RICE UNIVERSITY

Parameter Estimation for Discretely Observed Continuous-Time Markov Chains

by

Roxy D. Cramer

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Katherine B. Ensor, Chair
Professor of Statistics

Keith A. Baggerly
Assistant Professor of Statistics

Richard R. Batsell
Associate Professor, Jones Graduate School of Management

Peter Olofsson
Lecturer, Department of Statistics

Houston, Texas

May, 2001

UMI Number: 3021108

UMI Microform 3021108

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Parameter Estimation for Discretely Observed Continuous-Time Markov Chains

by

Roxy D. Cramer

This thesis develops a method for estimating the parameters of continuous-time Markov chains discretely observed by Poisson sampling. The inference problem in this context is usually simplified by assuming the process to be time-homogeneous and that the process can be observed continuously for some observation period. But many real problems are not homogeneous; moreover, in practice it is often difficult to observe random processes continuously. In this work, the Dynkin Identity motivates a martingale estimating equation which is no more complicated a function of the parameters than the infinitesimal generator of the chain. The time-dependent generators of inhomogeneous chains therefore present no new obstacles. The Dynkin Martingale estimating equation derived here applies to processes discretely observed according to an independent Poisson process. Random observation of this kind alleviates the so-called aliasing problem, which can arise when continuous-time processes are observed discretely. Theoretical arguments exploit the martingale structure to obtain conditions ensuring strong consistency and asymptotic normality of the estimators. Simulation studies of a single-server Markov queue with sinusoidal arrivals test the performance of the estimators under different sampling schemes and against the benchmark maximum likelihood estimators based on continuous observation.

Acknowledgments

Many people made this work possible. I am grateful to my thesis advisor, Dr. Katherine Ensor, for her indispensable guidance and advice during this research. For their valuable input and consideration, I thank my committee members, Drs. Keith Baggerly, Peter Olofsson, and Randy Batsell. To Dr. Peter W. Glynn I extend sincere thanks for graciously providing invaluable correspondence at the planning stages and over the course of this work. I also owe many thanks to Dr. Jan Hewitt of the Cain Project for her vigorous assistance and attention to the early drafts of my thesis.

I wish to thank the faculty and staff of the Department of Statistics, with special thanks to Dr. James Thompson for his support and encouragement during my years here, and to Katherine Ensor, who time and again brought wonderful opportunities my way. I also thank the department secretary, Diane Brown, for assisting me in many ways during my graduate studies. I thank my fellow students for many enriching and lasting friendships. Finally and always, I thank my family. Their enduring love and patience has made everything possible.

This work was supported in part by NSF Grant DMS-9704648 and funded in part through the Statistical Consulting Lab (SCL), Rice University.

Contents

Illustrations

Tables

Chapter 1

Introduction

This work develops a new method of parameter estimation for inhomogeneous continuous-time Markov chains. The method applies to processes discretely observed at times randomly generated by a Poisson process. It thereby addresses two standard problems in statistical inference for stochastic processes: the aliasing problem and the often over-simplifying assumption of time homogeneity.

A continuous-time stochastic process is indexed by time t ranging over a continuous interval (e.g., $t \in [0, \infty)$). Thus, to every time point $t \geq 0$ there corresponds a random variable X_t whose probability distribution P_t^θ depends on t and on a collection of other parameters, $\theta \in \mathbb{R}^d$. The method developed in this thesis estimates the unknown parameters based on a sample from a given continuous-time process.

A process is continuously observed over an interval $[T_0, T]$ if the state $X(t)$ is known for any $t \in [T_0, T]$. A sample obtained in this way is the actual realization of the process over the observation period. A continuously observed sample is most desirable for the purposes of inference, since it gives maximal information about the underlying process. In practice, however, it is often difficult to obtain continuously observed samples; more often, data points are taken at discrete-time points, often at fixed intervals. The aliasing problem stems from the fact that on regularly spaced observation points, distinct continuous-time processes may look the same in the sense that they have the same joint distribution. In this way, the true underlying process may well be *aliased* with false models. Based on a sample, an estimation method must be able to distinguish between the many possible processes in a statistically reliable sense; if the true model is aliased, then full identification is not possible.

An appealing way to alleviate the aliasing problem in discrete samples is to take observations at randomly determined time points. The general idea is that randomly generated observation times support all times on a continuous interval, thereby precluding many *candidate* processes. When identification is possible, a random observation scheme reduces the inference problem to that of a discrete-time process. Moreover, a random sampling scheme offers a means of reducing data size. This is appealing since data collection, storage, and manipulation are all costly endeavors, especially as it is now the age of the massive data set. This thesis embraces the idea of random observation, basing the new estimators on samples drawn at the event epochs of a Poisson process.

A second problem in inference for stochastic processes is the often over-simplifying assumption of time-homogeneity. The behavior of a homogeneous process is probabilistically the same over time intervals of equal length. Homogeneous models are mathematically simpler and therefore attractive, but the assumption can be an over-simplification that in effect applies the wrong model. In modeling a queueing process, for instance, the assumption of homogeneity implies that arrival rates and service rates are constant; this would suggest that, say, over any hour during the day the queue exhibits a similar pattern. For many real queueing applications this kind of model is inadequate. As simple examples, consider the line at a lunch counter at noon, or cars at a traffic light at rush hour; during these hours, the pattern of arriving customers is characteristically different from that an hour earlier or an hour later. Peak and lull times are common in more interesting applications, such as in the demands for cellular phone connections and Internet access. These applications call for models that account for time-dependent rates.

The homogeneity assumption is prevalent in applications because homogeneous models are more tractable than their inhomogeneous counterparts. The assumption of homogeneity makes it easier to prove asymptotic properties of estimators and many important results have already been established. But since homogeneous models are inappropriate for many applications, there is strong motivation to find tractable and effective methods that do not rely on the assumption of time homogeneity. A strategy that has emerged in the literature is to base inference on the infinitesimal generator of a continuous-time process rather than on the transition probability matrix, which is usually difficult to find. The method developed here adopts this strategy as a way to accommodate inhomogeneity in a Markov chain.

The estimation method developed in this thesis uses a martingale-based estimating equation. The Dynkin martingale, which applies generally to Markov processes, motivates a sample analog that inherits the martingale property for samples discretely observed at the event epochs of an independent Poisson process. By then applying well-known results for discrete-time martingales, further arguments suggest that the estimators have good statistical properties such as strong consistency and asymptotic normality.

An implementation demonstrates estimator behavior for a queueing model simulated with time-dependent arrivals. In particular, the example model is a single server queue with time dependent Poisson arrivals and exponential service times. The arrival rate function is a linear function of θ and depends on time through a sinusoidal term with period 24 (e.g., a queue with a 24 hour cycle). The simulation study tests the performance of the estimators for varying sampling rates and test functions, and against maximum likelihood estimators based on continuously observed samples. The results indicate that the performance of the new estimators is on par with maximum likelihood, showing that the new method offers a tractable and convenient way of obtaining decent estimates.

Chapter 2 presents the theoretical and statistical context of this research. Section 2.1.1 reviews the structure and properties of inhomogeneous, continuous-time Markov chains. The Dynkin martingale is defined in Section 2.1.2. Sections 2.2 and 2.3 describe the inference problem and provide an overview of parametric estimation for continuous-time stochastic processes. A review of the literature is given in Section 2.4.

Chapter 3 presents the estimating equation based on the Dynkin Martingale. Section 3.1 presents the sample analog, based on samples drawn according to an independent Poisson process. Section 3.2 provides a review of martingale-based estimating functions and a strategy for establishing asymptotic properties of estimators derived from martingale-based estimating equations is outlined in Section 3.3. The connection to Dynkin Martingale estimators is made in Section 3.4.

Chapter 4 provides an example of the Dynkin Martingale estimators derived for the $M_t/M/1$ queueing model with time-dependent arrival rates. The model and method of simulation is described in Section 4.1. The estimating equation is derived and preliminary results are shown in Section 4.2. and Section 4.2.2 contains a comparison of the new estimators with maximum likelihood estimators. Section 4.3 considers a weighted version of the estimating equation. Finally, Chapter 5 provides conclusions and directions for further study.

Chapter 2

Theoretical and Statistical Context

A continuous-time Markov chain, or CTMC, is a Markov process discrete in space and continuous in time. Under regularity conditions, the sample path $X(t)$ of such a process is right-continuous with left-hand limits. In effect, this means that the process remains in a given state for a length of time and then "jumps" instantaneously to a new state, with an intensity defined by its *infinitesimal generator*. Under general conditions, the infinitesimal generator of a Markov process satisfies the so-called *martingale problem*. The result, known as the Dynkin martingale, forms the basis of the new estimation method developed in this thesis.

This chapter first reviews the mathematical structure of an inhomogeneous CTMC. Section 2.1.2 reviews the Dynkin Martingale and how it arises in continuous-time for the inhomogeneous Markov chain. Section 2.2 follows, with a summary of estimation methods used prevalently in the context of parametric inference problems for stochastic processes. Section 2.3 provides an overview of methods of sampling from continuous-time processes. A review of the literature is given in Section 2.4.

2.1 Theoretical Context

2.1.1 Structure of a Continuous-Time Markov Chain

This section outlines the mathematical and probabilistic structure of a continuous-time Markov chain (CTMC). These processes are widely studied: the focus here is on "well-behaved" processes with infinitesimal generators depending on time. More thorough treatments are plentiful, see for example Barucha-Reid (1988) [5].

Let $\{X(t), t \geq 0\}$ be a continuous-time stochastic process taking values in a countable state space, S. Technically, $X(\cdot)$ denotes the sample path of the random variable $X(\omega, \cdot)$ defined on some probability space $(\Omega, \mathcal{F}, P)$ where $\mathcal{F}$ is the sigma-field generated by the open sets of Ω. That is, for every $\omega \in \Omega$, $X(\omega, \cdot)$ is a function from $[0, \infty)$ into S and for every $t \geq 0$, $X(\cdot, t)$ is an $\mathcal{F}$-measurable function from Ω to S. Measure-theoretic details are omitted for the most part in this thesis, deferring the interested reader to the vast literature on the theory of stochastic processes.

Transition Probabilities and the Markov Property

A right-continuous process with left-hand limits remains in a given state for a random length of time and then jumps to a new state. The probability measure governing this behavior is the *transition probability function*,

$$P(s, i; t, j) = p_{ij}(s, t) = P[X(t) = j | X(s) = i] \quad s < t$$

defined for each $i, j \in S$ as the probability that the process assumes state j at time t given that it is in state i at time s. Standard regularity conditions on the transition probabilities are that $0 \leq p_{ij}(s,t) \leq 1$, $\lim_{t \downarrow s} p_{ii}(s,t) = 1$, and that for any $i \in S$ and $0 \leq s < t$,

$$\sum_{j \in S} p_{ij}(s,t) = 1.$$

A process evolving continuously in time according to the transition probabilities, $p_{ij(s,t)}$, is said to be a *Markov* process if it has the Markov property. The Markov property is the property that given the states, $j, i_1, \ldots, i_{n-1} \in S$ and the times, $0 \leq t_1 < t_2 < \cdots < t_n$, the probability of transitioning to a state j at time t_n depends only on the most recent information, $X(t_{n-1}) = i_{n-1}$; that is,

$$P(X(t_n) = j | X(t_1) = i_1, \ldots, X(t_{n-1}) = i_{n-1}) = P(X(t_n) = j | X(t_{n-1}) = i_{n-1}).$$

Thus a Markov process has the property that the future of the process, given the present, is independent of the past. In this sense, a Markov process is *memoryless*. Further, if for any $t, s \geq 0$ and $h > 0$,

$$p_{ij}(t, t+h) = p_{ij}(s, s+h),$$

so that the probability of transition from state i to state j depends only on the length of the interval, h, the process is *homogeneous* in time. Otherwise, it is *inhomogeneous*.

The Infinitesimal Generator and Kolmogorov Equations

The *intensity function*, $q_i(t)$ is defined so that

$$q_i(t)\Delta t + o(\Delta t)$$

is the probability that $X(t)$ will undergo a random change in Δt given that $X(t) = i$. Then

$$p_{ii}(t, t + \Delta t) = 1 - q_i(t)\Delta t + o(\Delta t) \qquad (2.1)$$

is the probability that $X(t)$ remains in state i at least through time $t + \Delta t$. Taking the limit as $\Delta t \to 0$, $q_i(t)$ is the intensity at time t with which the process will jump out of i into some other state. In particular, for any state i,

$$\lim_{\Delta t \to 0} \frac{1 - p_{ii}(t, t + \Delta t)}{\Delta t} = q_i(t).$$

The *relative transition probability function*,

$$Q_{ij}(t + \Delta t) = P[X(t + \Delta t) = j | X(t) = i \text{ and } X(t + \Delta t) \neq i],$$

is the conditional probability that the process will be in j at time $t + \Delta t$ given that the current state is i and that $X(\cdot)$ undergoes a change in Δt. As $\Delta t \to 0$, $Q_{ij}(t)$ behaves as a transition probability at the jump instants. That is,

$$\lim_{\Delta t \to 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t} = q_i(t)Q_{ij}(t).$$

The intensity and relative transition functions determine the infinitesimal generator.

The *infinitesimal generator*, $A(t)$, is the (infinite) matrix indexed by the states, $i, j \in S$, having elements

$$a_{ij}(t),$$

where

$$\begin{aligned} a_{ii}(t) &= -q_i(t) \text{ and} \\ a_{ij}(t) &= q_i(t)Q_{ij}(t),\ i \neq j. \end{aligned} \tag{2.2}$$

If the process is in state i in the instant just before time t, then at time t it will either stay in state i or jump to some other state, j. In view of the definitions of $q_i(t)$ and $Q_{ij}(t)$, the infinitesimal generator governs the instantaneous behavior of the process, $X(t)$.

The *Kolmogorov* differential equations relate $p_{ij}(s,t)$ to the relative transition function and intensity functions, $q_i(t)$, and $Q_{ij}(t)$, and therefore, to the infinitesimal generator. The *forward* Kolmogorov equation involves the derivative of the transition probability with respect to the (forward) time t:

$$\frac{\partial p_{ij}(s,t)}{\partial t} = -q_j(t)p_{ij}(s,t) + \sum_{k=0}^{\infty} q_k(t)Q_{kj}(t)p_{ik}(s,t);$$

the *backward* equation involves the derivative with respect to the backward time, s:

$$\frac{\partial p_{ij}(s,t)}{\partial s} = q_i(s)[p_{ij}(s,t) - \sum_{k=0}^{\infty} Q_{ik}(s)p_{kj}(s,t)].$$

Both equations have the initial condition,

$$p_{ij}(s,s) = \delta_{ij} \quad = \quad 1 \text{ if } i = j$$

$$= \ 0 \text{ otherwise.}$$

The forward and backward differential equations derive from the *Chapman-Kolmogorov* functional equation.

$$p_{ij}(s,t) = \sum_{k=0}^{\infty} p_{ik}(s,\tau)p_{kj}(\tau,t) \quad s \leq \tau \leq t.$$

The forward, backward, and Chapman-Kolmogorov equations in matrix form are, respectively,

$$\frac{\partial}{\partial s}P(s,t) = -A(s)P(s,t) \tag{2.3}$$

$$\frac{\partial}{\partial t}P(s,t) = P(s,t)A(t) \tag{2.4}$$

and

$$P(\tau,t) = P(\tau,s)P(s,t) \quad 0 \leq \tau \leq s \leq t.$$

Equations 2.3 and 2.4 relate the transition probability function to its infinitesimal generator. If $A(t) = (a_{ij}(t))$ where for any $i,j \in S$, $a_{ij}(t)$ is a bounded continuous function of t satisfying

$$\begin{aligned} & a_{ii}(t) < 0, \\ & a_{ij} \geq 0, \text{ for } j \neq i \text{ and} \\ & \textstyle\sum_{j \in S} a_{ij}(t) = 0, \end{aligned} \tag{2.5}$$

and if for any $0 \leq s \leq t$,

$$\sum_{j=0}^{\infty} p_{ij}(s,t) = 1,$$

then the derivative in 2.3 exists and exists in 2.4 for almost all t. See Feller (1940) [14], or the discussion in Barucha-Reid (1988) [5], Chapter 2 for further details.

Ergodicity of Inhomogeneous CTMCs

The term *ergodicity* means the existence of a long-run or *invariant* distribution. The implication of an invariant distribution is that the sequences $\{X(t_1), X(t_2), \ldots\}$ and $\{X(t_1+s), X(t_2+s), \ldots\}$ will have the same probabilistic structure. (Many authors call the invariant distribution a *stationary* distribution and the terms stationarity and ergodicity are often interchanged.) Based on the papers, Johnson and Isaacson (1988) [30] and Zeifman and Isaacson (1994) [44], this section reviews ergodicity for inhomogeneous continuous-time Markov chains.

For the following, let $\{X(t), t \geq 0\}$ be an inhomogeneous continuous-time Markov chain with transition probability matrix $P(s,t)$ and associated infinitesimal generator $A(t)$ satisfying the conditions in 2.5. Define the norm $||\cdot||$ for a matrix B with elements (b_{ij}) as $||B|| = \sup_i \sum_{j=0}^{\infty} |b_{ij}|$.

Definition 2.1.1 Uniform Ergodicity.

$X(t)$ is uniformly ergodic if there exists a vector $\pi = (\pi_0, \pi_1, \cdots)$, $\pi_i \geq 0$ and $\sum_{j=0}^{\infty} \pi_j = 1$, such that $\lim_{t\to\infty} p_{ij}(s,t) = \pi_j$ uniformly in s for every j and i.

The vector π is called the *long-run distribution* for the chain.

Definition 2.1.2 Strong Uniform Ergodicity.

$X(t)$ is uniformly strongly ergodic if there exists a row constant stochastic matrix L such that

$$\lim_{t\to\infty} ||P(s,t) - L|| = 0$$

uniformly in s.

Strong uniform ergodicity implies uniform ergodicity, and the long-run distribution is any row of the (row-constant) matrix L.

Definition 2.1.3 Weak Ergodicity.

$X(t)$ is weakly ergodic if for all $t \geq 0$ there exists a (stochastic) vector function $Q(t)$ such that if $Q(t) = (q(t), q(t), \ldots)'$, then for all $s \geq 0$,

$$\lim_{t\to\infty} ||P(s,t) - Q(t)|| = 0.$$

It is sufficient to show that the limit is 0 for some $s \geq 0$.

Weak ergodicity is a *loss of memory* property, in the sense that the process forgets the starting distribution as $t \to \infty$. $P(s,t)$ and $Q(t)$ are in some sense close, but $P(s,t)$ and $P(s,t+h)$ need not be, even for large t's. This sense of closeness is implied by strong ergodicity.

Johnson and Isaacson (1988) [30] give sufficient conditions for the strong ergodicity of an inhomogeneous continuous-time Markov chain in terms of the infinitesimal generator, $A(t), t \geq 0$:

Theorem 2.1.1 Strong Ergodicity using A(t).

Suppose for every $t \geq 0$ there exist stochastic vectors π and $\pi(t)$ such that

$\pi(t)A(t) = 0$ and that $\int_0^\infty ||\pi(t) - \pi||dt < \infty$. Then if $\{X(t)\}$ is weakly ergodic, it is strongly ergodic.

This is an important result since finding $P(s,t)$ directly is usually difficult. The authors also cite a result that provides a way of testing for weak ergodicity in terms of $A(t)$. Thus it is possible to answer the question of ergodicity through the infinitesimal generator alone.

The Birth and Death Process

Birth and death processes comprise a large class of continuous-time processes having discrete state spaces. In a birth-death process, $X(t)$ represents the number of members in a population at time t. New members join the population either through birth or immigration; members leave the population by death or emigration. In general, given that $X(t) = i$, transitions in the small time interval Δt satisfy

$$p_{ij}(t,t+\Delta t) = \begin{cases} \lambda(i,t)\Delta t + o(\Delta t) & \text{if } j = i+1 \\ 1 - (\lambda(i,t) + \mu(i,t))\Delta t + o(\Delta) & \text{if } j = i \\ \mu(i,t)\Delta t + o(\Delta t) & \text{if } j = i-1 \\ o(\Delta t) & \text{otherwise.} \end{cases} \tag{2.6}$$

where $\lambda(x,t)$ and $\mu(x,t)$ are the birth and death rates at time t when $X(t) = x$. The model incorporates an immigration process by replacing $\lambda(x,t)$ with $\alpha(t) + \lambda(x,t)$, where $\alpha(t)$ is the rate of immigration.

From equation 2.6, the infinitesimal generator matrix,

$$A(t) = (a_{ij}(t)) \quad i, j \in S$$

is

$$a_{ij}(t) = q_i(t)Q_{ij}(t) = \begin{cases} \mu(i,t) & \text{if } j = i-1, i \geq 1 \\ \lambda(i,t) & \text{if } j = i+1, i \geq 1 \\ -(\lambda(i,t) + \mu(i,t)) & \text{if } j = i, i \geq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

Note that the infinitesimal generator of a birth-death process is *tridiagonal*; this is because transitions in the small time interval Δt to other than a neighboring state $\{i-1, i, i+1\}$ has probability $o(\Delta t)$. The process is inhomogeneous if the birth and death rates are functions of time t. There are many good references for birth and death processes; Bharucha-Reid (1997) [5] is in particular an excellent reference for theory and applications.

