(\theta)f(x_t)|\mathcal{Y}_t] \\ =\ & \int_{\Re}\int_{\Re} h(\theta)f(x)q_t(\theta, x)d\theta dx \qquad (7.3.6) \\ =\ & \bar{E}[\Lambda_{t-1} \times \frac{\phi_\sigma(y_t - x_t)}{\phi_\sigma(y_t)}h(\theta)f(x_t)|\mathcal{Y}_t] \\ =\ & \bar{E}[\Lambda_{t-1} \times \frac{\phi_\sigma(y_t - g(\theta, x_{t-1}))}{\phi_\sigma(y_t)}h(\theta)f(g(\theta, x_{t-1}))|\mathcal{Y}_t] \\ =\ & \int_{\Re}\int_{\Re} h(\theta)f(g(\theta, x))\frac{\phi_\sigma(y_t - g(\theta, x))}{\phi_\sigma(y_t)}q_{t-1}(\theta, x)d\theta dx. \end{aligned}$$

Substituting $x = g(\theta, z)$ in (7.3.6) we have that

$$\sigma_t(h(\theta)f(x_t)) = \int_{\Re}\int_{\Re} h(\theta)f(g(\theta, z))q_t(\theta, g(\theta, z))g_z(\theta, z)d\theta dz.$$

As h and f are arbitrary functions we see that

$$q_t(\theta, g(\theta, x))g_x(\theta, x) = \frac{\phi_\sigma(y_t - g(\theta, x))}{\phi_\sigma(y_t)} q_{t-1}(\theta, x)$$

and the result follows. □

7.4 Numerics

We shall discretize the recursion given in theorem 7.1. Suppose initially the joint probability distribution of (θ, x_0) is approximated by probability masses located at points (θ_i, x_{ij}), $1 \leq i \leq M$, $1 \leq j \leq N$.

Write

$$P_0(\theta_i, x_{ij}) = P(\theta = \theta_i, x = x_{ij}), \tag{7.4.1}$$

$$x_{ij}(0) = x_{ij}, \tag{7.4.2}$$

$$x_{ij}(t) = g(\theta_i, x_{ij}(k-1)), k = 1, 2, ..., \tag{7.4.3}$$

$$q_0(\theta_i, x_{ij}(0)) = \frac{\phi_\sigma(y_0 - x_{ij}(0))}{\phi_\sigma(y_0)} \delta_{x_{ij}}(x). \tag{7.4.4}$$

For $t = 1, 2, ...,$

$$q_t(\theta_i, x_{ij}(t)) = \frac{\phi_\sigma(y_t - x_{ij}(t))q_{t-1}(\theta_i, x_{ij}(t-1))}{\phi_\sigma(y_t)g_x(\theta_i, x_{ij}(t-1))}. \tag{7.4.5}$$

Given the observations $y_0, y_1, ..., y_t$, that is, given $\mathcal{Y}_t$, the unnormalized probability that $\theta = \theta_i$ and $x(t) = x_{ij}(t)$ is

$$q_t(\theta_i, x_{ij}(t))g_x(\theta_i, x_{ij}(t-1)). \tag{7.4.6}$$

The normalized probability that $\theta = \theta_i$ and $x(t) = x_{ij}(t)$ is

$$P_t(\theta_i, x_{ij}(t)) = \frac{q_t(\theta_i, x_{ij}(t))g_x(\theta_i, x_{ij}(t-1))}{\sum_{l=1}^{M}\sum_{\delta=1}^{N} q_t(\theta_l, x_{l\delta}(t))g_x(\theta_l, x_{l\delta}(t-1))}. \tag{7.4.7}$$

Given $\mathcal{Y}_t$ the expected value of θ is:

$$\sum_{l=1}^{M}\sum_{\delta=1}^{N}\theta_l P_t(\theta_l, x_{l\delta}(t)). \tag{7.4.8}$$

Given $\mathcal{Y}_t$ the expected value of $x(t)$ is:

$$\sum_{l=1}^{M}\sum_{\delta=1}^{N}x_{l\delta}(k)P_t(\theta_l, x_{l\delta}(t)). \tag{7.4.9}$$

Note that care must be taken to avoid the zeros of $g_x(\theta, x)$.

