1 of/de 1

MICRO

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Services des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

**ASYMPTOTIC PARAMETER ESTIMATION THEORY FOR
STOCHASTIC DIFFERENTIAL EQUATIONS**

by

© **Raphael Abel Kasonga, B.Sc., Dip. App. Math., M.Sc.**

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University

Ottawa, Canada

December, 1986

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-35862-9

The undersigned recommend to the Faculty of Graduate Studies and Research acceptance
of the thesis

**Asymptotic Parameter Estimation Theory for
Stochastic Differential Equations**

submitted by Raphael Abel Kasonga, B.Sc., Dip. App. Math., M.Sc., in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Dr. D. Dawson, Thesis Supervisor

Dr. W.J. Anderson, External Examiner

Dr. B. M. Puttaswamaiah, Chairman

Department of Mathematics and Statistics

Carleton University

December, 1986

To the memory of my brother

Martin Abel Kasonga

ABSTRACT

We study the asymptotic behaviour of Maximum Likelihood (ML) and Least Squares (LS)-type estimators of parameters from Stochastic Differential Equations (SDE's). First we give a brief review of contributions to the ML theory based on continuous sampling. Then we consider repeated sampling on the time interval $[0, T]$. The observation period T therefore, does not go to infinity as usually is the case, but instead, we allow the number of replications of the process to grow without bound. We prove that this new procedure yields consistent and asymptotically normal estimates. An example is given where the asymptotic covariance matrix is calculated explicitly. In addition, these results are valid not only for independent replications but also for exchangeable interacting systems. An application to testing the hypothesis of non-interaction for this particular example is also considered. Next, we discuss briefly the contribution of Le Breton on discrete sampling

On the LS methods, our main contribution is the proof of the strong consistency of the LS estimator for stationary processes. We also consider estimators derived from non-stationary processes. The special case of non-stationary Ornstein-Uhlenbeck process is dealt with in detail and we obtain the strong consistency of the LS estimator.

Finally, we report on simulation studies which yield the empirical Mean Square Errors of the various procedures presented in this study.

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Professor D.A. Dawson for his invaluable guidance and encouragement. His modesty and patience always rekindled my self-confidence and positive attitude.

I am also grateful to the administrators of the exchange program between Carleton University and the University of Dar-es-salaam for providing me with the opportunity to do graduate studies at Carleton. Special thanks to the Department of Mathematics and Statistics of Carleton University for the excellent academic environment.

Last, but not the least, I wish to thank my wife Nora and daughter Linda for their love, patience and support.

TABLE OF CONTENTS

	Page
Abstract	(ii)
Acknowledgements	(iii)
Notation	(vi)
Introduction	(vii)
 Chapter 1: PRELIMINARIES FROM THE THEORY OF SDE's	
1.1 Stochastic Integrals	1
1.2 Stochastic Differential Equations	5
1.3 Itô's Formula and Applications	7
1.4 Absolute Continuity of Measures Induced by Solutions of SDE's	10
1.5 Other Preliminaries	11
 Chapter 2: MAXIMUM LIKELIHOOD ESTIMATION	
2.1 Introduction	13
2.2 General Martingale Approach	14
2.3 SDE's with Linear Parametrization	17
2.4 Estimation for Large Systems of Equations	21
2.5 Exchangeable Interacting Diffusions	27
2.6 Testing of Hypothesis	33
2.7 Non-Linear Stochastic Differential Equations	34
2.8 Estimation from Discrete Data	36

Chapter 3: LEAST SQUARES METHODS

3.1	Introduction	38
3.2	Estimation from Stationary Processes	38
3.3	Strong Consistency of the LS Estimator	43
3.4	Non-Stationary Processes	53
3.5	The Ornstein-Uhlenbeck Process	57

Chapter 4: SIMULATION STUDIES

4.1	Introduction	64
4.2	Description of the Simulation Procedure	64
4.3	The Mean Field Model	66
4.4	The Ornstein-Uhlenbeck Process	70
4.5	Directions for Further Research	77

References	79
-------------------------	----

NOTATION

R^d ... d -dimensional Euclidean space

$R^{d \times d}$... The space of $d \times d$ -matrices with real entries

u^* ... The transpose of the vector u

Z_+^d ... The set of all d -tuples $(\alpha_1, \dots, \alpha_d)$ where $\alpha_j, j = 1, \dots, d$ are non-negative integers.

$|x|^{2m} = \sum_{j=1}^d x_j^{2m}$ for $x \in R^d$ and $m > 0$.

