



Extensions of
Markov Modulated Poisson Processes
and Their Applications to
Deep Earthquakes

by

Shaochuan Lu

A Thesis

submitted to the Victoria University of Wellington
in fulfillment of the requirements for the degree of
Doctor of Philosophy in Statistics

September 25, 2009

Abstract

The focus of this thesis is on the Markov modulated Poisson process (MMPP) and its extensions, aiming to propose appropriate statistical models for the occurrence patterns of main New Zealand deep earthquakes. Such an attempt might be beyond the scope of the MMPP and its extensions, however we hope its main patterns can be characterized by current models proposed in three parts of the thesis.

The first part of the thesis is concerned with introductions and preliminaries of discrete time hidden Markov models (HMMs) and MMPP. The flexibility in model formulation and openness in model framework of HMMs are reviewed in this part, suggesting also possible extensions of MMPP.

The second part of the thesis is mainly about several extensions of MMPP. One extension of MMPP is by associating each occurrence of MMPP with a mark. Such an extension is potentially useful for spatial-temporal modelling or other point processes with marks. A special case of this type of extension is by allowing the multiple observations of MMPP synchronized together under the same Markov chain. This extension opens the possibility of modelling multiple point process observations with weak dependence. The third extension is motivated by the attempt to describe small scale temporal clustering existing in the deep earthquakes via treating the recognized aftershocks as marks which itself forms a finite point process. The rest of the second part focuses on some information theoretical aspects of MMPPs such as the entropy rate of the underlying Markov chain and observed point process respectively and their mutual information rate. A conjecture on the possible links between mutual

information rate of MMPP and the Fisher information of the estimated parameters is suggested. The second part on extensions of MMPP is featured by the derivation of the likelihood and complete likelihood, parameter estimation via EM algorithm, state smoothing estimation and model evaluation through systematic applications of rescaling theory of multivariate point processes and marked point processes.

The third part of the thesis includes the applications of these methods to the deep earthquakes in New Zealand. We first evaluate the data coverage, catalogue completeness and explore its descriptive characteristics and empirical properties such as epicentral distributions, depth distributions and magnitude distributions. Clustering behavior is studied via the second order moment analysis of point processes in the chapter 8. We also apply, the stress release models and the ETAS models which are usually used for shallow earthquakes, to the New Zealand deep earthquakes and provide tentative explanations of why they are not satisfactory for the deep earthquakes. The chapter 9 is on the applications of MMPP and its extensions to the New Zealand deep earthquakes. Conclusions and future studies are presented in chapter 10.

Statement of Originality

As far as the author knows, the application of MMPP to the deep earthquakes (chapter 9), several extensions of MMPP such as MMPP with state-dependent marks (chapter 4) and MMPP with a subsidiary Neyman-Scott type point process (chapter 6) and the information theoretical aspects of MMPP in chapter 7 are original contributions. Chapter 8, on exploratory data analysis, is also author's own work except the literature review part.

Acknowledgements

Many thanks to my supervisor David Vere-Jones for his over years mentoring on this thesis. I was most benefited from his foresight, thoroughness and encouragement which makes my PhD study on statistical seismology challenging though, exciting at many moments for the existing possibilities of exploring some extraordinary complex pattern appearing in earthquakes through statistical methods.

I am also indebted to my supervisor David Harte for his many very constructive suggestions for my thesis. His well-written codes of hidden Markov models, particularly SSLib package cooperated with Ray Brownrigg and others, greatly facilitated my study of New Zealand deep earthquake modeling.

I thank Prof. Ma Li and Prof. Li Yong when I was a master student in Beijing Normal University for their guidance and recommendations for joining in this project.

I also thank Dr. Stefanka Chukova for her help in preparation of the examinations and her very good suggestions for modifying the thesis. Suggestions of Dr. Richard Arnold, Dr. David Bryant and Dr. Dimitris Karlis are acknowledged.

I am also in gratitude to discussants, seminar attendees, visitors who are working on statistical seismology, among them, Wang Ting, Mark Bebbington, Zhuang Jianchang and others. The financial support from New Zealand Institute of Mathematics and its Applications (NZIMA) is particularly acknowledged. I am also in gratitude to the support of School of Mathematics, Statistics and Computer Science

of Victoria University of Wellington.

At last, but not least, I express my thanks to my family for their support, particularly my wife Chen Bo and my parents.

Contents

I	INTRODUCTION AND PRELIMINARIES	1
1	Objectives and Motivations	2
2	Introduction	6
2.1	Preliminaries of earthquakes	6
2.2	Notations and Definitions	12
2.3	Basic conceptions in Point processes	13
2.4	Hidden Markov Models in Discrete Time	16
2.4.1	The Basic Hidden Markov Model	16
2.4.2	Hidden Markov model with covariates	21
2.4.3	Hidden semi-Markov models	24
3	Markov Modulated Poisson Process (MMPP)	26

<i>CONTENTS</i>	vii
-----------------	-----

3.1	Introduction	26
3.2	Parameter Estimation	29
3.3	Statistical Inference and Model Evaluation	32
3.4	Second-Order Moment Properties	33
3.5	Relatives of MMPP: MAP, BMAP	35
3.6	Chapter Appendix	38

II EXTENSIONS AND INFORMATION THEORETI- CAL ASPECTS 39

4	Markov Modulated Poisson Process with Marks	40
4.1	Introduction	40
4.2	The Likelihood	42
4.3	Parameter Estimation of MMPP with Exponential Family Marks . .	45
4.3.1	Parameter estimation through EM algorithm	45
4.3.2	Examples	48
4.4	Inference on State Process and Observed Point Process	50
4.4.1	Estimate hidden state and intensity rate of observed point process	50

4.4.2	Assessing the goodness-of-fit	51
4.5	Simulation study	53
5	Multivariate MMPP	59
5.1	Introduction	59
5.2	The Likelihood	62
5.3	Parameter Estimation through EM Algorithm	63
5.4	Second Order Moment Analysis	64
5.5	Inference on State Process and Hidden Process, Model Evaluation . .	68
6	MMPP with a subsidiary Neyman-Scott type Process	70
6.1	Introduction	70
6.2	MMPP with compounding marks	71
6.3	MMPP with a Subsidiary Neyman-Scott cluster process	72
7	Information Theoretical Aspects of MMPP	76
7.1	Introduction	76
7.2	The Entropy of the Underlying Markov Chain	78
7.3	The Entropy Rate of MMPP	80
7.3.1	The Entropy of Point Process	80

7.3.2	The Entropy of Stationary Point Process and Marked Point Process	82
7.3.3	The Entropy Rate of MMPP	83
7.4	The Mutual Information Rate of the MMPPs	83
7.5	Fisher Information and Observed Fisher Information	87
7.6	Simulation Studies	90
7.7	Chapter Appendix	91
III	APPLICATION TO DEEP EARTHQUAKES	95
8	Exploratory Data Analysis for Deep Earthquakes	96
8.1	Introduction	96
8.2	Descriptive Features and Empirical Properties	99
8.2.1	Data coverage and completeness	99
8.2.2	Epicentral and depth distributions	102
8.2.3	Magnitude distributions	107
8.2.4	Clustering properties	119
8.3	Models and model-fitting	123
8.3.1	Poisson, stress release and simple ETAS models	123

9	Deep Earthquakes Modeling	127
9.1	Application of MMPP to the Deep Earthquakes	127
9.2	Application of MMPP with Marks to Deep Earthquakes	141
9.3	Application of Multivariate MMPP to Deep Earthquakes	154
9.4	Evaluations of Estimation Errors	161
10	Concluding Remarks and Further Studies	167
10.1	Concluding Remarks	167
10.2	Questions and Future Studies	170

List of Tables

4.1	Standard deviation and mean of estimates for Q and Λ (1).	57
4.2	Standard deviation of estimates for Q and Λ (2).	57
7.1	Standard deviation of estimations.	91
8.1	The b-values of earthquakes in individual groupings.	117
8.2	Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 6$, $4.4 \leq m_J \leq 4.8$ and $0 < h < 15$ (km).	121
8.3	Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 6$, $4 \leq m_J \leq 4.4$ and $-15 < h < 15$ (km).	122
8.4	Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 6$, $4.4 \leq m_J \leq 4.8$ and $0 < h < 15$ (km) in an extended time period.	122
8.5	Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 5.2$, $4 \leq m_J \leq 5.2$ and $-90 < h < 90$ (km).	123

9.1	Estimated parameters of MMPPs for deep earthquakes.	136
9.2	Model selection between MMPP and Poisson model.	136
9.3	Order determination of MMPP.	137
9.4	Standard error of estimates.	138
9.5	Estimated parameters of MMPP with state-dependent marks for deep earthquakes.	142
9.6	Model selection between MMPP with state-dependent marks and MMPP with state-independent marks.	143
9.7	Model selection between the second order MMPP with marks and the third order MMPP with marks.	150
9.8	Model selection between bivariate MMPP and independent MMPPs (1).	156
9.9	Bootstrap Confidence Intervals for Parameters in Table 9.5.	166
9.10	Bootstrap estimation of the covariance matrix for the MMPP param- eters.	166

List of Figures

2.1	Map of New Zealand.	11
2.2	Conditional independence graph of hidden Markov model.	18
2.3	Conditional independence graph of non-homogeneous hidden Markov model.	23
4.1	A switching Poisson process.	41
4.2	Histograms of estimates (1).	55
4.3	Histograms of estimates (2).	56
4.4	Histograms of estimates (3).	58
7.1	Histograms of estimated parameters for simulated data from MMPP(1).	92
7.2	Histograms of estimated parameters for simulated data from MMPP(2).	93
8.1	Frequency, magnitude and cusum statistics ($M > 4$).	103
8.2	Frequency, magnitude and cusum statistics ($M > 4.5$).	104

8.3	Frequency, magnitude and cusum statistics ($M > 3.5$).	105
8.4	Epicenter distribution of main deep earthquakes in the North Island.	108
8.5	Section plot of deep earthquakes at azimuth 45 degrees.	109
8.6	Section plot of deep earthquakes at azimuth 135 degrees.	110
8.7	Section plot of deep earthquakes with latitude greater than -37 at azimuth 135 degrees.	111
8.8	Smoothed estimation of events frequencies (1).	112
8.9	Smoothed estimation of events frequencies (2).	113
8.10	Kernel density estimation of the depth distribution.	114
8.11	Histogram of the depth in main groupings.	115
8.12	Cusum magnitude plot for deep events around the North Island.	118
9.1	Partitions of main deep groupings.	129
9.2	Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for all groupings.	132
9.3	Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for Cook Strait grouping.	133
9.4	Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for Taranaki grouping.	134

9.5	Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for Taupo-BOP grouping.	135
9.6	K-S tests for the cumulative rescaled occurrence times for all groupings.	140
9.7	Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for all groupings.	146
9.8	Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for Cook Strait grouping.	147
9.9	Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for Taranaki grouping.	148
9.10	Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for Taupo-BOP grouping.	149
9.11	Doubly rescaled marked point process.	152
9.12	K-test for the doubly rescaled MMPP with marks.	153
9.13	The estimated probabilities of $X(t)$ in the state 2.	157
9.14	Taupo-BOP grouping fitted via a bivariate MMPP.	158
9.15	Taranaki grouping fitted via a bivariate MMPP.	159
9.16	Second moments of the bivariate MMPP.	160
9.17	Histograms of the Bootstrap Replicates (1).	162
9.18	Histograms of the Bootstrap Replicates (2).	163

9.19 Histograms of the Bootstrap Replicates (3).	164
9.20 Histograms of the Bootstrap Replicates (4).	165

Notations and Abbreviations

$A \triangleq B$	A is denoted by B
\mathbb{Z}	Space of integers
\mathcal{R}	The set of real number
$P(\cdot), P\{\cdot\}$	Probability measure or Probability distribution
\mathbf{P}	Transition probability matrix
$\mathbf{1}\{\cdot\}, \mathbf{1}_{\{\cdot\}}, \mathbf{1}(\cdot)$	Indicator function
$N(t)$	A counting process
\mathcal{N}	Space of positive integers
$\mathbf{1}$	A column vector with all entries being unity
e_i, \mathbf{e}_i	A unit column vector with ith entry being unity
$X_1^n, \{X\}_1^n$	(X_1, \dots, X_n)
$\#$	Cardinality of a set
i.i.d.	Independent and Identically distributed
$o(\delta)$	Higher order infinitesimal quantity of δ
$E\{\cdot\}, E(\cdot)$	Expectation value
$\ \cdot\ $	the norm of a Matrix, the largest absolute value of all entries
HMMs	Hidden Markov models
MMPP	Markov modulated Poisson process
MAP	Markov arrival process
BMAP	Batch Markov arrival process
BOP	Bay of Plenty, New Zealand
AIC	Akaike information criterion
BIC	Bayes information criterion

Part I

INTRODUCTION AND PRELIMINARIES

Chapter 1

Objectives and Motivations

Deep Earthquakes, which form a large portion of observations in New Zealand earthquake catalogue, are not as well studied as shallow earthquakes, particularly on the side of its statistical properties. It might be ascribed for several reasons. Firstly, their occurrence mechanism is not yet clear which complicates the attempt to propose a physically well-based model from a statistical point of view. Secondly, deep earthquakes are generally less destructive as shallow earthquakes. Hence, it is less urgent to call for thorough studies. Thirdly, unlike shallow earthquakes, most deep earthquakes occur only in subduction zones associated with ocean trenches which form only part of the seismic active zones around the world. So, they do not evoke as much attention as shallow earthquakes from geophysicists all over the world. However, the deep earthquakes are important in that they give indication of the structure of the earth, the dynamics of the crust and mantle and may have links with shallow earthquakes and volcano activities, see Frohlich (2006) for a compre-

hensive discussion of deep earthquakes.

Although statistics has a long history in applications to seismology as early as Omori's aftershock decay law (1894) and Gutenberg-Richter (1956) magnitude frequency relationship, it is only in recent decades that point process models have been suggested for the description of the occurrence patterns of earthquakes based on the catalogue data, see Vere-Jones (1970). The purpose of this thesis is an attempt to bridge the big gap existing between the catalogue observations of deep earthquakes and possible point process models. The existing models such as the ETAS model and the stress release model which are usually used for shallow earthquakes are also applied to the deep earthquakes. Their lack of fit strongly suggests that the occurrence pattern of the deep earthquakes is essentially different from that of the shallow earthquakes. To propose an appropriate model for deep earthquakes, exploratory data analysis is essential to understand the characteristics of deep earthquakes. After that, analytically tractable models are suggested for those characteristics in which we are interested. The main model we propose is the Markov modulated Poisson process (MMPP). Even though detailed patterns of deep earthquakes are probably beyond the scope of MMPP and the extensions given here to model, however, the present initial study should be sufficient to single out some of the main patterns, especially its time-varying activity characteristic in a relatively large time scale via MMPP and its variations.

This thesis is composed of three parts. The first part of the thesis is an introduction to hidden Markov models (HMMs) and their variations. The purpose of this part is to give a literature review of hidden Markov models and to illustrate

a variety of ways of associating hidden Markov models with covariates in practice. However, the literature review is not exhaustive in nature and tends to be brief as much as possible in the introductory part. Extension of the original HMMs for a more detailed description or a better fit to the real data is one of the main objectives of the thesis that author pursues.

Due to its openness and flexibility as one case of continuous time hidden Markov model, there also exist several different ways to extend the MMPP. This forms the main focus of the second part. One extension of MMPP is by associating each occurrence time of the observed process with a variable which could be spatial coordinate, magnitude or any other covariate of interest. Another extension is by simultaneously accommodating several observed point processes with an identical underlying Markov chain. Such a synchronization of multiple processes is potentially useful for modeling a multivariate process with weak dependence. The third extension is motivated by the attempt to describe small scale temporal clustering existing in the deep earthquakes by treating the set of recognized aftershocks as a mark which itself forms a finite point process. In this theoretical part, we also include some information theoretical aspects of MMPP such as the entropy rate of the underlying Markov chain, the entropy rate of the observed point process and the mutual information rate between them. Tentative remarks on the possible relationship between the mutual information rate and Fisher information of the parameters in MMPP are also included.

The third part of this thesis is on applications of these methods to the deep earthquakes in New Zealand. The motivation to use MMPP for deep earthquake

modeling is based on the fact that one of the apparent characters of deep seismicity is that it varies over time, active in one period, quiescent in another. Reasons and mechanisms behind this variation are not completely well-understood. Such a lack of interpretability in mechanism for a time-varying evolving system forms one key justification of applying "hidden" Markov models. Before suggesting any statistical models for New Zealand earthquakes, we first evaluate the data coverage, catalogue completeness and explore descriptive characteristics and empirical properties such as epicentral distributions, depth distributions and magnitude distributions. Clustering behavior is studied via second order moment analysis of point processes in the 8th chapter. We also apply the stress release models and ETAS models which are usually used for shallow earthquakes to the New Zealand deep earthquakes and provide tentative explanations of why they are not satisfactory for the deep earthquakes. The last chapter is on the applications of MMPP and its extensions to the deep earthquakes, and features applications of likelihood analysis via the EM algorithm, state smoothing, model selection and rescaling theorems of multivariate point process and marked point process for model evaluation, topics which are largely neglected in the original MMPP contexts.

Chapter 2

Introduction

2.1 Preliminaries of earthquakes

An **earthquake** is the result of a sudden release of energy in the Earth's crust that creates seismic waves. They are caused mostly by rupture of geological **faults** which are large fractures in Earth's crust, but also by volcanic activity, landslides, mine blasts, and nuclear experiments. An earthquake's point of initial rupture is called its focus or **hypocenter**. The term **epicenter** refers to the point at ground level directly above the hypocenter. After the 1906 San Francisco earthquake, a theory relating faulting and a tremor was enunciated by Reid's **elastic rebound theory**. It states that the crustal stresses, generally resulting from large scale regional crustal shearing motions, cause strain to accumulate in the immediate vicinity of faults. When the strain accumulation reaches a threshold imposed by the material properties of the rock and the fault surface, abrupt frictional sliding occurs, releasing

the accumulated strain energy.

To study earthquakes, **seismometers** are deployed to measure and record motions of the ground, including those of seismic waves generated by earthquakes, nuclear explosions, and other seismic sources. Almost all seismometers are based on damped inertial pendulum systems of one form or another. Modern instruments use electronic sensors, amplifiers, and recording instruments to record earthquake waves. Most of them are broadband covering a wide range of frequencies. However, in early postwar period, the seismometers deployed in New Zealand are only the **Wood-Anderson seismometers**, or the torsion seismographs which involve a copper cylinder attached to a vertical suspension wire. Shaking causes the cylinder to rotate slightly, moving a mirror that reflects a light signal to a photographic recorder. It is less sensitive and accurate as modern seismometers. Seismologists have employed a variety of methods to estimate the amount of energy release at the source of the earthquake. In 1935, Richter developed the magnitude scale by using seismic records to estimate the relative sizes of earthquakes. The **Richter scale** is based on the amplitude of the largest seismic waves (P, S, or surface wave). The comparative proportion of events number with different magnitude is determined by **Gutenberg-Richter law** (1956) which asserts that

$$\log_{10} N(M) \approx a - bM, \quad N(M) = \text{number of events with magnitude} \geq M.$$

It is often observed that after a major earthquake (**main shock**), there are many additional movements formed as the crust around the displaced fault plane adjusts to the effects of the main shock. The adjustments that follow a major earthquake

often generate smaller earthquakes called **aftershocks**. Typically, the frequency of occurrence of aftershocks decay rapidly. Omori (1894) studied aftershocks in Japan and developed an empirical formula for the aftershock activity. The modified **Omori's law** suggests that

$$n = \frac{k}{(t + c)^p},$$

where n is the frequencies of aftershocks at time t after mainshock and k, c, p are constants. The constant p called **p -value** is usually close to 1.0-1.4. If an aftershock is larger than the main shock, the aftershock is redesignated as the main shock and the original main shock is redesignated as a **foreshock**.

The majority of tectonic earthquakes originate in depths not exceeding tens of kilometers. Earthquakes occurring at a depth of less than 40 km are classified as 'shallow-focus' earthquakes or **shallow earthquakes**, while those with a focal-depth between 40 and 300 km are commonly termed 'mid-focus' or 'intermediate-depth' earthquakes. Deep-focus earthquakes may occur at much greater depths, ranging from 300 up to 700 kilometers. We call both the intermediate-depth and deep-focus earthquakes the **deep earthquakes**. Most of the deep earthquakes occur at boundaries of tectonic plates.

In plate tectonics, a **convergent plate boundary** is an actively deforming region where two (or more) tectonic plates or fragments of lithosphere move toward one another and collide. **Subduction** is the process that takes place at convergent boundaries by which one oceanic plate moves under another tectonic plate sinking into the Earth's mantle, forming **subduction zones**. Subduction zones mark sites of downwelling of the Earth's lithosphere. It is at subduction zones that the Earth's

lithosphere, oceanic crust, sedimentary layers, and some trapped water are recycled into the deep mantle. In the subduction process, the down-going **slab**, the leading edge of the subducting plate, is overridden by leading edge of the other plate. The slab begins to sink at increasing angle to the surface of the Earth. We call the deep active seismic area in a subduction zone the **Wadati-Benioff zone**. The deep earthquakes along the zone allow seismologists to map the three-dimensional surface of a subducting slab of oceanic crust and mantle since the angle of dip of the zone is the same as that of the subducting slab.

Wadati (1928) first convinced the scientific community that the deep earthquake occurs. Immediately after Wadati's paper, some research focused on confirming the existence of deep earthquakes and analyzing their property in other places besides Japan. Several researchers noticed that the association of deep earthquake, deep ocean trench and volcanos. As the data cumulated, it allows more thorough studies of the geometry of Wadati-Benioff zones. An excellent introduction of the deep earthquake in a global scale is given by Frohlich (2006). More conceptions about earthquakes at an entrance level are given by Bolt (1993).

The pattern of deep earthquake occurrence in New Zealand is determined by the tectonic structure, where there are distinct subduction zones. The first of these stretches along the east coast of the North Island from the Tonga-Kermadec trench, bends westwards underneath Cook Strait, and terminates approximately around the latitude of the Chatham Rise. In this zone, the Pacific Plate is subducting beneath the Australian Plate. In the second zone, however, the Australian Plate is subducting beneath the Pacific Plate. Both zones are associated with intermediate

earthquakes, see following Figure 2.1 (Reyners, 1989). In New Zealand map (Figure 2.1), heavy lines denote major faults. BOP and T denote Bay of Plenty and Lake Taupo, WI denotes White Island. Arrows indicate the velocity, in mm/year, of the Pacific plate relative to the Australian plate.

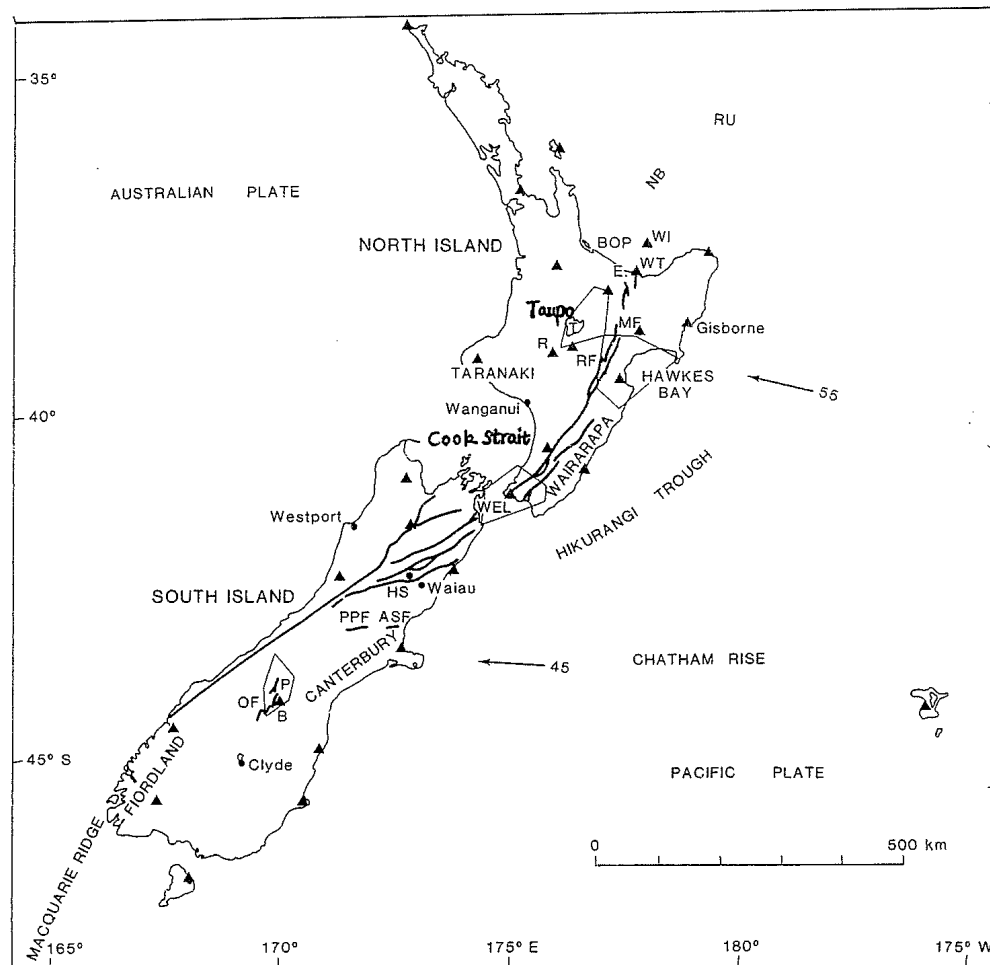


Figure 2.1: Tectonic setting of New Zealand.

2.2 Notations and Definitions

Definition 2.1 A **stochastic process** $\{X(t), t \in \mathcal{T}\}$ is a collection of jointly distributed random variables parameterized by an index t , where the set \mathcal{T} is called an index set.

It is customary to think of the index t as time. If \mathcal{T} is a countable set, then the process is called a **discrete-time stochastic process**. If \mathcal{T} is the whole real line, the process is called a **continuous-time stochastic process**. The discrete-time stochastic process is indexed by $X_i \triangleq X(i), i \in \mathcal{Z}$ throughout the thesis, where \mathcal{Z} is the integer set.

Definition 2.2 Let $\{X(t), t \in \mathcal{T}\}$ be a stochastic process, then for each $t \in \mathcal{T}$, $X(t)$ is called the state of the process at time t and the set of all possible values of $X(t)$ is called the **state space**, denoted by \mathcal{S} .

If \mathcal{S} is a countable set, then the state space of the process is called a **discrete state space**. Otherwise, the state space of the process is a **continuous state space**.

We call $\{X(t), t \in \mathcal{T}\}$ a **time-homogeneous** stochastic process if $P\{X(t+r) = j | X(r) = i\} = P\{X(t) = j | X(0) = i\}$, for all $t, r \geq 0$.

Definition 2.3 A discrete-time stochastic process $\{X_i, i = 0, 1, 2, \dots\}$ with discrete state space $\mathcal{S} = \{1, 2, \dots, r, \dots\}$ is called a **discrete-time Markov chain**, if for all $l, m \in \{0, 1, 2, \dots\}$ such that $l < m$ it satisfies the **Markov property**

$$P\{X_m = j | X_l = i, X_1^{l-1}\} = P\{X_m = j | X_l = i\}, \text{ for all } i, j \in \mathcal{S},$$

where $X_1^{l-1} = (X_1, \dots, X_{l-1})$.

Let $\{X_n, n \geq 0\}$ be a homogeneous discrete-time Markov chain. We call $\mathbf{P} = (p_{ij})$, $p_{ij} = P\{X_{n+1} = j | X_n = i\}$ the **transition probability matrix** of $\{X_n\}$.

Definition 2.4 A continuous-time stochastic process $\{X(t) : t \geq 0\}$ with discrete state space $\mathcal{S} = \{1, 2, \dots, r, \dots\}$ is called a **continuous-time Markov chain**, if for all $s > 0$ and $t \geq 0$ it satisfies the Markov property

$$P\{X(t+s) = j | X(t) = i, X(u), 0 \leq u < t\} = P\{X(t+s) = j | X(t) = i\} \text{ for all } i, j \in \mathcal{S}.$$

Let $\{X(t)\}$ be a homogeneous continuous-time Markov chain with state space $\mathcal{S} = \{1, 2, \dots\}$. Define $p_{ij}(h) = P(X(t+h) = j | X(t) = i)$. The **infinitesimal generator** of $\{X(t)\}$ is the matrix $Q = (q_{ij})$ such that for $i, j \in \mathcal{S}$,

$$q_{ij} \hat{=} \lim_{h \downarrow 0} \frac{p_{ij}(h) - \delta_i(j)}{h}$$

where $\delta_i(j) = 1$ if $i = j$ and $\delta_i(j) = 0$ if $i \neq j$. These limits exist when $\mathbf{P}(t)$ satisfies $\lim_{t \downarrow 0} \mathbf{P}(t) = I$, where I is the identity matrix, see Cinlar (1975), .

If the state space of a continuous-time or discrete-time Markov chain is finite, we call it a **finite Markov chain**.

2.3 Basic conceptions in Point processes

Point process can be viewed as stochastic processes with realizations consisting of collections of points, in which each of the points has a well specified position. A point process defined on one-dimensional space \mathcal{R} is described in several equivalent

ways by:

1. a non-decreasing integer-valued step function, indicating the cumulative number of points from a starting point.
2. sequences of intervals between points.
3. sequences of points.
4. **random counting measures** (random measures with unit mass at the points of the process).

For higher dimensional point process, there is no similar descriptions such as 1-3 because of lack of a natural generalization in higher dimensions. Description 4 is the only natural approach to generalize the definition of point process in higher dimensions by resorting to the random measure which has non-negative integer value.

Let $\hat{\mathcal{N}}_{\mathcal{S}}$ is the space of all boundedly finite integer-valued measures N called counting measure for short and $\hat{\mathcal{N}}_{\mathcal{S} \times \mathcal{K}}$ is the family of all boundedly finite counting measures defined on the product space $\mathcal{S} \times \mathcal{K}$, where \mathcal{K} is a c.s.m.s. (complete separable metric space) of marks, subject to the additional requirement that the **ground measure** N_g defined by $N_g(A) \triangleq N(A \times \mathcal{K})$ for all $A \in \mathcal{B}(\mathcal{S})$.

Definition 2.5 A **point process** N on a c.s.m.s. \mathcal{S} is a measurable mapping from a probability space (Ω, \mathcal{E}, P) into a measure space $(\hat{\mathcal{N}}_{\mathcal{S}}, \mathcal{B}(\hat{\mathcal{N}}_{\mathcal{S}}))$ where $\mathcal{B}(\hat{\mathcal{N}}_{\mathcal{S}})$ is the σ -field generated by all events of the form $\{N \in \hat{\mathcal{N}}_{\mathcal{S}} : N(A) \leq n\}$ for bounded $A \in \mathcal{B}(\mathcal{S})$. A point process N is **simple** when $P\{N \in \hat{\mathcal{N}}_{\mathcal{S}}\} = 1$. A **marked point process** on \mathcal{S} with marks in \mathcal{K} is a point process N on $\mathcal{B}(\mathcal{S} \times \mathcal{K})$ for which

$P\{N \in \hat{\mathcal{N}}_{\mathcal{S} \times \mathcal{K}}\} = 1$ and its **ground process** is given by $N_g(\cdot) \triangleq N(\cdot \times \mathcal{K})$.

Definition 2.6 Let $N(\cdot)$ be a point process adapted to the history \mathcal{F} , its **compensator** is the unique \mathcal{F} -predictable cumulative process A satisfying

$$E \left\{ \int_{\mathcal{R}^+} C(t, \omega) N(dt, \omega) \right\} = E \left\{ \int_{\mathcal{R}^+} C(t, \omega) A(dt, \omega) \right\},$$

for all nonnegative \mathcal{F} -predictable cumulative process C .

Alternatively, the counting process $N(t)$ is a submartingale with respect to $\{\mathcal{F}\}_{t \geq 0}$ and compensator can be realized as the predictable cumulative process in the **Doob-Meyer decomposition** of $N(\cdot)$ such as $N(t) = M(t) + A(t)$, where $\{M(t)\}_{t \geq 0}$ is a zero mean \mathcal{F} -martingale and $A(t)$ is the unique \mathcal{F} -predictable cumulative process.

Definition 2.7 The \mathcal{F} -**compensator** of an marked point process on \mathcal{S} with marks in \mathcal{K} is any **mark-predictable**, cumulative process $A(t, K, \omega)$ such that, for each $K \in \mathcal{B}(\mathcal{K})$, $A(t, K, \omega)$ is the \mathcal{F} -compensator for the **simple** point process $N_K(t) \triangleq N((0, t] \times K)$, where the mark-predictable σ -algebra is the product of the predictable σ -algebra $\Psi^{\mathcal{F}}$ with $\mathcal{B}(\mathcal{K})$.

Definition 2.8 (a) Let N be a point process adapted to the history \mathcal{F} . An \mathcal{F} -**intensity** for N is any \mathcal{F} -adapted process $\lambda(t, \omega)$, measurable with respect to the product σ -algebra $\mathcal{B}(\mathcal{R}_0^+) \times \mathcal{E}$, and such that a.s. for all t ,

$$A(t, \omega) = \int_0^t \lambda(u, \omega) du.$$

(b) Let N be a marked point process, with mark space $(\mathcal{K}, \mathcal{B}(\mathcal{K}))$ equipped with a reference measure $l_{\mathcal{K}}$ and \mathcal{F} -compensator A . An \mathcal{F} -**intensity** for N is any \mathcal{F} -

adapted process $\lambda(t, k, \omega)$, measurable with respect to the three-fold product σ -algebra $\mathcal{B}(\mathcal{R}_0^+) \times \mathcal{K} \times \mathcal{E}$, and such that a.s. for all t and $K \in \mathcal{B}(\mathcal{K})$,

$$A(t, k, \omega) = \int_{(0,t] \times K} \lambda(u, k, \omega) \, du \, l_{\mathcal{K}}(dk).$$

For one-dimensional point process defined on \mathcal{R}_0^+ , we denote the number of points of $N(\cdot)$ in $[s, t]$ by $N(s, t)$ and denote $N(0, t)$ by $N(t)$ for short. The definition of conditional intensity function suggests its interpretation as the instantaneous occurrence rate of events at time t conditioned on the history \mathcal{F}_t of the point process $N(t)$ such as $\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{E\{N(t, t+\Delta t)\}}{\Delta t}$ or equivalently $\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{N(t, t+\Delta t) > 0\}}{\Delta t}$. Given a realization (t_1, t_2, \dots, t_n) of the point process $N(\cdot)$ on $[0, T]$ with conditional intensity function $\lambda_{\theta}(t)$, the corresponding **likelihood function** is written by

$$L(\theta) = \prod_{i=1}^n \lambda_{\theta}(t_i) \exp \left\{ - \int_0^T \lambda_{\theta}(t) \, dt \right\}.$$

More about the general theory of the point process can be found in Daley and Vere-Jones (2003, 2008), Karr (1991) and Bremaud (1980).

2.4 Hidden Markov Models in Discrete Time

2.4.1 The Basic Hidden Markov Model

The last couple of decades, have seen an extensive application of hidden Markov models (HMM) in speech processing, bioscience, finance, hydrology, climatology and others. It was first introduced by Baum and others through a series of papers

to exploit related statistical inference procedures and algorithms. Since then, HMM has evoked significant attention in signal processing and speech recognition. An excellent introduction to these methods and applications is given by Rabiner (1989). Boosted by the development of computation technology and statistical computation methods, particularly the EM algorithm, HMM now receives significant attention in many fields. Several books have been published on its various extensions and applications, among them, the books written by MacDonald and Zucchini (1997) and Elliot et al. (1995) demonstrate different approaches in dealing with HMM modelling.

Definition and notation Generally speaking, the HMM is a discrete-time stochastic process $\{(s_k, x_k)\}_{k=1}^n$ which satisfies several assumptions. It is assumed that $\{s_k\}$ forms an irreducible homogeneous discrete-time finite **Markov chain** on the state space $\mathcal{S} = \{1, \dots, r\}$ with transition probability matrix $\mathbf{P} = (p_{ij})_{r \times r}$, $p_{ij} = P(s_{k+1} = j | s_k = i)$ and initial distribution vector $\pi = (\pi_1, \dots, \pi_r)$. Given the process $\{s_k\}$, $\{x_k\}$ are **conditionally independent** random variables whose probability distributions $f(x_k | s_k, \theta_{s_k})$ with parameter θ_{s_k} are dependent only on the current state s_k . Later on, we refer to $f(x_k | s_k, \theta_{s_k})$ as **state-dependent distribution**. Since $\{s_k\}$ forms an irreducible homogeneous discrete-time finite Markov chain, there exists a unique, strictly positive, stationary distribution. The initial distribution is usually taken to be equal to this stationary distribution.

The model assumptions are demonstrated according to the conditional independence graph. In such a graph the absence of an edge between two vertices indicates

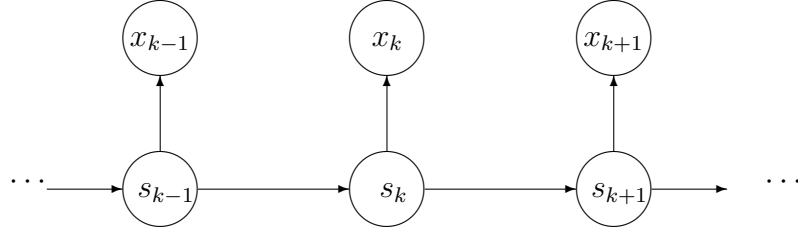


Figure 2.2: Conditional independence graph of hidden Markov model.

the two variables are independent given the other variables. Figure 2.2 demonstrates the conditional independence of $\{x_k\}$ conditioned on $\{s_k\}$ and the conditional independence of s_{k-1} and s_{k+1} given s_k , i.e. the Markov property.

In the HMM context, $\{s_k\}$ is called the state process or Markov regime and the name HMM is motivated by the assumption that $\{s_k\}$ is not observable, so the inference is instead based only on the information of $\{x_k\}$. Once the conditional distribution of x_k given s_k is specified parametrically, the observations are generated according to the state-dependent probability distribution $f(x_k|s_k, \theta_{s_k})$ whose parameters vary according to the underlying Markov chain. Inference of an HMM is typically likelihood-based as the likelihood can be written as

$$L(\pi, \mathbf{P}, \theta) = \sum_{s_1=1}^r \cdots \sum_{s_n=1}^r \pi_{s_1} f(x_1|s_1, \theta_{s_1}) \cdots p_{s_{n-1}s_n} f(x_n|s_n, \theta_{s_n}) \quad (2.1)$$

where $\theta = (\theta_1, \dots, \theta_r)$ and the summation is over n^r permutations. It is worth noting that the number of terms increase exponentially fast with respect to the order r of

the underlying Markov chain. To avoid such a direct likelihood evaluation, which tends to be intractable even for a moderate r , a more efficient algorithm is introduced in the parameter estimation procedures by Baum et al. (1970).

In the standard HMM contexts, e.g. Baum et al. (1970) and Rabiner (1989), the **forward probability** $\alpha_t(i)$ determines the distribution of the current state $s_t = i$ jointly with the observations from time 0 up to t . It can be written by

$$\alpha_t(i) = \sum_{s_1, \dots, s_{t-1}} \pi_{s_1} f(x_1 | s_1, \theta_{s_1}) \prod_{k=2}^t \{p_{s_{k-1}s_k} f(x_k | s_k, \theta_{s_k})\} \mathbf{1}\{s_t = i\}, \quad (2.2)$$

where $\mathbf{1}\{.\}$ is the indication function. Correspondingly, the **backward probability** $\beta_t(.)$ defines the probability of observations from $t+1$ to the end conditional on the Markov chain sojourn in a given state, i.e. j at t ; then it can be written by

$$\beta_t(j) = \sum_{s_{t+1}, \dots, s_n} \mathbf{1}\{s_t = j\} \prod_{k=t+1}^n p_{s_{k-1}s_k} f(x_k | s_k, \theta_{s_k}). \quad (2.3)$$

From this device, the likelihood is obviously written by $L = \sum_{i=1}^r \alpha_t(i) \beta_t(i)$. Note that the forward and backward probabilities can be updated recursively, hence the likelihood is evaluated in a much more efficient fashion.

Since the log-likelihood function can be evaluated routinely, even for long sequences of observations, it is feasible to perform parameter estimation by direct numerical maximization. Parameter inference through EM algorithm and its variations, Newton type algorithms or Monte Carlo optimization methods are discussed in detail by Cappé et al. (2005). Here, we only focus on the Baum-Welch algorithm (Baum et al., 1970), which is an early version of the EM algorithm (Dempster et al., 1977). The EM algorithm is an iterative method for obtaining the maximum likelihood estimation in incomplete data problems which has several appealing properties

in comparison to other methods such as Newton-Raphson or quasi-Newton methods. It is easily implemented since it only involves two steps in each iteration: 1) taking expectations of the complete data likelihood with respect to missing data, 2) maximizing the conditional expectation which in many cases is in simple closed form, particularly for the exponential family. Furthermore, it is numerically stable in that each iteration returns an improved likelihood and convergence is nearly always to a local maximum.

Let $\log L^c(\theta'|\mathbf{X}_1^n, \mathbf{Y}_1^m)$ denotes the complete data log-likelihood, in which $\mathbf{X}_1^n = (X_1, \dots, X_n)$ and $\mathbf{Y}_1^m = (Y_1, \dots, Y_m)$ are observations and missing data respectively and θ' is the parameter vector to be estimated. The E-step requires taking expectation of the complete data log-likelihood given the observations such that

$$Q(\theta'|\theta) = E_\theta \left(\log L^c(\theta'|\mathbf{X}_1^n, \mathbf{Y}_1^m) \middle| \mathbf{X}_1^n \right) \quad (2.4)$$

which gives the best predictor of $\log L^c(\theta'|\mathbf{X}_1^n, \mathbf{Y}_1^m)$ from \mathbf{X}_1^n in the sense of mean square error. The M-step follows by maximizing $Q(\theta'|\theta)$ with respect to θ' . Then the iteration steps are repeated until some convergence criterion is met.

Although it is possible to directly maximize the likelihood, it is often preferred to obtain the parameter estimates via the Baum-Welch algorithm (Baum et al., 1970), due to some of its appealing properties. Treating the hidden state sequence $\{s_k\}$ as missing data \mathbf{Y}_1^m , the complete likelihood for observations \mathbf{X}_1^n can be written as

$$L^c(\pi, \mathbf{P}, \theta | s_1^n, x_1^n) = \pi_{s_1} f(x_1 | s_1, \theta_{s_1}) \prod_{k=2}^n p_{s_{k-1}s_k} f(x_k | s_k, \theta_{s_k}). \quad (2.5)$$

After taking logarithms and executing the EM steps, the parameters are updated

by

$$\hat{\pi}_i = \gamma_1(i), \quad (2.6)$$

$$\hat{p}_{ij} = \frac{\sum_{m=1}^{n-1} \xi_m(i, j)}{\sum_{m=1}^{n-1} \gamma_m(i)}, \quad (2.7)$$

$$\hat{f}(\omega_j|i) = \frac{\sum_{m=1}^n \gamma_m(i) \delta_{\omega_j}(x_m)}{\sum_{m=1}^n \gamma_m(i)}, \quad (2.8)$$

where $\xi_m(i, j) = P(s_m = i, s_{m+1} = j | \mathbf{X}_1^n)$, $\gamma_m(i) = \sum_{j=1}^r \xi_m(i, j)$ and δ is the Kronecker delta function, see Baum et al. (1970). A comprehensive review of the hidden Markov process is given by Ephraim and Merhav (2002).

2.4.2 Hidden Markov model with covariates

Due to the flexibility and universality of the framework provided by hidden Markov models, there exists a variety of extensions adapted to particular applications. One type of modification for the hidden Markov model is based on variations in the underlying Markov process. For example, the Markov chain could be treated as a higher order Markov chain, allowing second order or higher order Markov dependence. Or the underlying Markov process could be a semi-Markov process or Markov renewal process, see definition 3.2. In this case, each state of the finite Markov chain is associated with a given sojourn time distribution once the Markov chain enters that state. Another example of this type of extension is based on relaxation of some assumptions on the underlying Markov chain. For example, the underlying Markov chain might be assumed non-stationary or non-homogeneous. Another type

of extension is based on variations in the state-dependent distribution. We will mainly illustrate some of the first type of extensions in the following subsections and illustrations.

Non-homogeneous Hidden Markov model

By dropping the assumption that the Markov chain is homogeneous, a time-dependent underlying Markov chain opens the possibility of incorporating time trend and seasonality in HMMs. Such an extension of HMMs has potential value for a better understanding of the underlying state process when the underlying Markov chain is associated with one or more covariates which pose measurement effects on the state transition probabilities. Hughes and Guttorp (1999) apply this type of model in meteorology. In the model, the weather variables c_t are treated as covariates and $x_t = \{x_t^1, \dots, x_t^m\}$ are multivariate observations for the rainfall occurrences at a network of m sites with observed value $x_t^i = 1$ if rain occurs on day t at station i and $x_t^i = 0$ otherwise. The model assumptions are given such as:

$$p(x_t | s_1^t, x_1^{t-1}, c_1^n) = p(x_t | s_t), \quad (2.9)$$

$$p(s_t | s_1^{t-1}, c_1^n) = p(s_t | s_{t-1}, c_t). \quad (2.10)$$

The first assumption requires that the distribution of rainfall depends only on current atmosphere state, i.e, the rainfall record is conditionally independent given the underlying Markov states. The second assumption requires that the transition probability of the current state depends not only on the previous state but also on the current weather measurements. It is this assumption that allows a

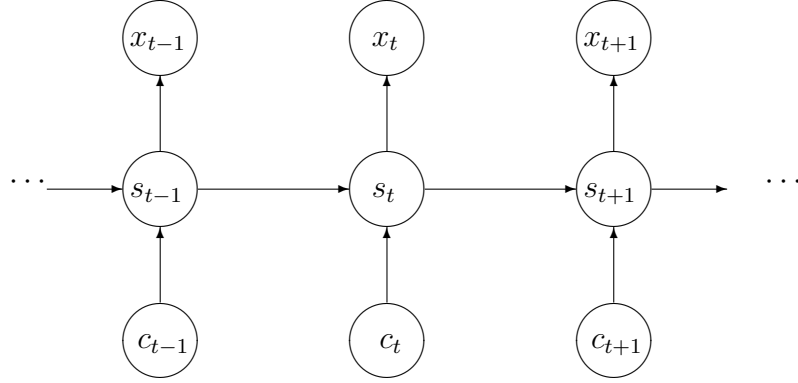


Figure 2.3: Conditional independence graph of non-homogeneous hidden Markov model.

non-homogeneous underlying Markov chain whose transition probability varies according to the weather measurements. These assumptions can be summarized by a conditional independence graph (Figure 2.3).

