STOCHASTIC MODELS FOR MRI LESION COUNT SEQUENCES FROM PATIENTS WITH RELAPSING REMITTING MULTIPLE SCLEROSIS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Xiaobai Li, M.S.

* * * * *

The Ohio State University

2006

Dissertation Committee:

Haikady N. Nagaraja, Adviser

Catherine A. Calder

Kottil W. Rammohan

Thomas J. Santner

Approved by

Adviser

Graduate Program in Statistics

ABSTRACT

Relapsing remitting multiple sclerosis (RRMS) is a chronic and autoimmune disease where the disease states alternate between the relapse and remission. Magnetic resonance imaging (MRI) is widely used to monitor the pathological progression of this disease. The longitudinal T1-weighted Gadolinium-enhancing MRI lesion count sequences provide information on the onset and sojourn time of the lesion enhancement. We construct biologically interpretable queueing models for the longitudinal data of these lesion counts that describe the natural evolution of the lesions. The infinite-server queue with Poisson arrival process and exponential service ($M/M/\infty$) is proposed for this purpose. The rate of the Poisson arrival process can also be allowed to be governed by a two-state hidden Markov chain. We describe the likelihood function for each model based on appropriate assumptions and fit these models to data from 9 RRMS patients. We obtain the maximum likelihood estimators of the parameters of interest arising from these models and study their asymptotic properties through simulation. We discuss the validation of the assumptions for the proposed models and examine the robustness of these estimators. We suggest the application of the models for characterizing the disease progression and testing treatment effect and discuss implication for planning of RRMS clinical trials.

ACKNOWLEDGMENTS

The years in the Ohio State University are very special for me. During the journey to my PhD, I have been very lucky that I have met many people who give me their support. When I try to write down the names of those people to whom I owe my gratitude, I find out that the list is such a long one!

First of all, I want to thank my adviser, mentor, Professor Haikady N. Nagaraja, for his academic as well as spiritual support during the whole research. The conversations with him not only broaden my horizon to statistics knowledge but also on how to become a more open-minded person. I want to thank all my committee members, Professor Kottil Rammohan from the Department of Neurology, Professor Thomas Santner and Professor Catherine Calder (both from the Department of Statistics) for their helpful discussions and feedback on my research. I want to express my special thanks to Professor Rachel Altman from the Department of Statistics at Simon Fraser University, Canada, for sharing the dataset and ideas of her own research. I want to thank Dr. Paul Albert from National Cancer Institute, Bethesda, Maryland, who generously allowed us to explore the modeling of the lesion count sequences from his data on 9 relapsing remitting multiple sclerosis patients. I want to thank Professor Narayan Bhat from Southern Methodist University, Dallas, Texas, for all the insightful discussions about queueing theory.

I have benefited a lot from other experiences in the graduate school. Many thanks would go to Professor Elizabeth Stasny first. In the summer of 2002, she allowed me to take a reading course with her on missing data analysis. This immediately triggered my interest in research. With her full recommendation, I got the internship in Food and Drug Administration. With the support from her and Professor Steve MacEachern, I worked on judgment post-stratification in finite population setting. This research project led to my two presentations in the joint statistical meetings. I also want to thank all the professors in the Department of Statistics for offering wonderful courses and a very competitive program. I want to thank the department for financial support for all these years.

I want to thank Professor Lee Hebert and Professor Daniel Birmingham from the Department of Internal Medicine for the valuable interaction during my association with the Lupus project. Working with them on that project was always pleasant and instructive.

I want to thank all my colleagues and my students from Beijing Jingshan School, Beijing, China. My teaching experience there is a wonderful inspiration for such a journey.

I am indebted to my parents, Bixia Peng and Hongbo Li, who always have confidence in me and always encourage me to look at things on the positive side. I want to thank my fiancé Sherab Chen, who takes good care of me when I work hard. I want my special thanks to go to my family like friends, John Lawrence, Peiling Yang and my 'adopted mother' Monica Lewis, for their loving kindness to me since I met them in 2003. I also feel blessed to have my friend Liang Liu, who has always been there with me through all the ups and downs in the research.

To my mother, Peng Bixia, and my father, Li Hongbo

To all the people who are doing extraordinary things

in the face of Multiple Sclerosis.

VITA

February 1975	Born, Changsha, China
June 1996	BS Probability and Mathematical Statistics Beijing Normal University Beijing, China

FIELDS OF STUDY

Major Field: Statistics

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

Introduction

Multiple sclerosis (MS) is a well known progressive, autoimmune disease affecting the human central nervous system. Relapsing remitting multiple sclerosis (RRMS) is mostly seen among the MS patients in their early disease stages. In clinical studies, other than clinical evidence such as the number of relapses or changes in Expanded Disease Status Scale (EDSS) scores, magnetic resonance imaging (MRI) has been applied to identify lesions in the brain and spinal cord. The MRI is also used to help the diagnosis of MS and track the subclinical activity. Natural evolution history of a lesion is a very complicated process. As the standard measures in clinical trials in RRMS, gadolinium-enhancing (Gd-enhancing) lesion counts provide an index of the inflammatory activity at the time of the MRI scan. Gd-enhancement occurs in almost all new lesions and it lasts for a while. The serial monthly MRI study has the information on the time course of the enhancement. The Gd-enhancing lesion counts have become the primary endpoints in phase II studies and secondary outcome measures in phase III studies for those treatments targeting the inflammatory aspects of the disease.

In Chapter 2, we introduce some basic knowledge about MS and MRI measurements. For Gd-enhancing lesion count data, a statistical review of the models proposed in the literature for the total lesion count sequence or the new lesion count sequence is given. The review covers the negative binomial model for the cross-sectional study proposed by Sormani et al. (1999) and the Markov regression model as well as the hidden Markov models for the longitudinal study proposed by Albert et al. (1994). As far as we know, no model in the literature has ever taken both the new and total lesion counts into account even though such data have been available.

Since the newly enhancing lesions form a portion of the total enhancing lesions, we propose to model the relationship between the new and the total counts across time using a special queueing process as a possible approximation to the biological process. This is done in Chapter 3. The patient is taken as the system, and the lesions are the customers who come in and get service (enhancement). We make assumptions about the arrival distribution and the service distribution. The likelihood functions are derived thereafter. We fit the models to Gd-enhancing MRI lesion count sequences from 9 RRMS patients. We discuss the estimation of the parameters in these models as well as their interpretation. We suggest settings where a particular model may be more appropriate than others.

In Chapter 4, we examine the long-run property for one of our proposed models. Simulation results are provided to show the asymptotic normality of the maximum likelihood estimators for the model parameters.

The validity of the model assumptions are addressed in Chapter 5. We provide formal inference such as likelihood ratio test for nested models. Simulations are used to illustrate the robustness of the model estimation when minor departures from the

assumptions exist. In Chapter 6, we discuss the application of the models. We illustrate how to test disease progression and how to take account of the heterogeneity among RRMS patients using our proposed models. Some power studies are given using simulation. Application issues related to sample size calculation for the experimental design of a RRMS clinical trial are addressed. We summarize the thesis work in Chapter 7 and discuss some potential future work.

CHAPTER 2

Medical and Statistical Literature Review

2.1 Multiple Sclerosis (MS)

Multiple sclerosis (MS) is a common chronic and progressive disease of the white matter in the central nervous system (CNS). In total there are about 1.1 to 2.5 million of MS patients in the world. More than 50% of the cases have occurred in Canada, North America, Australia, Russia and northern Europe. Approximately 250,000 to 350,000 patients are in the United States. This disease tends to be more prevalent in higher latitudes (40 to 60 degrees south and north).

2.1.1 Clinical Characteristics of MS

In general, most of the MS patients start at a relapsing-remitting phase, experiencing different combinations of symptoms occasionally. The symptoms include problems with speech, walking, vision loss, bladder or bowel dysfunctions, dizziness, seizures, etc. Acute attacks or exacerbations are called relapses. Here exacerbation refers to increasing severity of previously observed symptoms. A relapse is defined as the appearance of new symptoms or the reactivation of old ones lasting more than 24 hours in the absence of change in core body temperature or infection (information

from *www.mssociety.org.uk*). Patients can stay in the relapse period for hours, days, even months. The recovery after the relapse is the remission, which may last for years.

The course of MS is very unpredictable for any particular person. Some individuals experience rapid progression to total disability and others may not be affected much by the disease for years. Even within an individual patient, the disease activity varies largely through time. Usually, younger patients with early onset will have a slower disability progression . However, the disease continues with or without clinical attacks as the disability becomes more visible and eventually the remission stops.

2.1.2 Types of MS

There are four basic types of MS which characterize the ongoing disease course:

1. Relapsing remitting multiple sclerosis (RRMS): Patients experience clearly defined relapse episodes followed by a partial or complete recovery (remission). The National MS Society Advisory Committee on Clinical Trials defined RRMS as "clearly defined disease relapses with full recovery or with sequelae and residual deficit upon recovery; periods between disease relapses characterized by a lack of disease progression" (Lublin and Reigold, 1996). For about 85% of MS patients start out with this type.

2. Secondary progressive multiple sclerosis (SPMS): Patients experience an initial period of RRMS followed by steadily worsening disease course with or without occasional relapses or minor remissions. Usually RRMS patients evolve into SPMS.

3. Primary progressive multiple sclerosis (PPMS): Patients experience a slow but nearly continuous worsening of the disease from the onset, with no distinct relapses or remissions. The rate of the progression might be different. It occurs only in about 10% of the MS patients and the patients tend to be older.

4. Progressive relapsing multiple sclerosis (PRMS): Patients experience a progression of the disease from the onset, with occasional clear relapses. About 5% of the MS patients are in this category.

Our research work focuses on the study of RRMS.

2.1.3 Pathological Features of MS

Although MS is widely accepted as an autoimmune disease, the pathological mechanisms are rather complicated (Noseworthy et al., 2000). The CNS consists of the brain and the spinal cord. It sends and receives communication signals through a network of nerves. Both the components of CNS contain white matters, where nerve fibers, also named as axons, are coated with myelin. Myelin is a substance made up of proteins and lipid fats. It forms layers around the nerve fibers, and acts as insulation. Normally, nerve impulses are transmitted as electronic signals through the nerve fibers efficiently in a healthy CNS. Myelin cannot only protect the signals from being lost by the insulation, but also speed up the transmission by containing the signals in a small space surrounding the axon.

In MS patients, demyelination happens as the myelin sheath is stripped off and leaves the axon unprotected. The natural immune cells, T cells, will bind with the lost protein from the demyelination and become pro-inflammatory. These harmful T cells will enter the CNS through the blood-brain barriers (BBBs). Then they start

attacking myelin proteins and that leads to new myelin breakdowns. However, what triggers the immune system to destroy its own myelin remains unknown. The damage to the tissue structure from demyelination is called as a lesion. The MS plaque is often seen randomly distributed in the CNS of the patients. Thus multiple sclerosis is also known as a disease with 'many scars'.

In the CNS, when inflammatory cells attack nerve fibers, edema (a fluid) will collect around the nerve fibers and compress them, which could slow or block the transmission of the signals. Specific functions related to the signals may not work and some symptoms show up as a response. Thus the relapses in the early stage of the disease may be the result of the axonal demyelination. However, the inflammation may not necessarily cause demyelination. It can possibly damage the axons directly. Even with the existence of the demyelination or axonal injury, the axons can regain their conducting ability by redistributing the transmitted electrical current. Thus we can see a lot of silent lesions in the patients. There is no clear-cut answer as to what brings on the relapses in the early phase of the disease.

It is possible for the CNS to repair the broken nerves since the oligodendrocytes (a kind of special cells) can produce myelin provided that they are not killed during the remyelination. As a consequence, some lesions may disappear after the healing process. It is uncertain whether demyelination or axonal injury occurs first. Frequent attacks on myelin make the axon vulnerable. However, axonal injuries can be seen in very early stages of MS patients, where demyelination is uncommon (Rammohan, 2003).

2.1.4 Methods of Monitoring the Progression of MS

In clinical studies, methods such as clinical evaluation and some imaging techniques to detect the subclinical activity have been used to monitor the disease progression of MS.

Clinical Evaluation

Several types of clinical evaluations are applied to monitor the progress of various types of MS. They consist of the degree of disability, occurrences of relapses and time to clinical deterioration. These have been used as the primary outcome measures in the RRMS clinical trials.

Expanded Disability Status Scale (EDSS) score (Kurtzke, 1983) is an ordinal (discontinuous and nonlinear) summary (0 to 10) from the standard neurological examination. It has been used to follow the progression of the MS disability at different regions of the scale. Although it has been accepted as one of the primary measurements in MS phase III clinical trials, the EDSS has some limitations. Firstly, it is insensitive to changes in other neurological functions because of its strong emphasis on ambulatory index. Secondly, it fails to detect the cognitive dysfunctions which can be commonly seen in MS patients. Thirdly, since the measure is not linear in the disease change, it is also relatively insensitive to the changes as the clinical disability deteriorates over time.

The occurrences of relapses are also considered as another clinical manifestation of the disease activity. Relapse rate is a common primary outcome measure in RRMS studies. However, the quantification of relapses suffers from the large variability among MS symptoms. Also, such a clinical manifestation is modest. In their paper,

Cohen and Rudick (2003) summarized that in large RRMS clinical trials with usual duration of 2 to 3 years, most of the patients experienced no relapses or only one relapse.

Magnetic Resonance Imaging

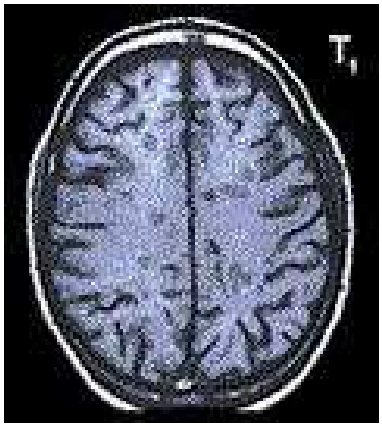
Magnetic resonance imaging (MRI) is an imaging technique to monitor the pathological progression of the disease. It has been increasingly used over the last 15 years in the diagnosis and prognosis of MS.

This evolving tool can produce high quality images of the inside of the human body. Patient is put into a tube of a scanning machine, where a huge magnet will force the protons inside the water molecules in the body to line up and react to the power with signals. The difference between the region-of-interest and the background would be recognized by different signal intensities. For example, a diseased tissue, such as an MS lesion, contains more water than normal tissue and thus gives stronger signals. Technical issues such as how to increase the detectability can be found in the review article by Miller et al. (1998). After the signal analysis based on the images, the neurologists gather the information for the site and size of the lesions. The recognition of the lesions is done manually. This may possibly introduce substantial intra-observer or inter-observer variations because of the subjectivity. Calculating the size or the total volume of the lesions is a semi-automated process. The delineation could be done either by manually outlining the lesion area or by a computer software.

The MRI technology helps researchers make great progress in the exploration of the pathogenesis and the treatment for MS. In 1986, the National Multiple Sclerosis Society (*www.nmss.org*) recommended the optional use of MRI in the diagnosis procedure. It also contributes to the understanding of MS for its sensitivity and the

reproducibility. The evolution of a lesion could be examined on serial MRI scans over time. The round or oval-shaped lesion can extend along the small or medium-sized vessels around them. During the disease course, the lesion appears, followed by subsequent enhancement, enlarges or shrinks. It could disappear but with residual effect left.

There are two common types of MRI scans, one is T1-weighted and the other is T2-weighted. The term 'weighted' is related to the type of contrast that dominates the image. New lesions are detected as bright spots on the enhanced image obtained by using a contrast agent such as gadolinium, whereas the old lesions remain dark in an ordinary T1 scan. In Figure 2.1, we present two T1-weighted MRI pictures taken both before and after the injection of gadolinium. These new lesions reflect

(a) 

(b) 

Figure 2.1: Two T1-weighted MRI scans; (a) without gadolinium (b) with gadolinium (source: *www.thjuland.tripod.com*).

the breakdowns of the BBB, which may be the early abnormality associated with the demyelination. Another type is the hypointense T1 lesions ('black holes'), which may

reveal the axon injury. Since it is suggested that axon loss plays a more important role in the relationship with the clinical disability, the hypointense T1 lesions are considered more harmful.

The standard T2-weighted images are very sensitive for the detection of the lesions. However, the lesions related to edema, inflammation, demyelination, remyelination, and axonal loss would appear to have similar hyperintensity on T2 images. This type of conventional scan lacks of specificity with regard to the heterogeneous pathology of the individual lesions. The T1-weighted images overcome this limitation partially by using the contrast to detect the disruptions of BBB corresponding to the ongoing inflammation.

It is well known that MRI can detect the subclinical activity 5 to 10 times more than the relapses (Miller et al., 1998). Its measures have been used as the primary outcome measures for phase II studies and the secondary outcome measures in phase III studies. However, many studies find out that the MRI abnormalities correlate weakly with the clinical disability. The reason may be that the conventional MRI measures lack the detailed quantification and pathology of the tissue damage directly related to the clinical manifestation.

2.1.5 MRI Data in RRMS Studies

The currently used MRI indices of disease activity include new or enlarging lesions on T1-weighted, T2-weighted, and Gd-enhancing T1-weighted images. These techniques can detect various active lesions and measure the size or volume of the lesion. The common MRI indices are the number of lesions and the total lesion volume.

In MS patients, the lesion activity changes over time. It is impossible to scan the patients frequently considering the cost and scheduling convenience. In most MRI-monitored studies, patients have the scans monthly or on a half-a-year basis. The follow-up often lasts months or years. The two main approaches to MRI are detecting active lesions, and measuring total lesion load. Data are often reported as the following:

- number of lesions seen on sequential scans (e.g., number of new enhancing lesions seen on T2-weighted scan), number of total lesions seen on T1-weighted scan.
- total volume of lesions seen on each scan (e.g., total volume of lesions on hypointense T1-weighted images, total volume of new lesions seen on T2-weighted images).

Here we need to clarify the definition of the new enhancing lesions. The Gd-enhancing T1-weighted images can identify the areas of 'leaky' BBB by comparing the images before and after the injection of the contrast agent. The recognition of the enhancement depends not only on the presence of the disrupting BBB, but also on the relationship between the timing of the scan and the flow of the contrast agent.

Different protocols may have been used to count the number of the new lesions in different studies. In the PRISMS Study (Li et al., 1999), new Gd-enhancing T1-weighted lesions are those that have never enhanced before. Recurrent enhancing T1 lesions are those enhancing and reappearing at the site at which an earlier lesion has disappeared. There are also persistent lesions that have enhanced on consecutive scans. In the study by McFarland et al. (1992), the newly enhancing lesions are those which have not enhanced on any of the previous scans and the recurrently enhancing

lesions. The number of enhancing lesions are used to define the cohorts for entry into clinical trials (Comi et al., 2001). The number of new enhancing lesions on Gd-enhancing images is often used to calculate the sample size for different experiment settings (McFarland et al., 1992; Sormani et al., 2000). The MRI parameters are also very important tools in monitoring the efficacy of the new therapies. Many drugs show their ability to reduce the rate of new lesions as well as the total volume of the lesions (Comi et al., 2001).

The imaging results have played either the major role or a supporting role in testing the drug efficacy in several clinical trials. For example, the total number of enhancing lesions is used as the primary outcome measures in the European/Canadian trial for the effects of GA (Glatiramer Acetate) (Comi et al., 2001). Other MRI results such as the total volume of enhancing lesions and number of new enhancing lesions are used as the secondary outcome endpoints in the PRISMS Study (Li et al., 1999). Our research is mainly about the application of Gd-enhancing MRI lesion count data on T1-weighted scans since RRMS patients show a high rate of inflammatory lesion activity.

2.2 Statistical Literature on MRI Data Analysis for MS

Several modeling methods have been proposed for RRMS lesion count data. We start our review with the methods for modeling cross-sectional lesion count data.

2.2.1 Stochastic Modeling of Cross-Sectional Lesion Count Data

Various common approaches have been used in the statistical inferences based on the lesion count data in cross-sectional studies. To test the drug efficacy, classical

approaches such as the analysis of variance (ANOVA) and the analysis of covariance (ANCOVA) with log or rank transformation are useful when the lesion counts are studied (Li et al., 1999; Comi et al., 2001). In the PRISMS Study (Li et al. 1999), an analysis of variance on the ranks model is used for the outcome of mean number of newly enhancing T1-lesions per patient per scan, incorporating the terms of the center and the treatment group as main effects. Covariates, such as the number of lesions at the baseline scan, age and gender, are also included in the model.

Sormani et al. (1999) propose a negative binomial (NB) model for RRMS studies with Gd-enhancing lesion counts. For illustration purposes, let N_t denote the number of new lesions counted over the fixed time interval $(0, t]$. Different models have been considered to describe the distribution of N_t across patients. The Poisson model is the simplest. It assumes that N_t follows the Poisson distribution with the same mean parameter μt for all the patients. Specifically, it requires that the occurrence of the new lesions should follow a homogeneous Poisson process. However, Sormani et al. (1999) argue that the Poisson assumption is most likely to be violated in the case of the MS lesion count data. Since there is large variability across the RRMS patients, it may not be the case that the observed counts from each individual patient follow the same Poisson distribution. One way to take account of the extra variability is to fit an over-dispersed Poisson model, where the variance of the Poisson variable is adjusted to be proportional to the mean. The dispersion parameter involved in the model produces more flexibility. However, it does not vary the model much from a probabilistic point of view. Thus, Sormani et al. (1999) employ a model where a random effect with a gamma distribution is used to describe the patient-specific variation. Then the marginal distribution of N_t would be negative binomial.

If necessary, an over-dispersed negative binomial model could be used by adding another dispersion parameter in the mean-variance relationship. Both the Poisson model and the negative binomial model (the NB model) are fitted to a cohort of 56 MS patients (36 RRMS and 18 SPMS). From the histogram of the distribution of the number of new lesions, there are many patients having zero new lesion counts. The probability distribution implied by the NB model seems to provide a better fit. It has a smaller deviance from the observed data compared to the Poisson model and the estimated variance for N, the total counts over time, is closer to the sample variance. Thus the NB model is more capable of describing the extra-variability shown in the data.

Sormani et al. (1999) apply their NB model in the sample size calculation for a parallel group design and two-period cross-over design with different duration of the follow-up. They fit the model to the dataset reported in the paper by Nauta et al. (1994), where a nonparametric approach is proposed. Simulation results show that the power estimated based on the NB model is less than the power estimated by the nonparametric technique used by Nauta et al. (1994) for the same experimental set up. This suggests that in order to achieve the same level of power and treatment effect in the same type of design, more patients are needed for the NB model. The nonparametric method shows an overestimation of the power where the lesion variability across subjects is ignored. Sormani et al. (2001) extend their work on sample size calculation for another cohort of RRMS patients for three types of designs: parallel groups design, parallel group with a single baseline correction scan design, and a baseline versus treatment design. The estimated sample size could be reduced when including MRI activity at study entry as a patient selection criterion.

Sormani et al. (2002) also assess the surrogacy of MRI-derived endpoints for the clinical relapse rate in RRMS patients in the context of a clinical trial of Copaxone. Their results suggest that the total number of new Gd-enhancing lesions over the follow-up period and the percentage change in T2 lesion volume over the follow-up period are valid surrogate markers of relapse rate when assessing some of the criteria proposed by Prentice (1989). However, there is no definitive answer available at the moment.

2.2.2 Stochastic Modeling of Longitudinal Lesion Count Data

Understanding the stochastic nature of the longitudinal T1-weighted Gd-enhancing lesion count data is important for the investigation of the disease progression. Unlike the models we reviewed for those cross sectional studies in the previous section, some stochastic models have been provided to get a better understanding of the evolution of the lesion activity, which may be related to the disease course. Albert et al. (1994) apply two classes of models in their paper on the lesion count data of three RRMS patients. Let $Y_1, Y_2, \ldots, Y_m$ denote the sequence of total number of lesions.

The first model they consider is the Markov regression model. Hereafter we use the abbreviation ALB1 as its name for convenience. Suppose μ_t is the Poisson mean of Y_t, ALB1 takes the form

$$\log \mu_t = \beta_0 + \beta_1 f(t, \omega, \alpha) + \sum_{i=1}^{q} \theta_i (\log (y^*_{t-i}) - \beta_0 - \beta_1 f(t-i, \omega, \alpha))$$

where $f(t, \omega, \alpha) = \cos(2\pi t\omega + \alpha)$ and ${y^*}_{t-i} = \max(c, y_{t-i})$ and $0 < c < 1$. The number c is used to adjust the count when y_t is zero. The model is observation-driven in the sense that the distribution for the observation at a given time is specified in terms of earlier observations. It tries to model the autocorrelation structure present in the

time series. This kind of model is an extension of the generalized linear model and the quasi-likelihood approach in longitudinal data analysis. As we see from their application, it has more flexibilities in terms of including covariates and temporal dependence structure into the model.

A harmonic term with a cosine function is used to account for the cyclical trend in the three time sequences. The two cycles may correspond to the relapsing and remitting behavior of the the disease process. For the estimation of the model parameters, they use a Taylor expansion to approximate the nonlinear function f. Some changes have been made on the re-weighted least squares algorithm originally proposed by Zeger and Qaqish (1988). A grid search over the parameter space is used to find the global maxima. McFarland et al. (1992) have applied this model to the same study but in a cross-sectional setting. They use parametric bootstrap to calculate the sample size for various design schemes.

The second model proposed by Albert et al. (1994) is a parameter-driven model (henceforth referred to by the abbreviation ALB2). This approach is more appropriate in the sense of reflecting the relapsing remitting nature of the disease. In this model, the lesion count follows a Poisson distribution where the rate is governed by an underlying unobservable Markov chain (the parameter process), i.e., the total lesion count follows a Poisson distribution where the mean increases or decreases by the same constant from the previous total given the chain. The parameter process here is assumed to be a two-state Markov chain. The Markov chain is an unobservable binary time series. Since the patients were untreated and observed at arbitrary time points in their disease process, it could be assumed that the chain starts in equilibrium. To make the model simpler, the transition probability matrix is also assumed to

be symmetric (i.e., the probability of transition from either state to the other is the same). Thus, the chain starts with equal probability assigned to each state. The mean μ_t can be expressed as a function of the baseline mean μ_0 because of the definition of the states. A modified EM algorithm is developed to find the maximum likelihood estimates.

The ALB1 and ALB2 models are fitted to each sequence from the 3 RRMS patients. The independent Poisson model (referred to by IPOI), which assumes independent Poisson distribution for Y_t, is also fitted. The Akaike Information Criterion (AIC) is used for comparing the models. It is defined as the following:

$$AIC = -2\log\left(\mathcal{L}(\hat{\theta}|x)\right) + 2k \tag{2.2.1}$$

where k is the number of estimated parameters in the model and $\mathcal{L}$ is the likelihood function value under the given MLEs for the parameters θ. For one patient, ALB2 is reduced to IPOI and IPOI is the best fit. Both ALB1 and ALB2 pick up the mean fluctuation in the other two patients, where they are much better than IPOI. Among them, ALB2 is the best for one. But, selecting the better-fitting model for the other is not easy since the AIC values for ALB1 and ALB2 are so close.

Altman and Petkau (2005) visit the same dataset and discuss several extensions of ALB2 in the setting of hidden Markov models (HMMs). In fact, as they state, ALB2 is not a standard HMM since the conditional mean for total lesion count at time t given the hidden state also depends on previous means. However, they suggest that the model could be rewritten as the usual form of an HMM when they rewrite the Markov chain into a two-dimensional Markov chain. The transition probabilities between the states of the new Markov chain depend on time. So it is a non-homogeneous Markov chain, which implies that ALB2 is a non-homogeneous HMM. It would be difficult

to make inferences for the nonhomogeneous HMM using the results available for the homogeneous HMM. They use the statistical analysis methods for homogeneous HMM and mention it to be informal.

Altman and Petkau (2005) generalize ALB2 in several directions:

ALT1 : Generalization of the transition probabilities for the underlying Markov chain: allow different transition probabilities between the two states.

ALT2 : Generalization of the mean structure: allow two different parameters for the amplitude of increase and decrease in the mean.

ALT3 : Incorporating the extensions considered both in ALT1 and ALT2.

ALT4 : Increasing the number of hidden states from two to three where the additional state represents the normal mode.

All the likelihood functions for HMM data can be expressed in a matrix form as MacDonald and Zucchini (1997) explain in their book. This leads to a possibility of direct maximization of the likelihood.

The likelihood ratio tests are implemented for the model comparison among the first three extensions and the original ALB2, since the likelihood ratio test (LRT) can be used when the models are nested. Testing the parameters on the boundary of the parameter space would be a problem, since the regularity conditions for the nice performance of MLE may not be satisfied.

When it concerns the decision of the number of states of the Markov chain (e.g., any comparison between the fit of ALT4 and the other three), LRT cannot be used. Not only the use of AIC, but also the use of the Bayesian Information Criterion (BIC),

which is defined as

$$BIC = -2\log\left(\mathcal{L}(\hat{\theta}|x) + k(\log m)\right) \tag{2.2.2}$$

where k is the number of parameters, $\mathcal{L}$ is the likelihood function and m is the length of the sequence.

2.2.3 Limitations of the Current Models

The nonhomogeneous HMM for the total lesion count sequence proposed by Albert et. al. (1994) and extensions by Altman and Petkau (2005), reviewed in the previous section, have nice structures and good flexibilities. Nevertheless, the Gd-enhancing lesion count data have a very special feature which should deserve our attention and thus may lead to different modeling approaches.

The Gd-enhancing lesions reflect the breakdowns of the BBBs. The harmful T-cells enter CNS through these breakdowns and become proinflammatory. It is well accepted that there is association between the enhancing lesions and the inflammation (Simon, 2003). The disruptions of the BBB are recognized as early events accompanying inflammatory demyelination. Thus the Gd-enhancing lesions are considered to mark the early inflammatory status of the lesions. Their natural history and the underlying pathology are important to the understanding of the subclinical disease activity in the early disease years of an RRMS patient.

Once a lesion is observed enhancing using the contrast agent, it remains visible for a period of time. In a series of weekly MRI scans (Lai et al., 1996), the enhancement

was present for less than 4 weeks in 29% of the lesions and 5% was seen on only one scan. In their monthly study, McFarland et al. (1992) reported that of all the new lesions, 68% had only been observed on one scan, 28% seen on only two consecutive scans, and there was no lesion lasting more than 3 months.

A lesion starts enhancing at some time point, stays in the status of enhancement for a while, then stops being active. For some of the lesions, the damage could be permanent (i.e., the corresponding BBB may not be repaired ever since it is disrupted). In the early stage of RRMS, patients may not expect such a severe damage since the remission possibility is still high. Also, some of the lesions may not recover because of frequent acute attacks. If such a lesion stays in the CNS permanently and is closely related to a functional ability, the long-term symptom may show up as a consequence, although it usually takes a long time before an RRMS patient has total disability.