Queueing models are examples of birth and death processes with immigration, birth-rates equal to 0, and death rates equal to the rates at which customers are served and leave the system. Queueing models describe systems that provide service to randomly arising demands ($\{X(t)\}$ is the number of customers in the system at time t). These models have broad applications; they emerged from problems in the telecommunications industry beginning in 1908 with A.K. Erlang's pioneering work with the Copenhagen Telephone Company. Recent applications include modeling the demands for internet access and cellular phone connections. Queueing models have applications in other areas: traffic control, financial transactions, biomedical and chemical problems. A queueing model of ATM transactional data was the initial motivation for studying parameter estimation for continuous-time processes discretely observed.

2.1.2 The Dynkin Martingale

The infinitesimal generator of a Markov process, under certain conditions, solves the so-called martingale problem. It is useful here to view the infinitesimal generator as a bounded operator on a space of functions. A solution to the martingale problem is essentially the pair $(A(u), \text{Domain}(A(u)))$, such that for any function of the sample space S in Domain$(A(u))$,

$$f(X(t)) - \int_0^t (A(u)f)(X(u))du$$

is a martingale with respect to $\mathcal{F}_t = \sigma(X(s), 0 \le u \le t)$. Let $P(s,t)$ be the matrix of transition probabilities,

$$p_{ij}(s,t) = P[X(t) = j | X(s) = i] \quad i,j \in S,$$

for the Markov process $X(t)$. For $s \le t$, the integrated forms of the backward and forward Kolmogorov equations are, respectively,

$$P(s,t) = I + \int_s^t P(s,u)A(u)du \quad \text{and} \tag{2.8}$$

$$P(s,t) = I + \int_s^t A(u)P(u,t)du. \tag{2.9}$$

Using the matrix norm defined by

$$\|B\| = \sup_i \sum_{j\in S} |B(i,j)|,$$

assume the condition

$$\sup_{t\geq 0} \|A(t)\| < \infty. \tag{2.10}$$

Condition 2.10 implies that for all states, $i, j \in S$, $a_{ij}(t)$ is bounded for all $t \geq 0$. This implies that there exists a unique solution, $P(s,t)$ to the integral equations 2.8 and 2.9 (see Ethier and Kurtz (1986) [13], pp. 221-222). Let the *test function* $g: S \rightarrow \mathbb{R}^m$, satisfying

$$\|g\| = \sup_{x\in S} |g(x)| < \infty,$$

be such that

$$(A(s)g)(X(s)) = \frac{d}{dt}E[g(X(t))|X(s)]\Big|_{t=s},$$

is defined. That is, g belongs to the domain of the generator, $A(t)$. For example, if S is discrete and assuming $X(s) = x$, then

$$(A(s)g)(x) = \sum_{j\in S} a_{xj}(t)g(j),$$

for any x and j in S. Now for $t \geq 0$, let

$$M(t) = g(X(t)) - \int_0^t (A(s)g)(X(s))ds. \tag{2.11}$$

Then $M(t)$ is a martingale adapted to the sigma-field

$$\mathcal{F}_t = \sigma(X(s) : 0 \leq s \leq t).$$

This follows since

$$
\begin{aligned}
E[M(t)|X(0)=x] &= [P(0,t)M(t)](x) \\
&= [P(0,t)g(X(t))](x) - [P(0,t)\int_0^t (A(u)g)(X(u))du](x) \\
&= [P(0,t)g(X(t))](x) - \int_0^t (P(0,u)A(u)g)(x)du \\
&= g(x)
\end{aligned}
$$

by equation 2.9. Thus, $E[M(t)|X(0)=x] = M(0)$. This result together with the Markov property establishes that $M(t)$ is a martingale adapted to $\mathcal{F}_t$. Then

$$E[M(t)|X(s)=x] = M(s)$$

is the Dynkin Identity and $M(t)$ (2.11) is a Dynkin Martingale. This continuous-time martingale forms the basis of the estimation method explored in this thesis. See Athreya and Kurtz (1973) [2] for generalizations to Dynkin's Identity for time-homogeneous Markov chains. For inhomogeneous processes, Ethier and Kurtz (1986) [13] provide greater detail and see also Pazy (1980) [38] for the theory in the context of two-parameter semigroups.

2.2 Statistical Context

The inference problem in this thesis is to estimate an unknown paramter $\theta \in \Theta \subset \Re^d, d \geq 1$ for an inhomogeneous continuous-time Markov chain $\{X(t), t \geq 0\}$ with state space S. The distribution of $X(t)$ will depend on θ through its infinitesimal generator, $A_\theta(t)$. Thus the model and form of the generator will be specified *a priori*. The estimation problem in this thesis is therefore one of *parametric* inference. The goal in *non-parametric* problems, on the other hand, is to estimate the distribution itself, usually for the purpose of prediction. While this is an important and rich problem for continuous-time processes, it will not be treated here. For a general reference, see Basawa and Rao (1980) [4]. The next Section provides an overview of estimation methods appropriate for the inference problem of this thesis. In particular, these are maximum likelihood and quasi-likelihood methods, generalized method of moments, and martingale estimating equations. These techniques overlap in certain respects; for more information regarding a unifying theory, see Godambe and Heyde (1987) [20].

2.2.1 Parametric Estimation

Parametric methods in general arrive at finding an estimator $\hat{\theta}_n$ which solves some kind of *estimating equation* as in

$$U(\theta; \{x_i\}) = 0. \tag{2.12}$$

where $\{x_i\}$, $i = 1, \ldots, n$ is the sample data. (If θ is a vector, equation 2.12 represents a system of equations.) An estimating equation often results from the minimization or maximization of some objective function. A *score* function is an estimating function $U(\theta; \{x_i\})$ that is the derivative (or gradient) with respect to θ of the objective function.

Suppose $\{x_i\}$ is a sample from the distribution with density, $f(x;\theta)$. The goal is to estimate θ based on the sample $\{x_i\}$. The Maximum Likelihood principle asserts that the value of θ which maximizes the likelihood function

$$L_n(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

is the one most probable based on the observations. The log-likelihood function has the same maximizer (if it exists) and so

$$\log(L_n(\theta)) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

yields the *likelihood score* function,

$$U(\theta;\{x_i\}) = \frac{\partial \log L_n(\theta)}{\partial \theta}.$$

Under regularity conditions, maximum likelihood estimators are asymptotically consistent, normal, and efficient, meaning they have least variance. They serve in parametric inference as the benchmark to any new method that is justifiable in an asymptotic sense.

Methods that generate score functions are known generally as *Quasi-Likelihood* methods. Quasi-likelihood thus includes maximum likelihood as a special case. Another special case is the method of Least Squares. The principle of least squares calls for the estimator that minimizes a sum of squared differences yielding the score function,

$$\begin{aligned} U(\{x_i\},\theta) &= \frac{\partial}{\partial\theta}\sum_{i}^{n}(x_i - h(x_i,\theta))^2 \\ &\propto \sum_{i}^{n} \frac{\partial h(x_i,\theta)}{\partial\theta}(x_i - h(x_i,\theta)). \end{aligned}$$

The function $h(\theta, \{x_i\})$, typically an expectation, represents the model; then $h(\hat{\theta}; \{x_i\})$ is the model "closest" to the data in the least squares (Euclidean distance) sense. Excellent references for Quasi-likelihood include Godambe and Heyde, (1987) [20], and Heyde (1997) [27].

Generalized Method of Moments

Standard method of moments works by equating sample moments to their population counterparts which are functions of the unknown parameter $\theta \in \Theta \subset \Re^d, d \geq 1$. This step constructs a system of *moment equations*. If a model implies d independent moment equations, the solution of the system is the method of moments estimator for θ. When a model generates more than d equations, i.e., when the system is *overdetermined*, the method is generalized from an equation-solving problem to that of a minimization problem. Borrowing from Greene, 1993 [21], Chapters 4 and 13, let $g_k(\cdot)$, $k = 1, \ldots, d$ be any continuous function of the sample space. A sample moment is then

$$\bar{g}_k = \frac{1}{n}\sum_i g_k(x_i), \quad k = 1, 2, \ldots, r.$$

Typically, a law of large numbers gives

$$\bar{g}_k \xrightarrow{P} E_\theta[g_k(x)] = \gamma_k(\theta)$$

from which follows the moment equation,

$$m_k(\theta) = \bar{g}_k - \gamma_k(\theta) = 0, \quad k = 1, 2, \ldots r.$$

where $\gamma_k(\theta)$ represents the population mean, $E_\theta[g_k(x)]$. The corresponding *moment conditions* are that

$$E[m_k(\theta^*)] = 0, \ k = 1, \ldots, r$$

when θ^* is the true parameter value. An identifiability condition is that the moment conditions hold only at the true value, θ^*. If a model gives d functionally independent moment equations (i.e., $r = d$), the method of moments estimator is identifiable as the the vector of the d solutions: $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_d)'$. In the simplest examples, the g_k's are based on powers of x, or on sample central moments as in $g_k(x_i) = (x_i - \bar{x})^k$. Assuming identifiability, method of moments estimators will be consistent though not efficient unless they coincide with maximum likelihood estimators. There are cases in which a model implies more moment equations than there are parameters to be estimated and method of moments estimators cannot be exactly identified. The method of moments is generalized to handle this situation.

If a model implies $r > d$ moment conditions,

$$E[m_j(x_i, \theta)] = 0, \ j = 1, \ldots, r,$$

yielding the r moment equations,

$$\overline{m}_j = \frac{1}{n} \sum_i m_j(x_i, \theta),$$

there are $\binom{r}{d}$ possible sets of method of moments estimators. In such cases, GMM proceeds by choosing θ which minimizes the criterion function,

$$Q(\theta) = \overline{\mathbf{m}}(\theta)' \mathbf{W}^{-1} \overline{\mathbf{m}}(\theta).$$

where

$$\overline{\mathbf{m}}(\theta) = (\overline{m}_1(\theta), \overline{m}_2(\theta), \ldots, \overline{m}_r(\theta))'.$$

and $\mathbf{W}$ is some appropriate weighting matrix. The minimizer of Q is the GMM estimator. If the underlying process is stationary and ergodic, and if other technical conditions hold (Hansen (1982) [23]), the GMM estimator $\hat{\theta}$ converges almost surely to θ^* and $\hat{\theta}$ is asymptotically normal $N(\theta^*, \mathbf{\Sigma})$ as $n \to \infty$, where

$$\mathbf{\Sigma} = [\mathbf{G}'\mathbf{W}^{-1}\mathbf{G}]^{-1}.$$

$\mathbf{W}$ is the asymptotic variance of $\overline{\mathbf{m}}(\theta)$, and $\mathbf{G}$ is a matrix of partial derivatives whose j^{th} row is

$$\mathbf{G}^j = \mathbf{G}^j(\theta^*) = \frac{\partial \overline{m}_j(\theta^*)}{\partial \theta'}.$$

Hansen [23] also establishes the conditions under which these results hold with consistent estimators for $\mathbf{W}$ and $\mathbf{G}$.

Martingale Estimating Equations

Another approach in parametric inference is to find estimating equations that are mean 0 martingales. If $M_n(\theta)$ is a mean 0 martingale with respect to the filtration $\{\mathcal{F}_n\}$ generated by the observations, then the *martingale difference*,

$$D_i(\theta) = M_i(\theta) - M_{i-1}(\theta)$$

has 0 mean and is measurable with respect to $\mathcal{F}_i$. That is, for any $k \geq 1$,

$$E[M_{n+k}(\theta)|\mathcal{F}_n] = M_n(\theta)$$

and $E[M_0(\theta)] = 0$ imply that $E[D_i(\theta)] = 0$. Moreover, since $E[D_i(\theta)|\mathcal{F}_{i-1}] = 0$, the D_i are uncorrelated: for $i < j$,

$$\begin{aligned} E[D_j D_i] &= E[E[D_j D_i|\mathcal{F}_{i-1}]] \\ &= E[D_j E[D_i|\mathcal{F}_{i-1}]] \\ &= 0. \end{aligned}$$

A *martingale estimating function* is then given by

$$U(\theta; \{x_i\}) = M_n(\theta) = \sum_{i=1}^{n} D_i(\theta),$$

with $E[M_n(\theta)] = 0$, and

$$Var[M_n(\theta)] = E[M_n^2(\theta)] = \sum_{i=1}^{n} E[D_i^2(\theta)].$$

As usual, the estimator based on $M_n(\theta)$ solves the estimating equation,

$$M_n(\theta) = 0.$$

More generally, for the martingale differences, $\{D_i(\theta)\}$, and $\mathcal{F}_{i-1}$ -measurable random variables, a_{i-1}, the estimating equation

$$M_n(\theta) = \sum_{i=1}^{n} a_{i-1} D_i(\theta) = 0 \tag{2.13}$$

has solution $\hat{\theta}_n$, which satisfies, under regularity conditions,

$$\frac{\sum_{i=1}^{n} a_{i-1} E[D_i'(\theta)|\mathcal{F}_{i-1}]}{\sqrt{\sum_{i=1}^{n} a_{i-1}^2 Var[D_i(\theta)|\mathcal{F}_{i-1}]}}(\hat{\theta}_n - \theta) \to N(0,1) \quad \text{as} \quad n \to \infty. \tag{2.14}$$

where $D_i'(\theta) = \frac{\partial}{\partial\theta} D_i(\theta)$. Godambe (1985) [18] establishes that 2.13 is optimal when the weight variables are set to

$$a_{i-1}^* = \frac{E[D_i'(\theta)|\mathcal{F}_{i-1}]}{E[D_i^2|\mathcal{F}_{i-1}]}.$$

That is, among all estimating functions constructed as in 2.13,

$$M_n^*(\theta) = \sum_{i=1}^{n} a_{i-1}^* D_i(\theta)$$

minimizes the ratio

$$\frac{E[M_n^2(\theta)]}{E[M_n'(\theta)]^2}. \tag{2.15}$$

which can be viewed as a measure of precision for a sample of size n. The criterion to minimize 2.15 asks simultaneously that $M_n(\theta)$ be as close as possible to 0 and that $M_n(\theta + \delta)$ be as large as possible (Godambe (1960) [17]).

2.3 Sampling from Continuous-Time Markov Chains

Estimation techniques for continuous-time Markov chains must take into account the manner by which the process is observed. Asymptotic properties in particular depend on how a sample is allowed to grow large. Methods in this context usually consider asymptotic properties as the length of the observation interval $T \to \infty$, although there are other situations. For instance, the observation interval can be fixed while the number of independent realizations over the interval increases. See Basawa and Rao (1980) [4] for further discussion. Given an observation interval, the process is observed continuously or discretely, either at fixed intervals or at randomly determined observation times. If the process is observed continuously, the likelihood function can be written in terms of the infinitesimal generator of the process; otherwise, it requires transition probabilities and other techniques are generally sought.

Continuous Observation and Maximum Likelihood

If the process is observed continuously on some interval $[T_0, T]$, so that every state and transition time throughout the interval is known, then the likelihood function is expressible in terms of the infinitesimal generator and maximum likelihood is straightforward. Under regularity conditions, maximum likelihood estimators attain optimal properties asymptotically as $T \to \infty$.

For a sample taken continuously on some observation period $[T_0, T]$, the inference problem essentially reduces to that of a discrete time process because the sample is a realization of the *imbedded* Markov chain. The imbedded Markov chain is the process $\{(z_k), k = 0, 1, \ldots, J(T)\}$, where

$$z_k = X(t) \text{ for } T_k \leq t < T_k + \tau_k \tag{2.16}$$

and $J(T)$ is the number of transitions made by the process by time T. T_k denotes the time of the kth transition (or jump) and τ_k is the length of time (or holding time) the process is in state z_k. Therefore,

$$T_k = \inf\{t \geq 0 : X(t) = z_k\} = T_0 + \tau_0 + \tau_1 + \ldots + \tau_{k-1}.$$

Then, as derived in Billingsley (1961) [7], the likelihood based on the sample,

$$\{(z_i, \tau_i), \quad i = 0, \ldots, J(T) - 1\},$$

for a (homogeneous) Markov chain is

$$L_{J(T)}(\theta) = \prod_{k=0}^{J(T)-1} q_\theta(z_k) Q_\theta(z_k, z_{k+1}) \exp^{-q_\theta(z_k)\tau_k}, \tag{2.17}$$

where q and Q are the transition intensity and relative transition intensity functions, respectively. Note that since $J(T)$ is the last jump observed on $[0, T]$, $\tau_{J(T)}$ is unknown. The log-likelihood is

$$\sum_{k=0}^{J(T)-1} \log(q_\theta(z_k) Q_\theta(z_k, z_{k+1})) - \sum_{k=0}^{J(T)-1} q_\theta(z_k)\tau_k. \tag{2.18}$$

Billingsley (1961) [7] Section II gives conditions necessary for maximum likelihood estimation in the case of homogeneous Markov chains. Under appropriate conditions, MLEs from 2.18 are strongly consistent, asymptotically normal, and asymptotically efficient as $T \to \infty$. Basawa and Rao (1980) [4] review maximum likelihood when n independent sample paths of the process are observed continuously over $[T_0, T]$; then asymptotic properties are as $n \to \infty$ with T fixed.

The inhomogeneous analog to 2.17 is

$$L_{J(T)}(\theta) = \prod_{k=0}^{J(T)-1} q_\theta(z_k, T_{k+1}) Q_\theta(z_k, z_{k+1}, T_{k+1}) \exp^{-\int_0^{\tau_k} q_\theta(z_k, s)ds}.$$

In terms of the infinitesimal generator $A(t) = (a_{ij}(t))$ defined in 2.2, this is seen to be

$$L_{J(T)}(\theta) = \prod_{k=0}^{J(T)-1} a(z_k, z_{k+1}, T_{k+1}) \exp^{\int_0^{\tau_k} a(z_k, z_k, s)ds}.$$

Observation at Fixed Intervals

In practice, it is common that a continuous-time process is observed at discrete, fixed interval time points. Samples look like $\{(z_k, t_k), k = 0, \ldots, n\}$, where

$$\begin{aligned} z_k &= X(t_k) \quad \text{and} \\ t_k &= kh, \end{aligned}$$

for some fixed positive value, h. These samples are fragmentary in the sense that they do not provide full information about the evolution of the process over the period of observation. Since the $\{t_k\}$ are not the jump (trasition) times of the process, the likelihood is inexpressible in terms of the infinitesimal generator. In this case, the likelihood is

$$\prod_{k=0}^{n-1} P_\vartheta(t_k, z_k; t_{k+1}, z_{k+1})$$

and the transition probabilities must be known. Finding the transition probabilities explicitly requires solving the system

$$\begin{aligned} \frac{\partial}{\partial s} P(s,t) &= -A(s)P(s,t) \\ P(s,s) &= I \end{aligned}$$

for each $s = t_k, t = t_{k+1}$ and all $i, j \in S$. Thus obtaining maximum likelihood estimators in this case presents a huge computational burden, especiall when S is countably infinite. Even in the homogeneous case, maximum likelihood based on fixed samples requires additional information or simplifying assumptions. For example, Keiding (1975) [32] derives maximum likelihood estimates of functionals of the birth and death parameters in the fixed sampling case by assuming that, in addition to knowing $X_{kh} = z_k$, one also knows a quantity C_k which equals the number of particles among $X_{(k-1)h} = z_{k+1}$ that have 0 offspring. Usually, however, the problem calls for other estimating functions that balance the loss in efficiency with computational feasibility.

Fixed Observation Intervals and Aliasing

Any method based on fixed sampling must contend with the so-called *aliasing* problem, which arises when a sample represents more than one underlying process. It is generally a difficult task to resolve the problem so that the true process is identifiable. One way to alleviate it is to take the interval size h sufficiently small, thereby reducing the class of candidate processes. Occasionally it is possible to put restrictions on the model. Phillips (1973) [39] brings the problem to light in the context of estimating the spectral density of a continuous-time process observed at fixed intervals, and provides conditions on the structural (parameter) matrix of a first-order linear time-series model that ensure identification. Hansen and Sargent (1983) [24] provide related work for covariance stationary models. Singer and Spilerman (1976) [43] provide a detailed account of the aliasing problem for homogeneous Markov processes with finite state spaces. Alternatively, other authors obtain identification results under certain random observation schemes (Shapiro and Silverman (1960) [41]. Duffie and Glynn (1996) [11]).

Observation at Random Times

In the general view, a process randomly observed is observed at time points

$$T_i = \inf\{t : N(t) = i\}, \quad i = 1, \ldots, n,$$

where T_i denotes the i^{th} event time of *doubly stochastic* point process with time-varying intensity $\{\beta(X(t)) : t \geq 0\}$. The sample is the collection, $\{(Z_i, T_i)\}$, where $Z_i = X(T_i)$ for $0 \leq i \leq n$, taking $T_0 = 0$. The technique is called Poisson sampling when $N(t)$ is a Poisson process. Note that since the observation times do not coincide with the jump times of the process, the likelihood function requires the transition probabilities $(P[X(T_k) = Z_k | X(T_{k-1}) = Z_{k-1}])$ which are generally unknown or difficult to find. Discrete samples based on random observation, as in the fixed observation case, usually call for methods other than a direct likelihood approach.

The next section provides a review of a number of papers on estimation for stochastic processes that were consulted for this research and which are not necessarily referenced elsewhere in this dissertation. The organization is by estimation method.