After parameterization of (2.9) and (2.10), the likelihood can be written as

$$L(x_1^n) = \sum_{s_1} \cdots \sum_{s_n} p(s_1|c_1) \prod_{t=1}^n p(s_t|s_{t-1}, c_t) p(x_t|s_t). \quad (2.11)$$

Maximum likelihood estimates for the model are obtained typically via EM algorithm by treating s_1^n as 'missing data'. However, no explicit M-step is available in general.

Another similar extension is to assume the state transition probability depends on the covariates $\{c_t\}$ through logistic regression, see Albert (1991), Wang and

Puterman (2001). For instance, define a two-state Markov chain by

$$p(s_t = 1 | s_{t-1} = 1) = p_{11}(c_t, \beta_1) = \text{expit}(\beta_1 c_t), \quad (2.12)$$

$$p(s_t = 2 | s_{t-1} = 1) = p_{21}(c_t, \beta_1) = 1 - p_{11}(c_t, \beta_1), \quad (2.13)$$

$$p(s_t = 2 | s_{t-1} = 2) = p_{22}(c_t, \beta_2) = \text{expit}(\beta_2 c_t), \quad (2.14)$$

$$p(s_t = 1 | s_{t-1} = 2) = p_{12}(c_t, \beta_2) = 1 - p_{22}(c_t, \beta_2), \quad (2.15)$$

where $\text{expit}(\beta c) = \frac{\exp\{\beta c\}}{1 + \exp\{\beta c\}}$.

The parameter estimation is carried out by numerical maximization of the likelihood or the EM algorithm, though in general without an explicit M-step.

HMM with state-dependent distribution depending on covariates

Similar extensions are obtained by incorporating covariates in the state-dependent distributions. For a hidden Markov model with Poisson observations, the Poisson rate parameter can be associated with covariates by a log-linear model such as $\log \lambda_i = \beta_i c_t'$, where c_t is a row vector of covariates. For a hidden Markov model with binomial or multinomial observations, the binomial probability can be similarly associated with covariates through a logit link function such as $\text{logit} p_i = \beta_i c_t'$. Parameter estimation is implemented by standard EM iteration steps.

2.4.3 Hidden semi-Markov models

By presuming the underlying process is a semi-Markov process, it yields a more general model, the so-called hidden semi-Markov model. Briefly speaking, in a hidden

semi-Markov model, each state of the Markov chain is associated with a holding time distribution and self-transition in the Markov transition matrix is forbidden, i.e., all diagonal elements in the transition probability matrix are zeros. Such an extension is important for a better understanding of different time scale stochastic behavior of a process and has important applications in meteorology modeling, see Sansom and Thompson (2001). In this scenario, the missing data include not only the state sequence $\{s_k\}$ but also the holding times in the states. The complete data log-likelihood thus can be written as:

$$\log \mathbf{L}^c = \log \pi_{s_1} + \sum_{k=2}^n \log p(s_k | s_{k-1}) + \sum_{k=1}^n \log p(D_k | s_k) + \sum_{k=1}^n \log f(x_k | s_k), \quad (2.16)$$

where D_k is the sojourn time of the underlying process in state s_k and $p(D_k | s_k)$ is the conditional density of the state holding time. After specifying the parametric distribution of the holding time and the state-dependent distribution, it is possible to obtain explicit EM iteration step, particularly for those distributions from the exponential family.

Chapter 3

Markov Modulated Poisson Process (MMPP)

3.1 Introduction

MMPP is particularly useful in modeling time-varying intensity rate processes such as traffic flows of communication networks, internet traffic flows and queuing systems. A collection of properties of ordinary MMPP is given in Fischer and Meier-Hellstern (1993). The parameter estimation for MMPP through the EM algorithm and its comparison with the downhill simplex algorithm is addressed by Rydén (1996a). Jensen (2005) addresses the likelihood process of MMPP with discrete type marks.

In this section, we define the Markov modulated Poisson process (MMPP) and

treat it as a Markov renewal process. Then the likelihood of a MMPP is derived.

Definition 3.1 Markov modulated Poisson process (MMPP) is a doubly stochastic Poisson process whose stochastic intensity rate is given by $\lambda_{X(t)}$ where $X(t)$ is a r -state irreducible homogeneous Markov process.

For a MMPP, the underlying Markov process $X(t)$ is characterized by the infinitesimal generator matrix Q where its (i, i) th element satisfies $-q_i \hat{=} q_{ii} = -\sum_{j:j \neq i} q_{ij}$, $q_{ij} > 0$ and r is the order of the Markov chain. Given $X(t)$, the observed point process is a Poisson process with intensity rate $\lambda_{X(t)}$.

Another approach to specify a MMPP is through its inter-arrival times and the associated states of the underlying Markov chain at arrivals. We use **events** or **arrivals** exchangeably throughout the discussion. It is this specification of MMPP that we will focus on for the derivation of the likelihood of a MMPP. Before this, we should review the definition of Markov renewal process.

For each $n \in \mathcal{N}$ (\mathcal{N} being the set of positive integers), define a random variable X_n taking values in a countable set \mathcal{S} and a random variable Y_n taking values in $[0, \infty)$ with $Y_0 = 0$.

Definition 3.2 The stochastic process $\{(X_n, Y_n), n \geq 0\}$ is a **Markov renewal process** with state space \mathcal{S} provided that $Pr\{X_{n+1} = j, Y_{n+1} \leq y | X_0, \dots, X_n, Y_0, \dots, Y_n\} = Pr\{X_{n+1} = j, Y_{n+1} \leq y | X_n\}$ for all $n \in \mathcal{N}$, $j \in \mathcal{S}$ and $y \in [0, \infty)$.

Note that the observed point process can be specified equivalently by occurrence times $\{t_i\}_{i=1}^n$ or inter-event times $Y_i = t_i - t_{i-1}$ between arrivals. The sequence $\{(X_k, Y_k), 0 \leq k \leq n, Y_0 = 0\}$ combining the embedded Markov chain $X_k = X(t_k)$

and the corresponding inter-event times Y_k forms a Markov renewal process. To clarify this, we consider the transition probability without arrivals $G_{ij}(t) \triangleq P\{X(t) = j, N(t) = 0 | X(0) = i\}$ where $N(t)$ is the counting process of MMPP from 0 up to t . For a small increment Δt ,

$$G_{ij}(t + \Delta t) = G_{ij}(t)(1 - q_j\Delta t - \lambda_j\Delta t) + \sum_{k \neq j} G_{ik}(t)q_{kj}\Delta t + o(\Delta t).$$

We obtain **Chapman-Kolmogorov** differential equations such as

$$\begin{cases} G'_{ij}(t) = G_{ij}(t)(-q_j - \lambda_j) + \sum_{k \neq j} G_{ik}(t)q_{kj} \\ G_{ij}(0) = \delta_{ij}, \end{cases}$$

which yields $G(t) = \exp\{(Q - \Lambda)t\}$ according to the definition of the matrix exponential in this chapter appendix, where $\Lambda \triangleq \text{diag}(\lambda_1, \dots, \lambda_r)$. The transition probability matrix of the sequence $\{(X_k, Y_k), 0 \leq k \leq n, Y_0 = 0\}$ is then given by $\mathbf{F}(y) = \int_0^y G(u)\Lambda du = \int_0^y \exp\{(Q - \Lambda)u\}\Lambda du$, for $y > 0$. It satisfies

$$\begin{aligned} \mathbf{F}(y) &= \int_0^y \exp\{(Q - \Lambda)t\}\Lambda dt \\ &= \{I - e^{(Q - \Lambda)y}\}(\Lambda - Q)^{-1}\Lambda \\ &= \{I - e^{(Q - \Lambda)y}\}\mathbf{F}(\infty), \end{aligned} \tag{3.1}$$

whose (i, j) th element is the probability $P\{X_k = j, Y_k \leq y | X_{k-1} = i\}$, see Fischer and Meier-Hellstern (1992) and refer to the properties of matrix exponential in the chapter appendix. The distribution of the inter-event times depends not only on the current state that the underlying Markov chain sojourn in but also on the previous state. Hence, the sequence $\{(X_k, Y_k), 0 \leq k \leq n, Y_0 = 0\}$ forms a Markov renewal sequence. The matrix $\mathbf{F}(\infty) \triangleq (\Lambda - Q)^{-1}\Lambda = P\{X_k = j | X_{k-1} = i\}$ in (3.1) is the

transition probability matrix of the Markov chain embedded at arrival times. If the initial distribution is chosen according to the stationary vector of transition matrix $\mathbf{P} = \mathbf{F}(\infty)$ which satisfies $\pi\mathbf{P} = \pi$, then the MMPP starts at an arbitrary arrival time. In this case, we call the MMPP interval-stationary. There exists another version of stationary Markov modulated Poisson process in a sense that the counting process itself is stationary when the initial distribution is selected according to the stationary vector of the underlying Markov chain satisfying $\pi Q = 0$. In this case, we call the MMPP environment-stationary, see Fischer and Meier-Hellstern (1992). We characterize a MMPP by its parameters (π, Q, Λ) . From the above discussions, the likelihood of a MMPP is given in the following theorem.

Theorem 3.1 For a MMPP (π, Q, Λ) , given the observed arrivals $0 = t_0 \leq t_1 \leq \dots \leq t_n = T$ or their corresponding inter-event times $Y_i = t_i - t_{i-1}, i = 1, \dots, n$, the likelihood is

$$L(\pi, Q, \Lambda) = \pi \exp\{(Q - \Lambda)Y_1\}\Lambda \cdots \exp\{(Q - \Lambda)Y_n\}\Lambda \mathbf{1}, \quad (3.2)$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$, $\exp\{(Q - \Lambda)t\}\Lambda$ is the transition density matrix of the Markov renewal sequence $(X_i, Y_i), i = 1, \dots, n$ and $\mathbf{1}$ is a $r \times 1$ column vector with all entries being unity.

3.2 Parameter Estimation

Parameter estimation approaches are roughly classified into two categories. One of them is likelihood-based method and another approach is moment-based. In this

section, we briefly review the likelihood-based method.

The main restriction of utilizing Newton type algorithms to obtain MLE is that it requires differentiation with respect to parameters from a product of matrices. Ramesh (1995) uses the downhill simplex optimization algorithm in MMPP to avoid the derivatives of the product of matrix exponential. Similar to discrete time hidden Markov models (HMMs) in which so called Baum-Welch iteration method, an earlier version of general EM algorithm is applied to obtain the MLE since the parameter estimation of MMPP is a missing data problem in nature with missing data formed by the underlying Markov process. Deng and Mark (1993) utilize a discrete Markov chain and approximate MMPP by a HMM with Poisson observations, assuming the state transitions occur only at the boundaries of divided time intervals.

As Asmussen et al. (1996) and Rydén (1996a) pointed out, the key point of utilizing EM algorithm is to consider the whole trajectory of the underlying Markov chain as missing data. Assume the Markov chain has jumps at $0 < u_1 < \dots < u_m < T$, denote $\Delta u_i = u_i - u_{i-1}$ where $u_0 = 0, u_{m+1} = T$ and let $s_k = X(u_k)$, the complete likelihood can then be written by

$$\begin{aligned} L^c(\pi, Q, \Lambda) &= \pi_{s_1} \left\{ \prod_{k=1}^m q_{s_k} e^{-q_{s_k} \Delta u_k} \times \frac{q_{s_k, s_{k+1}}}{q_{s_k}} \right\} e^{-q_{s_{m+1}} \Delta u_{m+1}} \\ &\times \left\{ \prod_{k=1}^{m+1} \frac{(\lambda_{s_k} \Delta u_k)^{z_k}}{z_k!} e^{-\lambda_{s_k} \Delta u_k} \times \frac{z_k!}{(\Delta u_k)^{z_k}} \right\}, \end{aligned}$$

where z_k is the number of arrivals within $[u_{k-1}, u_k]$. In this equation, the first term of the product is the likelihood of the underlying Markov chain and the second term gives the conditional likelihood of the number of arrivals and their locations which are uniformly distributed order statistics over $[u_{k-1}, u_k], i = 1, \dots, m+1$ conditioned

on the trajectory of the underlying Markov chain. Note that the complete likelihood is in a form of scalar product which is much easier for numerical evaluation. After taking logarithms, the complete likelihood is summarized by

$$\log L^c(\pi, Q, \Lambda) = \sum_{i=1}^r \left\{ \log \pi_i \mathbf{1}\{X(0) = i\} - (q_i + \lambda_i)T_i + \sum_{j:j \neq i} N_{ij} \log q_{ij} + N_i \log \lambda_i \right\}$$

, where $T_i \triangleq \int_0^T \mathbf{1}\{X(t) = i\} dt$ is the sojourn time of $X(t)$ in state i , $N_{i,j} \triangleq \#\{t : 0 < t \leq T, X(t-) = i, X(t) = j\}$ is the transition times of $X(t)$ from state i to state j and $N_i \triangleq \int_0^T \mathbf{1}\{X(t) = i\} dN(t)$ is the arrival times upon $X(t)$ in state i . They form a group of sufficient statistics of the parameters (Q, Λ) . The E-step only requires the conditional expectations of the sufficient statistics (T_i, N_{ij}, N_i) given the internal history \mathcal{F}_t of the observed point process since the complete likelihood only depends on the sufficient statistics (T_i, N_{ij}, N_i) . To facilitate the evaluation of the likelihood and other statistics in a MMPP much as in the context of a discrete time HMM, we introduce the forward and backward probabilities $\alpha_t(i) = \pi L_1 \cdots L_{N(t)} \exp\{(Q - \Lambda)(t - t_{N(t)})\} \mathbf{e}_i$ and $\beta_t(j) = \mathbf{e}'_j \exp\{(Q - \Lambda)(t_{N(t)+1} - t)\} \Lambda L_{N(t)+2} \cdots L_n \mathbf{1}$ where $L_k = \exp\{(Q - \Lambda)Y_k\} \Lambda$. By using the above notations, it turns out that the conditional expectation of the sufficient statistics (N_i, N_{ij}, T_i) given \mathcal{F}_t can be written as

$$N_i^* = E\{N_i | \mathcal{F}_T\} = \sum_{k=1}^n \frac{\alpha_{t_k}(i) \beta_{t_k}(i)}{\sum_{j=1}^m \alpha_t(j) \beta_t(j)}, \quad (3.3)$$

$$T_i^* = E\{T_i | \mathcal{F}_T\} = \int_0^T \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^m \alpha_t(j) \beta_t(j)} dt, \quad (3.4)$$

$$N_{ij}^* = E\{N_{ij} | \mathcal{F}_T\} = \int_0^T \frac{\alpha_t(i) q_{ij} \beta_t(j)}{\sum_{j=1}^m \alpha_t(j) \beta_t(j)} dt, \quad (3.5)$$

see Asmussen et al. (1996). After an explicit M-step, the parameter are updated by

$$\hat{q}_{ij} = \frac{N_{ij}^*}{T_i^*} (i \neq j), \quad \hat{\lambda}_i = \frac{N_i^*}{T_i^*}. \quad (3.6)$$

3.3 Statistical Inference and Model Evaluation

Given a realization of a MMPP, the probability of the underlying process in a particular state at a specific time $P\{X(t) = i | \mathcal{F}_T\}$ is estimated by

$$\hat{p}_t(i) \hat{=} E\{\mathbf{1}(X(t) = i) | \mathcal{F}_T\} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^r \alpha_t(j)\beta_t(j)}, \quad (3.7)$$

see also equation (3.3) or Asmussen et al. (1996) for the derivation of this equation. Clearly, $\hat{p}_t(i)$ in (3.7) is continuous with respect to t . So by evaluating (3.7) at many points taken within $[0, T]$ and joining them together by lines, it is sufficient to mimic the evolution of the latent Markov process. In order to simulate, evaluate and forecast based on the model, one needs an estimate of the conditional intensity rates $\lambda_{X(t)}$ which, as suggested above, is given by

$$\hat{\lambda}(t) \hat{=} E\left\{\sum_{i=1}^r \lambda_i \mathbf{1}\{X(t) = i\} | \mathcal{F}_T\right\} = \sum_{i=1}^r \lambda_i \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^r \alpha_t(j)\beta_t(j)}. \quad (3.8)$$

Papangelou (1972) points out that by rescaling point process from $\{t_i\}$ to the compensator $\{\int_0^{t_i} \lambda(t) dt\}$, we obtain a stationary Poisson process with unit rate. The collection of the rescaled points which are supposed to be a unit rate Poisson process when the proposed model is a real one are so-called residual point process (Ogata, 1988). By this conception, a deviation of rescaled points from unit rate

Poisson process may suggest some deficiencies of the proposed model and even illuminate some improvements in some cases. Numerical evaluation of the estimated compensator calls for the integration of matrix exponential which can be implemented by matrix eigenvalue decomposition (Rydén, 1996a), Poisson randomization (Klemm et al., 2003) or some matrix exponential in higher order (Van Loan, 1978). Then standard Kolmogorov-Smirnov statistics can be applied to verify whether or not the cumulative distribution of the transformed occurrence times forms a uniform distribution.

3.4 Second-Order Moment Properties

Ramesh (1995) studies the moment properties of MMPP. The moment property of observed point process are studied through that of underlying Markov chain. Note that the intensity rate process is a Markov process, and when all possible intensity rates are different, we have

$$\begin{aligned}
 E\{\lambda(0)\lambda(t)\} &= \sum_{i=1}^r P\left\{\lambda(0) = \lambda_i\right\} \lambda_i E\left\{\lambda(t) | \lambda(0) = \lambda_i\right\} \\
 &= \sum_{i=1}^r \pi_i \lambda_i \sum_{j=1}^r p_{ij}(t) \lambda_j \\
 &= \pi \Lambda p(t) \Lambda \mathbf{1}
 \end{aligned} \tag{3.9}$$

where $p(t) = e^{Qt}$. From the relationship $E\{\lambda(t)\} = E\{\frac{dN(t)}{dt}\}$ and $E\{\lambda(t_1)\lambda(t_2)\} = E\{\frac{dN(t_1)dN(t_2)}{dt_1dt_2}\}$ etc., the **auto-covariance density** of the MMPP is

$$\begin{aligned}
 c(t) &= \lim_{\Delta t_1, \Delta t_2 \rightarrow 0+} Cov\left\{N(0, \Delta t_1), N(t, t + \Delta t_2)\right\} / \Delta t_1 \Delta t_2 \\
 &= E\left\{\lambda(0)\lambda(t)\right\} - E\left\{\lambda(0)\right\}E\left\{\lambda(t)\right\} \\
 &= \pi \Lambda p(t) \Lambda \mathbf{1} - \pi \Lambda \mathbf{1} \pi \Lambda \mathbf{1} \\
 &= \pi \Lambda \left\{p(t) - \mathbf{1} \pi\right\} \Lambda \mathbf{1}.
 \end{aligned} \tag{3.10}$$

For a stationary point process, an important second moment associated with the conditional intensity rate given an arbitrary arrival epoch at origin is the so called **Palm intensity** defined as $h(t) = \lim_{\Delta t_1, \Delta t_2 \rightarrow 0+} Pr\{N(t, t + \Delta t_2) > 0 | N[0, \Delta t_1] > 0\} / \Delta t_2$. The Palm intensity of a MMPP is given by

$$\begin{aligned}
 h(t) &= \lim_{\Delta t_1, \Delta t_2 \rightarrow 0+} Pr\left\{N(t, t + \Delta t_2) > 0 \middle| N[0, \Delta t_1] > 0\right\} / \Delta t_2 \\
 &= \frac{\pi \Lambda p(t) \Lambda \mathbf{1}}{\pi \Lambda \mathbf{1}}.
 \end{aligned} \tag{3.11}$$

In the theory of stationary time series, it is valuable to consider frequency domain analysis such as the power spectrum. We therefore consider the spectra corresponding to the analysis of counting properties, see Bartlett (1963, 1964). For a stationary point process $N(t)$, its **counting spectral density** is defined by the Fourier transform of the covariance density $c(t)$ of counts such as $g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c(u) e^{-i\omega u} du$. The counting spectral density of the MMPP is formulated by

$$g(\omega) = \frac{1}{\pi} \left\{ \pi \Lambda \mathbf{1} + \pi \Lambda P^*(i\omega) \Lambda \mathbf{1} + \pi \Lambda P^*(-i\omega) \Lambda \mathbf{1} \right\}, \omega \geq 0, \tag{3.12}$$

where $P^*(t)$ is the Laplace transform of the matrix $p(t)$. For a second order MMPP, the auto-covariance, the Palm intensity and spectral density have the explicit form

as follows (Ramesh, 1995):

$$c(t) = \frac{q_1 q_2 (\lambda_2 - \lambda_1)^2 e^{-(q_1 + q_2)t}}{(q_1 + q_2)^2}, \quad (3.13)$$

$$h(t) = \frac{q_1 \lambda_2 + q_2 \lambda_1}{q_1 + q_2} + \frac{q_1 q_2 (\lambda_2 - \lambda_1)^2 e^{-(q_1 + q_2)t}}{(q_1 + q_2)(q_1 \lambda_2 + q_2 \lambda_1)}, \quad (3.14)$$

$$g(\omega) = \frac{\pi \Lambda \mathbf{1}}{2\pi} \left\{ \frac{2q_1 q_2 (\lambda_2 - \lambda_1)^2}{((q_1 + q_2)^2 + \omega^2)(q_1 \lambda_2 + q_2 \lambda_1)} \right\}. \quad (3.15)$$

See also the chapter on the multivariate MMPP for the derivation of these second order moments.

3.5 Relatives of MMPP: MAP, BMAP

The Markov arrival process (MAP) is defined by a background finite Markov chain $X(t)$ with r states and infinitesimal generator matrix Q . On the time interval where $X(t) = i$, the arrivals are Poisson process with rate β_i . At the same time, there is a probability a_{ij} that an arrival occurs at a jump from state i to state j . The point process defined by the epochs of these two kinds of arrivals is a Markov arrival process, see Neuts (1979) and Rudemo (1973) for the seminal ideas of point process determined by a Markov chain.

Definition 3.3 Consider a finite Markov process $X(t)$ with infinitesimal generator $Q = C + D$, where all the off-diagonal elements of C and all the elements of D are nonnegative. The transitions associated with D are called type I transitions. A **Markov arrival process (MAP)** with parameters (C, D) , $MAP(C, D)$, is a point process where an event occurs when a type I transition occurs in the Markov chain.

Hence, for a $MAP(C, D)$, the infinitesimal generator is decomposed into two

components in which D gives the intensities of state change with arrivals and C are associated with those state changes without arrivals. That is

$$d_{ij} = \begin{cases} \beta_i, & i = j \\ q_{ij}a_{ij}, & i \neq j \end{cases}, \quad c_{ij} = \begin{cases} -\sum_{k \neq i} c_{ik} - \sum_{k=1}^r d_{ik} & i = j \\ q_{ij}(1 - a_{ij}), & i \neq j \end{cases}.$$

Note that "state change with an arrival" allows transitions like $i \rightarrow i$. In addition to the matrices C, D , a complete specification of a MAP also requires the specification of the initial distribution of $X(t)$. We always stipulate the initial distribution π is selected according to the stationary distribution which satisfies $\pi Q = \pi(C + D) = \theta$, $\pi \mathbf{1} = 1$, where θ is a zero vector and $\mathbf{1}$ is a vector with all entries being unity.

By using this notation given above, a Markov modulated Poisson process is a special case when $d_{ij} = 0$ for $i \neq j$, i.e. D is a diagonal matrix.

Theorem 3.2 For a MAP with parameter (π, C, D) , the joint density at Y_1, \dots, Y_n of the n interarrival times is $\pi e^{CY_1} D e^{CY_2} D \dots e^{CY_n} D \mathbf{1}$.

We omit the derivation of the likelihood of a MAP due to its similarity to MMPP. See Asmussen (2000) or Asmussen (2003) for a good review of the properties of MAP.

For good reasons in practical applications, it is natural to allow each arrival to be a doublet, triplet or multiplet. For this case, there exists a particular type of state transition associated with group arrivals or multiple rewards. Hence, the infinitesimal generator Q of $X(t)$ is decomposed into multiple components such as $Q = D(0) + D(1) + \dots + D(m)$ in which $D(k), k = 1, \dots, m$ gives the intensities

of state change with batch arrivals of size k and $D(0)$ are associated with those state changes without arrivals. The batch Markovian arrival process (BMAP) is a generalisation of the Markovian arrival process by having arrivals of size greater than one. The homogeneous case, the rate matrix exists and is given by the matrix Q in the definition 2.4.

Definition 3.4 Let Q be the infinitesimal generator of a finite Markov chain satisfying $Q = D(0) + D(1) + \dots + D(m)$, where all the off-diagonal elements of $D(0)$ and all the elements of $D(k), k = 1, \dots, m$ are nonnegative. We call a transition associated with $D(k), k = 1, \dots, m$ an arrival with batch size k . A point process defined by this type of arrival associated with $D(k), k = 1, \dots, m$ is called a **batch Markov arrival process (BMAP)** with parameters $(D(0), D(1), \dots, D(k))$.

Given a sequence of arrival times t_i with batch size b_i , the likelihood for the observations (t_i, b_i) is given by

$$\mathbf{L}(\theta) = \pi \prod_{i=1}^n \{e^{D(0)\Delta t_k} D(b_i)\} \mathbf{1} \quad (3.16)$$

where π is the initial distribution of $X(t)$ and $\Delta t_k = t_k - t_{k-1}$.

There are merits in obtaining parameter estimates via the EM algorithm, much the same as in the MMPP context, since an explicit iteration solution is obtainable for BMAP once the whole trajectory of the Markov chain is treated as missing data and hence the complete likelihood or its logarithm can be written down. After utilizing several sufficient statistics such as the sojourn time in each state, the transition times of the Markov chain from one state to another and the number of arrivals with batch size m to summarize the likelihood, the E-step requires only the evaluation

of the conditional expectation of those sufficient statistics given the observations, which can be written up neatly in terms of forward and backward equations as in the context of general hidden Markov model. Then an explicit M-step follows, see Rydén (1996a) and Klemm et al. (2003) for a detailed description of the algorithm.

3.6 Chapter Appendix

The matrix exponential e^{At} can be defined through the convergent power series $e^{At} = \sum_{k=1}^{\infty} \frac{(At)^k}{k!}$. It is the solution to the matrix differential equation $\frac{dX}{dt} = AX$ with initial condition $X(0) = I$. Using the series expansion for e^{At} , the following properties can be derived:

$$(1) \quad e^{A(s+t)} = e^{As}e^{At}; \quad (3.17)$$

$$(2) \quad e^{(A+B)t} = e^{As}e^{Bt} \text{ if and only if } AB = BA; \quad (3.18)$$

$$(3) \quad \frac{de^{At}}{dt} = Ae^{At} = e^{At}A. \quad (3.19)$$

Methods for computing e^{At} can be found in Moler and Van Loan (2003).

Part II

EXTENSIONS AND INFORMATION THEORETICAL ASPECTS

Chapter 4

Markov Modulated Poisson Process with Marks

4.1 Introduction

The topic of the current chapter is a type of Markov Modulated Poisson Process (MMPP), in which the occurrence times of the point process are attached to marks. More accurately, it is a marked doubly stochastic point process for which the stochastic intensity of the ground process and the mark distribution are determined by an underlying continuous time finite Markov chain. Such an extension of MMPP has potential applications in modeling spatial-temporal point patterns, multivariate point processes or other point processes with attached marks, see Figure 4.1 for an example.

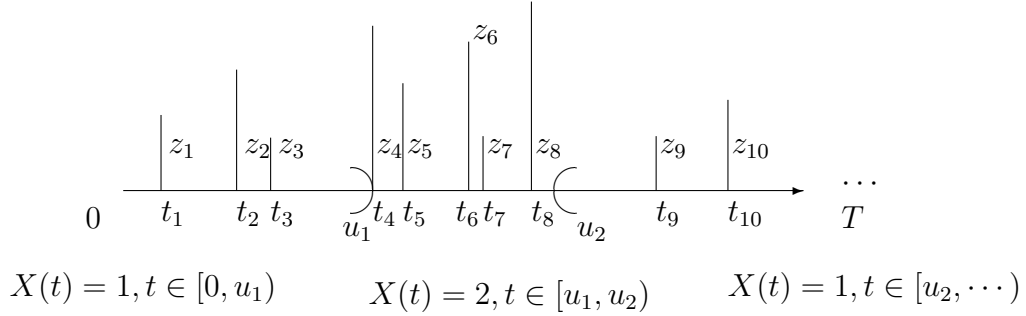


Figure 4.1: An example of a switching Poisson model (second order MMPP) associated with non-negative marks, where $\{t_i, z_i\}$ are the arrival times and associated marks, the brackets at $\{u_i\}$ are the transition times of the underlying Markov chain $X(t)$.

Definition 4.1 Markov modulated Poisson process with state-dependent marks is a marked doubly stochastic Poisson process with the conditional intensity $\lambda_{X(t)} f_{X(t)}(z)$ where $X(t)$ is a r -state irreducible Markov process, $\lambda_{X(t)}$ is the arrival rate of the ground process and $f_{X(t)}(z)$ is the probability density of the mark with respect to a reference measure μ on mark space \mathcal{Z} conditional on the current state of $X(t)$.

In this chapter, we obtain an extension for the likelihood of MMPP with marks and outline a procedure for parameter estimation by the EM algorithm. We pay special attention to the case when the marks come from an exponential family distribution. Then we discuss the statistical inference of the state process and estimation of the intensity rate of the observed point process. A method of assessing the goodness-of-fit for MMPP with marks based on the residual point process is also suggested.

Then we present some simulations to show the effects of the additional marks on the estimation efficiency of MLE for the infinitesimal generator of the underlying Markov chain and the intensity rates of observed point process. The effect is demonstrated by comparing the bootstrap variances of the estimates for a MMPP and a MMPP with marks whose infinitesimal generator and intensity rates are assumed identical to the former so that both are comparable to each other in accuracy of the estimates.

4.2 The Likelihood

We denote the infinitesimal generator matrix of the underlying Markov chain $X(t)$ by Q with its (i, i) -th element $q_{ii} \hat{=} -q_i$ satisfying $q_i = \sum_{j:j \neq i} q_{ij}, i = 1, \dots, r$ and $q_{ij} > 0$ for $i \neq j$. The observed marked point process is specified by the conditional intensity rate $\lambda_{X(t)} f_{X(t)}(z)$, where $\lambda_{X(t)}$ is the conditional intensity rate of the ground process $N(t)$ and $f_{X(t)}(z)$ is the probability density of the mark with respect to a reference measure μ on mark space \mathcal{Z} conditional on the current state of $X(t)$. The initial distribution π of the Markov chain is chosen according to the stationary vector which satisfies $\pi Q = 0$ and $\pi \mathbf{1} = 1$, where $\mathbf{1}$ is a column vector with unit entries. The internal history generated by this process is denoted by $\mathcal{F} = \{\mathcal{F}_t : 0 \leq t\}$. Given the observations $(t_i, z_i), i = 0, 1, \dots, n$ over $[0, T]$ and assuming $t_0 = 0$ without losing generality, the likelihood is obtained as below.

Let $X_k = X(t_k)$ and $Y_k = t_k - t_{k-1}$. The sequence $\{(X_i, Y_i, z_i)\}_{i=1}^n$ which forms a Markov sequence is equivalent to the sequence $\{(X_i, t_i, z_i)\}_{i=1}^n$. For any $y > 0$ and

Borel set $B \in \mathcal{B}(\mathcal{Z})$, the transition probability for this sequence is

$$P \{X_k = j, Y_k \leq y, Z_k \in B | X_{k-1} = i\} \quad (4.1)$$

$$= P \{X_k = j, Y_k \leq y | X_{k-1} = i\} P \{Z_k \in B | X_k = j\} \quad (4.2)$$

$$= \int_0^y \int_B \mathbf{e}'_i \exp\{(Q - \Lambda)t\} \Lambda \Upsilon(z) \mathbf{e}_j \mu(dz) dt, \quad (4.3)$$

where \mathbf{e}_i is a unit column vector with unit entry in the i th element, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$ and $\Upsilon(z) = \text{diag}(f_1(z), \dots, f_r(z))$. The first term of (4.2) is derived according to a result of Meier-Hellstern (1987) which gives the transition probability of the Markov renewal sequence $(X_i, Y_i)_{i=1}^n$, see also equation (3.1) in the previous chapter for the derivation of the likelihood of MMPP. According to the transition probability density of Markov sequence formed by the observations (X_i, Y_i, z_i) given in (4.3), the likelihood of the observations is obviously given in the following Theorem 4.1.

Theorem 4.1 Given that the initial distribution is stationary vector π and the transition density matrix Q , the likelihood for the inter-event times and associated marks $(Y_i, z_i)_{i=1}^n$ is

$$L(\pi, Q, \Lambda, \theta) = \pi \exp\{(Q - \Lambda)Y_1\} \Lambda \Upsilon(z_1) \cdots \exp\{(Q - \Lambda)Y_n\} \Lambda \Upsilon(z_n) \mathbf{1}, \quad (4.4)$$

where θ is the parameters in mark distributions.

To facilitate the evaluation of likelihood and other statistics involved in MMPP with marks, we introduce the forward and backward probabilities as in the context of discrete time hidden Markov models. The forward and backward probability densities are defined by $\alpha_t(i) \triangleq \Pr\{\text{points occurrences up to } t \text{ with } X(t) = i\}$ and $\beta_t(j) \triangleq \Pr\{\text{points occurrences from } t \text{ up to } T \text{ given } X(t)=j\}$ respectively. Denote $L_k = \exp\{(Q - \Lambda)Y_k\} \Lambda \Upsilon(z_k)$. For $0 < t < T$, the forward and backward

probabilities are written by $\alpha_t(i) = \pi L_1 \cdots L_{N(t)} \exp\{(Q - \Lambda)(t - t_{N(t)})\} \mathbf{e}_i$ and $\beta_t(j) = \mathbf{e}'_j \exp\{(Q - \Lambda)(t_{N(t)+1} - t)\} \Lambda \Upsilon(z_{N(t)+1}) L_{N(t)+2} \cdots L_n \mathbf{1}$, respectively. The likelihood of the observations in terms of this device is obviously $L = \sum_{i=1}^r \alpha_t(i) \beta_t(i)$.

The likelihood is invariant under the permutation of the states. But it is still possible that seemingly different MMPPs may have the same laws. The question of under what conditions the MMPPs are equivalent is the so-called identifiability problem. For the MMPP with marks, two point processes $N(t, z)$ and $\bar{N}(t, z)$ are equivalent if and only if their conditional intensity rates $\lambda_{X(t)} f_{X(t)}(z)$ and $\bar{\lambda}_{X(t)} \bar{f}_{X(t)}(z)$ have the same laws. The question can be translated into the identifiability of discrete time Markov chains by Poisson randomization of the continuous time Markov chain $X(t)$ and utilizing results of Ito et al. (1992) on identifiability of discrete Markov chain, see also Rydén (1996b) for the methods.

Poisson randomization (uniformization) is implemented by defining a Poisson process $N_P(t)$ with intensity rate $q \geq \|Q\| \triangleq \max_i q_i$ (assuming all diagonal elements of Q are uniformly bounded) and let $\{Y_n, n = 0, 1, \dots\}$ be a discrete time Markov chain with transition probability matrix $\mathbf{P} = \frac{1}{q}Q + I$ which itself is independent of $N_P(t)$. Then it is well-known that $\{Y_{N(t)}\}$ and $X(t)$ have same finite dimensional distributions, hence probabilistically identical, see Cinlar (1975).

4.3 Parameter Estimation of MMPP with Exponential Family Marks

4.3.1 Parameter estimation through EM algorithm

The key point for obtaining explicit E-steps and M-steps within the EM algorithm, as Asmussen et al. (1996) and Rydén (1996a) point out, is taking the whole trajectory of the Markov chain $X(t)$ as missing data. When this is done, much the same as Rydén (1996a) except for the conditional likelihood of the marks given the underlying Markov process, the complete data log-likelihood of this extended MMPP can be summarized in the form:

$$\begin{aligned} \log L^C(\pi, Q, \Lambda, \theta) &= \sum_{i=1}^r \left\{ \log \pi_i \mathbf{1}_{X(0)=i} - (q_i + \lambda_i) T_i + \sum_{j:j \neq i} N_{ij} \log q_{ij} + N_i \log \lambda_i \right\} \\ &\quad + \sum_{k=1}^n \log f_{X_k}(z_k) \end{aligned} \quad (4.5)$$

$$= \log L_1^C(\pi, Q, \Lambda) + \log L_2^C(\theta), \quad (4.6)$$

in which $T_i \triangleq \int_0^T \mathbf{1}\{X(t) = i\} dt$ is the sojourn time of $X(t)$ in state i , $N_{i,j} \triangleq \#\{t : 0 < t \leq T, X(t-) = i, X(t) = j\}$ is the number of transition times of $X(t)$ from i to j and $N_i \triangleq \int_0^T \mathbf{1}\{X(t) = i\} dN(t)$ is the number of observations occurred in state i , leaving the mark densities unspecified temporally. See Rydén (1996a) for the derivation of the complete likelihood of MMPP in $\log L_1^C$. We limit our attention to those marks whose distributions have no common parameters with $\log L_1^C$. Such a simplification is sufficient for most practical problems of interest at this stage. Therefore, $\log L_1^C$ and $\log L_2^C$ can be maximized separately within the M-step after taking expectation

conditional on the overall observations (the arrival times and associated marks). Note that the expectation value of $\log L_1^C$ conditional on the history \mathcal{F}_T in the E-step depends only on the expectation of the sufficient statistics (T_i, N_{ij}, N_i) for parameters Q and λ 's. They can be written as $E\{N_i|\mathcal{F}_T\} = \sum_{k=1}^n \alpha_{t_k}(i)\beta_{t_k}(i)/L$, $E\{T_i|\mathcal{F}_T\} = \int_0^T \alpha_t(i)\beta_t(i)dt/L$ and $E\{N_{ij}|\mathcal{F}_T\} = \int_0^T \alpha_t(i)q_{ij}\beta_t(j)dt/L$ which are denoted respectively by N_i^* , T_i^* and N_{ij}^* throughout later discussion and $L = \sum_{i=1}^r \alpha_t(i)\beta_t(i)$. After an explicit M-step, the parameters in $\log L_1^C(\pi, Q, \Lambda)$ are updated by

$$\hat{q}_{ij} = \frac{N_{ij}^*}{T_i^*}, \quad \hat{\lambda}_i = \frac{N_i^*}{T_i^*}, \quad i \neq j, 1 \leq i, j \leq r. \quad (4.7)$$

Note that equation (4.7) is nothing different from (3.6) except that extra information of the marks is incorporated into the group of sufficient statistics. Now suppose the mark distribution $f_{X_i}(z)$ comes from an Exponential family distribution in its canonical form $f(z) = h(z) \exp\{\theta^T t(z)\}/c(\theta)$, where the sufficient statistic $t(z)$ and parameter vector θ are d dimensional vectors, $h(z)$ and $c(\theta)$ are scalar functions. The parameter space Ω is a d -dimensional convex set such that $\Omega = \{\theta : \int h(z) \exp\{\theta^T t(z)\}dz < \infty\}$. Taking conditional expectation value of $\log L_2^C$ given the history of the observed process and ignoring terms without θ , we have

$$\begin{aligned} Q_2(\theta) &\hat{=} E\{\log L_2^C(\theta)|\mathcal{F}_T\} \\ &= \sum_{k=1}^r E\left\{\theta_k^T \left(\sum_{i=1}^n t(z_i) \mathbf{1}(X_i = k)\right) - N_k \log c(\theta_k) \middle| \mathcal{F}_T\right\} + \sum_{i=1}^n \log h(z_i) \\ &= \sum_{k=1}^r \left\{\theta_k^T \left(\sum_{i=1}^n \frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L} t(z_i)\right) - N_k^* \log c(\theta_k)\right\} + \sum_{i=1}^n \log h(z_i). \end{aligned}$$

On differentiation of $Q_2(\theta)$ with respect to θ in Ω , it follows that

$$\frac{\partial Q_2(\theta)}{\partial \theta_k} = \sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L} t(z_i) - \frac{N_k^*}{c(\theta_k)} \frac{\partial c(\theta_k)}{\partial \theta_k}.$$

Note that in exponential family $E_\theta\{t(z)\} = \frac{\partial \log c(\theta)}{\partial \theta}$, the M-step requires θ to be chosen by solving the equation

$$E_{\theta_k}(t(z)) = \frac{1}{N_k^*} \sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L} t(z_i). \quad (4.8)$$

If the above equation can be solved in Ω , then it is uniquely solvable due to the convexity property of minus the log likelihood of the regular exponential family.

Such a procedure is not automatically applicable for those marks from non-exponential family distribution. For instance, the t -distribution does not belong to the exponential family, but it has many applications in applied statistics. Assume the mark variable Z comes from the multivariate t -distribution $t_p(\mu, \Sigma, \nu)$ with location parameter μ , positive definite inner matrix Σ and ν degrees of freedom. Given weight u , $Z|u \sim N(\mu, \Sigma/u)$, where the random variable U corresponding weight u is distributed as $\Gamma(\frac{1}{2}\nu, \frac{1}{2}\nu)$. Hence, the multivariate t distribution is defined as the marginal distribution of the bivariate exponential family (U, Z) in Z . We omit the definition of multivariate t according to its probability density and instead treat it as a derivation distribution from the bivariate exponential family (U, Z) in the discussion of parameter estimation via EM algorithms. For known ν , a closed form EM step can be obtained on the T-type mark since $Z|u$ and u are from exponential family distribution. We treat not only the whole trajectory of the Markov chain but also the mark Z as missing data. Then the second part $\log L_2^C$ of the complete-data

log likelihood can be written by

$$\begin{aligned} \log L_2^C(\theta) = & c - \sum_{i=1}^n \sum_{k=1}^r \frac{N_k}{2} \log |\Sigma_k| \mathbf{1}\{X_i = k\} - \sum_{i=1}^n \sum_{k=1}^r \frac{u_i}{2} \mu'_k \Sigma_k^{-1} \mu_k \mathbf{1}\{X_i = k\} \\ & - \sum_{i=1}^n \sum_{k=1}^r \frac{u_i}{2} (Z'_i \Sigma_k^{-1} Z_i - 2\mu'_k \Sigma_k^{-1} Z_i) \mathbf{1}\{X_i = k\} + \sum_{i=1}^n f(u_i), \end{aligned} \quad (4.9)$$

where c and $f(u_i)$ are items without unknown parameters. In the E-step, the conditional expectation will be taken both over $(Z_i Z'_i, Z_i) \mathbf{1}(X_i = k)$ and u_i given the observed marked point process and current value of parameters. It turns out that the conditional distribution of U_i is distributed as $U_i | Z_i \sim \Gamma(m_1, m_2)$, where $m_1 = \frac{1}{2}(\nu + p)$ and $m_2 = \frac{1}{2}(\nu + (Z_i - \mu)' \Sigma^{-1} (Z_i - \mu))$, see page 63-68 in McLachlan and Krishnan(1997) for a detailed description. So we have

$U_i^* \triangleq E(U_i | Z_i) = \frac{\nu + p}{\nu + (Z_i - \mu)' \Sigma^{-1} (Z_i - \mu)}$. After the execution of an explicit E-step, each U_i is replaced by U_i^* and if we denote $\frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L}$ by $p_{t_i}(k)$, the M-step yields by:

$$\begin{cases} \hat{\mu}_k = \frac{\sum_{i=1}^n U_i^* Z_i p_{t_i}(k)}{\sum_{i=1}^n U_i^* p_{t_i}(k)}; \\ \hat{\Sigma}_k = \frac{\sum_{i=1}^n U_i^* (Z_i Z'_i - 2\hat{\mu}_k Z'_i + \hat{\mu}_k \hat{\mu}_k') p_{t_i}(k)}{N_k^*}. \end{cases} \quad (4.10)$$

4.3.2 Examples

Example 1: Assume each attached mark is an indicator of the class to which the point belongs, then such a MMPP with discrete marks forms a multivariate MMPP with intensity rate $\lambda_{X(t)} \prod_j (p_k^{(j)})^{\mathbf{1}(z=j)}$, where $\sum_j p_k^{(j)} = 1$, $p_k^{(j)} > 0$, $1 \leq j \leq M$, $1 \leq k \leq r$. Rewrite the mark distribution in its canonical form $f_k(z) = \exp\{\sum_j \mathbf{1}(z = j) \log p_k^{(j)}\}$ and note that sufficient statistic $\mathbf{1}(z = j)$ satisfies $E_{\theta_k}\{\mathbf{1}(z = j)\} = p_k^{(j)}$,

One obtains

$$\hat{p}_k^{(j)} = \frac{1}{N_k^*} \sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L} \mathbf{1}(z_i = j). \quad (4.11)$$

Example 2: When the mark variable of the MMPP is exponentially distributed with probability density $f(z) = \theta \exp\{-\theta z\} \mathbf{1}\{z > 0\}$, the expectation value of the sufficient statistic z given $\theta = \theta_k$ is $E_{\theta_k}(z) = \frac{1}{\theta_k}$. The explicit EM solution is given by

$$\hat{\theta}_k = N_k^* \left(\sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L} z_i \right)^{-1}. \quad (4.12)$$

Example 3: Suppose the mark variable comes from p -variate normal distribution $N_p(\mu, \Sigma)$ with canonical probability density $\exp\left\{-\frac{1}{2}(\mathbf{V}(zz')\mathbf{V}(\Sigma^{-1}) - 2\mu'\Sigma^{-1}z)\right\} \frac{1}{c(\mu, \Sigma)}$, where \mathbf{V} is the vectorization of a matrix and $c(\mu, \Sigma) = (2\pi)^{p/2} |\Sigma|^{1/2} \exp\{\frac{1}{2}\mu'\Sigma^{-1}\mu\}$. Then the expectation of the sufficient statistics (z, zz') for parameter (μ, Σ) are

$$\begin{cases} E_{\theta_k}(z) = \mu_k; \\ E_{\theta_k}(zz') = \Sigma_k + \mu_k \mu_k'. \end{cases}$$

Thus, in terms of the equation (4.8), the closed form of EM iteration step is given by

$$\begin{cases} \hat{\mu}_k = \frac{1}{N_k^*} \sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L} z_i; \\ \hat{\Sigma}_k = \sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L N_k^*} z_i z_i' - \left(\sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L N_k^*} z_i \right) \left(\sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L N_k^*} z_i' \right). \end{cases} \quad (4.13)$$

4.4 Inference on State Process and Observed Point Process

4.4.1 Estimate hidden state and intensity rate of observed point process

There exist state smoothers for MMPP such as those given by Elliott and Malcolm (2005). Much the same as in the context of discrete time hidden Markov models, a computationally efficient algorithm, fixed point smoothing, is available to estimate the probability of the underlying Markov chain in a given state at a specific time conditioned on all available observations, which only involves the forward and backward probabilities, see MacDonald and Zucchini (1997, pp 85). Let $\hat{p}_t(i) = P\{X(t) = i | \mathcal{F}_T\}$. After taking conditional expectation of $\mathbf{1}\{X(t) = i\}$ conditioned on observations, one obtains $\hat{p}_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^r \alpha_t(j)\beta_t(j)}$. The fixed point smoothing estimate $p_t(i)$ is obviously continuous with respect to t . Therefore, the fixed point smoothing estimates obtained at many grid points and connected by lines should be sufficient to mimic the evolution of the underlying process.