Generally speaking, it is impossible to continuously monitor the evolution and progression of these lesions. Information such as the exact time of the occurrences of the enhancement and how long the enhancement persists are unknown. The observed sequence of count data is just a discretized form of the true process. The total number of lesions counted at a specific scan consist of two parts: one is the 'leftovers' from the previous scan (i.e., lesions which have been observed enhancing at the previous scan can still be seen enhancing on the current scan); and the other is the new lesions which come between the two consecutive scans and continue their enhancement through. The re-enhancement of a previously seen lesion can be observed if the follow up is more than a year (McFarland et al., 1992) and thus counted as a newly enhancing lesion. The percentage of the number of newly enhancing lesions to the number of total lesions are reported for each patient.

In McFarland et al. (1992), the mean number of the total lesions and the newly enhancing lesions vary a lot among the patients, with a range of 0.5 to 8 and 0.2 to 5.9 respectively. Of the total counts over time, there is a substantial percentage of the 'leftovers' in every patient. This poses a question: does it make sense to discard this piece of information when modeling the stochastic process?

So far to our knowledge, no models have been proposed to take account of the above special feature of the data. The models we have reviewed in Section 2.2.1 and 2.2.2 do not consider the enhancing duration information. Meanwhile, the HMM proposed by Albert et al. (1994) needs some justification for their informal inference. What would be the appropriate analysis methods? This is still an open problem. The accumulation of the hidden states over time may cause the rate of the Poisson distribution of the total count to be too large because $\mu_t = \mu_0 \theta^{(2\sum_{i=1}^{t} z_i - t)}$ where the z_i's are the underlying states of the Markov chain and θ is the change in the mean according to the status of the chain. This accumulation effect would be magnified when θ is moderately large, which appears very unlikely with the available MRI data.

CHAPTER 3

Models With $M/M/\infty$ Structure

When constructing stochastic models for the longitudinal MRI Gd-enhancing lesion count sequences, we want to find a reasonable way of describing the nature of the process. The evolution process of the lesions suggests the following: a lesion first appears enhancing, stays enhanced for a while (the duration of enhancement), and then goes away. In this way, an individual sequence of the total and new lesion counts of an RRMS patient can be viewed as a queueing process.

3.1 Queueing Theory

Queueing theory plays an important role in areas such as industrial engineering, operational research and management science, as it forms mathematical idealizations of systems and predicts systems behavior in the future. In a queueing system, customers arrive, wait for service if it is not immediate, and leave the system after being served. Such a system can be characterized by the following six elements:

- the arrival pattern of customers
- the service pattern
- the queue discipline

- the number of service stages
- the number of servers
- the system capacity (the limitation on the queue length).

A well-known notation, which consists of the characteristics of the arrival pattern, the service pattern and the number of servers, is widely used to describe the main features of the queueing system (Gross, 1974). The following symbols are common in the literature:

- M: the inter-arrival distribution/service distribution is exponential, which corresponds to Poisson arrival process and exponential service duration.
- G: general distribution.
- M^X: batch arrival as a Poisson process where the batch size is denoted by X. Here X could be random.

For example, the notation $M/G/1$ denotes a queue with homogeneous Poisson arrival, general service distribution and only a single server in the system. The notation $M/M/\infty$ represents a queueing system which has a Poisson arrival, exponential service and infinite servers.

3.2 The Queueing Process Analogy

In the queueing systems terminology, our process can be described as follows:

- the patient: the service system
- the lesions: the customers

- the patches of the white matters in CNS: the server location
- the lesion enhancement duration: the service time.

Each potential site of a coming lesion is taken as a server. When a new lesion occurs in the Gd-enhancing scan, there is a corresponding patch in the CNS of the patient. If we can divide the CNS into small evenly distributed blocks, any new lesion will have roughly the same chance to fall into one of them. This makes sense when the RRMS patient is in the early stages of the disease since there are many blocks available and new lesions seem to occur randomly in the CNS. However, in reality it is difficult to define such a block since lesion may occur across two blocks, or several lesions may be crowded in the same block. Thus the definition of the server here is vague. However, most of the new lesions will start enhancing upon the arrivals. The assumption of infinite-server queue would be appropriate since there are abundant sites in the CNS available for new lesions and the new lesions need no waiting period before being served; that is, enhancement occurs immediately. In a monthly MRI Gd-enhancing lesion count sequence, each month except for the first one, the MRI scans will report the total number of Gd-enhancing lesions (i.e., how many 'customers' are in service at the time). The subset of these 'customers' would be the newly-enhancing lesions. It is obvious that the total lesion count is always bigger than the new lesion count. The above discussion enables us to pursue an approach where the queueing processes has infinite servers.

3.2.1 Infinite-Server Queues

Infinite-server queues play an important role in queueing theory since they could be used to represent a mathematical idealization of systems with many servers. If

the many-server system has a small fraction of time for all servers being busy, the infinite-server queue provides a good approximation. The infinite-server queueing models are well known for their application in the telecommunication environment, where most of the systems have ample servers. Most of the customers enter service upon their arrivals immediately. Therefore the chance of a waiting line for service is rare. Many results for finite-server queues can be generalized to the infinite-server queues (Gross, 1974). Newell (1982) considered an infinite-server queue as a stochastic system and argued that the way to analyze it could be quite different from the traditional methods used in queueing theory. However, his methods provide information about different aspects of the system and do not conflict with others. Commonly seen infinite-server queues are $M/M/\infty$, $M/G/\infty$ and $M^X/M/\infty$. The extension includes $MMPP/M/\infty$, where $MMPP$ denotes the Markovian-modulated Poisson process, where the arrival rate of the Poisson process varies according to a finite state irreducible continuous time Markov chain. Statistical inference methods include method-of-moment and maximum likelihood method. Most of the time, the properties of the queueing systems are explored by simulations. Bhat et al. (1997) provide an overview of the applications of these methods in queueing literature.

3.2.2 Observations from the Queueing Process

How the queueing process is observed is important as it affects the statistical inference procedures. For a queueing system, the data could be observed in the following two ways: the event history data and the time slicing data. The former refers to systems that are continuously observed. The inter-arrival time between the

customers and the duration of each service are recorded. Parametric methods are helpful in the search of the appropriate distributions for the inter-arrival time and the service time. The time slicing datasets are more difficult to work with.

There are two types of time slicing depending on the observed time points. The system could be observed at any arbitrary time point by an 'outsider' who is irrelevant to the system. Or the system could be observed by an 'insider', i.e., at each arrival or each departure time point. If the inter-arrival distribution is i.i.d. exponential, which is the case of Poisson arrival process, these two time-slicing ways would have no difference. The property of PASTA (Poisson arrivals see times average) is guaranteed when the arrival process is a Poisson process. See Cooper (1981) for a good discussion of this property.

Our Gd-enhancing MRI dataset is unique in terms of the way it is observed. It falls into the category of time slicing data where the queueing process is observed at regularly spaced time intervals (i.e., monthly MRI scans). Each time, the number of lesions in the service is observed and each lesion is also identified as a newly enhancing lesion or a persistently enhancing lesion when compared to the preceding scan.

3.2.3 System Size Distribution of the Queueing Process

In the application of the infinite-server queues, the system size process $\mathbf{Q} = \{Q(t), t > 0\}$ is of utmost interest, where $Q(t)$ is the number of busy servers (or the number of customers in service) at time t. In telecommunication applications the tail probability of $\mathbf{Q}$ is used to measure the performance of the system since such a probability represents the chance of the loss of potential customers when all the servers are busy.

In our MRI Gd-enhancing lesion count sequence, the longitudinal total counts denotes a discretized observation of $\mathbf{Q}$. Thus the system size distribution is also relevant to our setting. In the literature, there have been results available for some of the queueing systems.

The System Size Distribution of the $M/G/\infty$ System

The system size distribution of an $M/G/\infty$ system is available in the queueing theory literature.

Theorem 3.2.1. *For an $M/G/\infty$ queue, suppose the rate for the arrival process is λ, G is the cumulative distribution function (CDF) for the service distribution, and the queue is observed from when the system is empty. Then $Q(t)$ follows a Poisson distribution with mean $\lambda \int_0^t \bar{G}(s)\, ds$, where $\bar{G}(s) = 1 - G(s)$.*

The proof can be found in Medhi (2003; 314-316). The following comments are made on the implications of Theorem 3.2.1.

Note 1: For $M/M/\infty$, suppose the service rate is μ (i.e., the mean service time is $\frac{1}{\mu}$). Then $Q(t)$ has a Poisson distribution with mean $\frac{\lambda}{\mu}(1 - e^{-\mu t})$.

Note 2: As $t \to \infty$, $Q(t)$ has a steady-state distribution. It is Poisson with mean $\frac{\lambda}{\mu}$. It seems that when the queue reaches its equilibrium state, only the mean of the service distribution remains important, but not the distributional form of G.

Note 3: Although $Q(t)$ has a Poisson distribution with mean depending on time t, $\mathbf{Q}$ is not a nonhomogeneous Poisson process as Medhi (2003) stated in his book. The process does not have independent increments, i.e., $Q(t_1)$ and $Q(t_2) - Q(t_1)$

are not independent for $t_1 < t_2$. The covariance between these two random variables is not zero. The theorem that characterizes the covariance structure of the process $\mathbf{Q}$ for $M/G/\infty$ queue can be found in Eick et al. (1993).

The System Size Distribution of the $M/M/\infty$ System

For $M/M/\infty$, the distribution for $Q(t)$ has closed form even when the queue starts at an arbitrary time point. The process $\{Q(t), t > 0\}$ is a Markov process because of the following: the service distribution is exponential henceforth it is memoryless as is the arrival process. Suppose an individual customer is observed in service at time t, the amount of remaining service time is independent of the elapsed service time so far. The services for each customer are independent. The distribution of $Q(t + \triangle t)$ would depend on the information from $Q(t)$ as well as the arrival rate and the service rate. The discretized form $\{Q(t), t \in \mathbf{Z}^+\}$ is a Markov chain with countable state space $\mathbf{S} = \{0, 1, 2, \ldots\}$. Here we use the notation $\mathbf{Z}^+$ for the set of all nonnegative integers. The Markovian nature of the process is helpful when making inference because we can write down the likelihood function based on the Markovian property.

If we start observing the queue when it is not empty, we still can get closed form for the distribution of $Q(t)$ in $M/G/\infty$ queues. However, there is no Markovian property in this case since the remaining service time distribution will depend on the general service distribution G and thus affects the length of the customer's stay in the system.

In both cases of $MMPP/M/\infty$ and $M^X/M/\infty$, the distribution of $Q(t)$ is also available but has more complicated forms. Since it is not easy to solve the set of Chapman-Kolmogorov differential equations, the probability generating function has

been used. The process $\mathbf{Q}$ does not have nice properties for such systems. Take the $M^X/M/\infty$ for example. Although the batches of customers arrive as a Poisson process, the distribution of the batch size will have an effect on the number of observed customers, which is not Poisson. We expect that the customers from the same batch will have similar activity compared to the customers from different batches. For the customers observed at t, the independence between the services may not be true. Without PASTA, methods based on embedded Markov chain or embedded Markov renewal process have been applied when the queue size process is viewed at the arrival or departure time points. For the distribution of $Q(t)$ in $M^X/M/\infty$ and $MMPP/M/\infty$ queues, the results can be found in Medhi (2003) and Fischer and Meier-Hellstern (1992), respectively.

3.2.4 Characterization of the Biological Queueing Process for the Gd-enhancing MRI Lesion Count Data

In order to tailor the notion of infinite-server queue to our Gd-enhancing MRI lesion count data, we have to specify the arrival process and the service pattern. Our total lesion count dataset serves as an observed $\{Q(t), t \in \mathbf{Z}^+\}$. Thus our interest focuses on the transient behavior of the queue, the system size distribution. Practically, it is unclear what type of queueing process would be appropriate for the biological process. The exponential distribution has been used for telephone-call durations for a long time not only because it is mathematically tractable, but also because real data support its use. The way that new lesions get enhancing service upon their arrivals is similar to the telephone-call setting. Thus it may be appropriate for us to use exponential distribution to approximate the enhancement duration. The Poisson

process is a viable model when the customers originate from a large population independently. The lesions can occur randomly in the white matter of the CNS. Thus Poisson process is taken as the approximate input process for our queueing model. On the one hand, the rate for the Poisson process could be constant over a period of time since some RRMS patients would have a relatively mild and stable disease courses with very low frequencies of relapse rate. During the 2-3 years of a large clinical trial, most RRMS patients experienced no relapse or one relapse (Cohen and Rudick, 2003). On the other hand, for the RRMS patients with a more active disease course, the rate of the Poisson process could be governed by the alternating disease status, remission and relapse.

Here, we make many idealizations (Poisson arrivals, exponential services) in the mathematical formulations. Our goal is to construct a model that provides a reasonably good approximation to the biological process and allows tractable mathematical computation. This sort of notion encourages us to adopt the idea of $M/M/\infty$ queue.

3.3 Models with $M/M/\infty$ Structure

In this section, we will develop the model which incorporates the $M/M/\infty$ structure. There are two types of such models: one is the discretized version of the $M/M/\infty$ queue, the other is the process which has an $M/M/\infty$ queue structure on the Markov regime, where the arrival rate of the queue is governed by a finite state space Markov chain. We write down the likelihood functions for each case. These models are fitted to each of the Gd-enhancing MRI lesion count sequences taken from 9 RRMS patients. Maximum likelihood estimates of the parameters are calculated. In Section 3.5, we provide a discussion of the model fitting results.

3.3.1 Model Notations

For the monthly MRI data from an individual RRMS patient, suppose at time $k, k = 2, 3, \ldots, m$, we observe $\mathbf{Y}_k = (Y_k^{(1)}, Y_k^{(2)})$, where

- $Y_k^{(1)}$: the number of lesions observed both at time k and $k-1$ (i.e., the number of 'old' lesions).
- $Y_k^{(2)}$: the number of 'new' lesions observed at time k (i.e., the lesions which start enhancing after time $k-1$ and are observed being enhancing at time k).

Also $Y_k = Y_k^{(1)} + Y_k^{(2)}$ denotes the total number of enhancing lesions observed at time k. At the baseline, we only observe the total number of enhancing lesions Y_1. Thus the observed data would be $(Y_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m)$.

3.3.2 The $M/M/\infty$ Model

As we discussed in Section 3.2.4, for those patients whose disease process is relatively stable and mild, the $M/M/\infty$ queue might be appropriate. we assume that the arrival distribution for the queue input is Poisson with monthly rate λ and the service distribution is exponential with monthly rate μ. We call this model Model 1. we can view the structure of the observed stochastic process in the following diagram (Figure 3.1).

From the diagram we can see that for each time, the number of new enhancing lesions $Y_k^{(2)}$ would be independent of the previous total number of enhancing lesions Y_{k-1} and the number of 'old' lesions $Y_k^{(1)}$. However it will affect the following number of 'old' lesions $Y_{k+1}^{(1)}$, since it is part of the total number of enhancing lesions at time k. Each time point, the number of 'old' lesions depends on the total number of enhancing lesions observed in the previous month.

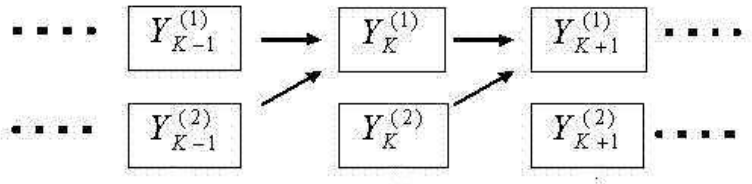


Figure 3.1: Diagram for the structure of the $M/M/\infty$ model.

The Likelihood Function

For the $M/M/\infty$ model, If we assume the process to be stationary at the beginning of the observed sequence, that is, its steady-state distribution has been achieved, we have the following:

$$Y_1 \backsim Poisson(\frac{\lambda}{\mu}). \tag{3.3.1}$$

This is a reasonable assumption as the starting time point of the disease process is quite remote when compared to the time period during which the MRI scans are taken. In addition, we have

$$Y_k^{(1)}|Y_{k-1} \backsim Binomial(Y_{k-1}, e^{-\mu}) \tag{3.3.2}$$

$$Y_k^{(2)} \backsim Poisson(\frac{\lambda}{\mu}(1-e^{-\mu})) \tag{3.3.3}$$

where $k = 2, 3, \ldots, m$, and these two random variables are independent. The binomial distribution is obvious since the exponential service distribution allows us to ignore the effect of previous service time on the remaining service duration. The probability that the service duration will be longer than s is

$$P(X > s) = 1 - (1 - e^{-\mu s}).$$

All of the enhancing lesions have a probability $e^{-\mu s}$ being observed enhanced in the next MRI scan. In our setting, s is 1 because the unit interval is 1 month. For the new enhancing lesions we can observe the following:

$$\begin{aligned}
&P(\text{a lesion arriving in } (0,t) \text{ is still enhancing at time } t) \\
&= \int_0^t P(\text{the service time} > t-s | \text{the lesion arrives at } s) f_T(s) ds \\
&= \int_0^t \frac{1}{t} P(\text{service time} \geq t-s) ds \\
&= \int_0^t \frac{1}{t} e^{-\mu(t-s)} ds \\
&= \frac{1-e^{-\mu t}}{\mu t}, \qquad (3.3.4)
\end{aligned}$$

where $f_T(s)$ is the is the density function for the arrival time of the lesion given that it has arrived in $(0,t)$. This random variable is uniform over $(0,t)$ (See Medhi, 2003, p. 29). The number of lesions that arrive between the interval of two consecutive monthly MRI scans and caught seen in the service will be Poisson with rate $\frac{\lambda}{\mu}(1-e^{-\mu t})$. Thus, with $t=1$, we obtain (3.3.3).

Now we are ready to derive the loglikelihood:

$$\begin{aligned}
&\log \mathcal{L}(\lambda, \mu | Y_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots, \mathbf{Y}_m) \qquad (3.3.5) \\
&= \log P(\mathbf{Y}_2, \mathbf{Y}_3, \ldots, \mathbf{Y}_m | \lambda, \mu, Y_1) P(Y_1 | \lambda, \mu) \\
&= \log \{ P(Y_1) P(Y_k^{(1)}) P(Y_2^{(2)} | Y_1) \prod_{k=3}^{m} [P(Y_k^{(1)}) P(Y_k^{(2)} | Y_{k-1})] \} \\
&= \sum_{k=2}^{m} \log \frac{\exp(-\frac{\lambda(1-e^{-\mu})}{\mu}) (\frac{\lambda(1-e^{-\mu})}{\mu})^{Y_k^{(2)}}}{Y_k^{(2)}!} \\
&+ \sum_{k=2}^{m} \log \binom{Y_{k-1}}{Y_k^{(1)}} e^{-\mu Y_k^{(1)}} (1-e^{-\mu})^{(Y_{k-1}-Y_k^{(1)})} \\
&+ \log \frac{e^{-\frac{\lambda}{\mu}} (\frac{\lambda}{\mu})^{Y_1}}{Y_1!}
\end{aligned}$$

Let

$$\begin{aligned}\frac{\lambda(1-e^{-\mu})}{\mu} &= u \\ 1-e^{-\mu} &= v,\end{aligned}$$

where $u > 0$ and $v \in (0,1)$. It is obvious that

$$\begin{aligned}\log \mathcal{L} =& \log \frac{e^{-\frac{u}{v}}(\frac{u}{v})^{Y_1}}{Y_1!} + \sum_{k=2}^{m} \log [\binom{Y_{k-1}}{Y_k^{(1)}} (1-v)^{Y_k^{(1)}} v^{Y_{k-1}-Y_k^{(1)}}] + \sum_{k=2}^{m} \log \frac{u^{Y_k^{(2)}} e^{-u}}{Y_k^{(2)}!} \\ =& -\frac{u}{v} + Y_1 \log(\frac{u}{v}) + \log(1-v) \sum_{k=2}^{m} Y_k^{(1)} + \log v \sum_{k=2}^{m} (Y_{k-1} - Y_k^{(1)}) + \\ & \log u \sum_{k=2}^{m} Y_k^{(2)} - (m-1)u - \log Y_1! + \sum_{k=2}^{m} \log \binom{Y_{k-1}}{Y_k^{(1)}} - \sum_{k=2}^{m} \log Y_k^{(2)}!.\end{aligned} \tag{3.3.6}$$

Parameter Estimation

We take the first derivative of the loglikelihood with respect to u and v to obtain

$$\begin{aligned}\frac{\partial \log \mathcal{L}}{\partial u} &= -\frac{1}{v} + \frac{1}{u} Y_1 + \frac{\sum_{k=2}^{m} Y_k^{(2)}}{u} - (m-1) \\ \frac{\partial \log \mathcal{L}}{\partial v} &= \frac{u}{v^2} - \frac{Y_1}{v} - \frac{\sum_{k=2}^{m} Y_k^{(1)}}{1-v} + \frac{\sum_{k=2}^{m} (Y_{k-1} - Y_k^{(1)})}{v}.\end{aligned}$$

Upon setting these two expressions to 0, we will have:

$$-u + v(Y_1 + \sum_{k=2}^{m} Y_k^{(2)}) = uv(m-1) \tag{3.3.7}$$

$$u + v[\sum_{k=3}^{m} Y_{k-1} - \sum_{k=2}^{m} Y_k^{(1)}] = \frac{v^2}{1-v} \sum_{k=2}^{m} Y_k^{(1)}. \tag{3.3.8}$$

Let

$$\begin{aligned}A &= Y_1 + \sum_{k=2}^{m} Y_k^{(2)} \\ B &= \sum_{k=3}^{m} Y_{k-1} - \sum_{k=2}^{m} Y_k^{(1)} = \sum_{k=2}^{m-1} Y_k^{(2)} - Y_k^{(1)} \\ C &= \sum_{k=2}^{m} Y_k^{(1)}.\end{aligned}$$

From (3.3.7), we conclude

$$u = \frac{Av}{v(m-1)+1}$$

and upon using this in (3.3.8), we obtain

$$(m-1)(B+C)v^2 - [B(m-2)-(A+C)]v - (A+B) = 0. \tag{3.3.9}$$

There are two roots for the above quadratic equation given by

$$v = \frac{B(m-2)-(A+C) \pm \sqrt{[B(m-2)-(A+C)]^2 + 4(m-1)(B+C)(A+B)}}{2(m-1)(B+C)}.$$

It is easy to choose the appropriate solution since v should be greater than 0. Note that while B can be negative, $A + B \geq Y_1 + \sum_{k=2}^{m-1} Y_k^{(2)} - Y_m^{(1)} \geq 0$ as the number of old lesions at time m cannot exceed the total number of new lesions arrived by time $(m-1)$. Clearly $(B+C) > 0$. Thus

$$|B(m-2)-(A+C)| \leq |\sqrt{[B(m-2)-(A+C)]^2 + 4(m-1)(B+C)(A+B)}$$

and consequently the larger root is the only eligible solution. In other words,

$$\hat{u} = \frac{A\hat{v}}{(m-1)\hat{v}+1} \tag{3.3.10}$$

and

$$\hat{v} = \frac{B(m-2)-(A+C) + \sqrt{[B(m-2)-(A+C)]^2 + 4(m-1)(B+C)(A+B)}}{2(m-1)(B+C)} \tag{3.3.11}$$

provide the solution to the likelihood equation.

We will check the conditions to see if such a set of $(\hat{u}, \hat{v})$ is the local maxima. From the following equations

$$\begin{aligned}
\frac{\partial \log^2 \mathcal{L}}{\partial u^2} &= -\frac{A}{u^2} \\
\frac{\partial \log^2 \mathcal{L}}{\partial v^2} &= -\frac{2u}{v^3} - \frac{B}{v^2} - \frac{C}{(1-v)^2} \\
\frac{\partial \log^2 \mathcal{L}}{\partial u \partial v} &= \frac{1}{v^2},
\end{aligned}$$

we can see that the second-order partial derivative with respect to u is negative. The Jacobian of the second-order partial derivatives evaluated at $(\hat{u}, \hat{v})$ is

$$\begin{aligned}
&\left| \begin{matrix} -\frac{A}{u^2} & \frac{1}{v^2} \\ \frac{1}{v^2} & -\frac{2u}{v^3} - \frac{B}{v^2} - \frac{C}{(1-v)^2} \end{matrix} \right|_{\hat{u},\hat{v}} \\
&= \frac{2A}{uv^3} + \frac{AB}{u^2v^2} + \frac{AC}{u^2(1-v)^2} - \frac{1}{v^4} \\
&= \frac{v(m-1)+1}{Av} \cdot \frac{2A}{v^3} + \frac{(v(m-1)+1)^2}{A^2v^2} \cdot \frac{AB}{v^2} + \frac{(v(m-1)+1)^2}{A^2v^2} \cdot \\
&\frac{AC}{(1-v)^2} - \frac{1}{v^4} \\
&= \frac{1}{v^4}(2(v(m-1)+1)-1) + \frac{AB(v(m-1)+1)^2}{A^2v^4} + \frac{AC(v(m-1)+1)^2}{A^2v^2(1-v)^2} \\
&> 0
\end{aligned}$$

since $2[v(m-1)+1]-1 > 0$ and all the other terms are nonnegative. Thus the needed sufficient conditions are satisfied and we have found the maximum. Since λ and μ are one-to-one transformations of u and v, the corresponding MLE for them would be

$$\hat{\lambda} = \frac{\hat{u}}{\hat{v}} \cdot \log(1-\hat{v}) \tag{3.3.12}$$

$$\hat{\mu} = \log(1-\hat{v}) \tag{3.3.13}$$

with $\hat{u}$ defined in (3.3.10) and $\hat{v}$ defined in (3.3.11). The corresponding variance covariance matrix $\hat{\Sigma}$ would be approximated by $MH^{-1}M^T$ where H is the observed information matrix for the bivariate random variables $(\hat{u}, \hat{v})$, i.e.,

$$H = -\begin{pmatrix} -\frac{A}{u^2} & \frac{1}{v^2} \\ \frac{1}{v^2} & -\frac{2u}{v^3} - \frac{B}{v^2} - \frac{C}{(1-v)^2} \end{pmatrix}_{\hat{u},\hat{v}},$$

and

$$M = \begin{pmatrix} \frac{\partial\hat{\lambda}}{\partial u} & \frac{\partial\hat{\lambda}}{\partial v} \\ \frac{\partial\hat{\mu}}{\partial u} & \frac{\partial\hat{\mu}}{\partial v} \end{pmatrix}_{\hat{u},\hat{v}}$$

$$= \begin{pmatrix} \frac{\log(1-v)}{v} & -\frac{\log(1-v)}{v^2} + \frac{1}{(1-v)v} \\ 0 & -\frac{1}{(1-v)} \end{pmatrix}_{\hat{u},\hat{v}}. \tag{3.3.14}$$

Under some circumstances, the estimator $\hat{v}$ given in (3.3.11) will fall on the boundary of the parameter space $(0, 1)$. For example, when $C = 0$, there are two possibilities. If $B = 0$, then the quadratic equation (3.3.9) reduces to a linear equation $Av = A$. If $B \neq 0$, then (3.3.11) reduces to $\hat{v} = 1$. Thus the corresponding $\hat{\mu}$ would be ∞ and $\hat{\lambda}$ would be 0. These estimates are on the boundary of the parameter space. But these extreme cases occur when $C = 0$ or the total number of old lesions observed is 0. In other words the lesions are cleared very fast.

3.3.3 The Model with $M/M/\infty$ Structure on the Markov Regime

Another way to incorporate the $M/M/\infty$ structure in the model is motivated from the hidden Markov models. The RRMS patients experience the relapse and remission in the disease course. When a patient is experiencing a relapse, we may expect more enhancing lesions. When he/she is in the remission, we may expect less lesions. Enhancing brain lesions occur more often during clinical relapse (Miller and Frank, 1998). Thus the lesion arrival rate may be governed by an unobserved Markov

chain $\mathbf{X} = \{X_k, k = 1, 2, \ldots, m\}$ with state space $\mathcal{B} = \{0, 1\}$. Here we need to clarify that this setting is different from the $MMPP/M/\infty$ queue. It is very difficult to deal with time-slicing data according to this specific type of queue compared to the event history data. There is no information about when the switching between the states for the Markov chain occurs and the Markov chain is hidden. We have to assume the time of transition between the states matches with the observing time point. As long as the changes between the states are not too frequent, we may still be able to get a good approximation.

The graphical representation of the model is illustrated in Figure 3.2. It shows that the underlying Markov chain $\mathbf{X}$ has a direct impact on the new lesions at the first level. At the second level, the current old $Y_t^{(1)}$ would be influenced by the previous total Y_{t-1}, which consists of the old $Y_{t-1}^{(1)}$ and the new $Y_{t-1}^{(2)}$.

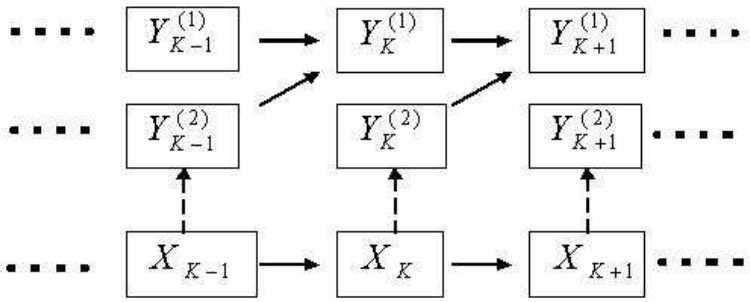


Figure 3.2: Diagram for the structure of the $M/M/\infty$ model with the Markov regime.