7.5 Simulation

To demonstrate the performance of the filter described in this chapter, we consider two examples with different nonlinear functions of the state dynamics. We find our method works well for nonlinear functions.

For the first example, we consider the logistic dynamics.

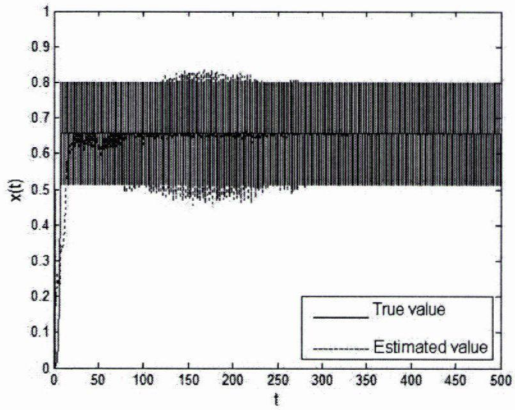
$$x_t = \theta x_{t-1}(1 - x_{t-1}),$$

where $\theta = 3.2$. The results for the state estimation and the convergence of the parameter are shown in figure 7.1.

Next, we consider another non-linear dynamics.

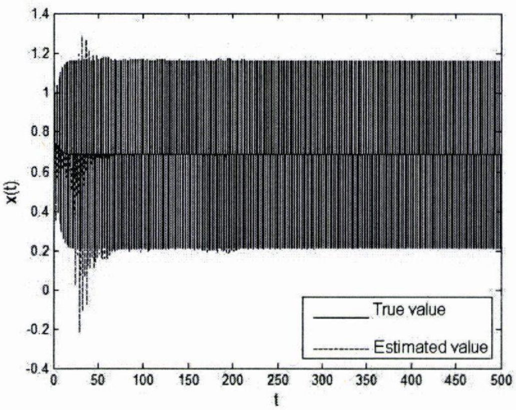
$$x_t = \theta cos(\theta x_{t-1}),$$

where $\theta = 1.2$. The results are shown in figure 7.2.



(a) (b)

Figure 7.1: Estimation for logistic function



(a) (b)

Figure 7.2: Estimation for Cosine function

Chapter 8

A NONLINEAR FILTER WITH FRACTIONAL GAUSSIAN NOISE

8.1 Introduction

Following Chapter 7, in this chapter, we consider a discrete time, state space model, where the signal has non-linear dynamics, and is observed through a real valued process which is corrupted by fractional Gaussian noise. We derive an exact estimate and an approximate recursive estimate for the conditional density of the hidden signal and the parameter, using the change of measure method.

This chapter is arranged as follows. In the next section, we give a brief description of fractional Gaussian noise and the model used in this chapter. In section 8.3, we derive the exact estimate for the conditional density of the hidden signal and the parameter. In section 8.4, we derive an approximate estimate for the conditional density of the hidden signal and a parameter. In the final section we give some conclusions.

8.2 State Space Model with Fractional Gaussian Noise

The dynamics of the scalar state process $x = \{x_t, t = 0, 1, ...\}$ is as described in Chapter 7. Similarly, x is not observed directly, but observed through another scalar process $y = \{y_t, t = 0, 1, ...\}$, whose values are corrupted by fractional Gaussian noise. Here, we consider only the case where the observations and hidden states have the simple linear relation

$$y_t = x_t + w_t^r, \tag{8.2.1}$$

where $w^r = \{w_t^r, t = 0, 1, ...\}$ is the measurement noise. This is now a sequence of fractional Gaussian random variables as described in Chapter 5.

As in Chapter 5, write

$$\begin{aligned} z_t &= (u^{-r} * y)(t), \\ \gamma_t(x_0, x_1, ..., x_t) &= (u^{-r} * x)(t). \end{aligned}$$

Then (8.2.1) implies the following equation.

$$z_t = \gamma_t(x_0, ..., x_t) + w_t. \tag{8.2.2}$$

These are still the dynamics of z under the 'real world' probability P.