2.4 Literature Review

2.4.1 Maximum Likelihood

Two early papers serve as background papers on the method of maximum likelihood for continuous-time processes. Keiding (1975) [32] derives maximum likelihood estimates for a (homogeneous) birth and death process observed continuously. He discusses the estimation problem when observations are obtained discretely: for such samples he uses the term *discrete skeleton*. An earlier work is Darwin [10], who discusses the discrete skeleton and under which circumstances maximum likelihood estimates of a linear birth and death process can be obtained.

Baba [3] considers maximum likelihood estimation of (homogeneous) birth and death processes by Poisson sampling. The Poisson sampling scheme in his paper, however, differs from the scheme considered in this work. His assumes that the number of transitions of states in the generated observation intervals, (T_k, T_{k+1}), is known, whereas in the version considered here, only the state at the observation time is known. The author applies the method to a homogeneous $M/M/1$ queueing model.

2.4.2 Conditional Least Squares and Quasi-Likelihood

Klimko and Nelson (1978) [33], derive *Conditional* Least Squares (CLS) estimators for discrete-time processes. Their estimating function is of the form

$$U(\theta; \{x_i\}) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n-1}(x_{i+1} - E^{\theta}_{x_i}(x_{i+1}))^2. \tag{2.19}$$

where $E^{\theta}_{x_i}[x_{i+1}] = E^{\theta}[x_{i+1}|x_i]$. Thus the CLS estimator minimizes the sum of squared differences between the observations and their expectations conditioned on the previous step. Though intended for discrete-time processes, it is relatively straightforward to apply CLS in the case of discretely sampled continuous-time processes. One complication is that the conditional expectation, $E^{\theta}[z_{i+1}|z_i]$ for $i = 1, \ldots, n-1$, may be difficult to compute. It is, however, generally easier to work with conditional expectations than it is to find transition probabilities necessary for the likelihood. Ensor and Glynn (1996) [12] work out the asymptotic results specifically for the continuous-time Markov process and in addition suggest a grid-based simulation method for computing conditional expectations.

Hutton and Nelson (1986) [28] develop quasi-likelihood estimation for samples taken under continuous observation. They construct a quasi-likelihood objective function based on a semi-martingale representation of the process. Their method applies to a broad class of processes including diffusions (Markov processes with continuous state space); furthermore, they do not require that the process be homogeneous, ergodic, or even Markov.

Hutton, Ogunyemi, and Nelson (1991) [29] propose a modified quasi-likelihood estimator using a simplified covariance matrix. They provide conditions for consistency and asymptotic normality. The authors apply the method to a branching process with immigration. The presentation is for discrete-time processes but results extend to continuous-time in conjunction with the work in [28]. Broader references for quasi-likelihood include the papers by Godambe and Heyde (1987) [20] and Heyde (1993) [26]. The text by Heyde (1997) [27] is comprehensive and should be consulted first for the general theory of quasi-likelihood.

2.4.3 Generalized Method of Moments

Hansen [23] studies large sample properties of Generalized Method of Moments (GMM) estimators. He establishes strong consistency and asymptotic normality under stationarity and ergodicity assumptions. This paper provides an important theoretical foundation for GMMs.

Hansen and Scheinkman (1995) [25] derive a class of GMM estimators for time-homogeneous, reversible Markov processes observed at fixed intervals. Their moment conditions involve the infinitesimal generator of the process and the generator of the associated reverse-time process. They establish conditions on the generators ensuring that their estimators attain the asymptotic properties of GMM estimators. In view of the aliasing problem, they conclude that their moment conditions based on fixed samples cannot distinguish between generators admitting the same stationary distributions except possibly for reversible processes.

As a follow-up to Hansen and Scheinkman, Shoji and Ozaki (1997) [42] compare the Hansen and Scheinkman estimators to other methods applied to diffusions sampled at fixed intervals. For time-homogeneous ergodic processes, Duffie and Glynn (1996) [11] derive GMM estimators based on randomly observed samples. Their idea is to first choose a test function $g : (\Theta \times S \times S) \mapsto \Re^m$ and then to define the function f via

$$f(\theta, x, y) = g(\theta, x, y) - \frac{A_\theta g^{(\theta, x)}(y)}{\beta(\theta, y)}, \quad (\theta, x, y) \in \Theta \times S \times S, \tag{2.20}$$

where

$$A_\theta g^{(\theta, x)}(y) = \frac{d}{dt} E_y[g(X(t))]_{|t=0^+} \quad x, y \in S, \tag{2.21}$$

E_x denoting expectation associated with a given initial condition $x \in S$. The basic moment condition is given by

$$E_\pi[f(\theta^*, Z_i, Z_{i+1}) - g(\theta^*, Z_i, Z_i)] = 0, \tag{2.22}$$

where π is the invariant probability measure of the sample process (Z_i, T_i). They also show that the generator is fully identified by their moment conditions for a constant sampling rate bounded away from zero, and for time-dependent sampling rates bounded away from zero as long as observation times are observable.

2.4.4 Martingale Estimating Functions

On the general theory of martingale-based estimating functions the main references overlap a good deal with the quasi-likelihood literature, namely Godambe (1985) [18], and Godambe and Heyde (1987) [20]. Lloyd (1987) [36] considers optimal martingale estimating equations derived in the presence of nuisance or "accessory" parameters. The collection edited by Godambe (1991) [19] includes many useful papers. Among them, Lloyd and Yip (1991) [35] derive estimating functions to estimate the size of the population in capture and recapture experiments. The authors provide an excellent overview of martingale-based estimation, apply it to a specific application, and discuss bias reduction and more complex models.

Chapter 3

Estimating Equation Based on a Dynkin Martingale

The Dynkin martingale motivates an estimating equation in the following way: if in fact a sample analog of the Dynkin martingale based on discrete observations inherits the martingale property, then it falls within the class of martingale estimating equations introduced in Chapter 2. In this way, the inference problem for a continuous-time process is cast in the context of discrete-time, martingale-based estimation procedures. This immediately outlines a strategy for proving asymptotic properties for the Dynkin martingale estimators. This strategy is to mimic the arguments already established for discrete-time martingale estimating equations.

This chapter first establishes the martingale property for the sample analog when discrete samples are taken via Poisson sampling. Section 3.2 contains a review of martingale estimating equations and Section 3.3 presents a discussion of the asymptotic properties of martingale estimating functions and how these properties translate to the estimator itself. Finally, conditions and implications for the Dynkin martingale estimator are summarized in Section 3.4.

Recall the inference problem in this thesis: $\{X(t), t \geq 0\}$ is an inhomogeneous Markov chain with (discrete) state space S and infinitesimal generator, $A_\theta(t)$. The underlying probability space is $\{\Omega, \mathcal{F}, P_\theta\}$, where $\mathcal{F}$ is the collection of Borel sets of

Ω. The purpose here is to estimate the unknown parameter $\theta \in \Theta \subset \mathbb{R}^d$, for $d \geq 1$. Suppose the generator satisfies

$$\sup_{t \geq 0} \|A_\theta(t)\| < \infty \tag{3.1}$$

(the matrix norm $\|B\| = \sup_i \sum_{j \in S} |B(i,j)|$) and that the real-valued function g defined on the sample space S satisfies

$$\|g\| = \sup_{x \in S} |g(x)| < \infty. \tag{3.2}$$

Then

$$M(t) = g(X(t)) - \int_0^t (A_\theta(u)g)(X(u))du \tag{3.3}$$

is a (Dynkin) martingale adapted to the filtration $\mathcal{F}_t = \sigma(X(u) : 0 \leq u \leq t)$. That is,

$$E[M(t)|\mathcal{F}_s] = M(s).$$

That $M(t)$ is a martingale follows directly from the integrated version of the forward equations as shown in Section 2.1.2. The condition on A_θ ensures that a unique process exists, and the condition on g ensures integrability.

3.1 Sample Analog

If the process cannot be observed for a continuous time period, it would appear that the martingale result 3.3 is not useful in constructing an estimating equation. However, if the process can be observed at random observation times, independently of the process, the martingale in 3.3 implies a sample analog. This is shown below when the process is sampled according to a homogeneous Poisson process independently of $X(t)$.

The process is observed at random times according to an independent Poisson process $N(t)$ with constant rate parameter β. Let $T_1, T_2, \ldots$ be the event times of this observation process. For a sample of size n,

$$T_i = \inf\{t : N(t) = i\} \text{ for } i = 1, 2, \ldots, n.$$

Let $\tau_i = T_i - T_{i-1}$ indicate the iid inter-arrival times, which are exponential with mean $1/\beta$. Therefore with $Z_i = X(T_i)$, the set

$$\{(Z_i, T_i), i = 1, \ldots, n\}$$

fully describes the sample. Let

$$\mathcal{F}_i = \sigma(T_i, X(s) : 0 \leq s \leq T_i).$$

That is, $\mathcal{F}_i$ is the sigma field generated by $X(t)$ on $[0, T_i]$. Suppose the function g satisfies condition 3.2, and consider the expectation

$$E[g(Z_{i+1})|\mathcal{F}_i] = E[g(Z_{i+1})|(Z_i, T_i)],$$

using the Markov property. From the martingale result 3.3, this is

$$\begin{aligned} E[g(Z_{i+1})|(Z_i, T_i)] &= M(T_i) + E[\int_0^{T_{i+1}} (A_\theta(u)g)(X(u))du|(Z_i, T_i)] \\ &= g(Z_i) - \int_0^{T_i} (A_\theta(u)g)(X(u))du \end{aligned}$$

$$
\begin{aligned}
& +E[\int_0^{T_{i+1}} (A_\theta(u)g)(X(u))du|(Z_i, T_i)] \\
= \;& g(Z_i) + E[\int_{T_i}^{T_{i+1}} (A_\theta(u)g)(X(u))du|(Z_i, T_i)] \\
= \;& g(Z_i) + E[\int_0^{\tau_{i+1}} (A_\theta(T_i + s)g)(X(T_i + s))ds|(Z_i, T_i)].
\end{aligned}
$$

Letting I_B denote the indicator of the set B, the expectation in the second term is equivalent to

$$
\begin{aligned}
& E[\int_0^\infty I_{\{\tau_{i+1}>s\}}(A_\theta(T_i + s)g)(X(T_i + s))ds|(Z_i, T_i)] \\
= \;& \int_0^\infty E[I_{\{\tau_{i+1}>s\}}(A_\theta(T_i + s)g)(X(T_i + s))ds|(Z_i, T_i)] \\
= \;& \int_0^\infty E[I_{\{\tau_{i+1}>s\}}]E[(A_\theta(T_i + s)g)(X(T_i + s))|(Z_i, T_i)]ds \\
= \;& \int_0^\infty Pr(\tau_{i+1} > s)E[(A_\theta(T_i + s)g)(X(T_i + s))|(Z_i, T_i)]ds. \qquad (3.4)
\end{aligned}
$$

using the independence of the inter-arrival times, τ_i. The τ_i follow an exponential distribution with mean $1/\beta$, so that expression 3.4

$$
\begin{aligned}
= \;& \frac{1}{\beta}\int_0^\infty \beta e^{-\beta s}E[(A_\theta(T_i + s)g)(X(T_i + s))|(Z_i, T_i)]ds \\
= \;& \frac{1}{\beta}E[(A_\theta(T_i + \tau_{i+1})g)(X(T_i + \tau_{i+1}))|(Z_i, T_i)] \\
= \;& \frac{1}{\beta}E[(A_\theta(T_{i+1})g)(X(T_{i+1}))|(Z_i, T_i)].
\end{aligned}
$$

Finally,

$$
E[g(Z_{i+1})|(Z_i, T_i)] = g(Z_i) + \frac{1}{\beta}E[(A_\theta(T_{i+1})g)(Z_{i+1})|(Z_i, T_i)].
$$

Therefore, a martingale is given by

$$
M_n(\theta) = \sum_{i=1}^n D_i(\theta).
$$

where the

$$D_i(\theta) = g(Z_i) - g(Z_{i-1}) - \frac{1}{\beta}(A_\theta(T_i)g)(Z_i) \tag{3.5}$$

are martingale differences with respect to the filtration

$$\mathcal{F}_{i-1} = \sigma(T_{i-1}, X(s) : 0 \leq s \leq T_{i-1}).$$

Using 3.5, the estimating function satisfies

$$M_n(\theta) = \sum_{i=1}^{n} D_i(\theta) = g(Z_n) - g(Z_0) - \frac{1}{\beta} = \sum_{i=1}^{n}(A_\theta(T_i)g)(Z_i)$$

and has these properties:

i. $M_n(\theta)$ is measurable $\mathcal{F}_n$.

ii. $E[|M_n(\theta)|] < \infty$.

iii. $E[M_n(\theta)|\mathcal{F}_m] = M_m(\theta)$, $\quad 0 \leq m \leq n$.

It follows that $E[E[M_n(\theta)|\mathcal{F}_m]] = E[M_m(\theta)] = E[M_0(\theta)] = 0$, so that $M_n(\theta)$ is a mean 0 martingale. For the true value θ^*, the moment condition

$$E[M_n(\theta^*)] = 0 \text{ for all } n \geq 0$$

implies the estimating equation

$$M_n(\theta) = 0,$$

so that the Dynkin Martingale estimator (DME) $\hat{\theta}_n$ satisfies

$$M_n(\hat{\theta}_n) = g(X(T_n)) - g(X(T_0)) - \frac{1}{\beta}\sum_{i=1}^{n}(A_{\hat{\theta}_n}(T_i)g)(X(T_i)) = 0.$$

The work that follows considers only scalar parameters θ, but it is straightforward to define the DME for a vector valued parameter: For $\theta \in \Theta \in \mathbb{R}^d$, $d > 1$, choose test functions $g_1, g_2, \ldots, g_d$, and then compute $\hat{\theta}_n$ as the root of

$$g_k(X(T_n)) - g_k(X(T_0)) - \frac{1}{\beta}\sum_{i=1}^{n}(A_\theta(T_i)g_k)(X(T_i))$$

by component, $k = 1, 2, \ldots d$. The vector case adds (at least) the additional concern of how to choose multiple test functions.

3.2 Martingale Estimating Equations

The goal of this thesis is to reduce the problem of inference for an inhomogeneous continuous-time Markov chain to that of a discrete-time process. The construction of the sample version of the Dynkin Martingale accomplishes this in part as it provides a martingale estimating equation for θ based on discrete samples. This is appealing since much of the asymptotic theory for discrete-time martingale estimating functions has been worked out. This section reviews the established theory. Recall from Section 2.2.1 that for martingale differences $D_i(\theta)$ and $\mathcal{F}_{i-1}$ measurable random variables a_{i-1}, the martingale estimating equation

$$M_n(\theta) = \sum_{i=1}^{n} a_{i-1} D_i(\theta) = 0$$

has solution $\hat{\theta}_n$ satisfying (under regularity conditions)

$$G_n(\hat{\theta}_n - \theta) \to N(0, 1).$$

where

$$G_n = \frac{\sum_{i=1}^n a_{i-1} E[D_i'(\theta)|\mathcal{F}_{i-1}]}{\sqrt{\sum_{i=1}^n a_{i-1}^2 Var[D_i(\theta)|\mathcal{F}_{i-1}]}}.$$

For example, if the $a_{i-1} \equiv 1$, then

$$G_n = \frac{K_n(\theta)}{\sqrt{I_n(\theta)}}.$$

where

$$K_n(\theta) = \sum_{i=1}^n E[D_i'(\theta)|\mathcal{F}_{i-1}]$$

and

$$I_n(\theta) = \sum_{i=1}^n E[D_i^2(\theta)|\mathcal{F}_{i-1}].$$

Alternatively, if $a_{i-1} = a_{i-1}^* = E[D_i'(\theta)|\mathcal{F}_{i-1}]/E[D_i^2(\theta)|\mathcal{F}_{i-1}]$, the optimal weights of Godambe, then

$$G_n = \left(\sum_{i=1}^n \frac{E^2[D_i'(\theta)|\mathcal{F}_{i-1}]}{E[D_i^2(\theta)|\mathcal{F}_{i-1}]}\right)^{1/2}.$$

Since $I_n(\theta)$ and $K_n(\theta)$ involve conditional expectations and are most likely unattainable in closed form, in practice the optimally weighted estimating equation is likely forsaken for the more tractable, unweighted version. However, one could try to find reasonable estimates of the optimal weights and investigate how the estimates improve over those based on an unweighted version.

3.3 Establishing Asymptotic Properties

A fundamental idea in the study of estimating equations is that properties of estimators derive from the estimating functions that generate them. To illustrate, suppose that θ^* is the true value of the parameter and consider a first order Taylor series expansion for $M_n(\theta^*)$ about $\theta' \in \Theta$:

$$\begin{aligned} M_n(\theta^*) &= M_n(\theta') + (\theta^* - \theta')\frac{\partial}{\partial\theta}M_n(\tilde{\theta}) \\ &= M_n(\theta') + (\theta^* - \theta')J_n(\tilde{\theta}) \\ &= M_n(\theta') + (\theta^* - \theta')[J_n(\tilde{\theta}) + I_n(\theta^*) - I_n(\theta^*))] \\ &= M_n(\theta') - (\theta^* - \theta')I_n(\theta^*) + (\theta^* - \theta')[J_n(\tilde{\theta}) + I_n(\theta^*)] \end{aligned}$$

where $\tilde{\theta} \in (\theta', \theta^*)$ and $I_n(\theta)$ is a quantity to be determined. Then

$$I_n^{-1}(\theta^*)M_n(\theta^*) = I_n^{-1}(\theta^*)M_n(\theta') - (\theta^* - \theta') + (\theta^* - \theta')I_n^{-1}(\theta^*)[J_n(\tilde{\theta}) + I_n(\theta^*)].$$

If $\theta' = \hat{\theta}_n$ where $M_n(\hat{\theta}_n) = 0$, this is

$$I_n^{-1}(\theta^*)M_n(\theta^*) = (\hat{\theta}_n - \theta^*) + (\theta^* - \hat{\theta}_n)I_n^{-1}(\theta^*)[J_n(\tilde{\theta}) + I_n(\theta^*)]. \tag{3.6}$$

A strong law of large numbers for $M_n(\theta^*)$ with the norming $I_n(\theta^*)$, as in

$$I_n^{-1}(\theta^*)M_n(\theta^*) \to 0 \text{ a.s.} \quad \text{as } n \to \infty,$$

and the condition that

$$\limsup_{n\to\infty}[I_n(\theta^*)]^{-1}|I_n(\theta^*) + J_n(\tilde{\theta})| < 1 \text{ a.s.} \tag{3.7}$$

would then imply that $\hat{\theta}_n$ is strongly consistent for θ^* ($\hat{\theta}_n \to \theta^*$ a.s.). Furthermore, if a central limit theorem (CLT) holds for $M_n(\theta^*)$, as in

$$I_n^{-1/2}(\theta^*)M_n(\theta^*) \xrightarrow{\mathcal{D}} N(0, 1) \quad \text{as } n \to \infty. \tag{3.8}$$

then

$$I_n^{1/2}(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{D}} N(0,1) \tag{3.9}$$

as long as

$$J_n(\tilde{\theta})^{-1}J_n(\theta^*) \rightarrow 1 \text{ and} \tag{3.10}$$

$$J_n(\theta^*)/I_n(\theta^*) \rightarrow -1 \text{ in probability, as } n \rightarrow \infty. \tag{3.11}$$

This parallels the argument in Hall and Heyde (1980) [22] for the maximum likelihood estimator, for which $D_i(\theta) = \frac{d}{d\theta}[\log L_i(\theta) - \log L_{i-1}(\theta)]$ and

$$I_n(\theta) = \sum_{i=1}^{n} E_\theta[D_i^2(\theta)|\mathcal{F}_{i-1}].$$

In Chapter 6 of [22], the authors give conditions ensuring the CLT result 3.8 for the log-likelihood score function. Note that if condition 3.11 does not hold, then (assuming conditions 3.10 and 3.8), the result in 3.9 generalizes to

$$I_n^{1/2}(\theta^*)J_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{D}} N(0,1). \tag{3.12}$$

In summary, the asymptotic properties of the $\hat{\theta}_n$ derive from the asymptotic behavior of the estimating function, $M_n(\theta)$. For consistency and asymptotic normality, the challenge is first to verify a strong law and a CLT for $M_n(\theta)$. The purpose of the next section is to sketch out the arguments for doing this.

3.3.1 Strong Law of Large Numbers and Central Limit Theorem

This section first treats consistency by reviewing some well-known convergence results for martingales and martingale differences which lead to the strong law of large numbers (SLLN) for martingales. Then the martingale CLT is presented. The discussion borrows from Glynn [16] and Hall and Heyde (1980) [22], Sections 1 and 2.

Consistency

The first result, the Martingale Convergence Theorem, states that, for instance, a square-integrable martingale has an almost sure limit as $n \to \infty$.

Theorem 3.3.1 Martingale Convergence Theorem.

Let $M_n, n \geq 0$ be a martingale with respect to $\mathcal{F}_n$. If there exists a $p > 1$ such that

$$\sup_{n \geq 1} E|M_n|^p < \infty,$$

then there exists M_∞ such that $M_n \to M_\infty$ a.s. as $n \to \infty$.