The Likelihood Function

Suppose the hidden Markov chain $\mathbf{X}$ has the following transition probability matrix (TPM):

$$\mathbf{P} = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}. \tag{3.3.15}$$

Since the model has the same loglikelihood values under the permutation of the states, we restrict state 1 to be the state with higher arrival rate of the lesions and state 0 to have the lower arrival rate. Thus a is the probability from a 'high' to 'high' and b is the probability from a 'low' to 'low'. Conditional on this chain, we have

$$\begin{aligned} Y_k^{(1)}|Y_{k-1} &\backsim Binomial(Y_{k-1}, e^{-\mu}) \\ Y_k^{(2)} &\backsim Poisson(\frac{\lambda(X_k)}{\mu}(1-e^{-\mu})) \end{aligned}$$

where

$$\lambda(X_k) = \begin{cases} \lambda_1, & X_k = 1 \\ \lambda_2, & X_k = 0. \end{cases}$$

Here λ_1 corresponds to the higher arrival rate and λ_2 corresponds to the lower arrival rate. The likelihood is

$$\begin{aligned} &\mathcal{L}(\lambda, \mu|Y_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots, \mathbf{Y}_m) \\ &= \sum_{X_1 \in \mathcal{B}} \sum_{X_2 \in \mathcal{B}} \cdots \sum_{X_m \in \mathcal{B}} [P(Y_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots, \mathbf{Y}_m, X_1, X_2, \ldots, X_m|\lambda, \theta, \mathbf{P})] \\ &= \sum_{X_1 \in \mathcal{B}} \sum_{X_2 \in \mathcal{B}} \cdots \sum_{X_m \in \mathcal{B}} [P(X_1)P(Y_1|X_1) \cdot \prod_{k=2}^{m} P(\mathbf{Y}_k|\mathbf{Y}_{k-1}, X_k) \cdot \\ &\prod_{k=2}^{m} P_{X_{k-1}, X_k}] \end{aligned} \tag{3.3.16}$$

$$= \sum_{X_1 \in \mathcal{B}} \sum_{X_2 \in \mathcal{B}} \cdots \sum_{X_m \in \mathcal{B}} [P(X_1) \prod_{k=2}^{m} P_{X_{k-1},X_k} \frac{e^{-\frac{\lambda(X_1)}{\mu}} (\frac{\lambda(X_1)}{\mu})^{Y_1}}{Y_1!}$$
$$\prod_{k=2}^{m} \frac{e^{-a_k} {a_k}^{Y_k^{(2)}}}{Y_k^{(2)}!} \binom{Y_{k-1}}{Y_k^{(1)}} e^{-\mu Y_k^{(1)}} (1 - e^{-\mu})^{(Y_{k-1} - Y_k^{(1)})}] \tag{3.3.17}$$

where

$$a_k = \frac{\lambda(X_k)(1 - e^{-\mu})}{\mu}.$$

In fact, we can write the likelihood in the form of a matrix product. This provides us a convenient way for the direct maximization of the loglikelihood. The matrix product form is

$$\mathcal{L}(\lambda, \mu | Y_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots, \mathbf{Y}_m) = \boldsymbol{\pi}_{x_1} \mathbf{D}(Y_1) \prod_{k=2}^{m} (\mathbf{P}\mathbf{G}(\mathbf{Y}_k | \mathbf{Y}_{k-1})) \mathbf{1}. \tag{3.3.18}$$

where $\boldsymbol{\pi}_{x_1} = (\pi_1, \pi_2)$ is the row vector of the equilibrium distribution for the hidden Markov chain which satisfies the following conditions:

$$\boldsymbol{\pi}_{x_1} \mathbf{P} = \boldsymbol{\pi}_{x_1}$$

$$\pi_1 + \pi_2 = 1,$$

$\mathbf{D}(Y_1)$ is a diagonal matrix of Poisson densities with mean parameters $\frac{\lambda_1}{\mu}$ and $\frac{\lambda_2}{\mu}$ and $\mathbf{G}(\mathbf{Y}_k | \mathbf{Y}_{k-1})$ is a diagonal matrix with the diagonal elements $g_1(\mathbf{Y}_k | \mathbf{Y}_{k-1})$ and $g_2(\mathbf{Y}_k | \mathbf{Y}_{k-1})$) defined as the following:

$$g_1(\mathbf{Y}_k | \mathbf{Y}_{k-1}) = \frac{\exp(-\frac{\lambda_1(1-e^{-\mu})}{\mu})(\frac{\lambda_1(1-e^{-\mu})}{\mu})^{Y_k^{(2)}}}{Y_k^{(2)}!} \binom{Y_{k-1}}{Y_k^{(1)}} e^{-\mu Y_k^{(1)}} (1 - e^{-\mu})^{(Y_{k-1} - Y_k^{(1)})}$$

$$g_2(\mathbf{Y}_k | \mathbf{Y}_{k-1}) = \frac{\exp(-\frac{\lambda_2(1-e^{-\mu})}{\mu})(\frac{\lambda_2(1-e^{-\mu})}{\mu})^{Y_k^{(2)}}}{Y_k^{(2)}!} \binom{Y_{k-1}}{Y_k^{(1)}} e^{-\mu Y_k^{(1)}} (1 - e^{-\mu})^{(Y_{k-1} - Y_k^{(1)})}.$$

In the above model setting, we can have several variations:

- The TPM $\mathbf{P}$ of the two-state Markov chain, defined in (3.3.15), could be symmetric (i.e., $a = b$). It implies that on the average, the patient would stay in relapse mode during half of the disease course. This could happen when the RRMS patient is in a relatively stable alternating disease course for a while. When Albert et al. (1994) fitted their Markov regression model to the individual total enhancing lesion count sequence, they identified the cyclical trend and used a sine curve to model the trend. This also gives support to the use of symmetric TPM.

- The state space $\mathcal{B}$ could have three components, $0, -1, 1$ instead of two. The three states correspond to three disease status: stable, remission and relapse respectively. Here 'stable' is the intermediate state between remission and relapse. The rate for the arrival process could be defined as the following:

$$\lambda(X_k) = \begin{cases} \lambda\theta, & X_k = 1 \\ \lambda, & X_k = 0 \\ \frac{\lambda}{\theta}, & X_k = -1. \end{cases}$$

From here after we use Model 2 to denote the model with an underlying two-state Markov chain with a symmetric TPM, and Model 3 to denote the one with an underlying two-state Markov chain with a nonsymmetric TPM. We do not pursue the model with an underlying three-state Markov chain which has more parameters and the computation is more complicated.

3.4 The MRI Lesion Count Data from 9 RRMS Patients

Figure 3.3–3.5 provide the plots of new and total lesion counts for 9 RRMS patients collected by National Institute of Neurological Disease and Stroke (NINDS) in the early 1990s.

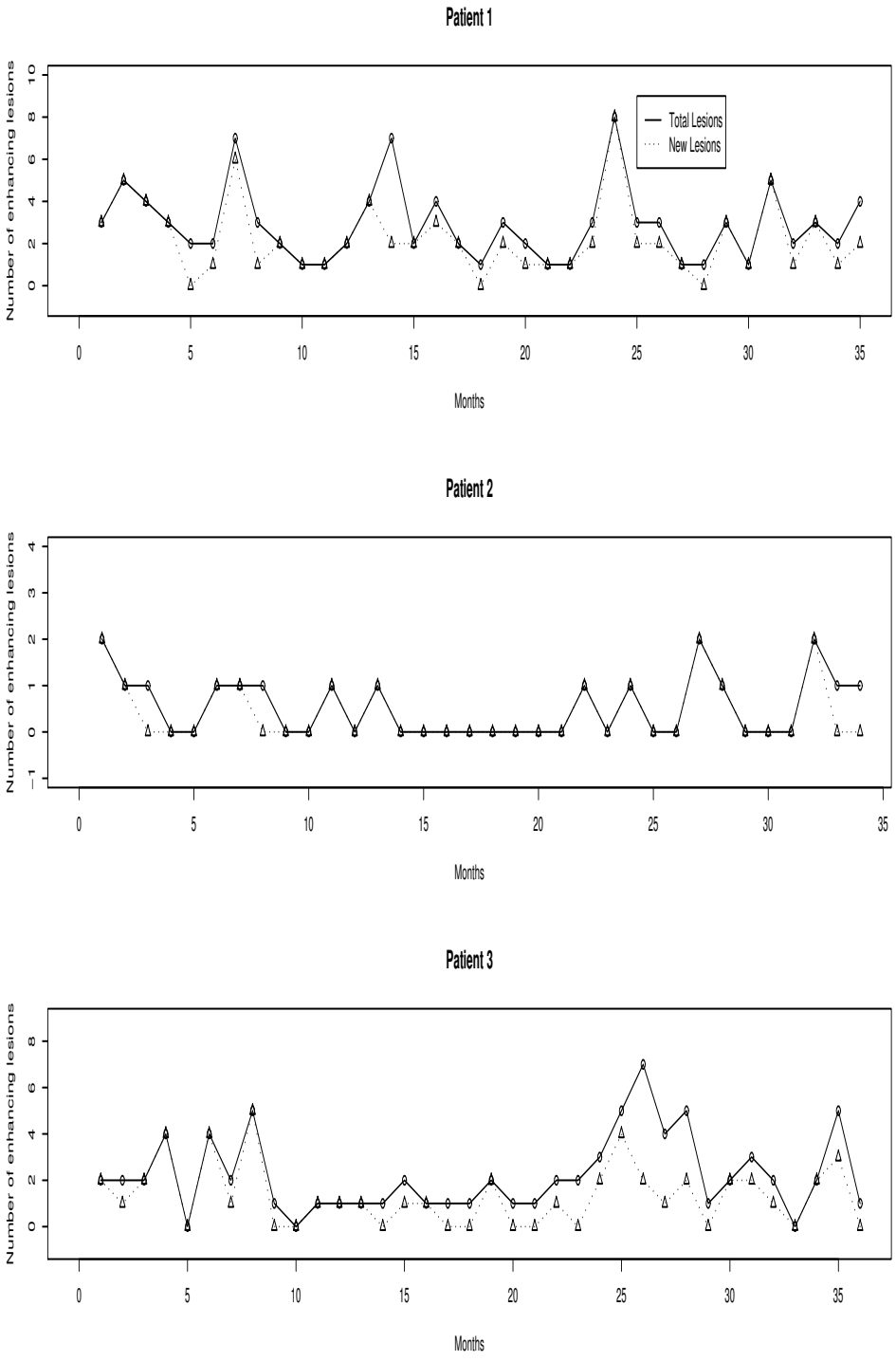


Figure 3.3: Monthly lesion counts for RRMS patients, Patient 1, 2, and 3.

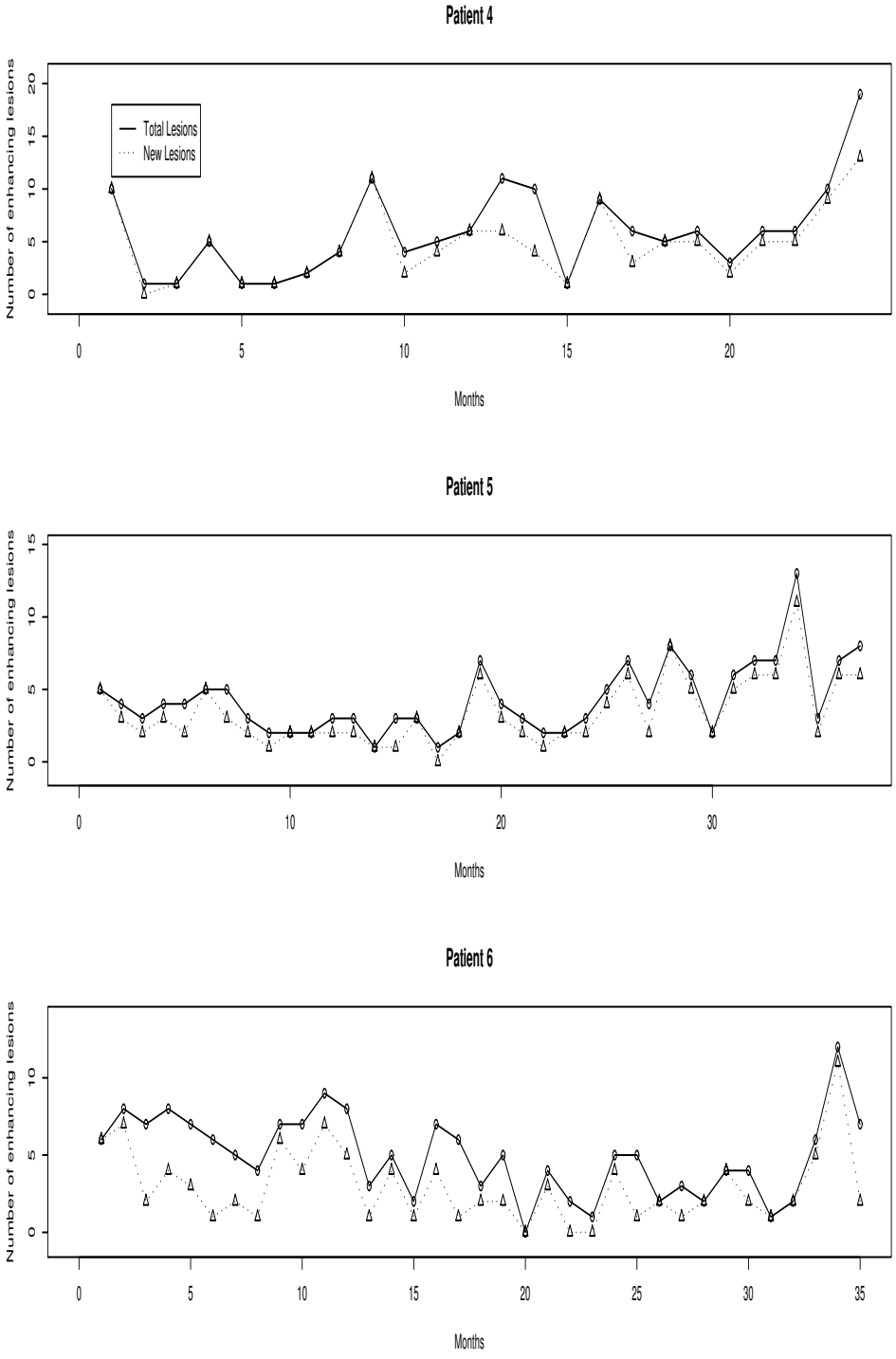


Figure 3.4: Monthly lesion counts for RRMS patients, Patient 4, 5, and 6.

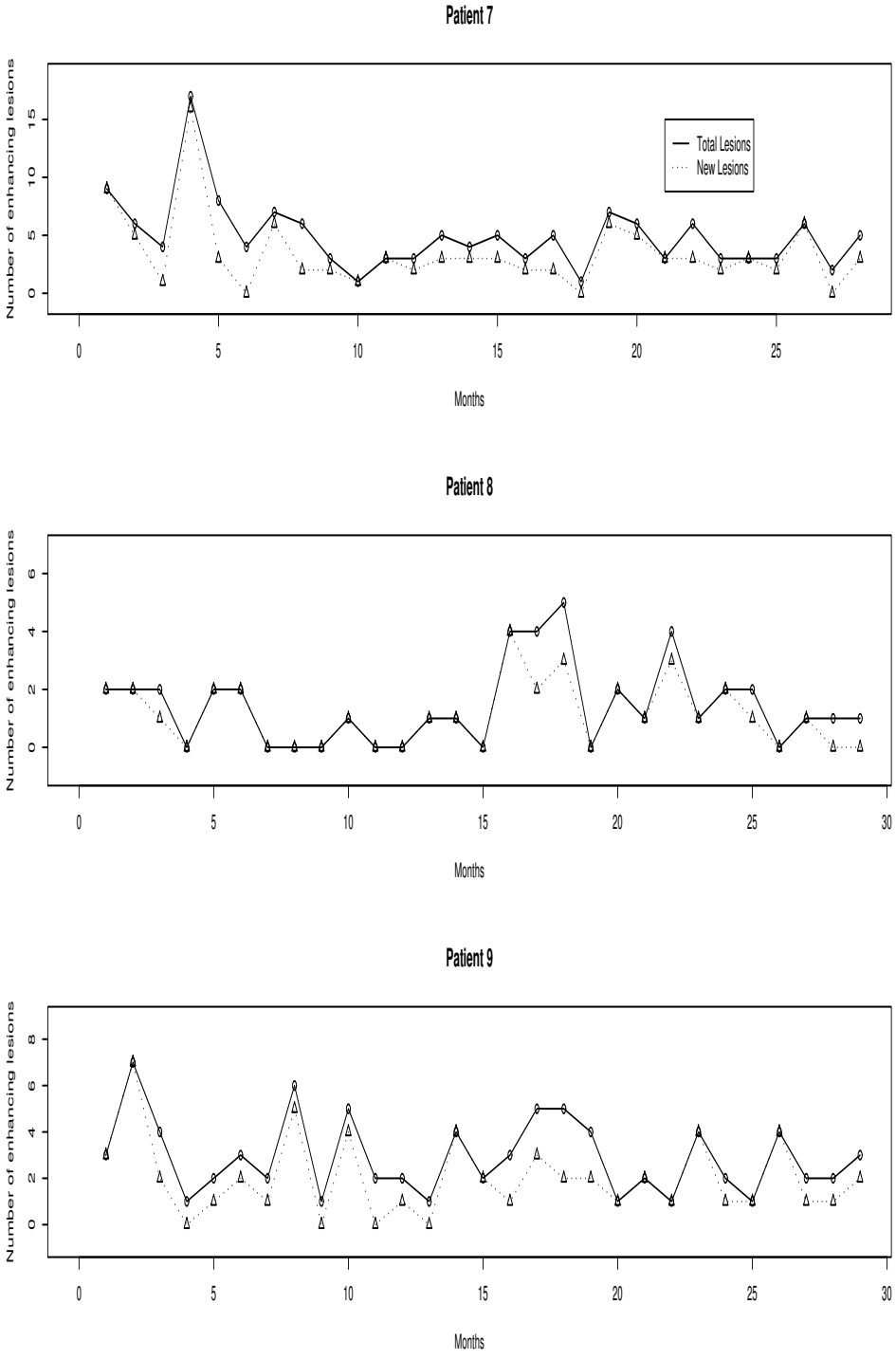


Figure 3.5: Monthly lesion counts for RRMS patients, Patient 7, 8, and 9.

These patients underwent the T1-weighted Gd-enhancing scans every month. Lesions were numbered sequentially, and new enhancing lesions were those that had not enhanced the previous month. Re-enhancing lesions were noticed from patients with long follow-up. Thus they were also considered as new lesions. Each monthly MRI scan except the first one reported the number of total enhancing lesions and new enhancing lesions. Accompanying the MRI lesion counts, the total volume of the Gd-enhancing lesions were measured. Monthly EDSS scores were given. The datasets are provided by Paul Albert (personal communication, 2005).

We present some summary measures of the lesion count data in Table 3.1. The table show that the lengths of the follow-up for these patients are not short. The average total enhancing lesion counts vary greatly across these patients. For some patients, the average is high (more than 4). For some patients, the average is intermediate (around 2 or 3). Patient 2 has a relatively low average (0.53). This patient has many zero counts both in the total lesion sequence and the new lesion count sequence. Patient 8 also has relatively more zero lesion counts. The maximum of the total lesion counts is smaller for these two patients when compared to others.

3.5 Model Fitting Results

For each of the MRI Gd-enhancing lesion count data from 9 RRMS patients, we fit both Model 1 and Model 2. For Model 1, we directly calculate the maximum likelihood estimates for the parameters using the formulas in Section 3.3.2. The standard errors are provided by taking the square root of the diagonal terms of the matrix $\hat{\Sigma}$. When fitting Model 2, since there is no analytical form available for the MLE, we use the 'optim' function in R as the procedure for the direct maximization of the loglikelihood

Patient	months studied	total counts 0 count	mean	range	new counts 0 count	mean	range	p^1
1	35	0	2.89	(1, 8)	3	2.24	(0, 8)	77.5%
2	34	19	0.53	(0, 2)	23	0.36	(0, 2)	67.9%
3	36	3	2.19	(0, 7)	12	1.31	(0, 5)	59.8%
4	24	0	5.96	(1, 19)	1	4.52	(0, 13)	75.8%
5	37	0	4.30	(1, 13)	1	3.39	(0, 11)	78.8%
6	35	1	4.94	(0, 12)	3	2.91	(0, 11)	58.9%
7	28	0	4.93	(1, 17)	3	3.22	(0, 16)	65.3%
8	29	9	1.41	(0, 5)	11	1.07	(0, 4)	75.9%
9	29	0	2.90	(1, 7)	4	1.96	(0, 7)	67.6%

p^1: percentage of the new to the total.

Table 3.1: Summary statistics of the Gd-enhancing lesion count sequences from 9 RRMS patients.

function. Multiple starting points are used to search for the MLE. The standard errors associated with the estimates are calculated from the inverted approximated Hessian matrix. We also give the negative loglikelihood function values, AIC (defined in (2.2.1)) and BIC (defined in (2.2.2)) for the purpose of model selection. In Table 3.2 and 3.3, we summarize the model fitting results.

For Model 1, the estimates for the arrival rates for the 9 patients, displayed in Table 3.2, are very different. This suggests the heterogeneity across the patients. From the estimates for the service rate μ, we can see that the mean enhancing duration time of the lesions is less than 1 month for most of the patients. Only Patient 3 and Patient 6 have the mean duration time more than 1 month. These estimates provide not only evidence to support the findings by McFarland et al. (1992), but also avoid the vague meaning of 'less than two months' they claimed.

Patient	$\hat{\lambda}$ ($\hat{se}(\hat{\lambda})$)	$\hat{\mu}$ ($\hat{se}(\hat{\mu})$)	-log L	AIC	BIC
1	4.293 (0.576)	1.483 (0.187)	102.710	209.42	212.531
2	0.745 (0.237)	1.358 (0.428)	36.653	77.306	80.359
3	2.007 (0.316)	0.925 (0.138)	92.688	193.376	192.543
4	8.784 (1.000)	1.422 (0.161)	100.353	204.706	207.062
5	6.697 (0.718)	1.547 (0.156)	115.740	235.48	238.702
6	4.435 (0.472)	0.886 (0.092)	129.749	263.498	266.609
7	5.608 (0.649)	1.122 (0.126)	109.897	223.794	226.458
8	2.091 (0.441)	1.477 (0.291)	56.628	117.256	119.991
9	3.292(0.488)	1.135 (0.160)	80.612	165.224	167.959

Table 3.2: The model fitting results: Model 1.

The estimates for μ in Model 2, given in Table 3.3, are close to those obtained in Model 1 and strikingly the standard errors for $\hat{\mu}$ are almost the same. This might be due to the fact that the hidden Markov chain affects only the arrival process. The constant λ in Model 1 is redistributed into two parts: λ_1 and λ_2 according to a symmetric two-state Markov chain. Thus this may not have much influence on the estimation of the service rate μ. We estimate the average service rate for this cohort of RRMS patients. The arithmetic mean of the estimates is 1.21 and the harmonic mean is 1.26. Since the difference is minor, we use the arithmetic mean to estimate the average service rate. Based on the exponential distribution with rate 1.21, about 70% of the lesions would finish the enhancement within one month, 8% would need more than two months. McFarland et al. (1992) reported that only 5% of the new lesions are seen enhancing in three consecutive monthly scans. Our estimate 8% is larger than that. These results do not support the findings in Lai et al. (1996), however, they were using a weekly study of the T1-weighted Gd-enhancing MRI scans from one RRMS patient.

Patient	$\hat{\lambda}_1$ ($\hat{se}(\hat{\lambda}_1)$) $\hat{\lambda}_2$ ($\hat{se}(\hat{\lambda}_2)$)	$\hat{\mu}$ ($\hat{se}(\hat{\mu})$) $\hat{a}$ ($\hat{se}(\hat{a})$)	-log L	AIC	BIC
1	5.579 (1.094) 3.033 (0.842)	1.481 (0.463) 0.297 (0.187)	102.175	212.35	218.571
2	0.745 (0.183) 0.745 (0.289)	1.358 (0.428) 0.266 (NA)	36.653	81.306	87.414
3	2.991 (0.649) 1.037 (0.462)	0.926 (0.139) 0.361 (0.253)	91.716	191.432	197.767
4	13.135 (1.948) 4.412 (1.262)	1.445 (0.161) 0.496 (0.155)	94.565	197.13	201.842
5	10.445 (1.533) 4.819 (0.717)	1.541 (0.156) 0.967 (0.034)	109.846	227.692	234.134
6	6.599 (0.965) 2.332 (0.637)	0.895 (0.093) 0.461 (0.271)	126.558	261.116	267.337
7	8.113 (1.223) 3.087 (0.777)	1.138 (0.126) 0.281 (0.262)	105.870	219.74	225.069
8	2.719 (0.684) 1.499 (0.473)	1.467 (0.291) 0 (NA)	55.921	119.842	125.311
9	4.659 (0.902) 1.964 (0.619)	1.126 (0.161) 0.146 (0.183)	79.043	166.086	171.555

Table 3.3: The model fitting results: Model 2.

For some of the patients, the negative loglikelihood has not decreased much by using the more complicated model. For example, the two λ parameters for Patient 2 are the same in Model 2. There is no gain in using Model 2 in such a case. Since the two arrival rates are the same, the transition probability can be any value. Thus the standard errors for $\hat{a}$ is not available (NA). When AIC is used as the model selection criteria, we can see that the fit for Model 2 is much better than that of Model 1 in Patient 3, 4, 5, 6 and 7. We also notice that the range for the new lesion counts and the total lesion counts is much wider in these patients than others (Table 3.1). This suggests that the $M/M/\infty$ with the hidden Markov regime is more appropriate taking account of the extra variability shown in the lesion counts. When BIC is used to compare these two models, we can see that only Patient 4, 5, and 7 stands out with a smaller BIC for Model 2 than Model 1. The criterion BIC tends to penalize on the number of parameters in the model as well as the length of the follow-up. Order estimation in HMM is an unsolved problem. Cappé et al. (2005) discuss about several methods of the Markov order estimation and seem to recommend BIC. The BIC estimator is consistent for any stationary irreducible Markov process under some restrictions. However, the maximum likelihood in formal HMM is not well-behaved compared to the Markov chain setting. Our situation is even more complicated than the formal HMM. Thus it is not clear if BIC is a good model selection criterion.

We also fit Model 3 (with unequal transition probabilities) individually to all the patients except for Patient 2 and Patient 8. These two patients are excluded here because it is unnecessary to fit the more complicated model when there is little gain using Model 2 than Model 1. The 'optim' function in R is used with multiple starting points searching for the MLE. The model fitting results are summarized in Table

3.4–3.6. The long-run probability for the patient being in a specific state can be derived using the estimates for the transition probabilities in the TPM. For example, for Patient 1, we can calculate the long run distribution for the hidden Markov chain. The chance that the patient would stay in the 'high' state and the 'low' state would respectively be:

$$\frac{1-0.868}{2-0-0.868}=0.117 \qquad\qquad \frac{1-0}{2-0-0.868}=0.883.$$

This means that given the patient's data, he/she would stay in the state that produces less new lesions longer than the other one. The data plot for Patient 1 in Figure 3.3 also supports this interpretation. Except for a few peaks, the number of total lesions and the number of new lesions are low most of the time.

There are a few zeros for the transition probabilities in the TPM in Table 3.4–3.6. This may suggest that the sequence is not long enough for us to accurately estimate the transition probabilities between the states. Since the estimate of 0 (or 1) is on the boundary of the parameter space, the inference therefore may not be informative. Thus we do not provide the standard errors of the estimates. Meanwhile, the loglikelihood, AIC and BIC scores do not improve much compared to Model 2 in most of the patients. Only Patient 7 shows a better fit using Model 3 compared to Model 2 and Model 1. From the lesion counts plot in Figure 3.5, we can see that he had a fairly stable period starting from Month 5. He has experienced 6.8 new lesions on the average for the first 5 months compared to an average of 2.7 for the following 23 months. This might be the reason that there is little chance to observe transitions from high to high in the underlying Markov chain. The estimate for a in Table 3.5 is 0.

Model 3	Patient 1	Patient 3	Patient 4
$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\mu})$	(9.764, 3.573, 1.479)	(4.003, 1.420, 0.922)	(12.449, 3.724, 1.442)
$\hat{\mathbf{P}}$	$\begin{pmatrix} 0 & 1 \\ 0.132 & 0.868 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 0.298 & 0.702 \end{pmatrix}$	$\begin{pmatrix} 0.590 & 0.410 \\ 0.575 & 0.425 \end{pmatrix}$
$-\log L$	101.354	91.563	94.542
AIC	212.708	193.126	199.084
BIC	220.484	201.044	204.974

Table 3.4: Model fitting results: Model 3, Patient 1, 3, and 4.

Model 3	Patient 5	Patient 6	Patient 7
$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\mu})$	(10.463, 4.826, 1.541)	(8.138, 3.135, 0.895)	(26.408, 4.831, 1.117)
$\hat{\mathbf{P}}$	$\begin{pmatrix} 0.958 & 0.042 \\ 0.030 & 0.970 \end{pmatrix}$	$\begin{pmatrix} 0.535 & 0.465 \\ 0.165 & 0.835 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 0.038 & 0.962 \end{pmatrix}$
$-\log L$	109.814	126.091	100.969
AIC	229.628	262.182	211.938
BIC	237.683	269.959	215.267

Table 3.5: Model fitting results: Model 3, Patient 5, 6, and 7.

Model 3	Patient 9
$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\mu})$	(5.645, 2.084, 1.122)
$\hat{\mathbf{P}}$	$\begin{pmatrix} 0 & 1 \\ 0.524 & 0.477 \end{pmatrix}$
$-\log L$	78.753
AIC	167.506
BIC	174.342

Table 3.6: Model fitting results: Model 3, Patient 9.

Since the RRMS patients usually present high variability in the lesion count sequences, longer follow-ups are needed to investigate if Model 3 would be better to describe the evolution process than Model 2.

So our conclusion, based on BIC, is that Model 1 (the $M/M/\infty$ model) is preferred for patients 1, 2, 3, 6, 8, and 9, and Model 2 (the $M/M/\infty$ model with the Markov regime and the equal transition probabilities between the two hidden states) is preferred in the other three cases.

CHAPTER 4

Asymptotic Normality of the MLE

In this chapter, we examine the asymptotic properties of the maximum likelihood estimators for Model 1 and Model 2. Although Model 1 is a special case of Model 2 (i.e., there is only one state for the underlying Markov chain), we want to study Model 1 separately for its special features.

4.1 Long Run Distribution of the Process $\{(Y_t^{(1)}, Y_t^{(2)}), t = 1, 2, \ldots\}$

In Model 1, the vector process $\{(Y_t^{(1)}, Y_t^{(2)}), t = 1, 2, \ldots\}$ has the Markovian property. The state space for the Markov chain is $\mathcal{G} = \mathbf{Z}^+ \times \mathbf{Z}^+$ where the notation $\mathbf{Z}^+$ is the set of all the nonnegative integers. This Markov chain is aperiodic and irreducible. It is aperiodic since each state has positive probability to return in one step transition. It is irreducible since each state needs at most two transitions to go to any other state. If $Y_k^{(1)} \leq Y_{k-1}$, then one transition is enough. Otherwise, we need some new lesions during the intermediate transition before we can have a larger number of total lesions.

When the queue $M/M/\infty$ starts with its equilibrium state, we can derive the analytical result for the long run distribution of the process $\{(Y_t^{(1)}, Y_t^{(2)}), t = 1, 2, \ldots\}$.