8.3 Filtering

Consider a probability measure $\bar{P}$ on the measurable space $(\Omega, \mathcal{F})$ such that, under $\bar{P}$, $\{z_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables. We call the measure $\bar{P}$ a "reference" probability.

We now construct the 'real world' probability P from $\bar{P}$, such that, under P, $\{w_t\}$ is a sequence of $N(0,1)$ i.i.d. random variables, where $w_t = z_t - \gamma_t(x_0, ..., x_t)$.

Write

$$\phi(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2}),$$

and for $l \geq 0$,

$$\lambda_l = \frac{\phi(z_l - \gamma_l(x_0, ..., x_l))}{\phi(z_l)}, \tag{8.3.1}$$

$$\Lambda_t = \prod_{l=0}^{t} \lambda_l, \quad t = 0, 1, 2, ... \tag{8.3.2}$$

Definition 8.1. *Define P by putting*

$$\frac{dP}{d\bar{P}}|_{\mathcal{G}^z_{t-1}} = \Lambda_t. \tag{8.3.3}$$

Theorem 8.1. *Under P the random variables $w_t := z_t - \gamma_t(x_0, ..., x_t)$ are i.i.d. and $N(0, 1)$.*

The proof of Theorem 8.1 is the same as the proof of Theorem 5.2.

Corollary 8.1. *Under P, $y_t := (u^r * z)(t) = (u^r * \gamma)(t) + (u^r * w)(t)$. That is,*

$$y_t = x_t + w_t^r.$$

Write $p_t(\theta, x)$ for the joint conditional density of (θ, x) given observations $y_0, y_1, ..., y_t$. In the remainder of this section, we shall obtain an exact estimate for $p_t(\theta, x)$.

Write (7.3.2) to be

$$\begin{aligned} \sigma_t(h(\theta)f(x_t)) &= \bar{E}[\Lambda_t h(\theta)f(x_t)|\mathcal{Y}_t] \\ &= \bar{E}[\Lambda_t h(\theta)f(x_t)|\mathcal{Z}_t]. \end{aligned}$$

Similarly, $\sigma_t(h(\theta)f(x_t))$ could be given by a density q_t such that

$$\sigma_t(h(\theta)f(x_t)) = \int_{\Re}\int_{\Re} h(\theta)f(x)q_t(\theta, x)d\theta dx,$$

where q_t is an unnormalized conditional density of (θ, x_t) given $\mathcal{Z}_t$ and

$$\begin{aligned} P(\theta \in d\theta, x_t \in dx|\mathcal{Z}_t) &= E[I_{\theta \in d\theta} I_{x_t \in dt}|\mathcal{Z}_t] \\ &= \frac{q_t(\theta, x)d\theta dx}{\int_{\Re}\int_{\Re} q_t(u, v)dudv}. \end{aligned}$$

Theorem 8.2.

$$q_0(\theta, x) = \frac{\phi(z_0 - x_0)}{\phi(z_0)} p_0(\theta, x),$$

where $p_0(\theta, x)$ *is the initial density of* (θ, x_0). *For* $t = 1, 2, ...,$ *the unnormalized conditional density* $q_t(\theta, \bullet)$ *is estimated by*

$$
\begin{aligned}
& q_t(\theta, g(\theta, u))\frac{\partial g(\theta, u)}{\partial u} \\
= & \int ... \int \frac{\phi(z_0 - \gamma_0(x_0))}{\phi(z_0)} ... \frac{\phi(z_t - \gamma_t(x_0, ..., g(\theta, x_{t-1})))}{\phi(z_t)} h(\theta) f(g(\theta, x_{t-1}))[\int q_0(\theta, x_0)d\theta] ... \\
& [\int q_{t-2}(\theta, x_{t-2})d\theta] q_{t-1}(\theta, u)dx_0...dx_{t-2}. \qquad (8.3.4)
\end{aligned}
$$