For the purpose of simulation, model evaluation and prediction (Daley and Vere-Jones, 2003, chapter 7.5) based on the intensity rate of MMPP, we use the conditional expectation of $\lambda_{X(t)}f_{X(t)}(z) = \sum_{i=1}^r \lambda_i f_i(z) \mathbf{1}\{X(t) = i\}$ conditioned on the observations (the arrival times and associated marks) as the estimate of the inten-

sity rate of the observed point process, say,

$$\hat{\lambda}(t, z) \triangleq \sum_{i=1}^r \lambda_i f_i(z) p_t(i) = \sum_{i=1}^r \lambda_i f_i(z) \frac{\alpha_t(i) \beta_t(i)}{\sum \alpha_t(i) \beta_t(i)}. \quad (4.14)$$

It is a weighted summation of all possible intensity rates $\lambda_i(t, z)$ which suggests an interpretation of MMPP as a mixing process with Markov dependence. Then the estimated ground intensity is given by $\hat{\lambda}(t) = \sum_{i=1}^r \lambda_i \frac{\alpha_t(i) \beta_t(i)}{L}$ and the conditional mark density is given by $\hat{f}(t, z) = \sum_{i=1}^r f_i(z) \frac{\alpha_t(i) \beta_t(i) \lambda_i}{\sum \alpha_t(k) \beta_t(k) \lambda_k}$.

To predict the occurrence of at least an event in a small time interval $[t, t + \Delta t]$ in terms of the intensity rate $\lambda(t, z | \mathcal{H}(t))$, one has to resort to point process simulation algorithms such as the inversion method or thinning method (Ogata, 1981) since the required prediction probability is rarely analytically obtainable. After the estimates of parameters are obtained, one can simulate an arrival time $t + \tau$ and a mark z according to $\lambda(t, z | \mathcal{H}(t))$ and add the new event to the history $\mathcal{H}(t)$. Then update $\lambda(t, z | \mathcal{H}(t))$ to $\lambda(t + \tau, z | \mathcal{H}(t))$. Repeat the above procedures until the time $t + \Delta t$ is exceeded. We repeat the above steps many times until the required quantities can be approximated by the simulation results satisfactorily. The incidence of at least one event occurring in $[t, t + \Delta t]$ may be taken as the estimate of the forecasting probability. See Daley and Vere-Jones (2003) in chapter 7.5 or Vere-Jones (1999) for the discussion of point process prediction.

4.4.2 Assessing the goodness-of-fit

Rescaling a multivariate or marked point process (Schoenberg, 1999; Vere-Jones and Schoenberg, 2004) technique is particularly useful for statistical model evaluation.

The detection of deviations of the rescaled residual point process (Ogata, 1988) from the standard process, a homogeneous unit rate Poisson process with stationary mark density in the circumstance of a marked point process, forms a key tool to suggest deficiencies and even illuminate possible improvements or alternatives of the current model. Vere-Jones and Schoenberg (2004) show that under quite general conditions by changing each point from (t_k, z_k) to $(\int_{t_0}^{t_k} \lambda(t, z_k) dt, z_k)$, the collection of these rescaled ones forms a stationary compound Poisson process with unit ground intensity and stationary mark distribution on the rescaled space. Especially when the mark admits a continuous univariate cumulative distribution function $F_\pi(z)$ with respect to a reference probability measure π , the doubly transformed points $(\int_{t_0}^{t_k} \lambda(t, z_k) dt, F_\pi(z_k))$ form a planar Poisson process with unit rate over $\mathcal{R} \times [0, 1]$, see also Daley and Vere-Jones (2008) for a proof. To numerically evaluate

$$\Lambda(t_k, z_k) \hat{=} \int_{t_0}^{t_k} \frac{1}{L} \sum_{i=1}^r \lambda_i f_i(z_k) \alpha_t(i) \beta_t(i) dt, \quad (4.15)$$

we use the computational methods of the integral of the matrix exponential involved in MMPP with marks, a numerical technique which can be implemented through matrix eigenvalue decomposition (Rydén, 1996a), Poisson randomization (Klemm et al., 2003) or a certain matrix exponential in higher order (Van Loan, 1978). Then the usual tests for Poisson processes such as the K -functions, L -functions and nearest neighbor tests etc. can be applied to evaluate the goodness-of-fit of the proposed hidden Markov models, see Cressie (1991). For instance, the estimated K -functions $K(d)$ which indicates the proportion of paired points per unit area within a specified distance d is a powerful test to detect clustering or regularity appearing in the spatial point pattern.

4.5 Simulation study

For MMPP without marks, it is noted that the better the intensity rates are separated, the better the estimates are close to the true values for certain length of observations, given that Q is fixed. Intuitively, the estimates of MMPP attached by state-dependent marks should be better than that of MMPP without marks in accuracies and efficiencies when the infinitesimal generator Q and the intensity rates λ_i remain fixed, due to an increase of the mutual information between the observed point process and the underlying Markov chain brought by the additional marks. To conform this, we illustrate several simulations to compare the accuracies and efficiencies of the estimates for several types of MMPPs, one of them is MMPP without marks, others are MMPP associated with several different types of state-dependent marks whose Q matrix and intensity rates λ_i are identical. We also demonstrate that among MMPP with marks whose Q matrix and λ_i s are same, the better the mark distributions are separated, the better the estimates are close to the true values of the model parameters even for different types of marks. How well the mark distributions are separated is measured by the Kullback-Leibler divergence.

To appraise how the accuracies and efficiencies of estimates vary according to different types of marks attached to MMPP, It is preferred to evaluate the observed Fisher information of the estimates as the MLE is consistent, asymptotic normal under mild conditions, see Bickel et al. (1998), see also section 7.5. However, evaluating the Fisher information or the observed Fisher information which requires the derivatives of matrix exponential and its products is numerically complicated.

Existing Monte Carlo methods or approximation methods are questioned by whether their accuracies are sufficient to the level to enable a delicate comparison for MMPP attached by different marks. We turn to the bootstrap methods.

Assume all types of MMPPs with or without marks in the simulation have identical infinitesimal generator $Q = \begin{pmatrix} -1 & 1 \\ 0.5 & -0.5 \end{pmatrix}$ and intensity rates $\Lambda = \text{diag}(5, 1)$. We attach 4 pairs of marks to MMPP observations, e.g. $(N(0, 1), N(1, 1))$, $(N(0, 1), N(1, 0.5^2))$, $(\exp(2.5), \exp(3))$ and $(\exp(1), \exp(10))$ which are denoted by MMPP(1), MMPP(2), MMPP(3) and MMPP(4) respectively throughout later discussions. The Kullback-Leibler divergences $d(., .)$ of these 4 pairs of marks comparatively satisfies $d(\exp(2.5), \exp(3)) < d(N(0, 1), N(1, 1)) < d(N(0, 1), N(1, 0.5^2)) < d(\exp(1), \exp(10))$. Then we generate many series of observations in same length (1000 observations) from MMPP and MMPP with the 4 pairs of marks, each one of them repeats 1000 times. The histograms of the estimates for each type of MMPP attached by marks and MMPP without marks clearly indicate the effects of different marks on the estimation error, see Figure 4.2 and Figure 4.3.

From the histograms, it is evident that the estimates are more and more centralized about the true values of the parameters upon the paired marks are more and more separated in terms of the Kullback-Leibler divergence, see also the standard deviations and means of estimates as listed in Table 4.1. Note that for MMPP(3), the leverage effect of the paired marks $(\exp(2.5), \exp(3))$ exerting on the estimates is very weak comparing with that of MMPP without marks. The simulation study for MMPP(3) is a similar to the real situation in applications of chapter 9.2 which

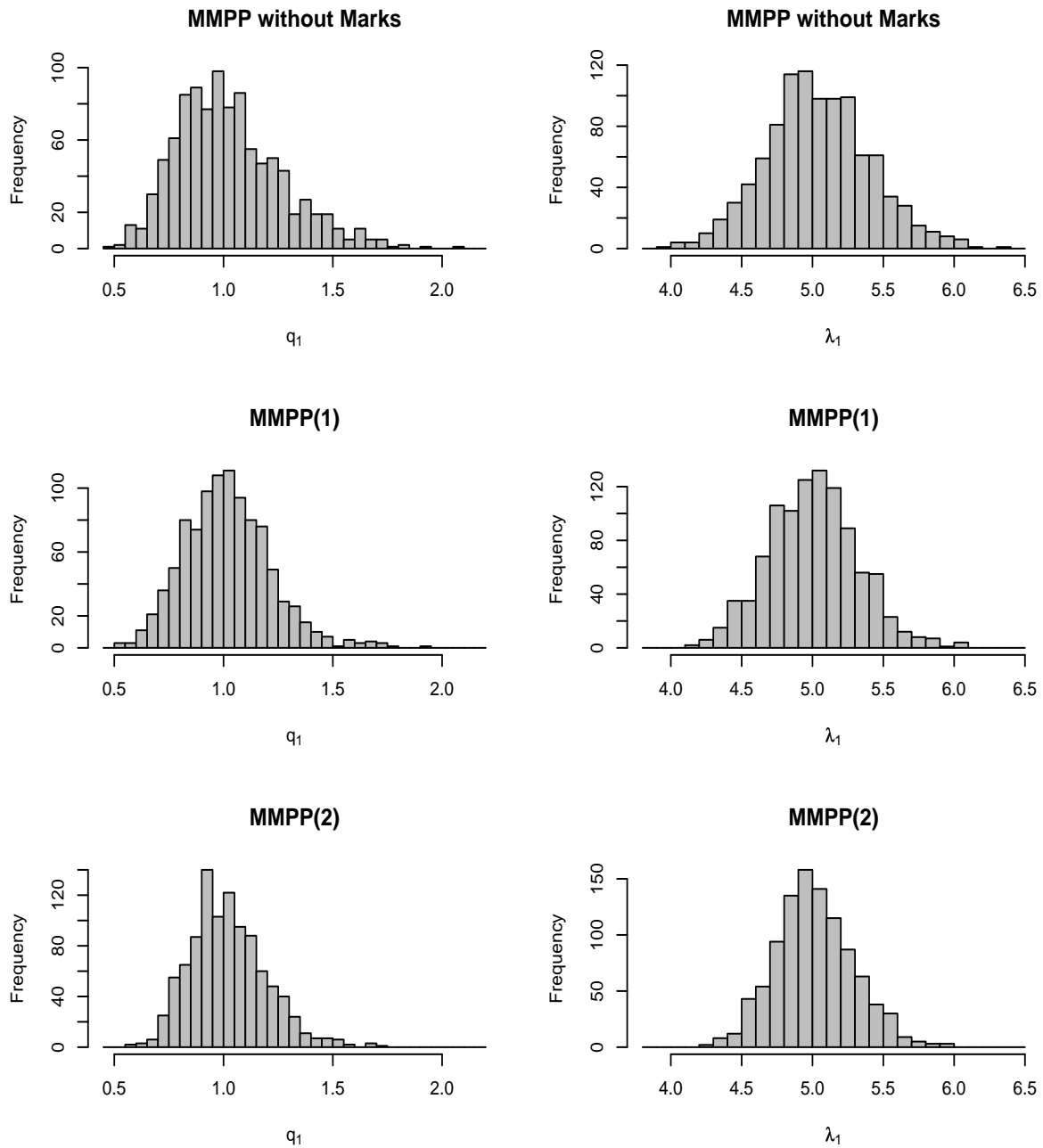


Figure 4.2: The left column of the plot gives the histograms of 1000 bootstrap replicates of estimated transition rates q_1 . The right column of the plot gives the histograms of 1000 bootstrap replicates of estimated intensity rates λ_1 .

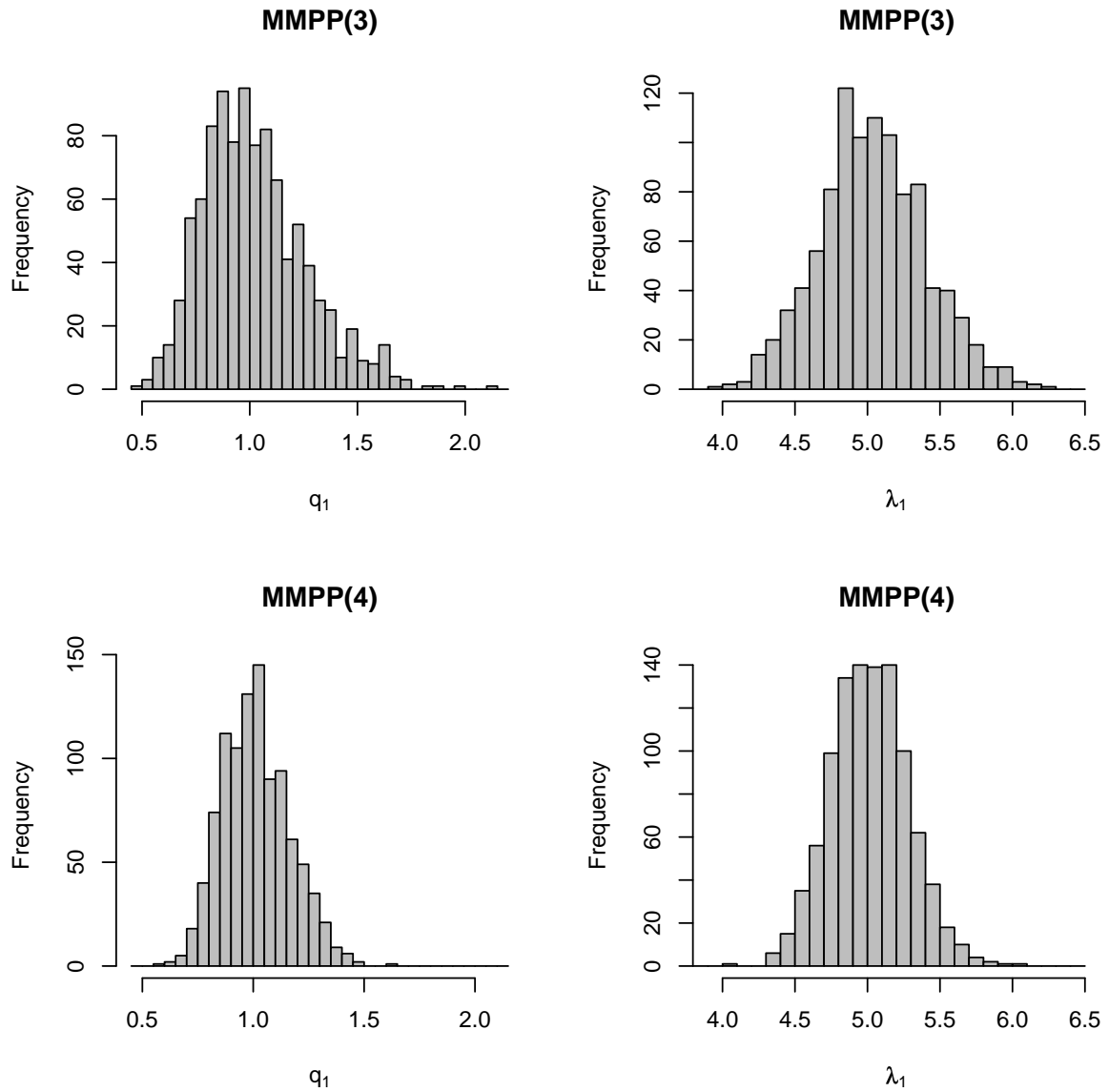


Figure 4.3: The left column of the plot gives the histograms of 1000 bootstrap replicates of estimated transition rates q_1 . The right column of the plot gives the histograms of 1000 bootstrap replicates of estimated intensity rates λ_1 .

Dev.	MMPP	MMPP(1)	MMPP(2)	MMPP(3)	MMPP(4)
q_1	-1.024 ± 0.242	-1.018 ± 0.198	-1.020 ± 0.174	-1.022 ± 0.240	-1.023 ± 0.152
q_2	-0.509 ± 0.123	-0.512 ± 0.102	-0.508 ± 0.076	-0.508 ± 0.122	-0.507 ± 0.067
λ_1	5.04 ± 0.368	5.01 ± 0.314	5.01 ± 0.276	5.04 ± 0.366	-0.504 ± 0.265
λ_2	1.00 ± 0.123	1.00 ± 0.096	1.00 ± 0.075	1.00 ± 0.122	1.00 ± 0.068

Table 4.1: Standard deviation and mean of estimates for Q and Λ according to 1000 bootstrap replicates obtained from simulated series of events with 1000 observations.

Std. Dev.	q_1	q_2	λ_1	λ_2
MMPP	0.1004	0.0525	0.1612	0.0558
MMPP(3)	0.0997	0.0521	0.1605	0.0554

Table 4.2: Standard deviation of estimates for Q and Λ according to 1000 bootstrap replicates obtained from simulated series of events with 5000 observations.

gives an indication of how much leverage effect the magnitude distribution can have on the estimates. The estimates are roughly convergent to the true values of the parameters at the rate of $1/\sqrt{n}$, see the histograms of estimates in Figure 4.4 and the standard deviations in Table 4.2 when the data sizes are increased to 5000 for MMPP and MMPP(3). When comparing MMPP and MMPP(3), It is more clear that the leverage effects from the marks $(\exp(2.5), \exp(3))$ exerting on the estimation are weakened with the increase of the data sizes..

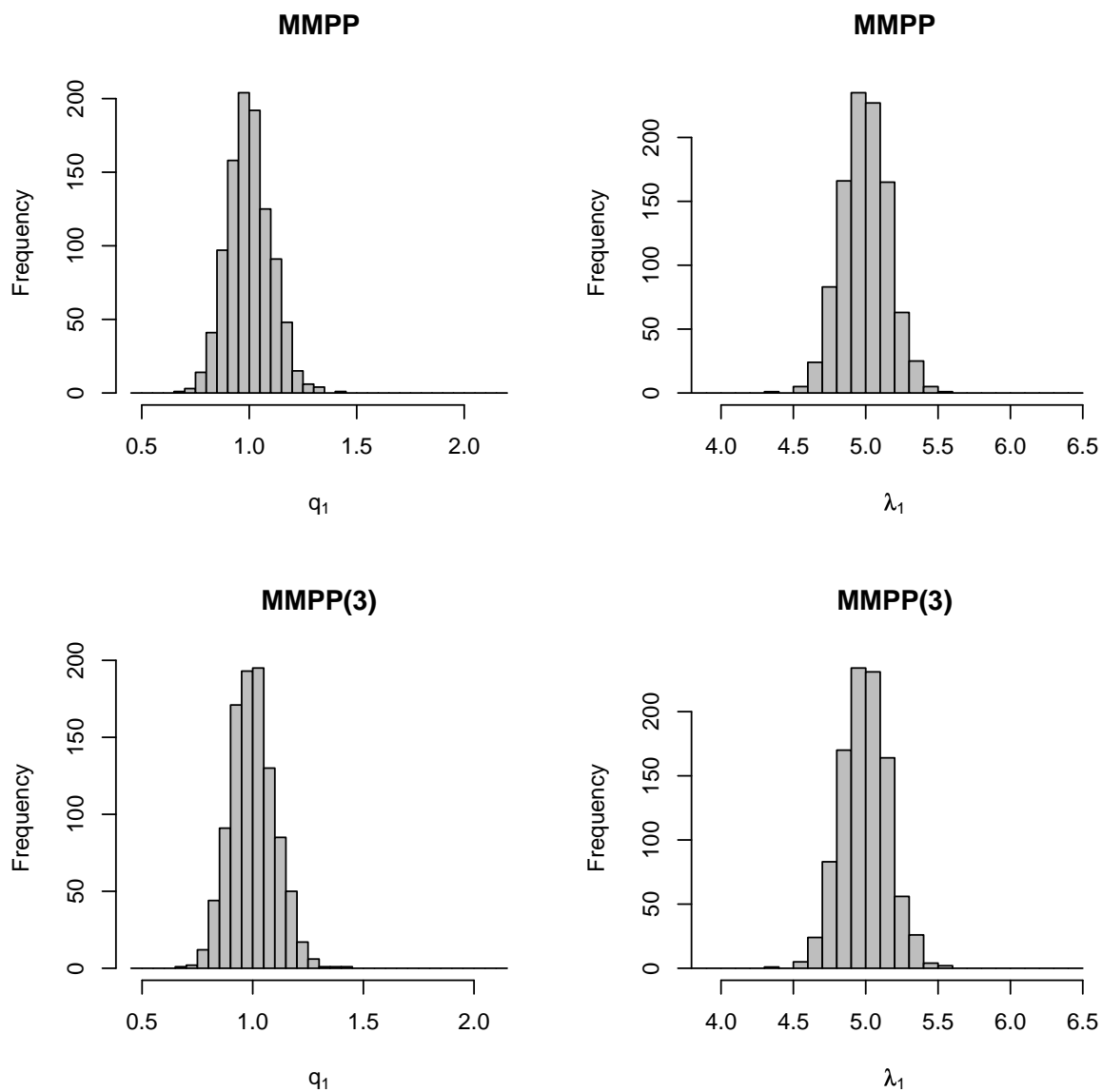


Figure 4.4: The left column of the plot gives the histograms of 1000 bootstrap replicates of estimated transition rates q_1 . The right column of the plot gives the histograms of 1000 bootstrap replicates of estimated intensity rates λ_1 .

Chapter 5

Multivariate MMPP

5.1 Introduction

This chapter is on a special case of MMPP with state-dependent marks, in which the mark indicate the class to which the point belongs. We focus on a detailed discussion of this type of MMPP including the second order moment analysis such as cross covariance density, cross intensity function etc. to characterize the correlation structure among component processes of a multivariate MMPP.

When two or more point processes evolve simultaneously, it is essential to describe not only the evolution of each marginal process individually but also the interplay of all processes, see Milne (1971). Much the same as the discrete time hidden Markov models, for Markov Modulated Poisson Process, it is natural to introduce the multivariate Markov Modulated Poisson Process, in which each marginal process is controlled by a common underlying continuous time finite state Markov

chain. The correlations among component processes are brought by the underlying Markov chain. To specify a multivariate point process exactly, we need some basic definitions.

Definition 5.1 In a univariate point process $N(t)$, the **regularity** is defined by requiring that $P\{N(t, t + \delta) > 1\} = o(\delta)$, for all $t \in \mathcal{R}$ as $\delta \rightarrow 0 +$.

Definition 5.2 For a multivariate point process, the process is **marginally regular** if its marginal processes are all regular. It is **regular** if the process of superposed marginal events is regular.

Other alternatives of the word regular are **simple** or **orderliness**. To specify a multivariate point process which is marginally simple by intensity rates of the process, it is necessary to specify not only the intensity rates of individual component process but also the simultaneous occurrence rate among marginal processes. One needs the following definition.

Definition 5.3 Denote the history of the process at time t by \mathcal{H}_t . The **complete intensity functions** (Cox and Lewis, 1972) of a bivariate point process $(N^{(1)}(t), N^{(2)}(t))$ are given by the conditional intensity rates of the marginal processes and the simultaneous occurrence rate among component processes such as

$$\lambda^{(1)}(t|\mathcal{H}_t) = \lim_{\Delta t \rightarrow 0+} \frac{P\{N^{(1)}(t, t + \Delta t) \geq 1|\mathcal{H}_t\}}{\Delta t}, \quad (5.1)$$

$$\lambda^{(2)}(t|\mathcal{H}_t) = \lim_{\Delta t \rightarrow 0+} \frac{P\{N^{(2)}(t, t + \Delta t) \geq 1|\mathcal{H}_t\}}{\Delta t}, \quad (5.2)$$

$$\text{and } \lambda^{(12)}(t|\mathcal{H}_t) = \lim_{\Delta t \rightarrow 0+} \frac{P\{N^{(1)}(t, t + \Delta t)N^{(2)}(t, t + \Delta t) \geq 1|\mathcal{H}_t\}}{\Delta t}. \quad (5.3)$$

However, in the case of a **simple bivariate point process**, it is sufficient to specify

only the conditional intensity rates of the marginal processes.

As a case in point, a bivariate MMPP $N(t) = (N^{(1)}(t), N^{(2)}(t))$ can be specified by the intensity functions of two marginal processes such as $\Lambda(t) = (\Lambda^{(1)}(t), \Lambda^{(2)}(t)) = (\lambda_{X(t)}^{(1)}, \lambda_{X(t)}^{(2)})$, in which $X(t)$ is a continuous time r th order Markov chain with the infinitesimal generator Q and $\lambda_{X(t)}^{(1)}, \lambda_{X(t)}^{(2)}$ are independent with each other conditioned on $X(t)$. Note that we use notations $\Lambda^{(i)}(t)$ and $\lambda^{(i)}(t)$ exchangably throughout the discussion. The simultaneous occurrence rate of the two components is obviously zero. Hence, a bivariate MMPP is a simple bivariate point process, which makes it sufficient to specify a bivariate MMPP only by the intensity rates of the marginal processes. It is also possible to specify the multivariate MMPP via counts, inter-event times or univariate point process associated with marks which indicate the class that it belongs to. The stationarity of a bivariate MMPP is studied through that of the underlying Markov chain. Particularly, when the underlying Markov process starts from stationary distribution π which satisfies $\pi Q = 0$ and $\pi \mathbf{1} = 1$, then correspondingly, the observed multivariate point process is stationary.

Definition 5.4 A **bivariate MMPP** is a **simple** bivariate doubly stochastic point process whose stochastic intensity rates for the marginal point process are given by $\Lambda(t) = (\lambda_{X(t)}^{(1)}, \lambda_{X(t)}^{(2)})$ where $X(t)$ is a r -state irreducible Markov chain.

Such an extension of MMPP is potentially useful for multivariate point process with Markov dependence, see Ramesh (1995). To fully exploit this device, it is necessary to derive its likelihood.

5.2 The Likelihood

We denote the internal history generated from the observed process by $\mathcal{F} = \{\mathcal{F}_t : t \geq 0\}$. Given the occurrence time t_k and the associated class z_k to which it belongs, the likelihood of the observations (Y_k, z_k) , $k = 1, \dots, n$ can be obtained as follows. Let $X_k = X(t_k)$ and $Y_k = t_k - t_{k-1}$. Then the sequence $(X_k, Y_k, z_k)_1^n$ which is equivalent to $(X_k, t_k, z_k)_1^n$, forms a Markov sequence. For any $y > 0$, the transition probability for this sequence is

$$\begin{aligned} & P\{X_k = j, Y_k \leq y, Z_k = z_k | X_{k-1} = i\} \\ &= \int_0^y e'_i \exp\{(Q - \Lambda^{(1)} - \Lambda^{(2)})t\} \Lambda^{(z_k)} e_j dt, \end{aligned}$$

where e_i is a unit column vector with unit entry in the i th element and $\Lambda^{(i)} = \text{diag}(\lambda_1^{(i)}, \dots, \lambda_r^{(i)})$. The derivation of the transition probability density is similar as (3.1) and (4.3). Hence, the likelihood of a bivariate MMPP is given by following theorem.

Theorem 5.1 (Ramesh, 1995) Given that the initial distribution is stationary vector π and the transition density matrix Q of Markov chain $X(t)$, the likelihood for the observations $(Y_i, z_i)_{i=1}^n$ is written by

$$L(\pi, Q, \Lambda) = \pi \exp\{(Q - \Lambda^{(1)} - \Lambda^{(2)})Y_1\} \Lambda^{(z_1)} \dots \exp\{(Q - \Lambda^{(1)} - \Lambda^{(2)})Y_n\} \Lambda^{(z_n)} \mathbf{1} \quad (5.4)$$

The evaluation of the likelihood and other statistics involved in the estimation and inference procedure is facilitated by introducing the forward and backward probabilities as in the previous context. Denote $L_k = \exp\{(Q - \Lambda^{(1)} - \Lambda^{(2)})Y_k\} \Lambda^{(z_k)}$. Respectively, for $0 < t < T$, the forward and backward probabilities are defined

by $\alpha_t(i) = \pi L_1 \cdots L_{N(t)} \exp\{(Q - \Lambda^{(1)} - \Lambda^{(2)})(t - t_{N(t)})\} e_i$ and $\beta_t(j) = e'_j \exp\{(Q - \Lambda^{(1)} - \Lambda^{(2)})(t_{N(t)+1} - t)\} \Lambda^{(z_{N(t)+1})} L_{N(t)+2} \cdots L_n \mathbf{1}$. The likelihood of the observations in terms of this device is $L = \sum_{i=1}^r \alpha_t(i) \beta_t(i)$. Note that when the two marginal processes are superposed, it forms a MMPP with parameters (Q, Λ) , where $\Lambda = \Lambda^{(1)} + \Lambda^{(2)}$, see equation (5.4) and (3.2).

5.3 Parameter Estimation through EM Algorithm

As discussed previously, the key point for applying the EM algorithm in this scenario is treating the whole trajectory of the underlying Markov chain as missing data. Then the complete data log-likelihood of the bivariate MMPP will be summarized by

$$\begin{aligned} \log L^c(\pi, Q, \Lambda) &= \sum_{i=1}^r \left\{ \log \pi_i \mathbf{1}\{X(0) = i\} - q_i T_i + \sum_{j:j \neq i} N_{ij} \log q_{ij} \right\} \\ &+ \sum_{i=1}^r \left\{ -(\lambda_i^{(1)} + \lambda_i^{(2)}) T_i + N_i^{(1)} \log \lambda_i^{(1)} + N_i^{(2)} \log \lambda_i^{(2)} \right\}, \quad (5.5) \end{aligned}$$

in which $T_i \triangleq \int_0^T \mathbf{1}\{X(t) = i\} dt$ is the sojourn time of $X(t)$ in state i , $N_{i,j} \triangleq \#\{t : 0 < t \leq T, X(t-) = i, X(t) = j\}$ is the number of transition times of $X(t)$ from i to j and $N_i^{(j)} \triangleq \int_0^T \mathbf{1}\{X(t) = i\} dN^{(j)}(t)$ is the number of type j arrivals upon $X(t)$ in state i . Note that the expectation value of $\log L^c(\pi, Q, \Lambda)$ conditional on the history \mathcal{F}_T in the E-step depends only on the expectations of the sufficient statistics $(T_i, N_{ij}, N_i^{(1)}, N_i^{(2)})$ for parameters Q and λ 's. They can be written as $E\{N_i^{(j)} | \mathcal{F}_T\} = \sum_{k=1}^n \alpha_{t_k}(i) \beta_{t_k}(i) \mathbf{1}(z_k = j) / L$, $E\{T_i | \mathcal{F}_T\} = \int_0^T \alpha_t(i) \beta_t(i) dt / L$ and $E\{N_{ij} | \mathcal{F}_T\} = \int_0^T \alpha_t(i) q_{ij} \beta_t(j) dt / L$ which are denoted respectively by $N_i^{(j)*}, T_i^*$ and

N_{ij}^* throughout later discussion. After replacing $N_i^{(j)}, T_i$ and N_{ij} in (5.5) by $N_i^{(j)*}, T_i^*$ and N_{ij}^* respectively and executing an explicit M-step, the parameters in $\log L^C$ are updated by

$$\hat{q}_{ij} = \frac{N_{ij}^*}{T_i^*}, \quad \hat{\lambda}_i^{(k)} = \frac{N_i^{(k)*}}{T_i^*}, \quad i \neq j, 1 \leq i, j \leq r, \quad k = 1, 2. \quad (5.6)$$

5.4 Second Order Moment Analysis

One important second order moment of multivariate point process is the second order cross intensity function which gives the occurrence rate of paired points from different component processes, assuming one of them at the origin. For instance, the **cross intensity function** of a bivariate MMPP is defined by $\lambda_2^{(1)}(t) = \lim_{\Delta t \rightarrow 0^+} \Pr\{N^{(1)}(t, t + \Delta t) \geq 1 | \text{a type 2 event occurs at } 0\}$ which is more easily computed in terms of intensity rates of the point process. Hence the equations relating the moment densities of the counting process and the moments of $\lambda(t)$ such as $E\{\lambda^{(i)}(t)\} = E\left\{\frac{dN^{(i)}(t)}{dt}\right\}$, $E\{\lambda^{(i)}(t_1)\lambda^{(i)}(t_2)\} = E\left\{\frac{dN^{(i)}(t_1)dN^{(i)}(t_2)}{dt_1 dt_2}\right\}$, $E\{\lambda^{(i)}(t_1)\lambda^{(j)}(t_2)\} = E\left\{\frac{dN^{(i)}(t_1)dN^{(j)}(t_2)}{dt_1 dt_2}\right\}$, $i \neq j$, $i, j = 1, 2$ are essential for obtaining these second order moments in terms of $\Lambda(t)$. These equations can be derived from the characteristic functionals of $N(t)$ defined as $E\left\{\exp \sum_{k=1}^2 \int_0^T i\theta_k(t) dN^{(k)}(t)\right\}$ where $\theta_k(t)$ is any bounded complex-valued Borel measurable functions, see Daley and Vere-Jones (2003, 2008) in chapter 5.5 and 9.4.

The equation relating the characteristic functional of a **simple** bivariate doubly stochastic point process and its complete intensity function $\Lambda(t) = (\Lambda^{(1)}(t), \Lambda^{(2)}(t))$

is given as

$$\begin{aligned} & E \left\{ \exp \left(\int_0^T i\theta_1(t) dN^{(1)}(t) + \int_0^T i\theta_2(t) dN^{(2)}(t) \right) \right\} \\ &= E_{\Lambda(t)} \left\{ \exp \left(\int_0^T (\lambda^{(1)}(t)[e^{i\theta_1(t)} - 1] + \lambda^{(2)}(t)[e^{i\theta_2(t)} - 1]) dt \right) \right\}. \quad (5.7) \end{aligned}$$

From this equation, complete relation between $N(t)$ and $\Lambda(t)$ such as identities relating the moment density function of the counting process and the moments of $\Lambda(t)$ illustrated above holds for **simple** bivariate doubly stochastic Poisson process.

This equation can be derived by following arguments.

$$\begin{aligned} & E \left\{ \exp \left(\int_0^T i\theta_1(t) dN^{(1)}(t) + \int_0^T i\theta_2(t) dN^{(2)}(t) \right) \right\} \\ &= E_{\Lambda(t)} \left\{ E \left\{ \exp \left(\int_0^T i\theta_1(t) dN^{(1)}(t) + \int_0^T i\theta_2(t) dN^{(2)}(t) \right) \middle| \Lambda(t) \right\} \right\} \\ &= E_{\Lambda(t)} \left\{ \sum_{\substack{M=0,1,\dots,m,\dots \\ N=0,1,\dots,n,\dots}} \frac{(\int_0^T \lambda^{(1)}(t) dt)^M}{M!} e^{-\int_0^T \lambda^{(1)}(t) dt} \frac{(\int_0^T \lambda^{(2)}(t) dt)^N}{N!} e^{-\int_0^T \lambda^{(2)}(t) dt} \right. \\ &\quad \left. \prod_{\substack{m=1,\dots,M \\ n=1,\dots,N}} \int_0^T \int_0^T \frac{e^{i\theta_1(t_m)+i\theta_2(t_n)} \lambda^{(1)}(t_m) \lambda^{(2)}(t_n)}{\int_0^T \lambda^{(1)}(t) dt \int_0^T \lambda^{(2)}(t) dt} dt_m dt_n \right\} \\ &= E_{\Lambda(t)} \left\{ \sum_{M=1}^{\infty} \sum_{N=1}^{\infty} e^{-\int_0^T (\lambda^{(1)}(t)+\lambda^{(2)}(t)) dt} \frac{(\int_0^T e^{i\theta_1(t)} \lambda^{(1)}(t) dt)^M}{M!} \frac{(\int_0^T e^{i\theta_2(t)} \lambda^{(2)}(t) dt)^N}{N!} \right\} \\ &= E_{\Lambda(t)} \left\{ \exp \int_0^T (\lambda^{(1)}(t)[e^{i\theta_1(t)} - 1] + \lambda^{(2)}(t)[e^{i\theta_2(t)} - 1]) dt \right\}. \end{aligned}$$

The characteristic functional contains the information of all moments. For example, putting $\theta_1(t) = \theta_1$ between t_1 and $t_1 + h$, and 0 otherwise, $\theta_2(t) = \theta_2$ between t_2 and

$t_2 + h$, and 0 otherwise, expansions of θ gives that

$$\begin{aligned} & E \left\{ \exp \left(\int_0^T i\theta_1(t) dN^{(1)}(t) + \int_0^T i\theta_2(t) dN^{(2)}(t) \right) \right\} \\ &= 1 + i\theta_1 E\{N^{(1)}(t_1, t_1 + h)\} - \frac{\theta_1^2}{2} E\{N^{(1)}(t_1, t_1 + h)^2\} + \dots \\ & \quad + i\theta_2 E\{N^{(2)}(t_2, t_2 + h)\} - \frac{\theta_2^2}{2} E\{N^{(2)}(t_2, t_2 + h)^2\} + \dots \end{aligned} \quad (5.8)$$

Similar expansion of $E_{\Lambda(t)} \left\{ \exp \left(\int_0^T (\lambda^{(1)}(t)[e^{i\theta_1(t)} - 1] + \lambda^{(2)}(t)[e^{i\theta_2(t)} - 1]) dt \right) \right\}$ derives that $E\{\lambda^{(i)}(t)\} = E\left\{\frac{dN^{(i)}(t)}{dt}\right\}, i = 1, 2$.

Based on a similar discussion of the expansions of the character functional, it also can be derived that

$$E\{\lambda^{(i)}(t_1)\lambda^{(j)}(t_2)\} = E\left\{\frac{dN^{(i)}(t_1)dN^{(j)}(t_2)}{dt_1 dt_2}\right\}, \quad i, j = 1, 2. \quad (5.9)$$

Now we use these equations to derive the **cross covariance density**, the **cross intensity function** and the **cross spectral** of a bivariate MMPP. Note that the intensity rate of the marginal MMPP itself is a Markov chain when all the intensity rates are different from each other. So we have the following equations:

$$\begin{aligned} E\{\Lambda^{(1)}(0)\Lambda^{(2)}(t)\} &= \sum_{i=1}^r P\left\{\Lambda^{(1)}(0) = \lambda_i^{(1)}\right\} \lambda_i^{(1)} E\left\{\Lambda^{(2)}(t) | \Lambda^{(1)}(0) = \lambda_i^{(1)}\right\} \\ &= \sum_{i=1}^r \pi_i \lambda_i^{(1)} \sum_{j=1}^r p_{ij}(t) \lambda_j^{(2)} \\ &= \pi \Lambda^{(1)} p(t) \Lambda^{(2)} \mathbf{1}, \end{aligned} \quad (5.10)$$

where $p(t) = e^{Qt}$.

The **cross covariance density** of the bivariate MMPP is

$$\begin{aligned}
\gamma_1^{(2)}(t) &= \lim_{\Delta t_1, \Delta t_2 \rightarrow 0+} \text{Cov}\left\{N^{(1)}(0, \Delta t_1), N^{(2)}(t, t + \Delta t_2)\right\} / \Delta t_1 \Delta t_2 \\
&= E\left\{\Lambda^{(1)}(0)\Lambda^{(2)}(t)\right\} - E\left\{\Lambda^{(1)}(0)\right\}E\left\{\Lambda^{(2)}(t)\right\} \\
&= \pi\Lambda^{(1)}p(t)\Lambda^{(2)}\mathbf{1} - \pi\Lambda^{(1)}\mathbf{1}\pi\Lambda^{(2)}\mathbf{1} \\
&= \pi\Lambda^{(1)}\left\{p(t) - \mathbf{1}\pi\right\}\Lambda^{(2)}\mathbf{1}.
\end{aligned} \tag{5.11}$$

Also the **cross-intensity function** of the bivariate MMPP is

$$\begin{aligned}
h_1^{(2)}(t) &= \lim_{\Delta t_1, \Delta t_2 \rightarrow 0+} \text{Pr}\left\{N^{(2)}(t, t + \Delta t_2) > 0 \mid N^{(1)}[0, \Delta t_1) > 0\right\} / \Delta t_2 \\
&= \frac{\pi\Lambda^{(1)}p(t)\Lambda^{(2)}\mathbf{1}}{\pi\Lambda^{(1)}\mathbf{1}}.
\end{aligned} \tag{5.12}$$

The **cross-spectral of counts** of the bivariate MMPP is defined by

$$g_1^{(2)}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \gamma_1^{(2)}(t) dt. \tag{5.13}$$

For a bivariate MMPP with infinitesimal generator $Q = \begin{pmatrix} -q_1 & q_1 \\ q_2 & -q_2 \end{pmatrix}$, note the stationary distribution is $\pi = (\frac{q_2}{q_1+q_2}, \frac{q_1}{q_1+q_2})$ and the transition probability of the Markov chain $X(t)$ equals

$$p(t) = \begin{pmatrix} \frac{q_2}{q_1+q_2} + \frac{q_1 \exp\{-(q_1+q_2)t\}}{q_1+q_2} & \frac{q_1}{q_1+q_2} - \frac{q_1 \exp\{-(q_1+q_2)t\}}{q_1+q_2} \\ \frac{q_2}{q_1+q_2} + \frac{-q_2 \exp\{-(q_1+q_2)t\}}{q_1+q_2} & \frac{q_1}{q_1+q_2} + \frac{q_2 \exp\{-(q_1+q_2)t\}}{q_1+q_2} \end{pmatrix},$$

the cross-covariance, cross-intensity and the cross-spectrum are exactly given as

follows (Ramesh, 1995):

$$\gamma_1^{(2)}(t) = \frac{q_1 q_2 (\lambda_2^{(1)} - \lambda_1^{(1)}) (\lambda_2^{(2)} - \lambda_1^{(2)}) e^{-(q_1 + q_2)t}}{(q_1 + q_2)^2}, \quad (5.14)$$

$$h_1^{(2)}(t) = \frac{q_1 \lambda_2^{(2)} + q_2 \lambda_1^{(2)}}{q_1 + q_2} + \frac{q_1 q_2 (\lambda_2^{(1)} - \lambda_1^{(1)}) (\lambda_2^{(2)} - \lambda_1^{(2)}) e^{-(q_1 + q_2)t}}{(q_1 + q_2) (q_1 \lambda_2^{(1)} + q_2 \lambda_1^{(1)})}, \quad (5.15)$$

$$g_1^{(2)}(\omega) = \frac{\pi \Lambda^{(1)} \mathbf{1}}{2\pi} \left\{ \frac{2q_1 q_2 (\lambda_2^{(1)} - \lambda_1^{(1)}) (\lambda_2^{(2)} - \lambda_1^{(2)})}{((q_1 + q_2)^2 + \omega^2) (q_1 \lambda_2^{(1)} + q_2 \lambda_1^{(1)})} \right\}. \quad (5.16)$$

5.5 Inference on State Process and Hidden Process, Model Evaluation

The probability of the underlying Markov chain in a given state at a specific time is estimated by the fixed point smoothing algorithm which requires only the forward and backward probabilities numerically as shown previously. Denote $\hat{p}_t(i) = P\{X(t) = i | \mathcal{F}_T\}$, then

$$\hat{p}_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^r \alpha_t(i) \beta_t(i)}. \quad (5.17)$$

The state process then can be visualized by using a grid-based fixed point smoothing algorithm as portrayed in chapter 3.3 and 4.4.

The estimated intensity rates of the observed point process are similarly given by

$$\hat{\lambda}^{(k)}(t) = \sum_{i=1}^r \hat{\lambda}_i^{(k)} \frac{\alpha_t(i) \beta_t(i)}{\sum \alpha_t(j) \beta_t(j)}, \quad k = 1, 2. \quad (5.18)$$

The estimated intensity rates allow the simulation of a multivariate MMPP and model evaluation directly in terms of the estimated intensity rates.

Meyer (1971) states that if a multivariate point process (N_1, N_2, \dots, N_k) is formed from times $\{t_{ij}, i = 1, \dots, k\}$ with continuous compensator $(\Lambda_1, \Lambda_2, \dots, \Lambda_k)$ such that $\Lambda_i(\infty) = \infty$ for each $i = 1, \dots, k$, then the multivariate point process forms from $\{\Lambda(t_{ij}), i = 1, \dots, k\}$ are independent Poisson process with unit rate. See Brown and Nair (1988) for a simplified proof of the theorem.

By rescaling points of each component process from its original time to its estimated compensator $\int_0^t \sum_{i=1}^r \lambda_i^{(k)} \alpha_t(i) \beta_t(i) dt / L$ according to the estimated intensity rates of each component process, we obtain a rescaled process called 'a residual point process' by Ogata (1988). If the proposed model fits the pattern well, each transformed component process should be a Poisson process with unit rate. Any deviation from it suggests some deficiencies of the proposed model for the real data.

Chapter 6

MMPP with a subsidiary

Neyman-Scott type Process

6.1 Introduction

As discussed in later chapters, deep earthquakes rarely have plenty of aftershocks which decay according to Omori's law (Omori, 1894), see chapter 8. Nevertheless, some deep earthquakes do have a small portion of aftershocks, the abundance of which may vary from time to time over depth. In some cases, the deep clustering forms as multiplets, triplets or doublets. To better understand the clustering behavior of deep earthquakes, we develop more sophisticated model to accommodate this small scale clustering behavior and at the same time reserve the characteristic of relatively large scale time-varying activity appearing in deep earthquakes within the framework of Markov modulated Poisson process.

The method behind this idea is to treat the mainshock as a baseline MMPP process and simultaneously associate each mainshock with a subsidiary clustering process following after the main shock. Thus various clustering properties such as spatial or (and) temporal clustering, along with magnitude will be treated as the mark of Markov modulated Poisson process. We will use the same notations as those in chapter 4 throughout this chapter.

6.2 MMPP with compounding marks

The simplest example allowing clustering is assuming the baseline process is MMPP and each arrival is associated with a Poisson variable representing multiple occurrences. Here the distribution of the Poisson variable might also be dependent on the underlying Markov process. We explain this multiple occurrences as the number of following aftershocks. In this case, the model is nothing other than a MMPP with Poisson marks.

Using the notations appeared in the chapter 4 on MMPP with marks and denoting the mark distribution by $f(z) = \frac{\beta^z}{z!}e^{-\beta}$ or in its canonical form $\frac{e^{z \log \beta}}{z!}e^{-\beta}$, the explicit iteration solution of β is given by $\hat{\beta}_k = \sum_{i=1}^n \frac{\alpha_{t_i(k)}\beta_{t_i(k)}}{L} z_i / N_k^*$ according to an application of the resulting equation (4.8) in the chapter of MMPP with marks.