$\|G\|^{2m} = \sum_{k=1}^d \sum_{l=1}^n G_{kl}^{2m}$ for $G \in R^{d \times n}$ and $m > 0$.

$G^{1/2} = A$ when G has a decomposition $G = AA^*$ where $G \in R^{d \times d}$ and some $A \in R^{d \times n}$

$C(\Theta, R^d)$... The space of continuous functions on Θ with values in R^d .

χ_A ... The indicator function of the set A defined as $\chi_A(t) = 1$ if $t \in A$, $\chi_A(t) = 0$ otherwise.

For a sequence of random elements $X_n, n \in \mathbb{Z}_+$ and a random element X , the following denote

$\bar{X}_n \xrightarrow{p} X$... Convergence in probability

$X_n \xrightarrow{a.s.} X$... Almost sure convergence

$X_n \xrightarrow{D} X$... Convergence in distribution

$X_n \Rightarrow X$... Weak convergence

INTRODUCTION

In the study of inference problems for stochastic processes it has been found that a special class of processes, that is, those which are solutions of parametric Stochastic Differential Equations (SDE's) are amenable to statistical analysis. Large sample theory for Maximum Likelihood (ML), Least Squares (LS) and Bayes estimators of parameters of the drift function of such processes is well developed (cf. Brown and Hewitt [9], [10], Dorogovcev [13]; Prakasa Rao [34] and [35]; Taraskin [41]; etc.) One of the basic assumptions in the study of estimation of parameters from SDE's is that the process can be observed continuously in time. This assumption is not always appropriate, in fact it is impossible to achieve in practice. In view of this, it would be interesting to know whether the asymptotic properties already enjoyed by these various types of estimators are preserved if this assumption is weakened. Moreover, very often properties of the estimators are studied for large values of the time interval T . There are numerous examples from the physical sciences of phenomena which can be observed and measured for restrictively short time intervals. An alternative asymptotic procedure in this case is possible if the underlying stochastic system is composed of a large number of subsystems. This problem is the main preoccupation of Chapter 2. There we consider SDE's of the form

$$dX_j(t) = \sum_{k=1}^d \theta_k f_{j,k}(X(t))dt + g_j(X(t))dW_j(t) \quad (0.1)$$

where $X(t) = (X_1(t), \dots, X_N(t))^*$, $(W_j(t); t \geq 0)$, $j = 1, \dots, N$ are N independent Wiener processes and $(\theta_1, \dots, \theta_d)$ are the unknown parameters. u^* denotes the transpose of the vector u . The functions $f_{j,k}$, g_j , $j = 1, \dots, N$, $k = 1, \dots, d$ are assumed to be

known and non-random. Under some appropriate conditions (cf. Chapter 1, Theorem 3) the solution $(X(t); t \geq 0)$ of (0.1) is a Markov diffusion process in R^N . Suppose $\hat{\theta}_N$ is the ML estimator of $\theta = (\theta_1, \dots, \theta_d)$ based on the observations $(X_j(t); 0 \leq t \leq T, j=1, \dots, N)$. Our main result in Chapter 2 is that for any fixed T , $\hat{\theta}_N$ is consistent and $N^{1/2}(\hat{\theta}_N - \theta)$ is asymptotically normal as $N \rightarrow \infty$. The proof of this result is accomplished via the identity

$$I_N(t) (\hat{\theta}_N - \theta) = M_N(t), \quad (0.2)$$

the martingale inequalities and Rebolledo's central limit theorem [36]. In (0.2) $I_N(t)$ is a $d \times d$ -matrix consisting of ordinary integrals of functions of $(X(s), s \leq t)$ while $(M_N(t), t \geq 0)$ is an R^d -valued martingale.

One interesting example of the system (0.1) is the case where the subsystems are interacting and exchangeable. For instance consider a simple mean-field model described by the solution of an Itô SDE

$$dX_j(t) = \alpha X_j(t) dt - \beta (X_j(t) - \bar{X}_N(t)) dt + dW_j(t) \quad j = 1, \dots, N \quad (0.3)$$

where $\bar{X}_N(t) = N^{-1} \sum_{j=1}^N X_j(t)$ and $\beta > 0$. The term containing the parameter β in (0.3) can be viewed as an interaction between subsystems which create a tendency for the subsystems to relax toward the center of gravity of the ensemble. Thus the system provides a simple example of a cooperative interaction. In section 2.5 we derive the ML estimator $\hat{\theta}_N$ of $\theta = (\alpha, \beta)$ and study its properties for large N . One possible application of the asymptotic distribution of $N^{1/2}(\hat{\theta}_N - \theta)$ could be to test the hypothesis of non-interaction among the subsystems, that is $H_0: \beta = 0$. This is briefly discussed in section 2.6.