Proof. It is easy to see $q_0(\theta, x) = \frac{\phi(z_0 - x_0)}{\phi(z_0)} p_0(\theta, x)$. When $t \geq 1$,

$$
\begin{aligned}
& \sigma_t(h(\theta) f(x_t)) \\
= & \bar{E}[\Lambda_t h(\theta) f(x_t) | \mathcal{Z}_t] \\
= & \bar{E}[\Lambda_{t-1} \frac{\phi(z_t - \gamma_t(x_0, ..., x_t))}{\phi(z_t)} h(\theta) f(x_t) | \mathcal{Z}_t] \\
= & \bar{E}[\Lambda_{t-2} \frac{\phi(z_{t-1} - \gamma_{t-1}(x_0, ..., x_{t-1}))}{\phi(z_{t-1})} \frac{\phi(z_t - \gamma_t(x_0, ..., x_t))}{\phi(z_t)} h(\theta) f(x_t) | \mathcal{Z}_t] \\
= & ...... \\
= & \bar{E}[\frac{\phi(z_0 - \gamma_0(x_0))}{\phi(z_0)} ... \frac{\phi(z_t - \gamma_t(x_0, ..., x_t))}{\phi(z_t)} h(\theta) f(x_t) | \mathcal{Z}_t] \\
= & \bar{E}[\frac{\phi(z_0 - \gamma_0(x_0))}{\phi(z_0)} ... \frac{\phi(z_t - \gamma_t(x_0, ..., g(\theta, x_{t-1})))}{\phi(z_t)} h(\theta) f(g(\theta, x_{t-1})) | \mathcal{Z}_t] \\
= & \int ... \int \frac{\phi(z_0 - \gamma_0(x_0))}{\phi(z_0)} ... \frac{\phi(z_t - \gamma_t(x_0, ..., g(\theta, x_{t-1})))}{\phi(z_t)} h(\theta) f(g(\theta, x_{t-1}))[\int q_0(\theta, x_0)d\theta] ... \\
& [\int q_{t-2}(\theta, x_{t-2})d\theta] q_{t-1}(\theta, x_{t-1})dx_0...dx_{t-1}d\theta.
\end{aligned}
$$

Also,

$$
\begin{aligned}
& \sigma_t(h(\theta) f(x_t)) \\
= & \int \int h(\theta) f(x) q_t(\theta, x) d\theta dx \\
= & \int \int h(\theta) f(g(\theta, x_{t-1})) q_t(\theta, g(\theta, x_{t-1})) \frac{\partial g(\theta, x_{t-1})}{\partial x_{t-1}} d\theta dx_{t-1}.
\end{aligned}
$$

As h and f are arbitrary functions, we see that

$$
\begin{aligned}
& q_t(\theta, g(\theta,u))\frac{\partial g(\theta,u)}{\partial u} \\
= & \int ... \int \frac{\phi(z_0-\gamma_0(x_0))}{\phi(z_0)} ... \frac{\phi(z_t-\gamma_t(x_0,...,g(\theta,x_{t-1})))}{\phi(z_t)} h(\theta)f(g(\theta,x_{t-1}))[\int q_0(\theta,x_0)d\theta]... \\
& [\int q_{t-2}(\theta,x_{t-2})d\theta]q_{t-1}(\theta,u)dx_1...dx_{t-2}.
\end{aligned}
$$

□

8.4 Approximate Estimation of States and Parameters

In this section, we derive recursive approximate estimates of $q_t(\theta, x_t)$.

Theorem 8.3. *The unnormalized density $q_t(\theta, \bullet)$ is approximately computed by the recursion*

$$
\tilde{q}_t(\theta, g(\theta,u))\frac{\partial g(\theta,u)}{\partial u} = \frac{\phi(z_t-\gamma_t(x_0,\tilde{x}_1...,\tilde{x}_{t-2},u,g(\theta,u)))}{\phi(z_t)}\tilde{q}_{t-1}(\theta,u).
$$

Proof.