6.3 MMPP with a Subsidiary Neyman-Scott cluster process

The Neyman-Scott model (Neyman and Scott, 1958) is a cluster model proposed for the distribution of galaxies and ecological applications. In this model, the process of cluster centers is supposed stationary, Poissonian and the cluster members are independently identically distributed about the cluster centers with some common distribution function once the size of each cluster is specified.

More generally, we assume the cluster centers form a baseline MMPP and each cluster center is associated with a finite point process which can be regarded as a mark. We suppose further that each cluster size is governed by a Poisson distribution whose intensity parameter is also dependent on the underlying Markov chain. For instance, we assume the (earthquake) magnitudes in each cluster are exponentially distributed i.i.d variables so that the distribution of such a finite point process is $f(z, m_1, \dots, m_z) \hat{=} \frac{\beta^z}{z!} e^{-\beta} \prod_{j=1}^z \alpha e^{-\alpha m_j}$, where z is the size of the cluster and (m_1, \dots, m_z) are the magnitudes of the cluster members. Note that z and $\sum_{j=1}^z m_j$ are sufficient statistics of parameters β and α , and that $E(z) = \beta$ and $E(\sum_{j=1}^z m_j) = z/\alpha$, then the explicit EM iteration solution can be written by $\hat{\beta}_k = \sum_{i=1}^n \frac{\alpha_{t_i(k)} \beta_{t_i(k)}}{L} Z_i / N_k^*$ and $\hat{\alpha}_k = N_k^* / \sum_{i=1}^n \frac{\alpha_{t_i(k)} \beta_{t_i(k)}}{L} \bar{m}_i$ where \bar{m}_i is the average magnitude of cluster members within the i th cluster.

For a clustering model, it is natural to specify not only the distribution of the cluster size but also the time delay distribution of the cluster member from the cluster

center. For instance, we may assume that the time delays of the cluster members, given the cluster size, are i.i.d. exponential variables, which therefore forms a finite point process with distribution $f(z, \tau_1, \dots, \tau_z)$. Leaving aside the magnitude for simplicity, the distribution $f(z, \tau_1, \dots, \tau_z)$ can be written by $\frac{\beta^z}{z!} e^{-\beta} \prod_{j=1}^z \alpha e^{-\alpha \tau_j}$. The explicit EM iteration solution is given as those similar to aforementioned. A more appropriate selection of the time distribution for the aftershock is Omori's law, i.e. the aftershocks decay in terms of the inverse power law $f(\tau) = (p-1)c^{p-1}/(\tau+c)^p$. In this case, the joint distribution of this finite point process can be written as $f(z, \tau_1, \dots, \tau_z) = \frac{\beta^z}{z!} e^{-\beta} \prod_{j=1}^z (p-1)c^{p-1}/(\tau_j+c)^p$. At this stage, we first assume the time scaling parameter c is given, then the joint distribution in its canonical form is $f(z, \tau_1, \dots, \tau_z) = \exp\{z \log \beta - \log z! - \beta + z \log(p-1) + z(p-1) \log c - p \log(\prod_{j=1}^z (\tau_j + c))\}$. Note z and $\log(\prod_{j=1}^z (\tau_j + c))$ are sufficient statistics of parameters β and p , after some trivial integrations, we obtain $E\{\sum_{j=1}^z \log(\tau_j + c)\} = z(\log c + \frac{1}{p-1})$. From the estimation procedure aforementioned in chapter on MMPP with marks (equation (4.8)), the estimation is given by solving the following equation

$$\sum_{i=1}^n \frac{\alpha_{t_i}(k) \beta_{t_i}(k)}{L} \frac{1}{z_i} \log \left(\prod_{j=1}^{z_i} (\tau_i^{(j)} + c) \right) \Big/ N_k^* = \log c + \frac{1}{p_k - 1} \quad (6.1)$$

where $\tau_i^{(j)}$ is the j th cluster member in the i th cluster and z_i is the number of cluster members in the i th cluster. Denoting the left side of the equation by A , then the estimated p_k is given by $\hat{p}_k = \frac{1}{A - \log c} + 1$.

When c is also an unknown parameter, the distribution of the clustering process forming in the mark obviously does not belong to the exponential family. The estimation procedure set out in the previous chapter is not automatically applicable

for this case. We suggest one extension of the EM algorithm, the ECM (Meng and Rubin, 1993) for the parameter estimation since explicit EM iteration steps for p depend on c which itself is also unknown.

Using the same notation as in the chapter on MMPP with marks, the complete data likelihood can be written by

$$\begin{aligned}
 \log L^c &= \sum_{k=1}^r \left\{ \log \pi_k \mathbf{1}(X(0) = k) - q_k T_k + \sum_{j:j \neq k} N_{kj} \log q_{kj} + \lambda_k T_k + N_k \log \lambda_k \right\} \\
 &+ \sum_{k=1}^r \left\{ \log \beta_k \left(\sum_{i=1}^n z_i \mathbf{1}(X_i = k) \right) - \sum_{i=1}^n \log z_i! - N_k \beta_k \right\} \\
 &+ \sum_{k=1}^r \left\{ \log(p_k - 1) \left(\sum_{i=1}^n z_i \mathbf{1}(X_i = k) \right) + (p_k - 1) \log c \left(\sum_{i=1}^n z_i \mathbf{1}(X_i = k) \right) \right. \\
 &\quad \left. - p_k \left(\sum_{i=1}^n \log \left(\prod_{j=1}^{z_i} \tau_j^{(i)} + c \right) \mathbf{1}(X_i = k) \right) \right\} \\
 &= \log L_1^c + \log L_2^c + \log L_3^c, \tag{6.2}
 \end{aligned}$$

where the three summations over terms within the large braces are denoted by $\log L_1^c, \log L_2^c, \log L_3^c$ respectively.

Note that the complete data log-likelihood is decomposed into three terms in large brackets, they can be maximized separately since they do not have common parameters. The maximization of first two terms within ECM algorithm has already been suggested beforehand. We focus on the maximization of the third term.

The ECM algorithm is a subclass of the GEM algorithm. Briefly speaking, the M-step in EM algorithm is executed by several conditional maximization steps, each of which utilizes the updated results in last conditional maximization step. The preferred properties such as monotonicity and convergence of EM still hold for the

ECM. In this case the conditional maximization step is executed by first maximizing $\log L_3^c$ with respect to c given p fixed in last M-step, then maximizing $\log L_3^c$ with respect to p conditioned on c updated from the last conditional maximization step. However, no explicit iteration solution is obtainable for c , a direct maximization of $\log L_3^c$ such as Newton-Raphson etc. should be evoked here. After that, an explicit conditional maximization step for parameter p is applicable as above mentioned. More generally, the parameter c might be also dependent on the current state of the underlying Markov chain. Evidently, the difference between the two scenarios of parameterization, in which c either remains constant or also varies according to the underlying Markov chain is trivial when evoking the ECM algorithm.

Chapter 7

Information Theoretical Aspects of MMPP

7.1 Introduction

The purpose of this chapter is to address the statistical inference problem involved in MMPP from the point of view of information theory. Simulation studies show that the better the intensity rates of the observed point process are separated, the better the estimation is close to the true values of the parameters for certain length of simulated observations. Here, we suggest the accuracy and efficiency of the MLE is controlled by the mutual information rate between the underlying Markov chain and the observed point process when the infinitesimal generator is given. We validate this supposition via simulation studies since at this stage any explicit relationship between the mutual information rate and the Fisher information of the estimated

parameters will be avoided at least in this study. In later sections, we will review the entropy rate of a continuous time finite state Markov chain and that of Markov Modulated Poisson process by so called **intrinsic conditional intensity** of a stationary point process, see Papangelou (1978), Daley and Vere-Jones (2004). Then the mutual information rate and the conditional entropy rate of the observed point process given the whole trajectory of the underlying Markov chain are obtained in section 4. After that, a supporting simulation study is presented to validate this conjecture.

First, we recall some basic conceptions of Shannon entropy. The Shannon entropy of a finite measure which is absolutely continuous with respect to a measure μ with Radon-Nikodym derivative f is defined as $H_f = - \int f(x) \log f(x) d\mu(x)$. For a discrete measure μ , it is $H_\mu = - \sum_i \mu_i \log \mu_i$.

The definition is also similarly extended to a stochastic process. For example, the entropy of a discrete time stochastic process $X_1^n = (x_1, \dots, x_n)$ is defined as the entropy of the corresponding n dimension distribution of X_1^n , namely $H(X_1^n) = E\{-\log f(x_1, \dots, x_n)\}$, where $f(x_1, \dots, x_n)$ is the probability function of X_1^n with respect to some reference measure. Particularly, when the limit of the entropy at time n divided by n exists, it is called the entropy rate of the process. It is first defined by Shannon for an ergodic Markov chain with r dimensional finite state space as summation of the entropies of the transition probability $P = (p_{ij})$ weighted by the stationary probability vector $\pi = (\pi_1, \dots, \pi_r)$, i.e., $R(X) = - \sum_i \pi_i \sum_j p_{ij} \log p_{ij}$. For a continuous-time process X_t , the entropy at time T is given by $H(X_0^T) = -E(\log f(X_0^T))$, where $f(X_0^T)$ is the likelihood of the process X_t with respect to some

reference measure. We shall use X_0^T to denote the stochastic process $\{X(t), 0 \leq t \leq T\}$ for short through this chapter. The entropy rate for a continuous time process is similarly defined as that of discrete time process except n replaced by T . A comprehensive introduction of the information theory was given by Billingsley (1965), Cover and Thomas (1991).

7.2 The Entropy of the Underlying Markov Chain

In this section, we review a version of the Shannon-McMillan-Breiman theorem for an irreducible finite Markov chain. Given the state transition times $0 = u_0 < u_1 < \dots < u_m < u_{m+1} = T$ and $X(u_k)$ denoted by s_k , the likelihood of the underlying Markov chain X_0^T with initial probability vector $\pi = (\pi_1, \dots, \pi_r)$ and infinitesimal generator $Q = (q_{ij})_{r \times r}$ can be written as

$$L(X_0^T) = \pi_{s_1} \left\{ \prod_{k=1}^m q_{s_k} e^{-q_{s_k} \Delta u_k} \times \frac{q_{s_k, s_{k+1}}}{q_{s_k}} \right\} e^{-q_{s_{m+1}} \Delta u_{m+1}},$$

where $\Delta u_k = u_k - u_{k-1}$. After taking logarithm, the Log-likelihood is summarized as

$$\log L(X_0^T) = \sum_{i=1}^r \left\{ \log \pi_i \mathbf{1}\{X(0) = i\} - q_i T_i + \sum_{j: j \neq i} N_{ij} \log q_{ij} \right\}, \quad (7.1)$$

where $T_i \triangleq \int_0^T \mathbf{1}\{X(t) = i\} dt$ and $N_{ij} \triangleq \#\{t : 0 < t \leq T, X(t-) = i, X(t) = j\}$ are the total sojourn time of $X(t)$ in state i and the transition times of $X(t)$ from state i to state j respectively. They form a group of sufficient statistics of the parameters.

The entropy of the Markov chain $X(t)$ is nonetheless the expectation of the negative log-likelihood and its time average $\frac{1}{T} E\{-\log L(X_0^T)\}$ gives average entropy

increment per unit time. If the limit of such a time average exists as T approaches to infinity, it is the entropy rate of the Markov chain $X(t)$. Note that the log-likelihood is linear with respect to the sufficient statistics T_i and N_{ij} , it requires the first moment of the sufficient statistics T_i , N_{ij} and the limitations $\frac{1}{T}E\{T_i\}$, $\frac{1}{T}E\{N_{ij}\}$ for obtaining the entropy of X_0^T .

Lemma 7.1 $E(T_i) = \int_0^T P_i(t) dt$, where $P_i(t) = \pi e^{Qt} \mathbf{e}_i$.

Proof: From Fubini's theorem, we have

$$\begin{aligned} E(T_i) &= E \left\{ \int_0^T \mathbf{1}\{X(t) = i\} dt \right\} \\ &= \int_0^T E \left\{ \mathbf{1}\{X(t) = i\} \right\} dt. \end{aligned}$$

Lemma 7.2 (Albert,1962) For $i \neq j, 1 \leq i, j \leq r$, $E(N_{ij}) = q_{ij} \int_0^T P_i(t) dt$.

Proof: see the chapter appendix.

Lemma 7.3 (Geometric convergence rate of Markov chain) Let $X(t)$ be an irreducible finite Markov chain. Then $\lim_{t \rightarrow \infty} P_{jk}(t) = \pi_k$ is independent of j for all k . Moreover, there exist non-negative numbers α and ρ such that $|P_{jk}(t) - \pi_k| \leq \alpha \rho^t, t > 0$.

Lemma 7.4 The matrix $A = \int_0^\infty \{P(t) - \pi\} dt = (\int_0^\infty \{p_{jk}(t) - \pi_k\} dt)_{jk}$ of an irreducible Markov chain has finite elements.

Proof: This is a straightforward consequence of the geometric rate of convergence of $P(t)$ to its limits π as $t \rightarrow \infty$.

Proposition 7.1 The entropy rate of an irreducible finite state Markov chain $X(t)$ is $\sum_{i=1}^r \left\{ \pi_i q_i - \sum_{j:i \neq j} \pi_i q_{ij} \log q_{ij} \right\}$.

Proof: From lemma 7.4, we have $\lim_{t \rightarrow \infty} \frac{1}{T} \int_0^T P_i(t) dt = \pi_i$. After taking the negative expectation of the log-likelihood, it follows that the entropy rate $R_X = \lim_{t \rightarrow \infty} \frac{H(X_0^t)}{t} = \lim_{t \rightarrow \infty} \frac{-E(\log L(X_0^t))}{t}$ of the Markov chain $X(t)$ is given by $\sum_{i=1}^r \left\{ \pi_i q_i - \sum_{j:i \neq j} \pi_i q_{ij} \log q_{ij} \right\}$.

7.3 The Entropy Rate of MMPP

7.3.1 The Entropy of Point Process

Generally speaking, the realization of a point process is specified by the count N and the location (x_1, \dots, x_N) . Given the probability distribution controlling count N and location (x_1, \dots, x_N) conditional on N , the likelihood of such a realization of point process can be written in terms of Jannossy density as

$$L = j_n(x_1, \dots, x_N) = p_N N! \pi_N^{sym}(x_1, \dots, x_N),$$

where $\pi_k^{sym}(x_1, \dots, x_N)$ is a symmetric probability density for the locations of N unordered points, see i.e. Daley and Vere-Jones (2003) in chapter 5.3. The entropy of a point process defined as the expectation of the negative log likelihood then can be partitioned as

$$\begin{aligned} H &= - \sum_{k=0}^{\infty} p_k \log p_k - \sum_{k=1}^{\infty} p_k \int \pi_k^{sym}(x_1, \dots, x_k) \log[k! \pi_k^{sym}(x_1, \dots, x_k)] dx_1 \cdots dx_k \\ &= H(N) + E\{H(x_1, \dots, x_N) | N\}, \end{aligned} \quad (7.2)$$

in which the $H(N)$ is the **numerical** entropy associated with the count of events and $E\{H(x_1, \dots, x_N) | N\}$ is the **location** entropy related to the location of events,

see McFadden (1965).

This representation of point process entropy can be further simplified when the corresponding conditional intensity function $\lambda(t)$ exists. In this case, the log-likelihood of point process is given by

$$\begin{aligned}\log L &= \sum_{i=1}^{N(T)} \log \lambda(t_i) - \int_0^T \lambda(t) dt \\ &= \int_0^T \{ \log \lambda(t) dN(t) - \lambda(t) dt \},\end{aligned}$$

where $\{t_i\}_{i=1}^n$ are the occurrence times of points over interval $[0, T]$.

By taking expectation of the negative log likelihood, the entropy of the point process in terms of the conditional intensity function is written as

$$H = -E \left\{ \int_0^T \lambda(t) \log \lambda(t) dt \right\} + \int_0^T m(t) dt, \quad (7.3)$$

where $m(t) = E[\lambda(t)] = E[dN(t)/dt]$.

For a Marked point process with conditional intensity function $\lambda(t, k)$, the entropy can be written as (Daley and Vere-Jones, 2003)

$$\begin{aligned}H &= -E \left\{ \int_0^T \int_{\mathcal{K}} \log \lambda(t, k) dN(t, k) - \int_0^T \int_{\mathcal{K}} \lambda(t, k) l(dk) dt \right\} \\ &= -E \left\{ \int_0^T \int_{\mathcal{K}} \lambda(t, k) \log \lambda(t, k) l(dk) dt - \int_0^T \int_{\mathcal{K}} \lambda(t, k) l(dk) dt \right\} \\ &= -E \left\{ \int_0^T \lambda_g(t) \log \lambda_g(t) dt \right\} + \int_0^T m_g(t) dt \\ &\quad - E \left\{ \int_0^T \left[\int_{\mathcal{K}} f(k|t) \log f(k|t) l(dk) \right] \lambda_g(t) dt \right\},\end{aligned} \quad (7.4)$$

where $\lambda_g(t)$ is the ground intensity function and $f(k|t)$ is the conditional mark density satisfying $\lambda(t, k) = \lambda_g(t) f(k|t)$ with reference probability measure $l(k)$.

7.3.2 The Entropy of Stationary Point Process and Marked Point Process

When the point process or the marked point process is stationary, the above formulas can be further simplified by considering so-called **intrinsic conditional intensity** $\lambda^\dagger(t)$ conditioned on some initial σ -algebra, which in this case is taken by the infinite past history of the point process. Note that for a stationary marked point process

$$E[\lambda^\dagger(t, k)] = E[\lambda^\dagger(0, k)] = m_g E[f^\dagger(k|0)]$$

with $m_g = E[\lambda_g^\dagger(0)]$ denoting the overall mean rate. Therefore, the entropy of a marked point process can be written as

$$H = -T \left\{ E \left\{ \int_{\mathcal{K}} \lambda^\dagger(0, k) \log \lambda^\dagger(0, k) l(\mathrm{d}k) \right\} - m_g \right\}. \quad (7.5)$$

The entropy rate of a stationary marked point process is given by

$$R = \frac{H}{T} = -E[\lambda_g^\dagger(0) \log \lambda_g^\dagger(0)] + m_g - E[\lambda_g^\dagger(0) \int_{\mathcal{K}} f^\dagger(k|0) \log f^\dagger(k|0) l(\mathrm{d}k)]. \quad (7.6)$$

Following a similar argument, the entropy rate of a stationary point process is then reduced to

$$R = -E[\lambda^\dagger(0) \log \lambda^\dagger(0)] + m \quad (7.7)$$

with $m = E[\lambda^\dagger(0)]$. See Daley and Vere-Jones (2003, 2008) in the chapter 7.6 and 14.8.

7.3.3 The Entropy Rate of MMPP

From the above arguments, it is obvious that the stationarity of point process $N(t)$ is inherited from the stationarity of underlying Markov chain $X(t)$ of a MMPP with time-varying intensity rate $\lambda_{X(t)}$. In this case, the **intrinsic conditional intensity** of MMPP conditioned on an initial σ -algebra generated from the infinite past history of the point process is given by $\lambda^\dagger(\mathbf{0}) = \sum_{i=1}^r \pi_i \lambda_i$ when the underlying finite Markov chain is irreducible, hence ergodic. According to equation (7.7), we obtain the entropy rate of a Markov modulated Poisson process.

Theorem 7.1 The entropy rate of a Markov modulated Poisson process when the underlying Markov chain is finite and irreducible is given by

$$R_N = -\left(\sum_{i=1}^r \pi_i \lambda_i\right) \left(\log\left(\sum_{i=1}^r \pi_i \lambda_i\right) - 1\right).$$

Proposition 7.2 The entropy rate of a Markov modulated Poisson process associated with marks when the underlying Markov chain is finite and irreducible is given by

$$R_N = -\left(\sum_{i=1}^r \pi_i \lambda_i\right) \left(\log\left(\sum_{i=1}^r \pi_i \lambda_i\right) - 1\right) - \left(\sum_{i=1}^r \pi_i \lambda_i\right) \int_{\mathcal{K}} \left(\sum \pi_i f_i(k) \log\left(\sum \pi_i f_i(k)\right)\right) l(dk).$$

7.4 The Mutual Information Rate of the MMPPs

It is suggested in the previous chapter that the complete data log likelihood can be summarized by sufficient statistics as

$$\log L^c = \sum_{i=1}^r \left\{ \log \pi_i \mathbf{1}_{X(0)=i} - q_i T_i + \sum_{j:j \neq i} N_{ij} \log q_{ij} \right\} + \sum_{i=1}^r \left\{ -\lambda_i T_i + N_i \log \lambda_i \right\},$$

which can be decomposed into the log-likelihood of the underlying Markov chain and the conditional log-likelihood of the observed point process conditioned on the underlying Markov chain. Thus the **joint entropy** of the underlying Markov chain $X(t)$ and the observed point process $N(t)$ can be written by

$$H(X_0^T, N_0^T) = H(X_0^T) + H(N_0^T | X_0^T), \quad (7.8)$$

i.e., the entropy of the Markov chain $X(t)$ and the **conditional entropy** of $N(t)$ given the history of $X(t)$.

Also note that

$$\begin{aligned} E\{N_i\} &= E \left\{ \int_0^T \mathbf{1}(X(t) = i) dN(t) \right\} \\ &= E \left\{ \int_0^T \mathbf{1}(X(t) = i) \lambda(t) dt \right\} \\ &= E \left\{ \int_0^T \mathbf{1}(X(t) = i) \lambda_i dt \right\} = \lambda_i E T_i, \end{aligned} \quad (7.9)$$

we obtain the conditional entropy rate of $N(t)$ conditional on $X(t)$ according to Lemma 7.4.

Proposition 7.3 The **conditional entropy rate** of a MMPP conditioned on the underlying irreducible finite state Markov chain is given by $R_{N|X} = \lim_{T \rightarrow \infty} \frac{H(N_0^T | X_0^T)}{T} = \sum_{i=1}^r \{\pi_i \lambda_i - \pi_i \lambda_i \log \lambda_i\}$.

According to the properties of **mutual information** $H(X_0^t : N_0^t) = H(N_0^t) - H(N_0^t | X_0^t)$, we obtain following theorem for the mutual information rate between the observed point process and the underlying irreducible finite Markov chain.

Theorem 7.2 The **mutual information rate** $R_{X:N}$ of the underlying irreducible finite Markov chain $X(t)$ and the observed point process of a MMPP is given by $\sum_{i=1}^r \{\pi_i \lambda_i \log \lambda_i\} - (\sum_{i=1}^r \pi_i \lambda_i) \log(\sum_{i=1}^r \pi_i \lambda_i)$. It equals zero if and only if all the intensity

rates of the point process equal almost surely.

Proof: The first statement is the direct result of the previous discussion. We assume Z a random variable with probability distribution $P\{Z = \lambda_i\} = \pi_i, i = 1, \dots, r$. To prove the inequality, it is equivalent to prove $E\{Z \log \frac{Z}{EZ}\} \geq 0$ or $E\{\frac{Z}{EZ} \log \frac{Z}{EZ}\} \geq 0$. note that $x \log x$ is a convex function, by Jessen inequality, we have $E\{(\frac{Z}{EZ}) \log(\frac{Z}{EZ})\} \geq E(\frac{Z}{EZ}) \log(E\{\frac{Z}{EZ}\}) = 0$, the equality holds if and only if $P\{Z = \lambda\} = 1$.

lemma 7.5 For a MMPP associated with marks, we have

$$E \left\{ \sum_{i=1}^n \log f_{X_i}(Z_i) \right\} = \sum_{j=1}^r E(N_j) \int_{\mathcal{K}} f_j(z) \log f_j(z) l(dz). \quad (7.10)$$

Proof:

$$\begin{aligned} E \left\{ \sum_{i=1}^n \log f_{X_i}(Z_i) \right\} &= E \left\{ E \left(\sum_{i=1}^n \log f_{X_i}(Z_i) \middle| X_1, X_2, \dots, X_n \right) \right\} \\ &= E \left\{ \sum_{j=1}^r N_j \int_{\mathcal{K}} f_j(z) \log f_j(z) l(dz) \right\} \\ &= \sum_{j=1}^r E(N_j) \int_{\mathcal{K}} f_j(z) \log f_j(z) l(dz). \end{aligned}$$

Proposition 7.4 The joint entropy rate of the underlying irreducible finite Markov chain $X(t)$ and the observed marked point process of a Markov modulated Poisson process with stationary distribution $\pi = (\pi_1, \dots, \pi_r)$, infinitesimal generator Q and intensity rate $\lambda_{X(t)}$ is

$$\sum_{i=1}^r (\pi_i q_i - \sum_{j:i \neq j} \pi_i q_{ij} \log q_{ij}) + \sum_{i=1}^r (\pi_i \lambda_i - \pi_i \lambda_i \log \lambda_i - \pi_i \lambda_i \int_{\mathcal{K}} f_i(z) \log f_i(z) l(dz)).$$

Proof: This is a direct conclusion of lemma 7.4 and $\lim_{t \rightarrow \infty} \frac{E(T_i)}{t} = \pi_i$.

The first summation is the entropy rate of the underlying Markov chain and the second summation is the conditional entropy rate of the marked point process conditioned on the state process $X(t)$.

From proposition 7.2 and 7.4, we obtain the following theorem.

Theorem 7.3 The mutual information rate between the underlying Markov chain $X(t)$ and the marked point process of a MMPP associated with state-dependent marks defined as aforementioned is

$$\begin{aligned} & \sum_{i=1}^r (\pi_i \lambda_i \log \lambda_i) - \left(\sum_{i=1}^r \pi_i \lambda_i \right) \log \left(\sum_{i=1}^r \pi_i \lambda_i \right) + \sum_{i=1}^r \int_{\mathcal{K}} \pi_i \lambda_i f_i(z) \log f_i(z) l(dz) \\ & - \left(\sum_{i=1}^r \pi_i \lambda_i \right) \int_{\mathcal{K}} \sum_{i=1}^r \pi_i f_i(z) \log \left(\sum_{i=1}^r \pi_i f_i(z) \right) l(dz). \end{aligned} \quad (7.11)$$

The latter term is the gain of the mutual information rate due to the additional marks associated with MMPP arrivals. It is zero when all distributions of the mark are identical.

The following sections are based on a conjecture that the accuracy or efficiency of the estimation for MMPPs with identical infinitesimal generator of the underlying Markov chain varies according to how well the intensity rates are separated with each other in terms of the mutual information rate. In other words, for the MMPPs with identical infinitesimal generator of the underlying Markov chain, the greater the mutual information rate is, the better the estimates are in terms of Fisher information matrix.

7.5 Fisher Information and Observed Fisher Information

It is well-known that under Cramér type regularity conditions, the maximum likelihood estimate is consistent and asymptotically normal, the inverse of the asymptotic variance matrix is the Fisher information matrix which is defined as the covariance matrix of the score vector. Although for general hidden Markov models including Markov modulated Poisson processes the observations are not independent and identically distributed, under similar regularity conditions and other mild conditions, the MLE $\hat{\theta}$ is still consistent, asymptotic normal and the asymptotic variance equals to the inverse of the Fisher information matrix \mathcal{I}_{θ} , see literatures by Bickel et al. (1998), Douc and Matias (2001). In practice, the traditional approximations to Fisher information are $\mathcal{I}_{\hat{\theta}}$, i.e. the Fisher information matrix at the value when real parameters are replaced by the estimated ones and the observed information $I(\theta)$, i.e. minus the second derivative of the log likelihood function at $\hat{\theta}$ given the observations. Efron and Hinkley (1978) suggest a justification for preferring $I(\theta)$ to $\mathcal{I}_{\hat{\theta}}$. It is suggested that $1/I(\theta)$ is a better approximation to the variance than $\mathcal{I}_{\hat{\theta}}$ by a higher order for a translation family. Another reason for preferring the estimated Fisher information is that $\mathcal{I}_{\hat{\theta}}$ is not so easily tractable in some circumstances such as hidden Markov models.

Although the EM algorithm is criticized for that it does not automatically produce an estimate of the covariance matrix of the maximum estimate as Newton-Raphson algorithm, Louis (1982) made an notable contribution by introducing the EM algo-

rithm to obtain the observed Fisher information. It is suggested that

$$\begin{aligned}
I(\theta_0) &= -\frac{\partial^2 \log L(\theta_0)}{\partial \theta^2} = -E_{\theta_0} \left\{ \frac{\partial^2 \log L^C(\theta_0)}{\partial \theta^2} \middle| \mathcal{F}_{N_0^T} \right\} \\
&\quad - E_{\theta_0} \left\{ \left(\frac{\partial \log L^C(\theta_0)}{\partial \theta} \right) \left(\frac{\partial \log L^C(\theta_0)}{\partial \theta} \right)^T \middle| \mathcal{F}_{N_0^T} \right\} \\
&\quad + \left(\frac{\partial \log L(\theta_0)}{\partial \theta} \right) \left(\frac{\partial \log L(\theta_0)}{\partial \theta} \right)^T,
\end{aligned} \tag{7.12}$$

where $\mathcal{F}_{N_0^T}$ is the internal history generated by the observed point process $N(t)$. The first term in the right side of the equation is the conditional expectation of the complete data observed Fisher information, while the last two terms form minus the expected information for the conditional distribution of the underlying Markov process given the trajectory of the observed point process which in a heuristic form can be written as

$$I(\theta) = I^C(\theta) - I^{X_0^T | N_0^T}(\theta) \tag{7.13}$$

according to an application of the missing information principle. In other words, the observed Fisher information is the complete information minus the missing information. The third term in equation (7.10) can be obtained by noting that

$$\begin{aligned}
\frac{\partial \log L(\theta_0)}{\partial \theta} &= E_{\theta_0} \left\{ \frac{\partial \log L^C(\theta_0)}{\partial \theta} \middle| \mathcal{F}_{N_0^T} \right\} \\
&= \frac{\partial Q(\theta)}{\partial \theta} \bigg|_{\theta=\theta_0},
\end{aligned} \tag{7.14}$$

where Q is the conditional expectation given the observed point process which is implemented through EM steps. Similarly, the observed complete information matrix

is given by

$$\begin{aligned} I^C(\theta_0) &= E_{\theta_0} \left\{ \frac{\partial^2 \log L^C(\theta_0)}{\partial \theta^2} \middle| \mathcal{F}_{N_0^T} \right\} \\ &= \frac{\partial^2 Q(\theta)}{\partial^2 \theta} \bigg|_{\theta=\theta_0}. \end{aligned} \quad (7.15)$$

The above two equalities are readily computed via EM algorithm as implemented in previous chapter. However evaluating the middle term of the right side of the equation (7.10) requires calculating the second order conditional moments such as $E\{T_i T_j | N_0^T\}$ which is not in an computationally efficient way, see Rydén (1996a).

Oakes (1999) provides an explicit formula of the matrix of second derivatives of the observed data log-likelihood in terms of EM algorithm. Following Dempster et al. (1977), the observed data y with likelihood $L(\theta, y)$ depending on the parameter vector θ maybe expressed as a many-to-one map of complete data with log-likelihood $L_0(\theta, x)$. The maximum likelihood estimation is achieved by considering $Q(\theta'|\theta) = E_{X|Y;\theta} L_0(\theta', x)$. Under the usual interchangeable condition of the operations of expectation with respect to x and differentiation in θ , it can be shown that

$$\frac{\partial^2 L(\theta, y)}{\partial \theta^2} = \left\{ \frac{\partial^2 Q(\theta'|\theta)}{\partial \theta'^2} + \frac{\partial^2 Q(\theta'|\theta)}{\partial \theta' \partial \theta} \right\}_{\theta'=\theta}. \quad (7.16)$$

Although the formula looks ready to implement, it generally requires an explicit M-step before differentiating with respect to parameters which hence requires the derivatives of the forward and backward equations of MMPPs.

Lystig and Hughes (2002) describes an iteration algorithm to directly implement the observed Fisher information without using the EM algorithm. For the Markov

modulated Poisson process, it turns out the iterated computation which again requires the derivatives of matrix exponential is not computationally tractable and also it involves lengthy code work.

7.6 Simulation Studies

In this section, we will compare the accuracies and efficiencies of the EM estimates for two MMPPs with identical infinitesimal generator $Q = \begin{pmatrix} -1 & 1 \\ 0.5 & -0.5 \end{pmatrix}$ and the intensity rates $\lambda = (1, 3)$ and $\lambda = (1, 5)$, which are denoted by MMPP(1) and MMPP(2) respectively within this chapter. The mutual information rate for the two MMPPs are about 0.22 and 0.60 respectively. Although several numerical methods are available for obtaining the observed Fisher information matrix, some of them are not numerically efficient since they require the derivatives of the matrix exponential and its products. Louis's method also requires some Monte Carlo simulations to evaluate the higher order conditional moments of the score statistic of the complete likelihood conditional on the history of the observed process which is not accurate to the degree that satisfying conclusions can be drawn in this study. We turn to apply the bootstrap method to compare the variance or the standard error of the EM estimates of the two simulated MMPPs with certain length of observations. We simulate two types of MMPPs each more than 10000 times with 5000 arrivals in every sequence. For each sequence of simulated MMPP observations, the EM iteration procedure is repeated until the convergence criterion for the likelihood meets. Then we list the bootstrap estimates of the standard deviation of all parameters in two

Std. Dev.	q_1	q_2	λ_1	λ_2
MMPP(1)	0.2676	0.1626	0.2078	0.1462
MMPP(2)	0.1319	0.0796	0.1309	0.1345

Table 7.1: Standard deviation of estimations.

MMPPs, see Table 7.1. The histograms of the estimated parameters in two MMPP are also followed, see Figure 7.1 and 7.2.

The histograms for the estimated parameters indicate that for given length of observations, the more the mutual information rate between the observed point process and the underlying Markov chain, the better the estimations are for MMPPs with given infinitesimal generator of the underlying Markov chain. In other words, the better the intensity rates are separated with each other, the better the EM estimations are for MMPP with identical infinitesimal generator and how well the intensities are separated with each other somehow is dependent on the mutual information rate between the underlying Markov process and the observed point process. Such a supposition is strengthened by the simulation results given here.

7.7 Chapter Appendix

Proposition 7.4 (Albert, 1962) Let $n(t)$ be the total number of jumps of the finite irreducible Markov process in $[0, T]$. Then exist non-negative constants α and β ,

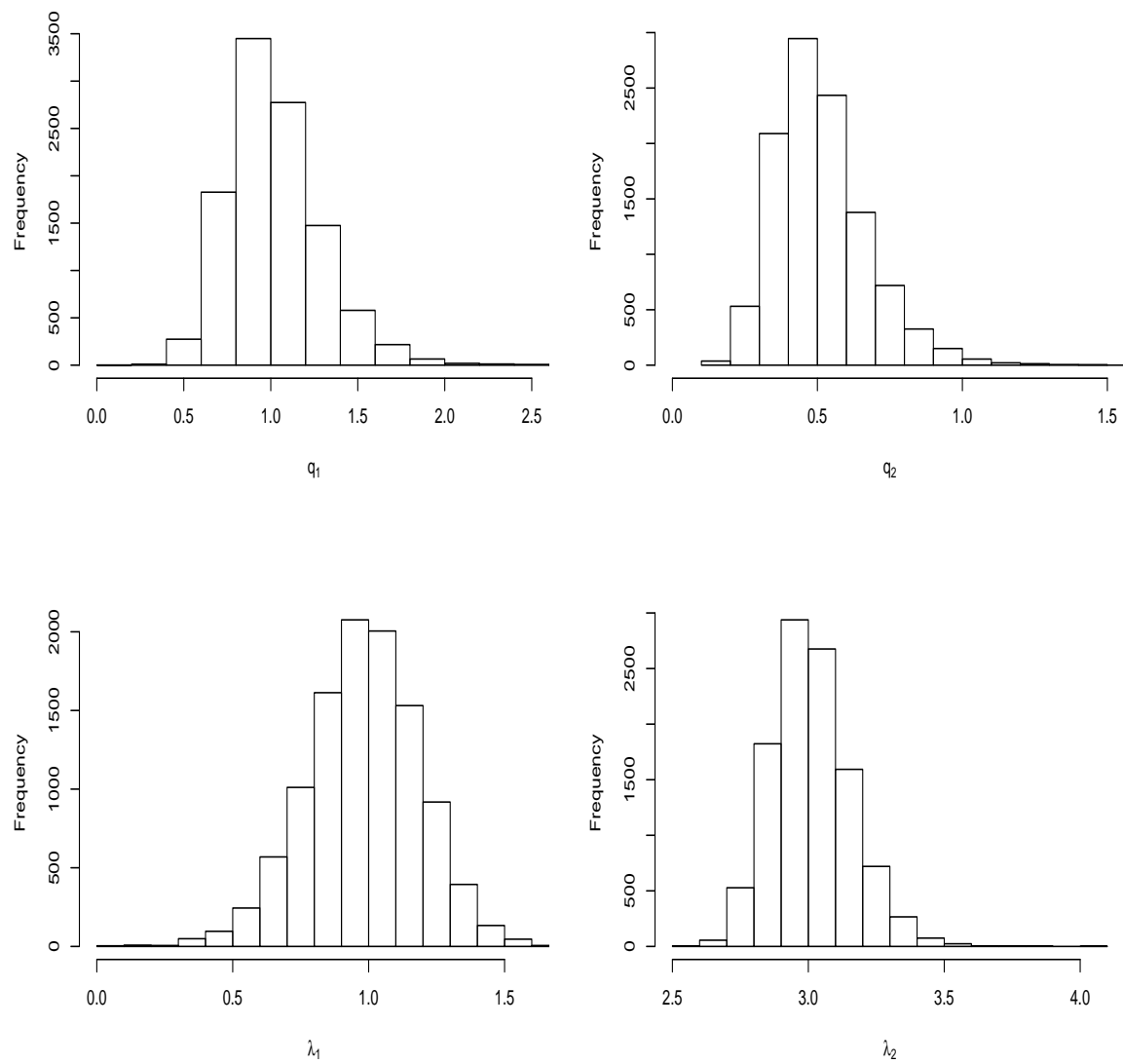


Figure 7.1: Histograms of estimated parameters for simulated data from MMPP(1).

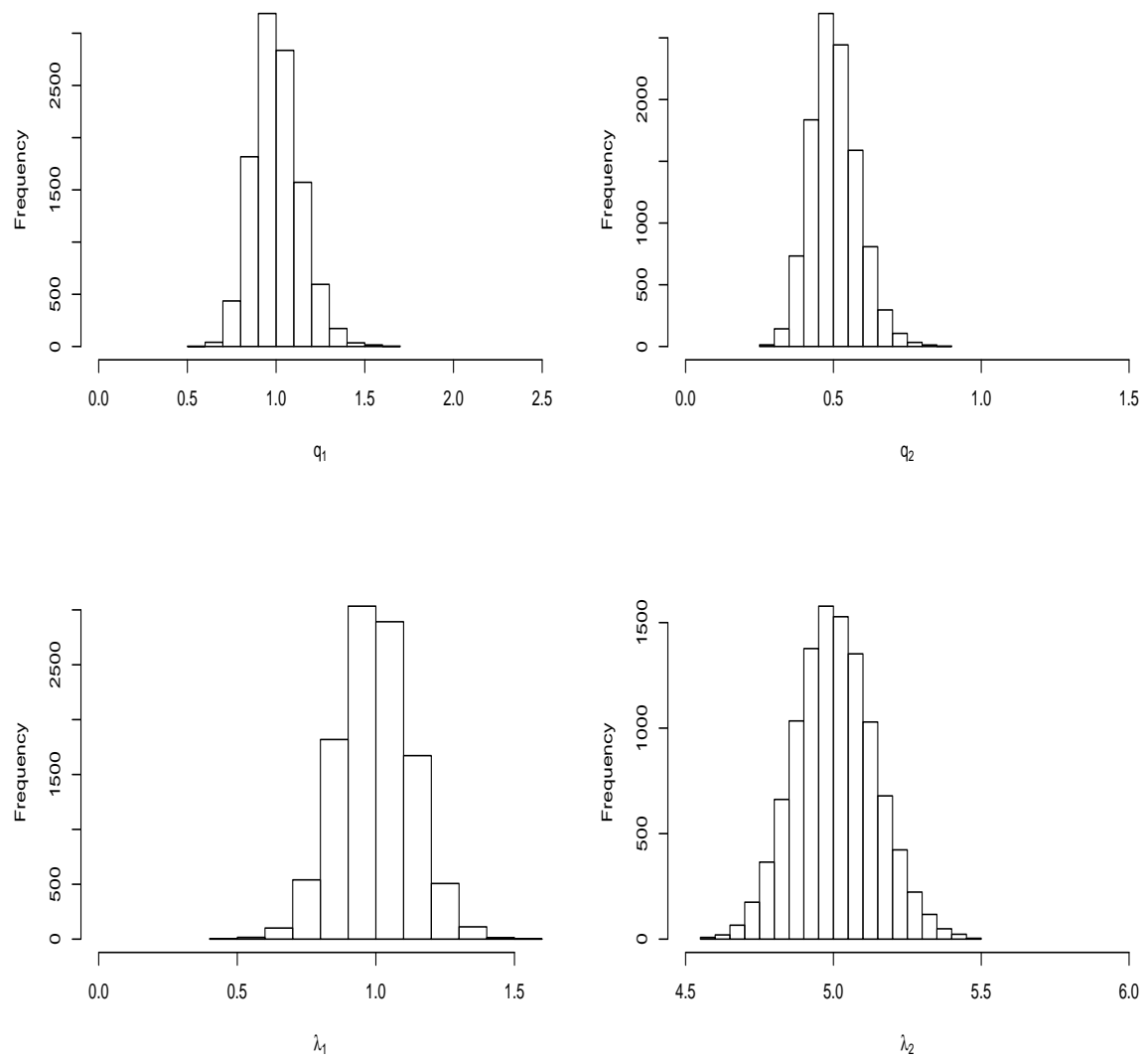


Figure 7.2: Histograms of estimated parameters for simulated data from MMPP(2).

$\beta > 0$, such that for all $h \in (0, 1), n \geq 1$,

$$Pr \{n(h) \geq n\} \leq \beta \int_0^h \frac{u^{n-1} e^{\alpha u}}{1-u} du.$$

Corollary 7.1

$$\sum_{n=2}^{\infty} Pr \{n(h) \geq n\} \leq \beta \int_0^t \frac{u e^{\alpha u}}{(1-u)^2} du = o(h).$$

Lemma 7.2 (Albert, 1962) For $i \neq j, 1 \leq i, j \leq r, E(N_{ij}) = q_{ij} \int_0^T P_i(t) dt$.

Proof: Let $[(k-1)h, kh), k = 1, \dots, s$ be a partition of the interval $[0, T)$ and $n_k(i, j)$ denotes the number of direct transitions from state i to state $j, i \neq j$, during the interval $[(k-1)h, kh), k = 1, \dots, s$, where $h = t/s$. It is easy to see that

$$N_{ij} = \sum_{k=1}^s n_k(i, j).$$

If s is large enough that $h < 1$, it implies that

$$Pr \{n_k(i, j) \geq 2\} = o(h)$$

$$\text{and } \sum_{n \geq 2} Pr \{n_k(i, j) \geq n\} = o(h).$$

Observe further that for $l \neq j$ and $r \neq i$,

$$\begin{aligned} & Pr \{n_k(i, j) = 1, X(kh) = l, X((k-1)h) = r\} \\ &= \{q_{ri}h + o(h)\} \{q_{ij}h + o(h)\} \{q_{jl}h + o(h)\} + o(h) \\ &= o(h). \end{aligned}$$

It follows, therefore, that

$$Pr \{n_k(i, j) = 1\} = Pr \{X((k-1)h) = i\} q_{ij}h + o(h),$$

which is Riemann sum approximation of the integral $q_{ij} \int_0^T P_i(u) du$.

Part III

APPLICATION TO DEEP EARTHQUAKES

Chapter 8

Exploratory Data Analysis for Deep Earthquakes

8.1 Introduction

A review of New Zealand deep earthquakes in 1964-87 by Reyners (1989) describes the setting of the subduction zones and the geometry of the Wadati-Beinoff zone and other aspects of deep earthquakes in New Zealand. After the up-gradation of the national seismograph network in the late 1980s, a further review of deep seismicity between 1990-93 by Anderson and Webb (1994) shows similar patterns for smaller events recognized by Reyners (1989). The catalogue quality is also evaluated in terms of accuracies of epicentres, depths and magnitudes. A comparison of the catalogue coverage between the PDE (Preliminary Determination of Epicenters) and New Zealand local earthquakes by Harte and Vere-Jones (1999) indicates that a

lot of events with magnitude greater than 5 in New Zealand catalogue are missing in PDE catalogue, probably due to sparse network coverage. It is also pointed out that the New Zealand catalogue misses many events with magnitude greater than 4.5 in the PDE catalogue occurring to the northeast of New Zealand. The difference of the epicentres, depths and magnitudes between two catalogues is also demonstrated in the study.

In earlier studies, Adams (1963) describes two cases of doublets located beneath the North Island at depth of about 230km and 600km. Gibowicz (1974) gives a description of the geometry of the Wadati-Beinoff zone in North Island. In the study the b-value (see 8.2.3) variations with depth are tabulated and associated with the maximum stress along the dipping slab. Vere-Jones et al. (1964) analyses the time trend of the seismicity in 1942-61. In that study, it is noted that an apparent increase in deep seismicity and a slight decrease in shallow seismicity. In the following study by Vere-Jones and Davies (1965), several statistical models are proposed for the clustering patterns appearing in the earthquakes and second moment properties such as variance time function, spectrum analysis are carried out for both shallow and deep earthquakes.

The seismograph network coverage has changed, from relatively sparse station coverage in the early postwar period to much improved coverage by digital seismograph after late 1980s. Within this period, there are two major upgrades of the national seismograph network respectively in 1960s and late 1980s, the latter of which is completed in mid 1990s, see Reyners (1989), page 308, Anderson and Webb (1994), page 477. The most recent seismograph network in operation is able to be

found at IGNS homepage (<http://www.gns.cri.nz/>). The coverage of the national seismograph network during 1987 can be found in Reyners (1989), page 308. The coverage information about the seismograph network in 1970s can be found in Adams and Ware (1977) in page 61. However, those in operation in early 1940s are only Wood-Anderson seismometers or torsion seismometer (see chapter 2.1) and teleseismic instruments with very sparse coverage around New Zealand, see Vere-Jones et al. (1964), page 723-725.

The purpose of this chapter is on the exploratory data analysis of the main New Zealand deep earthquakes around the North Island and north of the South Island. We first evaluate the catalogue coverage and completeness according to early literature, then confirm it by using internal statistical evidence. The epicentral or hypocentral distribution of these deep events and the characteristics of the Wadati-Benioff zones geometry is recognized again as in the early literature. Variations in depth distribution over time and variations in magnitude distribution over depth are pointed out in this study. We also explore the clustering properties of deep seismicity via a certain of statistical second order moment function. Then we give a tentative explanation why the stress release model and ETAS model are not satisfying in the situation of deep earthquakes.