For any state $(l,k) \in \mathcal{G}$, we want to find $x_{l,k}$ to satisfy the following equation

$$x_{l,k} = \sum_{(j,i)\in\mathcal{G}} x_{j,i} P_{(j,i)(l,k)} \tag{4.1.1}$$

where

$$P_{(j,i)(l,k)} = P(Y_2^{(1)} = l, Y_2^{(2)} = k | Y_1^{(1)} = j, Y_1^{(2)} = i).$$

Take

$$x_{j,i} = \frac{e^{-\lambda_1}{\lambda_1}^i}{i!} \cdot \frac{e^{-\lambda_2}{\lambda_2}^j}{j!}$$

where

$$\lambda_1 = \frac{\lambda}{\mu}(1 - e^{-\mu}) \tag{4.1.2}$$

$$\lambda_2 = \frac{\lambda}{\mu} \cdot e^{-\mu}. \tag{4.1.3}$$

Consider the right hand side of (4.1.1),

$$\begin{aligned}
R.H.S. &= \sum_{(j,i)\in S} \frac{e^{-\lambda_1}{\lambda_1}^i}{i!} \cdot \frac{e^{-\lambda_2}{\lambda_2}^j}{j!} \cdot P(Y_2^{(1)} = l, Y_2^{(2)} = k | Y_1^{(1)} = j, Y_1^{(2)} = i) \\
&= \sum_{i+j\geq l} \frac{e^{-\lambda_1}{\lambda_1}^i}{i!} \cdot \frac{e^{-\lambda_2}{\lambda_2}^j}{j!} \cdot P(Y_2^{(2)} = k) \cdot \binom{i+j}{l} p^l (1-p)^{i+j-l} \\
&\quad \text{where } p = e^{-\mu} \\
&= P(Y_2^{(1)} = l) \sum_{t=l}^{\infty} e^{-(\lambda_1+\lambda_1)} \binom{t}{l} p^l (1-p)^{t-l} \cdot \sum_{i=0}^{t} \frac{\lambda_1^i \lambda_2^{t-i}}{i!(t-i)!} \\
&= P(Y_2^{(2)} = k) \sum_{t=l}^{\infty} e^{-(\lambda_1+\lambda_2)} \binom{t}{l} p^l (1-p)^{t-l} \cdot \frac{(\lambda_1+\lambda_2)^t}{t!} \\
&= P(Y_2^{(2)} = k) \sum_{t=l}^{\infty} e^{-(\lambda_1+\lambda_2)} \frac{1}{l!(t-l)!} p^l (1-p)^{t-l} \cdot (\lambda_1+\lambda_2)^t \\
&= P(Y_2^{(2)} = k) \cdot \frac{1}{l!}[(\lambda_1+\lambda_2)p]^l e^{-(\lambda_1+\lambda_2)} \cdot \sum_{t=l}^{\infty} \frac{1}{(t-l)!}[(1-p)(\lambda_1+\lambda_2)]^{t-l} \\
&= P(Y_2^{(1)} = k) \cdot \frac{1}{l!}[(\lambda_1+\lambda_2)p]^l e^{-(\lambda_1+\lambda_2)} \cdot \sum_{m=0}^{\infty} \frac{1}{m!}[(1-p)(\lambda_1+\lambda_2)]^m \\
&= P(Y_2^{(1)} = k) \cdot \frac{1}{l!}[(\lambda_1+\lambda_2)p]^l \cdot e^{-p(\lambda_1+\lambda_2)}.
\end{aligned}$$

Since

$$P(Y_2^{(2)} = k) = \frac{e^{-\lambda_1}{\lambda_1}^k}{k!}$$

and

$$\begin{aligned} p \cdot (\lambda_1 + \lambda_2) &= \frac{\lambda e^{-\mu}}{\mu} \\ &= \lambda_2, \end{aligned}$$

we show that the choice of $x_{j,i}$ satisfies (4.1.1). Since $x_{j,i}$ is the product of two Poisson probabilities, we have $\sum_{(j,i)\in\mathcal{G}} |x_{j,i}| < \infty$. Then by Theorem 3.1 in Karlin (1966, p. 132), the Markov chain $\{(Y_t^{(1)}, Y_t^{(2)}), t = 1, 2, \ldots\}$ is positive recurrent. Its long-run distribution is given by

$$P(Y_2^{(1)} = k, Y_2(2) = l) \;\; = \;\; \frac{e^{-\lambda_1}{\lambda_1}^k}{k!} \cdot \frac{e^{-\lambda_2}{\lambda_2}^l}{l!} \tag{4.1.4}$$

with λ_1 and λ_2 defined by equation (4.1.2) and (4.1.3). Further, this Markov chain is α-mixing because for a strictly stationary Markov chain with countable state space, the irreducibility and aperiodicity are equivalent to α-mixing (Robert and Rosenthal, 2004).

Properties of the Process $\{Y_k, k = 1, 2, \ldots\}$

If the queue $M/M/\infty$ starts without any server doing service at time 0, the total lesion count Y_t at time t is distributed as Poisson with mean $\frac{\lambda(1-e^{-\mu t})}{\mu}$. The process $\{Y_k, k = 1, 2, \ldots\}$ is an irreducible aperiodic Markov chain with $\mathbf{Z}^+$ as its countable state space. The transitional probability from $Y_{k-1} = i$ to $Y_k = j$ is:

$$P(Y_k = i | Y_{k-1} = j) = \sum_{l=0}^{min(i,j)} \binom{j}{l} p^l (1-p)^{(j-l)} \frac{\lambda_* (i-l) e^{-\lambda_*}}{(i-l)!} \tag{4.1.5}$$

with $p = e^{-\mu t}$ and $\lambda_* = \frac{\lambda(1-e^{-\mu t})}{\mu t}$. It is obvious that this Markov chain is irreducible and aperiodic. Also, for $n > 0$, suppose the queue starts from its equilibrium distribution, which is Poisson with mean $\frac{\lambda}{\mu}$. Then

$$cov(Y_k, Y_{k+n}) = \frac{\lambda}{\mu} e^{-\mu n}, \tag{4.1.6}$$

since

$$\begin{aligned}
cov(Y_k, Y_{k+n}) &= E(Y_k Y_{k+n}) - E(Y_k)E(Y_{k+n}) \\
&= E[E(Y_k Y_{k+n})|Y_k)] - \frac{\lambda^2}{\mu^2} \\
&= E[Y_k(Y_{k+n}^{(2)} + Y_k e^{-\mu})] - \frac{\lambda^2}{\mu^2} \\
&= E(Y_k)E(Y_{k+n}^{(2)}) + e^{-\mu}E((Y_k)^2) - \frac{\lambda^2}{\mu^2} \\
&= \frac{\lambda}{\mu}(1 - e^{-\mu})\frac{\lambda}{\mu} + (\frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2})e^{-\mu} - \frac{\lambda^2}{\mu^2} \\
&= \frac{\lambda}{\mu} e^{-\mu n}.
\end{aligned}$$

As n goes to infinity, the covariance goes to zero. The derivation of the marginal distribution and the covariance structure can be found in queueing theory literature. See, for example, Eick et al. (1993).

The total count sequence cannot be used alone for the inference although we can write down the likelihood function based on the Markovian structure. This is because λ and μ are not identifiable when using only the total lesion count data. A different set of λ and μ might give the same ratio and thus we are not able to tell the total lesion count comes from which set unless we have the information about the new lesion count.

4.2 Numerical Results: Model 1

In this section, we study the asymptotic distribution of the MLE of parameters λ and μ for Model 1 using simulation. The steps are:

1. For a fixed set of (λ, μ), using the likelihood function (3.3.5) for Model 1, we simulate 100 sequences, each with the same specified length

2. Fit Model 1 to each simulated sequence to get MLEs ($\hat{\lambda}$ and $\hat{\mu}$)

3. Draw the diagnostic plots to assess the asymptotic property.

Repeat the steps at various lengths of the sequences. For bivariate normality, we are looking for the elliptical shape in the scatter plot of $\hat{\lambda}$ versus $\hat{\mu}$. The *mshapiro.test* procedure in the package of *mvnormtest* in R has been used to test the bivariate normality. This package is the generalization of the Shapiro-Wilk test for multivariate variables (see *http://cran.r-project.org/doc/packages/mvnormtest.pdf* for reference). The quantile-quantile (Q-Q) plots are given to assess the marginal normality. The p-value is calculated from the standard Shapiro-Wilk test procedure in R. Histograms of $\hat{\lambda}$ and $\hat{\mu}$ are also provided. For the likelihood ratio test statistics,

$$\begin{aligned} D(\mathcal{L}) =& 2[(\log \mathcal{L}(\hat{\lambda}, \hat{\mu} \mid \text{simulated data})) - \\ & (\log \mathcal{L}(\lambda, \mu \mid \text{simulated data}))], \end{aligned}$$

the histogram overlaid with a density curve of χ^2 distribution with 2 degrees of freedom is created. The p-value is reported by using the Kolmogorov-Smirnov test. We select two sets of estimates as the fixed parameters in the simulation study.

Simulation Results for $(\lambda, \mu) = (4.293, 1.483)$

The first set comes from the estimates for Patient 1. The diagnostic plots are shown in Figure 4.1–4.8. The lengths we use for the study are 10, 35, 96 and 180. Included in these figures are the scatter plots, histograms, and Q-Q plots of the estimators and the likelihood ration test statistic.

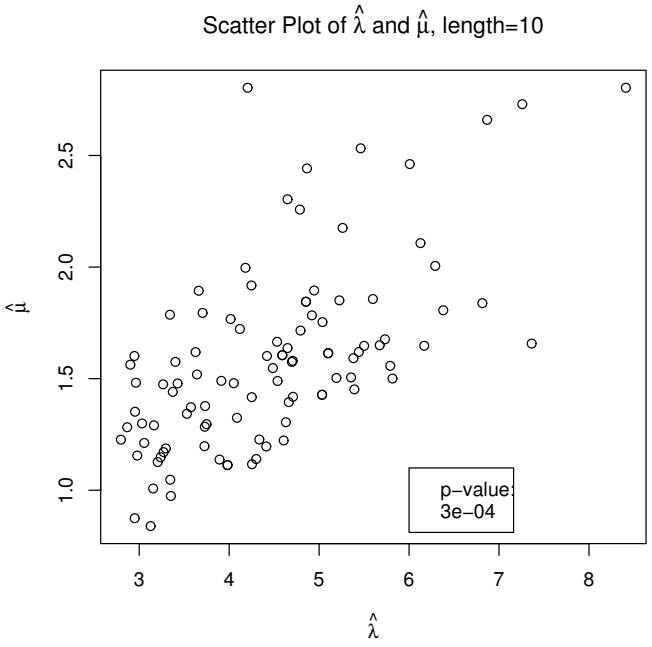
From these plots, we can see that $(\hat{\lambda}, \hat{\mu})$ appears to converge to bivariate normal as the length of the follow-up increases. The asymptotic χ^2 distribution with two degrees of freedom works well to approximate $D(\mathcal{L})$. The marginal normality of $\hat{\lambda}$ and $\hat{\mu}$ are both getting better as the sequence length increases. It takes longer for $\hat{\mu}$ to converge than for $\hat{\lambda}$.

The parameter set we have used is a well separated set of λ and μ (i.e., λ is much greater than μ). We have examined other well separated sets, the estimates from other patients except for Patient 2. The results agree with what we have found in the current case. The corresponding figures are not provided here.

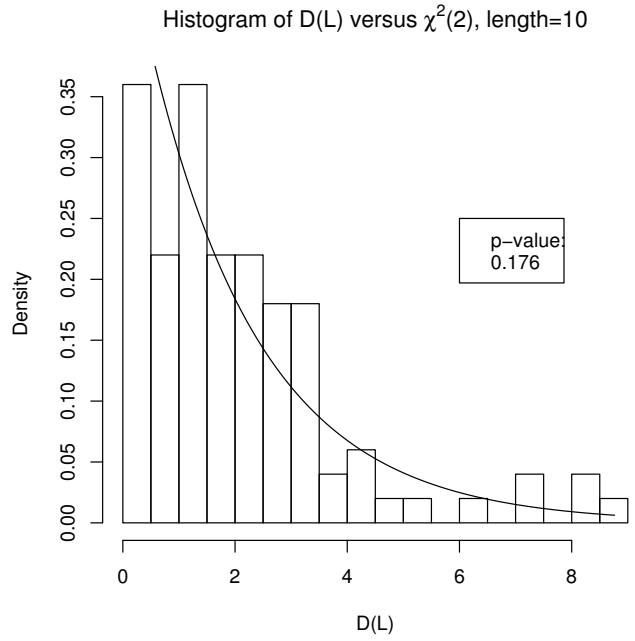
Simulation Results for $(\lambda, \mu) = (0.745, 1.358)$

The second set of fixed parameters are the estimates for Patient 2, $(0.745, 1.358)$. The diagnostic plots are given in Figure 4.9–4.16. The study lengths of the sequences are 10, 35, 96 and 180.

The estimates for Patient 2 draw our attention since they are quite different from the other estimates shown in Table 3.2. The arrival rate is fairly small compared to the service rate. The mean inter-arrival time $\frac{1}{0.745} = 1.342$ is longer than the mean service duration $\frac{1}{1.358} = 0.736$.



(a)



(b)

Figure 4.1: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=10, $(\lambda, \mu) = (4.293, 1.483)$.

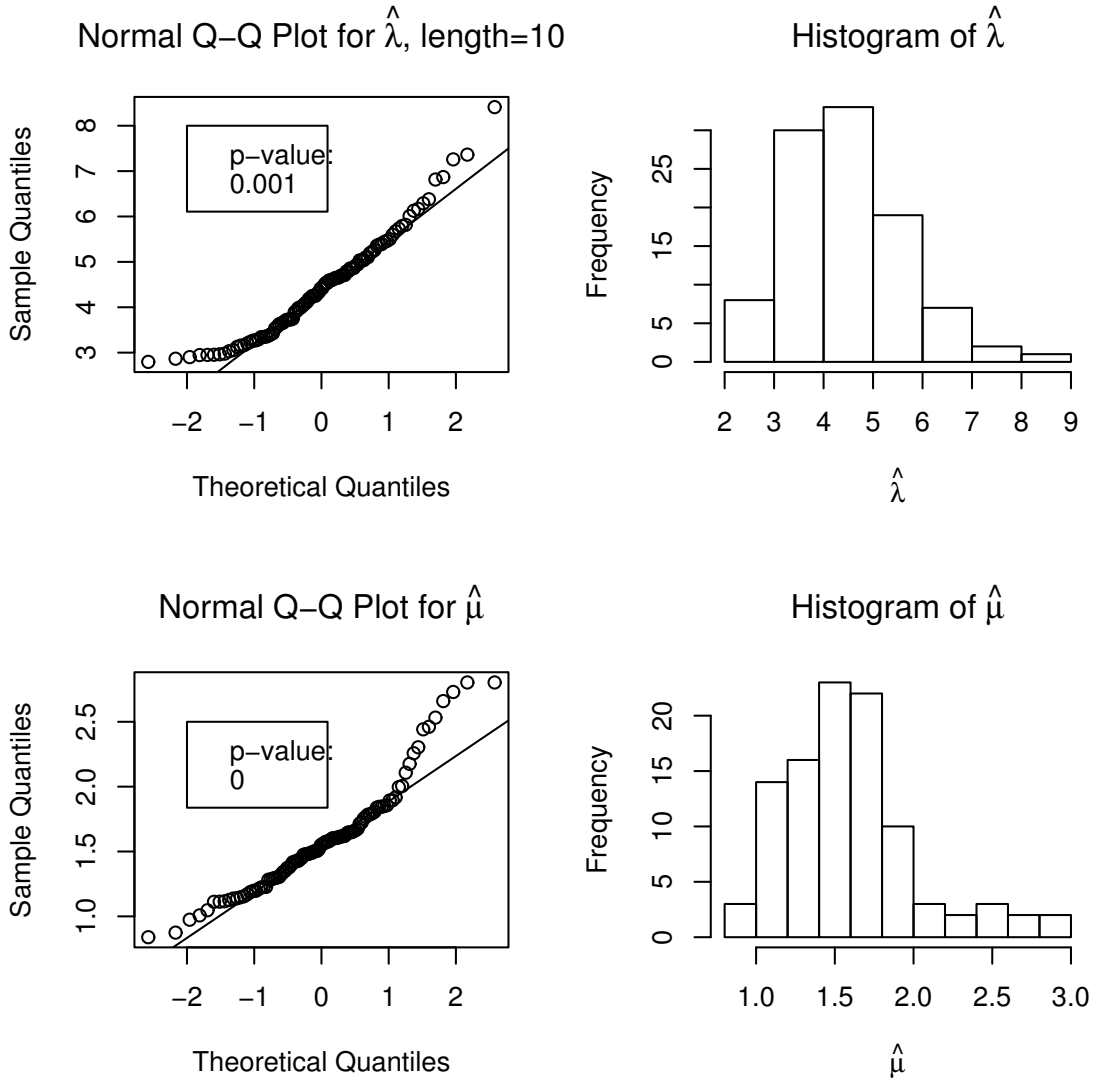
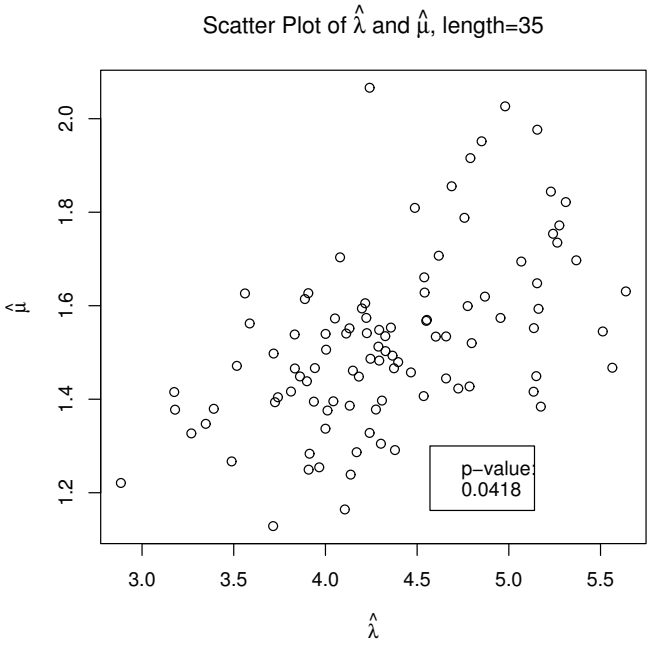
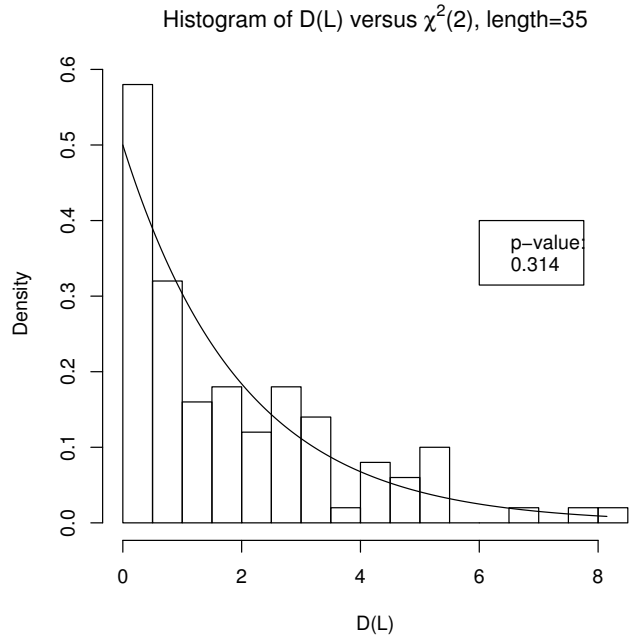


Figure 4.2: Normal Q-Q plots and histograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=10, $(\lambda, \mu) = (4.293, 1.483)$.



(a)



(b)

Figure 4.3: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=35, $(\lambda, \mu) = (4.293, 1.483)$.

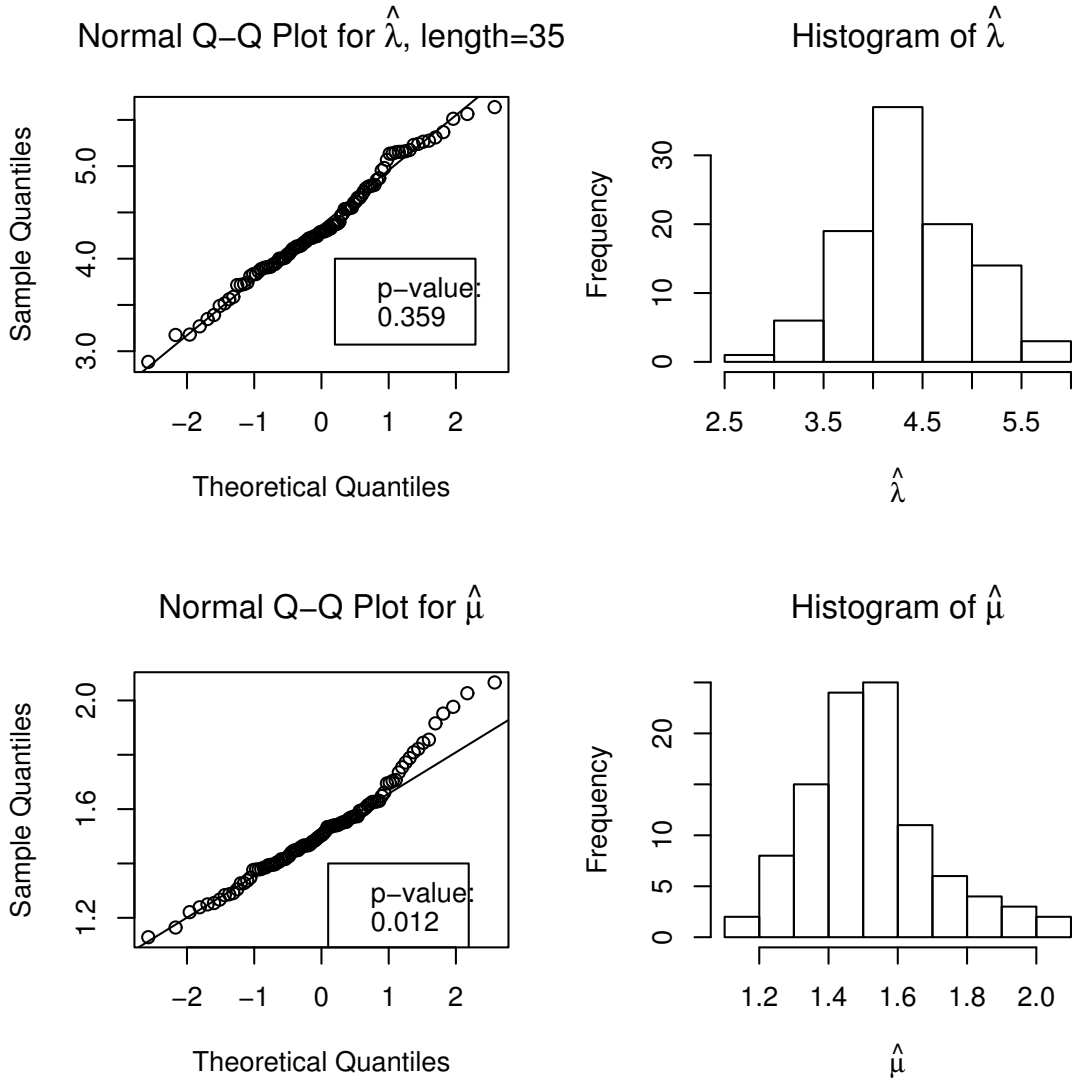
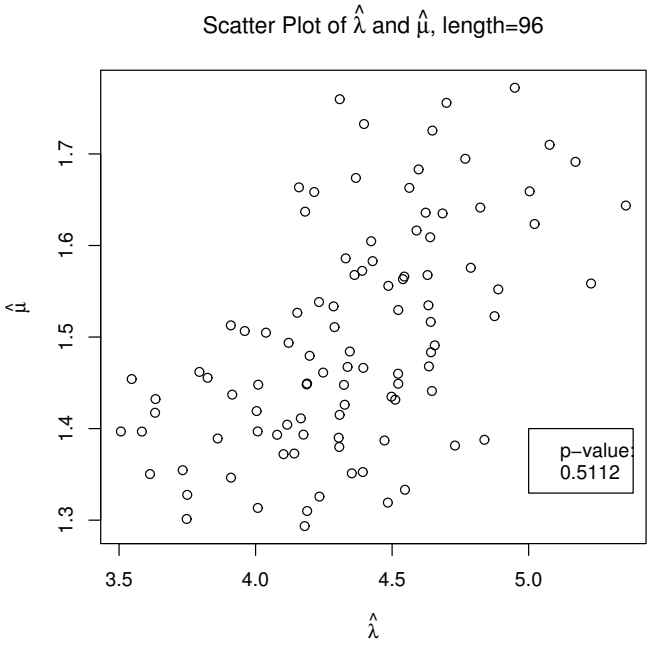
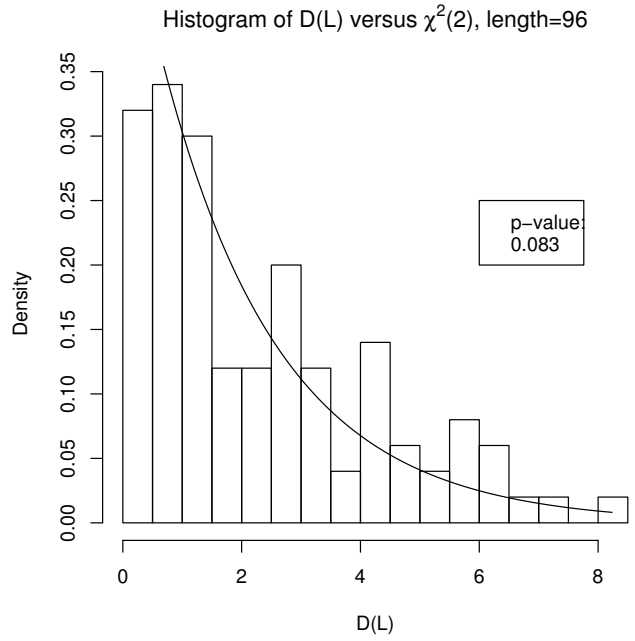


Figure 4.4: Normal Q-Q plots and histograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=35, $(\lambda, \mu) = (4.293, 1.483)$.



(a)



(b)

Figure 4.5: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=96, $(\lambda, \mu) = (4.293, 1.483)$.

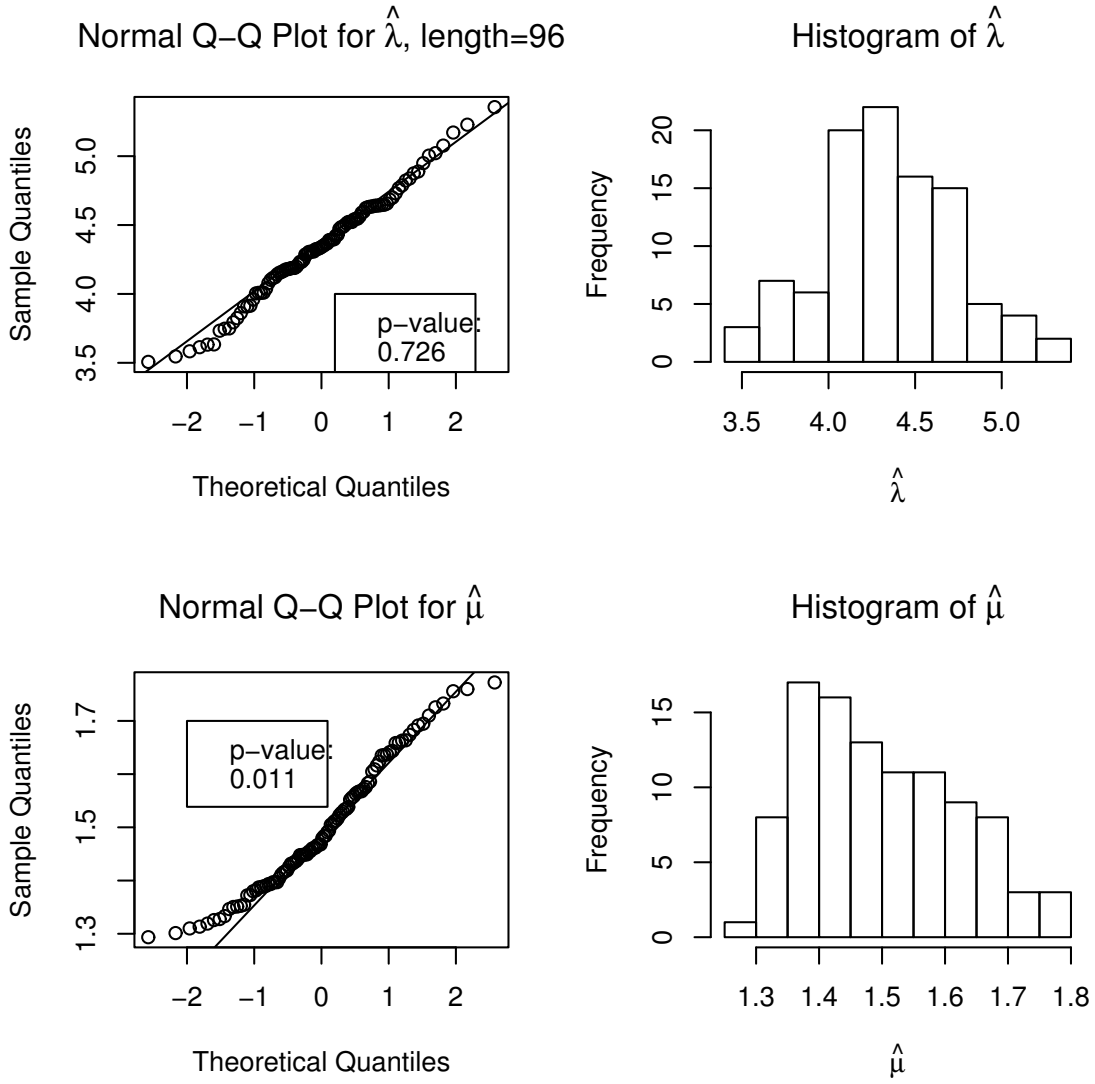
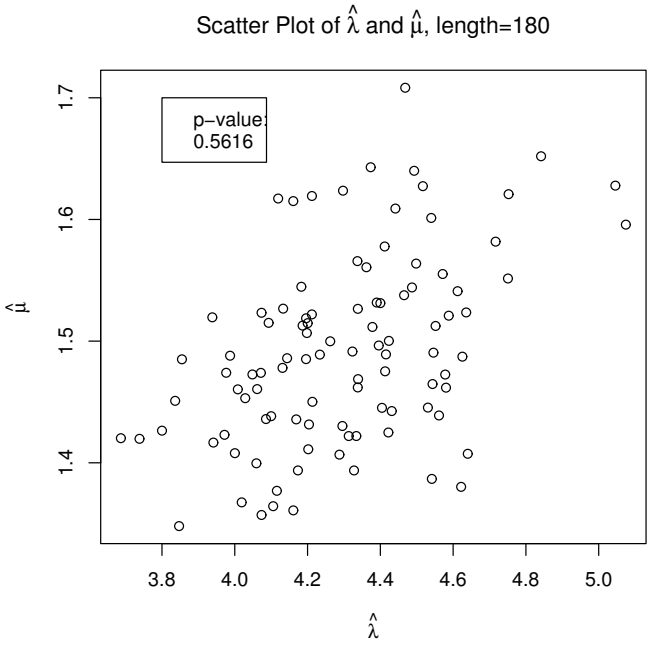
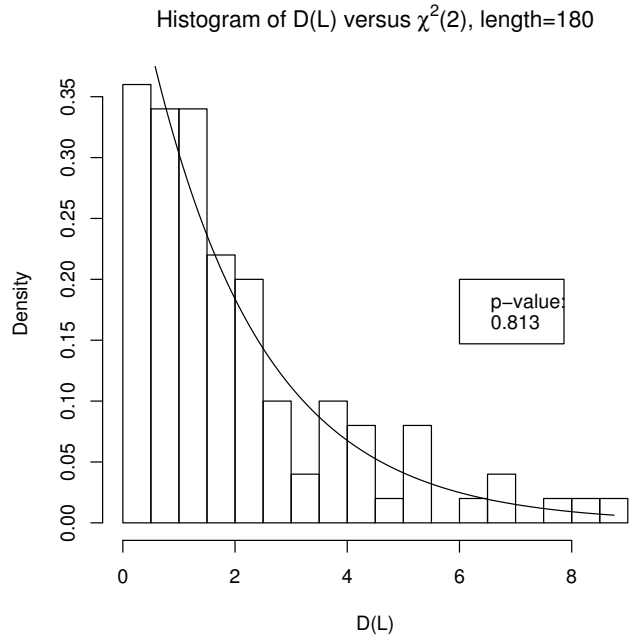


Figure 4.6: Normal Q-Q plots and histograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=96, $(\lambda, \mu) = (4.293, 1.483)$.