Returning to the question of relaxing the assumption of continuous observation, the first attempt in this direction was that of Le Breton [8]. He considered ML-type estimators

derived from observations of a linear system at discrete time points. He showed that by taking observations on a finer partition of the time interval, these estimators converge to the ML estimators. A more detailed discussion of this work is given in section 2.8

In Chapter 3 we study Least Squares type procedures based on continuous-discrete data. By continuous-discrete data we mean observations $X(t)$ made at times $t \in \mathcal{W}$, where $\mathcal{W}$ is some discrete subset of $R^+ = [0, \infty)$ and $X(t)$ is continuous in t (cf. [40]). This estimation problem has been studied by Dorogovcev [13], Prakasa Rao [35] and others (see section 3.1). In sections 3.1-3.3 we consider stochastic equations of the form

$$dX(t) = f(\theta, X(t))dt + dW(t), \quad (0.4)$$

where $(W(t), t \geq 0)$ is the standard Wiener process. The function f is non-random and such that (0.4) yields a stationary solution $(X(t), t \geq 0)$. The assumption of stationarity is very common in literature basically because it simplifies the asymptotic theory through the use of ergodic theorems. Throughout Chapter 3, $\mathcal{W} = \{0, \Delta, 2\Delta, \dots, n\Delta\}$ where $\Delta = T/n$ for some real number T and a positive integer n . For a special case of (0.4) where $f(\theta, X(t)) = \theta$, Stoyanov [40] examined estimators based on a variety of schemes $\mathcal{W}$. One of our main contributions in Chapter 3 is the proof of consistency of the LS estimator $\hat{\theta}_{n,T}$, as $T \rightarrow \infty$ and $T/n \rightarrow 0$.

Since the assumption of stationarity is not always appropriate we propose in section 3.4 an alternative estimation procedure. We shall use the term *modified LS* for this new procedure. The key idea there is the approximation of the solution $(X(t), 0 \leq t \leq T)$ of (0.4) by a piecewise deterministic analogue $(U(\theta, t), 0 \leq t \leq T)$. The latter is a solution of the ordinary differential equation

$$dU_k(\theta, t) = f(\theta, U_k(\theta, t))dt \quad t \in [t_{k-1}, t_k)$$

$$U_k(\theta, t_{k-1}) = X(t_{k-1}) \quad k = 1, \dots, n \quad (0.5)$$

$$\text{and} \quad U(\theta, t) = \sum_{k=1}^n U_k(\theta, t) \chi_{(t_{k-1}, t_k)}(t)$$

where $t_k = k\Delta$, $k = 0, \dots, n$ and $\chi_A(\cdot)$ is the indicator function of the set A , that is $\chi_A(t) = 1$ if $t \in A$ and $\chi_A(t) = 0$ otherwise. We prove that the modified LS estimator is consistent. This result does not require equally spaced observation times. In section 3.5 we consider a non-stationary Ornstein-Uhlenbeck process, a SDE described by (0.4) with $f(\theta, X(t)) = \theta X(t)$ and $\theta > 0$. In this case we prove that the modified LS estimator is strongly consistent.

The thesis is organized as follows. In Chapter 1 we summarise some basic concepts and results from the theory of SDE's. The presentation is selectively done to suit the needs of the subsequent chapters. Chapter 2 deals with ML procedures based on continuous sampling. A brief review of literature is done before embarking on the main topic of estimation for large systems of equations. In Chapter 3 we study some LS-type procedures based on continuous-discrete data. This chapter is divided into two parts. The first one deals with stochastic systems which are stationary. The second deals with non-stationary systems. Finally in Chapter 4, we report on simulation studies which yield the empirical mean square errors of the various procedures presented in this study.

Chapter 1

PRELIMINARIES FROM THE THEORY OF SDE's

This Chapter is mostly a collection of known facts from the theory of stochastic differential equations that will be needed in the subsequent Chapters. In a few cases we shall prove results which follow easily from the general theory but are not stated explicitly anywhere in the literature.