Given the form of the estimating equation, $M_n(\theta) = 0$, if the a.s. limit M_∞ is 0, then strong consistency holds for $M_n(\theta)$. Kronecker's Lemma can be used to establish that the limit is 0.

Theorem 3.3.2 Kronecker's Lemma.

Let $(x_n, n \geq 1)$ and $(a_n, n \geq 1)$ be two real-valued sequences. Suppose $a_k > 0$ and $a_k \uparrow \infty$. Let $B_n = \sum_{j=1}^{n} x_j / a_j$. If $B_n \to B < \infty$ as $n \to \infty$, then

$$\frac{1}{a_n} \sum_{j=1}^{n} x_j \to 0.$$

If the a_j are deterministic, then

$$\tilde{M}_n = \sum_{j=1}^{n} D_j/a_j$$

has the martingale property. If there exists a $p > 1$ such that

$$\sup_{n\geq 1} E[|\sum_{j=1}^{n} D_j/a_j|^p] < \infty$$

then by the martingale convergence theorem, $\tilde{M}_n$ has an $a.s.$ limit, $\tilde{M}_\infty$. Now set $p = 2$: since $E[D_i] = 0$ and $E[D_iD_j] = 0$, $j \neq i$,

$$E[|\sum_{j=1}^{n} D_j/a_j|^2] = \sum_{j=1}^{n} E[D_j^2/a_j^2]$$

and so, if $a_n \uparrow \infty$, $\sum_{i=1}^{\infty} E[D_j^2]/a_j^2 < \infty$, then

$$\frac{M_n}{a_n} \to 0 \ a.s. \text{ as } n \to \infty.$$

A similar argument proves the next theorem (cf. Theorem 2.18 of Hall and Heyde (1980) [22]).

Theorem 3.3.3 Martingale Strong Law with Random Norming. Let $\{M_n = \sum_{i=1}^{n} D_i, \mathcal{F}_n\}$ be a martingale and $\{U_n, n \geq 1\}$ a non-decreasing sequence of positive random variables such that U_n is measurable $\mathcal{F}_{n-1}$ for each $n \geq 1$. Then for $1 \leq p \leq 2$,

$$\sum_{i=1}^{\infty} U_i^{-1} D_i \text{ converges } a.s.$$

on the set $\{\sum_{i=1}^{\infty} U_i^{-p} E(|D_i|^p|\mathcal{F}_{i-1}) < \infty\}$ and

$$\lim_{n\to\infty} U_n^{-1} M_n = 0 \ a.s.$$

on the set

$$\{\lim_{n\to\infty} U_n = \infty, \quad \sum_{i=1}^{\infty} U_i^{-p} E(|D_i|^p|\mathcal{F}_{i-1}) < \infty\}.$$

To see how Theorem 3.3.3 applies to $M_n(\theta) = \sum_{i=1}^{n} D_i(\theta)$, set $U_n = I_n$ where

$$I_n(\theta) = \sum_{i=1}^{n} E[D_i^2(\theta)|\mathcal{F}_{i-1}].$$

Then if $I_n(\theta^*) \to \infty$ a.s.,

$$\sum_{1}^{\infty}[I_n(\theta^*)]^{-2} E(D_n^2(\theta^*)|\mathcal{F}_{n-1}) < \infty \text{ a.s.} \tag{3.13}$$

and the martingale SLLN gives

$$[I_n(\theta^*)]^{-1} \sum_{i=1}^{n} D_i(\theta^*) \xrightarrow{a.s.} 0. \tag{3.14}$$

This again parallels the argument in Hall and Heyde(1980) [22] for the MLE. The added condition 3.7 involving $I_n(\theta)$ and $J_n(\theta)$ would then ensure the strong consistency of the estimator $\hat{\theta}_n$.

Alternatively, Heyde (1997) [27], Chapter 12, provides a criterion for consistency in terms of a sequence of estimating functions $\{M_n(\theta)\}$ which are not necessarily the derivative with respect to θ of a scalar objective function. Rewritten in the current notation, this is:

Theorem 3.3.4 Consistency Criterion. Let $\{M_n(\theta)\}$ be a sequence of estimating functions that are continuous in θ a.e. on $E \subset \Omega$ for $n \geq 1$. If for all $\delta > 0$ a.e. on E there exists an $\epsilon > 0$ so that

$$\limsup_{n\to\infty}\left(\sup_{\|\theta-\theta^*\|=\delta}(\theta-\theta^*)' M_n(\theta)\right) < -\epsilon,$$

then there exists a sequence of estimators $\hat{\theta}_n$ such that for any $\omega \in E$, $\hat{\theta}_n \to \theta^*$ and $M_n(\hat{\theta}_n(\omega)) = 0$ when $n > N_\omega$.

Heyde admits some difficulty in checking this criterion in practice, particularly in the case when θ is a vector. Theorem 12.1 of Hutton, Ogunyemi, and Nelson (1991) [29] provides sufficient conditions in terms of the eigenvalues of the *quadratic characteristic* of $M_n(\theta)$, denoted by $\langle M(\theta)\rangle_n$. For example, with univariate θ, the quadratic characteristic for the DM estimating function $M_n(\theta) = \sum_{i=1}^n D_i(\theta)$ is

$$\langle M(\theta)\rangle_n = I_n(\theta) = \sum_{i=1}^n E[D_i^2(\theta)|\mathcal{F}_{i-1}].$$

More generally, Lin (1994) [34] gives sufficient conditions for the strong law to hold for multivariate martingales in terms of the eigenvalues of random norming matrices.

3.3.2 A Martingale Central Limit Theorem

Next is a martingale central limit theorem due to Brown (1971) [8] and following Section 1.6 of [22]:

Theorem 3.3.5 Martingale CLT. Let $\{M_n = \sum_{i=1}^n D_i, \mathcal{F}_n\}$ be a 0 mean martingale whose increments D_i have finite variance. Set

$$V_n^2 = \sum_{i=1}^n E(D_i^2|\mathcal{F}_{i-1}) \quad \text{and } s_n^2 = E[V_n^2] = E[M_n^2].$$

If $s_n^{-2}V_n^2 \xrightarrow{P} 1$ and if for all $\epsilon > 0$

$$s_n^{-2}\sum_{i=1}^{n} E(D_i^2 I(|D_i| \geq \epsilon s_n)) \to 0 \text{ as } n \to \infty$$

then

$$s_n^{-1}M_n \xrightarrow{D} N(0,1).$$

Billingsley (1961) [6] proves a CLT for homogeneous (stationary), ergodic processes. For example, he proves the condition $s_n^{-2}V_n^2 \xrightarrow{P} 1$ by exploiting stationarity and ergodicity in the following way: stationarity gives that

$$E[D_k^2] = E[D_1^2], \quad k \geq 1 \tag{3.15}$$

$$\Rightarrow s_n^2 = nE[D_1^2] = n\sigma^2.$$

Then the ergodic theorem gives

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E(D_i^2|\mathcal{F}_{i-1}) = \sigma^2 = E[D_1^2]$$

almost surely, which further implies

$$V_n^2 = \sum_{i=1}^{n} E(D_i^2|\mathcal{F}_{i-1}) \to \infty$$

as $n \to \infty$. This last result would assist in establishing consistency for the estimator in view of 3.13. Since relation 3.15 does not hold for an inhomogeneous chain, the very same argument cannot be used for the Dynkin martingale differences based on discrete samples. However, if the chain is in some sense ergodic (Section 2.1.1), an appropriate ergodic theorem could be used to establish the conditions of the CLT. Chapter 3 of Hall and Heyde (1980) [22] contains several generalizations of Theorem 3.3.5.

3.4 Asymptotic Results for the DME

Recall the inference setting for the DME. $\{X(t), t \geq 0\}$ is an inhomogeneous continuous-time Markov chain with infinitesimal generator, $A_{\theta^*}(t)$, depending on the parameter $\theta^* \in \Theta \in \mathbb{R}$. The sample, $\{Z_k = X(T_k), \quad k = 1, \ldots, n\}$, where $\{T_i, i \geq 0\}$ are exponentially distributed observation times with mean $1/\beta$, is drawn by Poisson sampling over the observation interval, $[T_0, T]$. For a bounded, continuous function $g : S \mapsto \mathbb{R}$, the DME $\hat{\theta}_n$ for θ^* satisfies

$$M_n(\hat{\theta}_n) = g(X(T_n)) - g(X(T_0)) - \frac{1}{\beta}\sum_{i=1}^{n}(A_{\hat{\theta}_n}(T_i)g)(X(T_i)) = 0. \tag{3.16}$$

As long as $n \to \infty$ coincides with $T \to \infty$, the arguments in the previous section yield the following two propositions:

Proposition 3.4.1 Let $\hat{\theta}_n$ be the DME for θ^* based on the Poisson sample, $\{Z_k = X(T_k)\}$. Let $\tilde{\theta} \in (\hat{\theta}_n, \theta^*)$ and define

$$V_n^2 = \sum_{i=1}^{n} E[D_i(\theta^*)^2|\mathcal{F}_{i-1}] \quad , \quad s_n^2 = E[V_n^2] = E[M_n^2]$$

$$I_n(\theta^*) = V_n^2 \quad , \quad J_n(\theta) = \sum_{i=1}^{n} D_i'(\theta).$$

If, as $n \to \infty$ ($\Leftrightarrow T \to \infty$),

(1) $V_n^2 \to \infty$ and

(2) $\limsup_{n\to\infty}[I_n(\theta^*)]^{-1}|I_n(\theta^*) + J_n(\tilde{\theta})| < 1 \quad$ a.s.,

then the DME is strongly consistent for θ^* ($\hat{\theta}_n \to \theta^*$ almost surely).

Furthermore,

Proposition 3.4.2 If

(3) $s_n^{-2} I_n(\theta^*) = s_n^{-2} V_n^2 \xrightarrow{P} 1$.

(4) For all $\epsilon > 0$, $s_n^{-2} \sum_{i=1}^n E(D_i^2 I(|D_i| \geq \epsilon s_n)) \to 0$, and

(5) For $\tilde{\theta} \in (\hat{\theta}_n, \theta^*)$,

$$\begin{aligned} J_n(\tilde{\theta})^{-1} J_n(\theta^*) &\to 1 \text{ and} \\ J_n(\theta^*)/I_n(\theta^*) &\to -1 \text{ in probability.} \end{aligned}$$

then the following holds:

$$I_n^{1/2}(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{D} N(0, 1).$$

Or, if (5) is not met but

(5*) For $\tilde{\theta} \in (\hat{\theta}_n, \theta^*)$, $J_n(\tilde{\theta})^{-1} J_n(\theta^*) \to 1$,

then the modified result holds:

$$I_n^{1/2}(\theta^*) J_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{D} N(0, 1).$$

The arguments in Section 3.3 support the above two propositions. There are, however, some technical difficulties in applying them. For instance, there exists a generalization to condition (3) for which a CLT for the estimating function still applies. Moreover, additional regularity assumptions on $I_n(\theta)$ are needed in order to use $I_n^{-1/2}(\theta^*)$ as a random norming in the CLT in place of its expectation, s_n^2 (e.g., see Hall and Heyde (1980) [22], page 160). Furthermore, there are the general results for martingale estimating equations that use

$$K_n(\theta) = \sum_{i=1}^{n} E[D_i'(\theta)|\mathcal{F}_{i-1}]$$

in place of $J_n(\theta)$, as given in Section 3.2.

In any case, it is clear that the sum of conditional variances $I_n(\theta)$, the sum of partial derivatives, $J_n(\theta)$, and the sum of conditional expected partial derivatives $K_n(\theta)$ figure prominently in obtaining asymptotic results for the DME. For this reason, the next sections explore these quantities in more detail for the inhomogeneous continuous-time Markov chain.

The Conditional Variance

The sum of conditional variances, $I_n(\theta)$ can be interpreted as the information contained in a sample of size n (e.g., in maximum likelihood estimation, $I_n(\theta)$ is Fisher's information.) For consistency, it is important to know under what conditions the sum of conditional variances diverges. For normality, it is important to know under what conditions $I_n(\theta)/E(I_n(\theta))$ converges to 1 in probability. The conditional variance for a given martingale difference, $D_i(\theta)$, $i \geq 1$ is considered below. Let $g_i = g(Z_i)$, $A_i = (A_\theta(T_i)g)(Z_i)$, and for $i = 1, \ldots, n$ consider

$$D_i(\theta)^2 \quad = \quad (g(Z_i) - g(Z_{i-1}) - \frac{1}{3}(A_\theta(T_i)g)(Z_i))^2 \qquad (3.17)$$

$$= (g_i - g_{i-1})^2 + (\frac{1}{\beta^2})A_i^2 - 2\frac{1}{\beta}(g_i - g_{i-1})A_i. \quad (3.18)$$

Taking conditional expectations with respect to the sigma-field

$$\mathcal{F}_{i-1} = \sigma\{T_{i-1}, X(s); 0 \le s \le T_{i-1}\}$$

and applying Jensen's inequality,

$$\begin{aligned} E[D_i(\theta)^2|\mathcal{F}_{i-1}] &= E[(g_i - g_{i-1})^2 + \frac{1}{\beta^2}A_i^2|\mathcal{F}_{i-1}] - 2E[\frac{1}{\beta}(g_i - g_{i-1})(A_i)|\mathcal{F}_{i-1}] \\ &\ge E[(g_i - g_{i-1})|\mathcal{F}_{i-1}]^2 + \frac{1}{\beta^2}E[A_i|\mathcal{F}_{i-1}]^2 - 2E[\frac{1}{\beta}(g_i - g_{i-1})(A_i)|\mathcal{F}_{i-1}]. \end{aligned}$$

Now using martingale property ($\frac{1}{\beta}E[A_i|\mathcal{F}_{i-1}] = E[(g_i - g_{i-1})|\mathcal{F}_{i-1}]$), the right hand side of the inequality is

$$2E[g_i - g_{i-1}|\mathcal{F}_{i-1}]^2 - 2E[\frac{1}{\beta}(g_i - g_{i-1})(A_i)|\mathcal{F}_{i-1}]. \quad (3.19)$$

Assuming $Z_{i-1} = x$ and $T_{i-1} = t$ and using the notation

$$E_x^t[\cdot] = E[\cdot|(Z_{i-1}, T_{i-1}) = (x, t)],$$

expression 3.19 simplifies to

$$\begin{aligned} &2(E_x^t[g_i] - g(x))^2 - \frac{2}{\beta}E_x^t[g_iA_i] + \frac{2}{\beta}g(x)E_x^t[A_i] \\ = \; &2(E_x^t[g_i] - g(x))^2 - \frac{2}{\beta}(g(x)E_x^t[A_i] - E_x^t[g_iA_i]). \end{aligned}$$

Expanding the first term and simplifying, the right hand side becomes

$$\begin{aligned} 2(\omega_i^2 - 2\omega_i g(x) &+ g(x)^2) + 2/\beta(g(x)\omega_i - g(x)^2) - (2/\beta)\rho_i \\ &= 2\omega_i^2 + (2/\beta - 4)\omega_i g(x) + (2 - 2/\beta)g(x)^2 - (2/\beta)\rho_i. \end{aligned}$$

where

$$\begin{aligned}
\omega_i &= E_x^t[g_i] = E_x^t[g(X(t+\tau_i)]\\
&= \int_0^\infty f(s)E_x[g(X(t+s))]ds\\
&= \int_0^\infty f(s)\sum_{j\in S} g(j)p_{xj}(t,t+s)ds\\
&= \sum_{j\in S} g(j)\int_0^\infty f(s)p_{xj}(t,t+s)ds.
\end{aligned}$$

and

$$\begin{aligned}
\phi_i &= E_x^t[g_iA_i] = E_x^t[g(X(t+\tau_i))(A(t+\tau_i)g)(X(t+\tau_i))]\\
&= \int_0^\infty f(s)E_x[g(X(t+s))(A(t+s)g)(X(t+s))]ds\\
&= \int_0^\infty f(s)[\sum_{j\in S} g(j)(\sum_{k\in S} a_{jk}(t+s)g(k))p_{xj}(t,t+s)]ds\\
&= \sum_{j\in S} g(j)\int_0^\infty [f(s)p_{xj}(t,t+s)\sum_{k\in S} a_{jk}(t+s)g(k)]ds\\
&= \sum_{j\in S} g(j)\sum_{k\in S} g(k)\int_0^\infty [f(s)p_{xj}(t,t+s)a_{jk}(t+s)]ds.
\end{aligned}$$

Thus,

$$E[D_i^2(\theta)|\mathcal{F}_{i-1}] \geq 2\omega_i^2 + (2/\beta - 4)\omega_i g(x) + (2 - 2/\beta)g(x)^2 - (2/\beta)\phi_i$$

for each i. The lower bound suggests the role of the sampling rate β in the growth of the sum of the conditional variances,

$$I_n(\theta) = \sum_i E[D_i^2(\theta)|\mathcal{F}_{i-1}].$$

For the divergence condition on $I_n(\theta)$, it would be enough to know when the bound is almost surely positive for each i. However, since this bound still depends on transition probabilities, it may not be of practical use in establishing that $I_n(\theta) \to \infty$. Still, the divergence condition seems a reasonable one, given that there is no apparent reason in the construction of $M_n(\theta)$ that the conditional variances would converge to 0 as $n \to \infty$ $(T \to \infty)$. In obtaining asymptotic confidence intervals, the dependence of $I_n(\theta)$ on $P(t, s)$ presents the major difficulty; in some cases the sum of squares $\sum_1^n D_i^2(\theta)$ is a reasonable approximation (Hall and Heyde, 1980 [22], pg. 55).

Conditional Expected Derivative

Next consider the conditional expected derivatives. For test functions free of θ and again setting $Z_{i-1} = x, T_{i-1} = t$, these are for each $i = 1, \ldots, n$,

$$
\begin{aligned}
E[D_i'(\theta)|\mathcal{F}_{i-1}] &= -\frac{1}{3}E[\frac{\partial}{\partial\theta}(A_\theta(Z_i)g)(Z_i)|\mathcal{F}_{i-1}] \\
&= -\frac{1}{3}\int_0^\infty f(s)E_x[\frac{\partial}{\partial\theta}(A_\theta(t+s)g)(X(t+s))]ds \\
&= -\frac{1}{3}\int_0^\infty f(s)\sum_{j\in S}(\frac{\partial}{\partial\theta}(A_\theta(t+s)g)(j))p_{xj}(t, t+s)ds \\
&= -\frac{1}{3}\int_0^\infty f(s)\sum_{j\in S}(\sum_{k\in S}(a'_{jk}(t+s)g(k))p_{xj}(t, t+s)ds \\
&= -\frac{1}{3}\sum_{j\in S}\sum_{k\in S} g(k)\int_0^\infty f(s)a'_{jk}(t+s)p_{xj}(t, t+s)ds
\end{aligned}
$$

where a' denotes the partial of a with respect to θ. For convenience, the dependence of a on θ is implied. Finally, for test functions free of θ,

$$
\begin{aligned}
J_n(\theta) &= \sum_{i=1}^n D_i'(\theta) \\
&= -\tfrac{1}{3} \sum_{i=1}^n \sum_{k\in S} a'_{z_i k}(T_i)(g(k)).
\end{aligned}
$$

Summary

In summary, through the vehicle of Poisson sampling of a continuous-time Markov chain, asymptotic results for the DME parallel those of discrete time martingale estimating equations. The problem remains to prove that the conditions hold for a particular process the DME is derived for. Ideally, a set of generally reasonable conditions on the process could be imposed to ensure that the conditions above hold (under Poisson sampling) for the estimator solving 3.16. This important goal is set in motion though not fully pursued in this thesis. However, a simulation study in the next chapter applies the DME to a particular Markovian queueing model and provides empirical evidence that the conditions of the proposition hold in that case.

Chapter 4

A Test Case for Dynkin Martingale Estimators

This chapter provides a fully-worked example of the Dynkin Martingale estimator (DME) applied to an $M_t/M/1$ queueing process with time-dependent arrival rates. The arrival rate $\lambda(\theta, t)$ depends on an unknown parameter, θ. The purpose here is to evaluate the behavior of the DME as an estimator for θ in a variety of sampling scenarios.

In a Poisson resampling scheme, the sampling parameters are the observation interval $[T_0, T]$, the Poisson sampling rate β, and m, the number of resamples taken from the interval at each rate. A second scheme sets the observation interval, $[T_0, T]$, the sampling rate, β, and R, the number of independent realizations (or independent segments) of the queue. Varying these parameters generates numerous sampling scenarios. Generally, however, small or large samples coincide with the length of the observation period, and asymptotic behavior is considered as $T \to \infty$.

Section 4.1 presents the $M_t/M/1$ test model and the method of simulation. Section 4.2 presents the Dynkin Martingale estimating equation for the test model and summarizes the results under the different sampling scenarios. The maximum likelihood estimator (MLE) for θ serves as a benchmark to the DME in Section 4.2.2. Finally, Section 4.3 discusses the weighted version of the DME.