(a)



(b)

Figure 4.7: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=180, $(\lambda, \mu) = (4.293, 1.483)$.

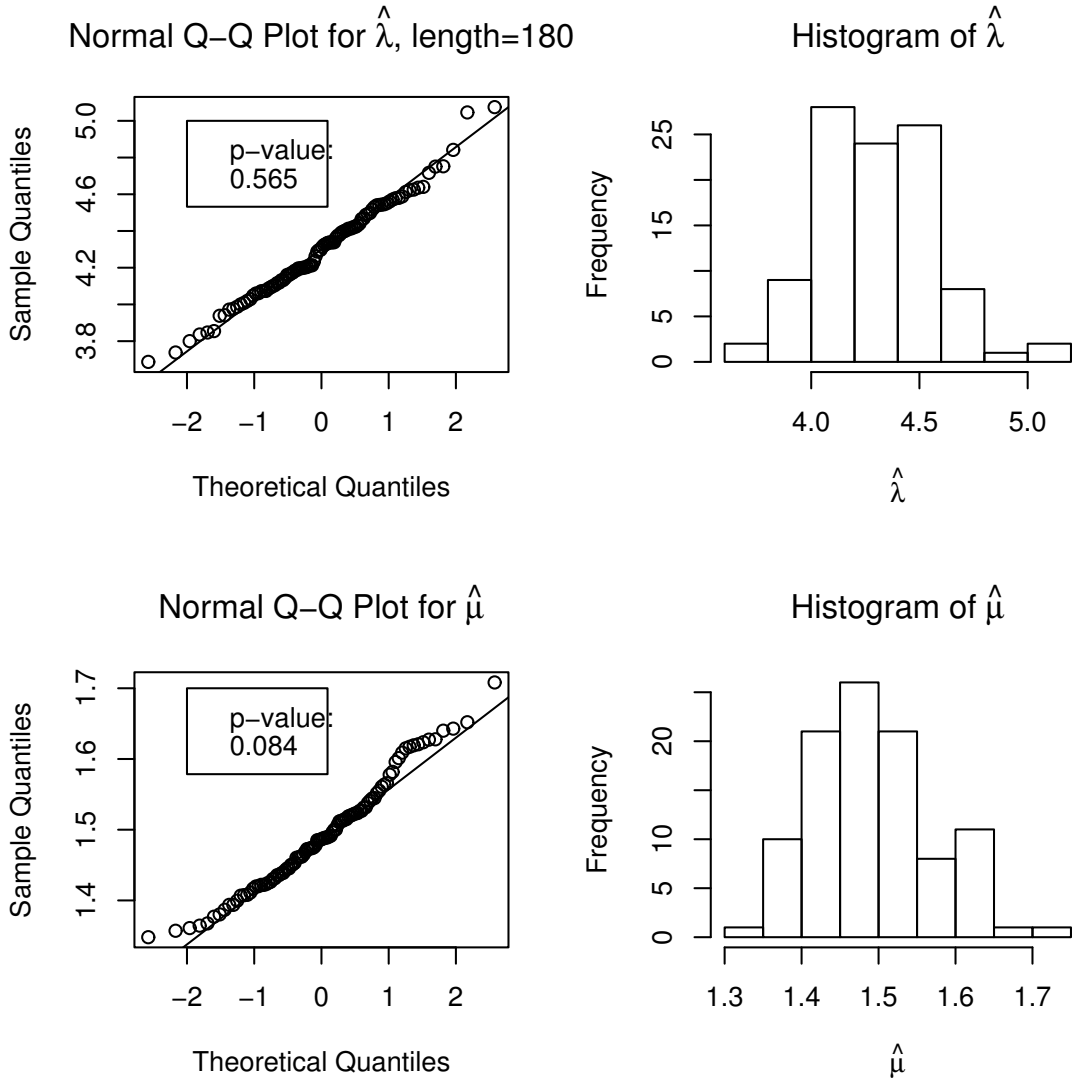
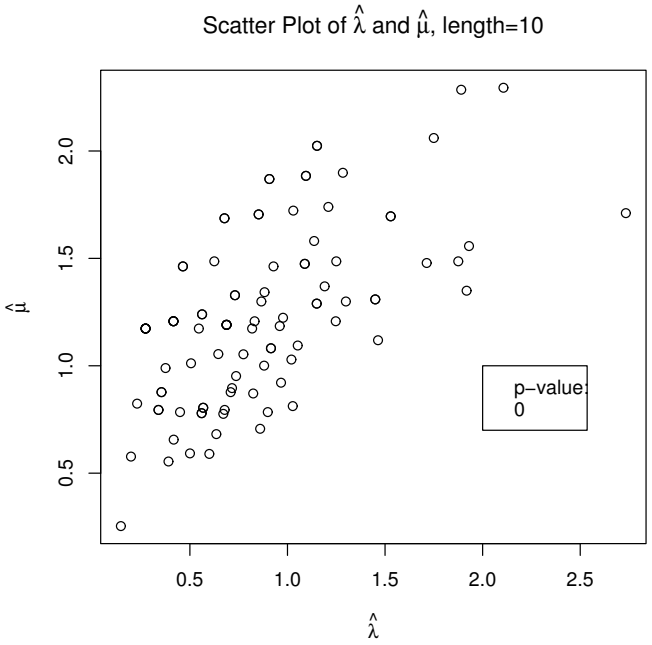


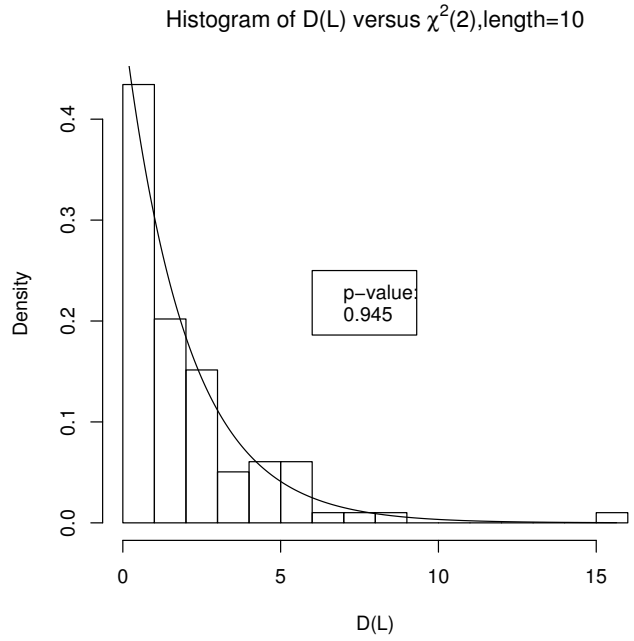
Figure 4.8: Normal Q-Q plots and histrograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=180, $(\lambda, \mu) = (4.293, 1.483)$.

For Patient 2 the lesions arrive slowly and service can be done in a short time. Henceforth, we would not observe many total lesions as well as the new from the monthly MRI scans. Out of 34 scans, we have 19 zeros for the total counts. For the number of new lesions, the situation is worse: 23 zeros out of 33 measurements. The maximum total number we observe in this patient is only 2. Such a big proportion of zeros may slow the convergence to normal distribution. Although the sequence for Patient 2 is as long as the one for Patient 1, the convergence to bivariate normality takes much longer. When the sequence length is 10, the estimates seem to concentrate on part of the elliptical shape. We also find out that there is quite a number of the simulated 'old' lesion count sequences having all zeros in it when sequence length is 10. First of all, the total lesion count would not be a big number because of the slow arrival rate. Secondly, the lesions are gone fast since the service rate is relatively high. Thus we hardly observe any old lesions left persistently enhancing.

The estimation encounters problems when the sum of the old lesion count is 0 as we have discussed in Section 3.3.2. The estimate for μ would be ∞ and the estimate for λ would be 0. This would bias the estimation. However, when we make the plots for a sequence length of 10, we ignore these boundary values of the parameter space. Thus the bias does not show up in the plots. It tells us that when the sequence length is short in this case, the estimation is not reliable. As the sequence length increases, the chance we have such a problem is smaller. The convergence for $\hat{\lambda}$ and $\hat{\mu}$ respectively are much slower compared to what we have observed from Patient 1. When the sequence length is 180, the marginal normality and the bivariate normality work much better than other shorter lengths. But one needs 15 years to collect such data!



(a)



(b)

Figure 4.9: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=10, $(\lambda, \mu) = (0.745, 1.358)$.

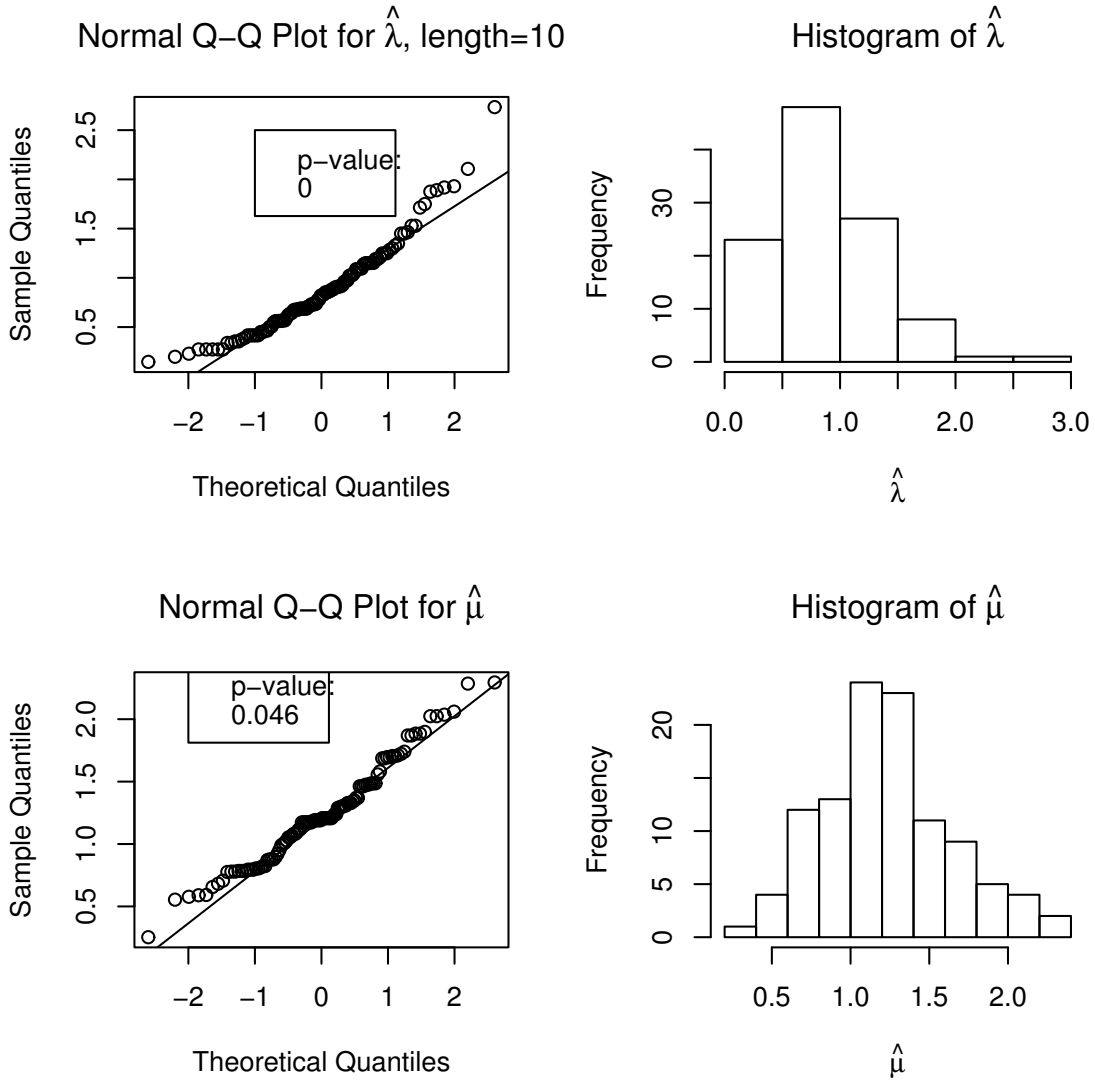
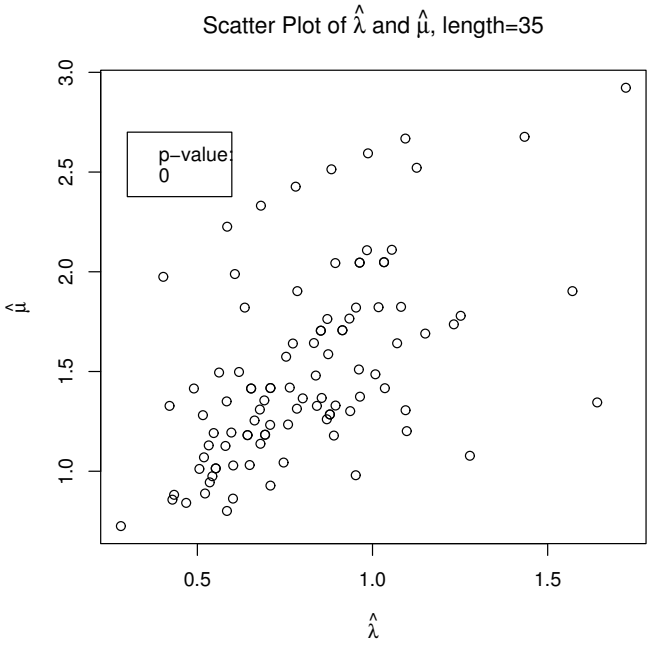
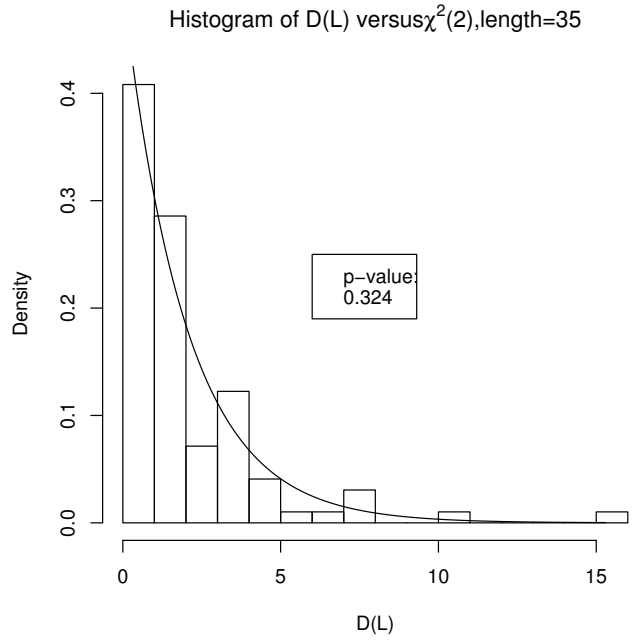


Figure 4.10: Normal Q-Q plots and histograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=10, (λ, μ) =$(0.745, 1.358)$.



(a)



(b)

Figure 4.11: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=35, $(\lambda, \mu) = (0.745, 1.358)$.

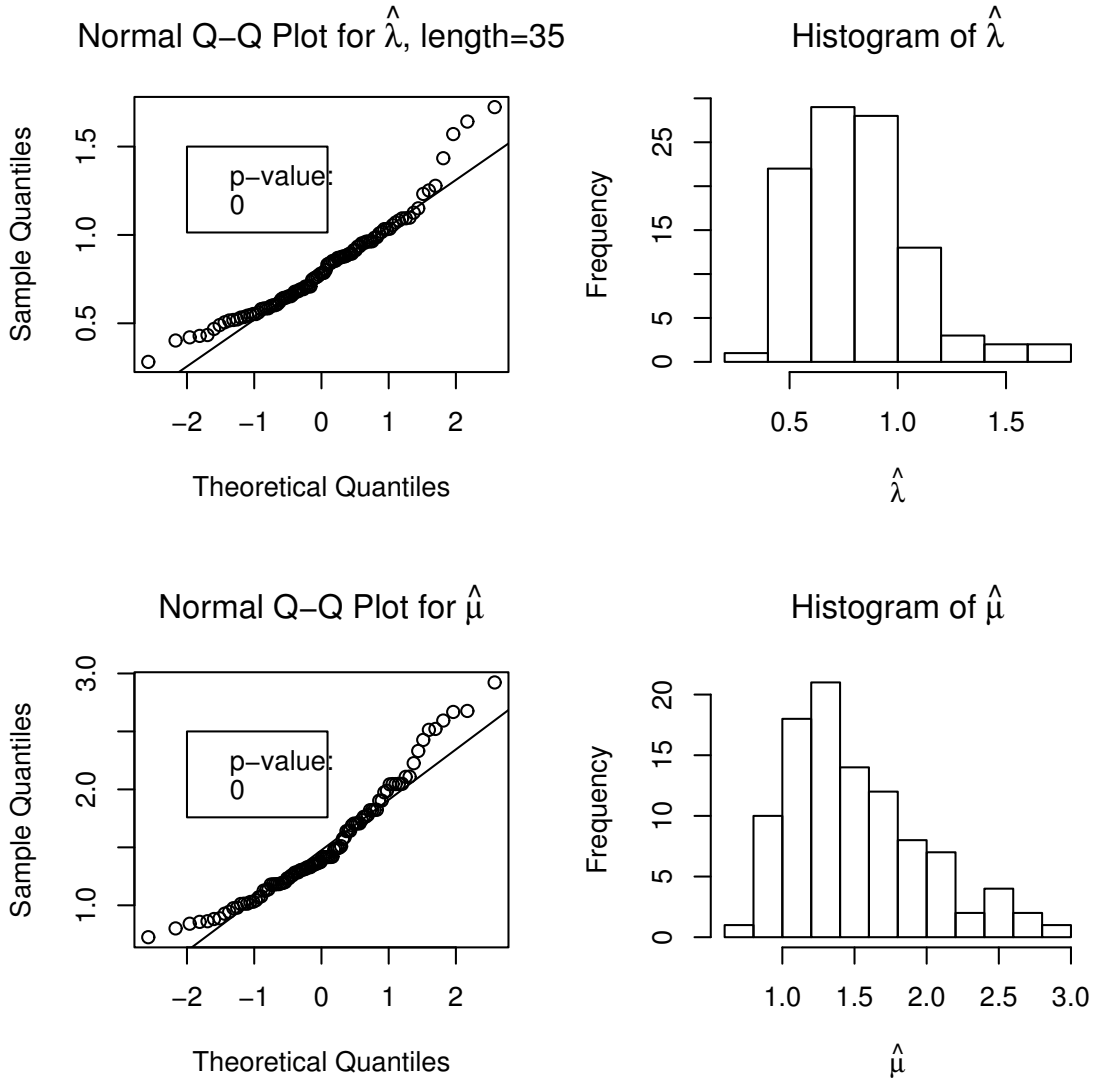
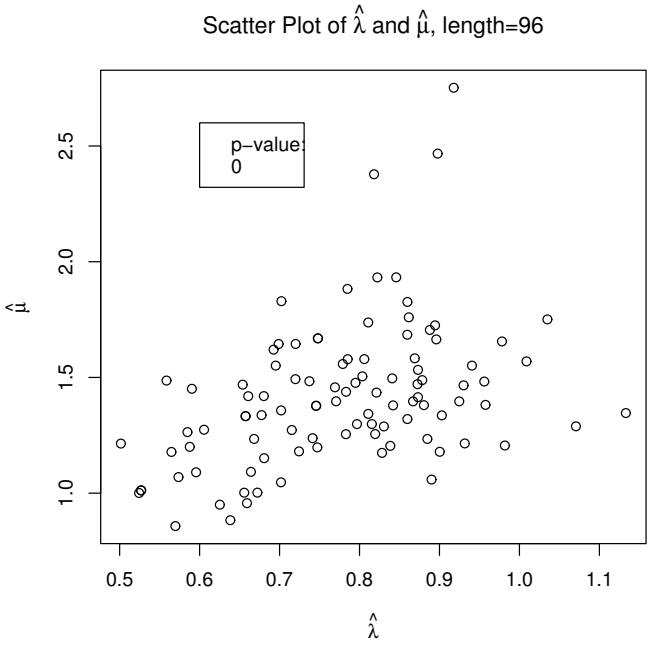
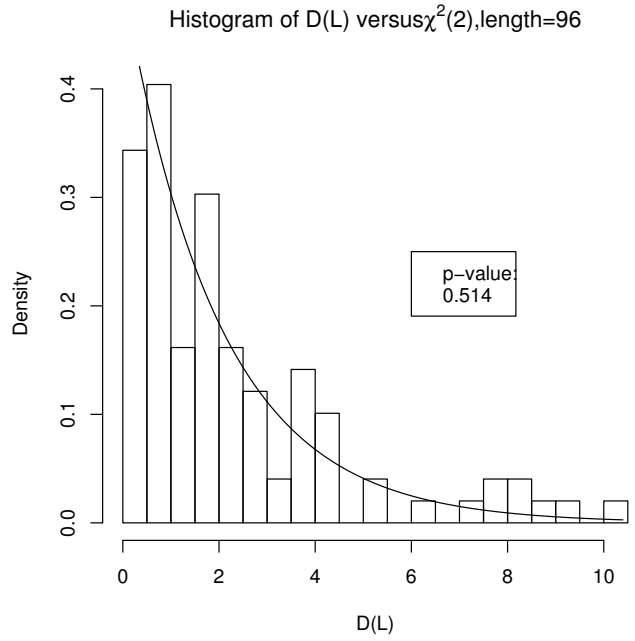


Figure 4.12: Normal Q-Q plots and histograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=35, (λ, μ) =$(0.745, 1.358)$.



(a)



(b)

Figure 4.13: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=96, $(\lambda, \mu) = (0.745, 1.358)$.

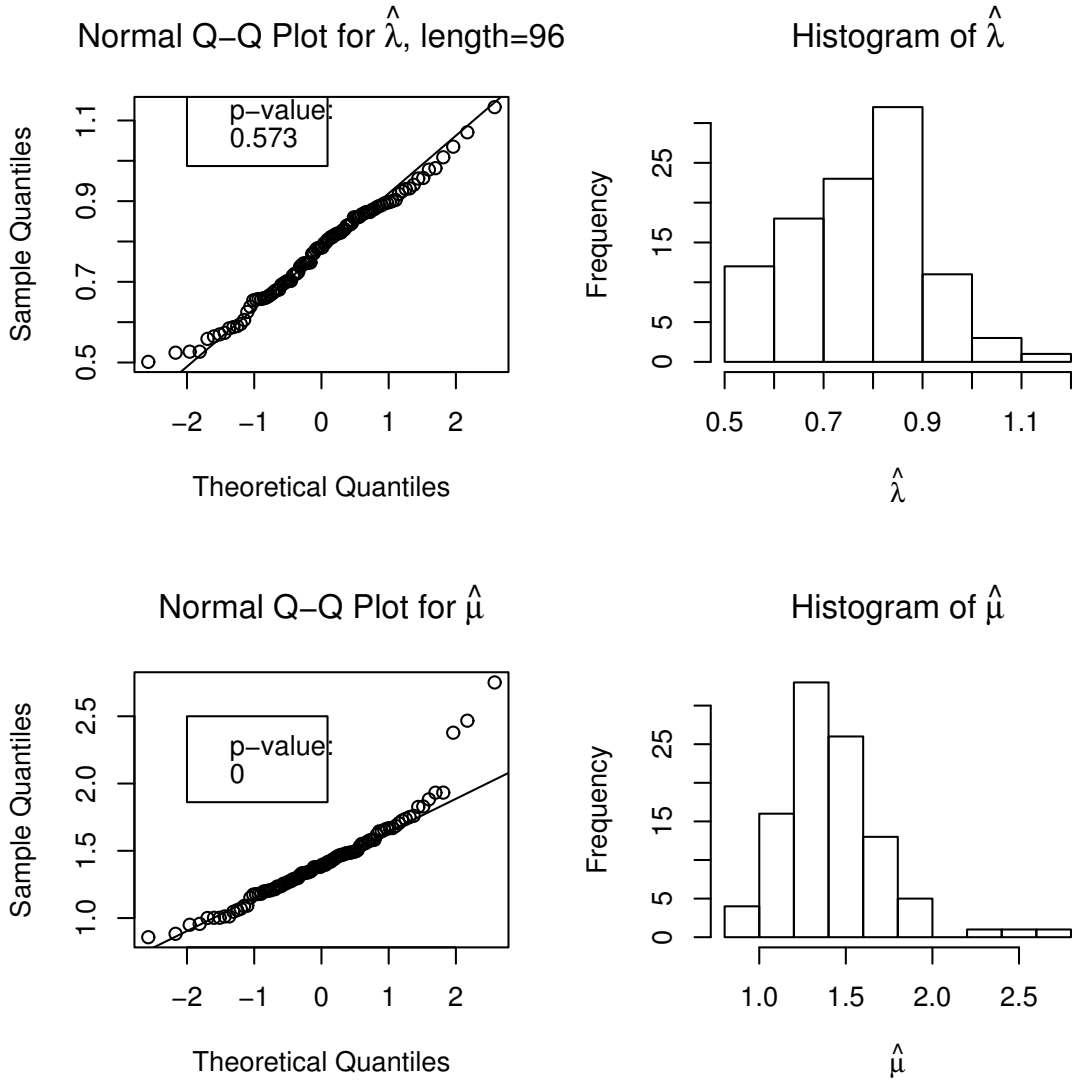
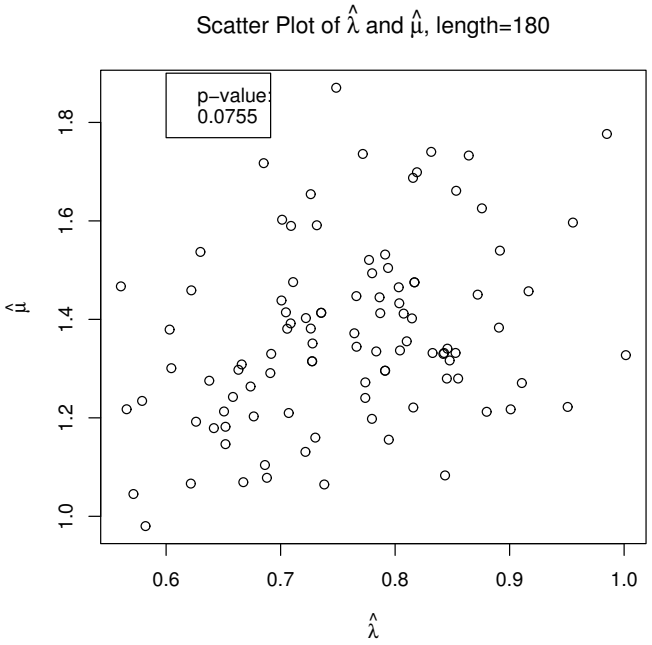
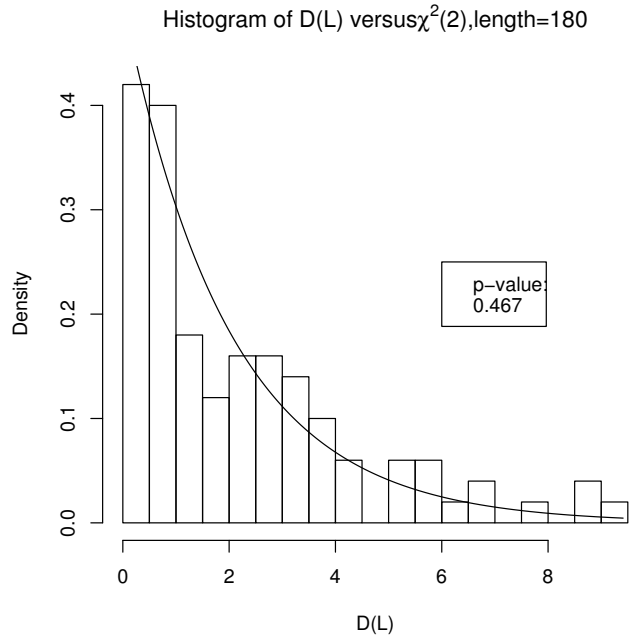


Figure 4.14: Normal Q-Q plots and histograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=96, (λ, μ) =$(0.745, 1.358)$.



(a)



(b)

Figure 4.15: (a) Scatter plot of $\hat{\lambda}$ vs $\hat{\mu}$. (b) Histogram of $D(\mathcal{L})$; length=180, $(\lambda, \mu) = (0.745, 1.358)$.