To be safe in any pre-analysis of the catalogue data, we restrict our studies of New Zealand data at least from 1945 onwards within the confines determined in the Figure 8.4. Afterwards, the catalogue consistency, catalogue coverage and catalogue completeness are still subjected to evaluation due to the up-gradation of the National Seismograph Network in the 1960s and late 1980s, see the next section.

8.2 Descriptive Features and Empirical Properties

8.2.1 Data coverage and completeness

Fortunately, the great bulk of deep earthquakes in the region studied occur under land areas, mainly the center of the North Island and the North-East of the South Island. Relatively speaking, therefore, the deeper events are well situated for good coverage with dry land based seismometers. However, this is offset by two key features. First, the three-dimensional distances between source and recording stations are commonly of the order of several hundred kilometers, which severely limits the detectability of smaller events, particularly in the early periods when the network was sparse and the instruments less sensitive. The second feature relates to the difficulty of establishing a reliable velocity model for seismic waves which traverse the regions beneath the volcanic zone. The result is that hypocentres and depths for the deeper events may be significantly in error. This problem is particularly acute for events in the region North-East of the North Island, where further complications arise from changing network coverage and instruments over the decades.

For the 1940s and 1960s an appraisal of the network coverage for deep earthquakes is given in Vere-Jones, et al. (1964). From that study it would appear that the coverage at depth 50km would be incomplete for magnitude below 5. For the period 1970s and late 1980s a reasonable lower threshold might be 4.5, and for the later periods a threshold of 4 or even 3.5.

Then we turn to the internal statistical evidence for the catalogue completeness. Here we use the technique of reverse cusum analysis. In essence, this means setting a reference level \bar{x} for either occurrence frequencies or mean magnitudes from the recent period when the data is assumed to be complete, and comparing the results from earlier periods with this. The cusum (cumulative sum) defined by $\sum_i (x_i - \bar{x})$ compares the slopes of the cumulative curves. Where a change of slope occurs, there should be some underlying change in the mean frequency or mean magnitude. Thus it is not the levels which are important, but the change in slope.

Unfortunately, one characteristic of the deep earthquakes over the last few decades is that there have been gross fluctuations in occurrence frequencies which can hardly be attributed to network effects. The fluctuations are not restricted to one geographical region or to one magnitude class, but appear to affect occurrence rates across the whole region. This means that recent occurrence rates are no guide to occurrence rates in earlier periods, and therefore cannot be made a meaningful reference point for cusum analysis.

The mean magnitude $\sum_i (m_i - \bar{m})$, however, has been much more stable, and it is reasonable to assume that an increase in mean magnitude as we go backwards in time is mainly to be attributed to incomplete coverage of the smaller events. By carrying out such studies for different magnitude classes, we can obtain a reasonable impression of the threshold magnitude level for reliable event detection in earlier periods.

For events with magnitude greater than 4 in the first data set referred in Fig-

Figure 8.1, the increasing trend of yearly counts displayed in the left top of Figure 8.1 indicates uncertainties of probable contributing factors for the seismicity escalation, a real change or simply catalogue incompleteness. The most probable factor of contributing this is a combination of both factors at different degrees in different periods. The right top plot gives the magnitude-frequency relationship. The cumulative plots in the bottom should be read in a reverse direction since the specified training period, from where an average seismicity either monthly mean frequency or magnitude in high quality data set is taken as a reference value \bar{x} and the cumulative deviation $\sum(x_i - \bar{x})$ is drawn from recent to the past. In the bottom of Figure 8.1, the dropping slope of cumulative frequency and the ascending cumulative magnitude strongly suggest the incompleteness of the catalogue at this magnitude threshold. Only after 1970, does the catalogue completeness look convincing. Note that the right bottom plot is based on event by event scale, hence the dash lines delimiting the calendar years are separated apart in different degrees due to fluctuations in occurrence frequencies.

When the magnitude cutoff comes to 4.5, the evidence of catalogue completeness is indicated through yearly occurrence number, magnitude frequency and the cumulative plots of Figure 8.2 which suggests a relatively quiet period in historical seismicity in New Zealand from 1960s to late 1980s (Reyners, 1989). However, the abruptly ascending slope appearing in the cusum magnitude in the earliest period, approximately from 1945 to 1950, suggests that the catalogue may be still incomplete at some seismic active zones in this early postwar period. Similar analysis in Figure 8.3 shows that after 1988, probably 1989, the catalogue is relatively homo-

geneous at magnitude threshold level $m_0 = 3.5$. These figures are easily produced by using SSLib package in R, see Statistical Research Associates Limited referred in the reference list. This is an optimistic viewpoint, and it is likely that some events are missing even within the periods indicated. However, such omissions are unlikely to be on a large enough scale to affect the broad conclusions made in the present study.

8.2.2 Epicentral and depth distributions

The non-seismic corridor at the northern part of the South Island marks the termination of deep seismicity of the Wadati-Beinioff zone, see Figure 8.4. The southwest-northeast depth section plot (Figure 8.5) of deep seismicity at section azimuth 45° indicates that the dipping seismicity progressively deepens north-easterly. A further northwest-southeast depth section plot (Figure 8.6) at section azimuth 135° indicate a bending Wadati-Beinioff zone which is nearly perpendicular at the greater depth. Additional section plot (Figure 8.7) also shows a more diffusing seismic zone at northeastern side of the North Island. Such patterns are well recognized by Reyners (1989), Anderson and Webb (1994) and others.

Epicenter distribution displays intensive spatial grouping characteristic by using a smoothing procedure, see Figure 8.8 and Figure 8.9. The smoothing estimation is obtained by using many windows centered at pre-selected grid points. They are partially superposed with surrounding windows. Then we count the number of events in each window and take the average values of the numbers over superposed

Frequency, Magnitude and Cusum Statistics

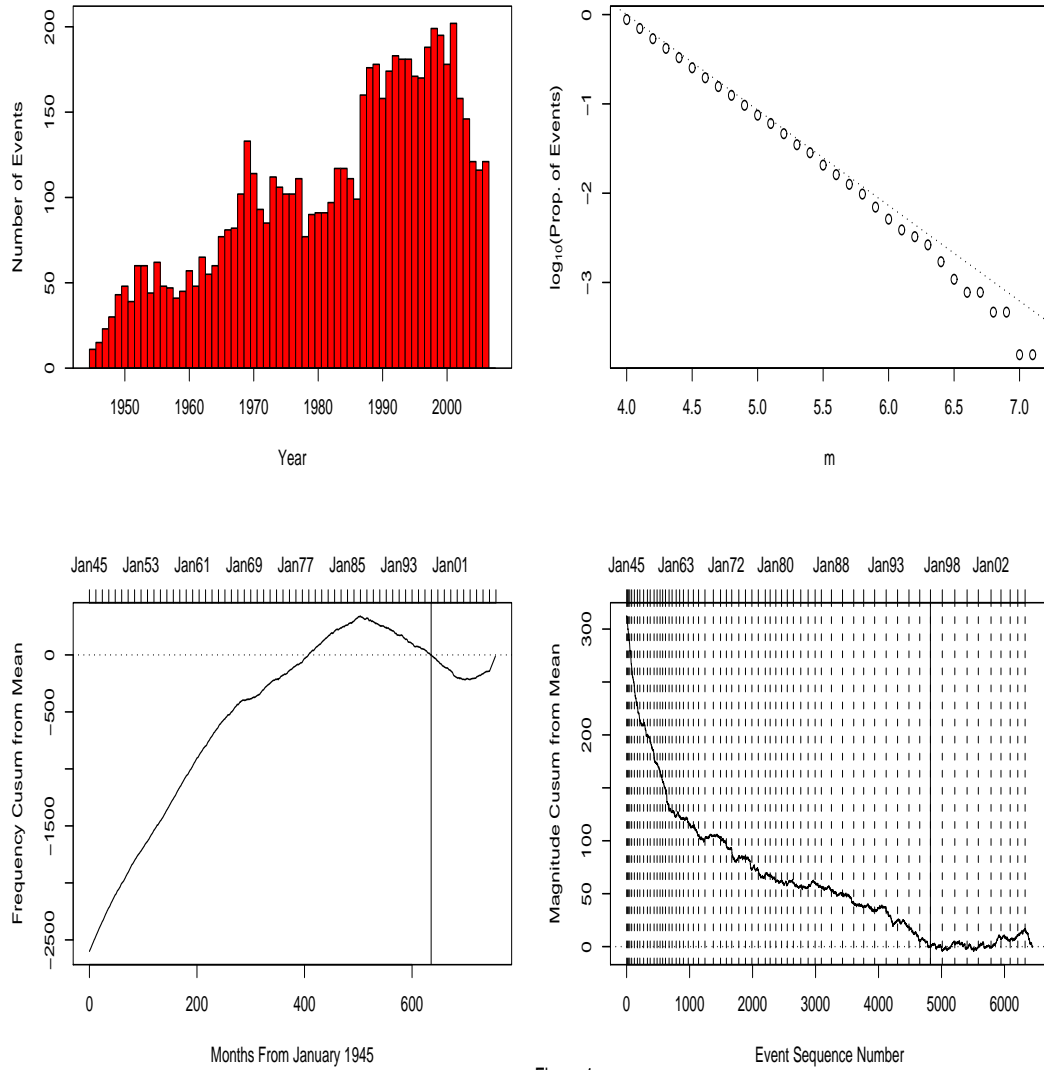


Figure 1

Figure 8.1: Events are selected from New Zealand catalogue within the confines defined in the Figure 8.4 and depth greater than 45km from Jan1 1950 up to Jan1 2007 with magnitude greater than 4. The solid lines mark the beginning of the reference period. Vertical lines mark the years.

Frequency, Magnitude and Cusum Statistics

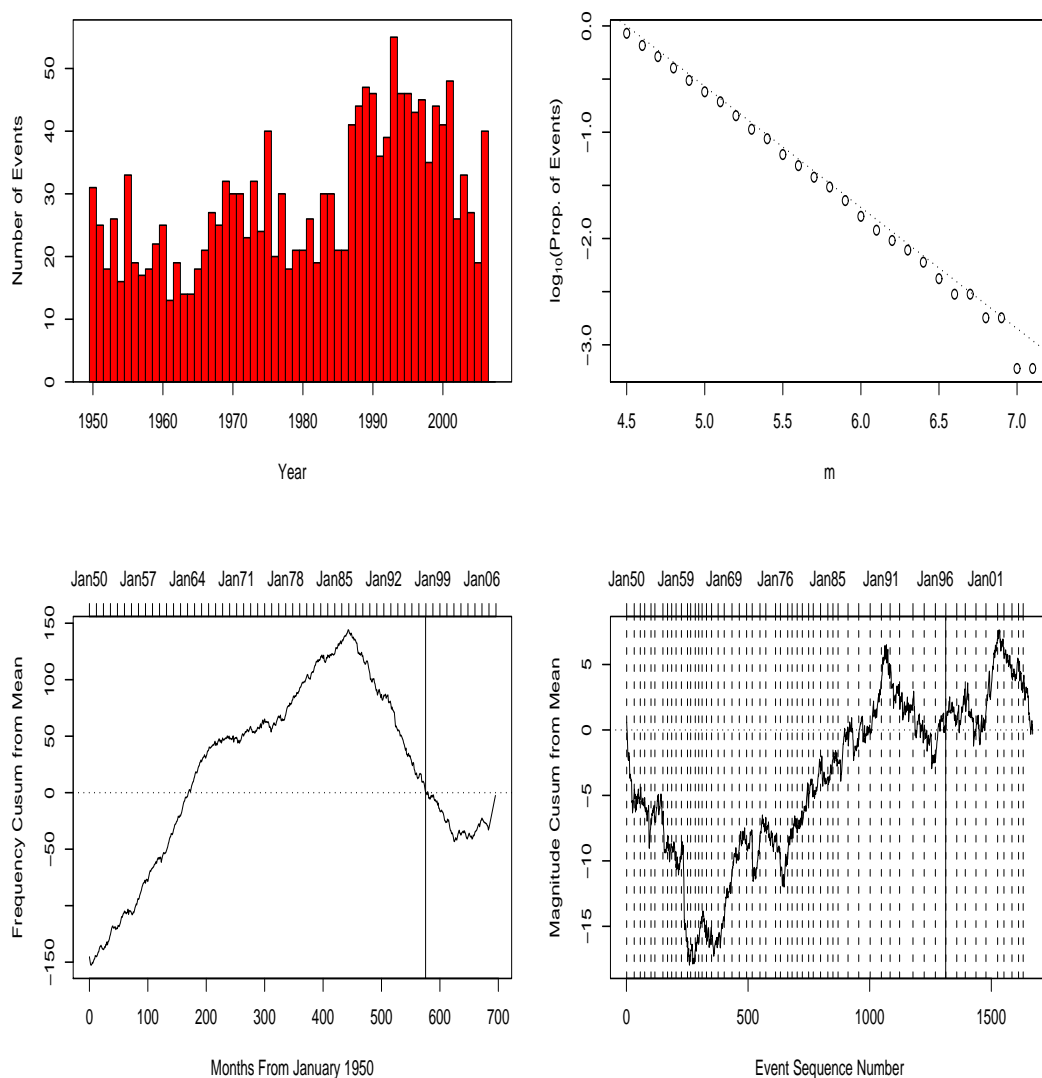


Figure 8.2: Events are selected from New Zealand catalogue within the confines defined in the Figure 8.4 and depth greater than 45km from Jan1 1950 up to Jan1 2007 with magnitude greater than 4.5. Vertical lines mark the years.

Frequency, Magnitude and Cusum Statistics

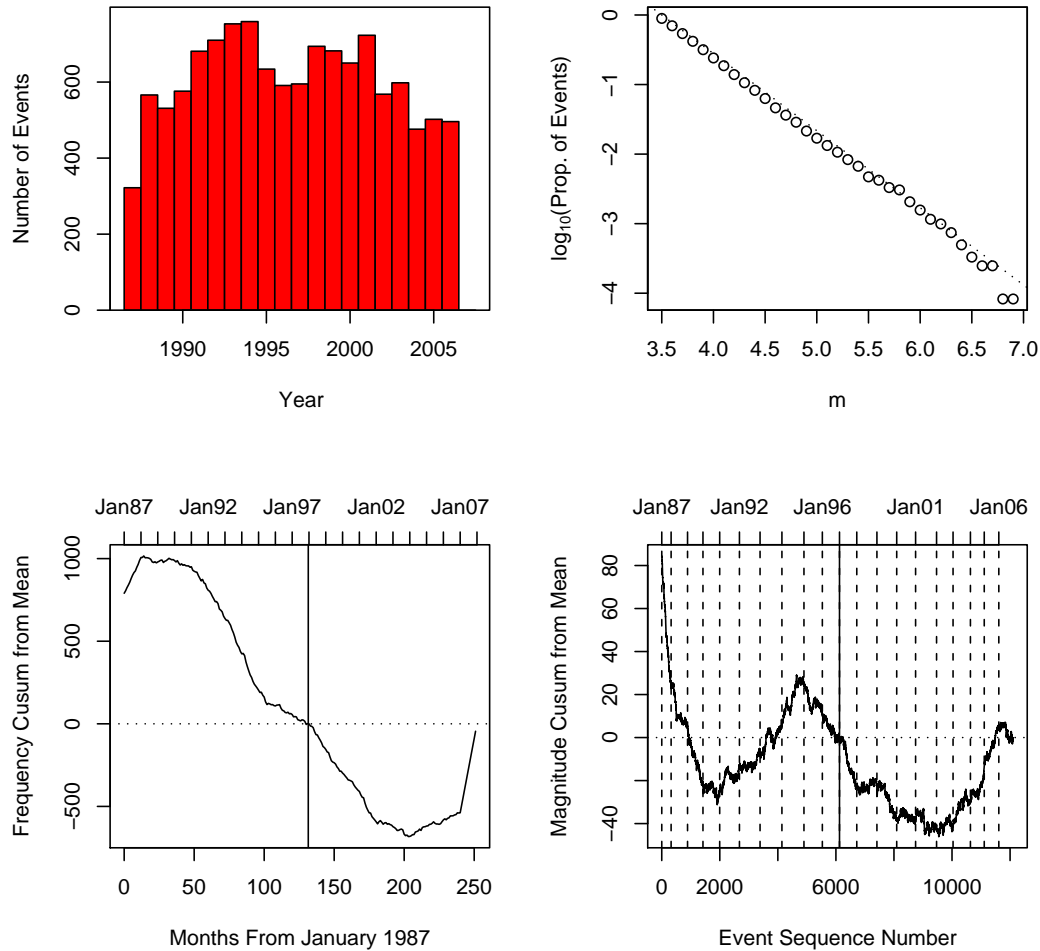


Figure 8.3: Events are selected from New Zealand catalogue within the confines defined in the Figure 8.4 and depth greater than 45km from Jan1 1987 up to Jan1 2007 with magnitude greater than 3.5. Vertical lines mark the years.

surrounding windows. These average values are assigned for each window. Among these groupings, the largest deep earthquake grouping extends from Taupo to the northeast of Bay of Plenty (BOP) with depth mainly ranging from 130km to 210km. Another conspicuous grouping is around the Taranaki region with depth between 200km and 250km. Other concentrations of deep seismicity are found at Cook Strait, south of Taranaki and elsewhere also, see Figure 8.4-5, left bottom of Figure 8.9 and 8.11. Readers who are unfamiliar with New Zealand geography can refer to Figure 2.1. Further exploratory data analysis shows that these deep groupings are subjected no obvious spatial migration and persist over time. However, temporally they vary from time to time at different levels over depth. For example, the frequency distribution of deep events shows an apparent increase in 1990-2007 at depth 40-170(km) and 170-230(km) mainly in the two largest groupings, Taupo-BOP grouping and Taranaki grouping, comparing the deep seismicity in 1970-1990, see Figure 8.8. Figure 8.9 shows that the frequency distribution of very deep earthquake with depth greater than 350km does not change very much over this two periods. One notable isolated group is beneath the Taranaki at great depth about 600km. However, for this very deep isolated group there still exist unsettled debates about whether the slab is continuous or it is a detached slab which separates from the frontier of the convergent plate and submerged into the mantel, see Figure 8.5.

At **most** subduction zones, the occurrence rates of deep earthquakes fall off roughly exponentially with depth until a small hump from nearly 500km to its deepest part which marks the termination of deep seismicity at this great depth. The scenario in New Zealand is a bit different for the apparent disproportionate in

intermediate depth seismicity which gradually comes to the culmination at nearly 170km and falls off abruptly at about 230km before 1980s. Afterwards, an active period of deep seismicity with depth less than 100km appears, see Figure 8.10. This trait also remains at the northeastern side of the North Island. However, the depth distribution of deep earthquakes in main groupings keeps constant, see Figure 8.11.

8.2.3 Magnitude distributions

The events number with varied magnitude is governed by the ubiquitous law, the Gutenberg-Richter (1956) frequency-magnitude relationship which asserts that the log number of earthquakes with magnitude greater than M is linear with respect to M , usually written as $\log_{10} N = a - bM$, where b is the so-called **b-value**.

Such a frequency magnitude log-linear relation sometimes deviates at certain magnitude ranges. For instance, it might fall off downwardly for relatively small events mainly due to the sparse monitoring station and limited detect-ability for small events in early period. Similar deviation often occurs for large events presumably due to some geophysical restrictions, such as finite fault width. A rectification of the magnitude-frequency relation at the tail via truncated Pareto-type distribution (Evans et al., 2000) gives more reasonable approximation to this deviation. Occasionally, a deviation from the frequency magnitude linear relation for large events with reverse hump appears in some local regions. This abnormal might be ascribed mainly for the sub-setting effect both temporally and spatially.

The comparative proportion of events with different magnitude is determined

**Epicenter Distribution of main deep earthquakes
in the North Island**

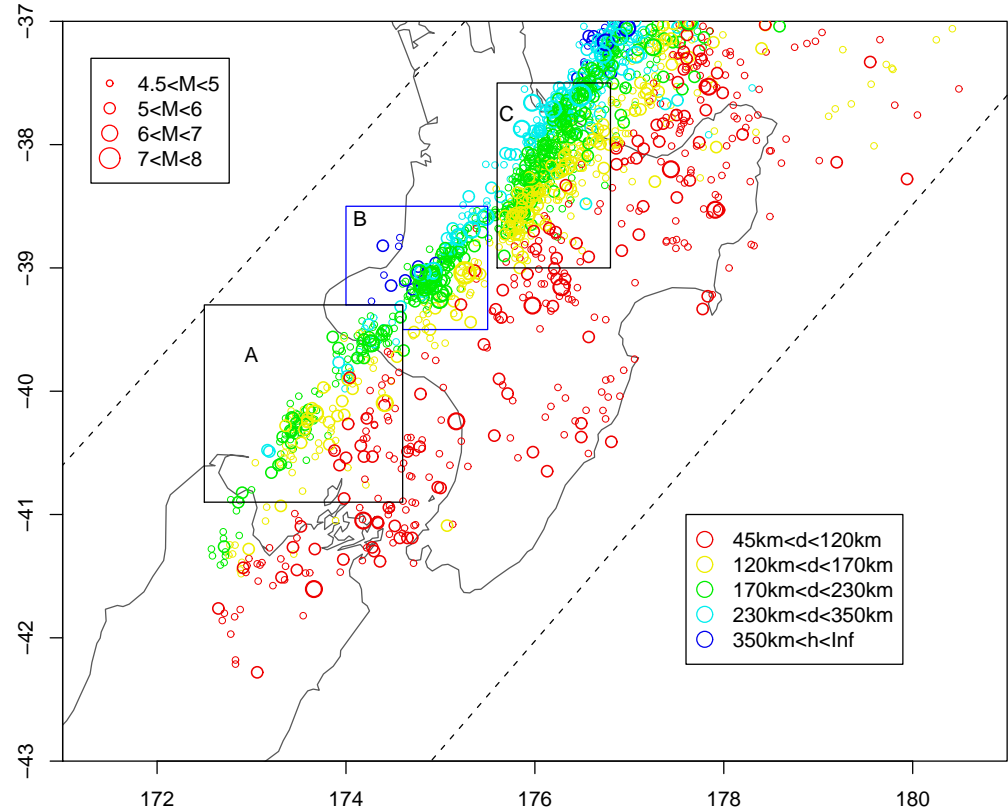


Figure 8.4: Epicenter distribution of deep earthquakes from 01Jan 1950 to 01Jan 2007 with magnitude greater than 4.5 around north of the South Island and the North Island. Events encircled by dashed line and map boundary are considered in good coverage. A, B and C denote Cook Strait deep grouping, Taranaki deep grouping and Taupo-BOP deep grouping respectively.

Section plot of Deep Earthquakes

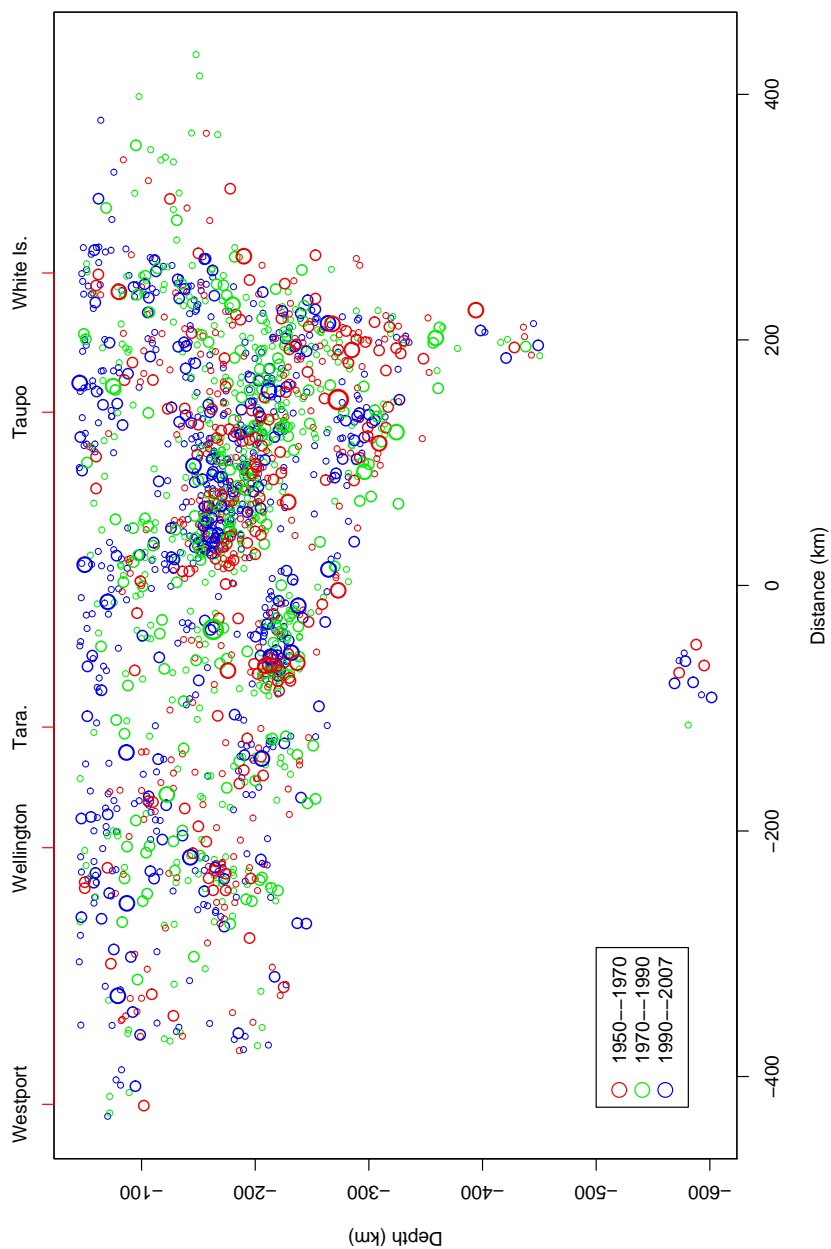


Figure 8.5: A southwest-northeast depth section plot of earthquakes deeper than 40km from 01Jan 1945 to 01Jan 2007 with magnitude greater than 4.5 around north of the South Island and the North Island. The section azimuth is 45° .

Section plot of Deep Earthquakes

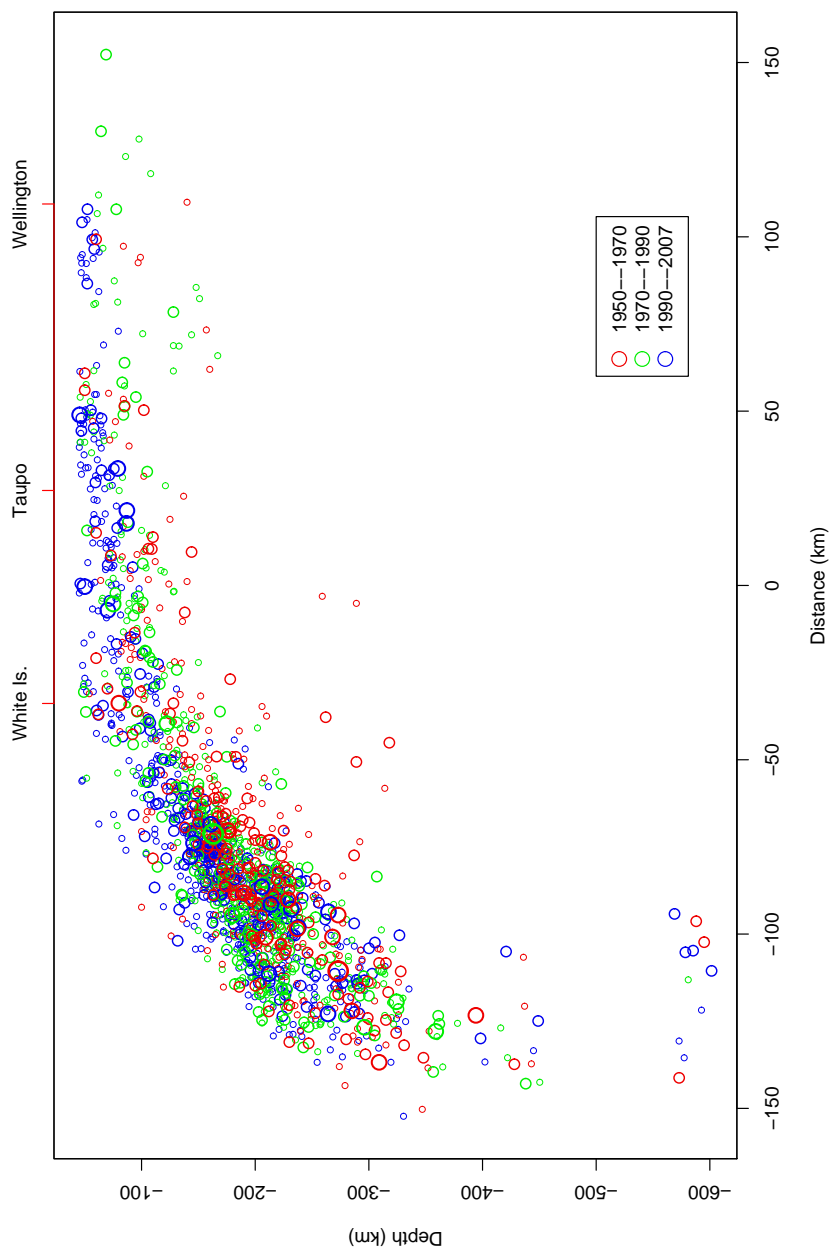


Figure 8.6: A Northwest-southeast depth section plot of deep events from 01Jan 1945 to 01Jan 2007 with magnitude greater than 4.5 around north of the South Island and the North Island. The section azimuth is 135° .

Section plot of Deep Earthquakes

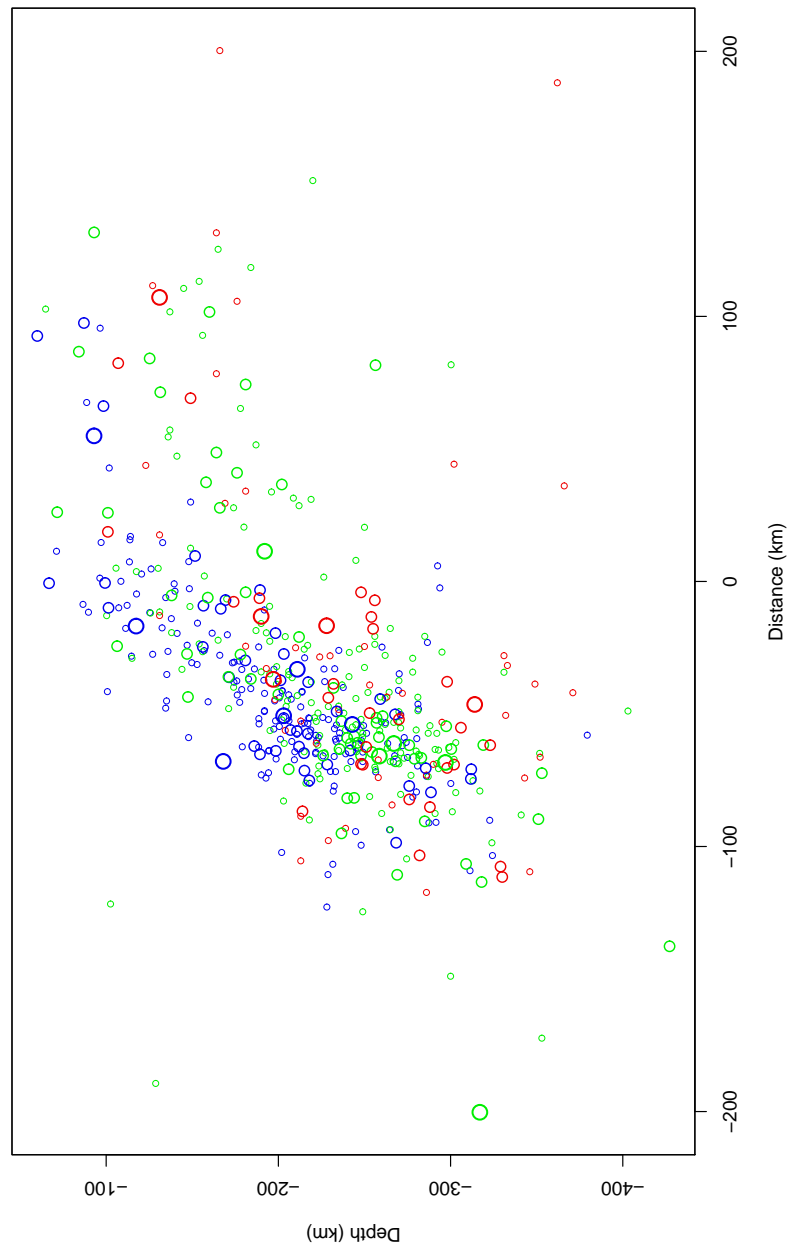


Figure 8.7: A Northwest-southeast depth section plot of deep events from 01Jan 1945 to 01Jan 2007 with magnitude greater than 4.5 and latitude greater than -37° . The section azimuth is 135° .

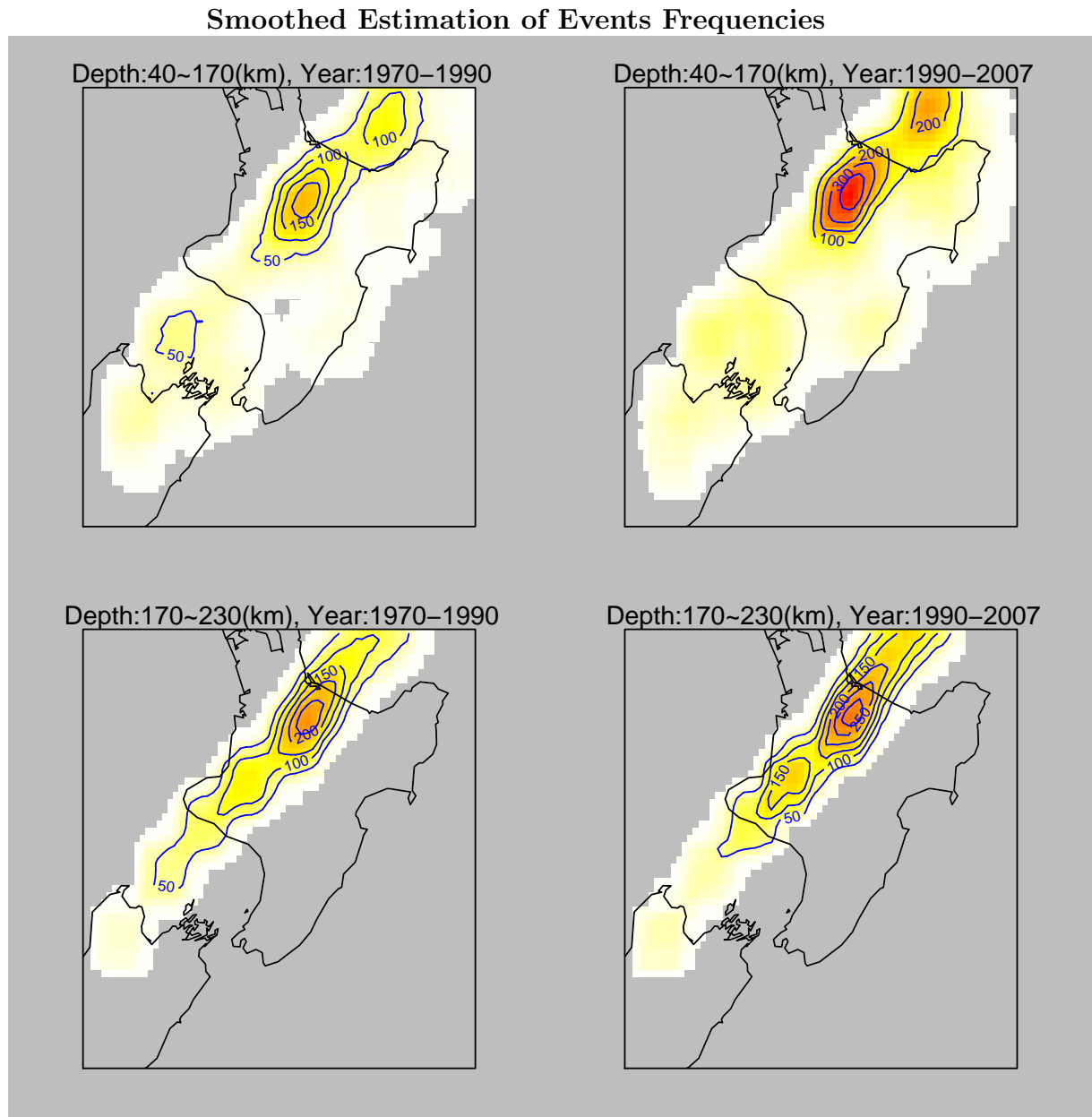


Figure 8.8: Smoothed estimation of event frequencies (total events number) for two periods 1970-1990, 1990-2007 with depth between $[45, 170](\text{km})$, $[170, 230](\text{km})$ and magnitude greater than 4.5.

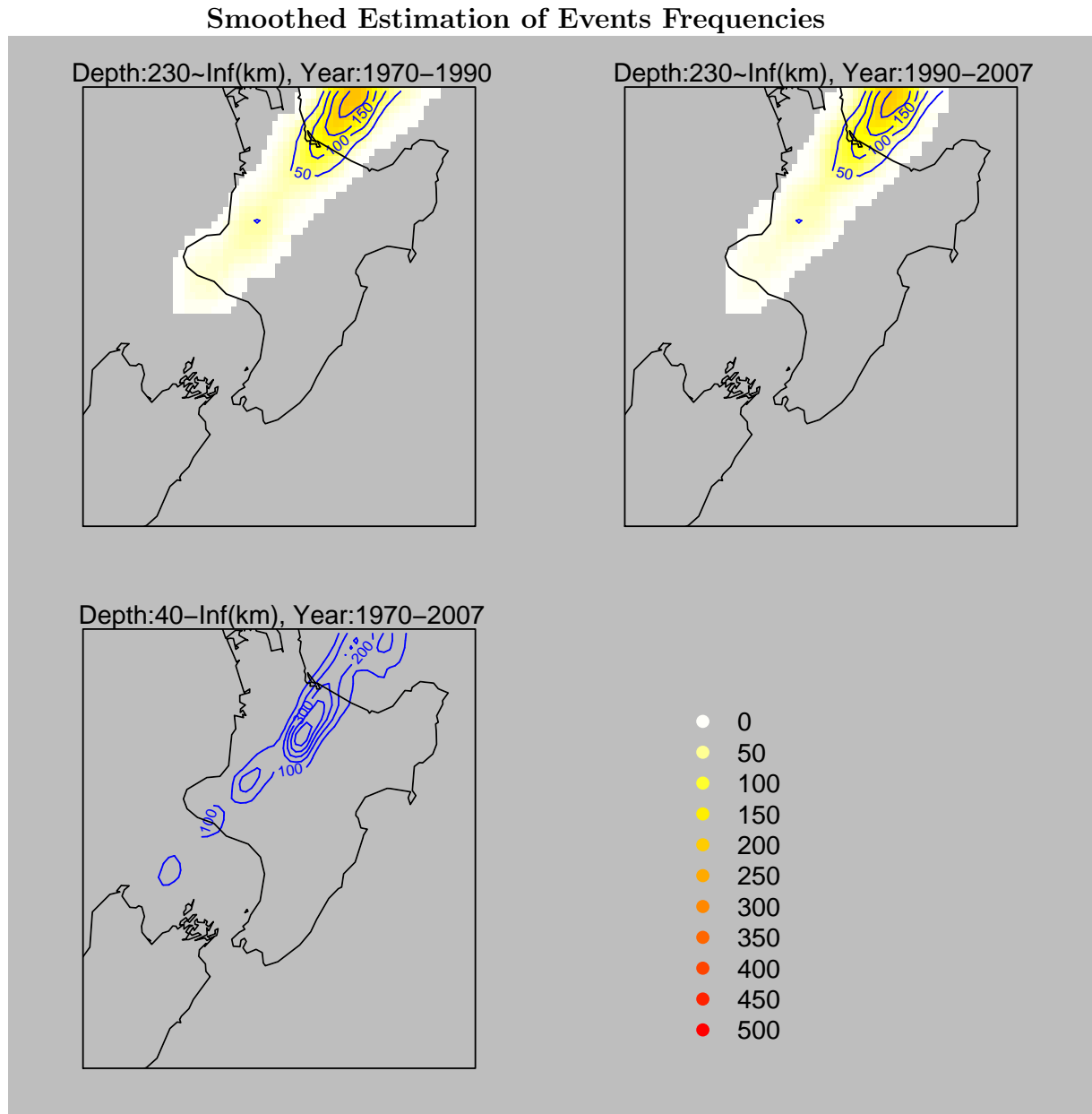


Figure 8.9: Smoothed estimation of event frequencies (total events number) for two periods 1970-1990, 1990-2007 with depth greater than 230km and magnitude greater than 4.5 in first row. The bottom plots show overall estimation of event frequencies and the palette.

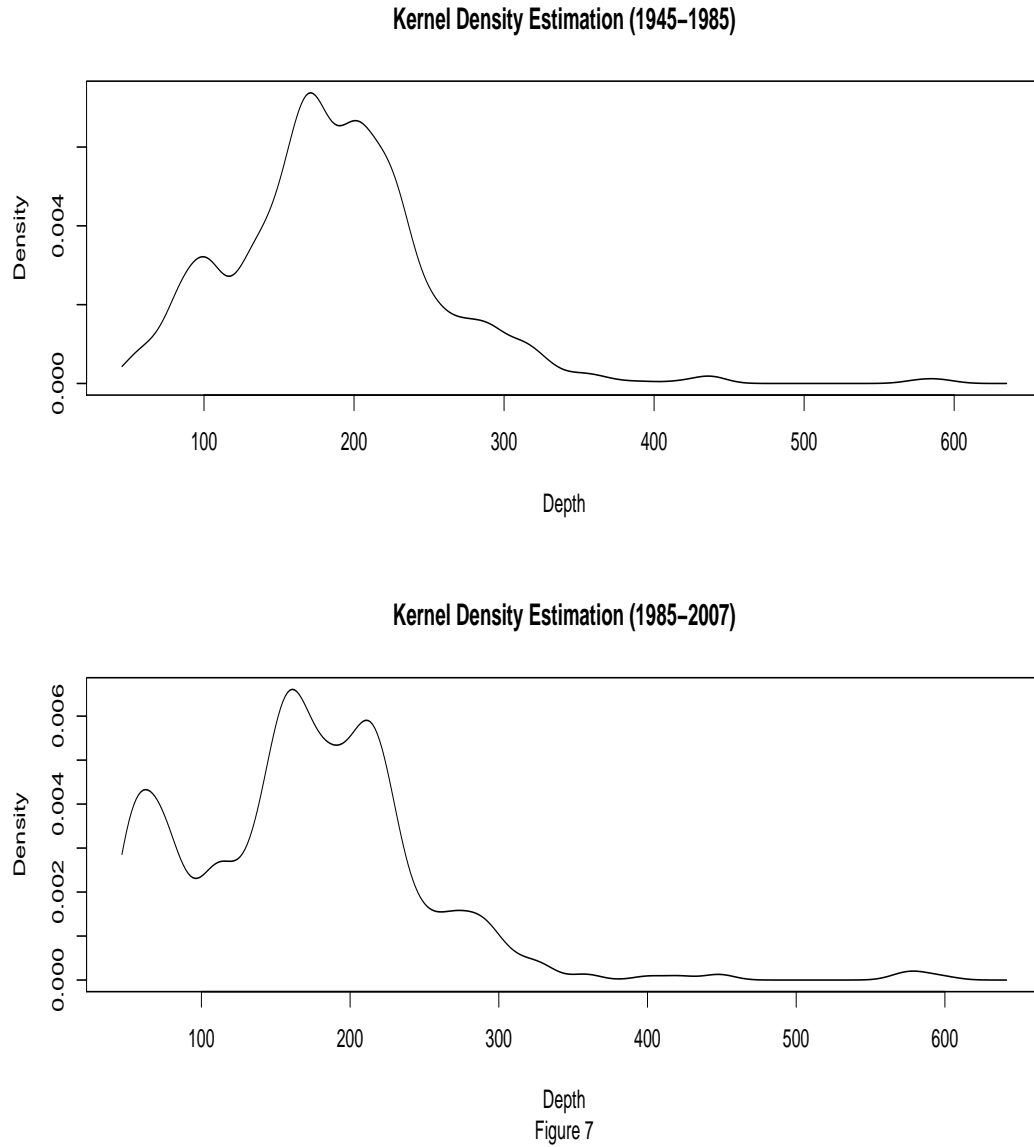


Figure 8.10: Kernel density estimation of the depth distribution of deep events with magnitude greater than 4.5 and latitude ranged in $[-43, -37]$. Normal kernel function is used in the smoothing procedure. The bandwidth used in the kernel function is $h = 1.06\sigma n^{-\frac{1}{5}}$, where σ is the standard error and n is the data size.