4.1 $M_t/M/1$ Queue with Time-Dependent Arrival Rates

This section presents the $M_t/M/1$ test model, its specification in the simulations, and details of the simulation and sampling methods. In the $M_t/M/1$ model, customers arrive according to an inhomogeneous Poisson process with rate function, $\lambda(t)$, and receive service from a single server in exponentially distributed service times with constant rate parameter, μ. The infinitesimal generator has the form

$$A(t) = (a_{ij}(t)),$$

where

$$a_{ii}(t) = -q_i(t) = \begin{cases} -(\lambda(t)+\mu) & \text{if } i \geq 1 \\ -\lambda(t) & \text{if } i = 0 \end{cases} \tag{4.1}$$

and

$$a_{ij}(t) = q_i(t)Q_{ij}(t) = \begin{cases} \mu & \text{if } j = i-1 \\ \lambda(t) & \text{if } j = i+1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

The intensity and relative transition probability function are denoted by $q_i(t)$ and $Q_{ij}(t)$, respectively. An event (arrival or departure) happens according to a Poisson process with rate $\lambda(t)+\mu$: that is,

$$Pr[\text{event occurs in } (t, t+\Delta t)] = (\lambda(t)+\mu)(\Delta t) + o(\Delta t) \tag{4.3}$$

and

$$Pr[\text{ two or more events in } (t, t+\Delta t)] = o(\Delta t). \tag{4.4}$$

These probabilities define the transition intensity function

$$q_i(t) \text{ for } i \geq 1.$$

For $i = 0$, the queue is empty and the only event possible is an arrival: therefore, $q_0(t) = \lambda(t)$. Given that an event occurs at time t, the probability that it is an arrival is

$$\frac{\lambda(t)}{(\lambda(t) + \mu)},$$

and the probability that it is a departure is

$$\frac{\mu}{(\lambda(t) + \mu)}.$$

These probabilities define the relative transition probability function,

$$Q_{ij}(t) \text{ for } i \neq j.$$

The above describes a general $M_t/M/1$ queueing process. In order to simulate the process and test the estimation method, it remains to specify the functional form of the arrival rate function and the dependence of the infinitesimal generator on the parameter θ. After specifying $\lambda(t)$ and introducing the parameter θ, the next step is to simulate the $M_t/M/1$ process. The simulation serves as a realization of the "true" process: samples are then drawn from such realizations, and estimates based on these samples are computed. Specifically, for each sample from a realization, the Dynkin Martingale estimator $\hat{\theta}_n$ solves

$$M_n(\theta) = 0$$

where

$$M_n(\theta) = \sum_{i=1}^{n} D_i(\theta) = g(X(T_n)) - g(X(T_0)) - \frac{1}{\beta}\sum_{i=1}^{n}(A_{\vartheta_n}(T_i)g)(X(T_i)).$$

4.1.1 Simulation

Let the arrival rate function depend upon t and on θ in the following way:

$$\lambda(\theta, t) = \theta + \sin(2\pi t/24).$$

The sinusoidal term puts a cyclic pattern with a period of 24 into the rate of arriving customers. In this case, the infinitesimal generator depends on θ through the arrival rate function.

A method known as *thinning* is used to simulate the queue. The thinning method simulates an inhomogeneous Poisson process with rate function $\omega(t)$ in the following way. Suppose there exists a constant ω^* so that $\omega(t) \leq \omega^*$ on the interval $[0, T]$. The first step is to simulate a homogeneous process with rate ω^*; these events are then 'thinned out' with probability $\omega(t)/\omega^*$. That is, an event is retained or rejected according to the probability $\omega(t)/\omega^*$. The events retained in such a way are simulated events from an inhomogeneous Poisson process with rate $\omega(t)$ over the time interval $[0, T]$. For more details and extensions, see Ross (1990) [40].

The event process of the queue described in equations 4.3 and 4.4 is an inhomogeneous Poisson process with a time-varying rate $q_i(t)$ given in 4.1. The thinning method is easy to use in this case, with

$$\omega^* = \sup_t \{q_i(t)\} = \sup_t \{\lambda(\theta, t)\} + \mu = \theta + 3$$

and rejection probabilities given by $q_i(t)/\omega^*$. With $\theta^* = 1$ and $\mu = 2$, Figure 4.1 shows a simulation of the queue on the interval $t \in [0, 140]$ with the arrival rate function superimposed. Since the service rate is constant, the traffic intensity

$$\rho(t) = \frac{\lambda(\theta, t)}{\mu},$$

varies as $\lambda(\theta, t)$ varies with t. With the choices $\theta^* = 1$ and $\mu = 2$, $\rho(\theta^*, t) \leq 1$ for all $t \geq 0$, so that the queue length will not explode. For the simulation over a longer period, [0,30K] (where 'K' represents 1000), Figures 4.2 and 4.3 show respectively the distribution of the queue length and the mean queue length as a function of time.

Checking The Simulation

A simple way to check the validity of the queue simulation without resorting to complicated queueing process formulas is to compare the number of events (arrivals and departures) on $[0, T]$ to the expected number of Poisson events. The expected number of events on $[0, T]$ for an inhomogeneous Poisson process with rate $\omega(t)$ is given by

$$m(T) = \int_0^T \omega(t)dt.$$

Setting $\omega(t) = q_i(t)$ for $i \geq 1$, $m(T)$ will in general overestimate the true expected number of events in the simulation, since while the queue is in state 0, $\omega(t)$ is really

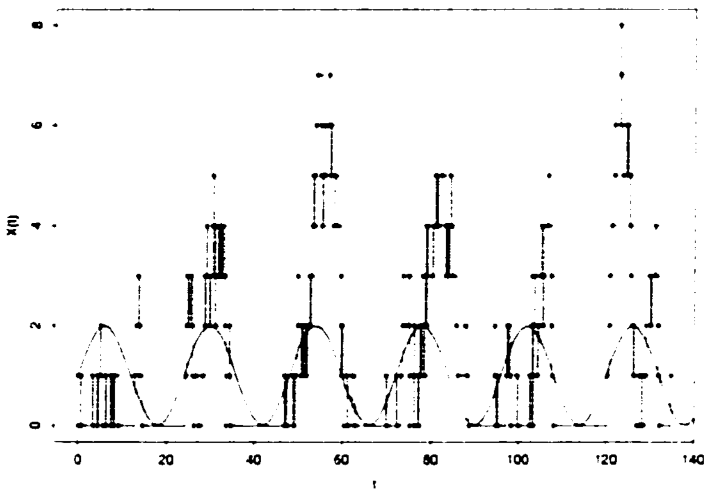

Figure 4.1 Realization of $M_t/M/1$ queue over $[0, 140]$ with arrival rate $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{\mu})$ overlaid. Simulation done with $\mu = 2$, $\theta^* = 1$.

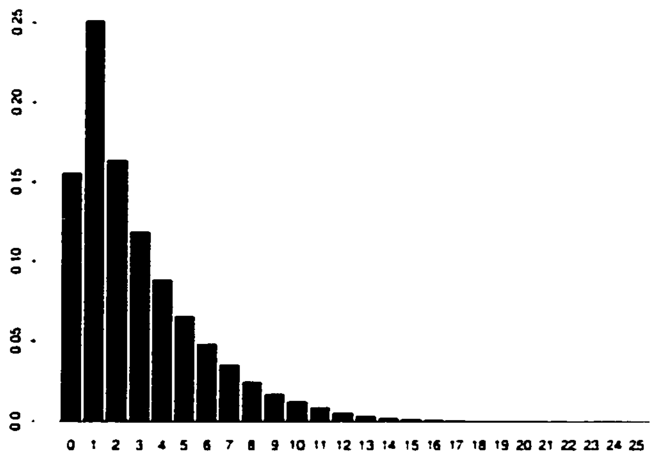

Figure 4.2 Relative frequency of states of $M_t/M/1$ simulated on $[0,30K]$ with $(\theta, \mu) = (1, 2)$, $\lambda(\theta, t) = \theta + \sin(2\pi t/24)$.

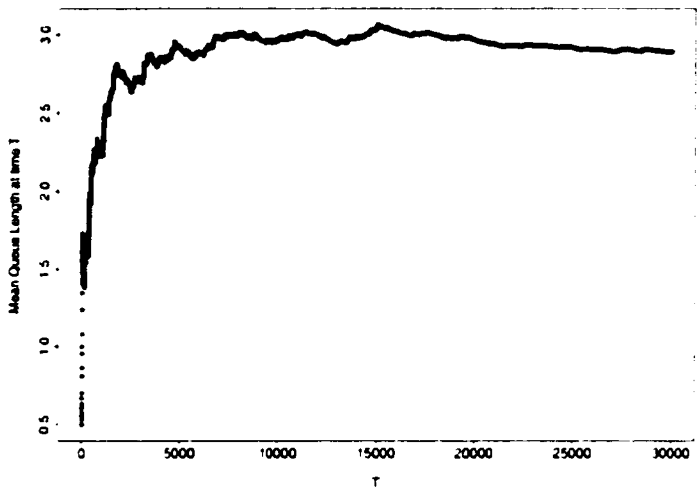

Figure 4.3 Mean Queue Length at time T of $M_t/M/1$ simulated on $[0.30\text{K}]$ with $(\theta, \mu) = (1, 2)$, $\lambda(\theta, t) = \theta + \sin(2\pi t/24)$.

$q_0(t) < q_i(t)$. Recall that for $i \geq 1$, $q_i(t) = (\lambda(\theta, t) + \mu)$, whereas $q_0(t) = \lambda(\theta, t)$. In particular, $m(T)$ computed with $\omega(t) = q_i(t)$ for $i \geq 1$ overestimates the expected number of events by $\mu\times$(length of time spent in state 0 on $[0, T]$). Therefore, the rough but easy way to check is to compute $m(T)$ and then subtract $\mu\times$(length of time *simulated* queue is in 0). For a queue simulated on $[0.30\text{K}]$ with $\theta^* = 1$ and $\mu = 2$, the expected number of events is $m(T) = m(30K) = 90K$. One realization contained 60315 events and spent about 14367 time units in state 0. Then $90K - 2(14367) = 60726$, indicating that the simulation procedure gives reasonable results.

More sophisticated approaches are available to test the validity of the simulation. Omosigho and Worthington (1988) [37] present a discrete time model to approximate certain functionals of a single server queueing process with time-dependent arrival rates and continuous service time distributions. They consider the mean and variance of queue length, idleness probability, and virtual waiting time. Clark (1981) [9] proposes the use of Polya distributions in approximating time-dependent probabilities for non-stationary (inhomogeneous) queues, in particular for the purpose of incorporating queueing delays in continuous simulations. Asmussen and Thorisson (1987) [1] consider general independent (non-Markov) queueing processes with periodic arrival rates and service times. The authors establish limit results for standard queue-related processes, such as waiting times and queue lengths, using the phase process which comprises a Markov chain.

Sampling

The next step is to sample from the queue according to an independent Poisson process with rate β. This means that the observation times, T_i are *i.i.d* exponential random variables with mean β. In practical terms, Poisson sampling with rate $\beta = .25$ would mean that on average, an observation is made every 4 time units. Higher rates would take observations more frequently, obviously. Figure 4.4 illustrates how the sampling rate affects the composition of the sample. (Observation times for the different rates are marked along the queue with an 'X'.)

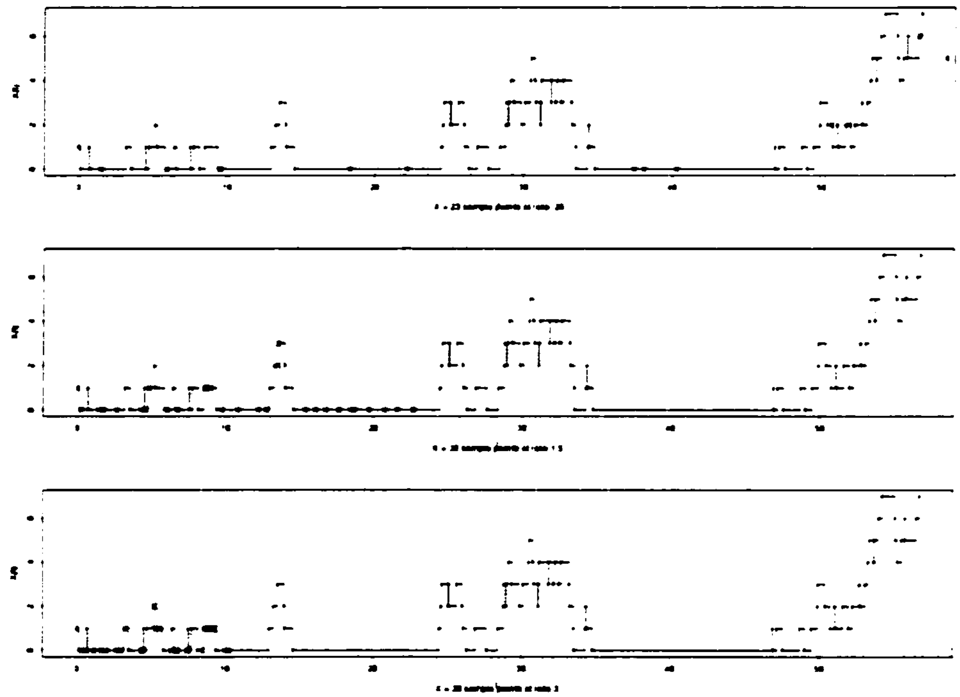

Figure 4.4 Sample segments at different sampling rates (.25. 1.5. 3) showing how the sampling rate affects the composition of the sample.

4.2 The Dynkin Martingale Estimating Equation

This section presents the derivation of the Dynkin Martingale estimating equation based on Poisson sample from the $M_t/M/1$ test model. Let $\{(Z_i, T_i), i = 1, 2, \ldots, n\}$ denote the collection of observations taken at Poisson rate β over some observation interval $[T_0, T]$. The (unweighted) Dynkin Martingale estimating function is

$$
\begin{aligned}
M_n(\theta) &= g(Z_n) - g(Z_0) - \frac{1}{\beta}\sum_{i=1}^{n}(A_\theta(T_i)g)(Z_i) \\
&= g(Z_n) - g(Z_0) \\
&\quad -\frac{1}{\beta}\sum_{i:Z_i \geq 1}\{\mu g(Z_i - 1) - (\mu + \lambda(\theta, T_i))g(Z_i) + \lambda(\theta, T_i)g(Z_i + 1)\} \\
&\quad -\frac{1}{\beta}\sum_{i:Z_i = 0}\lambda(\theta, T_i)(g(Z_i + 1) - g(Z_i)).
\end{aligned}
$$

In this case, $M_n(\theta)$ is linear in θ and the estimating equation, $M_n(\theta) = 0$, has solution

$$
\begin{aligned}
\hat{\theta}_n &= \frac{1}{\sum_{i=1}^{n}(g(Z_i + 1) - g(Z_i))} \times \\
&\quad \times \left[\beta(g(Z_n) - g(Z_0)) - \mu\sum_{i:Z_i \geq 1}(g(Z_i - 1) - g(Z_i)) \right. \\
&\quad \left. - \mu\sum_{i:Z_i \geq 1}\sum_{i=1}^{n}\sin(2\pi T_i/24)(g(Z_i + 1) - g(Z_i))\right]. \qquad (4.5)
\end{aligned}
$$

Once the test function g is specified*, equation 4.5 gives the DME $\hat{\theta}_n$ based on a sample, $\{(T_i, Z_i), i = 0, 1, \ldots, n\}$.

*The requirement for g as a function of the sample space is only that it be defined and bounded in supnorm for every $x \in S$ (a restriction which possibly can be relaxed). If g is also a function of the parameter θ, it must be continuous and differentiable. The simulation here considers functions g free of θ.

The Consistency Criterion

Proposition 3.4.1 gave conditions for the consistency of the DME in terms of the sum of the conditional variances, $I_n(\theta)$. For univariate θ, the alternative consistency criterion (Heyde (1997) [27]) may prove easier to check. Recall,

Theorem 4.2.1 Consistency Criterion. Let $\{M_n(\theta)\}$ be a sequence of estimating functions that are continuous in θ a.e. on $E \subset \Omega$ for $n \geq 1$. If for all $\delta > 0$ a.e. on E there exists an $\epsilon > 0$ so that

$$\limsup_{n\to\infty} \left(\sup_{\|\theta-\theta^*\|=\delta} (\theta - \theta^*)' M_n(\theta) \right) < -\epsilon,$$

then there exists a sequence of estimators $\hat{\theta}_n$ such that for any $\omega \in E$, $\hat{\theta}_n \to \theta^*$ and $M_n(\hat{\theta}_n(\omega)) = 0$ when $n > N_\omega$.

To test the criterion empirically for the $M_t/M/1$, Figures 4.5 and 4.6 display plots of

$$\theta \quad \text{vs.} \quad (\theta - \theta^*) M_n(\theta)$$

for values θ ranging about the true value, θ^*, where $M_n(\theta)$ is based on Poisson samples drawn with rate $\beta = 1$ over observation intervals of length 1K and 10K. The vertical line marks $\theta^* = 1.5$ and the horizontal line marks 0. According to the criterion, for θ close to θ^*, the graph should fall below the horizontal line at 0 as samples grow large enough. The plots do show that the criterion clusters near zero for θ close to θ^*, and do not preclude the possibility that in the limit the condition is satisfied. In the following sections, bias and error are investigated empirically under different sampling scenarios.

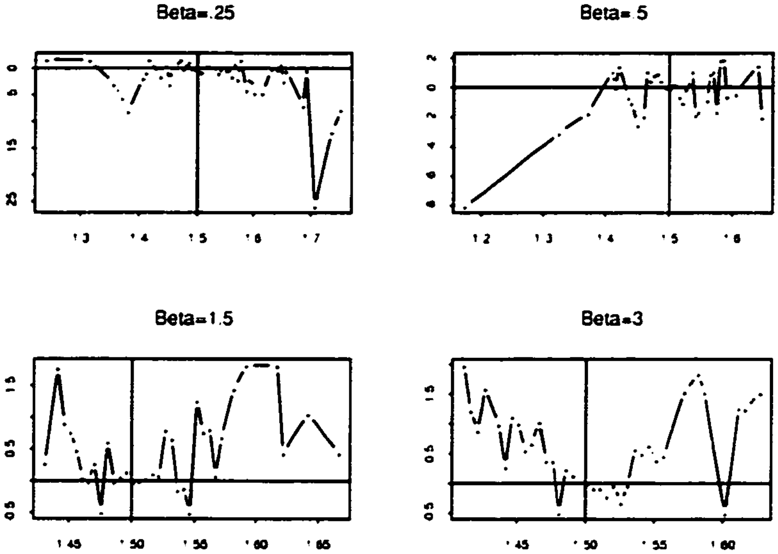

Figure 4.5 Plot of θ vs. $(\theta - \theta^*)M_n(\theta)$ for $\beta = .25, .5, 1.5, 3$ computed over an observation interval of length 1K for a single realization of the queue with $\theta^* = 1.5$, $\mu = 2.25$, $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^{-x}$.

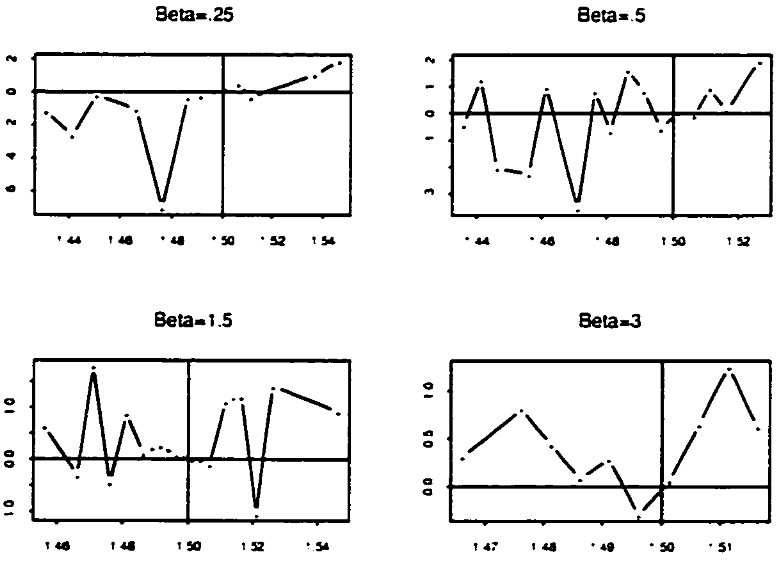

Figure 4.6 Plot of θ vs. $(\theta - \theta^*)M_n(\theta)$ for $\beta = .25, .5, 1.5, 3$ computed over an observation interval of length 10K for a single realization of the queue with $\theta^* = 1.5$, $\mu = 2.25$, $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^{-x}$.

4.2.1 Preliminary Results

The initial simulations fix the observation interval $[T_0, T]$ of a single realization of the queue. The sample is the collection of states observed at the event times of an independent Poisson process with constant rate β. The results here reflect the choices $\beta \in \{.25, .5, .75, 1, 1.5, 2, 2.5, 3\}$. For each β, Figure 4.7 shows boxplots of 1000 estimates based on samples taken on the interval [0,5K]. The distributions become tighter as the sampling rates increase, but indicate a bias. Taking larger samples (through higher sampling rates) will reduce the sampling variance but will not alleviate the bias problem (Figure 4.8). Similar plots for samples drawn at higher intervals along the queue are shown in figures 4.9 and 4.10. Figure 4.10 demonstrates that on certain intervals the DME performs well.

In this first scenario, the observation period is fixed and sample sizes vary with sampling rate. The higher rate samples contain more observations but this offers no apparent advantage in terms of the bias of the estimators. The sampling rate did not affect bias since each rate captured the same segment of the queue. These results indicate that the length of the observation period is a more important design parameter than (strict) sample size in terms of bias.