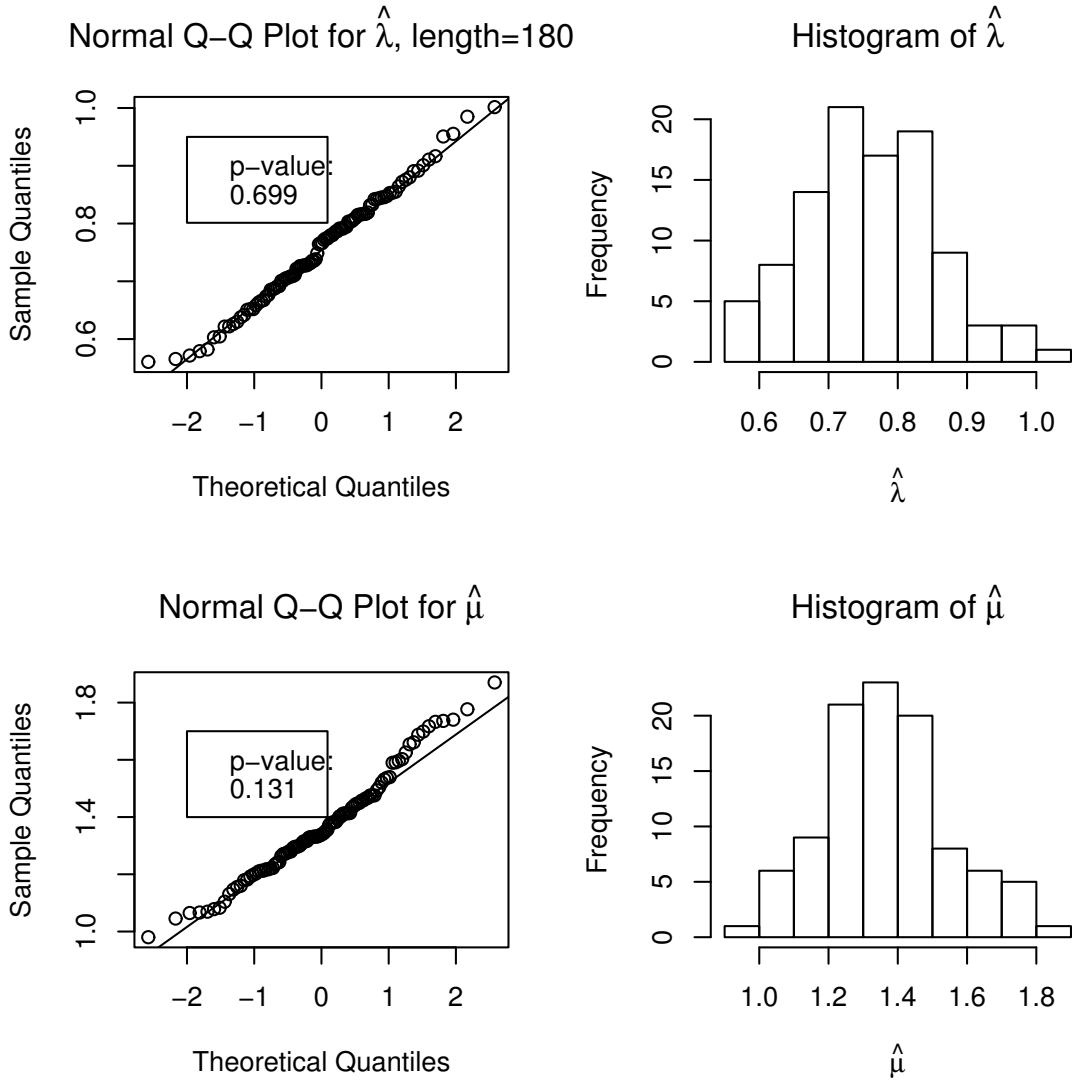


Figure 4.16: Normal Q-Q plots and histograms for estimators $\hat{\lambda}$ and $\hat{\mu}$; length=180, $(\lambda, \mu) = (0.745, 1.358)$.

From the above two simulation studies, we can see that in the first case, $\hat{\lambda}$ converges to normal distribution faster than $\hat{\mu}$. When the length is not long enough to achieve normality, log transformation can be done on the estimators. For example, for the first case at length 10, $\hat{\mu}$ cannot be approximated by normal well. The log transformation turns out to work better (Figure 4.17).

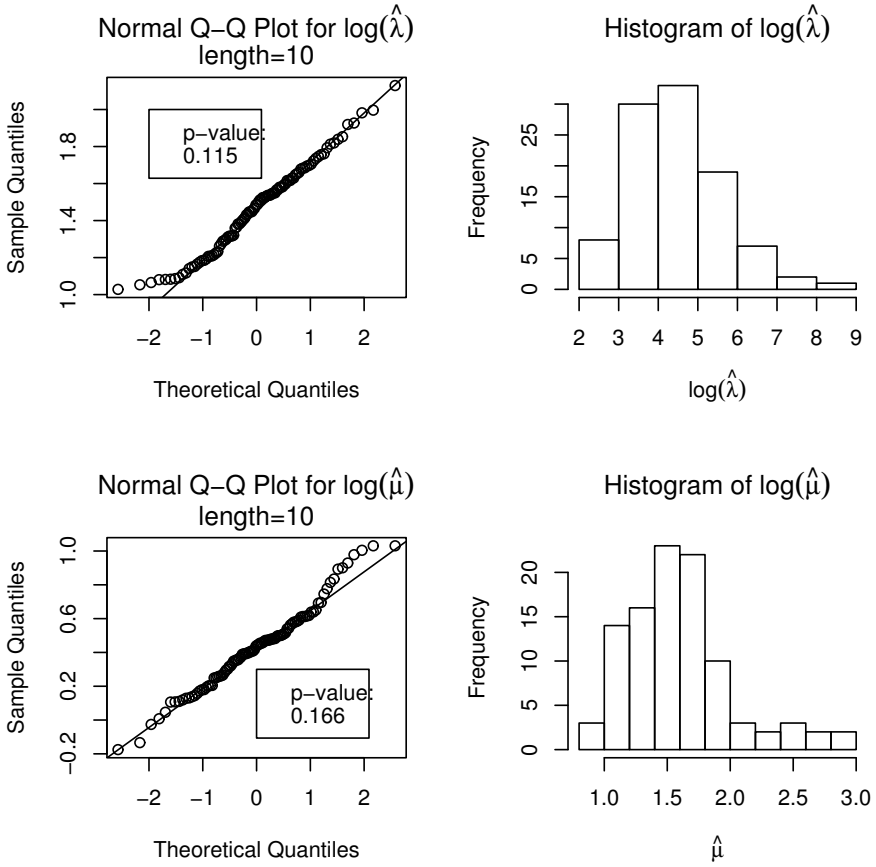


Figure 4.17: Normal Q-Q plots and the histograms for $\log \hat{\lambda}$ and $\log \hat{\mu}$; length=10, $(\lambda, \mu) = (4.293, 1.483)$.

For some of the RRMS patients, the disease course could be as long as 10-20 years before it gradually evolute to a worse situation: SPMS. However, as sequence gets longer, the stationary assumption for the process may not hold since the RRMS patient is more likely to develop into SPMS. Also, the patient may not feel comfortable with frequent MRI scans. The cost of the study will increase as well. Therefore, if we want to do analysis on the shorter sequence, the log transformation on the MLE may work well.

4.3 The Process in Model 2

In Model 2, there is a hidden Markov chain which governs the rate of the Poisson distribution for new lesions. However, as shown in Figure 3.2, this model is not a standard HMM. The observed vector process $(Y_t^{(1)}, Y_t^{(2)})$ has a Markovian structure because the observed number of new lesions $Y_t^{(1)}$ depends on the observed previous total Y_{t-1}. Generally, the HMMs have conditional independence between the observed data given the hidden state. Model 2 seems similar to the so called autoregressive model on the Markov regime (Douc et al., 2004). In contrast to the abundant theoretical results for general state space hidden Markov models, much less is known about the asymptotic properties of the MLE in autoregressive models with Markov regime. Douc et al. (2004) develop a very rigorous proof of the consistency and asymptotic normality of the MLE in such models. They try to show the geometrically decaying bound on the mixing rate of the conditional chain, which is the hidden Markov chain given the observations. Many assumptions are needed for the properties to hold. There is one particular assumption that is of concern for us. They state that the hidden Markov chain should be 1-small (i.e., for a chain with finite state space, all

the entries in the TPM should be bounded away from zero). Thus, when we fit Model 3 to the lesion data, we do not calculate the standard errors for those zero estimates.

4.4 Numerical Results: Model 2

The situation is complicated when we examine the asymptotic normality for the MLE of the parameters in Model 2. The dimension of the parameter space is much higher. The multivariate normality should be checked for the MLE. At least the marginal normal approximation is desired to hold for longer sequences. Simulation is run according to the following scheme:

- For a fixed set of $(\lambda_1, \lambda_2, \mu, a)$, simulate 200 sequences following the way we establish the likelihood function for Model 2.
- Fit Model 2 to each single sequence to get maximum likelihood estimates, $(\hat{\lambda_1}, \hat{\lambda_2}, \hat{a}, \hat{\mu})$.
- Draw the diagnostic plots and carry out tests to assess the asymptotic property.
- Repeat the steps at various lengths of the sequences.

The *mshapiro.test* procedure from the package *mvnormtest* in R has been used to test the multivariate normality. The quantile-quantile (Q-Q) plots are given to assess the marginal normality. P-value is calculated from Shapiro-Wilk test. Histograms of all the estimates are also provided. For the likelihood ratio test statistics $D(\mathcal{L})$, the histogram overlaid with a density curve of χ^2 distribution with 4 degrees of freedom is created. The p-value is reported by using the Kolmogorov-Smirnov test.

The parameter set we use is $(\lambda_1, \lambda_2, a, \mu) = (5.58, 3.03, 0.3, 1.48)$. This corresponds to Patient 1. The sequence lengths used in this study are 10, 35, 96 and 180. The plots are given in Figure 4.18–4.22.

Simulation Results for $(\lambda_1, \lambda_2, a, \mu) = (5.58, 3.03, 0.3, 1.48)$

The p-values from the multivariate normality test procedure are all less than 0.001 for all these cases. From the plots, we find out that the estimates for arrival rates (λ_1, λ_2) and the departure rate μ seem to achieve normality much faster than the one for the transition probability a in the underlying Markov chain. The shape in the Q-Q plot for $\hat{a}$ is highly right skewed even when the sequence length is 35 (Figure 4.18. However, as the sequence becomes longer, the shape gets back to a more 'familiar' pattern compared to the other three components. The histogram of $D(\mathcal{L})$ shows some departure from the χ^2 distribution with 4 degrees of freedom when sequence is 10 or 35. This may imply that the distribution of the LRT statistic is not well approximated by a χ^2 distribution for shorter sequences. Also, compared to the simulation results for Model 1, the length of the sequence required to achieve a moderate asymptotic normality is higher in the multivariate situation. This may be due to the existence of the additional hidden Markov chain. In application the increased number of parameters, the addition of the hidden Markov model appear to slow down the rate of convergence.

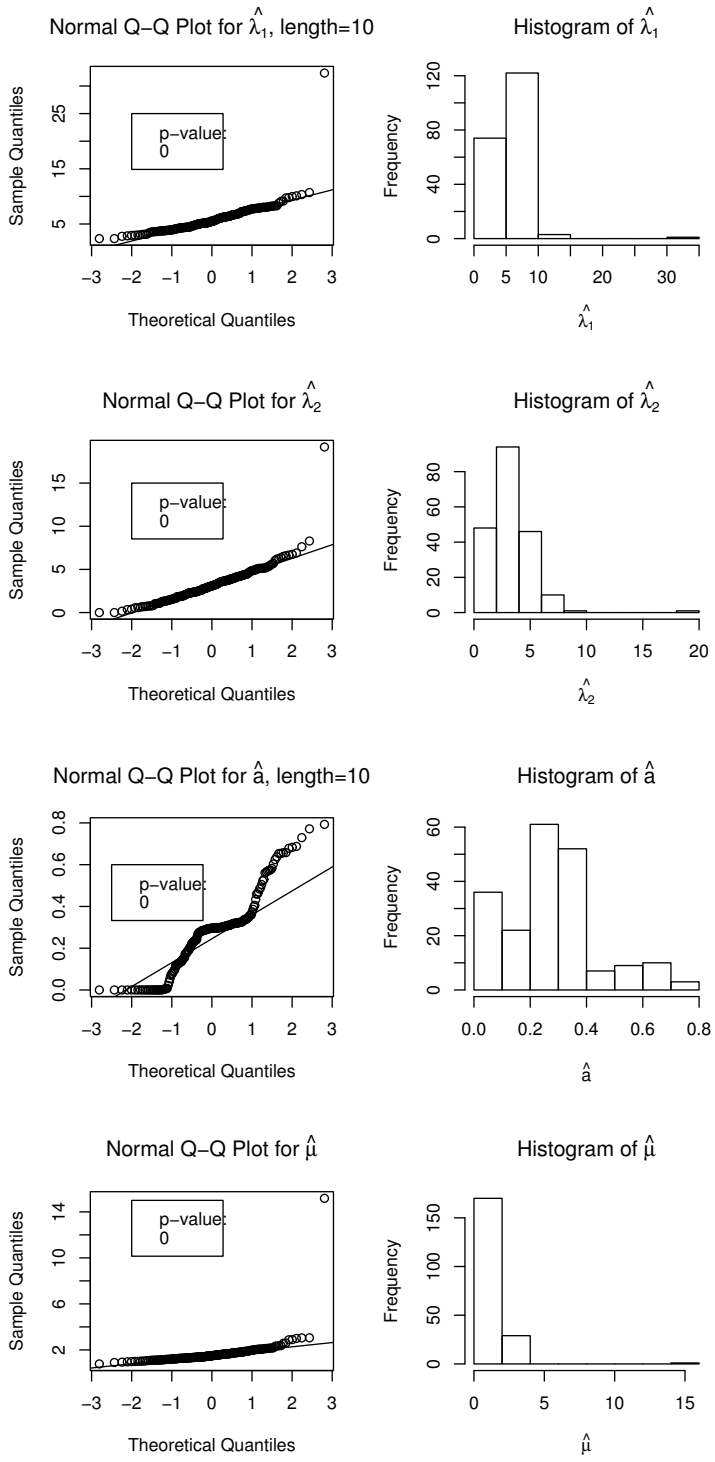


Figure 4.18: Normal Q-Q plots and histograms for estimators $\hat{\lambda}_1, \hat{\lambda}_2, \hat{a}$, and $\hat{\mu}$; length=10, $(\lambda_1, \lambda_2, a, \mu) = (5.579, 3.033, 0.297, 1.481)$.

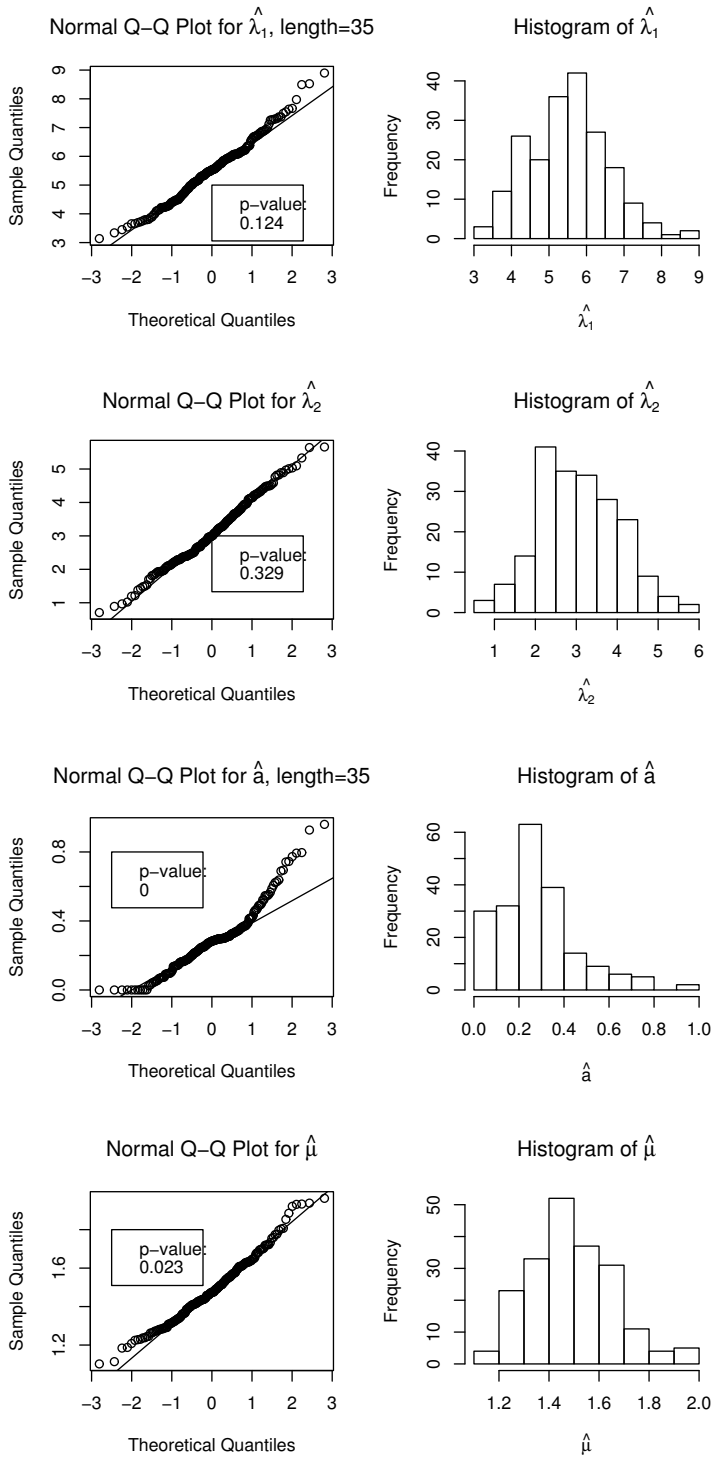


Figure 4.19: Normal Q-Q plots and histograms for estimators $\hat{\lambda}_1, \hat{\lambda}_2, \hat{a}$, and $\hat{\mu}$; length=35, $(\lambda_1, \lambda_2, s, \mu) = (5.579, 3.033, 0.297, 1.481)$.

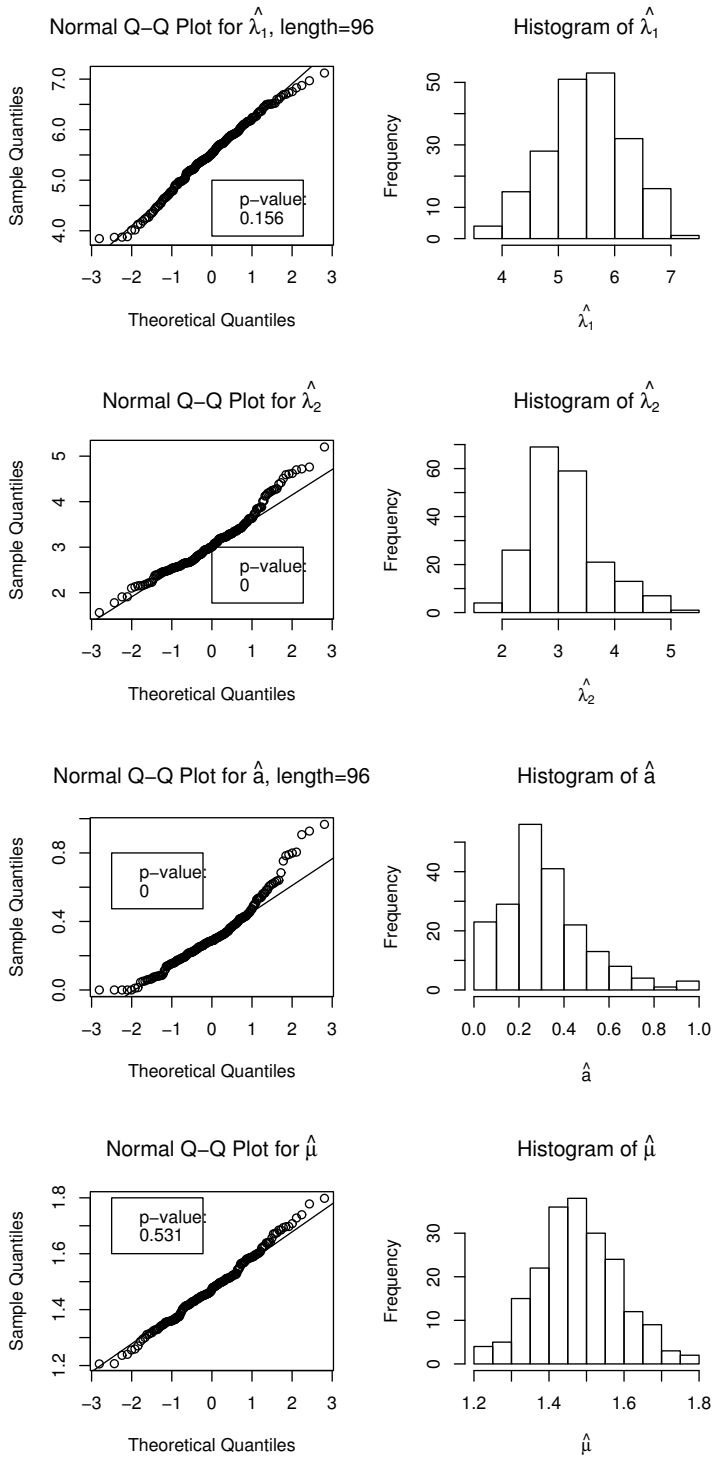


Figure 4.20: Normal Q-Q plots and histograms for estimators $\hat{\lambda}_1, \hat{\lambda}_2, \hat{a}$, and $\hat{\mu}$; length=96, $(\lambda_1, \lambda_2, s, \mu) = (5.579, 3.033, 0.297, 1.481)$.

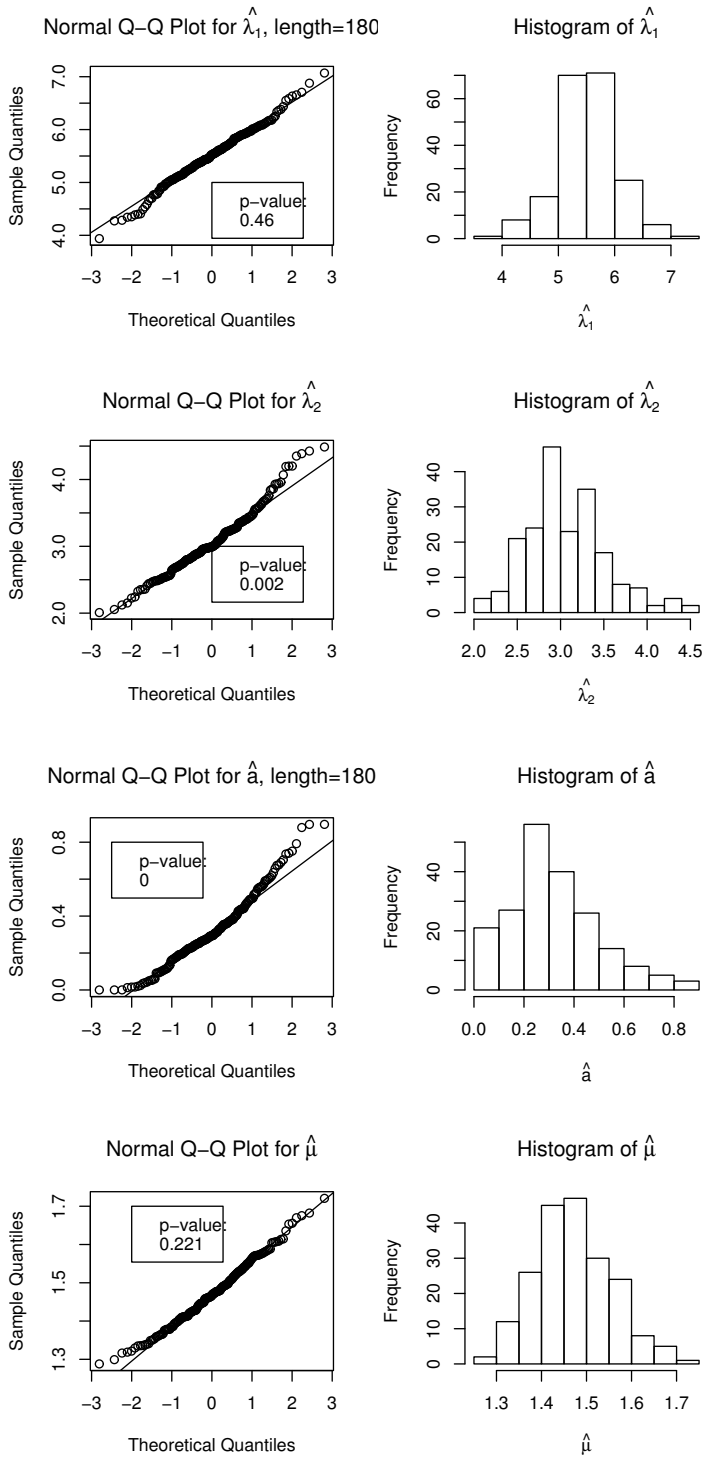


Figure 4.21: Normal Q-Q plots and histograms for estimators $\hat{\lambda}_1, \hat{\lambda}_2, \hat{a}$, and $\hat{\mu}$; length=180, $(\lambda_1, \lambda_2, a, \mu) = (5.579, 3.033, 0.297, 1.481)$.

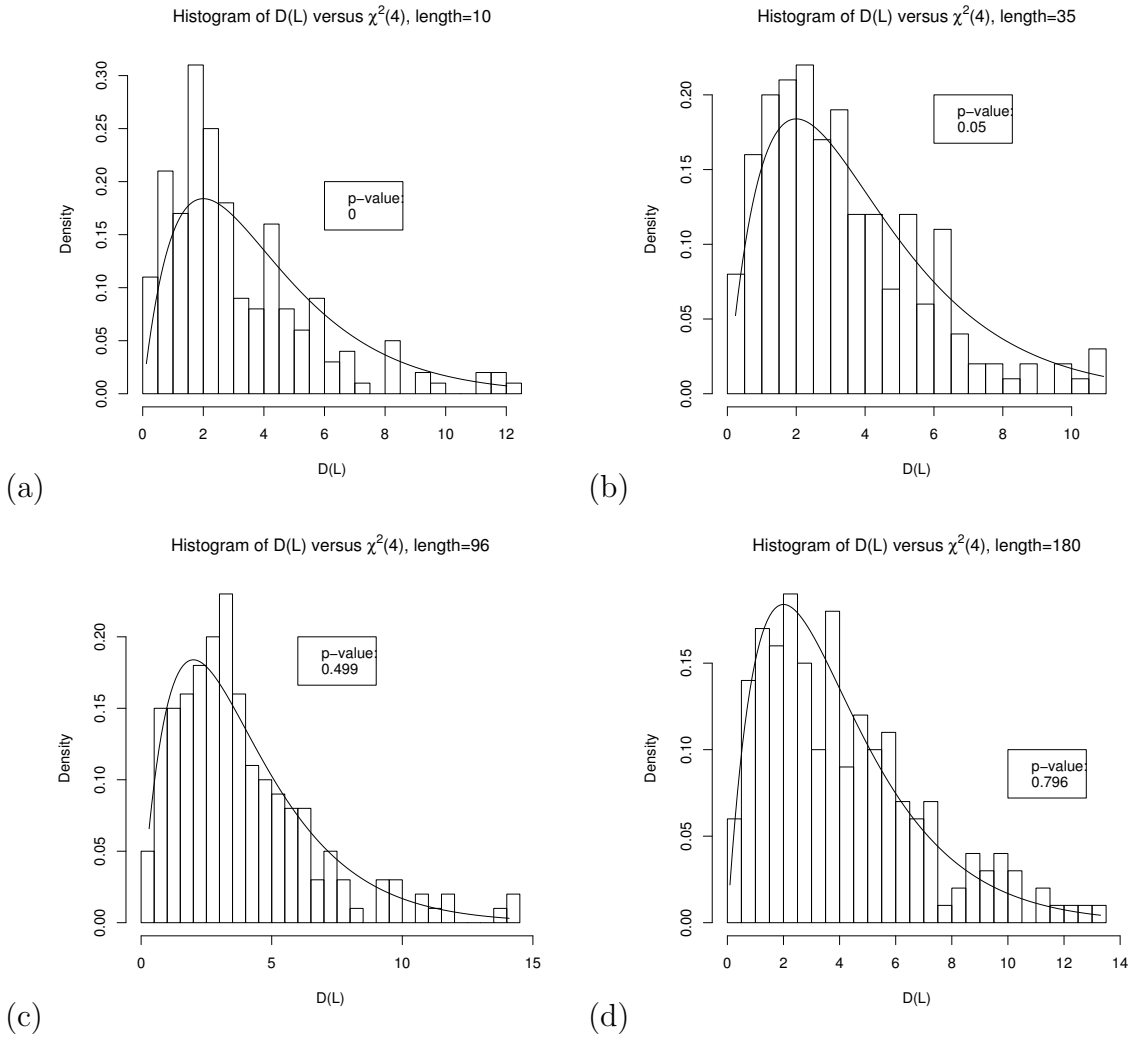


Figure 4.22: Histogram of $D(\mathcal{L})$ with sequence lengths at (a) 10, (b) 35, (c) 96, (d) 180, $(\lambda_1, \lambda_2, a, \mu) = (5.579, 3.033, 0.297, 1.481)$.

CHAPTER 5

Model Validation and Robustness Studies

In this chapter, we discuss model validation. The validity of a model has several aspects. In the following sections, we talk about checking model assumptions for both Model 1 and Model 2. This can be done using informal methods, such as graphical representation or formal methods, such as goodness-of-fit hypothesis tests. Usually a model is not a perfect match for the real process. However, if the proposed model is a good approximation as to the reality, we expect that some minor departures from the assumptions of the model would not influence the estimators greatly. In Section 5.1.1 and 5.1.2, we assess the robustness of the estimation through some simulation studies.

5.1 Validation of Model 1

In Model 1, we have made several specific assumptions about the biological process of the lesion evolution. For the $M/M/\infty$ model, we have the following assumptions:

(C1) The process starts with the first total number of lesions having a Poisson distribution, i.e., the stationary distribution derived from an $M/M/\infty$ queue.

(C2) The $M/M/\infty$ queue assumption.

The $M/M/\infty$ structure we put in the model has fairly strong features. The arrival process of the lesions is Poisson and the service distribution is exponential, which is independent of the arrival process as well. The parameters of these two are constant over time. The motivation of the proposed model we state in Chapter 3 is closely related to our discussions here. It provides some justification of the approach we adopt.

Were we able to measure the exact onset time and the service duration for each lesion, we could have carried out appropriate hypothesis tests to check the above two assumptions. However, the time-slicing data limit our ability to do that. We anticipate that the estimates from fitting Model 1 to a single sequence would not be affected greatly if minor deviations from the assumptions exist.

5.1.1 Assumptions for the Arrival Process

Assumption (C1) is plausible if we assume that the patient has the relapsing-remitting disease for a while. Otherwise we have to turn to the conditional likelihood which is conditioned on the first observed total Y_1 without specifying its distribution. Albert et al. (1994) compared the fitting results and found out that in one of the three patients' total count sequence, the hidden Markov model they proposed worked worse than the independent Poisson model. When Sormani et al. (1999) dealt with the new lesion counts, they also used Poisson distribution for the count of new lesions over a fixed time period. The Poisson assumption has never been formally justified or rejected. However, it might not be true that the arrival process is Poisson with constant rate. For example, if we observe a lot of lesions in the patient's MRI of the brain, there may not be enough space to have new lesions coming in with the usual

rate. Such a situation is very likely to happen when the arrival rate is much larger than the service rate. In the long run, more 'customers' will be accumulating in the system and others will be 'turned down' by the servers. Thus the lesion arrival rate will be dragged down to a lower level, say λ_1. On the other hand, if there are few lesions, the arrival rate may be pulled up from the usual level by the system, say λ_2, since there is more space available for the lesions. However, we expect that, based on our Model 1, the estimators for λ will not be far from the average of the high arrival rate and the low arrival rate, $\frac{\lambda_1+\lambda_2}{2}$. The estimator we acquire in Model 1 should be capable of reflecting the average arrival rate. Also, the estimators for μ should not be influenced much by the increasing and decreasing of the arrival rate, given a constant average arrival rate.