$$
\begin{aligned}
& \sigma_t(h(\theta)f(x_t)) \\
= & \int\int h(\theta)f(x)q_t(\theta,x)d\theta dx \\
= & \bar{E}[\Lambda_t h(\theta)f(x_t)|\mathcal{Z}_t] \\
= & \bar{E}[\Lambda_{t-1}\frac{\phi(z_t-\gamma_t(x_0,...,x_t))}{\phi(z_t)}h(\theta)f(x_t)|\mathcal{Z}_t].
\end{aligned} \tag{8.4.1}
$$

Suppose $\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_{t-1}$ have been defined, then we define $\tilde{q}_t(\theta, \bullet)$ such that

$$
\begin{aligned}
& \int\int h(\theta)f(x)\tilde{q}_t(\theta,x)d\theta dx \\
= & \bar{E}[\Lambda_{t-1}\frac{\phi(z_t-\gamma_t(x_0,\tilde{x}_1,...,\tilde{x}_{t-2}.x_{t-1},x_t))}{\phi(z_t)}h(\theta)f(x_t)|\mathcal{Z}_t] \\
= & \bar{E}[\Lambda_{t-1}\frac{\phi(z_t-\gamma_t(x_0,\tilde{x}_1...,\tilde{x}_{t-2},x_{t-1},g(\theta,x_{t-1})))}{\phi(z_t)}h(\theta)f(g(\theta,x_{t-1}))|\mathcal{Z}_k] \\
= & \int\int h(\theta)f(g(\theta,u))\frac{\phi(z_t-\gamma_t(x_0,\tilde{x}_1...,\tilde{x}_{t-2},u,g(\theta,u)))}{\phi(z_t)}\tilde{q}_{t-1}(\theta,u)d\theta du.
\end{aligned}
$$

Here, for $0 < l \leq t-2$, $\tilde{x}_l = \frac{\int\int x\tilde{q}_l(\theta,x)d\theta dx}{\int \tilde{q}_l(\theta,x)d\theta dx}$.

Also,

$$\begin{aligned} & \int\int h(\theta)f(x)\tilde{q}_t(\theta,x)d\theta dx \\ = & \int\int h(\theta)f(g(\theta,u))\tilde{q}_t(\theta,g(\theta,u))\frac{\partial g(\theta,u)}{\partial u}d\theta du. \end{aligned}$$

As h and f are arbitrary functions, we have

$$\tilde{q}_t(\theta,g(\theta,u))\frac{\partial g(\theta,u)}{\partial u} = \frac{\phi(z_t - \gamma_t(x_0,\tilde{x}_1...,\tilde{x}_{t-2},u,g(\theta,u)))}{\phi(z_t)}\tilde{q}_{t-1}(\theta,u).$$

□

8.5 Conclusions

In this chapter, we considered a state space model where the hidden states have a non-linear relationship, and the noise in the observations is fractional Gaussian noise. We introduced a joint density of the hidden states and an unknown parameter, and derived both an exact estimate and an approximate recursive estimate of the conditional joint density of the hidden state and the unknown parameter, based on the history of the observations.

Bibliography

[1] A. Dembo and O. Zeitouni, "Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm", *Stochastic processes and their applications*, **23**, 1986, pp. 91 - 113.

[2] A. J. Aguirre, C. Brenman, G. Bailey, R. Sinha, B. Feng, C. Leo, et al., "High-resolution characterization of the pancreatic adenocarcinoma genome", *PNAS 101*, 2004, pp. 9067 - 9072.

[3] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society, Series B (Methodological)*, **39 (1)**, 1977, pp. 1 - 38.

[4] B. K. Ray and R. S. Tsay, "Long-range denpendence in daily stock volatilities", *Journal of Business & Economics Statistics*, **18**, 2000, pp. 254 - 262.

[5] C. Chang and R. Ansari, "Kernel particle filter: iterative sampling for efficient visual tracking", *Proceeding of 2003 International Conference on Image Processing*, 2003, pp. III-977 - III-980.

[6] C. Chang and R. Ansari, "Kernel particle filter for visual tracking", *IEEE signal processing letters*, **12 (3)**, 2005, pp. 242 - 245.