Table 4.1 provides the bias, standard deviation, and mean squared error for 50 estimates based on intervals of length 5K with varying starting points. Table 4.2 similarly shows statistics for 50 estimates based on intervals of length 10K, and one set for an interval of length 60K.

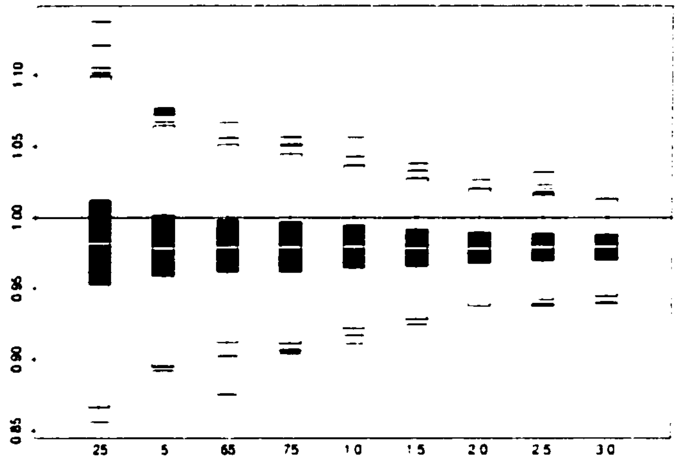

Figure 4.7 Boxplots of 1000 estimates $\hat{\theta}_n$ based on samples taken at different sampling rates β on the interval [0.5K].

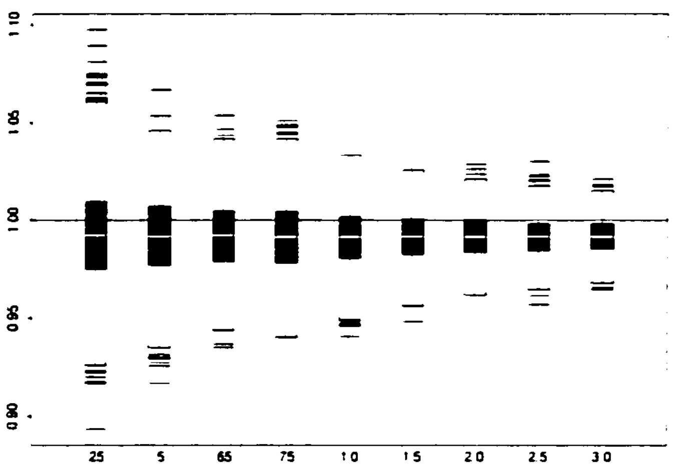

Figure 4.8 Boxplots of 1000 estimates $\hat{\theta}_n$ based on samples taken at different sampling rates β on the interval [0.10K].

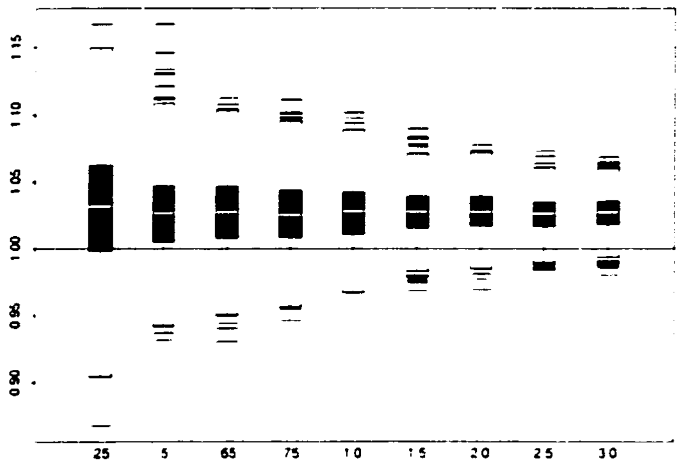

Figure 4.9 Boxplots of 1000 estimates $\hat{\theta}_n$ based on samples taken at different sampling rates β on the interval [5K,10K].

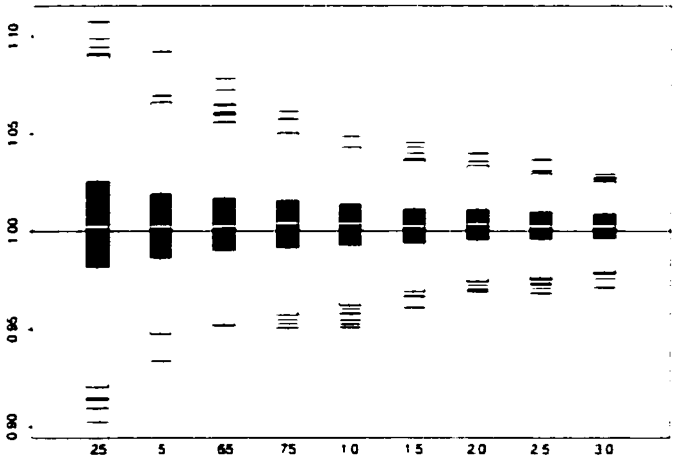

Figure 4.10 Boxplots of 1000 estimates $\hat{\theta}_n$ based on samples taken at different sampling rates β on the interval [5K,15K].

Table 4.1 Estimates of $\theta^* = 1$ based on intervals of length 5K from $M_t/M/1$, $\lambda(\theta,t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1+e^x)$

[$tstart, tfinish$]	β	Bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
[15K,20K]	.25	-.0319	.0351	.0022
	.75	-.0355	.0223	.0017
	1.5	-.0334	.0168	.0014
	2.5	-.0346	.0138	.0014
[25K,30K]	.25	.0101	.0376	.0015
	.75	.0029	.0213	.0005
	1.5	-.0008	.0167	.0003
	2.5	2.7e-05	.0124	.0001
[45K,50K]	.25	-.0052	.0409	.0017
	.75	-.0091	.0243	.0007
	1.5	-.0041	.0159	.0002
	2.5	-.0075	.0115	.0002

Table 4.2 Estimates of $\theta^* = 1$ based on intervals of length 10K (and of 60K) from $M_t/M/1$, $\lambda(\theta,t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1+e^x)$

[$tstart, tfinish$]	β	Bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
[5K,15K]	.25	.0056	.0351	.0007
	.75	.0053	.0223	.0002
	1.5	.0063	.0168	.0002
	2.5	.0036	.0138	6.0e-05
[35K,45K]	.25	.0286	.0326	.0019
	.75	.0219	.0182	.0008
	1.5	.0263	.0109	.0008
	2.5	.0246	.0062	.0007
[45K,55K]	.25	-.0040	.0039	.0007
	.75	-.0053	.0154	.0003
	1.5	-.0063	.0119	.0002
	2.5	-.0073	.0080	.0001
[0,60K]	.25	.0049	.0096	.0001
	.75	.0076	.0071	.0001
	1.5	.0077	.0044	7.0e-5
	2.5	.0064	.0064	5.0e-5

Independent Intervals

If each estimate is based on a different (independent) interval of the queue, the sampling distributions for the most part center on 1, the true value of the parameter. Figures 4.11 and 4.12 illustrate this fact for 50 estimates based on intervals of lengths 1K, 2K, 3K, and 5K.

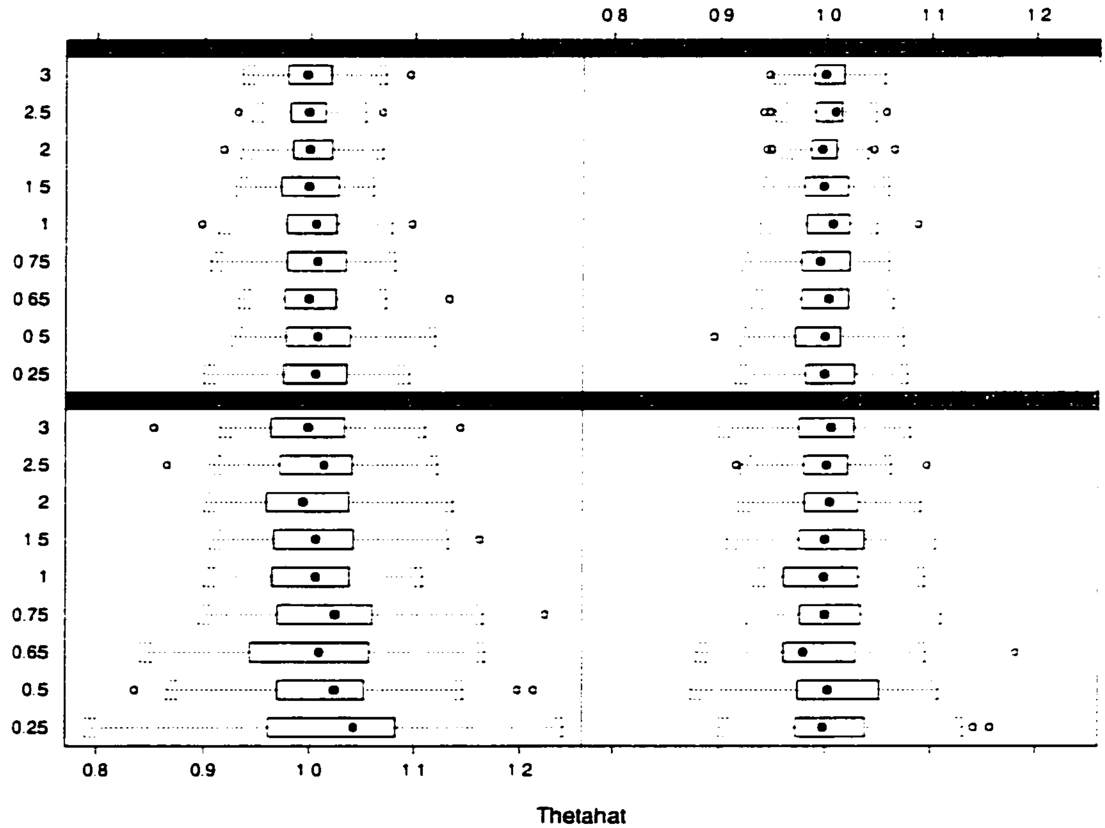

Figure 4.11 Boxplots of $\hat{\theta}_n$ based on 50 intervals of length 1K, 2K, 3K, and 5K (over panels) for different sampling rates (y-axis within panels). Panels read left to right. $\theta^* = 1$, $\mu = 2$, $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1 + e^x)$.

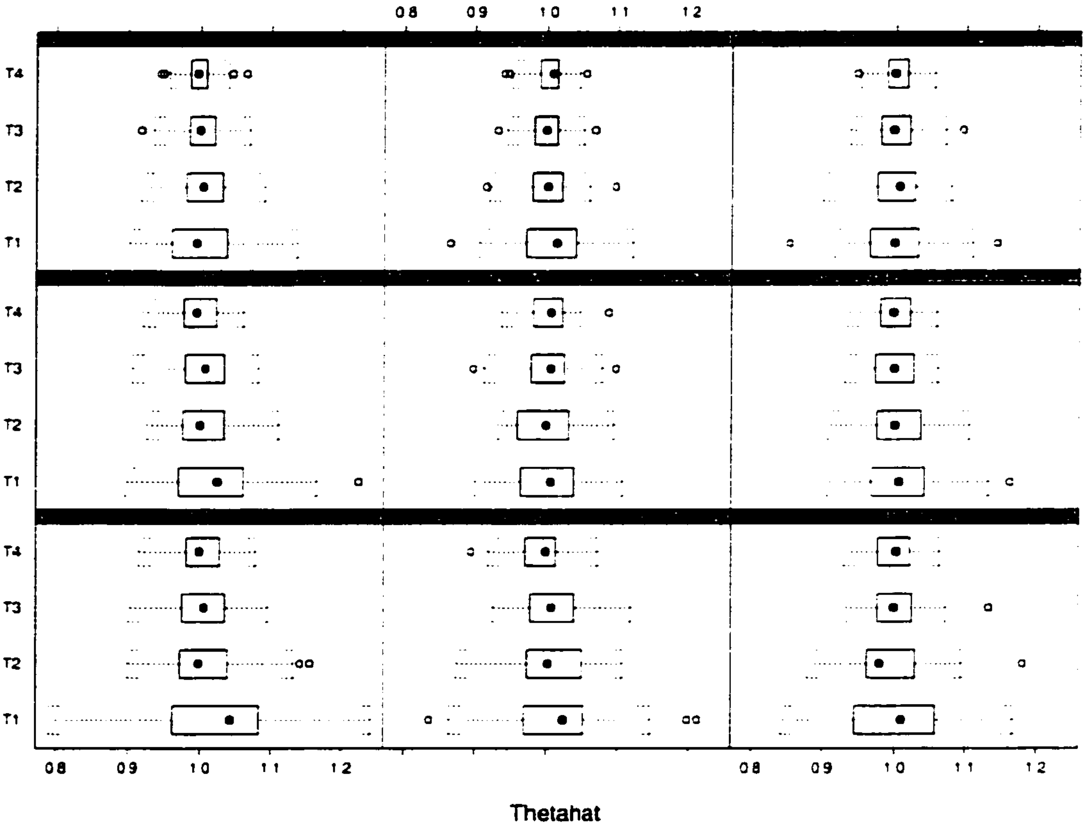

Figure 4.12 Boxplots of $\hat{\theta}_n$ based on 50 intervals of length 1K, 2K, 3K, and 5K (within panels) for different sampling rates (over panels). Panels read left to right. $\theta^* = 1$, $\mu = 2$, $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1 + e^x)$).

These plots also demonstrate that matters improve for the larger intervals, as one expects since asymptotic properties in this context obtain as $T \to \infty$. Table 4.3 contains the corresponding statistics for these runs.

Table 4.3 Summary statistics for 50 estimates of $\theta^* = 1$ based on intervals of length 1K, 2K, 3K, and 5K from the $M_t/M/1, \lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1+e^x)$

T (length)	β	bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
1K	.25	.0248	.0969	.0098
	.75	.0227	.0793	.0067
	1.5	.0116	.0603	.0037
	2.5	.0041	.0530	.0028
2K	.25	.0095	.0588	.0035
	.75	.0065	.0455	.0021
	1.5	.0038	.0448	.0020
	2.5	.0001	.0378	.0014
3K	.25	.0046	.0537	.0028
	.75	.0075	.0367	.0014
	1.5	-.0016	.0348	.0012
	2.5	-.0015	.0288	.0008
5K	.25	.0036	.0360	.0013
	.75	-.0014	.0314	.0010
	1.5	-.0008	.0272	.0007
	2.5	.0026	.0250	.0006

Based on $R = 200$ independent realizations of the queue, DMEs exhibit no detectable bias problems. Figure 4.13 displays boxplots of the corresponding estimates for two choices of the test function and each sampling rate over intervals of varying length.

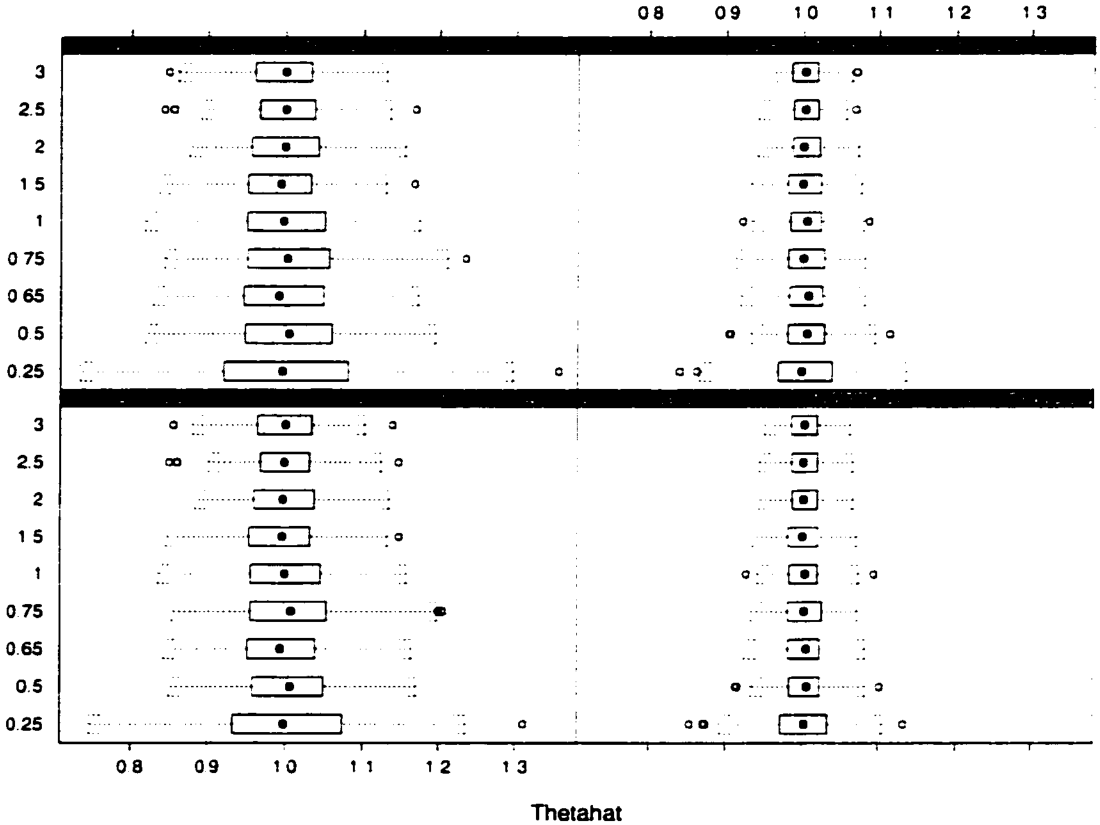

Figure 4.13 Estimates $\hat{\theta}_n$ of $\theta^* = 1$ based on samples over intervals of length 1K, and 5K (left to right) from 200 independent realizations. Test functions are $g(x) = e^{-x}$ and $g(x) = e^x/(1 + e^x)$ (bottom to top).

A Note on the Choice of Test Function

There are no restrictions besides boundedness on the choice of $g(x)$. This is imposed to ensure integrability in the derivation of the sample martingale (although in a practical sense, the boundedness condition may be unnecessarily restrictive). In the tests presented here, the condition poses no impediment. The logistic function is a natural choice for $g(x)$ since it is bounded on the state space, as is the negative exponential, $g(x) = e^{-x}$; the logistic is monotone increasing and concave while the negative exponential is monotone decreasing and convex. As a rough comparison of the two test functions, Table 4.4 shows statistics for DMEs using $g(x) = e^{-x}$ based on the same samples that were used above in Table 4.3 for the logistic. The bias in the smaller intervals ($T = 1K$) is better by an order of magnitude; the advantage of e^{-x} appears to diminish as T grows larger, however. The comparison suggests that there may be better choices of test functions in small samples. If further the test function is allowed to depend on the parameter, differentiability and tractability become important criteria. It is not clear at this point how a test function should depend on θ or what advantage it would offer.

Table 4.4 Estimates of $\theta^* = 1$ based on 50 intervals of length 1K, 2K, 3K, and 5K from $M_t/M/1$, $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^{-x}$

T (length)	β	bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
1K	.25	.0028	.1198	.0141
	.75	.0051	.0742	.0054
	1.5	.0013	.0664	.0043
	2.5	-.0009	.0573	.0032
2K	.25	.0093	.0653	.0043
	.75	.0089	.0507	.0026
	1.5	.0060	.0461	.0021
	2.5	.0009	.0392	.0015
3K	.25	.0061	.0625	.0039
	.75	.0090	.0392	.0016
	1.5	-.0017	.0371	.0014
	2.5	-.0009	.0317	.0010
5K	.25	.0052	.0360	.0019
	.75	-.0008	.0314	.0011
	1.5	-.0009	.0272	.0008
	2.5	.0032	.0250	.0006

4.2.2 MLE Comparison

Because of its optimal asymptotic properties, the maximum likelihood estimator serves as a benchmark in comparing different estimators. This section compares the MLE based on continuous observation with the DME based on Poisson samples.

Recall that the likelihood function for a time-inhomogenous Markov chain in terms of the infinitesimal generator $A(t) = (a_{ij}(t) = a(i, j, t))$, $i, j \in S$, is

$$L_{J(T)}(\theta) = \prod_{k=0}^{J(T)-1} a(z_k, z_{k+1}, T_{k+1}) \exp^{\int_0^{\tau_k} a(z_k, z_k, s) ds}$$

where $J(T)$ is the number of jumps on $[0, T]$, and $\{(z_k, T_k), k = 0, \ldots, J(T) - 1\}$ is the collection of states and transition times based on continuous observation.

Then the likelihood score function is

$$\frac{\partial \log L_{J(T)}(\theta)}{\partial \theta} = \sum_{k=0}^{J(T)-1} \frac{\frac{\partial}{\partial \theta} a(z_k, z_{k+1}, T_{k+1})}{(a(z_k, z_{k+1}, T_{k+1}))} + \int_0^{\tau_k} \frac{\partial}{\partial \theta} a(z_k, z_k, s)ds.$$

(It is assumed that the $a(i, j, t)$ depend on the parameter, θ.) For the $M_t/M/1$ test model, it is not hard to show that the MLE solves the equation,

$$\sum_{Z_i = Z_{i-1}+1} \frac{1}{\lambda(\theta, T_i)} = \sum_{Z_i = Z_{i-1}+1} \frac{1}{\theta + \sin(2\pi T_i/24)} = T. \tag{4.6}$$

Though the likelihood equation is simple, solving for the MLE requires a root-finding algorithm. The simulations with $\theta^* = 1$ and $\mu = 2$ present some practical difficulties because for $\theta < 1$, $\lambda(\theta, t)$ is negative and the likelihood itself is undefined. A constrained optimization procedure would solve this difficulty but would add substantially to the computational time and would not offer any greater insight about the new estimators. There is no loss of generality if instead the MLE is based on simulations of the queue with $\theta^* = 1.5$, $\mu = 2.5$. Then the likelihood is defined on an open interval about the true value and a basic root-finding algorithm is sufficient.