In a small simulation study, we consider the following two schemes to generate the data:

- Scheme 1: The $M/M/\infty$ model with fixed arrival rate λ and service rate μ.
- Scheme 2: The choice of arrival rate λ_1 or λ_2 depends on the number of previous total. If the total is below $\frac{\lambda}{\mu}$, we use the higher arrival rate λ_1, otherwise, we use the lower arrival rate λ_2. Here λ is the simple average of λ_1 and λ_2.

For each of the two schemes, we generate 100 sequences with length 50. The average arrival rate λ is fixed at 6 and the service rate is fixed at 1.3. Such a set of parameters is chosen to ensure that the normal approximation for the MLE would be appropriate. The higher rate λ_1 and the lower rate λ_2 are varied according to the

fixed average. The misspecified loglikelihood function (3.3.6) is evaluated based on each single sequence. The MLE for λ and μ are calculated as well as their standard errors. The bias and MSE (mean squared error) are given. The corresponding 95% confidence intervals are calculated assuming normality of the MLE. The proportion that the confidence intervals cover the true parameter is reported under each scheme. The results are listed in Table 5.1.

Cases	λ			μ		
$(\lambda_1, \lambda_2, \mu)$	$CP(\lambda)$	$Bias(\hat{\lambda})$	$MSE(\hat{\lambda})$	$CP(\mu)$	$Bias(\hat{\mu})$	$MSE(\hat{\mu})$
(7, 5, 1.3)	97%	-0.01	0.53	93%	0	0.11
(6, 6, 1.3)	94%	0.07	0.54	93%	0.03	0.11
(8, 4, 1.3)	95%	0.01	0.53	93%	0	0.11
(9, 3, 1.3)	97%	0.11	0.55	95%	0.01	0.11
(10, 2, 1.3)	93%	0.16	0.57	92%	0.01	0.11
(11, 1, 1.3)	95%	0.21	0.58	94%	0	0.11

$CP(\lambda)$ and $CP(\mu)$: the proportion of confidence intervals which cover the true parameters, $\lambda = 6$ and $\mu = 1.3$.

Table 5.1: Proportion of confidence intervals which cover true λ and μ using Model 1, when the arrival process has average rate $\lambda = (\lambda_1 + \lambda_2)/2$.

We can see that the coverage proportions for both λ and μ are not far from the desired confidence level 95%. The bias and the MSE for the estimators are moderate in all the cases. The estimate $\hat{\mu}$ does not change much when the underlying arrival process changes a lot (i.e., from constant rate to two different rates).

Also, in Chapter 3, when we compared the fitting results for Model 1 and Model 2 in Table 3.2 and 3.3, for each sequence, the estimates for μ seem to be stable. The

estimated service rates for the 7 patients we analyzed using Model 3 are not that different from what we get in Model 2. This suggests that the estimation for the service rate is robust when the constant arrival rate is replaced by different Poisson arrival rates which are governed by a Markov chain.

5.1.2 Assumptions for the Service Distribution

If it is possible to monitor the disease process more thoroughly, we would know whether the true 'service' distribution for the lesion is exponential or not. In our model, this assumption is very important since the memoryless property is crucial. Otherwise we have to specify the remaining service time distribution for every lesion observed at each MRI scan. Thus we want to see the changes in the properties of the estimates if the exponential assumption is not satisfied. This is examined in the following simulation.

The arrival rate for the Poisson process is fixed at 4. We select 6 Gamma alternatives that have the same mean as the exponential distribution with mean 0.83 (rate 1.2). The shape parameters and the scale parameters are denoted by α and β respectively. The density plots for three of them are given in Figure 5.1 to depict the variation among these distributions. As the shape parameter α increases, the tail of the density plot becomes heavier. When α is less than one, the density plot has a convex shape. When α is greater than one, the density plot turns into a concave shape.

Taking each of the gamma distributions as the service distribution, we generate 100 $M/G/\infty$ queueing system datasets each with sequence length 50. For the first observed total lesion count, exponential service is used such that we could ignore

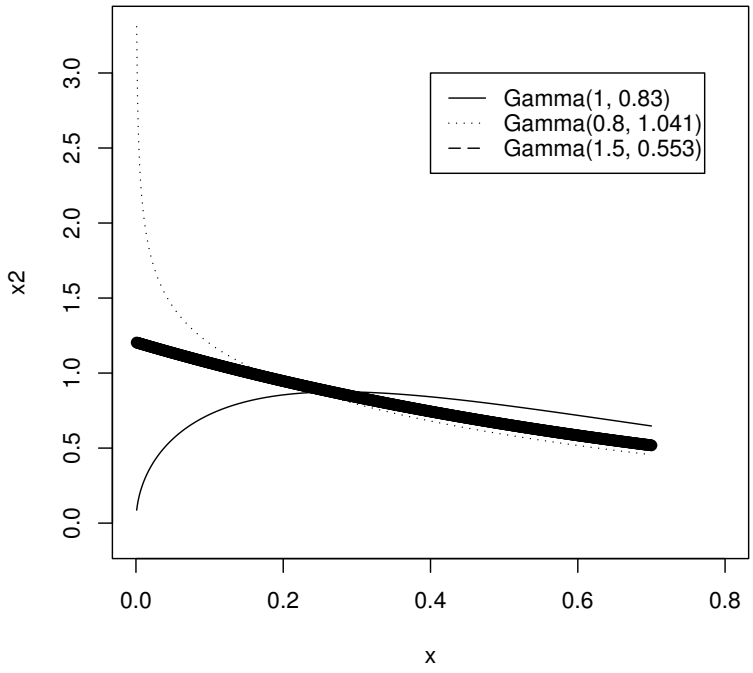


Figure 5.1: Density plots of Gamma(0.8,1.041), Gamma(1, 0.83), and Gamma(1.5, 0.553).

the remaining service time distribution in the simulation. The perturbation on the service distribution is allowed to start after the first observed total. Then we calculate the coverage proportions for the parameters by fitting the $M/M/\infty$ model to the simulated data. The results are listed in Table 5.2.

The coverage proportion for λ decreases from the nominal confidence level of 95% as the service distribution deviates further from the exponential distribution. A similar result holds for the estimator for μ. We notice that the very low coverage proportion when the service distribution is gamma with the shape and the scale

(λ, α, β)	λ	μ
(4, 1, 0.833)	96%	94%
(4, 0.95, 0.877)	97%	93%
(4, 0.9, 0.926)	95%	91%
(4, 0.8, 1.041)	88%	74%
(4, 0.7, 1.19)	71%	61%
(4, 1.5, 0.553)	52%	62%
(4, 2, 0.417)	24%	17%

Table 5.2: Proportion of confidence intervals which cover true λ and μ using Model 1, with the Gamma service distribution.

parameters as 2 and 0.417. It seems that the model does not work well if the service distribution has a form other than exponential. The applicability of our proposed model may be limited because of this. More information for the onset time and the duration of the lesions require a more precise measuring process. If it were possible to track the evolution of each of the observed lesions, it will enable us to model with better approximation to the true biological process.

5.2 Validation of Model 2

For Model 2, the assumptions are:

(D1) The Markov assumption for the process of the underlying states.

(D2) Stationarity of the transition probabilities and the same transition probabilities between the two states (i.e., the symmetric TPM for the Markov chain).

(D3) Homogeneity of the Poisson rate parameters λ_0 and λ_1 and the service rate parameter μ for the $M/M/\infty$ queue.

(D4) The assumption of the $M/M/\infty$ structure.

It does not seem possible for us to check the Markovian assumption of the underlying chain and if it is a stationary Markov chain since the exact transition times are not observable. Informally speaking, the biological process would be relatively stable in the early years of the disease so that we are able to approximate the true process with a much more idealistic one.

We can formulate a test on the equivalence of the transition probabilities between two states. This can be done by comparing the model fitting results from Model 3 to the fitting results from Model 2, since Model 2 is nested in Model 3. The likelihood ratio test has been formulated for the nested standard hidden Markov models (Giudici et al., 2000) with a common known number of the states for the underlying Markov chain. We can follow this idea here.

Assuming that the LRT statistic for the two models has an asymptotic χ^2 distribution with 1 degree of freedom, we can test the validity of the assumption that $a = b$. We calculate the LRT statistic $-2(\log \mathcal{L}_2 - \log \mathcal{L}_3)$ ($\mathcal{L}_2$ and $\mathcal{L}_3$ are the loglikelihood scores for Model 2 and Model 3 respectively) for all the patients who have been fitted with both models in Section 3.5. It seems that that only Patient 7 has a significant p-value 0 with the LRT statistic $9.82 = -2(100.96 - 105.870)$. The AIC and BIC scores also show a better fit using Model 3. The first 5 months with a much higher average new lesion counts may suggest that the symmetric assumption for the hidden Markov chain is not appropriate.

The above comparison needs the assumption that $D(\mathcal{L})$ (Chapter 4) for Model 3 should be approximately χ^2 distribution with 5 degrees of freedom. We find out through a simulation study that it is the case when the sequence is long. It takes

much longer for the estimators to achieve asymptotic normality in Model 3 than in Model 2. However, the approximation may not be good when the sequence is as 28 months long as for Patient 7. Thus the inference should be made with caution.

5.3 Goodness-of-fit

Methods have been proposed to assess the performance of stochastic models in the literature. Usually for time series models, after computing the estimated conditional mean based on the parameter estimates, we can create a diagnostic plot by overlaying the expected mean response on the observed data. This provides a graphical view of the fit. In both Model 1 and Model 2, we can calculate the conditional mean for the total lesion count Y_t and new lesion count $Y_t^{(2)}$ given the past information.

In Model 1, the arrival of the new lesions is independent of the number of persistent enhancing lesions. Thus

$$E(Y_t^{(2)}) = \frac{\lambda(1 - e^{-\mu})}{\mu} \tag{5.3.1}$$

and

$$\begin{aligned} E(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots, Y_1) &= E(Y_t^{(1)} + Y_t(2) \mid Y_{t-1}, Y_{t-2}, \ldots, Y_1) \\ &= Y_{t-1}e^{-\mu} + \frac{\lambda(1 - e^{-\mu})}{\mu}. \end{aligned} \tag{5.3.2}$$

From (5.3.2) we can see a linear association between Y_t and Y_{t-1}. Since both AIC and BIC scores support the use of Model 1 in Patient 1, 2, 3, 8 and 9, we plot Y_t versus Y_{t-1} in these patients. We can see the linear trend in all the plots although sometimes it is moderate.

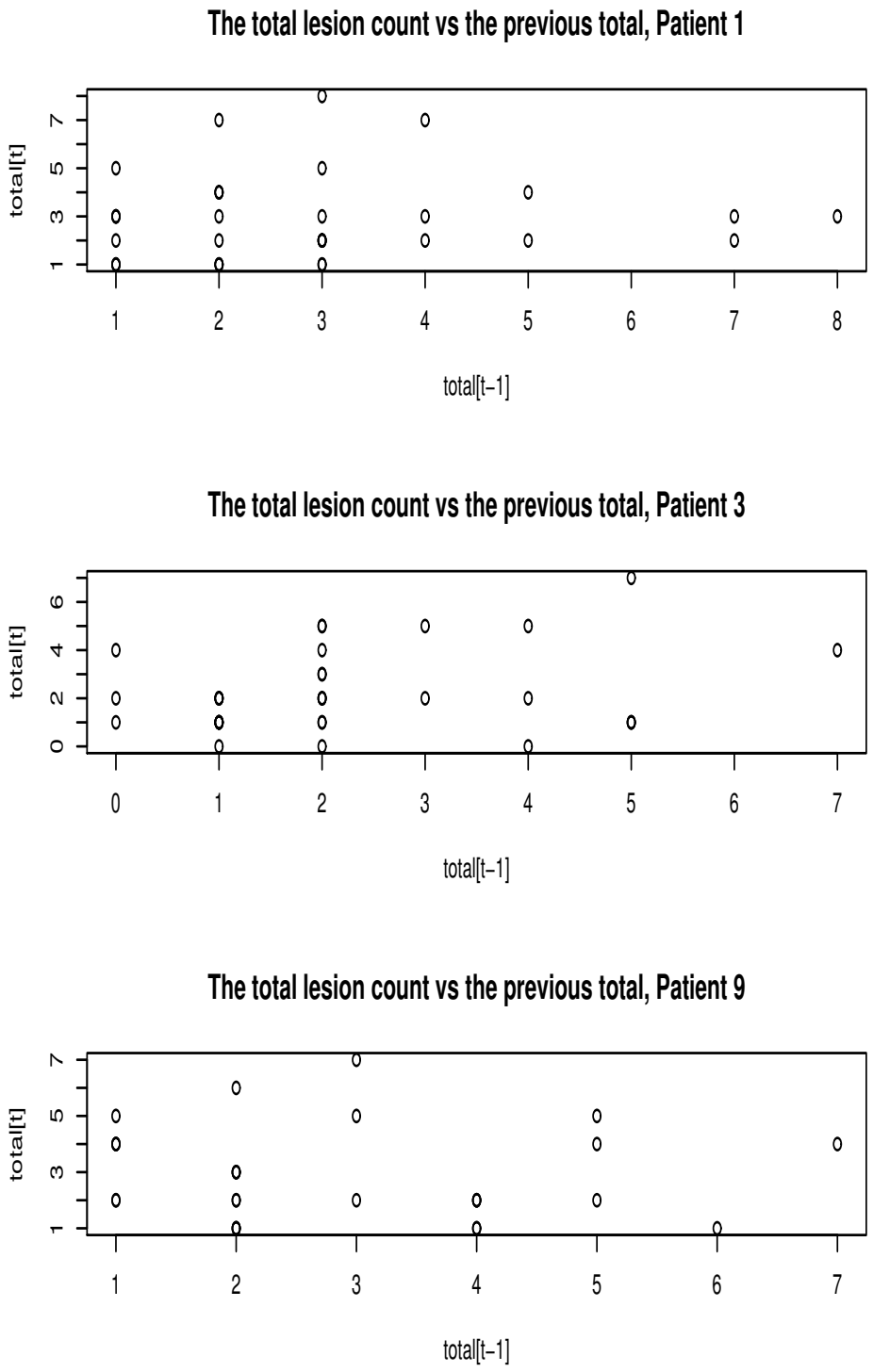


Figure 5.2: Current total lesion count Y_t versus the preceding total Y_{t-1}, Patient 1, 3, and 9.

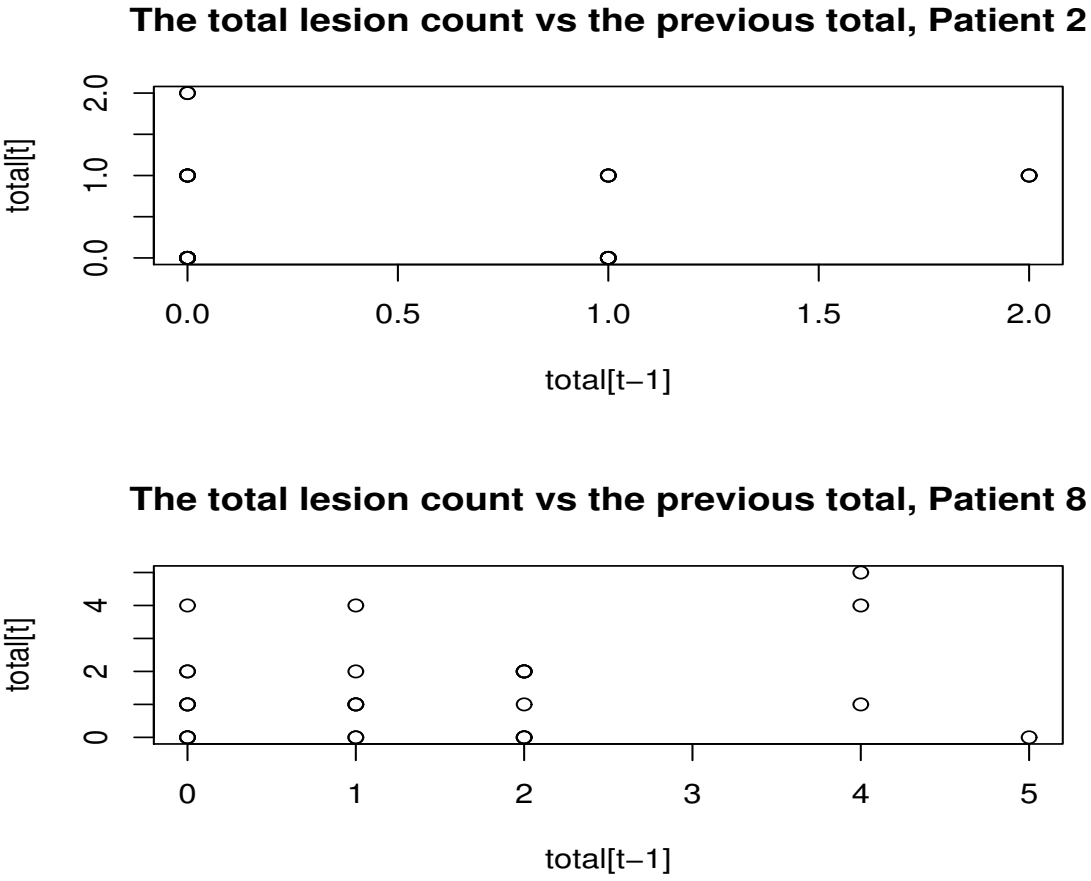


Figure 5.3: Current total lesion count Y_t versus the preceding total Y_{t-1}, Patient 2 and 8.

Since $E(Y_1) = \frac{\lambda}{\mu}$, we have constant mean $\frac{\lambda}{\mu}$ for the marginal total lesion count. It seems natural that we can fit a Poisson distribution with mean r_1 to the total lesion count sequence, and we can fit another independent Poisson distribution with mean r_2 to the new lesion count sequence. Model 1 considers to fit these two sequences together in one model. We can calculate the corresponding estimates for $\frac{\lambda}{\mu}$ and $\frac{\lambda(1-e^{-\mu})}{\mu}$ respectively using the estimates for λ and μ we get by fitting Model 1 to the data. The results are listed in Table 5.3.

	Poisson Model		Model 1	
Patient	r_1	r_2	r_3	r_4
1	2.89	2.24	2.895	2.238
2	0.53	0.36	0.549	0.408
3	2.19	1.32	2.170	1.309
4	5.96	4.73	6.177	4.687
5	4.30	3.40	4.329	3.407
6	4.94	2.79	5.006	2.942
7	4.93	3.15	4.998	3.371
8	1.41	1.04	1.416	1.092
9	2.90	1.78	2.90	1.968

r_1: Poisson rate for the total lesion count.
r_2: Poisson rate for the new lesion count.
r_3: estimates for $\frac{\lambda}{\mu}$ from Model 1.
r_4: estimates for $\frac{\lambda(1-e^{-\mu})}{\mu}$ from Model 1.

Table 5.3: Comparisons of the estimated Poisson rates for the new and the total lesions for Model 1 and the Poisson Model.

We notice that the rates for the total as well as for the new calculated from the two model fitting situations do not differ substantially for all the patients. The biggest difference resides in Patient 4 who has the widest range and the largest average for the total and the new (Table 3.1). When we formulate the likelihood function for

Model 1, the number of persistently enhancing lesions (as a part of the total counts) is modeled by a binomial distribution. When we substitute the Poisson probabilities for the binomial probabilities during the calculation, we found only small difference in the loglikelihood value. Again, the biggest difference corresponds to Patient 4. There is a possibility that the Poisson approximation for the binomial probabilities may work well, which makes the correlation between Y_t and Y_{t-1} weak. Although we can derive the arrival rate and the service rate for the queueing process mathematically from the two Poisson rates from fitting the independent Poisson distributions for the total and the new, the interpretation is not as straightforward as the one from Model 1. Moreover, the MLE computation is fairly easy in Model 1.

For Model 2, the conditional mean of the new is given by

$$\begin{aligned}
&E(Y_t^{(2)} \mid Y_{t-1}^{(2)}, X_{t-1}, Y_{t-2}^{(2)}, X_{t-2}, \ldots Y_1, X_1) \\
&= E(E(Y_t^{(2)} \mid X_t, X_{t-1}, \ldots, X_1)) \\
&= \frac{(1-e^{-\mu})}{\mu} \cdot E({\lambda_1}^{X_t} + {\lambda_2}^{1-X_t} \mid X_{t-1}, \ldots, X_1) \\
&= \frac{(1-e^{-\mu})}{\mu} \cdot ({\lambda_1}^{X_{t-1}}(1-a)^{1-X_{t-1}}a^{X_{t-1}} + {\lambda_2}^{X_{t-1}}(1-a)^{X_{t-1}}a^{1-X_{t-1}}). \qquad (5.3.3)
\end{aligned}$$

The conditional mean of the total is given by:

$$\begin{aligned}
&E(Y_t \mid Y_{t-1}, X_{t-1}, Y_{t-2}, X_{t-2}, \ldots Y_1, X_1) \\
&= Y_{t-1}e^{-\mu} + E(Y_t^{(2)} \mid Y_{t-1}^{(2)}, X_{t-1}, Y_{t-2}^{(2)}, X_{t-2}, \ldots Y_1, X_1) \qquad (5.3.4)
\end{aligned}$$

where the second term is given by (5.3.3).

Other approaches such as qualitatively comparing the observed and expected frequencies of each value can be applied here as well. However, the comparison can only be done based on the total lesion counts or new lesion counts separately. Altman and

Petkau (2002) indicate in their paper that the methods involving the expected mean quantitatively or qualitatively focuse on means rather than on distributions, which might not be appropriate to detect violations of the Poisson assumption.

Altman and Petkau develop a graphical method to evaluate the goodness-of-fit: plot of the estimated univariate/bivariate distribution against the empirical univariate/bivariate distribution. They state that the new method can detect a lack of fit as sequence length gets longer, when the marginal distribution or the correlation structure is misspecified. When fitted to the 3 RRMS total lesion count sequences, the standard Poisson HMM seems to provide a good fit in terms of its marginal feature. However their plots of the estimated bivariate probabilities versus the empirical bivariate probabilities suggest that the two-state HMM is unable to capture the correlation structure. In our opinion, the standard Poisson HMM is not adequate perhaps due to the correlation between the consecutive total lesion counts. The total lesion counts at time t include those old lesions who have been seen enhancing at time $t-1$. This feature has been incorporated by the $M/M/\infty$ structure in our models.

CHAPTER 6

Model Applications

In this chapter, we describe some of the possible applications for the model we have developed here. The Gd-enhancing MRI lesion sequence from an individual RRMS patient could be used to describe the lesion evolution with the enhancing information. We set up hypothesis tests for the disease progression which could be based on the MRI lesion counts using the proposed models. In Section 6.2, we talk about taking into account the heterogeneity among the patients for cross-sectional studies. Finally, in Section 6.3, we discuss how the models could be applied to the planning of RRMS clinical trials.

6.1 Testing Disease Progression

Testing for the disease progression is an important application if a neurologist would like to know how an individual patient is doing during the disease course. Before we answer this question, it is important for us to have some idea about the definition for disease progression in MS patients. Clinically, neurologists are using acute measures such as the clinical relapse rate, number of relapses for disability progression, while a chronic measure such as the change of the EDSS score has been used. For example, in the trial for Avonex, the sustained disability progression is

defined as deterioration from baseline evaluation by at least 1.0 point on the EDSS persisting for at least six months. Survival analysis using time to the first clinical event as the response is the common analytic technique in phase III studies. Many studies suggest that either the enhancing new lesions or the total lesion volume correlates with clinical impairment poorly despite the fact that MRI is useful for detecting sub-clinical disease activity. However, the frequency of active scans and active lesions increases shortly before and during clinical relapses (Comi et al., 1998). If a large number of Gd-enhancing lesions are seen in a period, the patient might be going through more severe inflammation, which may result in a higher chance of demyelination. However, this can only be tested for patients with RRMS since the Gd-enhancing lesions are less seen in other MS types (Comi et al., 1998).

In the following we describe an application of Model 1 to test the disease progression in a specific RRMS patient. Patient 5 is chosen for illustration. This patient seems to have more lesions in the last 15 months than the first 15 months (see Figure 3.4). We can formulate a test to see whether the difference is significant. The total follow-up is 37 months. Given the same patient, we assume that the estimates from the first period is independent of the estimates from the second period. This assumption may be appropriate when the 7-month interval between these two periods is viewed as the washout time. Based on the derivations in Section 3.3.2, we find out that the estimate for (λ, μ) using the first period data is $(4.312, 1.306)$ with standard error $(0.839, 0.243)$. The estimate using the second period data is $(10.054, 1.721)$ with standard error $(1.475, 0.232)$. The second-period lesion arrival rate is significantly larger than the first-period rate. The one-sided p-value is 0.0004 when we compare the observed standardized test statistic to a standard normal variable. Note

that the test statistic here is to standardize $\hat{\lambda}_{11} - \hat{\lambda}_{12}$ by its approximate standard error:

$$\hat{\lambda}_{11} - \hat{\lambda}_{12} \quad \overset{a}{\sim} \quad N(\lambda_{11} - \lambda_{12},\ \hat{se^2}(\hat{\lambda}_{11}) + \hat{se^2}(\hat{\lambda}_{12})). \tag{6.1.1}$$

where λ_{11} and λ_{12} are the arrival rates corresponding to the two periods respectively. On the other hand, there is no evidence that the service rates between these two periods differ. The two-sided p-value here is 0.218.

If λ increases, it implies that on the average, there are more lesions enhancing in the brain MRI, which makes the relapse more possible. This increase does have support from the clinical point of view. We examine the monthly EDSS scores for this patient in Figure 6.1. The EDSS scores for the second period is much higher and more fluctuating compared to the first one. This might suggest that with more inflammatory activity shown in MRI Gd-enhancing lesion counts, the neurological examination findings are consistent.

Model 3 can also be used to test the disease progression in the following way. As we discussed in Section 3.5, the estimates of for a and b can be used to calculate the long-run probability of the Markov chain in the states (i.e., $\frac{1-b}{2-a-b}$ for the higher rate state and $\frac{1-a}{2-a-b}$ for the lower rate state). Overall, If a is significantly bigger than b, that means the patient stays longer in the higher rate state. For example, from Patient 4, the estimate for a is 0.590, which is bigger than the estimate for b, 0.426. Using asymptotic normality, we can see that the test statistic has a value of $\frac{(0.590-0.426)}{\sqrt{0.1273+0.0818+2*0.0568}} = 0.288$. The standard normal p-value is 0.39.

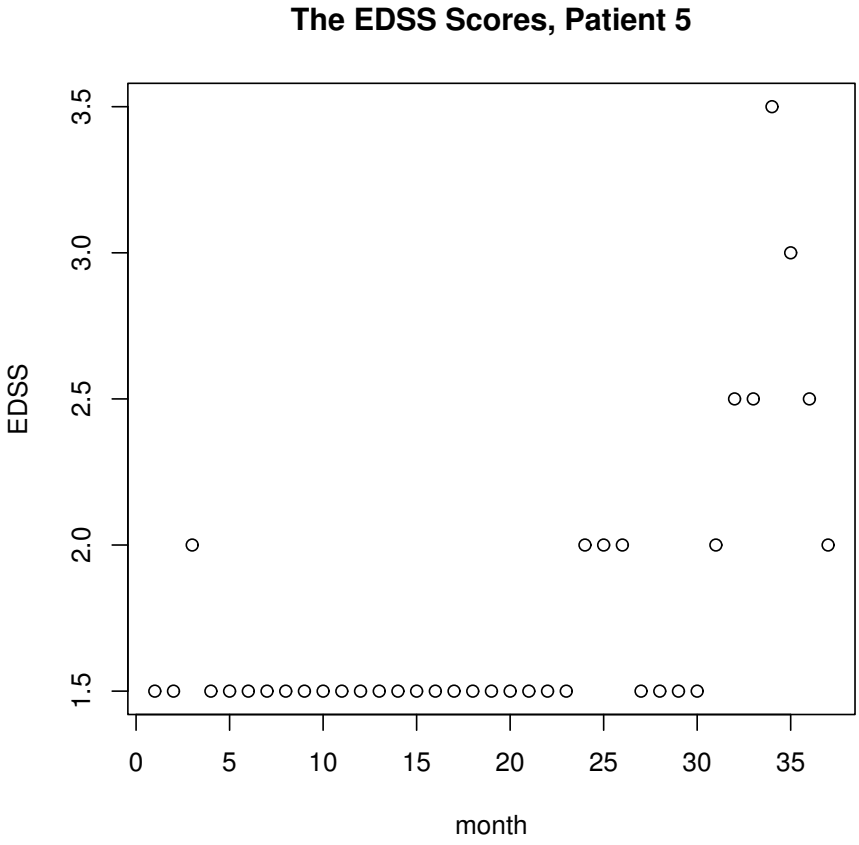


Figure 6.1: Monthly EDSS scores for Patient 5.

There is no evidence that this patient stays longer in the state having more enhancing lesions for this patient. The comparison of the fitting results between Model 2 and Model 3 also supports the claim. The LRT statistic is $-2(94.452 - 94.565) = 0.046$ and the χ^2_1 p-value is 0.84. Thus there is not sufficient evidence that the patient remains in the higher rate state longer given the observed lesion count sequence.

Since there are two arrival rates corresponding to the two underlying different states in Model 2 and Model 3, the disease progression could be shown by the increase of the two rates simultaneously. For example, when the patient has been followed long enough and Model 2 or 3 are more appropriate to describe the lesion evolution process. If there is suspicion that during one period this patient tends to have more lesions than in another period, we can fit the models to these two periods' sequences separately and compare the estimates for λ's. Tests regarding to the increase in λ_1 and in λ_2 can be done respectively and the Bonferroni correction can be made to adjust for multiplicity.

Testing Drug Effect

The Gd-enhancing T1-weighted MRI lesion counts have been considered as an outcome measure in the clinical trials to test the treatment efficacy. The enhancement is more related to BBB damage associated with intense inflammation and it may precede the onset of clinical symptoms although the inflammation may not directly be responsible for disease progression. In most RRMS clinical trials, a positive treatment needs to have an effect on the clinical disease progression, such as relapse rate. Subclinically, the therapeutic strategy that is targeting the inflammation should demonstrate its strength in reducing the lesion arrival rate. When McFarland et al.

(1992) illustrated the application of the Markov regression model they proposed, the number of new lesions, in contrast to the number of total lesions was chosen as the most appropriate measure because an effective treatment would more likely to stop new lesion development than to shorten the duration of the lesion enhancement. Usually the anti-inflammatory drug may inhibit the occurrences of new enhancing lesions.

If a patient serves as his own control, the lesion count data under treatment could be used to estimate the corresponding parameters. These estimates could be compared with those calculated from his natural history data or from the controlled/placebo period. If the drug is effective in reducing the newly enhancing lesion counts, an approach similar to the one in Equation (6.1.1) can be used to do the hypothesis test using Model 1. We illustrate the application in a simulation study.