[7] C. C. Wen, Y. J. Wu, Y. H. Huang, W. C. Chen, S. C. Liu, S. S. Jiang, J. L. Juang, C. Y. Lin, W. T. Fang, C. A. Hsiung and I. S. Chang, "A Bayes regression approach to array-CGH data", *Statistical Applications in Genetics and Molecular Biology*, **5 (1)**, 2006.

[8] C. Erdman and J. W. Emerson, "A fast Bayesian change point analysis for the segmentation of microarray data", *Bioinformatics*, **24**, 2008, pp. 2143 - 2148.

[9] C. Yau, O. Papaspiliopoulos, G. O. Roberts and C. Holmes, "Bayesian nonparametric hidden Markov models with application to the analysis of copy-number-variation in mammalian genomes", *J. Roy. Stat. Soc. Series B*, 2009, Under review.

[10] E. A. Wan and R. van der Merwe, "The unscented Kalman filter for nonlinear estimation", *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*, 2000, pp. 153 - 158.

[11] E. Cuevas, D. Zaldivar and R. Rojas, "Kalman filter for vision tracking", *Technical report*, 2005.

[12] E. Perrin, R. Harba, C. Berzin-Joseph, I. Iribarren and Aline Bonami, "nth-Order fractional Brownian motion and fractional Gaussian noises", *IEEE Transactions on Signal Processing*, **49 (5)**, 2001, pp. 1049 - 1059.

[13] F. Picard, S. Robin, M. Lavielle, C. Vaisse and J. Daudin, "A statistical approach for array CGH data analysis", *BMC Bioinformatics*, **6 (27)**, 2005.

[14] G. Welch and G. Bishop, "An introduction to the Kalman filter", *Technical report*, 2006.

[15] H. Leung, Z. Zhu and Z Ding, "An aperiodic phenomenon of the extended Kalman filter in filtering noisy chaotic signals", *IEEE Trans. Signal Processing*, **48 (6)**, 2000, pp. 1807-1810.

[16] H. Qian, "Fractional Brownian motion and fractional Gaussian noise", *Processes*

with Long-Range Correlations Theory and Applications, Edited by G. Rangarajan and M. Ding, Lecture Notes in Physics, **621**, 2003, pp.22 - 33.

[17] I. N. Lobato and N. E. Savin, "Real and spurious long-memory properties of stock-market data", *Journal of Business & Economics Statistics*, **16**, 1998, pp. 261 - 267.

[18] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models", *TR-97-021*, 1998.

[19] J. Martin, J. Gibrat and F. Rodolphe, "Hidden Markov Model for protein secondary structure", *Proceedings of International Symposium on Applied Stochastic Models and Data Analysis: Part. III. Bioinformatics and Statistics*, 2005.

[20] K. Asai, S. Hayamizu and K. Onizuka, "HMM with protein structure grammar", *Proceeding of the Twenty-Sixth Hawaii International Conference on System Sciences*, 1993, pp. 783 - 791.

[21] L. Aggoun and R. J. Elliott, "Measure theory and filtering —- introduction with applications", *Cambridge University press*, 2004.

[22] L. R. Rabiner and B. H. Juang, "An introduction to Hidden Markov Models", *IEEE ASSP Magazine*, **3**, 1986, pp. 4 - 16.

[23] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, 1989, pp. 257 - 286.

[24] M. Khojasteh, B. Coel, S. Shah, R. K. Ward, W. L. Lam and C. MacAulay, "A novel algorithm for the analysis Of array CGH data", *Proceeding of 2006 IEEE*

International Conference on Acoustics, Speech and Signal Processing, 2006, pp. II-1020 - II-1023.

[25] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear non-Gaussian Bayesian tracking", *IEEE transactions on signal processing*, **50** (**2**), 2002, pp. 174 - 188.

[26] N. G. Polson, J. R. Stroud and Peter Muller, "Practical filtering with sequential parameter learning", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**, 2008, pp. 413 - 428.

[27] O. Kallenberg, "Foundations of modern probability (Second edition)", *Springer*, 2002, pp. 119 - 159.

[28] R. J. Elliott, "Exact adaptive filters for Markov chains observed in Gaussian noise", *Automatica*, **30** (**9**), 1994, pp. 1399 - 1408.