From the likelihood equation 4.6, Fisher information is

$$\begin{aligned}
I_n(\theta) &= -E[\frac{\partial^2 \log L_{J(T)}(\theta)}{\partial \theta^2}] = \sum_{Z_i = Z_{i-1}+1} E[\frac{1}{\lambda^2(\theta, T_i)}] \\
&= \sum_{Z_i = Z_{i-1}+1} \frac{1}{\lambda^2(\theta, T_i)} P[Z_i = Z_{i-1} + 1 | (Z_{i-1}, T_{i-1})] \\
&= \sum_{Z_i = Z_{i-1}+1} \left(\frac{1}{\lambda^2(\theta, T_i)}\right) \left(\frac{\lambda(\theta, T_i)}{(\lambda(\theta, T_i) + \mu)}\right) \\
&= \sum_{Z_i = Z_{i-1}+1} \frac{1}{\lambda(\theta, T_i)(\lambda(\theta, T_i) + \mu)}.
\end{aligned} \tag{4.7}$$

Then the standard error of the MLE is $I_n^{-1/2}(\theta)$. The MLE (based on continuous observation) is shown along with the DME based on Poisson resampling and based on Poisson sampling from independent realizations of the queue. The standard deviation of the estimates obtained according to each sampling scenario roughly approximate standard errors for the DME.

Poisson Resampling

In the first comparison, the DME is based on m Poisson resamples over a fixed observation interval. The standard deviation of the DME over these m samples provides an estimate of the DME standard error. The MLE is based on continuous observation over the same interval, with its standard error given by the inverse square root of Fisher information, 4.7. Table 4.5 shows the bias and standard error for the MLE on intervals of length 100, 1K,5K, and 10K next to the bias, standard deviation, and root MSE for the DME based on $m = 100$ resamples over the same intervals.

The performance of resampled DMEs appears to be on par with the MLE, and even better in terms of bias for the smaller observation intervals. As expected, the MLE by all measures improves as the length of the observation interval grows large. For the DME, the standard deviation and MSE improve as T grows large, but the bias for all but the small interval ($T = 100$) remains about the same for each sampling rate. Table 4.5 is typical of the results obtained over several different realizations of the queue. Among the worst cases, however, the DME has bias over double that of the

MLE (table 4.6). For the longer observation intervals. DME standard deviations are typically lower than the MLE standard error. This is misleading, since the standard deviation over resamples is only a rough estimate of the true standard error. The resampling method suggests using better. bootstrap type estimators and this kind of investigation is among the topics for future work. A valid comparison can be made between the MLE standard deviation and the DME standard deviation over R independent realizations. The next subsection covers this case.

Table 4.5 MLEs and DMEs of $\theta^* = 1.5$ based on 100 resamples along intervals of varying length from $M_t/M/1$. $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$. $g(x) = e^x/(1+e^x)$.

T (length)	MLE	DME β	Bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
100	1.315	.25	.1570	.4566	.4949
	bias: -.1846	.75	.0996	.2478	.2659
	SE: .2186	1.5	.0881	.1936	.2119
		2.5	.0633	.1271	.1411
1K	1.474	.25	.0006	.1242	.1237
	bias: -.0259	.75	-.0103	.0672	.0678
	SE: .0631	1.5	-.0221	.0549	.0592
		2.5	-.0193	.0419	.0459
5K	1.521	.25	-.0196	.0586	.0616
	bias: .0210	.75	.0174	.0325	.0361
	SE: .0276	1.5	.0226	.0241	.0332
		2.5	.0195	.0177	.0265
10K	1.489	.25	.0287	.0489	.0566
	bias: -.0107	.75	.0226	.0177	.0286
	SE: .0198	1.5	.0151	.0219	.0265
		2.5	.0174	.0088	.0195

Table 4.6 DME of $\theta^* = 1.5$ based on 100 resamples over [50K,55K] from $M_t/M/1$, $\lambda(\theta,t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1+e^x)$.

T (length)	MLE	DME			
		β	Bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
	1.498	.25	.0411	.0606	.0054
5K	bias: -.0016	.75	.0358	.0345	.0025
	SE: .0278	1.5	.0380	.0225	.0019
		2.5	.0362	.0196	.0017

Independent Realizations

Table 4.7 summarizes the performance of the DME against the MLE over 200 independent realizations of the queue based on samples of lengths 100, 1K, 5K, and 10K. In some instances, the DME does substantially better in terms of bias than the MLE. Table 4.8 summarizes the performance of the DME based on samples taken over the interval [30K,40K]. Note also that, as expected, the standard deviations of the MLE are uniformly smaller than those of the DME.

Table 4.7 Summary statistics of MLE and DME of $\theta^* = 1.5$ based on samples from $R = 200$ independent realizations of the $M_t/M/1$ queue with $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1+e^x)$.

T (length)	β	bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
100	(MLE)	.0039	.1077	.0116
	.25	.0177	.4376	.1918
	.75	.0341	.3319	.1113
	1.5	.0144	.2571	.0663
	2.5	.0105	.2366	.0561
500	(MLE)	-.0045	.0464	.0022
	.25	.0149	.2065	.0429
	.75	-.0073	.1298	.0169
	1.5	-.0060	.1169	.0137
	2.5	-.0040	.1049	.0110
1K	(MLE)	-.0023	.0343	.0012
	.25	.0132	.1445	.0211
	.75	.0003	.0927	.0086
	1.5	-.0077	.0817	.0067
	2.5	-.0121	.0701	.0051
10K	(MLE)	.0006	.0109	.0001
	.25	.0031	.0434	.0019
	.75	.0019	.0308	.0009
	1.5	.0024	.0256	.0006
	2.5	.0023	.0265	.0007

Table 4.8 Summary statistics of the best performing DME of $\theta^* = 1.5$ based on samples over [30K,40K] from $R = 200$ independent realizations of $M_t/M/1$, $\lambda(\theta, t) = \theta + \sin(\frac{2\pi t}{24})$, $g(x) = e^x/(1+e^x)$.

T (length)	β	bias $\hat{\theta}_n$	sdev $\hat{\theta}_n$	MSE
10K	(MLE)	-.0014	.0100	.0001
	.25	.0003	.0474	.0022
	.75	.0006	.0315	.0009
	1.5	-.0002	.0269	.0007
	2.5	-.0011	.0226	.0005

4.3 The Weighted Version

Recall that the Dynkin Martingale estimating function constructed above falls in the class of the so-called unbiased martingale estimating function first presented in Chapter 2 and for such estimating functions there are optimal weights.

$$a^*_{i-1} = \frac{E[D_i'(\theta)|\mathcal{F}_{i-1}]}{E[D_i^2|\mathcal{F}_{i-1}]}$$

which minimize the precision measure.

$$\frac{E[M_n^2(\theta)]}{E[M_n'(\theta)]^2}.$$

The DME based on a weighted version solves

$$M_n(\theta) = \sum_{i=1}^{n} a_{i-1}(\theta) D_i(\theta)$$

where the $a_{i-1}(\theta)$ are measurable $\mathcal{F}_{i-1}$. Choosing the a_{i-1} to be optimal requires computing $E[D_i'(\theta)|\mathcal{F}_{i-1}]$ and $E[D_i^2(\theta)|\mathcal{F}_{i-1}]$ for each $i = 1, \ldots, n$. To find these in closed form requires the transition probabilities: this negates the advantage of basing inference for inhomogeneous processes on the infinitesimal generator. Case by case. one could investigate approximating the weights. perhaps by estimating the transition probabilities. (See Fleming (1978) [15] for an approach to estimating the transition probabilities of a continuous-time. inhomogeneous Markov chain.) Even then. however. the solution to the estimating equation likely would require an iterative root-finding algorithm. At this point. it is at least worthwhile to explore the quantities in more detail for the $M_t/M/1$ test model.

Conditional Variances of $D_i(\theta)$ for the $M_t/M/1$

Recall from Section 3.4 that when $X(t)$ is observed at times T_i following an exponential distribution with parameter β and density function. $f(t)$. the conditional variance for $D_i(\theta)$ satisfies

$$E[D_i^2(\theta)|\mathcal{F}_{i-1}] \geq 2\omega_i^2 + (2/\beta - 4)\omega_i g(x) + (2 - 2/\beta)g(x)^2 - (2/\beta)\phi_i$$

where

$$\omega_i = \sum_{j\in S} g(j) \int_0^\infty f(s) p_{xj}(t.t+s) ds$$

and

$$\phi_i = \sum_{j\in S} g(j) \sum_{k\in S} g(k) \int_0^\infty [f(s) p_{xj}(t.t+s) a_{jk}(t+s)] ds.$$

For the $M_t/M/1$ test model. first let

$$B_x^t(j.k) = \int_0^\infty [f(s) p_{xj}(t.t+s) a_{jk}(t+s)] ds.$$

then for the tridiagonal generator of a birth and death process

$$\begin{aligned} \phi_i &= \sum_{j\in S\setminus\{0\}} [g(j)g(j-1)B_x^t(j.j-1) + g(j)g(j)B_x^t(j.j) + g(j)g(j+1)B_x^t(j.j+1)] \\ &+ g(0)g(0)B_x^t(0.0) + g(0)g(1)B_x^t(0.1). \end{aligned}$$

For the $M_t/M/1$ test model. $A(t) = (a_{ij}(t))$ with

$$a_{ii}(t) = -q_i(t) = \begin{cases} -\lambda(\theta.t) + \mu) & \text{if } i \geq 1 \\ -\lambda(\theta.t)) & \text{if } i = 0 \end{cases} \tag{4.8}$$

and

$$a_{ij}(t) = q_i(t)Q_{ij}(t) = \begin{cases} \mu & \text{if } j = i - 1 \\ \lambda(\theta, t) & \text{if } j = i + 1 \\ 0 & \text{otherwise,} \end{cases} \tag{4.9}$$

so that

$$a'_{jk}(t+s) = \begin{cases} -1 & \text{if } k = j \\ 1 & \text{if } k = j + 1 \\ 0 & \text{otherwise,} \end{cases} \tag{4.10}$$

where a' denotes the partial derivative of a with respect to θ and the dependence of a on θ is assumed. Denoting by $p_{xj}(t, t+s)$ the transition probabilities,

$$P[X(t+s) = j | X(t) = x],$$

$$B^t_x(j,k) = \begin{cases} -\int_0^\infty [f(s)p_{xj}(t,t+s)(\lambda(\theta,t+s)+\mu)]ds & \text{if } k = j \geq 1 \\ -\int_0^\infty [f(s)p_{xj}(t,t+s)\lambda(\theta,t+s)]ds & \text{if } k = j = 0 \\ \int_0^\infty [f(s)p_{xj}(t,t+s)\lambda(\theta,t+s)]ds & \text{if } k = j+1 \\ \int_0^\infty [f(s)p_{xj}(t,t+s)(\mu)]ds & \text{if } k = j-1 \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

which leads to

$$\begin{aligned} \sigma_i &= \sum_{j\in S} \left[[g(j)g(j+1) - g(j)^2] \int_0^\infty [f(s)p_{xj}(t,t+s)\lambda(\theta,t+s)]ds \right] \\ &+ \sum_{j\in S\setminus\{0\}} \left[[g(j)g(j-1) - g(j)^2] \int_0^\infty f(s)p_{xj}(t,t+s)\mu ds \right] \end{aligned}$$

Conditional Expected Derivative

Recall from Section 3.4.

$$E[D_i'(\theta)|\mathcal{F}_{i-1}] = -\frac{1}{3}\sum_{j\in S}\sum_{k\in S} g(k)\int_0^\infty f(s)a_{jk}'(t+s)p_{xj}(t.t+s)ds$$

where a' denotes the partial of a with respect to θ. For convenience. the dependence of a on θ is implied. Let

$$\tilde{B}_x^t(j.k) = \int_0^\infty [f(s)p_{xj}(t.t+s)a_{jk}'(t+s)]ds.$$

Then for a birth-death process.

$$\begin{aligned} E[D_i'(\theta)|\mathcal{F}_{i-1}] &= -\frac{1}{3}[\sum_{j\in S\setminus\{0\}} g(j-1)\tilde{B}_x^t(j.j-1) + g(j)\tilde{B}_x^t(j.j) + g(j+1)\tilde{B}_x^t(j.j+1) \\ &+ g(0)\tilde{B}_x^t(0.0) + g(1)\tilde{B}_x^t(0.1)]. \end{aligned} \tag{4.11}$$

Using 4.10.

$$\tilde{B}_x^t(j.k) = \begin{cases} -\int_0^\infty[f(s)p_{xj}(t.t+s)]ds & \text{if } k=j \\ \int_0^\infty[f(s)p_{xj}(t.t+s)]ds & \text{if } k=j+1 \\ 0 & \text{otherwise.} \end{cases}$$

so that $E[D_i'(\theta)|\mathcal{F}_{i-1}]$ for the $M_t/M/1$ model is

$$\begin{aligned} &= -\frac{1}{3}\sum_{j\in S} -g(j)\int_0^\infty [f(s)p_{xj}(t.t+s)]ds + g(j+1)\int_0^\infty [f(s)p_{xj}(t.t+s)]ds \\ &= -\frac{1}{3}\sum_{j\in S}\int_0^\infty [f(s)p_{xj}(t.t+s)]ds[g(j+1)-g(j)]. \end{aligned}$$

If g is logistic (and therefore increasing on S).

$$0 \geq E[D'(\theta)|\mathcal{F}_{i-1}] \;\geq\; -\frac{1}{3}\sum_{j\in S}\left(\frac{e^{j+1}}{1+e^{j+1}} - \frac{e^j}{1+e^j}\right)$$

$$= -\frac{1}{\beta}\lim_{j\to\infty}\frac{e^j}{1+e^j} - \frac{e^0}{1+e^0}$$
$$= -\frac{1}{2\beta}.$$

On the other hand, if $g(x) = e^{-x}$, it is easy to show that

$$0 \le E[D_i'(\theta)|\mathcal{F}_{i-1}] \le \frac{1}{\beta}.$$

For the general test function g,

$$|E[D_i'(\theta)|\mathcal{F}_{i-1}]| \le \frac{1}{\beta}\sum_{j=0}^{\infty}(g(j+1)-g(j)) = \frac{1}{\beta}[\lim_{j\to\infty} g(j) - g(0)],$$

and therefore the sum of conditional expected derivatives,

$$|\dot{K}_n(\theta)| \le \frac{n}{\beta}[\lim_{j\to\infty} g(j) - g(0)].$$

Thus for the $M_t/M/1$, the conditional expected derivative of $D_i(\theta)$ has simple bounds depending on β and the test function g. However, in this case the infinitesimal parameters are linear in θ; in general, one should not expect to find bounds for the conditional expected derivative free of θ. In summary, these results for the $M_t/M/1$ may be used to approximate or bound asymptotic variances of the DME, perhaps by simulating the conditional expectations using the estimate $\dot{\theta}_n$ which solves the unweighted version of the estimating equations. This and many other interesting questions are left for future work. The next chapter outlines further strategies for continuing the investigation of Dynkin Martingale estimators for inhomogeneous continuous-time Markov chains.

Chapter 5

Conclusions and Future Work

The goal of this thesis was to derive and test a new estimating method for inhomogeneous, continuous-time Markov chains parameterized by their infinitesimal generators and discretely observed. Through the device of Poisson sampling, an estimating equation based on the Dynkin Identity belongs to the class of discrete-time martingale estimating equations. It was in this context that the new method and the estimators were cast. The simulation study showed that once the generator is specified, the method is easy to apply and the estimators perform reasonably well in terms of bias and error and against the benchmark maximum likelihood estimators. In Section 3.4, Propositions 3.4.1 and 3.4.2 establish conditions ensuring consistency and normality of the new estimators.

There remain important theoretical questions which stem from the asymptotic arguments in Chapter 3. The main goal would be to characterize the processes for which the DME satisfies the conditions in Propositions 3.4.1 and 3.4.2. In the test case considered in this thesis, the new estimators perform well and appear to attain the asymptotic properties expected of them. The $M_t/M/1$ queueing model with the sinusoidal arrival rate function and constant service rate function actually has what is called shift-stationarity. There is recent work in ergodic theory for shift-stationary processes. There appear to be key results in Kallenberg (1999) [31], who proves a mean ergodic theorem and a pointwise ergodic theorem for shift-stationary processes randomly observed. The author establishes specific results when the process is observed by Poisson sampling. The conjecture that the DME is consistent and asymptotically normal when applied to shift-stationary processes thus has substantial merit. Specific technical conditions require further investigation. There is also the question as to whether or not and under what conditions Poisson sampling resolves the aliasing problem in this context. Thus, future work would establish identification conditions for the new estimators.

Further steps involve extending the estimation method to the multivariate case and making larger scale comparisons with other parametric estimation methods that base inference on the infinitesimal generator. Extending the new estimators to the multivariate case requires specifying a test function for each dimension and then using a Newton-Raphson type iteration to solve the estimating equation. One issue is the choice of test functions. (There was no problem finding viable test functions in the univariate case.) Furthermore, there may be cases in which it is advantageous to specify more test functions than parameters and then to proceed in the context of GMM.

Future work also calls for a comparison study with other estimating techniques. The closest in spirit are the estimators of Hansen and Scheinkman (1995) [25] and Duffie and Glynn (1996) [11] in the sense that they use a test-function, generator approach. This would require in the first place making the connections between the techniques. The moment conditions of Duffie and Glynn fit the framework of the DME in the case of homogenous processes. The Hansen and Scheinkman estimators extended to samples based on random observation rather than fixed intervals may be special cases of DMEs as well. A DME under continuous observation would, presumably, use the continuous version of the Dynkin Identity. Then a comparison with the quasi-likelihood approach of Hutton and Nelson (1986) [28] could be made. More interesting perhaps would be to amend their quasi-score functions for discrete samples based on random observation and then to make a comparison.

Finally, the martingale property established for the sample-analog relies only on the fact that inter-observation times are independent: in fact, the martingale result holds if the Poisson rate is allowed to depend on time. This and more general sampling schemes could be considered, particularly as motivated by the application. In some applications, for example, it may not be possible to apply a Poisson sampling scheme. Based on other observation schemes, the main question is whether or not the Dynkin Martingale leads to a martingale estimating equation.