1: Stochastic Integrals

Consider a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t, t \in [0, T]\}$. A stochastic process $X = \{X_t; t \in [0, T]\}$ is said to be measurable if the function $(t, \omega) \mapsto X_t(\omega)$ is $\mathcal{B}[0, T] \times \mathcal{F}$ -measurable, where $\mathcal{B}[0, T]$ is the Borel σ -algebra on $[0, T]$. The process X is adapted if for each $t \in [0, T]$, the random variable $X_t : \Omega \rightarrow X_t(\omega)$ is $\mathcal{F}_t$ -measurable. A non-anticipating stochastic process is the one which is both measurable and adapted. We shall define the stochastic integral with respect to the Wiener process for non-anticipating functions G which satisfy;

$$\Pr \left\{ \omega; \int_0^T \|G(s, \omega)\|^2 ds < \infty \right\} = 1 \quad (1.1)$$

Denote by $H_2[0, T]$ the collection of all such functions G . Throughout this section, G will be a $d \times n$ -matrix-valued function so that

$$\|G\| = \left(\sum_{k=1}^d \sum_{l=1}^n G_{kl}^2 \right)^{1/2} = (tr GG^*)^{1/2},$$

where $tr A$ = trace of the matrix A , and $G = (G_{kl})$, $k = 1, \dots, d$, $l = 1, \dots, n$. Denote by $M_2[0, T]$ the class of non-anticipating functions G such that

$$\int_0^T E \|G(s, \omega)\|^2 ds < \infty \quad (1.2)$$

Obviously $M_2[0, T] \subset H_2[0, T]$. Let $G \in H_2[0, T]$ be a step function, that is, there exists a partition $0 = t_0 < t_1 < \dots < t_m = T$ of $[0, T]$ such that

$$G(t, \omega) = G_{t_{k-1}}(\omega) \quad \text{if } t \in [t_{k-1}, t_k), \quad k = 1, \dots, m.$$

In this case we define

$$\int_0^T G_s dW_s = \sum_{k=1}^m G_{t_{k-1}} (W_{t_k} - W_{t_{k-1}}) \quad (1.3)$$

Where $W = \{W_t; t \in [0, T]\}$ is a Wiener process in R^n . The following Theorem states the basic properties of the stochastic integral defined by (1.3).

Theorem 1: Suppose $G, G_1, G_2 \in H_2[0, T]$ and let α, β be real numbers. Then

$$(a) \quad \int_0^T (\alpha G_1 + \beta G_2) dW = \alpha \int_0^T G_1 dW + \beta \int_0^T G_2 dW$$

$$(b) \quad \int_0^T G dW = \begin{pmatrix} \sum_{k=1}^n \int_0^T G_{1k}(s) dW_k(s) \\ \vdots \\ \sum_{k=1}^n \int_0^T G_{dk}(s) dW_k(s) \end{pmatrix}$$

(c) If $E\|G(s)\| < \infty$ for all $s \in [0, T]$ we have

$$E \left(\int_0^T G dW \right) = 0$$

(d) If $E\|G(s)\|^2 < \infty$, $s \in [0, T]$ the following holds

$$E \left(\int_0^T G dW \right) \left(\int_0^T G dW \right)^* = \int_0^T E[G(s)G(s)^*] ds$$

and in particular

$$E \left| \int_0^T G dW \right|^2 = \int_0^T E\|G_s\|^2 ds$$

$$(e) \quad \Pr \left\{ \left| \int_0^T G dW \right| > C \right\} \leq \frac{N}{C^2} + \Pr \left\{ \int_0^T \|G_s\|^2 ds > N \right\}$$

The proof of Theorem 1 can be found in [3]. The step functions are dense in $H_2[0, T]$ in the sense that for each $G \in H_2[0, T]$, there exist a sequence of step functions $G_n \in H_2[0, T]$, $n = 1, 2, \dots$, such that:

$$\int_0^T \|G(s) - G_n(s)\|^2 ds \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty \quad (1.4)$$

By (1.4) it can be shown that (e.g. [3]) there exists a random variable $I_T(G)$ such that

$$\lim_{n \rightarrow \infty} \int_0^T G_n(s) dW_s = I_T(G) \quad (1.5)$$

in probability. $I_T(G)$ is unique with probability-1, and independent of the choice of the sequence G_n which satisfies (1.4). Hence for $G \in H_2[0, T]$ we define

$$\int_0^T G(s) dW_s = I_T(G) \quad t \in [0, T] \quad (1.6)$$

The integral defined by (1.6) also enjoys properties (a)-(e) from Theorem 1. Moreover, if $G \in M_2[0, T]$ then there exists a sequence of step functions $G_n \in M_2[0, T]$, $n = 1, 2, \dots$, such that

$$\lim_{n \rightarrow \infty} \int_0^T E \|G_n(s) - G(s)\|^2 ds = 0$$

and

$$\lim_{n \rightarrow \infty} E \left| \int_0^T G_n(s) dW_s - \int_0^T G(s) dW_s \right|^2 = 0$$