Histogram of the depth in main groupings

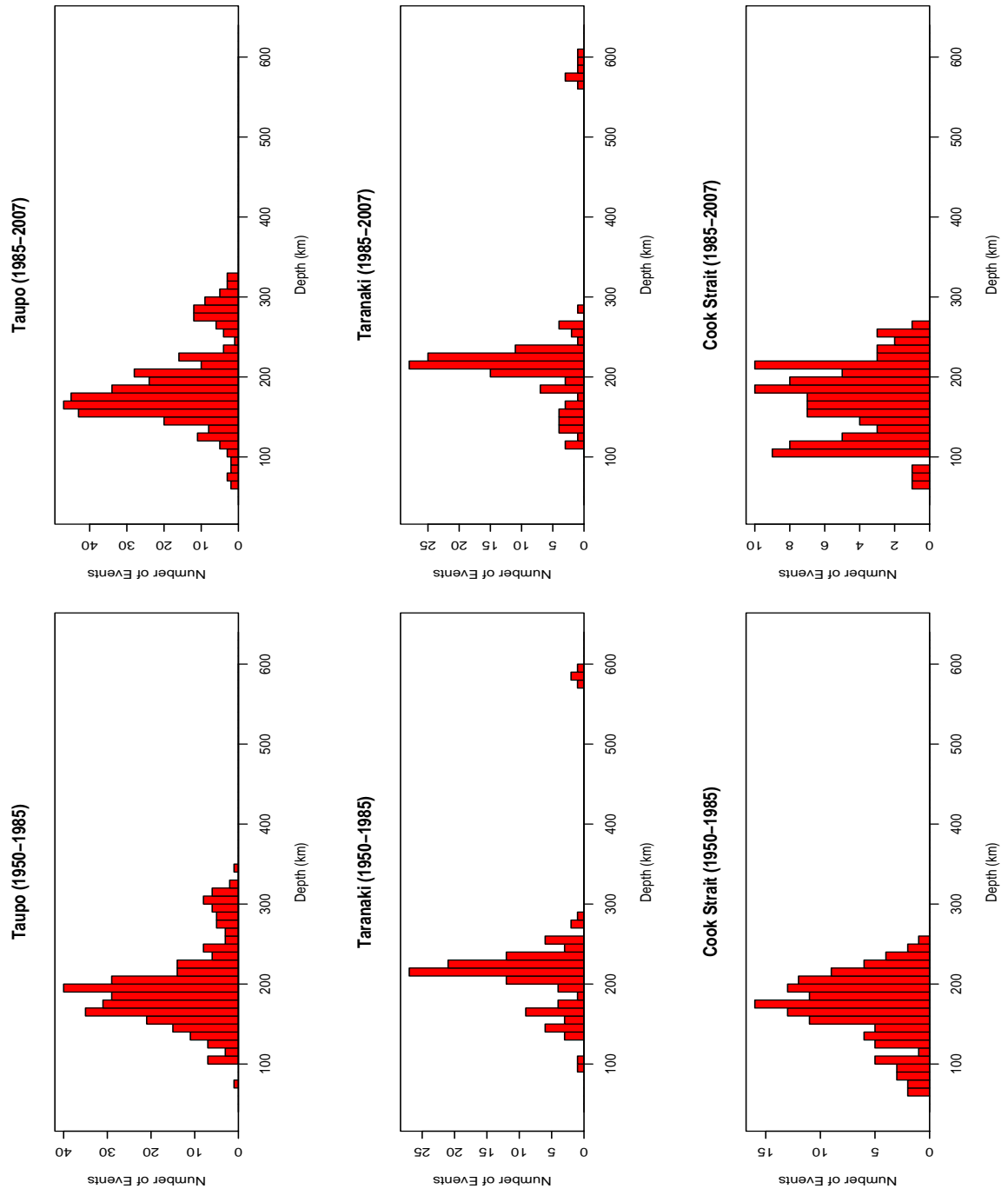


Figure 8.11: Depth distribution of deep events in three groupings in different periods.

by the b-value statistics in the Gutenberg-Richter law. The systematic variations of b-value over different regions and its relationship with the thermal parameter (the product of age of the subducted lithosphere and the vertical component of convergence rate of the plate) of the slab is given by the Frohlich (1989). The local variations of b-value with respect to depth in the North Island has been tabulated by Gibowicz (1974) for 50 km depth intervals, in steps of 25 km during 1955-1967 with magnitude greater than 3.9. The b-value tends to vary from 1 for shallow events to the culmination at about 100km and falls off gradually below 1 at greater depth. Such a trend of b-value variation over depth has also been found in other places outside of New Zealand. Studies over b-value anomaly at about 100 km depth on the upper surface of Wadati-Benioff zone and its association with volcanism is addressed by Wiemer and Benoit (1996). They also suggest an explanation of this b-value increasing by the slab dehydration and its associated effect, such as increasing pore pressure and lowering effective stress.

We tabulate the b-value variations over depth for deep events from 1989 to 2006 with magnitude greater than 3.5 in three groupings, roughly Cook Strait grouping, Taranaki grouping and Taupo-BOP grouping defined in Figure 8.4, see Table 8.1. The b-value tends to decrease over depth systematically. We also list the 95 percent error bound of estimated b-values according to Aki's (1995) method by utilizing the standard asymptotic normal theory of the maximum likelihood estimates for the b-value. Actual error bounds might be greater due to other errors such as rounding error, measurement error and even instrumental effect, see following Figure 8.12. The slope of cumulative magnitude is steeper than before at late 1960s and 1980s

b-value	Cook Strait	Taranaki	Taupo-BOP
40–100(km)	1.52 ± 0.22	NA	1.49 ± 0.23
100–170(km)	1.17 ± 0.1	1.34 ± 0.16	1.22 ± 0.05
170–230(km)	1.10 ± 0.09	0.97 ± 0.07	1.09 ± 0.05
230–350(km)	1.11 ± 0.27	1.03 ± 0.15	0.88 ± 0.09

Table 8.1: The b-values of earthquakes ($M \geq 3.5$) in individual groupings between [1989, 2006] with 95 percent confidence bounds. The b-values change with respect to depth systematically.

right at the upgradation periods of monitoring networks.

Generally speaking, the largest deep earthquakes are smaller than the largest shallow earthquake and there is no obvious indication that the size of the largest earthquakes varies over depth, see Frohlich (2006). However, the number of the very large deep earthquakes with magnitude greater than 7 in New Zealand catalogue is much more than that of the shallow earthquakes, an apparently disproportionate in contrast to the relatively small earthquakes. Nearly all of them are located beneath the surface at least 150 km. For instance, among 20 earthquakes with magnitude greater than 7, only 3 events are shallow earthquakes.

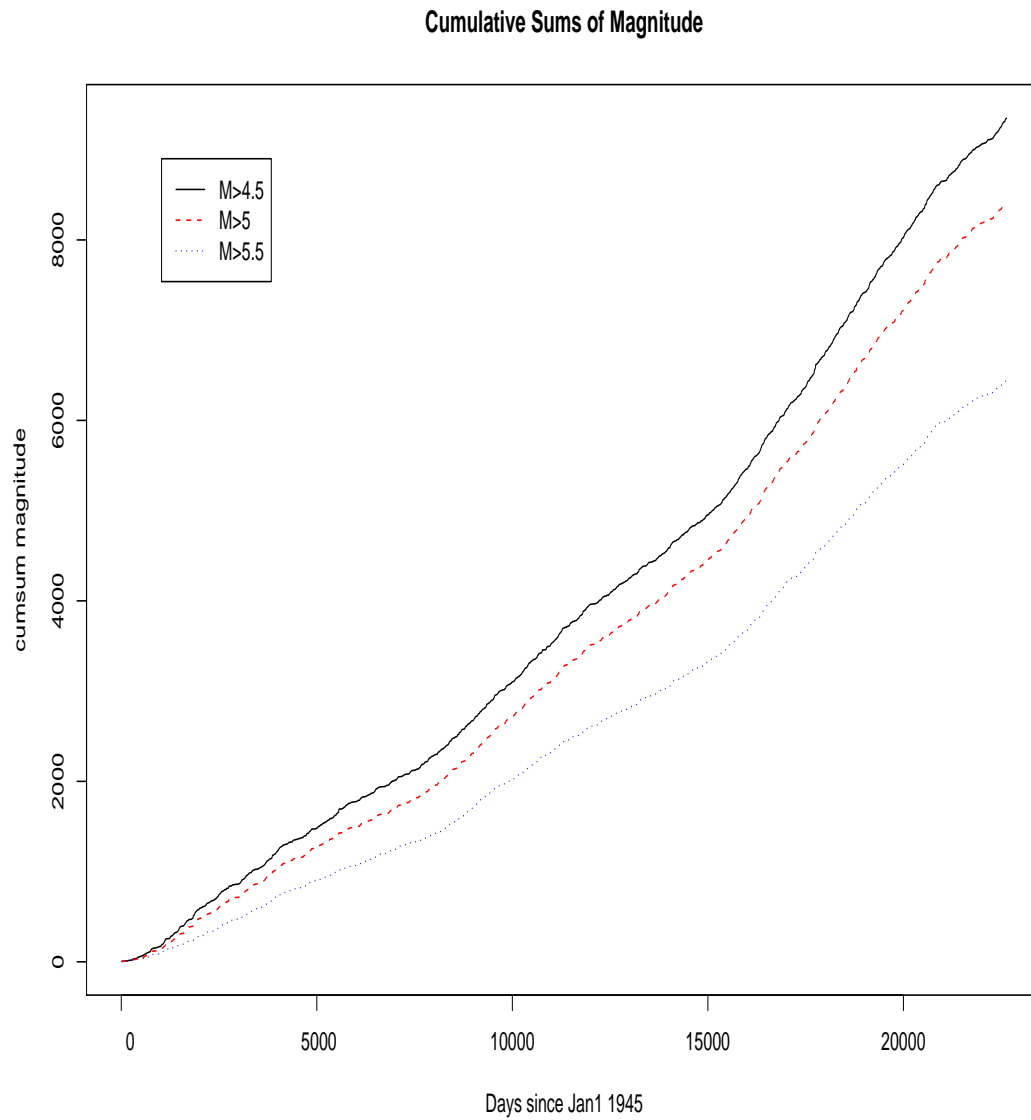


Figure 8.12: Cusum magnitude plot for deep events around the North Island from 01Jan 1945.

8.2.4 Clustering properties

It is well-known that the deep earthquakes are generally lack of fully developed aftershocks which decay in terms of Omori's law. Furthermore, It is more restrictive to effectively analyze aftershock clustering due to lack of statistical models for the deep earthquake clustering. Fortunately, a number of techniques are available to analyze the clustering properties of deep earthquakes through variance-time function, spectrum et al.. Among these second order moments, the two-point correlation function which is defined as the local intensity function at y given a point occurring at x such as $h(y|x)dy \triangleq E\{N(dy)|N(dx) = 1\}$ is very useful. When the two-point correlation function $h(y|x)$ is divided by the unconditional intensity $m(y)$, it yields the relative conditional intensity $p(x, y) \triangleq \frac{1}{m(y)}h(y|x) = 1 + \frac{c(x, y)}{m(x)m(y)}$, where $c(x, y)$ is the covariance density of the counting process $N(x)$. It equals unity for a Poisson process. The clustering or regular pattern is characterized by the deviations from unity, either greater than 1 or less than 1 respectively, see Vere-Jones (1978).

Let $N(I)$ denote the number of earthquakes with magnitude in category I and $N(I, J, K, L, M)$ denote the number of paired events with magnitude of reference events in category I and magnitude of related post events in category J within K th time delay category, L th distance category and M th depth category. Also let $A(K, L, M)$ and A denote the volume of (K, L, M) th cell and the total volume respectively. Then, a discrete version of the relative conditional intensity when assuming the seismicity is stationary in time and homogeneous in space is given by

$$p(m_I, m_J, t_K, r_L, h_M) = \frac{N(I, J, K, L, M)}{N(I)N(J)} \times \frac{A}{A(K, L, M)},$$

where $\frac{N(I,J,K,L,M)}{N(I)A(K,L,M)}$ gives the average number of post-events with magnitude in category J per unit volume in (K, L, M) th cell from a reference event with magnitude in category I and $\frac{N(J)}{A}$ gives the average number of events per unit volume in general.

The range of magnitude and time delay from a reference event are partitioned into (4, 4.4, 4.8, 5.2, 6, Infinity) and (0, 15, 30, 90, 180, 360, 1080, 1826)(days) respectively. Similarly, the spherical distance obtained from Haversine formula and the distance in depth h from a reference event both are categorized into (0, 15, 45, 90, 120)(km) in two directions, upwards and downwards.

We apply the methods to the deep events around the North of the South Island and most of the North Island. The general characteristic indicated in this analysis is lack of post-event effect (clustering effect) comparing that of the shallow earthquakes at the same period. A search for the post-event effect occurring among the cells indicates that the deep events do have small portion of aftershocks in the cell where the reference event magnitude is greater than 6 and the post-event magnitude is between 4.4 and 4.8 within 15 days time delay and 15km spherical distance, suggesting rather weak post-event effects comparing that of shallow earthquakes, see Table 8.2. A similar analysis for the same reference events but smaller post-events with magnitude between 4-4.4 shows no significant clustering effect appearing, see Table 8.3. Also, similar analysis for deep events in an extended period from 1970 to 2007 suggests that even such a small scale clustering effect is not time homogeneous, probably only limited in some individual hot spots after 1990, see Table 8.4. Such a clustering effect in deep events is also magnitude dependent for the smaller reference events with magnitude less than 6, the estimated relative conditional intensities in

time distance	0-15(km)	15-45(km)	45-90(km)	90-120(km)
0-15(days)	3.85	0.00	0.00	0.64
15-30(days)	0.00	0.00	0.00	0.00
30-90(days)	0.00	0.16	0.11	0.00
90-180(days)	0.21	0.21	0.14	0.00
180-360(days)	0.00	0.05	0.14	0.00
360-1080(days)	0.13	0.07	0.08	0.07

Table 8.2: Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 6$, $4.4 \leq m_J \leq 4.8$ and $0 < h < 15$ (km), where h is the distance in depth. The event selected are from 01Jan1991 to 01Jan2007 with latitude between $[-44, -36]$ and depth greater than 45km.

nearly all cells are less than 1, see Table 8.5. In other cells, no obvious clustering effect is presented.

Occasionally, some deep earthquake clusterings form as multiplets, triplets or doublets. For instance, intermediate large deep doublets are found beneath Taranaki and elsewhere in 23th and 27th March 1960 (Adams et al., 1963), 12th January, 1950 and 5th May, 1968. Among these intermediate or large doublets, the inter-occurrence time ranges from only seconds to a couple of minutes with very small magnitude difference and spatial distance.

time distance	0-15(km)	15-45(km)	45-90(km)	90-120(km)
0-15(days)	0.42	0.11	0.00	0.11
15-30(days)	0.21	0.21	0.21	0.00
30-90(days)	0.11	0.11	0.12	0.03
90-180(days)	0.04	0.18	0.07	0.05
180-360(days)	0.12	0.08	0.10	0.07
360-1080(days)	0.08	0.11	0.08	0.06

Table 8.3: Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 6$, $4 \leq m_J \leq 4.4$ and $-15 < h < 15(\text{km})$, where h is the distance in depth. The event selected are from 01Jan1991 to 01Jan2007 with latitude between $[-44, -36]$ and depth greater than 45km.

time distance	0-15(km)	15-45(km)	45-90(km)	90-120(km)
0-15(days)	1.22	0.00	0.00	0.20
15-30(days)	0.00	0.00	0.00	0.00
30-90(days)	0.00	0.10	0.03	0.00
90-180(days)	0.07	0.07	0.05	0.00
180-360(days)	0.07	0.05	0.08	0.00
360-1080(days)	0.05	0.04	0.04	0.03

Table 8.4: Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 6$, $4.4 \leq m_J \leq 4.8$ and $0 < h < 15(\text{km})$ in an extended time period. The event selected are from 01Jan1970 to 01Jan2007 with latitude between $[-44, -36]$ and depth greater than 45km.

time distance	0-15(km)	15-45(km)	45-90(km)	90-120(km)
0-15(days)	0.032	0.039	0.038	0.041
15-30(days)	0.021	0.039	0.037	0.036
30-90(days)	0.017	0.038	0.037	0.036
90-180(days)	0.017	0.035	0.039	0.038
180-360(days)	0.017	0.035	0.037	0.038
360-1080(days)	0.016	0.034	0.037	0.037

Table 8.5: Estimated $p(m_I, m_J, t_K, r_L, h_M)$ for $m_I \geq 5.2$, $4 \leq m_J \leq 5.2$ and $-90 < h < 90(\text{km})$, where h is the distance in depth. The event selected are from 01Jan1970 to 01Jan2007 with latitude between $[-44, -36]$ and depth greater than 45km.

8.3 Models and model-fitting

We assume the deep seismicity is stationary throughout the discussion, i.e., the stochastic laws do not change for an arbitrary time translation.

8.3.1 Poisson, stress release and simple ETAS models

It is worth to apply Poisson model to the deep earthquakes, the simplest model for sequential occurrence events which is usually treated as a null hypothesis compared with some alternatives according to their forecasting performance or goodness-of-fit test etc. However, the occurrence frequency plots in previous section shows that the yearly counts of deep seismicity varies from time to time which should not be able

to be characterized by Poisson process. Later analysis will confirm this viewpoint.

The stress release model (Xiaogu and Vere-jones, 1991) is a stochastic version of elastic rebound model which utilizes a notational regional stress level $Z(t)$ at time t as a state variable to describe the time-varying risk of earthquake occurrence, i.e., the conditional intensity function $\lambda(t)$ in the context of point process. In a more detailed description, the controlling role in stress release model is a process which indicate the regional stress level. The evolution of the process is controlled by two components. The deterministic component is given by the loading stress which is supposed to be linear over time for lack of prior information in the stress field, i.e. $Z(t) = Z(0) + \rho t$ and ρ is the loading rate of the tectonic stress. The stochastic component is given by the accumulative stress release through large earthquakes given by $S(t) = \sum_{i:t_i \leq t} S_i$, where t_i is the occurrence time of the i th earthquake and S_i is the amount of released stress through the i th earthquake. The stochastic behavior of the observed occurrences is determined by the conditional intensity function of the point process which is assumed to be exponential with respect to the underlying stress process, i.e. $\lambda(t) = \exp\{Z(0) + \rho t - \sum_{i:t_i < t} S_i\}$. Such a statistical simplification of elastic rebound model is useful for evaluating the long term or middle term risks of large shallow earthquake occurrence.

However, the interpretation of stress release model is complicated by some occasionally large doublets, triplets or multiplets when it is applied to the deep events. In other words, the occurrence pattern and underlying geophysical process of large deep doublets, triplets and multiplets is hard to interpret by original Stress Release model since it generally requires relatively long time to accumulate enough tectonic

stress energy to trigger a large shallow earthquake. It also remains doubtful that by looking only at large events the dynamics of main stress field and deep seismicity on the slab can be effectively described by the stress release model, since the great mass of small and medium size deep events are not apparent aftershocks of large deep earthquakes in the sense of the shallow earthquakes. More importantly, it is unlikely that the main factor of deep earthquake occurrence is only stress state rather than other factors such as temperature, age, composition of the slab and convergence rate of the tectonic plates etc..

ETAS (Epidemic Type Aftershock Sequence) model (Ogata, 1988, 1998) is a self-exciting point process model based on the classical aftershock statistics and several simple model assumptions. It assumes that total seismicity is classified into two types, the background activity and aftershock activity. The background seismicity is a stationary Poisson process with intensity rate μ and the triggering intensity of the aftershocks in the model is parameterized according to Omori's law (1894) with the assumption that every earthquake, whether a main shock or an aftershock, has same triggering power to trigger aftershocks. Hence, the triggering intensity rate of the model is parameterized by $\sum_{i:t_i < t} \frac{Ae^{\alpha m_i}}{(t-t_i+c)^p}$, where t_i and m_i are the occurrence time and magnitude of the i th earthquake respectively, c and p are time-delaying constant and the decaying rate in the Omori's law respectively. A is a constant. Hence, the intensity rate of the ETAS model is written by $\lambda(t) = \mu + \sum_{i:t_i < t} \frac{Ae^{\alpha m_i}}{(t-t_i+c)^p}$.

However, one of the prominent characteristics of the deep earthquakes is that they seldom have a great number of small aftershocks which decay according to the Omori's law. The deep earthquakes which have more than one aftershock are not so

common as a rule and the abundance of aftershocks of the deep earthquakes tends to vary over depth, see (2006). Also, another typical clustering behavior reported in the deep earthquakes is multiplets, mostly doublets rather than aftershock clusters. Such factors do not reasonably justify the utilization of ETAS model to describe the temporal and spatial clustering properties of the deep earthquakes.

When we apply the ETAS model to the deep earthquakes in several main groupings such as Taupo-BOP grouping and Taranaki grouping or as a whole in a longer time period or a shorter time period with smaller magnitude cutoff, in many cases, no satisfying convergence results can be obtained for the parameters of the simple ETAS model when utilizing numerical optimization methods to maximize the likelihood. When the numerical optimization procedure converges, some estimated parameters are not in reasonable domains. For example, in most cases, the estimated p values associated with Omori's law are less than 1 which strongly suggests the lack of aftershocks in the sense of Omori's law.

All in all, the lack-of-fit of the models usually used for the shallow earthquakes strongly suggests that the occurrence patterns of the deep earthquakes are essentially different from that of the shallow earthquakes.

Chapter 9

Deep Earthquakes Modeling

9.1 Application of MMPP to the Deep Earthquakes

From chapter 8, it is suggested that the widely used models for the shallow earthquakes such as the ETAS models (Ogata, 1988, 1998) and the stress release models (Zheng and Vere-Jones, 1991) fit the occurrence patterns of the deep earthquakes unsatisfactorily. The lack of fit of these models is partly due to the fact that the deep earthquakes rarely have a following sequence with numerous small aftershocks. Some deep earthquakes even do not have aftershocks at all. It is hard to say whether the time-decaying behavior of the aftershocks in the deep events conforms to Omori's law as the shallow earthquakes do. Instead, the main evolution feature of the deep earthquakes is that the deep seismicity varies from time to time, active in one period,

relatively quiescent in another. We suggest a switching Poisson model (the second order MMPP) to characterize the time-varying behavior of the deep earthquakes. The two levels of the deep seismicity are designated as the seismic active state and the seismic quiescent state respectively in terms of two levels of Poisson rates in a MMPP.

The data used in this study starts from 1 Jan 1945 to 1 Jan 2007 covering main deep groupings such as Taupo-BOP grouping, Taranaki grouping and Cook Strait grouping as defined in Figure 9.1 with focal depth greater than 40km and magnitude greater than 4.5. Generally, The catalogue data used in this analysis are in good coverage by the monitoring networks, see exploratory data analysis of the deep earthquakes in the section 8.2.

In this section, the main groupings of the deep earthquakes as a whole around the North Island is fitted by a switching Poisson process (the second order MMPP). When the deep seismicity is active or relatively quiescent is indicated by using a smoothing procedure for the inference of the state process of the MMPP. Keep in mind that when the deep seismicity is in a seismic active state, the second state in this model, at least one deep grouping must be in seismic active state. Otherwise, the deep seismicity in most deep groupings should be in seismic quiescent state (the first state). Then, a detailed analysis is carried out to investigate where and when the seismicity changes in each grouping. Here, we will not analyze the events in region D (Bay of Plenty grouping) through MMPP because the data might be incomplete and the magnitudes are generally inaccurate in the early period.

Main Groupings of Deep Earthquakes

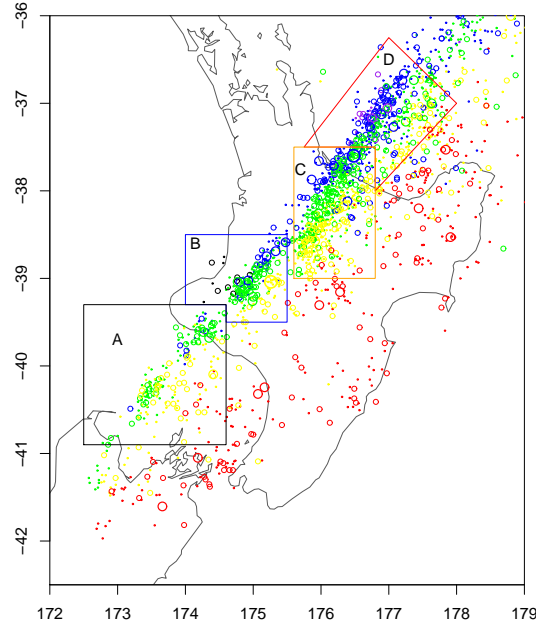


Figure 9.1: Partitions of main deep groupings. A, B, C and D represent Cook Strait grouping, Taranaki grouping, Taupo grouping and BOP (Bay of Plenty) grouping respectively.

The initial values used in the EM iteration procedure set out in chapter 3 should be selected carefully. A natural approach is to approximate the continuous time process by time discretization, see Deng and Mark (1993). In this procedure, the time interval is divided into small bins and the number of arrivals in each bin is counted. By assuming the state transitions occur only at bin boundaries, the model can be treated as a discrete-time hidden Markov model with Poisson observations. Then, the transition rates q_{ij} and the intensity rates λ_i can be straightly converted

from the discrete-time hidden Markov model, i.e. $q_{ij} = p_{ij}/h$ and $\lambda_i = b_i/h$, where p_{ij} is the state transition probability of the discrete-time Markov chain, b_i is the Poisson parameter of the discrete-time hidden Markov model and h is the bin width. In the procedure, the EM iterations terminate when the increments of the likelihood is less than 10^{-8} . We also use different initial values in the EM iteration steps to validate the convergence results. Generally speaking, the initial values have no significant influence on the convergence results and the iteration solutions begin to converge within hundreds steps. The execution of the algorithm depends on CPU and the codes, usually taking several minutes for a personal computer.

The probabilities of the underlying Markov chain in the second state (the seismic active state) are obtained by using the so-called fixed point smoothing algorithm which gives the conditional probabilities of $X(t)$ in state 2 at time t conditioned on all available observations, see section 3.3 for the derivation of the algorithm. These probabilities are evaluated at many pre-selected grid points and are connected by straight lines. Since the smoothing estimation of the probabilities of $X(t)$ in a specific state is continuous with respect to t , this approach is sufficient to demonstrate the evolution of $X(t)$. The magnitudes versus occurrence times, the estimated probabilities of $X(t)$ in state 2 and the estimated intensity rates for events from all groupings by using equation (3.8) are demonstrated in Figure 9.2. The top part of Figure 9.2 shows the occurrence times and magnitudes of the deep earthquakes. The middle part indicates when the seismicity is active or relatively quiescent. The bottom of the figure gives the estimated occurrence rates of the series of events appearing in the top of the figure. The deep seismicity demonstrates greater variability

before 10,000 days since Jan 01, 1945 in comparison to the seismicity in the later period. The deep seismicity after late 1980s shows an unusual active period without ending signs. Similar pictorial representation of the deep seismicity in individual groupings A, B, C is presented in Figure 9.3-9.5. From Figure 9.3-9.5, it is clear that the deep seismicity shows similar trends in individual groupings after 5000 days since 01 Jan, 1945, see the estimated probabilities of the underlying Markov chain in state 2 (seismic active state) and the estimated intensity rates of the observed point processes for the three individual groupings. The deep seismicity around 10,000 days and after 15,000 days since Jan 01, 1945 is relatively active in all the three groupings. The deep seismicity in grouping B (Taranaki grouping) shows less variations than the other two groupings. The deep seismicity in grouping C (Taupo grouping) varies a bit more frequently than the other two groupings, see also Table 9.2. Note that according to equation (3.8), the estimated intensity rate is $\hat{\lambda}(t) = \hat{\lambda}_1 \hat{p}_t(1) + \hat{\lambda}_2 \hat{p}_t(2)$. It is equal to $(\hat{\lambda}_2 - \hat{\lambda}_1) \hat{p}_t(2) + \hat{\lambda}_1$ since $\hat{p}_t(1) + \hat{p}_t(2) = 1$. If we raise the horizontal axis to λ_1 and rescale $\hat{\lambda}_2 - \hat{\lambda}_1$ to 1, the estimated intensity rate $\hat{\lambda}(t)$ should look exactly same as $\hat{p}_t(2)$. So, the last two figures in Figure 9.2 and others seem identical.

The estimated \mathbf{Q} and Λ are listed in Table 9.1. The time scale in the estimation is approximately one year (365 days). From Table 9.1, it is suggested that the deep seismicity in grouping C (Taupo-BOP region) varies most frequently among the three groupings as the estimated transition rates in \mathbf{Q} matrix are the largest ones, which is also indicated in Figure 9.5. Seismicity in grouping B (Taranaki region) has the lowest variability among the three groupings as the estimated transition rates in \mathbf{Q} matrix are the smallest ones, which is also suggested by Figure 9.4.

Deep Earthquakes Fitted through MMPP

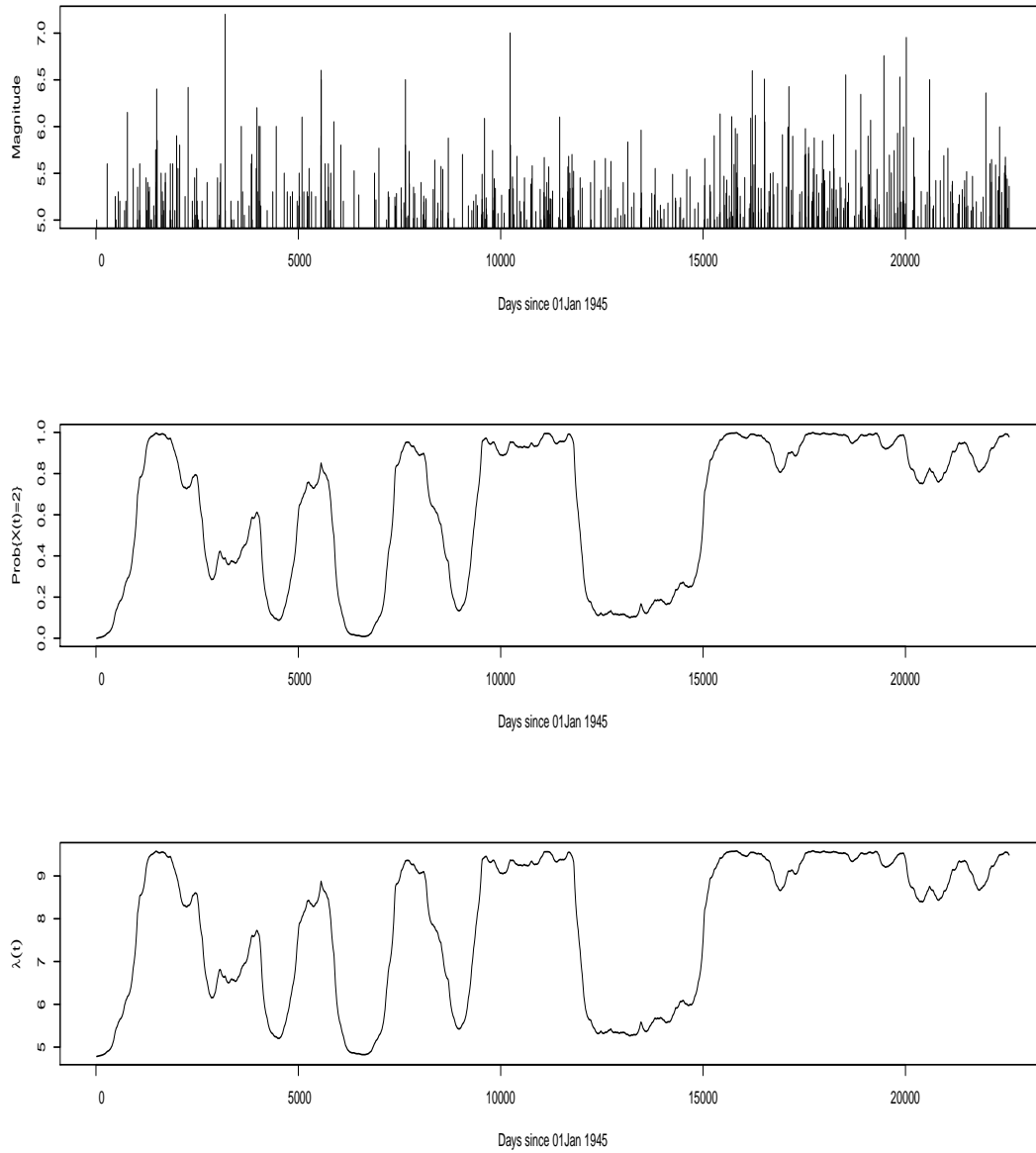


Figure 9.2: Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for the events with magnitude greater than 5 from all four main deep groupings in region A, B, C and D defined in Figure 9.1.

Deep Events in Cook Strait Grouping Fitted through MMPP

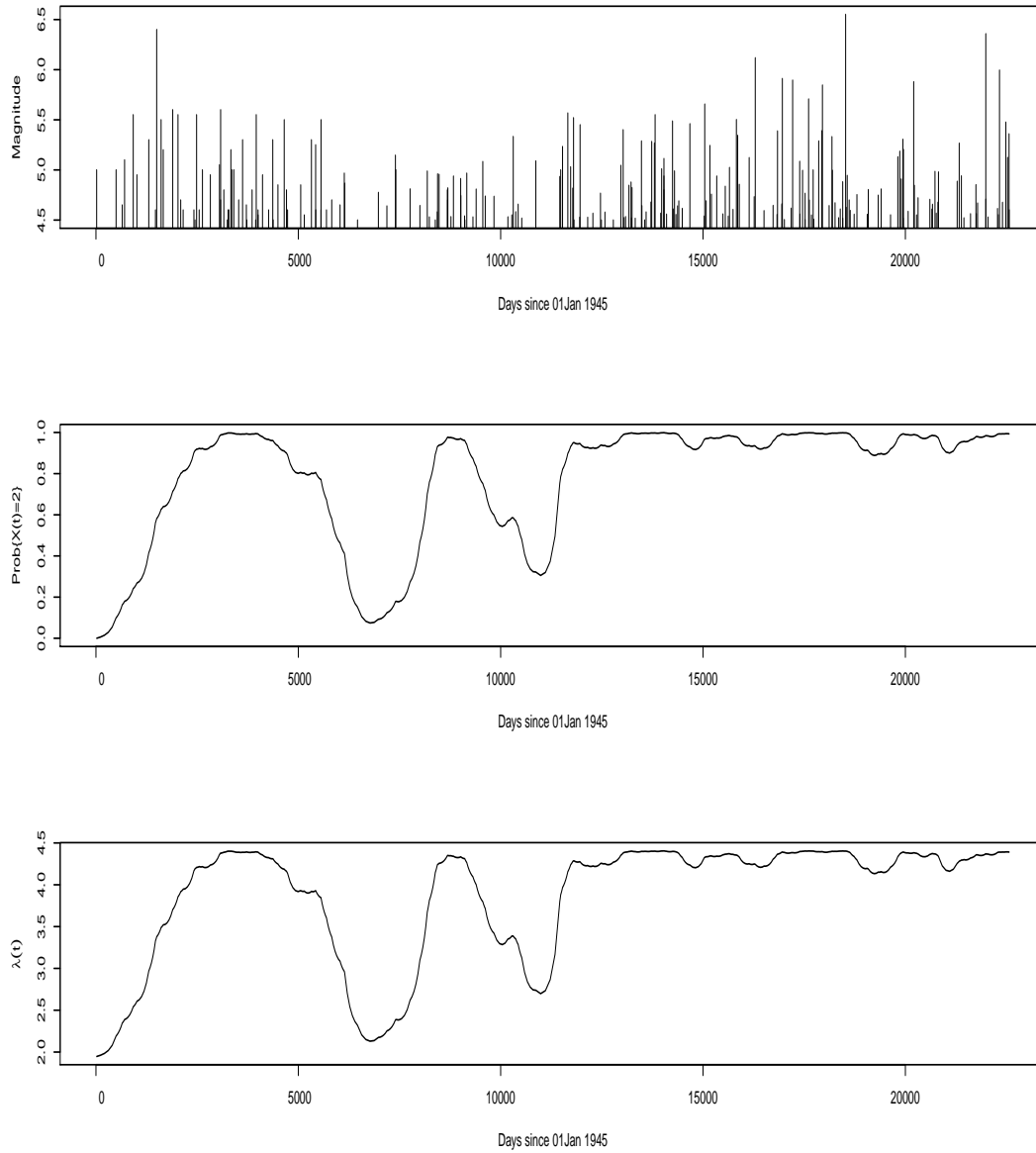


Figure 9.3: Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for the events from Cook Strait grouping in region A defined in Figure 9.1.

Deep Events in Taranaki Grouping Fitted through MMPP

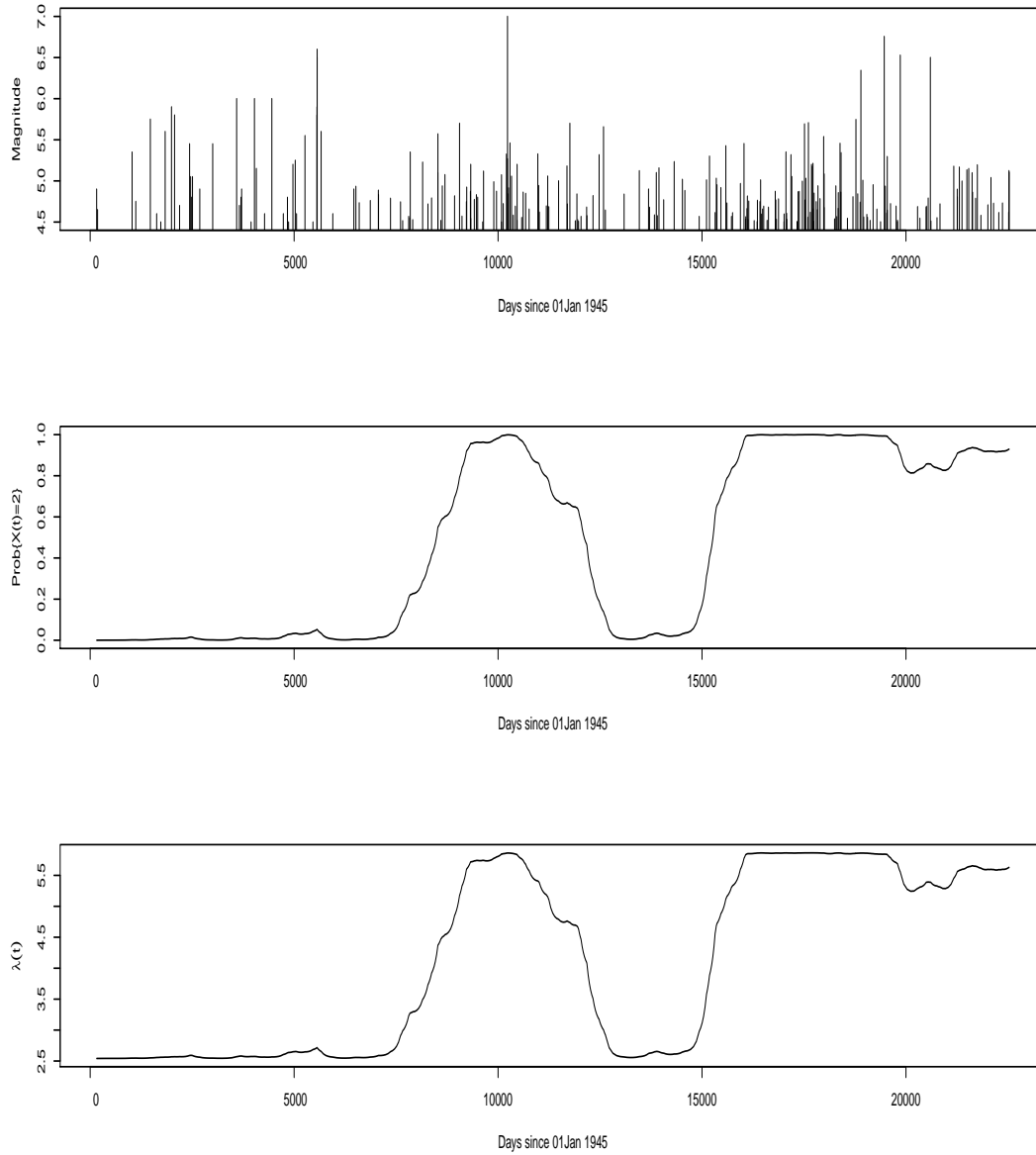


Figure 9.4: Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for the events from Taranaki grouping in region B defined in Figure 9.1.

Deep Events in Taupo-BOP Grouping Fitted through MMPP

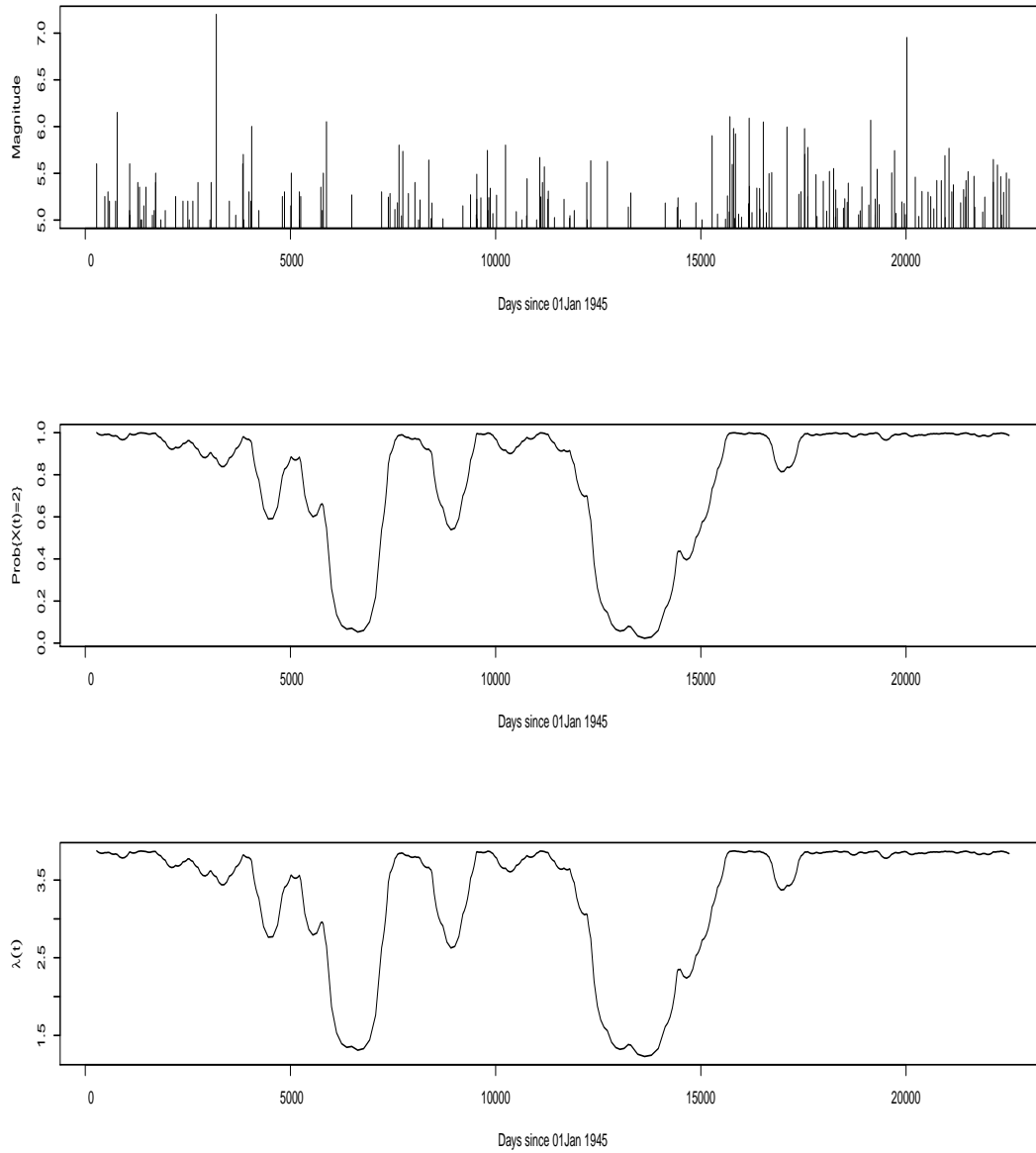


Figure 9.5: Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate for the events from Taupo-BOP grouping in region C defined in Figure 9.1.

Grouping	q_1	q_2	λ_1	λ_2	LogL
No grouping	0.403	0.206	4.78	9.59	520.23
Grouping A	0.307	0.0686	1.95	4.41	86.605
Grouping B	0.078	0.058	2.54	5.86	115.13
Grouping C	0.356	0.104	1.17	3.88	41.05

Table 9.1: Estimated parameters of MMPPs for deep earthquakes around the North Island and individual groupings A, B and C defined in Figure 9.1.

Grouping	MMPP-AIC	Poisson-AIC	MMPP-BIC	Poisson-BIC
No grouping	-1032.45	-1018.55	-1015.73	-1014.37
Grouping A	-165.21	-161.97	-151.32	-158.50
Grouping B	-222.27	-194.83	-208.19	-191.31
Grouping C	-74.11	-63.87	-60.94	-60.558

Table 9.2: Model selection between MMPP and Poisson model for deep earthquakes around the North Island and individual groupings A, B and C defined in Figure 9.1. There are four parameters in a second order MMPP.

Table 9.2 suggests that either by AIC or BIC, the second order MMPP outperforms Poisson model for all cases except for grouping A (Cook Strait grouping). For grouping A, which model is preferred is not conclusive as the conclusion depends on which model selection criterion is referred here. In this case, neither the switching

Poisson model nor the Poisson model outperforms one another. We also determine the order of the model by comparing AIC or BIC between a second order MMPP and a third order MMPP for events from all main groupings, denoted by MMPP_{2nd} and MMPP_{3rd} respectively. It turns out that a second order MMPP is sufficient to characterize the time-varying behaviour of the deep earthquakes, see Table 9.3. Higher order models will not be considered here due to limited available amount of data.

$\text{MMPP}_{2nd}\text{-AIC}$	$\text{MMPP}_{3rd}\text{-AIC}$	$\text{MMPP}_{2nd}\text{-BIC}$	$\text{MMPP}_{3rd}\text{-BIC}$
-1032.45	-1026.61	-1015.73	-988.97

Table 9.3: Model selection between the second order MMPP and the third order MMPP. There are 9 parameters in a third order MMPP.

Next, we will test whether our model provide a good fit for the time-varying behavior of the deep earthquakes. As suggested in section 3.3, we will consider using the rescaling theory of the point process for the goodness-of-test in this situation. By rescaling point process from $\{t_i\}$ to the compensator $\{\int_0^{t_i} \lambda(t) dt\}$, we obtain so-called residual point process which is a stationary Poisson process with unit rate when the proposed model is a real one. Kolmogorov-Smirnov goodness-of-fit statistic (e.g. the textbook of Stuart et al., 1999) is used to test whether the cumulative distribution of the transformed points is uniformly distributed as supposed if the suggested model fits the real pattern well, see section 3.3. In Figure 9.6, the dotted lines and dash lines give 99 and 95 percent confidence bands for the cumulative

number of arrivals by a unit rate Poisson process, the solid straight lines give the expected values of the cumulative numbers of a unit rate Poisson process. Although the intensity rates are slightly over-estimated at some periods, in all cases, the cumulative number of occurrences are within the 95 percent confidence bands.

The estimation errors of \mathbf{Q} and λ 's are estimated by parametric bootstrap methods. First, 1000 series of events with the same length of observations as the real data are simulated according to the estimated parameters listed in Table 9.1. Then the parameters are estimated via the EM algorithm for each series of simulated data. From 1000 bootstrap replications, we obtain the standard errors of the estimated parameters, see Table 9.4. From Table 9.4, it suggests that the estimates are relatively unstable for these small data sets, particularly for the transition rates in the \mathbf{Q} matrix.

Std. error	q_1	q_2	λ_1	λ_2
No grouping	1.07	1.12	1.22	1.08
Grouping A	1.56	3.32	0.87	1.13
Grouping B	0.48	0.37	0.58	0.96
Grouping C	1.23	0.65	0.97	0.89

Table 9.4: Standard error of estimates for \mathbf{Q} and λ 's corresponding to Table 9.1.

Considering that the estimates of parameters in MMPP in section 9.1 and MMPP with marks in section 9.2 are quite similar to each other and the marks can only exert very small leverage effects on the estimation for this small data sets, we will discuss the estimation errors of MMPP and MMPP with marks in a general approach, see

section 4 in this chapter.