To simulate the sequences under the null hypothesis, we set the parameters λ and μ equal to 7 and 1.3 respectively, which may allow us to use a reasonably good normal approximation for MLE even the sequence is not very long. We set different values for λ where the treatment effect is demonstrated by the proportional reduction in the arrival rate. The sequences under the alternative hypothesis are simulated using those different λ's and the same μ.

Under the specified treatment effect and the sequence length, we simulate 100 sequences under null hypothesis and another 100 sequences under the alternative. Then these sequences are fitted by Model 1 and the MLEs are determined. Using normal approximation, the rejection region for the test is $\{\hat{\lambda} : \frac{\hat{\lambda}-\lambda_0}{\hat{se}(\hat{\lambda})} < -1.65\}$. Here we use 1.65 for the desired type I error rate 0.05. We also calculate the percentage of the MLE for λ falling in the rejection region based on the simulation under the alternative

	length (α^0)	length (α^0)	length (α^0)
effect*	12 (0.06)	16 (0.04)	20 (0.03)
20%	29%	45%	49%
30%	54%	64%	82%
40%	78%	87%	97%

*: Proportional reduction in lesion arrival rate.
α^0: Estimated significance level by simulation.

Table 6.1: Statistical power for a two-period cross-over study for one patient for a reduction in the lesion arrival rate.

hypothesis. The results are listed in Table 6.1. We can see that ss the sequence length gets longer, we are able to achieve higher power for the same treatment effect.

From the immunologic point of view, the mean duration time of the enhancement may play an important role. Giovannoni et al. (1997) claimed that Gd enhancement may not only detect lesions in their pro-inflammatory phase but also during regression, a phase in which production of ICAM-1 would be decreased. During a longitudinal study, a hierarchy related to the inflammation has been constructed where new enhancing lesions have higher ICAM levels than the persistently enhancing lesions. The temporal profile of Gd-enhancing lesions was suggested for consideration when attempting to correlate inflammatory markers with Gd-enhancement.

Currently there is not a drug targeting the healing rate or shortening the mean duration of the enhancement. However, hypotheses related to the parameter μ can be established using Model 1 as well. Following the same idea for the previous study, we work on another simulation study to demonstrate the drug effect in μ. The same pair of λ and μ is used for the null hypothesis. Since the drug effect is shown by the increase of the μ values, the rejection region using normal approximation would be

$\{\hat{\mu} : \frac{\hat{\mu}-\mu_0}{\hat{se}(\hat{\mu})} \geq 1.65\}$. The desired type I error rate is still fixed at 0.05. Results are reported in Table 6.2. We can see that in order to achieve larger power, longer study is needed.

	length (α^0)	length (α^0)	length (α^0)
effect*	12 (0.03)	16 (0.04)	20 (0.04)
20%	15%	31%	38%
30%	32%	42%	55%
40%	39%	52%	69%

*: Proportional increase in healing rate.
α^0: Estimated Significance level by simulation.

Table 6.2: Statistical power for a two-period cross-over study for one patient for an increase in the lesion healing rate.

We can also test drug efficacy using Model 2 or Model 3. The formulation of the hypothesis depends on where the treatment effect is demonstrated. As most of the RRMS drugs do, the effect is to reduce the newly enhancing lesions no matter the patient is in relapse or remission. Thus the estimate for λ_1 would be expected to decrease in the same degree as the one for λ_2 when the patient takes the drug compared to the estimates from the placebo period.

6.2 Heterogeneity Among Patients

On the one hand, each total lesion count sequence and the new lesion count sequence from T1-weighted Gd-enhancing MRI scans could be used to describe the lesion evolution process for an individual patient. On the other hand, all the patients sequences can be considered together if we want to know characteristics about the

patient population. From Table 3.2 and 3.3, we have noticed that estimates for lesion arrival rates and the enhancing rates are highly variable among that RRMS cohort. It is known that the rate of development of new lesions vary greatly between patients and to a lesser extent, within the same patient over time. In our models, factors which may take account of the patient variation can be included.

Models with Covariates

The number of disease years at the entry of the study may be a potential source for the large patient-to-patient variation. The regional effect might be another relevant factor. Our proposed models can be modified to allow for the influence of such covariates. Borrowing the ideas from the generalized linear models (GLM), we can formulate our parameters incorporating these covariates. For example, in Model 2 or Model 3, the two arrival rates can be formulated as

$$\log \lambda_i = \lambda_{0i} + \beta_i \mathbf{c}_i^{'}, \quad i = 1, 2$$

where $\mathbf{c}_i$ is a row vector of the covariates. We can also formulate our parameter μ in such a way that if we suspect some specific covariate may influence the enhancing mean duration time. For a cross-sectional study, hypothesis tests could be applied using the estimated coefficients to test the covariates effect.

Models with Random Effects

Another way to account for the heterogeneity in our models is to consider the patients as a random sample from a population. There may great variability in the number of new enhancing lesions and the number of total enhancing lesions within RRMS patients over time. The magnitude of the peak number and the frequency

of the peaks vary among patients. Some patients such as Patient 2 in our dataset in Section 3.4 show little or no activity (usually one lesion or none) over the full follow-up and others tending to have active lesions on most monthly MRI scans. Those covariates we have mentioned above may partially explain the variability among patients. However, the factors underlying the high variability are not well understood. Since the disease course could be so heterogeneous across patients, random effects may be added to the modeling of the cross-sectional data.

Suppose we have observed $(Y_{1,i}, \mathbf{Y}_{2,i}, \ldots, \mathbf{Y}_{m_i,i})$ where $i = 1, 2, \ldots, n$, and m_i is the sequence length for Patient i. For example, in Model 1, we could assume that for each Patient i, the observed lesion count sequences $(Y_{1,i}, \mathbf{Y}_{2,i}, \ldots, \mathbf{Y}_{m_i,i})$ follows the $M/M/\infty$ structure with arrival rate λ_i and enhancing rate μ. Here we can assume that $\lambda_i \sim p_0$ where p_0 is some informative prior distribution. For estimation, we have to integrate the random effect out. As we have reviewed in Section 2.2.1, Sormani et al. (1999) include the random effect for the rate of the Poisson distribution of the new enhancing lesion counts over a fixed time period. The prior distribution they used for the Poisson mean is Gamma. Since the resulting negative binomial distribution turns out to give a better fit than the Poisson distribution in their study, a Gamma prior may be one of our candidates. The prior could also be placed on the parameter μ when the patient population has more heterogeneity where the enhancement pattern varies among the patients. Or we can also put random effect both on the parameters λ and μ using some joint prior distribution. Biological evidences are needed for a specific choice. We also have to bear in mind that the more complicated prior we choose, the more computational difficulty we may encounter in our analyses.

6.3 Implications of the Models in the Design of Clinical Trials for RRMS

Many therapeutic drugs have been developed to modify the disease course of RRMS. Successful drugs include the beta-interferon (Betseron, Rebif, and Avonex), anti-inflammatory agents that suppress cell migration into the central nervous system (CNS); and Copaxone, a mixture of peptide fragments thought to act as a decoy for the immune system to spare myelin from further attack. Clinical outcomes such as relapse rate and EDSS score are the primary outcome measures used in phase III clinical trials.

Although Sormani et al. (2002) implement a formal validation of MRI metrics as surrogate markers for clinical outcome, which is the clinical relapse rate in their case, the MRI measures they choose do not satisfy all the criteria for the surrogacy. Usually, if a potential anti-inflammatory therapeutic agent fails to reduce MRI enhancing lesions, it is not promising. Many drugs show their efficacy in the reduction or even cessation of the lesions (Comi et al., 2001). The long term effect of those drugs to control the disease progression is still unknown. Clinical improvement may not follow and new lesions may show up again even after the treatment continues. However, many trial studies still suggest the use of MRI. That is why we can see the MRI outcomes as the primary endpoint in some clinical trials. Overall, the use of MRI measures are limited.

Gd-enhancing MRI outcome measures such as lesion counts and lesion volume are the primary endpoints in phase II studies. For the application of our models in planning of RRMS trials, we also recommend its use in planning phase II studies. Although the correlation between the MRI outcome measures and the clinical outcome

measures is moderate, it is a fact that more frequent acute enhancing lesions are in a patient, the more likely that one of these lesions will involve an area in the CNS that will result in a clinical symptom. The pathology associated with Gd-enhancing lesions is more specific about the inflammatory aspect of the disease, especially the disruptions of the BBB. Those pharmaceutical agents with anti-inflammatory effects such as beta-interferon products should decrease the number of enhancing lesions (Simon, 2003).

We need to demonstrate how we can detect the treatment effect using our proposed models before we start calculating the sample size with desired power. Let us illustrate a possible approach using Model 1 when the treatment effect is to decrease the lesion arrival rate. For a parallel group design, suppose there are n patients in each group and they are followed for the same length m months. Denote the observed sequences for Patient i in group j by $\mathbf{S}_{ij} = (Y_{ij1}, \mathbf{Y}_{ij2}, \ldots, , \mathbf{Y}_{ijm})$ where $\mathbf{Y}_{ijk} = (Y^{(1)}_{ijk}, Y^{(2)}_{ijk})$. Notice that $i = 1, 2, \ldots, m$, and $j = 0, 1$, which represent the placebo and the treatment group, respectively. Assume that

$$\mathbf{S}_{ij}|\lambda_{ij}, \mu \backsim M/M/\infty\ (\lambda_{ij}, \mu) \tag{6.3.1}$$

$$\lambda_{ij} = \lambda_{0j}\epsilon_{ij} \tag{6.3.2}$$

$$\epsilon_{ij}|\theta \backsim \Gamma(\theta, \frac{1}{\theta}). \tag{6.3.3}$$

Now λ_{ij} is the arrival rate for Patient i in group j. We put a random effect on this parameter since the RRMS patients usually present large heterogeneity in the lesion counts. Notice that λ_{ij} follows a gamma distribution with mean λ_{0j}. Our purpose of the experiment is to test whether the absolute difference of the arrival rates between the two groups, i.e., $|\lambda_{00} - \lambda_{01}|$, is significant. The next step is to estimate these two

λ's based on the observed sequences. The likelihood function based on $\mathbf{S}_{ij}$ is:

$$\mathcal{L}(\mathbf{S}_{ij}|\lambda_{0j}, \mu, \theta) \tag{6.3.4}$$

$$= \int_{\epsilon_{ij}} \mathcal{L}(\mathbf{S}_{ij}|\lambda_{0j}\epsilon_{ij}, \mu) d\mathcal{L}(\epsilon_{ij}|\theta). \tag{6.3.5}$$

$$\tag{6.3.6}$$

The computation could be more complicated due to the integraion regarded to evaluate the likelihood function. We can use the likelihood ratio test statistic. If the asymptotic χ^2 distribution works in this case, the rejection region is easy to locate. We can simulate $2m$ observed data, with half of them from some fixed λ_{00} for the placebo group and the other half from $\lambda_{00}(1-\delta)$ where δ is the treatment effect. For this experimental design 100 trials can be generated and the power is calculated as the proportion of trials which yield a significant result based on the LRT statistic.

If Model 2 and Model 3 are considered to design RRMS trials, the effect from the treatment (such as the anti-inflammatory drugs) could be demonstrated through the reduction in both of the two arrival rates corresponding to the two underlying states of the Markov chain. Suppose λ_{ijk} is the arrival rate when the underlying state is k for Patient i in group j. There are two values for k, 0 or 1. We can assume

$$\lambda_{ijk} = \tau_k + \beta_k c_j$$

where c_j is the indicator for treatment. The coefficient β_0 can represent the treatment effect in the reduction of the lesion arrival rate associated with the lower state and β_1 with the higher state. A positive treatment may be expected to show the same amount of decrease in both β's. Following the same idea as presented above, for a parallel group design, we can specify the length of the follow-up (same for the patients in each group), and use simulation to find out the power for a given sample size. By

fitting the model to a data from a natural history study or a pilot study, we can estimate the parameters for the control group. Based on the estimates, the data for the control group is simulated. The same thing can be done for the treatment group by varying β's associated with a specified treatment effect. Then the power would be the proportion of significant tests reported by the LRT statistic, if the asymptotic distribution works well. Or one can use simulation to determine a cut-off value under the null hypothesis of no difference and estimate the power values.

If in future there is a therapeutic agent put to test to see if it has the effect to reduce the duration of the enhancement of the lesions, our model could be the one helping with the experimental design. If the drug can facilitate the service and thus shorten the stay of the lesions, we may assume the treatment is able to increase the service rate μ. And similarly, a fixed effect or a random effect can be included in this parameter.

Above methods are just simple illustrations of how we are going to calculate the sample size for RRMS trials using our proposed models. Many assumptions associated with the drug efficacy have to be put on the design. The practical situation would not be that easy. For the planning of clinical trials for RRMS, there are a lot of issues going on, such as the constraints of time, patient toleration as well as the lack of agent-free patients left from all the previous trials. Efficient use of the information from the limited number of patients requires a judicious choice of appropriate MRI outcome measures that are closely related to the clinical manifestations.

CHAPTER 7

Conclusion and Future Work

In previous chapters we have explored a new modeling scheme for the total and new Gd-enhancing MRI lesion count sequences at the same time. By incorporating the $M/M/\infty$ structure in the models, the lesion evolution at the early stage involving inflammation is viewed as a whole process. Based on the assumptions of $M/M/\infty$, we are able to formulate the arrival rate and the service (enhancement) rate in our models and estimate them with the MLE. The arrival rate could be constant or controlled by a hidden Markov chain with states representing the disease status (relapse and remission). We fit the models to the MRI lesion sequences from 9 RRMS patients and discuss the model fitting results in Chapter 3. For those lesion count sequences where substantial variability is present, the model with $M/M/\infty$ structure on the Markov regime is more appropriate on the basis of the AIC or BIC criteria for model selection.

We have discussed the asymptotic properties of the MLE in Chapter 4. The long-run distribution of the process with the new lesion counts and the persistent lesion counts is derived. Simulation studies show that when the sequence length gets longer, the MLE for the parameters will be asymptotically normal. In Chapter 5, we focus on the model validation and robustness. Simulation results show that the estimators

are robust to minor deviations from the assumptions for the arrival process. However, this is not the case when the exponential service distribution assumption is violated. In Chapter 6, we provide discussions about how the models can be used to test the disease progression, to take the patient-to-patient variation, and to plan future RRMS clinical trials.

Although we have described thoroughly the new modeling approach in this dissertation, some interesting questions need to be considered in future work. Firstly, we need more data to help with the model validation and application. For example, the model with $M/M/\infty$ structure on the Markov regime shows a better fit in some of the patients when we compare the BIC and AIC scores. It seems that the underlying Markov chain is more likely to account of the substantial variations in the lesion counts. However, it is not clear whether the variation is striking in most of the RRMS patients or not.

Secondly, the models we propose originally are motivated from the queueing theory. For application, they have been adapted to approximate a biological process. It is straightforward that the models could also be applied in the context of queueing process. In fact, Gafarian and Ancker (1966) discussed the pros and cons of observing a queueing process via the way of time-slicing and event-history. They compared these two using the mean queue system size in various types of queues. Our approach presents a parametric method to make inference when the underlying process is $M/M/\infty$ queue or the $MMPP/M/\infty$ queue with transitions between the underlying states happen at the regularly spaced time points. The $M/M/\infty$ structure can also be used when the data are from unequally spaced observing time points. The likelihood function in that case would be a little bit more complicated with the

interval length in it. However, since our model depends heavily on the assumptions of $M/M/\infty$, determining how to model the time-slicing data from $M/G/\infty$ queue or $G/G/\infty$ queue raises a tough question. By assuming that the service distribution has a specified form other than exponential, it is important to figure out the remaining service time in the system for the lesions. It would be interesting to examine how these concerns can be included in the model and to propose relevant inference procedures.

Thirdly, rigorous proof for the asymptotic normality of the MLE for the models would be another challenging task. The conditions for the existing central limit theorems in the setting of Markov chain (or mixing properties of the stochastic process with countable state space) may provide insight into establishing the result in this situation.

Fourthly, as we have mentioned in Section 3.5, it is not clear what is the best criteria for the order selection among hidden Markov models. Our models are involved in this topic. Unlike a general hidden Markov model, it has correlated observations given the underlying Markov chain. It may not be straightforward to adopt the approach for HMM. Meanwhile, the graphical method to detect lack-of-fit we mentioned in Section 5.3 could be explored in our models.

The lesion evolution process in RRMS patients is a very complicated biological process. The MRI images have discovered many pathological features of the disease. More advanced technology in the future will bring better understanding of this disease process. To describe the process, stochastic modeling is still a very important tool researchers should consider to work on.

Appendix

A.1 The 9 RRMS MRI Gd-enhancing Lesion Count Sequences

The datasets we are analyzing are sequences of total lesion counts and new lesion counts from nine RRMS patients in a study by NINDS (National Institute of Neurological Disorders and Stroke). These patients were followed for peroids of time for an average of 30 months. Lesion counts were recorded during the MRI scans.

- **Total lesion counts**:

  [Patient 1]: 3,5,4,3,2,2,7,3,2,1,1,2,4,7,2,4,2,1,3,2,1,1,3,8,3,3,1,1,3,1,5,2,3,2,4

  [Patient 2]: 2,1,1,0,0,1,1,1,0,0,1,0,1,0,0,0,0,0,0,0,0,1,0,1,0,0,2,1,0,0,0,2,1,1

  [Patient 3]: 2,2,2,4,0,4,2,5,1,0,1,1,1,1,2,1,1,1,2,1,1,2,2,3,5,7,4,5,1,2,3,2,0,2,5,1

  [Patient 4]: 10,1,1,5,1,1,2,4,11,4,5,6,11,10,1,9,6,5,6,3,6,6,10,19

  [Patient 5]: 5,4,3,4,4,5,5,3,2,2,2,3,3,1,3,3,1,2,7,4,3,2,2,3,5,7,4,8,6,2,6,7,7,13,3,7,8

  [Patient 6]: 6,8,7,8,7,6,5,4,7,7,9,8,3,5,2,7,6,3,5,0,4,2,1,5,5,2,3,2,4,4,1,2,6,12,7

  [Patient 7]: 9,6,4,17,8,4,7,6,3,1,3,3,5,4,5,3,5,1,7,6,3,6,3,3,3,6,2,5

  [Patient 8]: 2,2,2,0,2,2,0,0,0,1,0,0,1,1,0,4,4,5,0,2,1,4,1,2,2,0,1,1,1

  [Patient 9]: 3,7,4,1,2,3,2,6,1,5,2,2,1,4,2,3,5,5,4,1,2,1,4,2,1,4,2,2,3

- **New lesion counts**:

  [Patient 1]: 3,5,4,3,0,1,6,1,2,1,1,2,4,3,2,3,2,0,2,1,1,1,2,8,2,2,1,0,3,1,5,1,3,1,2

  [Patient 2]: 2,1,0,0,0,1,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,1,0,1,0,0,2,1,0,0,0,2,0,0

[Patient 3]: 2,1,2,4,0,4,1,5,0,0,1,1,1,0,1,1,0,0,2,0,0,1,0,2,4,2,1,2,0,2,2,1,0,2,3,0

[patient 4]: 10,0,1,5,1,1,2,4,11,2,4,6,6,4,1,9,3,5,5,2,5,5,9,13

[Patient 5]: 5,3,2,3,2,5,3,2,1,2,2,2,2,1,2,3,0,2,6,3,2,1,2,2,4,6,2,8,5,2,5,6,6,11,2,6,6

[Patient 6]: 6,7,2,4,3,1,2,1,6,4,7,5,1,4,1,5,1,2,2,0,4,0,0,4,1,2,1,2,4,2,1,2,5,11,2

[Patient 7]: 9,5,1,16,3,0,6,2,2,1,3,2,3,3,3,2,2,0,6,5,3,3,2,3,2,6,0,3

[Patient 8]: 2,2,1,0,2,2,0,0,0,1,0,0,1,1,0,4,2,3,0,2,1,3,1,2,1,0,1,0,0

[Patient 9]: 3,7,2,0,1,2,1,5,0,4,0,1,0,4,2,1,3,2,2,1,2,1,4,1,1,4,1,1,2

A.2 Basic R Codes for Fitting the Models

The following are the R codes we have used for the model fitting.

1. **Model 1**:

```
#total is the total lesion count sequence
#new is the new lesion count sequence
m <- length(total)
a <- total[1]+sum(new[2:m])
b <- sum(total[2:(m-1)])-sum(total-new)
c <- sum(total-new)
rootup1 <- -(b*(m-2)-a-c)+sqrt((b*(m-2)-a-c)^2+4*(m-1)*(b+c)*(a+b))
rootup2 <- -(b*(m-2)-a-c)-sqrt((b*(m-2)-a-c)^2+4*(m-1)*(b+c)*(a+b))
rootdn <- -2*(m-1)*(b+c)
broot1 <- rootup1/rootdn
broot2 <- rootup2/rootdn
aroot1 <- a/((m-1)+1/broot1)
```

```
aroot2 <- a/((m-1)+1/broot2)
#now check the hessian matrix
hes <- jac <- matrix(0, ncol=2, nrow=2)
hes[1,1] <- -a/(aroot2)^2
hes[1,2] <- hes[2,1] <- 1/(broot2)^2
hes[2,2] <- -2*(aroot2)/(broot2)^3-b/(broot2)^2-c/(1-broot2)^2
sh <- solve(-hes)
jac[1,1] <- log(1-broot2)/broot2
jac[1,2] <- -aroot2*log(1-broot2)/(broot2)^2-aroot2/((1-broot2)
*broot2)
jac[2,2] <- -1/(1-broot2)
temp <- sqrt(diag(jac%*%sh%*%t(jac)))
mu2 <- -log(1-broot2)
lamda2 <- aroot2/broot2*mu2
```

2. **Model 2**:

```
floglk2 <- function(par)
  {
l1 <- exp(par[1])
l2 <- exp(par[2])
a <- 1/(1+exp(par[3]))
mu <- exp(par[4])

pat1 <- c(0.5, 0.5)
```

```
pat2 <- diag(c(exp(obs[1]*log(l1)-l1/mu), exp(obs[1]*log(l2)
-l2/mu)))
bt <- t(pat1)%*%pat2
bt2s <- mean(bt)
bt <- bt/bt2s

temp1 <- exp(-l1*(1-exp(-mu))/mu)
temp2 <- exp(-l2*(1-exp(-mu))/mu)
gmat <- c(a*temp1, (1-a)*temp1, (1-a)*temp2, a*temp2)

fac <- rep(1, m-1)
for (j in 2:m)
  {
    bt <- bt%*%matrix(c(gmat[1:2]*(l1^nobs[j]),
            %*%gmat[3:4]*(l2^nobs[j])), nrow=2)
    fac[j-1] <- sum(bt)/2
    bt <- bt/fac[j-1]
     }

logl <-  log(bt2s) - log(factorial(obs[1])) -obs[1]*log(mu) +
sum(log(choose(obs[1:m-1], (obs[2:m]-nobs[2:m])))) -
sum(log(factorial(nobs[2:m]))) - log(mu)*sum(nobs[2:m]) -
mu*sum(obs[2:m]-nobs[2:m]) + log(1-exp(-mu))*sum(2*nobs[2:m]+
obs[1:m-1]-obs[2:m]) + log(bt%*%c(1,1)) + sum(log(fac))
```

```
return(-log)
}
obs <- total
nobs <- new
m <- length(obs)
minf <- floglk2(sk22)
#sk22 is the starting point
resm2 <- optim(par=sk22, fn=floglk2, method = "L-BFGS-B",hessian=T,
control=list(maxit=6000, lmm=6,trace=0))
v <- resm2$par
fpar <- c(exp(v[1]), exp(v[2]), 1/(1+exp(v[3])),  exp(v[4]))
est2[i, ] <- c(i,fpar, resm2$value, minf)
```

BIBLIOGRAPHY

[1] Albert, P. S., McFarland, H. F., and Smith, M. E. (1994). Repeated measures time series for modeling counts from a relapsing-remitting disease: application to modeling disease activity in multiple sclerosis. *Statistics in Medicine*, **13**, 453-466.

[2] Altman, R. M. (2004). Assessing the goodness-of-fit of hidden Markov models. *Biometrics*, **60**, 444-450.

[3] Altman, R. M. and Petkau, J. A. (2005). Application of hidden Markov models to multiple sclerosis lesion count data. *Statistics in Medicine*, **24**, 2335-2344.

[4] Bickel, P. J., Ritov, Y., and Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Annals of Statistics*, **26**,1614-1635.

[5] Bhat, U. N., Miller, G. K., and Rao, S. S. (1997). Statistical Analysis of Queueing Systems, 351-394 In: Dshalalow J. H. , ed., *Frontiers in Queueing*, CRC Press, Boca Raton, FL.

[6] Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models.*, Springer, New York.

[7] Cohen, A. J. and Rudick, A. R. (2003). Aspects of multiple sclerosis that relate to clinical trial design and treatment, 3-20 In: Cohen A. J. and Rudick A. R., eds., *Multiple Sclerosis Therapeutics*, 2nd ed., Martin Dunitz, Malden, MA.

[8] Comi, G., Filippi, M., Wolinsky, J. S., the European/Canadian Glatiramer Acetate Study Group. (2001). European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of Glatiramer Acetate on magnetic resonance imaging-measured disease activity and burden in patients with relapsing multiple sclerosis. *Annals of Neurology*, **49**, 290-297.

[9] Comi, G., Filippi, M., Rovaris, M., Leocani, L., Medaglini, S., Locatelli, T. (1998). Clinical, neurophysiological, and magnetic resonance imaging correlations in multiple sclerosis. *Journal of Neurology, Neurosurgery, and psychiatry*, **64**, (Supplement), s21-s25.

[10] Cooper, B. R. (1981). *Introduction to Queueing Theory*, Elsevier North Holland.

[11] Eick, S. G., Massey, W. A. and Whitt, W. (1993). The physics of the $M_t/G/\infty$ queue. *Operations Research*, **41**, 731-742.

[12] Fischer, W. and Meier-Hellstern, K. (1992). The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation*, **18**, 149-171.

[13] Gafarian, A. V. and Ancker Jr., C. J. (1966). Mean value estimation from digital computer simulation. *Operations Research*, **14**, 25-44.

[14] Giovannoni, G., Lai, M. H., Kidd, D., Chamoun, V., Thompson, A. J., Miller, D. H., Feldmann, M., Thompson, E. J. (1997). Longitudinal study of soluble adhesion molecules in multiple sclerosis: correlation with gadolinium enhanced magnetic resonance imaging. *Neurology*, **48**, 1557-1565.

[15] Giudici, P. L., Rydén, T., and Vandekerkhove, P. (2000). Likelihood-ratio tests for hidden Markov models. *Biometrics*, **56**, 742-747.

[16] Gross, D. and Harris, C. M. (1974). *Fundamentals of Queueing Theory*, John Wiley and Sons, New York.

[17] Karlin, S. (1966). *A First Course in Stochastic Processes*, Academic Press, New York and London.

[18] Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, **33**, 1444-1452.

[19] Lai, M. H., Hodgson, T.,and Gawne-Cain, M. (1996). A preliminary study into the sensitivity of disease activity detection by serial weekly magnetic resonance imaging in multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, **60**, 339-341.

[20] Li, D. K. B., Paty, D. W., the UBC MS/MRI Analysis Research Group, the PRISMS Study Group. (1999). Magnetic resonance imaging results of the PRISMS trial: a randomized, double-blind, placebo-controlled study of interferon-β1a in relapsing-remitting multiple sclerosis. *Annals of Neurology*, **46**, 197-206.

[21] Lublin, F. D. and Reigold, S. C. (1996). Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology*, **46**, 907-911.

[22] MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*, Chapman and Hall, New York.

[23] McFarland, H. F., Frank, J. A., Albert, P. S., Smith, M. E., Martin, R., Harris, J. O., Patronas, N., Maloni, H., McFarlin, D. E. (1992). Using gadolinium-enhanced magnetic resonance imaging to monitor disease activity in multiple sclerosis. *Annals of Neurology*, **32**, 758-766.

[24] Medhi, J. (2003). *Stochastic Models in Queueing Theory*, 2nd ed., Academic Press, Amsterdam.

[25] Miller, D. H. and Frank, J. A. (1998) Magnetic resonance imaging techniques to monitor short term evolution of multiple sclerosis and to use in preliminary trials. *Journal of Neurology, Neurosurgery and Psychiatry*, **64**, supplement, 44-46.

[26] Miller, D. H., Grossman, R. I., Reingold S. C., McFarland, H. F. (1998). The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain*, **121**, 3-24.

[27] Molyneux, P. D., Filippi, M., Barkhof, F., Gasperini, C., Yousry, T. A., Truyen, L., Lai, H. M., Rocca, M. A., Moseley, I. F., Miller, D. H. (1998). Correlations between monthly enhanced MRI lesion rate and changes in T2 lesion volume in multiple sclerosis. *Annals of Neurology*, **43**, 332-339.

[28] Nauta, J. J. P., Thompson, A. J., Arkhof, F., Miller, D. H. (1994). Magnetic resonance imaging in monitoring the treatment of multiple sclerosis patients, statistical power of parallel-groups and crossover designs. *Journal of Neurological Science*, **122**, 6-14.

[29] Newell, G. F. (1982). *Applications of Queueuing Theory*, Chapman and Hall, London.

[30] Noseworthy, J. H., Lucchinette, C., Rodriguez, M., Weinshenker, B. G. (2000). Multiple sclerosis. *New England Journal of Medicine*, **343**, 938-952.

[31] Prentice, R. L. (1989) Surrogate markers in clinical trials: definition and operational criteria. *Statistics in Medicine*, **8**, 431-440.

[32] Robert, O. G. and Rosenthal, S. J. (2004), General state space Markov chains and MCMC algorithms. *Probability Surveys*, **1**, 20-71.

[33] Rammohan, K. W. (2003). Axonal injury in multiple sclerosis. *Current Neurology and Neuroscience Reports*, **3**, 231-237.

[34] Simon, J. H. (2003). Measures of gadolinium enhancement in multiple sclerosis, 97-124 In: Cohen, J. and Rudick, R., eds., *Multiple Sclerosis Therapeutics*, Martin Dunitz, Malden, MA.

[35] Sormani, M. P., Bruzzi, P. and Miller, D. H. (1999). Modeling MRI enhancing lesion counts in multiple sclerosis using a negative binomial model: implications for clinical trials. *Journal of the Neurological Sciences*, **163**, 74-80.

[36] Sormani, M. P., Miller, D. H., Comi, G., Barkhof, F., Rovaris, M., Bruzzi, P., Filippi, M. (2001). Clinical trials of multiple sclerosis monitored with enhanced MRI: new sample size calculations based on large data sets. *Journal of Neurology, Neurosurgery and Psychiatry*, **70**, 494-499.

[37] Sormani, M. P., Bruzzi, P., Beckmann, K., Wagner, K., Miller, D. H., Kappos, L., Filippi, M. (2002). MRI metrics as surrogate markers for clinical relapse rate in relapsing-remitting MS patients. *Neurology*, **58**, 417-421.

[38] Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, **44**, 1019-1031.