Appendix A

C-program for Queue Simulation and Estimation

Main Program

```
#include <stdlib.h>
#include <stdio.h>
#include <math.h>
#include <time.h>

/*** Function Prototype Declarations ***/

/* Queue specific functions */
// Arrival rate function
double lmdaf(double t, double theta);
// Service rate function
double muf(double t);
// Simulates the queue
double simQf(double theta, double tmax, double t0,int x);
// Samples from the queue
double ssmpQf(double tstart,double tfinish,double tmax,double beta);

/* DME specific functions */
// Defines test function
double gf(int z,double theta);
// Generator elements a(t,x,y,theta)
double axyf(double t, int x, int y,double theta);
// Returns [A(t)g](z)
double Agf(int z,double t,double theta);
// Returns DME for sampling rate beta
double sol1f(double beta);

/* MLE specific functions */
// Derivative of arrival rate function wrt theta integrated 0 to t
double lmdaprimef(double theta,double t);
// Derivative of log-likelihood
double  objectivef(double theta, double tstart, double tfinish,double truetheta);
// Finds MLE
double bisolvef(double start, double tstart, double tfinish, double truetheta);

/************************************************************************
Program main offers simulation and estimation for the M_t/M/1
queue. It prompts for repeated realizations or resampling and asks also
whether user wants only MLE results
************************************************************************/
```

```
void main()
{
FILE *inmle,*ininfo,*inseed,*outmle2,*outdme,*outperf,*outMn;
int j,k,x0;
int seed,optionflag,mleonlyflag;
int ntheta;
double tstart,tfinish;
/* initialize sampling rates, beta */
double betas[8]={.25,.5,.75,1.0,1.5,2.0,2.5,3.0};
double tmax,beta,t0,DME;
double tsim,Mn,MLE,mlet,mlex;
double sdevmle,biasmle,msemle,ssdmle,meanmle;
double semle,tmpspace;
double sdev,bias,mse,ssdme,meandme;
double sdeva[8],biasa[8],ssdmea[8],meandmea[8],msedmea[8];
double truetheta;

/* set parameter values */
truetheta=1.5;
/* open io files */
outperf=fopen("ergperf1","a+");
outdme=fopen("ergbetadmes1","a+");
outMn=fopen("ergMn1","a+");
printf("Enter 1 for different simulations, 0 for resampling:\n");
scanf("%d",&optionflag);
printf("Enter 1 for MLE only, 0 otherwise:\n");
scanf("%d", &mleonlyflag);

if(optionflag==1){
/* file seedfile.txt contains random seeds (integers 1 to RAND_MAX) */
inseed=fopen("seedfile.txt","r+");
ntheta=0;
for(j=0;j<8;j++){
sdeva[j]=0;
biasa[j]=0;
ssdmea[j]=0;
meandmea[j]=0;
msedmea[j]=0;
}
sdevmle=biasmle=msemle=ssdmle=meanmle=0;
/* set starting state, time, and tmax here */
x0=1;
t0=0;
tmax=60000;
/* set sampling or observation interval here */
tstart=30000;
tfinish=40000;
// titles
fprintf(outdme,"%lf %lf\n", tstart,tfinish);
fprintf(outMn," %lf %lf\n", tstart,tfinish);
```

```
fprintf(outperf,"%lf %lf\n", tstart,tfinish);

while(fscanf(inseed,"%d",&seed) != EOF){
ntheta+=1;
srand( (unsigned) seed);
outmle2=fopen("fmlesim2","w+");
tsim=simQf(truetheta,tmax,t0,x0);
inmle=fopen("fsim2","r+");
while(fscanf(inmle,"%lf",&mlet) != EOF){
fscanf(inmle,"%d",&mlex);
if((mlet >= tstart) & (mlet <=tfinish)){
fprintf(outmle2,"%lf %d\n",mlet,mlex);
}
}
fclose(outmle2);
MLE=bisolvef(1.1,tstart,tfinish,truetheta);
meanmle+=MLE;
ssdmle+=pow(MLE,2);
msemle+=pow(MLE-truetheta,2);
biasmle+=(MLE-truetheta);
printf("%lf MLE \n",MLE);
printf("%lf %lf %lf %lf %d\n",meanmle,sdevmle,biasmle,msemle,ntheta);
for(j =0;j<8;j++){
beta=betas[j];
printf("%lf Beta \n",beta);
Mn=ssmpQf(tstart,tfinish,tmax,beta);
DME=sol1f(beta);
meandmea[j]+=DME;
ssdmea[j]+=pow(DME,2);
msedmea[j]+=pow(DME-truetheta,2);
biasa[j]+=(DME-truetheta);
fprintf(outMn,"%lf %lf \n",beta,Mn);
fprintf(outdme,"%lf  %lf \n",beta,DME);
}
} //matches while(fscanf(inseed...)

meanmle=meanmle/(double) ntheta;
msemle=msemle/(double) ntheta;
biasmle=biasmle/(double) ntheta;
sdevmle=ssdmle/(double) ntheta - pow(meanmle,2);
sdevmle=pow(sdevmle,.5);
fprintf(outperf,"%lf %lf %lf %lf %d\n",meanmle,sdevmle,biasmle,msemle,ntheta);

for(j =0;j<8;j++){
meandmea[j]=meandmea[j]/(double) ntheta;
msedmea[j]=msedmea[j]/(double) ntheta;
biasa[j]=biasa[j]/(double) ntheta;
sdeva[j]=ssdmea[j]/(double) ntheta - pow(meandmea[j],2);
sdeva[j]=pow(sdeva[j],.5);
fprintf(outperf,"%lf %lf %lf %lf %lf %d\n",betas[j],meandmea[j],sdeva[j],
```

```
biasa[j],msedmea[j],ntheta);
}
    } // matches if
else{ // else do resampling
/* set a seed */
seed=22;
srand( (unsigned) seed);
sdev=bias=mse=ssdme=meandme=0;
ntheta=1;
outmle2=fopen("fmlesim2","w+");
//start state
x0=1;
// Queue simulation interval
t0=0;
tmax=60000;
/* Sampling or observation interval */
tstart=5000;
tfinish=8000;
/* 20,20100,5000-6000 */
tsim=simQf(truetheta,tmax,t0,x0);
/* make continuous sample for MLE */
inmle=fopen("fsim2","r+");
while(fscanf(inmle,"%lf",&mlet) != EOF){
fscanf(inmle,"%d",&mlex);
if((mlet >= tstart) & (mlet <=tfinish)){
fprintf(outmle2,"%lf %d\n",mlet,mlex);
}
}
fclose(outmle2);

/* Get MLE */
MLE=bisolvef(1.1,tstart,tfinish,truetheta);
ininfo=fopen("infofile3","r+");
fscanf(ininfo,"%lf %lf",&semle,&tmpspace);
printf("%lf MLE  %lf\n",MLE,semle);
fprintf(outdme,"%d %lf %lf\n",seed, tstart,tfinish);
fprintf(outMn,"%d %lf %lf\n",seed, tstart,tfinish);
fprintf(outperf,"sim seed  MLE  Bias\n");
fprintf(outperf,"%d %lf %lf %lf\n",seed,MLE,MLE-truetheta,semle);
fprintf(outperf,"%.0f %.0f",tstart,tfinish);
fprintf(outperf,"\n beta mean sdev avgbias mse ntheta\n");
fclose(ininfo);

if(mleonlyflag==0){
/* Get DME for each beta; Resample m=ntheta times on tstart to tfinish
Compute mean,sdev,and mse*/
for(k=0;k<8;k++){
beta=betas[k];
printf("%lf Beta \n",beta);
for(j=0;j<ntheta;j++){
```

```
Mn=ssmpQf(tstart,tfinish,tmax,beta);
DME=sol1f(beta);
meandme+=DME;
ssdme+=pow(DME,2);
mse+=pow(DME-truetheta,2);
bias+=(DME-truetheta);
fprintf(outMn,"%lf %lf \n",beta,Mn);
fprintf(outdme,"%lf  %lf \n",beta,DME);
}
meandme=meandme/(double) ntheta;
mse=mse/(double) ntheta;
bias=bias/(double) ntheta;
sdev=ssdme/(double) ntheta - pow(meandme,2);
sdev=pow(sdev,.5);
/* output to outperformance file named above in io statements */
fprintf(outperf,"%lf %lf %lf %lf %lf %d\n",beta,meandme,
         sdev,bias,mse,ntheta);
meandme=mse=bias=sdev=ssdme=0;
}
fclose(outperf);
fclose(outMn);
}// matches mleonly flag
} //matches else
} // End Main()
```

Functions

Queue Specific Functions

```
/* lmdaf: defines the arrival rate function.  In this case,
      lambda(t,theta)= theta + sin(alpha*t)
     where alpha = 2*pi/24
*/
double lmdaf(t,theta)
double t,theta;
{
double rate,alpha;
double pi,sinarg;
pi= 3.1415926535897932;
alpha=2.0*pi/24.0;
sinarg = alpha*t;
rate=theta+sin(sinarg);
return rate;
}

/* muf: returns a constant service rate.  The t is dummy argument
in this implementation */
double muf(t)
double t;
{
```

```
return 2.25;
}

/* simQf: simulates the queue on the time interval (t0, tmax) starting
in state x0  with arrival rate lmdaf(t,theta)
*/
double simQf(theta,tmax,t0,x0)
double theta, tmax, t0;
int x0;
{
FILE *outf1;
float rnum,BIG;
int nevents,event,cx;
double runif,qmax,t;
/* Names queue output file */
outf1=fopen("fsim2","w+");
fprintf(outf1,"%.6f\t%d\n",t0,x0);
nevents=0;
t=t0;
BIG=(float) RAND_MAX;
/*  from theta+max(sin(*))+mu */
qmax=theta+1+muf(t);
runif=0;
while(runif==0){
rnum=(float) rand();
runif=(double)(rnum/BIG);
}
t=t-log(runif)/qmax;
cx=x0;
while(t < tmax){
runif=0;
while(runif==0){
rnum=(float) rand();
runif=(double)(rnum/BIG);
}
if(runif*qmax <= -axyf(t,cx,cx,theta)){
nevents+=1;
runif=0;
while(runif==0){
rnum=(float) rand();
runif=(double)(rnum/BIG);
}
if(runif <=lmdaf(t,theta)/-axyf(t,cx,cx,theta)){
event=1;}
else{
event=-1;
}
cx+=event;
/* output event time and state */
fprintf(outf1,"%.6f\t%d\n",t,cx);
```

```
    }
runif=0;
while(runif==0){
rnum=(float) rand();
runif=(double)(rnum/BIG);
}
t=t-1/qmax*log(runif);
} /* matches while t< tmax*/
fclose(outf1);
return t;
}

/* ssmpQf:  samples the queue by Poisson sampling with rate beta */
double ssmpQf(tstart,tfinish,tmax,beta)
double tstart, tfinish,tmax,beta;
{
FILE *inp1,*outp1,*outp2;
int readx,state,zprev;
double readt,randt,t,runif;
double tprev,Mn,di;
    double theta=1.5; //dummy
float rnum,BIG;
/* Queue realization input file */
inp1=fopen("fsim2","r+");
/* Queue sample output file */
outp1=fopen("fsmple2","w+");
/* This output file receives the Dynkin Martingale differences, D_i */
outp2=fopen("betadi","a+");
BIG=(float)RAND_MAX;
randt=tstart;
fscanf(inp1,"%lf %d",&readt,&readx);
t=readt;
state=readx;
while(randt>t && randt < tmax){
state=readx;
fscanf(inp1,"%lf %d",&readt,&readx);
t=readt;}
fprintf(outp1,"%.6f\t%d\n",randt,state);
Mn=0;
    tprev=randt;
    zprev=state;
/*this is based on tfinish */
while(randt < tfinish){
runif=0;
while(runif==0){
rnum=(float) rand();
runif=(double)(rnum/BIG);
}
randt=randt-log(runif)/beta;
while(randt > t && fscanf(inp1,"%lf",&readt) != EOF){
```

```
state=readx;
fscanf(inp1,"%d",&readx);
t=readt;}
di=gf(state,theta)-gf(zprev,theta)-(1/beta)*Agf(state,randt,theta);
Mn+=di;
fprintf(outp2,"%lf  %lf\n",beta,di);
tprev=randt;
zprev=state;
fprintf(outp1,"%.6f\t%d\n",randt,state);
   }
fclose(inp1);
fclose(outp1);
fclose(outp2);
return Mn;
}
```

DME Specific Functions

```
/* gf: defines test function g.  This implementation g does not
depend on theta though it has it as an argument
*/
double gf(z,theta)
int z;
double theta; //dummy
{
double gfval;
gfval=exp(- (double) z);
/* other choices commented out */
    /*gfval=exp( (double) z)/(1.0+exp((double) z));*/
return gfval;
}

/* axyf: returns generator elements, a(t,x,y,theta) */
double axyf(t,x,y,theta)
double theta,t;
int x,y;
{
double axy;
if(x==y){
if(x>0){
axy=-(lmdaf(t,theta) + muf(t));
}else{
axy=-lmdaf(t,theta);}
}
if(y==(x+1)){
axy=lmdaf(t,theta);}
if(y==(x-1)){
axy=muf(t);
}
return axy;
```

```
}

/* Agf: returns [A(t,theta)g](z) */
double Agf(z,t,theta)
double t, theta;
int z;
{
    double a0,a1,a2,Agval;
a2=axyf(t,z,z+1,theta);
a1=axyf(t,z,z,theta);
if(z > 0){
a0=axyf(t,z,z-1,theta);
}
else{
a0=0;
}
Agval=a0*gf(z-1,theta)+a1*gf(z,theta)+a2*gf(z+1,theta);
return Agval;
}

/* sol1f: gets the DME as solution to M_n(theta)=0. Receives sampling
 rate beta.DOES NOT generalize to differently defined arrival rate functions
 or vector parameters theta
*/
double sol1f(beta)
double beta;
{
FILE *in4,*out4;
int zi,n,x0;
double t,pi,ti,cn,bn,kn,mu;
double DME,sinarg,alpha;
in4=fopen("fsmple2","r+");
out4=fopen("Mn","a+");
n=0;
pi= 3.1415926535897932;
t=0.0;
mu=muf(t);
cn=bn=kn=0;
alpha=2*pi/24;
fscanf(in4,"%lf%d",&ti,&zi);
x0=zi;
cn+=gf(zi+1,0)-gf(zi,0);
if(zi >=1){
kn+=gf(zi-1,0)-gf(zi,0);}
    sinarg=alpha*ti;
    bn+=sin(sinarg)*(gf(zi+1,0)-gf(zi,0));
while(fscanf(in4,"%lf",&ti)!=EOF){
fscanf(in4,"%d",&zi);
cn+=gf(zi+1,0)-gf(zi,0);
if(zi >=1){
```

```
kn+=gf(zi-1,0)-gf(zi,0);
}
sinarg=alpha*ti;
bn+=sin(sinarg)*(gf(zi+1,0)-gf(zi,0));
}
DME=beta*(gf(zi,0)-gf(x0,0))-mu*kn-bn;
DME=DME/cn;
fclose(out4);
fclose(in4);
return DME;
}
```

MLE Specific Functions

```
/* lmdaprimef:  computes integral 0 to t of the derivative of
   of the arrival rate function wrt to theta */
double lmdaprimef(theta,t)
double theta,t;
{
return t;
}

/* Computes the derivative of the log-likelihood */
double objectivef(theta,tstart,tfinish,truetheta)
double theta,tstart,tfinish,truetheta;
{
FILE *inmle1,*info;
double infoterm,term1,term2,term3;
double tk1,tk,t,dloglike;
double standerr;
int zk1,zk,count,z;
inmle1=fopen("fmlesim2","r+");
/* this file is used to calculate fishers information for MLE */
info=fopen("infofile3","w+");
term1=0;
term2=0;
term3=0;
count=0;
infoterm=0;
fscanf(inmle1,"%lf%d",&t,&z);
tk=t;
zk=z;
while(fscanf(inmle1,"%lf",&t) != EOF){
fscanf(inmle1,"%d",&z);
tk1=t;
zk1=z;
count+=1;
if(zk1==(zk+1)){
infoterm+=1.0/(lmdaf(tk1,truetheta)*(lmdaf(tk1,truetheta)+muf(tk1)));
```

```
term1+=pow(lmdaf(tk1,theta),-1);
term2+=lmdaprimef(theta,(tk1-tk));
}
if(zk1==(zk-1))
{
term3+=lmdaprimef(theta,(tk1-tk));
}
zk=zk1;
tk=tk1;
    }
fclose(inmle1);
dloglike=term1-term2-term3;
standerr=pow(infoterm,-.5);
fprintf(info,"%lf %lf\n",standerr,theta);
    fclose(info);
return dloglike;
}

/* bisolvef: solves the likelihood equation */
double bisolvef(start,tstart,tfinish,truetheta)
double start,tstart,tfinish,truetheta;
{
int iter;
double tol,next,objval,increment,oldnext,err;
iter=0;
tol = 0.00000001;
next = start;
increment = 0.1;
err = 5;
while((err > tol) &  (iter < 100)) {
objval=objectivef(start,tstart,tfinish,truetheta);
if(objval<0){
oldnext=next;
while(objval< 0) {
oldnext=next;
next = next - increment;
objval=objectivef(next,tstart,tfinish,truetheta);
}
next = oldnext - (next - oldnext)/2.0;
increment = (next - oldnext)/4.0;
}
if(objval > 0) {
while(objval > 0) {
oldnext=next;
next = next + increment;
objval=objectivef(next,tstart,tfinish,truetheta);
}
next = oldnext + (oldnext - next)/2.0;
increment = (oldnext - next)/4.0;
}
```

```
err = pow(objval*objval,.5);
printf("%lf \n",objval);
iter+=1;
}
return(next);
}
```

Bibliography

[1] Soren Asumssen and Hermann Thorisson. A Markov chain approach to periodic queues. *Journal of Applied Probability*, 24:215–225, 1987.

[2] Krishna B. Athreya and Thomas G. Kurtz. A generalization of Dynkin's identity and some applications. *The Annals of Probability*, 1:570–599, 1973.

[3] Y. Baba. Maximum likelihood estimation of parameters in birth and death processes by Poisson sampling. *Journal of the Operations Research Society of Japan*, 25:99–112, 1982.

[4] I.V. Basawa and B.L.S. Prakasa Rao. *Statistical Inference for Stochastic Processes*. Academic Press, New York, 1980.

[5] A.T. Bharucha-Reid. *Elements of the Theory of Markov Processes and their Applications*. Dover, 1997. Originally published: New York: McGraw-Hill, 1960.

[6] Patrick Billingsley. The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12:788–792, 1961.

[7] Patrick Billingsley. *Statistical Inference for Markov Processes*. University of Chicago Press, 1961.

[8] B.M. Brown. Martingale central limit theorems. *Annals of Mathematical Statistics*, 42:59–66, 1971.

[9] Gordon M. Clark. Use of Polya distributions in approximate solutions to non-stationary $M/M/s$ queues. *Communications of the ACM*, 24(4):206–217, 1981.

[10] J.H. Darwin. The behaviour of an estimator for a simple birth and death process. *Biometrika*, 43:23–31, 1956.

[11] Darrell Duffie and Peter Glynn. Estimation of continuous-time Markov processes sampled at random time intervals. Working manuscript, October 1996.

[12] Katherine B. Ensor and Peter W. Glynn. Grid-based simulation and the method of conditional least squares. Working manuscript.

[13] S. Ethier and T. Kurtz. *Markov processes: Characterization and convergence*. Wiley and Sons, New York, 1986.

[14] Willy Feller. On the integro-differential equations of purely discontinuous Markoff processes. *Transactions of the American Mathematical Society*, 48:488–515, 1940.

[15] Thomas R. Fleming. Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. *The Annals of Statistics*, 6:1057–1070, 1978.

[16] Peter Glynn. Personal communication.

[17] V.P. Godambe. An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31(4):1208–1211, 1960.

[18] V.P. Godambe. Estimation in stochastic processes. *Biometrika*, 72(2):419–428, 1985.

[19] V.P. Godambe, editor. *Estimating Functions*. Oxford University Press, 1991.

[20] V.P. Godambe and C.C. Heyde. Quasi-likelihood and optimal estimation. *International Statistical Review*, 55(3):231–244, 1987.

[21] William H. Greene. *Econometric Analysis*. Macmillan, New York, 2nd edition, 1993.

[22] P. Hall and C.C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980.

[23] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[24] Lars Peter Hansen and Thomas J. Sargent. The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica*, 51:377–387, 1983.

[25] Lars Peter Hansen and Jose Alexandre Scheinkman. Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica*, 63(4):767–804, 1995.

[26] C.C. Heyde. Quasi-likelihood and general theory of inference for stochastic processes. In A. Obretenov and V.T. Stefanov, editors, *7th International Summer School on Probability Theory and Mathematical Statistics, Lecture Notes*, pages 122–152, Singapore, 1993. Science Culture Technology Publishing.

[27] Christopher C. Heyde. *Quasi-Likelihood and its Application*. Springer-Verlag, New York, 1997.

[28] James E. Hutton and Paul I. Nelson. Quasi-likelihood estimation for semimartingales. *Stochastic Processes and their Applications*, 22:245–257, 1986.

[29] J.E. Hutton, O.T. Ogunyemi, and Paul I. Nelson. Simplified and two-stage quasi-likelihood estimators. In V.P. Godambe, editor, *Estimating Functions*, pages 169–187. Oxford University Press, 1991.

[30] Jean Johnson and Dean Isaacson. Conditions for strong ergodicity using intensity matrices. *Journal of Applied Probability*, 25:34–42, 1988.

[31] Olav Kallenberg. Asymptotically invariant sampling and averaging from stationary-like processes. *Stochastic Processes and their Applications*, 82:195–204, 1999.

[32] Niels Keiding. Maximum likelihood estimation in the birth-and-death process. *The Annals of Statistics*, 3(2):363–372, 1975.

[33] Lawrence A. Klimko and Paul I. Nelson. On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6(3):629–642, 1978.

[34] Yan-Xia Lin. On the strong law of large numbers of multivariate martingales with random norming. *Stochastic Processes and their Applications*, 54:355–360, 1994.

[35] Chris J. Lloyd and Paul Yip. A unification of inference from capture-recapture studies through martingale estimating functions. In V.P. Godambe, editor, *Estimating Functions*, Oxford Statistical Science Series, pages 65–88. Oxford University Press, New York, 1991.

[36] Christopher J. Lloyd. Optimal martingale estimating equations in a stochastic process. *Statistics & Probability Letters*, 5:381–387, 1987.

[37] S.E. Omosigho and D.J. Worthington. An approximation of known accuracy for single server queues with inhomogeneous arrival rate and continuous service time distribution. *European Journal of Operational Research*, 33(3):223–366, 1988.

[38] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer-Verlag, New York, 1983.

[39] P.C.B. Phillips. The problem of identification in finite parameter continuous time models. *Journal of Econometrics*, 1:351–362, 1973.

[40] Sheldon M. Ross. *A Course in Simulation*. Macmillan, New York, 1990.

[41] H.A. Shapiro and R.A. Silverman. Alias-free sampling of random noise. *Journal of The Society of Industrial and Applied Mathematics*, 8:225–248, 1960.

[42] Isao Shoji and Tohru Ozaki. Comparative study of estimation methods for continuous time stochastic processes. *Journal of Time Series Analysis*. 18(5):485-506. 1997.

[43] Burton Singer and Seymour Spilerman. The representation of social process by Markov models. *American Journal of Sociology*. 82:1-54. 1976.

[44] A. I. Zeifman and Dean L. Isaacson. On strong ergodicity for nonhomogeneous continuous-time Markov chains. *Stochastic Processes and Their Applications*. 50:263-273. 1994.