K-S Tests for the Cumulative Rescaled Occurrence Times

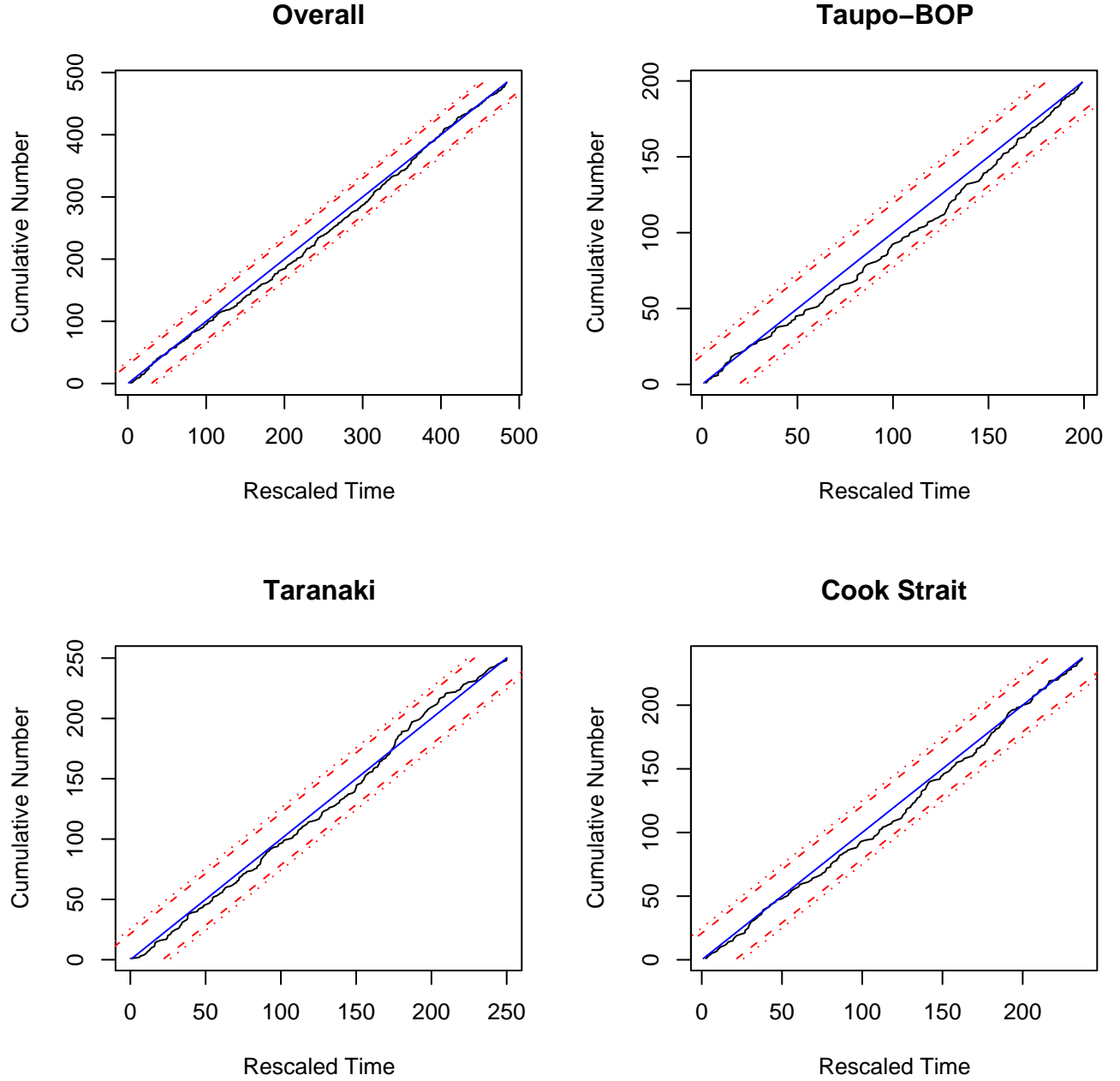


Figure 9.6: Kolmogorov-Smirnov test for the cumulative distribution of rescaled residual point process with 95 and 99 percent confidence bands by dash lines and dotted lines respectively for the deep earthquakes around the North Island and individual groupings defined in Figure 9.1.

9.2 Application of MMPP with Marks to Deep Earthquakes

The "b-values" (see section 8.2.3) governing the relative proportion of large and small events often varies over time and depth at many seismic active zones. Table 8.1 demonstrates that the b-value tends to decrease systematically with depth, suggesting the increasing possibilities for large tremors at greater depth. The b-value also tends to vary with the type of earthquake sequence. Generally speaking, the aftershock sequence tends to have a relatively high b-value as most of the earthquakes in the sequence are small aftershocks. The main-shock tends to have the lowest b-value. Some researchers suggest the b-value is associated with faulting types and its corresponding stress levels (Schorlemmer et al., 2005).

In previous section, we suggest the main evolution patterns of the deep seismicity are well characterized by a switching Poisson model. However, whether the magnitude distributions are also correlated with different levels of the deep seismicity occurrence rate is still unclear at this stage. The properties associated with the magnitude distribution when the deep seismicity comes to different levels merit further exploration. Motivated by a better understanding of the time-varying deep seismicity, we propose MMPP with events size taken as state-dependent marks which are exponentially distributed in terms of Gutenberg-Richter law. As done in previous section, we first apply a switching Poisson model with state-dependent marks to those groupings as a whole. Then a detailed analysis is carried out for each grouping. The estimated \mathbf{Q} matrix and intensity rates are listed in Table 9.5. The time scale in

the estimation is approximately one year (365 days). The b-value b and the decaying parameter α of the magnitude distribution are related by $b = \alpha/\log(10)$. The initial values used in the EM iteration steps for \mathbf{Q} and λ 's is determined as mentioned in section 9.1. The initial values used in the EM iteration steps for the decaying parameters of the magnitude distributions are selected by the following method. We first carry out the standard EM procedure for the discrete time MMPP without marks, a discrete-time hidden Markov model with Poisson observations as discussed in section 9.1. Then the Viterbi algorithm will yield the most likely state sequence for the underlying process at arrival times. So, the initial values of the parameters in the mark distributions can be obtained by standard MLE according to the Viterbi sequence which specifies the distribution of each mark.

Grouping	q_1	q_2	λ_1	λ_2	α_1	α_2	LogL
Overall	0.400	0.199	4.72	9.58	3.13	2.58	514.92
Grouping A	0.275	0.052	1.98	4.34	2.02	2.61	69.43
Grouping B	0.080	0.044	2.40	5.65	1.77	2.53	71.56
Grouping C	0.322	0.10	1.25	3.89	3.78	3.00	64.34

Table 9.5: Estimated parameters of MMPP with state-dependent marks for deep earthquakes around the North Island and individual groupings.

The transition rates of \mathbf{Q} matrix and intensity rates of the observed point process in Table 9.5 are very close to those in Table 9.1. From the simulation study in chapter 4.5, this similarity is most likely ascribed for that the mark distributions can only

exert very small leverage effects on the estimates. Note that in Table 9.5 the b-values of earthquakes in grouping A (Cook Strait grouping) and grouping B (Taranaki grouping) are less than or greater than unity as $X(t)$ in the seismic quiescent state (the first state) or the seismic active state (the second state) respectively. However, the b-values of earthquakes in the geothermal and volcanic active region C (Taupo-BOP region) are greater than unity as $X(t)$ is in both states. Furthermore, the b-values of deep earthquakes in region C show a reverse trend in comparison with that in region A and B, i.e. the smaller b-value is associated with the seismic active state (the second state) and the larger one is associated with the seismic quiescent state (the first state). These features need more geophysical insights to interpret.

Grouping	MMPP(1) AIC	MMPP(2) AIC	MMPP(1) BIC	MMPP(2) BIC
Overall	-1017.85	-1017.91	-992.75	-996.999
Grouping A	-126.86	-128.47	-106.27	-111.11
Grouping B	-131.12	-126.74	-109.99	-109.13
Grouping C	-116.69	-118.05	-96.93	-101.58

Table 9.6: Model selection between MMPP with state-dependent marks (MMPP(1)) and MMPP with state-independent marks (MMPP(2)). There are 5 parameters for a second order MMPP associated with state-independent marks.

Table 9.6 lists both the AIC and BIC values of the two nested models, namely MMPP with state-dependent marks and MMPP with state-independent marks, denoted by MMPP(1) and MMPP(2) respectively in the table. No obvious justifi-

cation has been established for preference of MMPP with state-dependent marks rather than MMPP with state-independent marks or vice versa in terms of AIC or BIC. However, the table shows that the b-values of the magnitude distribution associated with two states in grouping A (Cook Strait grouping) and grouping B (Taranaki grouping) vary between two values, less than unity when $X(t)$ in the first state and greater than unity when $X(t)$ in the second state. The difference of the estimated parameters of the magnitude distribution associated with two states might be significant. The largest grouping, Taupo-BOP which is a volcanic active zone, indicates rather high b-values in both states.

The probabilities of the underlying Markov chain in the second state (the seismic active state) are implemented by using the fixed point smoothing algorithm which gives the conditional probabilities of the underlying Markov chain in a state at a specific time conditioned on all available observations including the marks, see section 4.4.1. Since the smoothing estimation of the probabilities of $X(t)$ in a state is continuous with respect to t , the probabilities evaluated at many grid points and connected by straight lines can mimic the evolution of the underlying Markov chain well. Similar procedure is applied to demonstrate the evolution of the ground process of the observed marked point process according to equation (4.14). The magnitudes versus occurrence times, estimated probabilities of the hidden Markov process in the second state and the estimated ground intensity rates for events from all groupings are given in Figure 9.7. Considering the marks can only exert very small effects on the estimates of the transition rate matrix \mathbf{Q} and the intensity rates λ 's even when the magnitude distributions are truly varying according to the

underlying Markov chain, the estimated probabilities of $X(t)$ in the second state and the estimated intensity rates of the observed point process should be very similar to those of MMPP without marks. Hence, Figure 9.7 seems very similar to Figure 9.2. Again, the deep seismicity shows greater variability before 10,000 days since Jan 01, 1945 in comparison to the later period and an unusual seismic active period persists since late 1980s with no ending signs. Similar pictorial representation of the deep seismicity in individual groupings such as A, B, C defined in Figure 9.1 is presented in Figure 9.8-9.10. From Figure 9.8-9.10, the deep seismicity still shows similar trends in each grouping after 5000 days since Jan 01, 1945. Again, the deep seismicity around 10,000 days and after 15,000 days since Jan 01, 1945 is relatively active in all three groupings.

Deep Earthquakes Fitted through MMPP with Marks

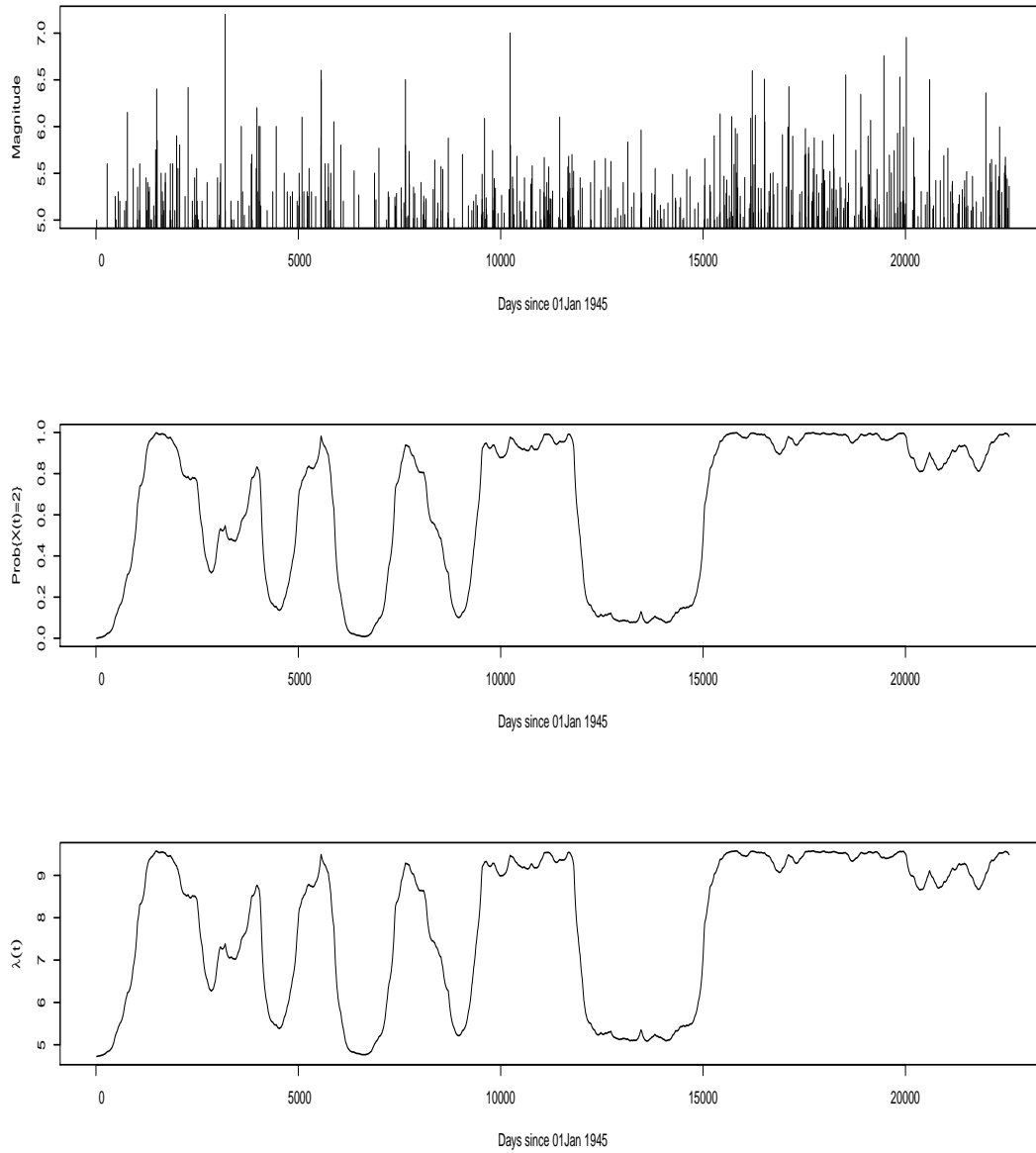


Figure 9.7: Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for the events from all four main groupings in region A, B, C and D defined in Figure 9.1.

Deep Events in Cook Strait Grouping Fitted through MMPP with Marks

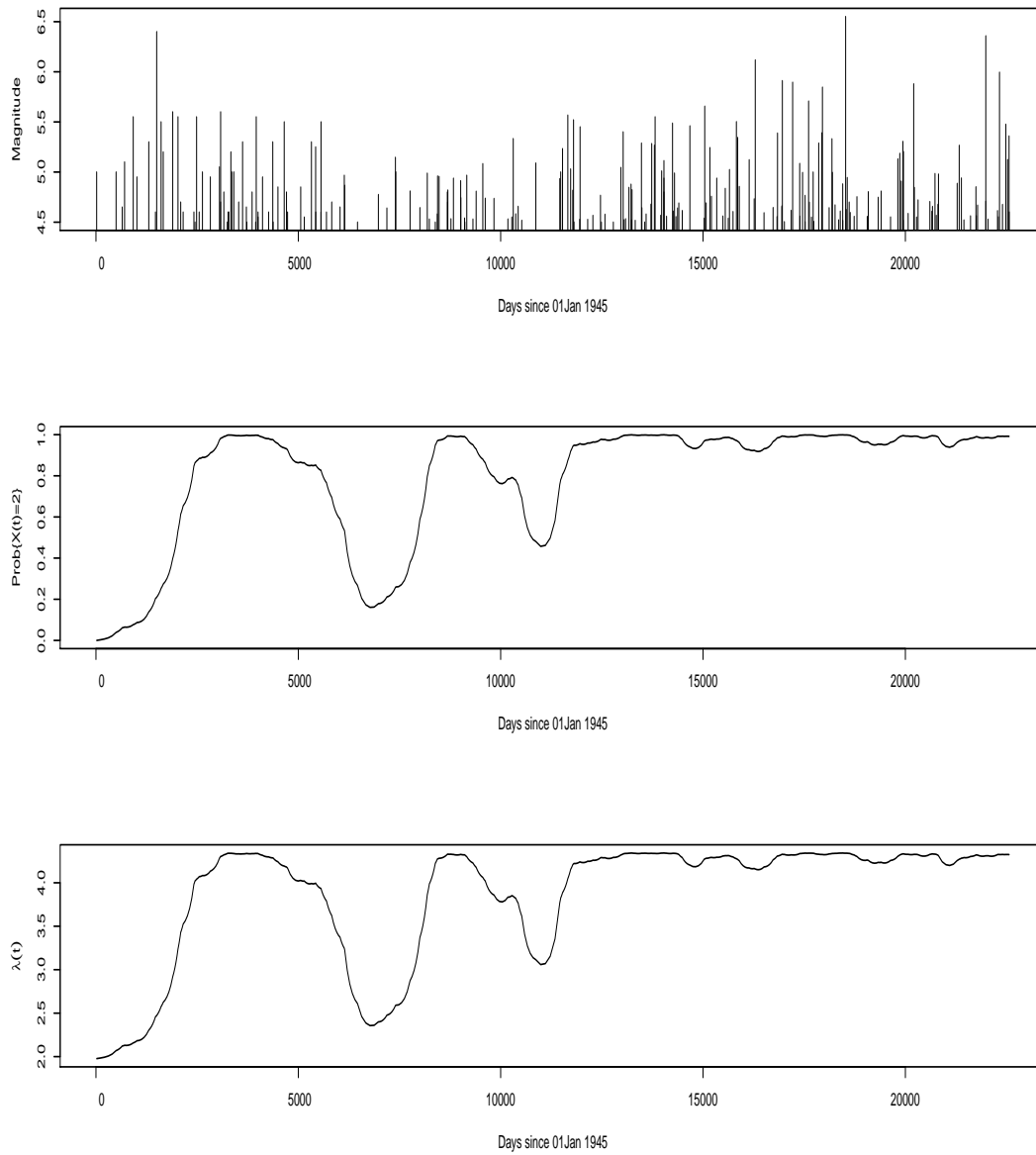


Figure 9.8: Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for the events from Cook Strait grouping in region A defined in Figure 9.1.

Deep Events in Taranaki Grouping Fitted through MMPP with Marks

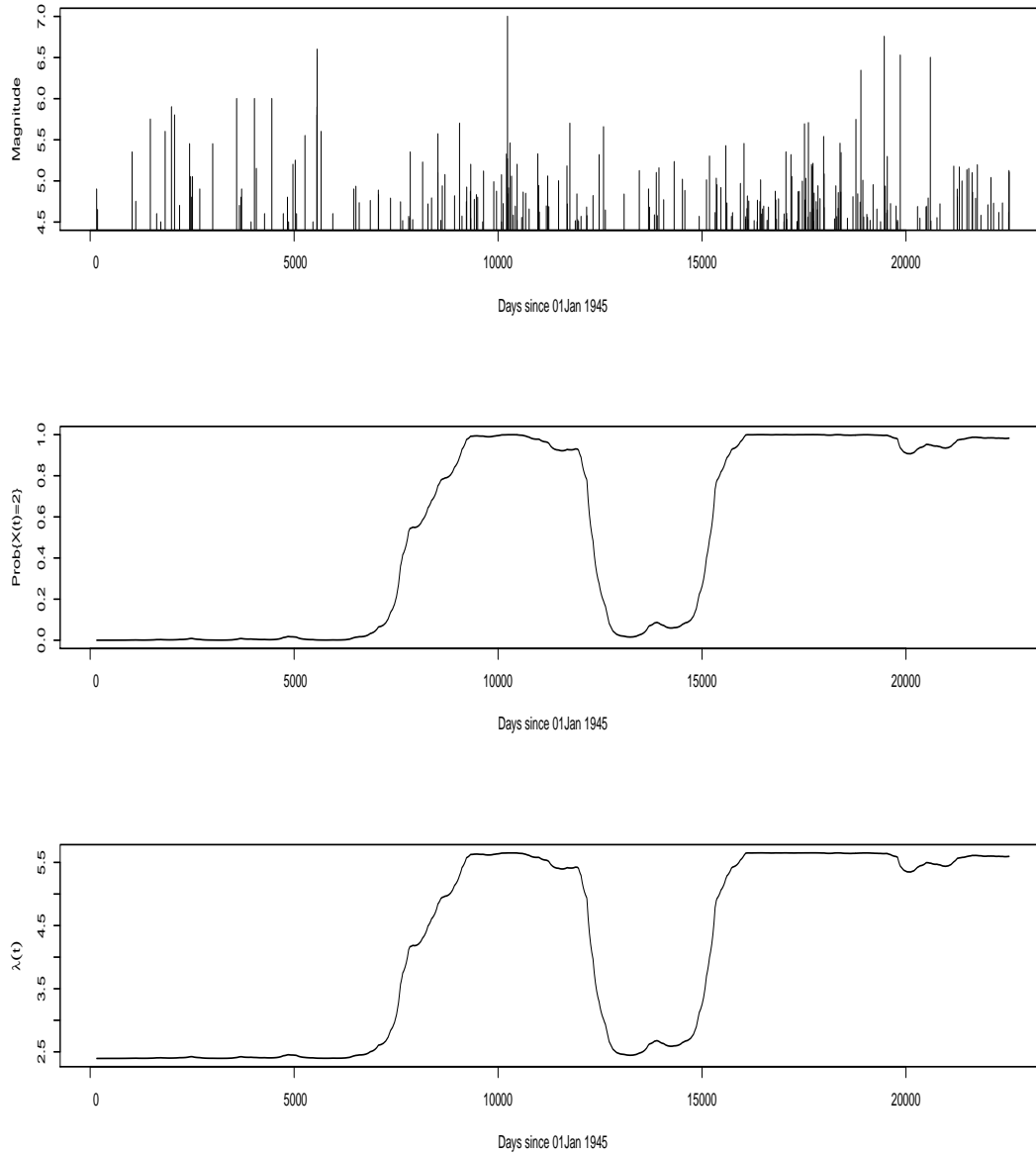


Figure 9.9: Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for the events from Taranaki grouping in region B defined in Figure 9.1.

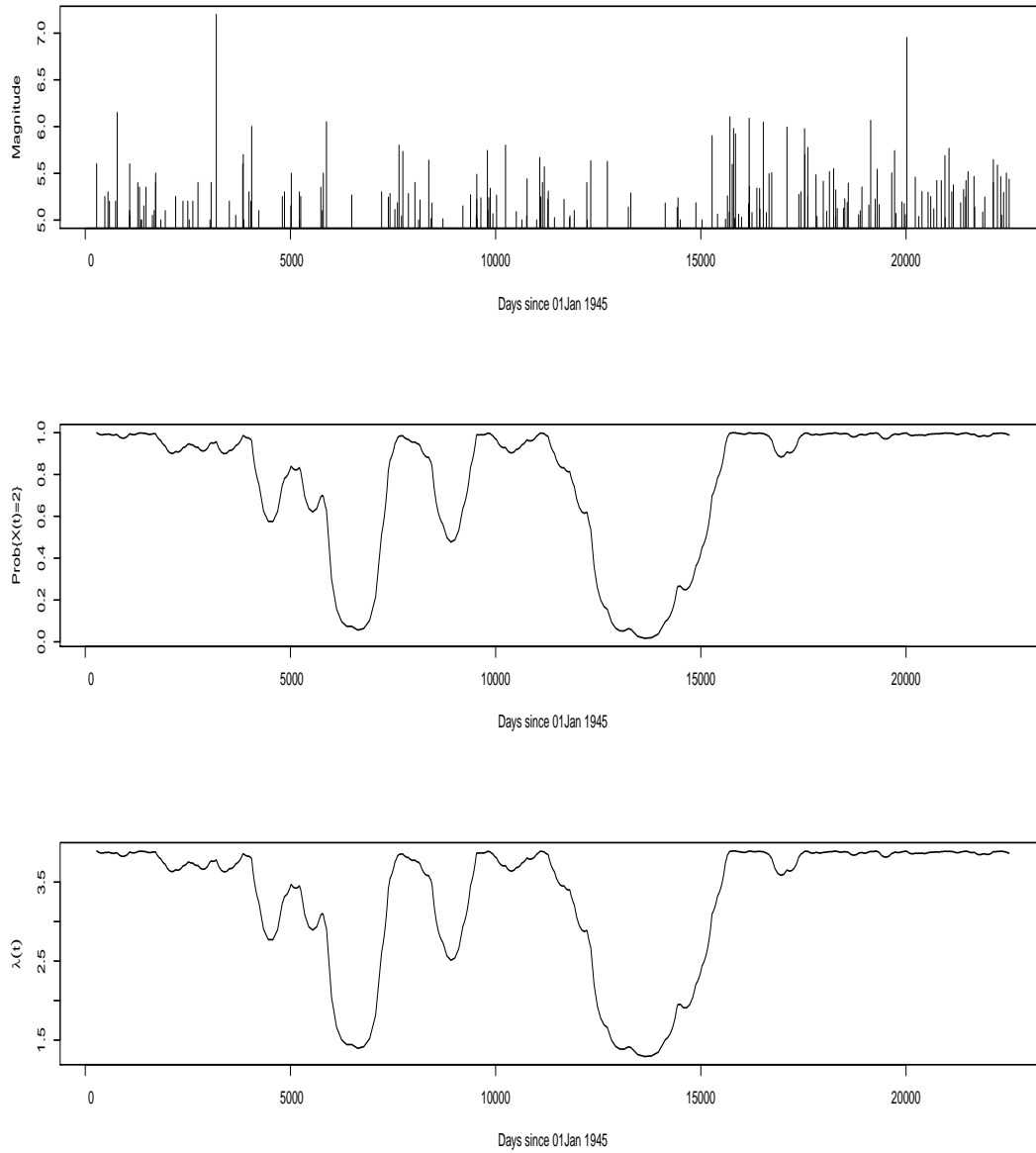
Taupo-BOP Deep Grouping Fitted through MMPP with Marks

Figure 9.10: Magnitude versus time plot, estimated probabilities of $X(t)$ in the state 2 and estimated intensity rate for the events from Taupo-BOP grouping in region C defined in Figure 9.1.

Note that to avoid the overflow or underflow problem in computation, we need scale the forward and backward probabilities by normalizing them to obtain the probabilities of the underlying Markov chain $X(t)$ in a specific state and the estimated intensity rates of the marked MMPP. Similar numerical procedures should be applied in the EM iterations by using the scaling factors determined by the order of the matrix exponential to update the Q matrix.

MMPP(1)-AIC	MMPP(3)-AIC	MMPP(1)-BIC	MMPP(3)-BIC
-1017.43	-1012.32	-992.33	-962.136

Table 9.7: Model selection between the second order MMPP with state-dependent marks (MMPP(1)) and the third order MMPP with state-dependent marks (MMPP(3)). There are 12 parameters in a full third order MMPP associated with state-dependent marks.

Table 9.7 suggests that a third order MMPP with state-dependent marks (MMPP(3)) is not better than a second order MMPP with state-dependent marks for the data from all main deep groupings. Higher order MMPP would not be considered due to limited data size.

The doubly rescaled marked point patterns based on a switching Poisson process with state-dependent magnitude distribution for events from all deep groupings as a whole and individual groupings are given in 4 plots of Figure 9.11. It is worth noting that there are some regular patterns at the left bottom of the doubly rescaled events in each plot of Figure 9.11 which obviously suggests the magnitude records

have rounding errors in early period for small events. We use estimated K-function $K(d)$ which indicates the proportion of paired points per unit area within a specified distance d to detect deviations of the proposed model from a unit rate planar Poisson process, see Cressie (1993). After performing K-test via Ripley's correction for the boundary effect over the doubly rescaled marked point process which is supposed to be Poissonian in the specified window if the proposed model fits the real pattern well, no obvious clustering or regularity pattern can be detected in the transformed process. Plots in Figure 9.12 show the estimated K-functions for the doubly rescaled process by solid lines and theoretical values of K-functions of Poisson processes by dash lines which is a quadratic function of the distance d . It suggests that the K-functions of the doubly rescaled MMPP with marks coincide with that of a Poisson process in all cases very well. In other words, the estimated intensity rates for the marked point process should be very close to the true values. However, considering K-function doesn't provide detail information of where the fit is unsatisfactory and there is no significance levels set out in the sense of a standard statistical test, it doesn't mean the model fits the pattern perfectly well.

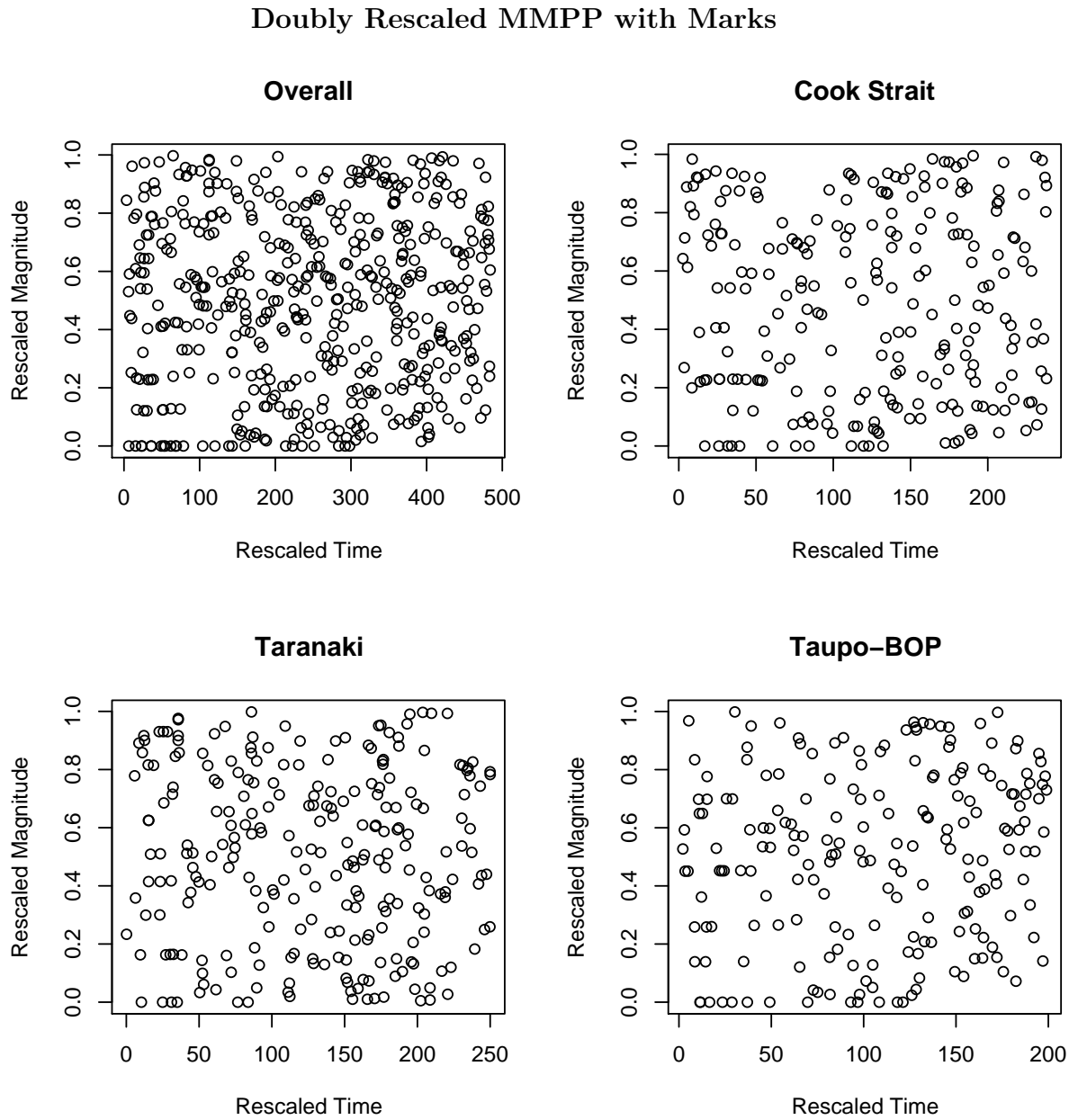


Figure 9.11: Doubly rescaled marked point process based on a switching Poisson process with state-dependent magnitude distribution for events from all deep groupings as a whole and Cook Strait grouping, Taranaki grouping, Taupo-BOP grouping.

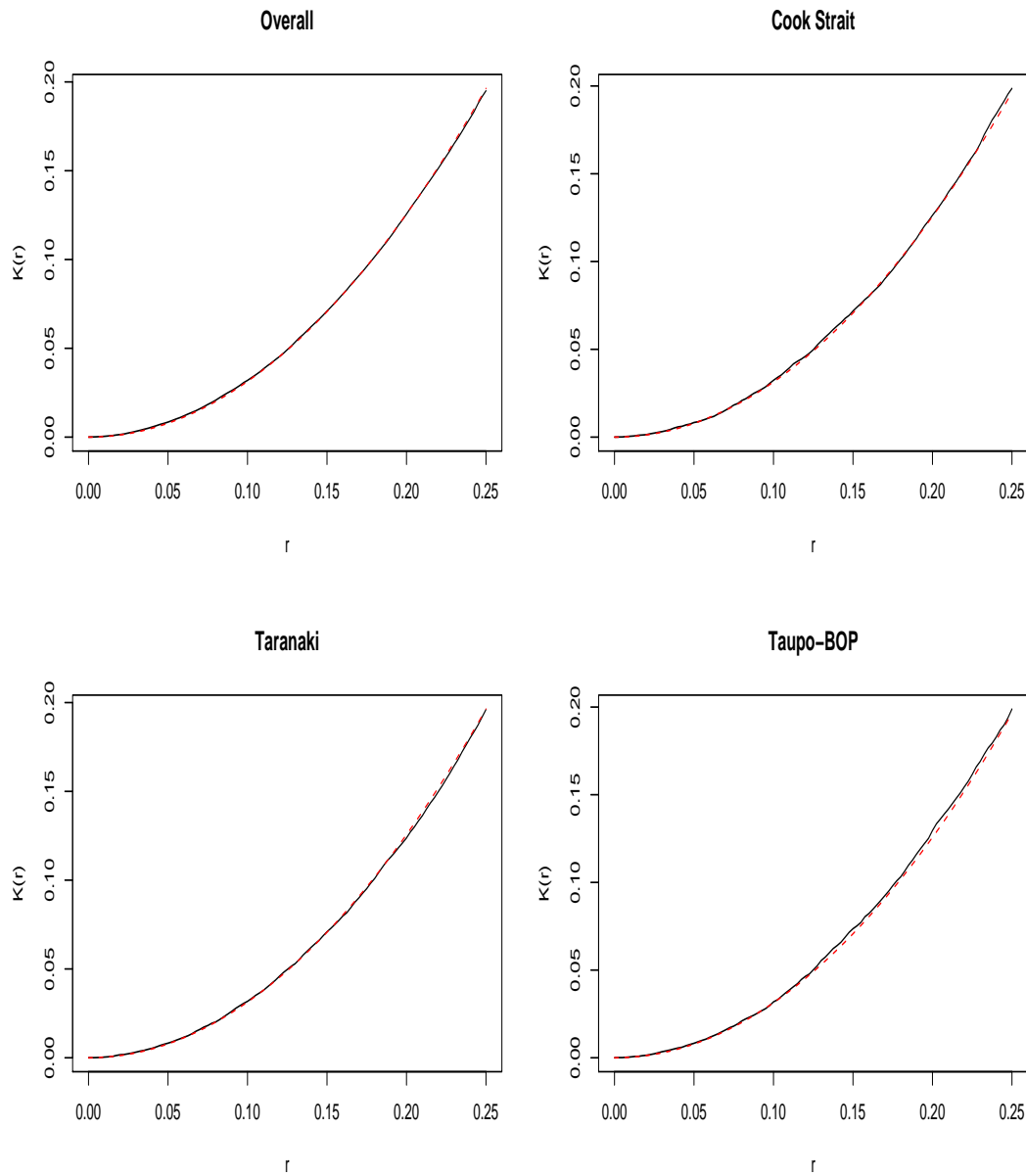
K-test for the Doubly Rescaled MMPP with Marks

Figure 9.12: K-tests via Ripley's correction for the boundary effect over the doubly rescaled marked point process for all four deep groupings as a whole and Cook Strait grouping, Taranaki grouping, Taupo-BOP grouping. The solid lines are estimated K-functions. The dash lines show the expected values of the K-function of a Poisson process.

9.3 Application of Multivariate MMPP to Deep Earthquakes

From previous analysis in section 9.1, it is suggested that the deep seismicity in all three individual groupings show similar time-varying behavior after 5000 days since 1 Jan 1945, see the estimated probabilities of the underlying process $X(t)$ in the second state and the estimated intensity rates in Figure 9.2-9.5. It is reasonable to speculate that the seismicity in individual groupings might be driven by a 'common' geophysical process associated with the subduction of the oceanic plate. We also speculate that the similar time-varying behaviors in seismicity might be associated with the time-varying behavior of the **thermal parameters** of the subduction zone which is defined as the products of the age of the slab and the vertical component of convergence rate of the subducted plate.

To validate the speculation that the occurrence patterns of the deep earthquakes in individual groupings might be associated with a common underlying process, we apply a bivariate MMPP to two conspicuous deep groupings, namely the Taupo-BOP grouping and the Taranaki grouping which are adjacent with each other geophysically. Then by comparing the bivariate MMPP with the individual MMPPs through the model selection criteria such as AIC or(and) BIC, whether the two individual MMPPs can be synchronized together under a common underlying process is concluded. Here, $AIC = -2 * \log L + 2k$ and $BIC = -2 * \log L + k * \log(n)$ respectively. In the definition, k is the number of the parameters and $\log L$ is the log-likelihood of the model.

The data used in this analysis are selected from New Zealand catalogue from 1950 to 2007 with depth greater than 40. The magnitude thresholds are respectively 4.5 and 5 in Richter scale for Taranaki grouping and Taupo-BOP grouping respectively. The two states of the bivariate MMPP are associated with the seismic quiescent state (the first state) and the seismic active state (the second state) respectively. We also apply the method to two other combinations of deep groupings, i.e. Cook Strait grouping and Taranaki grouping, Cook Strait grouping and Taupo-BOP grouping. Table 9.8 suggests that a bivariate MMPP outperforms two independent MMPPs for the Taranaki grouping and Taupo-BOP grouping. Either by AIC or BIC, it confirms that two individual MMPPs are positively correlated with each other through an underlying Markov process rather than independent. In other words, the synchronization of two MMPPs through a common underlying process fits the occurrence patterns of the two deep groupings much better than two independent MMPPs. However, for the other two combinations of the deep groupings, i.e. Cook Strait grouping and Taranaki grouping, Cook Strait grouping and Taupo-BOP grouping, whether a bivariate MMPP outperforms two independent MMPPs or vis versa are not conclusive. So, further synchronization of all three deep groupings via a three variate MMPP will not be considered here since Cook Strait grouping obviously does not evolve simultaneously with others.

Figure 9.13 shows the probabilities of the baseline process $X(t)$ of the bivariate MMPP in the second state. In both deep groupings (The Taupo-BOP grouping and Taranaki grouping), the seismicity is relatively active since late 1980s without ending signs. The probabilities of $X(t)$ in the state 2 are evaluated at many pre-selected grid

	Independent MMPPs	Bivariate MMPP
Tara-BOP AIC	-291.73	-306.31
Tara-BOP BIC	-264.99	-282.01
Cook-Tara AIC	-405.63	-407.46
Cook-BOP AIC	-229.10	-230.25

Table 9.8: Model selection between bivariate MMPP and independent MMPPs for 3 deep groupings. There are 6 parameters in a bivariate MMPP.

points according to equation (5.17) in section 5.5 and connected together by straight lines. Note that to avoid the overflow or underflow problem in computation, we need to scale the forward and backward probabilities by normalizing them to obtain the probabilities of the underlying Markov chain $X(t)$ in a specific state. Similar scaling procedure should also be deployed to evaluate other quantities as shown in following figures.

Figure 9.14 and 9.15 demonstrate the estimated intensity rates of the marginal processes of a bivariate MMPP and perform K-S tests for the marginal processes. From these two figures, it is evident that two marginal processes are positively correlated since the intensity rates of the marginal processes are both high or low at the same periods. In the bottom of these two figures, the dot lines give 95 and 99 percent confidence bands of the K-S tests. The solid lines are the cumulative number of arrivals in the marginal process which is rescaled to form a unit rate Poisson process if the proposed model fits the real pattern well and the dash lines

Hidden State Probabilities of Bivariate MMPP

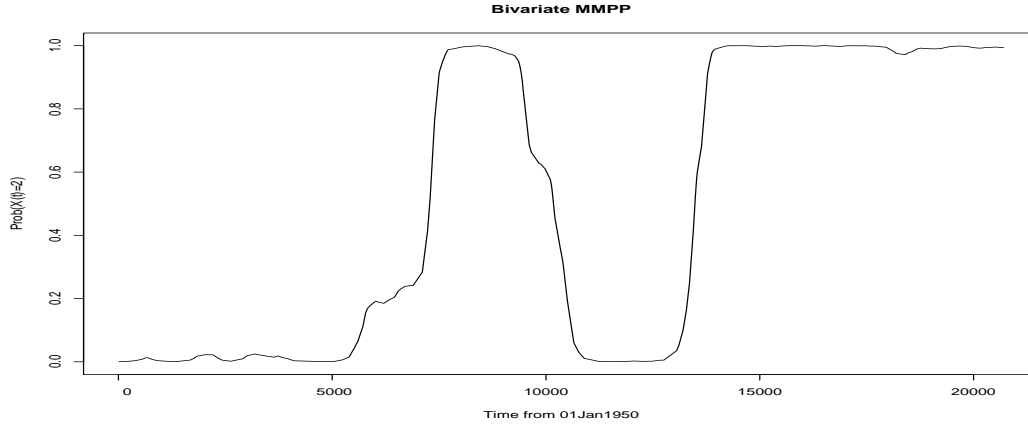


Figure 9.13: The estimated probabilities of $X(t)$ in the state 2 of a bivariate MMPP.

are the expected cumulative numbers of a unit rate Poisson process. The K-S tests suggest the bivariate MMPP fits the occurrence patterns of Taupo-BOP grouping and Taranaki grouping well.

Figure 9.16 gives the estimated cross-intensity, cross-variance and cross-spectrum $h_1^{(2)}(t)$, $\gamma_1^{(2)}(t)$ and $g_1^{(2)}(\omega)$. Both the cross intensity $h_1^{(2)}(t)$ and the cross variance $\gamma_1^{(2)}(t)$ indicate that the two marginal processes are positively correlated. The positive correlation reaches highest amount at the beginning and decreases slowly to nearly zero after 20 years, see the middle of the Figure 9.16 of the cross variance. There is no periodic effect existing between the marginal processes according to the cross spectrum $g_1^{(2)}(\omega)$.

The conclusions are subjected to the influence from the estimation errors of the parameters since the estimation errors of the parameters are not negligible, particularly for the parameters of the transition rates in Q matrix, see section 9.4.

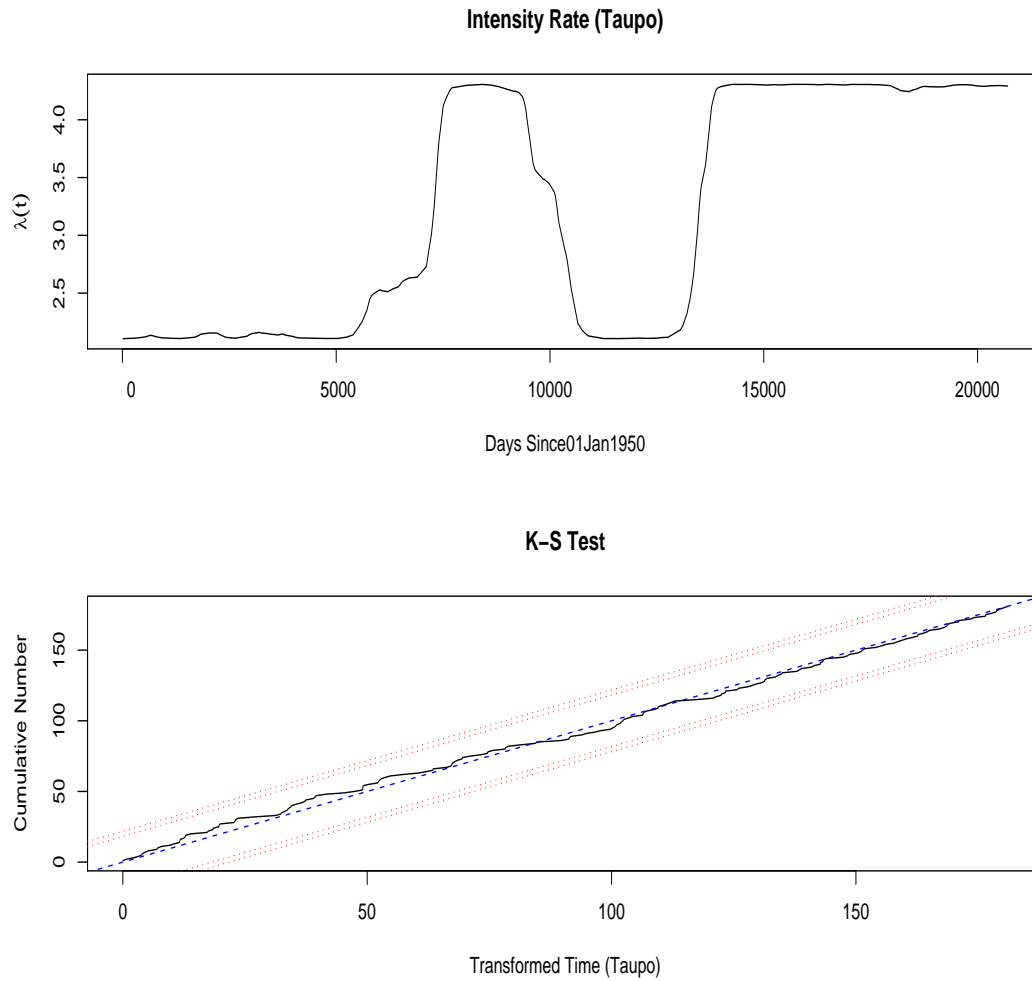
Taupo-BOP Grouping Fitted via a Bivariate MMPP

Figure 9.14: Taupo-BOP grouping fitted via a bivariate MMPP. The top of the figure is the estimated intensity rates of the marginal process for Taupo-BOP grouping. The quantities are evaluated at many pre-selected grid points according to equation (5.18) in chapter 5.5 and connected together by straight lines. The bottom of the figure performs the Kolmogorov-Smirnov test for the empirical cumulative distribution of the rescaled process.