[29] R. J. Elliott, H. Leung and J. Deng, "A non-Linear filter", *Stochastic Analysis and Applications*, **26** (**4**), 2008, pp. 856 - 862.

[30] R. J. Elliott and H. Miao, "Stochastic volatility model with filtering", *Stochastic Analysis and Applications*, **24**, 2006, pp. 661 - 683.

[31] R. J. Elliott and H. Miao, "Fractional differencing in discrete time", Preprint.

[32] R. J. Elliott and J. Deng, "A Viterbi smoother for discrete state space model", *Systems & Control Letters*, **58** (**6**), 2009, pp. 400 - 405.

[33] R. J. Elliott and J. Deng, "A filter for a state space model with fractional Gaussian noise", *Automatica*, **46** (**10**), 2010, pp. 1689 - 1695.

[34] R. J. Elliott and J. Deng, "A filter for hidden Markov model with fractional Gaussian noise", *Systems & Control Letters*, **60** **(2)**, 2011, pp. 93 - 100.

[35] R. J. Elliott and J. Deng, "A nonlinear filter with fractional Gaussian noise", *Stochastic Analysis and Applications*, In press.

[36] R. J. Elliott, L. Aggoun and J. B. Moore, " Hidden Markov models: estimation and control", *Springer - Verlag(1995), Second printing(1997), Berlin Heidelberg - New York.*

[37] R. J. Elliott and W. P. Malcolm, "Data-recursive smoother formulae for partially observed discrete-time Markov chains", *Stochastic Analysis and Applications*, **24**, 2006, pp. 579-597.

[38] S. Bregni and W. Erangoli, "Fractional noise in experimental measurements of IP traffic in a metropolitan area network", *Proc. of IEEE GLOBECOM 2005*, **2**, 2005, pp. 781 - 785.

[39] S. Chow, E. Ferrer and J. R. Nesselroade, "An unscented Kalman filter approach to the estimation of nonlinear dynamical systems models", *Multivariate Behavioral Research*, **42** **(2)**, 2007, pp. 283 - 321.

[40] S. Colella, C. Yau, J. M. Taylor, G. Mirza, H. Butler, P. Clouston, A. S. Bassett, A. Seller, C. C. Holmes and J. Ragoussis, "QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data", *Nucleic Acids Research, 35 (6)*, 2007, pp. 2013 - 2025.

[41] S. Guha, Y. Li and D. Neuberg, "Bayesian hidden Markov modeling of array CGH data", *Journal of the American Statistical Association.*, **103**, 2008, pp. 485 - 497.

[42] S. J. Julier, J. K. Uhlrnann and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems", *Proceeding of the American Control Conference*, 1995.

[43] S. Maybank, "Finite-dimensional filters", *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **354 (1710)**, 1996, pp. 1099 - 1123.

[44] S. E. Shreve, "Stochastic calculus for finance I: The Binomial asset pricing model", *Springer Finance*, 2005, pp. 25 - 52.

[45] S. E. Shreve, "Stochastic calculus for finance II: Continuous-time models", *Springer Finance*, 2004, pp. 107 - 108, 266 - 268.

[46] S. Stjernqvist, T. Ryden, M. Skold and J. Staaf, "Continuous-index hidden Markov modelling of array CGH copy number data", *Bioinformatics*, **23 (8)**, 2007, pp. 1006 - 1014.

[47] W. Tsay and W. K. Hardle, "A generalized ARFIMA process with Markov-switching fractional differencing parameter", *Journal of Statistical Computation and Simulation*, **79 (5)**, 2009, pp. 731 - 745.

[48] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson, "Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements", *Statistical Science*, **10**, 1995, pp. 67 - 85.

[49] X. Tang, "Autoregressive hidden Markov model with application in an EL NINO study", *A thesis for the degree of Master of Science, Department of Mathematics and Statistics, University of Saskatchewan*, 2004.

[50] Z. Li, H. Leung, and E. Bosse, "Joint data association, registration and fusion using EM-EKF", Working paper.