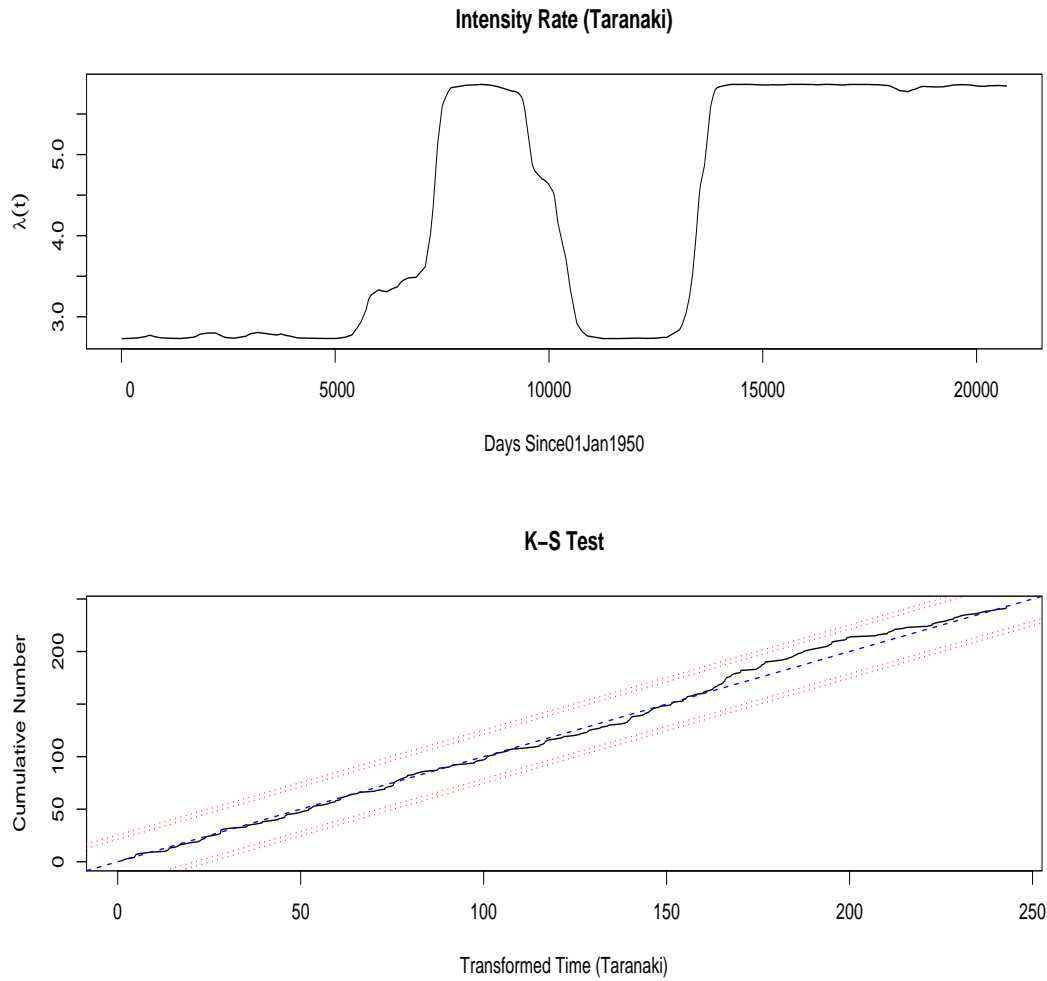
Taranaki Grouping Fitted via a Bivariate MMPP

Figure 9.15: Taranaki grouping fitted via a bivariate MMPP. The top of the figure is the estimated intensity rates of the marginal process for Taranaki grouping. The quantities are evaluated at many pre-selected grid points according to equation (5.18) in section 5.5 and connected together by straight lines. The bottom of the figure depicts the result of the Kolmogorov-Smirnov test for the empirical cumulative distribution of the rescaled process.

Second Moments of the Bivariate MMPP

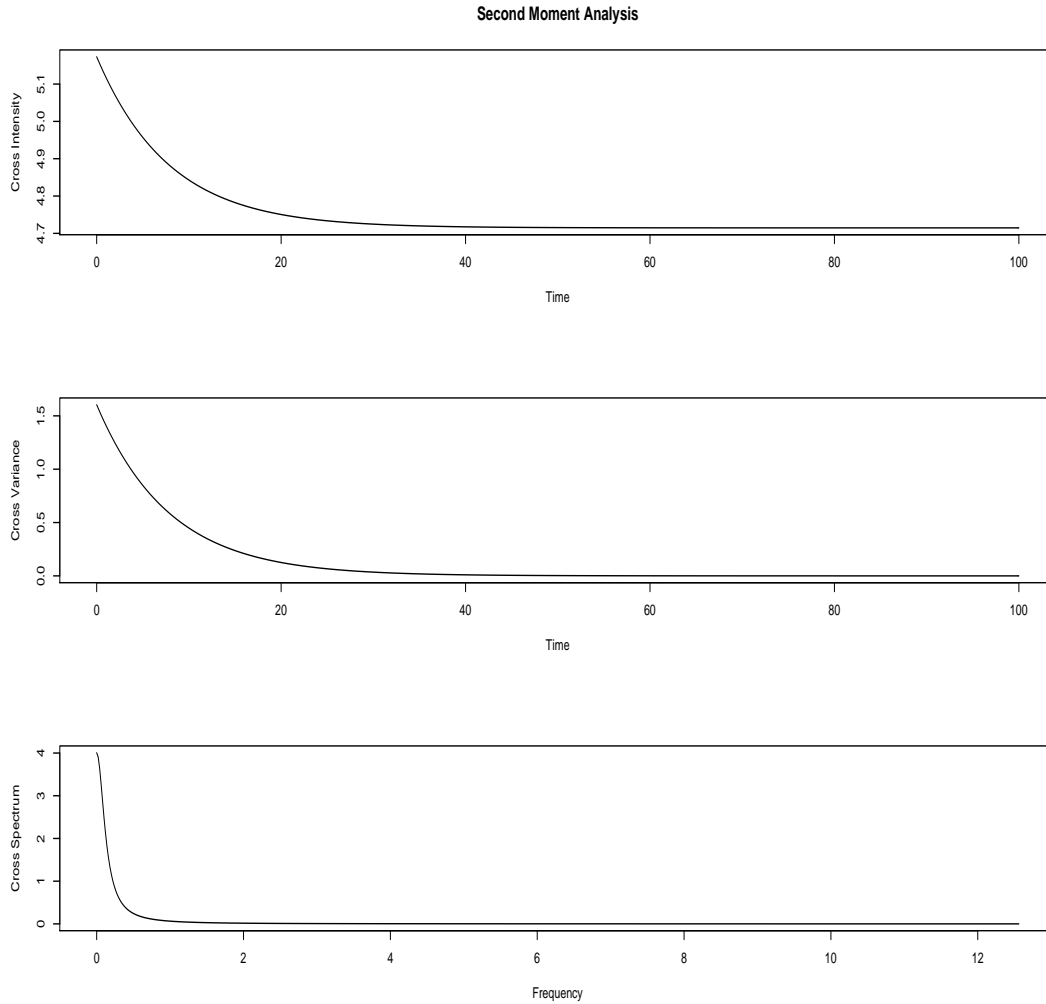


Figure 9.16: Second moments of the bivariate MMPP. The figures demonstrate the estimated cross intensity rate in the top, the estimated cross variance in the middle and the cross spectrum in the bottom. These quantities are defined in section 5.4. These figures suggest the marginal point processes are positively correlated.

9.4 Evaluations of Estimation Errors

In this section, we evaluate the estimation errors of the model parameters appearing in the Table 9.1 and 9.5. From the simulation studies in section 4.5, it is clear that the exponential type of marks with decaying parameters (b-values) can only exert very small effect on the estimation, see Table 4.1, 4.2 and Figure 4.4. The estimates in Table 9.1 and 9.5 are very close to each other. The model comparison between MMPP with state-dependent marks and state-independent marks suggests that the model preference is not conclusive for this small data sets. Due to these factors, we focus only on the estimation errors appearing in the Table 9.5.

The estimation errors are estimated by parametric bootstrap methods. We simulate 1000 series of events with the same length of observations as the real data according to the estimated parameters in Table 9.5. Then the parameters are estimated via the EM algorithm for each series of simulated data. From the 1000 bootstrap replicates, we obtain the 95 percent bootstrap percentile confidence intervals given in Table 9.9 and the bootstrap estimate of the covariance matrix for overall grouping as listed in Table 9.10. Histograms of bootstrap replicates for all parameters appearing in the Table 9.5 are listed in the Figure 9.17-9.20.

It is observed that the estimates are unstable for this small data set, particularly for the transition rate q_1 in Q matrix and α_1 in the magnitude distribution.

Histograms of the Bootstrap Replicates (Overall)

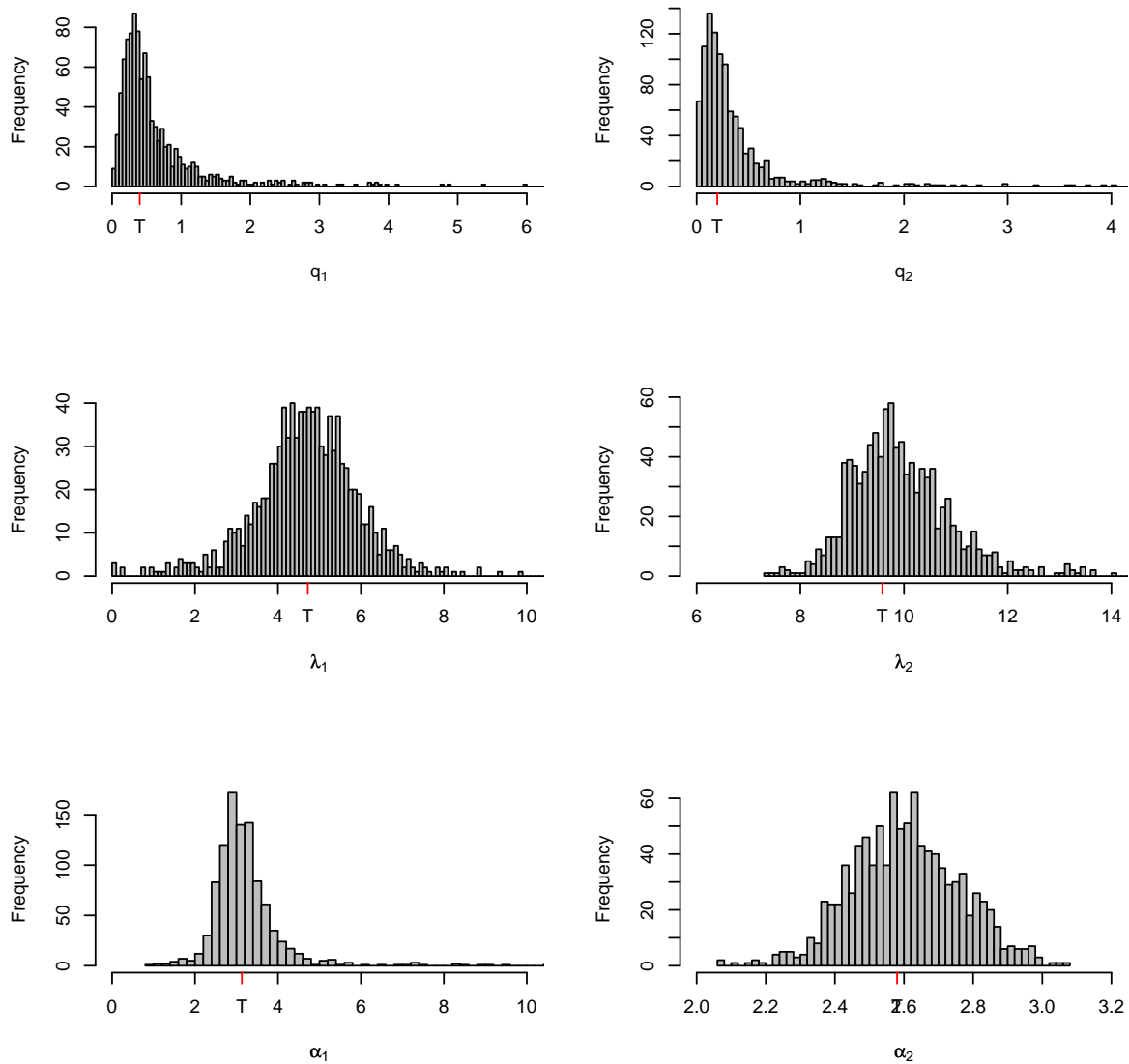


Figure 9.17: Histograms of the estimates for 1000 simulated MMPP series with given parameters for overall deep events. The tick marks T in the histograms give the true values of the parameters.

Histograms of the Bootstrap Replicates (Cook Strait)

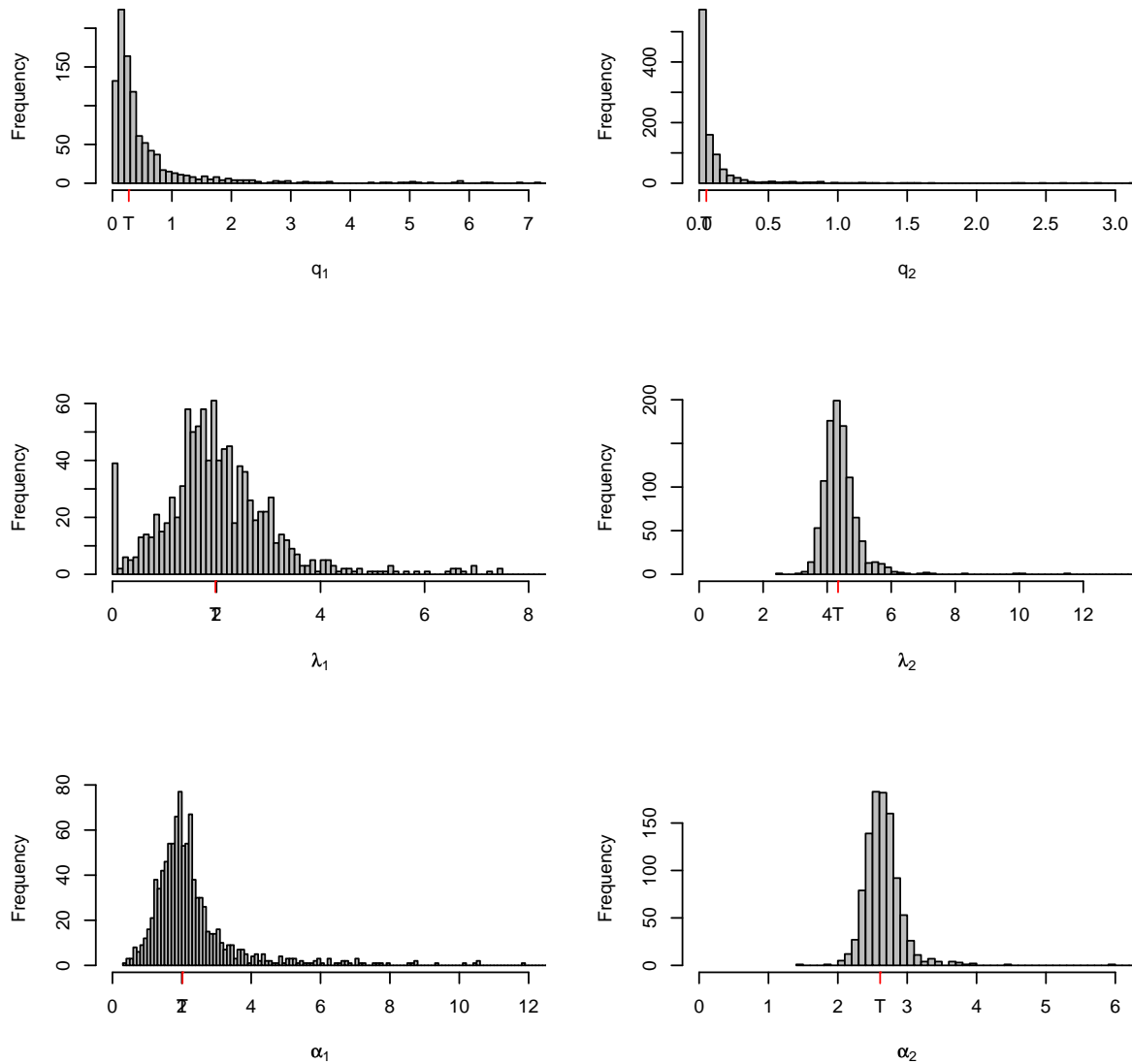


Figure 9.18: Histograms of the estimates for 1000 simulated MMPP series with given parameters for Cook Strait grouping. The tick marks T in the histograms give the true values of the parameters.

Histograms of the Bootstrap Replicates (Taranaki)

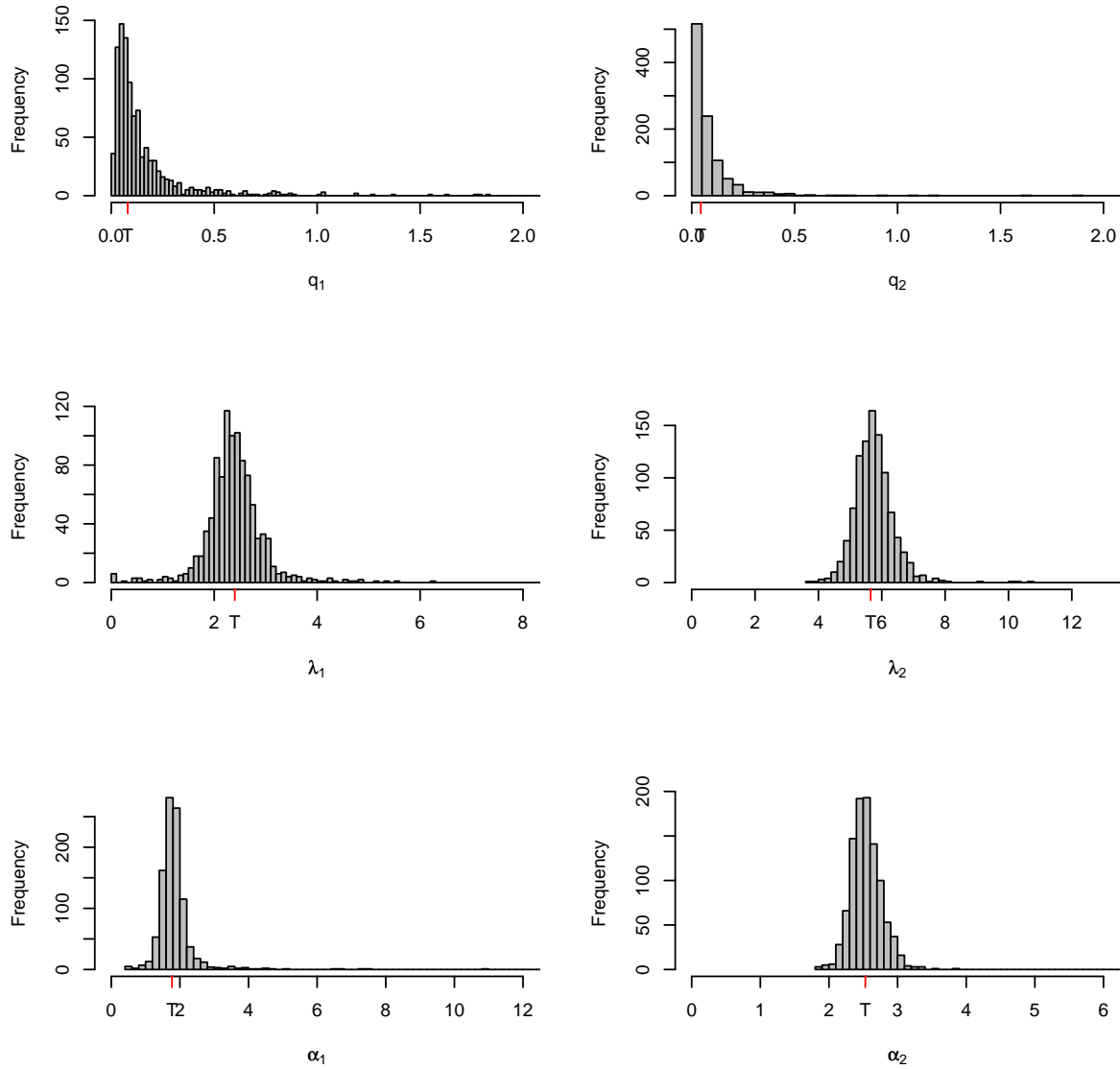


Figure 9.19: Histograms of the estimates for 1000 simulated MMPP series with given parameters for Taranaki grouping. The tick marks T in the histograms give the true values of the parameters.

Histograms of the Bootstrap Replicates (Taupo-BOP)

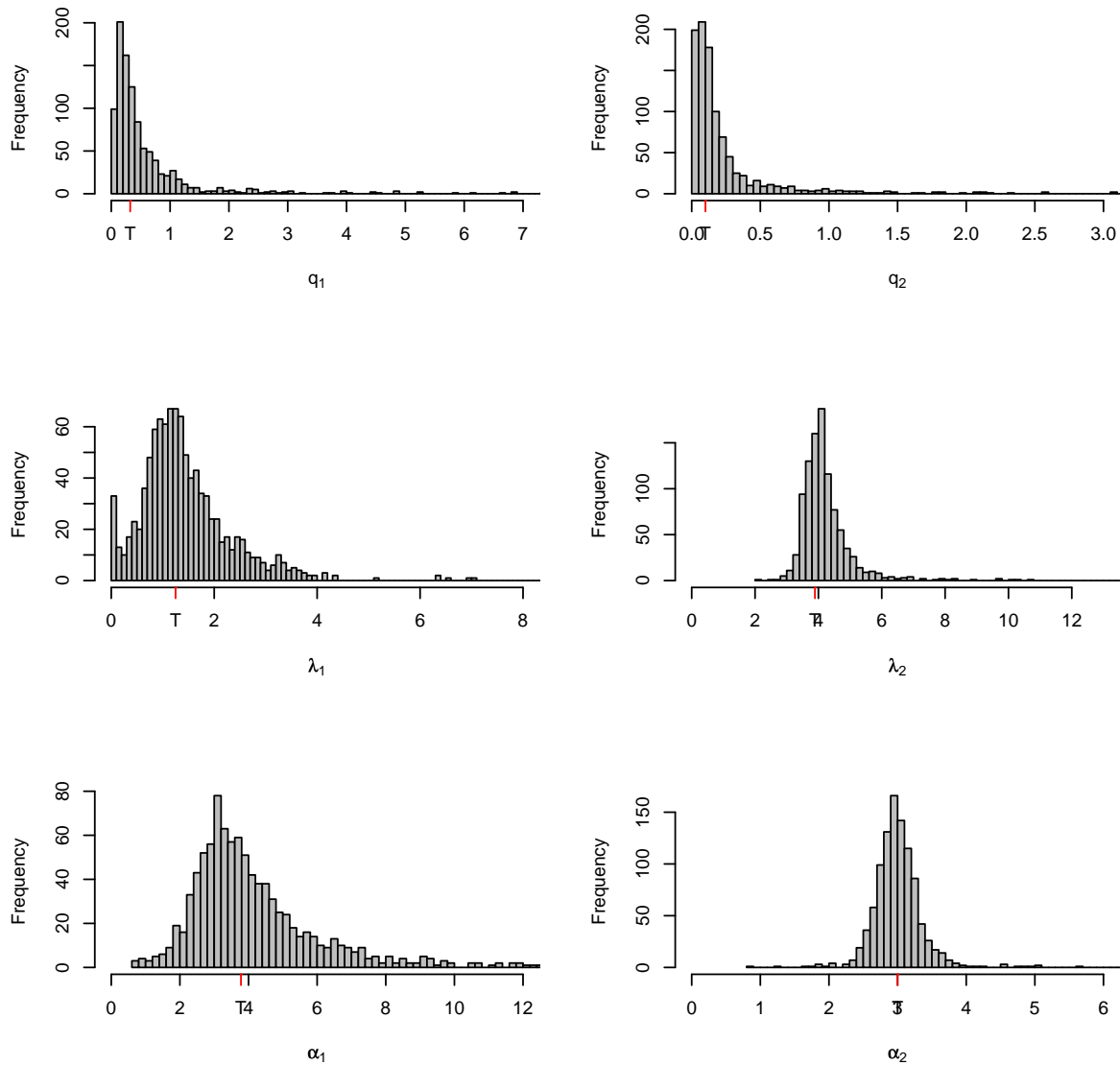


Figure 9.20: Histograms of the estimates for 1000 simulated MMPP series with given parameters for Taupo-BOP grouping. The tick marks T in the histograms give the true values of the parameters.

Grouping	q_1	q_2	λ_1	λ_2	α_1	α_2
Overall	[0.093,3.07]	[6e-05 2.00]	[1.81,7.17]	[8.34,12.61]	[2.10,5.70]	[2.30,2.92]
Grouping A	[0.041,3.63]	[4e-08 1.00]	[8e-06,5.18]	[3.67, 5.73]	[0.83,8.71]	[2.23,3.30]
Grouping B	[0.018,0.85]	[2e-08,0.40]	[1.29,3.62]	[4.69,7.03]	[1.18,3.11]	[2.15,3.03]
Grouping C	[1e-05 3.62]	[0.02,2.19]	[1e-3,3.53]	[3.27,6.37]	[0.69,14.7]	[2.43,3.71]

Table 9.9: 95 percent bootstrap confidence intervals for parameters in Table 9.5.

Covariance	q_1	q_2	λ_1	λ_2	α_1	α_2
q_1	1.31	0.44	-0.32	0.05	0.23	0.0039
q_2	0.44	0.395	-0.0075	0.389	0.005	-0.0027
λ_1	-0.32	-0.0075	2.55	0.46	-0.294	-0.032
λ_2	0.05	0.389	0.46	1.30	-0.175	-0.0115
α_1	0.23	0.005	-0.294	-0.175	1.428	-0.0118
α_2	0.0039	-0.0027	-0.032	-0.0115	-0.0118	0.0249

Table 9.10: Bootstrap estimation of the covariance matrix of the MMPP parameters for overall grouping in the first row of the Table 9.5.

Chapter 10

Concluding Remarks and Further Studies

10.1 Concluding Remarks

Among many differences between the deep earthquakes and the shallow earthquakes, one well-known fact is that the deep earthquakes rarely have plenty of aftershocks which decay in terms of Omori's law. A second order moment analysis carried out for the spatial and temporal clustering of the New Zealand deep earthquakes suggests that the deep earthquakes do have very small scale clustering, presumably only limited in some hot spots, along with occasional large doublets. Partly because of these factors, those models widely used for the occurrence patterns of the shallow earthquakes fit the main deep earthquakes in New Zealand poorly, see section 8.3. The main occurrence pattern of the New Zealand deep earthquakes, in a relatively

large time scale, is the time-varying seismic activities, active in one period, relatively quiescent in another, see Figure 8.1 and 8.2. The mechanisms and reasons behind this time-varying behavior are still not well understood. Hence, we suggest a hidden Markov model, in this case a second order MMPP or a switching Poisson model, to characterize the time-varying occurrence rates. The two states are designated to stand as two levels of seismicity, namely the seismic active state and the seismic quiescent state.

We apply a second order MMPP to the deep earthquakes as a whole. How and when the deep seismicity is changed is indicated through a smoothing procedure giving the probabilities of the underlying process in a specific state. Then, a detailed analysis is carried out to investigate where and when the deep seismicity is changed among the three deep groupings, namely Taupo-BOP grouping, Taranaki grouping and Cook Strait grouping. The estimated occurrence rates in individual groupings show similar patterns after 6000 days since 01 Jan 1945. We conjecture that the co-seismic features appearing among individual groupings are most likely driven by a 'common' process, which forms the motivation of suggesting a bivariate MMPP to synchronize the evolution of the seismicity in individual groupings under a common underlying process. There exists strong statistical evidence that the deep seismicity in individual groupings are positively correlated, particularly for the two largest groupings, Taupo-BOP grouping and Taranaki grouping, see the model comparison between the bivariate MMPP and independent MMPPs in Table 9.8. Goodness-of-fit tests based on the rescaling theory show that the estimation of the occurrence rates or the conditional intensity rates are well consistent with the true values. However,

these conclusions are partly offset by relatively large estimation errors, particularly for the transition rates in the Q matrix.

Another statistically interesting, geophysically important question is whether the magnitude distributions also vary simultaneously with the occurrence rates of the deep seismicity. Motivated by this question, we propose an extension of MMPP, in which each arrival is attached by additional variables or marks. However, any conclusion of whether the magnitude distributions vary simultaneously with the arrival rates is offset by two factors. Firstly, available quality data set is relatively small. Secondly, the simulation studies in chapter 4.5 suggest that the effects posed by the exponential type of marks with b-values taken in the usual range is rather small, which makes it difficult to draw a conclusion even when the b-values truly vary simultaneously with the occurrence rates, see Table 4.2 and Figure 4.4. Generally speaking, except for Taranaki grouping, the results of whether the b-values vary with the occurrence rates is not conclusive, either for all deep grouping together or for individual groupings. This is clearly demonstrated in Table 9.6. It is worth noting that the b-values vary either greater than 1 or less than 1 in different states for two groupings, i.e. Taranaki grouping and Cook Strait grouping. Another important fact found not only for New Zealand deep earthquake but also for those occurred in other subduction zones is that the b-values tend to decrease systematically with depth within individual groupings or altogether. What all this means and what is the geophysical importance need more interpretations from point of view of geophysics. Schorlemmer et al. (2005) suggest the b-values are associated with faulting types and its corresponding stress levels.

In all these analysis, we suggest a second order MMPP is sufficient to characterize the time-varying activities of deep seismicity. However, we don't rule out the possibility that the deep seismicity might be able to be characterized by a higher order MMPP when quality data with smaller events included in the catalogue accumulate to a longer period and a detailed analysis can be carried out for lower magnitude threshold.

Theoretically, we suggest that in a hidden Markov environment, e.g. MMPPs, the Fisher information has links with the mutual information rate between the observed process and the underlying Markov process. We conjecture that for given length of observations, the better the intensity rates are separated or the better the marks distributions are separated, the better the estimates will be close to the true values of the parameters when the transition rate matrix Q keeps intact. How well the intensity rates or the mark distributions are separated are measured by the mutual information rate between the observed (marked) point process and the underlying Markov process. The explicit formulas of the mutual information rates for MMPP or MMPP with marks are obtained, see Theorem 7.2 and 7.3. The conjectures are validated through simulation studies in 7.6 and 4.5.

10.2 Questions and Future Studies

We conclude that in a relatively large time scale, the main occurrence patterns of New Zealand deep seismicity are the time-varying behavior of the occurrence frequencies. However, detailed features such as the spatial, temporal clustering and

occasional occurrences of large doublets, which are also crucial for understanding the differences of the mechanisms between the shallow earthquakes and deep earthquakes, are beyond the current framework of MMPP to accommodate.

To enable a more detailed description of the occurrence pattern of the deep earthquakes and at the same time reserve the feature of relatively large time scale time-varying behavior in occurrence rate within the framework of MMPP, the model framework of original MMPP needs to be extended on its own right. Toward this direction, we suggest an extension of MMPP by treating the set of interested variables, e.g. the temporal and(or) spatial clustering distributions as a mark which itself forms a subsidiary point process as illustrated in chapter 6. To apply this model characterizing the clustering properties of the deep earthquakes, it is necessary to classify the events into two categories, the mainshocks and the aftershocks, and designate the event which triggers a specific aftershock. The shortcoming of this method is that it might evoke controversial argument due to that there always exist the possibilities that an event is wrongly designated as an aftershock or an event is falsely designated as an aftershock from a specific event rather than another by any deterministic declustering approach. Another possible approach allowing both large scale time-varying seismic activities and simultaneously together with detailed temporal and spatial clustering is by treating the mainshocks and aftershocks as a mixing process. However, this approach poses a challenge on whether the model is analytically tractable. Recent developments in dealing with parameter estimation problem involved in MMPP through point process martingale, measure change and Clark's transformations by Elliott and Malcolm (2008) is an interesting

extension in methodology. Nevertheless, before any clustering model for the deep aftershocks is proposed, far more than current studies of clustering properties by the second order moment analysis of point process are needed to characterize the decaying properties and the abundance of the deep aftershocks. Again, the abundance of the deep aftershocks may range from totally devoid of any clear following aftershocks to occasionally fully developed aftershock sequence decaying according to Omori's law. The spatial clustering properties are generally not homogeneous and isotropic. These factors complicate those attempts to propose a parametric model for the spatial-temporal clustering of the deep earthquakes.

There are also problems on the possible links between the deep earthquakes and other geological hazards near the surface such as the shallow earthquakes and volcanic activities. Since earthquakes near trenches are expression of plates motion and regional stress release, it is plausible that the occurrence of a large shallow quake might produce significant stress changes both in the outer rise and downdip portions of a plate, thus causing seismic activity there. Mogi (1973) concludes that seismicity tends to migrate downward in the years following, and sometimes proceeding, large subduction zone earthquakes. Lay et al. (1989) and others review the occurrence of relatively large intermediate quakes in the year prior to great shallow earthquakes. However, none of them actually apply any statistical tests to assess the relationship between the shallow and deep activities, see Frohlich (2006).

At last, as we suggested in the previous chapter, the estimation errors for the parameters in the model are not negligible, particularly for the transition rates in the Q matrix. More robust methods are required in this situation.

Bibliography

- [1] Adams, R. D. (1963) Source characteristics of some deep New Zealand earthquakes, New Zealand J. Geol. and Geophys., 6, 209–220.
- [2] Adams, R. D. and Ware, D. E. (1977) Subcrustal earthquakes beneath New Zealand: locations determined with a laterally inhomogeneous velocity model, New Zealand J. Geol. and Geophys., 20, 59–83.
- [3] Aki, K. (1965) Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, Bull. Earthquake Res. Inst. Tokyo Univ. 43, 237–239.
- [4] Albert, A. (1962) Estimating the infinitesimal generator of a continuous time, finite state Markov process, Ann. Math. Statist, 33, 2, 727–753.
- [5] Albert, P. S. (1991) A two-state Markov mixture model for a time series of epileptic seizure counts, Biometrics, 47, 4, 1371–1381.
- [6] Anderson, H. and Webb, T. (1994) New Zealand seismicity: patterns revealed by the upgraded National Seismograph Network, New Zealand J. Geol. and Geophys., 37, 477–493.

- [7] Asmussen, S., Nerman, O. and Olsson, M.(1996) Fitting phase-type distribution via the EM algorithm, *Scandinavian Journal of Statistics*, 23, 419–441.
- [8] Asmussen, S. (2000) Matrix-analytic Model and their Analysis, *Scandinavian Journal of Statistics*, 27, 193–226.
- [9] Asmussen, S. (2003) *Applied Probability and Queues*. Second Edition, Springer.
- [10] Bartlett, M. S. (1963) The Spectral Analysis of Point Processes, *J. R. Statist. Soc., Series B*, 25, 264–296.
- [11] Bartlett, M. S. (1964) The Spectral Analysis of Two-Dimensional Point Processes, *Biometrika*, 51, 299–311.
- [12] Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chain, *Ann. Statist.*, 41, 164–171.
- [13] Bickel, P. J., Ritov, Y. and Rydén, T. (1998) Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models, *Ann. Statist.*, 26, 4, 1614–1635.
- [14] Billingsley, P. (1965) *Ergodic Theory and Information*. Wiley.
- [15] Bolt, B. A. (1993) *Earthquake and Geological Discovery*, Scientific American Library.

- [16] Bremaud, P. (1980) Point Processes and Queues, Martingale Dynamics. Springer.
- [17] Brown, T. C. and Nair, M.G. (1988) A simple Proof of the Multivariate Random Time Change Theorem for Point Processes, J. Appl. Prob, 25, 210–214.
- [18] Capp, O., Moulines, E. and Ryden, T. (2005) Inference in Hidden Markov Models. Springe.
- [19] Cinlar, E. (1975) Introduction to Stochastic Processes. Prentice-Hall.
- [20] Cover, T. M. and Thomas, J. A. (1991) Elements of Information Theory. John Wiley.
- [21] Cox, D. R. and Lewis, P. A. W. (1972) Multivariate point processes, Proceedings of the 6th Berkeley symposium, 401–448.
- [22] Cressie, N. A. C.(1993) Statistics for Spatial Data. John Wiley Sons, Inc..
- [23] Daley, D. J. and Vere-Jones, D. (2003) An Introduction to the Theory of Point Processes. Vol 1, Springer.
- [24] Daley, D. J. and Vere-Jones, D. (2004) Scoring probability forecasts for point processes: the entropy score and information gain, J. Appl. Prob., 41A, 297–312.
- [25] Daley, D. J. and Vere-Jones, D. (2008) An Introduction to the Theory of Point Processes. Vol 2, Springer.

- [26] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *J. R. Statist. Soc., Series B*, 39, 1–38.
- [27] Deng, L. and Mark, J. W. (1993) Parameter estimation for Markov modulated Poisson processes via the EM algorithm with time discretization, *Telecommunication Systems*, 321–338, 1, 1993.
- [28] Douc, R. and Matias, C. (2001) Asymptotics of the maximum likelihood estimator for general hidden Markov models, *Bernoulli*, 7, 3, 381–420.
- [29] Efron, B. and Hinkley, D. V. (1978) Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika*, 65, 3, 457–482.
- [30] Elliott, R. J., Aggoun, L. and Moore, J. B. (1995) *Hidden Markov Models: Estimation and Control*. Springer-Verlag.
- [31] Elliott, R. J. and Malcolm, W. P. (2005) General smoothing formulas for Markov-modulated Poisson observations, *IEEE Trans. Automatic Control*, 50, 1123–1134.
- [32] Elliott, R. J. and Malcolm, W. P. (2008) Discrete-Time Expectation Maximization Algorithms for Markov-Modulated Poisson Process, *IEEE Trans. Automatic Control*, 53, 247–256.
- [33] Evans, M., Hastings, N. and Peacock, B. (2000) *Statistical Distributions*, New York: Wiley-Interscience, Third edition.

- [34] Ephraim, Y. and Merhav, N. (2002) Hidden Markov Processes, IEEE Transactions on Information Theory, 48, 6, 1518–1569.
- [35] Fischer, W. and Meier-Hellstern, K. (1992) The Markov-modulated Poisson process(MMPP) cookbook, Performance Evaluation, 18, 149–171.
- [36] Frohlich, C. (1989) The Nature of Deep-Focus Earthquakes, Ann. Rev. Earth Planet. Sci., 17, 227–254.
- [37] Frohlich, C. (2006) Deep Earthquakes. Cambridge University Press, UK.
- [38] Gibowicz, S. J. (1974) Frequency-Magnitude, Depth, and Time Relations for Earthquakes in an Island Arc: North Island, New Zealand, Tectonophysics, 23, 283–297.
- [39] Gutenberg, B. and Richter, C. F. (1956), Magnitude and energy of earthquakes, Ann. Geophys., 9, 1.
- [40] Harte, D. and Vere-Jones, D. (1999) Differences in Coverage between the PDE and New Zealand local earthquake catalogues, New Zealand J. Geol. and Geophys., 42, 237–253.
- [41] Hughes, J. P. and Guttorp, P. (1999) A non-homogeneous hidden Markov model for precipitation occurrence, J. R. Statist. Soc., Ser. C, 48, 15–30.
- [42] Ito, H., Amari, S. I. and Kobayashi, K. (1992) Identifiability of hidden Markov information sources and their minimum degrees of freedom, IEEE Transactions on Information Theory, 38(2), 324–333.

- [43] Jensen, A. T. (2005) Statistical Inference for Doubly Stochastic Poisson Processes. PhD thesis, University of Copenhagen.
- [44] Karr, A. F. (1991) Point Processes and Their Statistical Inference. Dekker, New York.
- [45] Klemm, A., Lindemann, C. and Lohmann, M. (2003) Traffic modeling of IP networks using the batch Markovian arrival process, *Performance Evaluation*, 54, 2, 149-173.
- [46] Louis, A. (1982) Finding the observed information matrix when using the EM algorithm, *J. R. Statist. Soc., Ser. B*, 44, 2, 226–233.
- [47] Lay, T., Astiz, L., Kanamori, H. and Christensen, D.H. (1989) Temporal variation of large intraplate earthquakes in coupled subduction zones, *Phys. Earth Planet. Int.*, 54, 258–312.
- [48] Lystig, T. C. and Hughes, J. P. (2002) Exact computation of the observed information matrix for hidden Markov models, *Journal of computational and graphical statistics*, 11, 3, 678–689.
- [49] MacDonald, I. L. and Zucchini, W. (1997) Hidden Markov and other models for discrete-valued time series. Chapman and Hall.
- [50] McFadden, J. A. (1965) The entropy of a point Process, *Journal of the society for industrial and applied mathematics*, 13, 4, 988–994.
- [51] McLachlan, G. J. and Krishnan, T. (1997) EM Algorithm and Extensions. Wiley, New York.

- [52] Meier-Hellstern, K. S. (1987) A fitting algorithm for Markov-modulated Poisson processes having two arrival rates, *European Journal of Operational Research*, 29, 370–377.
- [53] Meng, X. L. and Rubin, D. B. (1993) Maximum likelihood estimation via the ECM algorithm: a general framework, *Biometrika*, 80, 267–278.
- [54] Meyer, P. (1971) *Démonstration simplifiée d'un théorème de Knight*, *Lecture Notes in Math.*, 191, 191–195.
- [55] Milne, R. K. (1971) *Stochastic Analysis of Multivariate Point Process*, PhD Thesis, Australian National University.
- [56] Mogi, K. (1973) Relationship between shallow and deep seismicity in the western Pacific region, *Tectonophysics*, 17, 1–22.
- [57] Moler, C. and Van Loan, C. (2003) Nineteen dubious ways to compute the exponential of a matrix, Twenty-Five years later, *Siam Review*, 45(1), 3–49.
- [58] Neuts, M. F. (1979) A Versatile Markovian Point Process, *J. Appl. Prob.*, 16, 4, 764–779.
- [59] Neyman, J. and Scott, E. L. (1958) A statistical approach to problems of cosmology, *J. R. Statist. Soc., Ser. B*, vol. 20, 1–43.
- [60] Oakes, D. (1999) Direct calculation of the information matrix via the EM algorithm, *J. R. Statist. Soc. Ser. B*, 61, 2, 479–482.

- [61] Ogata, Y. (1981) On Lewis' simulation method for point process. *IEEE Transactions on Information Theory*, IT-27, 23–31.
- [62] Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statist. Soc.*, 83, 401, 9–27.
- [63] Ogata, Y. (1998) Space-Time Point-Process Models for Earthquake Occurrences, *Annals of the Institute of Statistical Mathematics*, 24, 379–402.
- [64] Omori, F. (1894) On the aftershocks of earthquakes, *Journal of the College of Science, Imperial University of Tokyo*, 7, 111–200.
- [65] Papangelou, F. (1972) Integrability of expected increments of point processes and a related change of scale, *Trans. Am. Math. Soc.*, 165, 483–506.
- [66] Papangelou, F. (1978) On the entropy rate of stationary point processes and its discrete approximation, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 44, 191–211.
- [67] R Development Core Team (2008) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [68] Rabiner, L. R. (1989), A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE* 77, 257–285.
- [69] Ramesh, N. I. (1995) Statistical Analysis on Markov-Modulated Poisson Processes, *Environmetrics*, 6, 165–179.

- [70] Reyners, M. (1989) New Zealand Seismicity 1964–87: an interpretation, *New Zealand J. Geol. Geophys.*, 32, 307–315.
- [71] Rudemo, M. (1973) Point Process Generated by Transitions of Markov Chains, *Advances in Applied Probability*, 5, 2, 262–286.
- [72] Rydén, T. (1996a) An EM algorithm for estimation in Markov-modulated Poisson processes, *Computational Statistics & Data Analysis*, 21, 431–447.
- [73] Rydén, T. (1996b) On identifiability and order of continuous-time aggregated Markov chains, Markov-modulated Poisson processes, and phase-type distributions, *J. Appl. Prob.*, 33, 640–653.
- [74] Sansom, J. and Thompson, P. J. (2001) Fitting hidden semi-Markov models to breakpoint rainfall data, *J. Appl. Probab.* 38A, 142–157.
- [75] Schoenberg, F. (1999) Transforming spatial point processes into Poisson processes, *Stoch. Proc. Appl.*, 81, 155–164.
- [76] Schorlemmer, D., Wiemer, S. and Wyss, M. (2005) Variations in earthquake-size distribution across different stress regimes, *Nature*, 437, 539–542.
- [77] Statistical Research Associates Limited, Statistical Seismology Library, URL <http://www.statsresearch.co.nz/software.html>.
- [78] Statistical Research Associates Limited, Hidden Markov Models, URL <http://www.statsresearch.co.nz/software.html>.

- [79] Stuart, A., Keith O. and Steven A. (1999) Kendall's Advanced Theory of Statistics. 2A. London: Arnold, a member of the Hodder Headline Group. pp. 25.37-25.43.
- [80] Van Loan, C. F. (1978) Computing integrals involving the matrix exponential, IEEE Trans. Autom. Control, Vol. AC-23, 3, 395-404.
- [81] Vere-Jones, D., Turnovsky, S. and Eiby, G. A. (1964) A statistical survey of earthquakes in the main seismic region of New Zealand, Part 1-time trends in the pattern of recorded activity, New Zealand J. Geol. and Geophys., 7, 4, 722-744.
- [82] Vere-Jones, D. and Davis, R. B. (1965) A statistical survey of earthquakes in the main seismic region of New Zealand, New Zealand J. Geol. and Geophys., 9, 251-284.
- [83] Vere-Jones, D. (1970) Stochastic models for earthquakes occurrence, J. R. Statist. Soc., Ser. B, 32, 1-45.
- [84] Vere-Jones, D. (1978) Space Time Correlations for Microearthquakes: A Pilot Study, Advances in Applied Probability, 10, 73-87.
- [85] Vere-Jones, D. and Schoenberg, F. (2004) Rescaling marked point processes, Aust. N. Z. J. Stat, 46(1), 133-143.
- [86] Wadati, K. (1928) Shallow and deep earthquakes. Geophys. Mag. 1, 161-202.

- [87] Wang, P. M. and Puterman, M. L. (2001) Analysis of longitude data of epileptic seizure counts—a two-state hidden Markov regression approach, *Biometrical Journal*, 43, 8, 941–962.
- [88] Wiemer, S. and Benoit, J. P. (1996) Mapping the b-value anomaly at 100 km depth in the Alaska and New Zealand subduction zones, *Geophysical Research Letters*, 23, 13, 1557–1560.
- [89] Zheng, X. G. and Vere-jones, D. (1991) Application of Stress Release Models to Historical Earthquakes from North China, *PAGEOPH*, 135, 4, 559–576.