The stochastic integral as a function of the upper limit

Again let $W = \{W_t; t \in [0, T]\}$ denote an n -dimensional Wiener process with a filtration $\{\mathcal{F}_t; t \in [0, T]\}$. Suppose $G \in H_2[0, T]$, $A \subset [0, T]$ and χ_A is an indicator function of the set A . Then $G\chi_A \in H_2[0, T]$. We therefore define

$$\int_A G dW = \int_0^T G\chi_A dW \quad (1.7)$$

By Theorem 1 part(a), if $A, B \subset [0, T]$ are disjoint sets, then

$$\int_{A \cup B} G dW = \int_A G dW + \int_B G dW$$

In particular, for $0 \leq a \leq b \leq c \leq T$

$$\int_a^c G dW = \int_a^b G dW + \int_b^c G dW.$$

Theorem 2: Let $G \in H_2[0, T]$ and suppose that

$$X_t = \int_0^t G_s dW_s, \quad t \in [0, T]$$

Then the following hold:

(a) X_t is $\mathcal{F}_t$ -measurable.

(b) If $G \in M_2[0, T]$ then $(X_t, \mathcal{F}_t; t \in [0, T])$ is an R^d -valued martingale. Furthermore

for $s, t \in [0, T], s < t$

$$E[X_t] = 0$$

$$E[X_t X_s^*] = \int_0^s E[G_u G_u^*] du,$$

In particular

$$E|X_t|^2 = \int_0^t E\|G_u\|^2 du.$$

(c) For all $c > 0$ and $0 \leq a \leq b \leq T, G \in M_2[0, T]$

$$\Pr \left\{ \sup_{a \leq t \leq b} |X_t - X_a| > c \right\} \leq \frac{1}{c^2} \int_a^b E\|G_s\|^2 ds,$$

and

$$E \left[\sup_{a \leq t \leq b} |X_t - X_a|^2 \right] \leq 4 \int_a^b E\|G_s\|^2 ds.$$

(d) X_t has continuous sample functions with probability-1.

(e) If for some natural number m

$$\int_a^T E\|G_s\|^{2m} ds < \infty$$

then

$$E|X_t - X_a|^{2m} \leq [m(2m-1)]^{m-1} (t-a)^{m-1} \int_a^t E\|G_s\|^{2m} ds \quad (1.8)$$

Proof: (a)-(d) are proved in Arnold [8], page 81. (e) is proved via Itô's Formula (below) and Hölder's inequality. (see for example Liptser and Shiriyayev [27], Lemma-4.12). ()

2: Stochastic Differential Equations

We shall consider Itô's stochastic differential equation which has the form

$$dX_t = f(t, X_t)dt + G(t, X_t)dW_t \quad X_0 = x_0, \quad t \in [0, T] \quad (1.9)$$

Where $X = \{X_t; t \in [0, T]\}$ is an R^d -valued stochastic process and $W = \{W_t; t \in [0, T]\}$ is an n -dimensional Wiener process. The R^d -valued function $f(\cdot, \cdot)$ and the $d \times n$ -matrix-valued function $G(\cdot, \cdot)$ are assumed to be defined and measurable on $[0, T] \times R^d$. We also assume that the functions a, G depend on the random parameter ω only through the process X as in (1.9).

A stochastic process X is called a solution of equation (1.9) if the following holds:

- (i) X_t is $\mathcal{F}_t$ -measurable
- (ii) The functions $\tilde{f}(t, \omega) = f(t, X_t(\omega))$ and $\tilde{G}(t, \omega) = G(t, X_t(\omega))$ are non-anticipating and satisfy

$$\int_0^T |\tilde{a}(s, \omega)| ds < \infty$$

and

$$\int_0^T \|\tilde{G}(s, \omega)\|^2 ds < \infty$$

with probability-1.

- (iii) The following integral equation holds for every $t \in [0, T]$ with probability-1

$$X_t = x_0 + \int_0^t f(s, X_s)ds + \int_0^t G(s, X_s)dW_s \quad (1.10)$$

The following Theorem gives sufficient conditions under which equation (1.9) has a unique continuous solution.

Theorem 3: Suppose that in (1.9), x_0 is independent of $W_t - W_0$ for $t \in [0, T]$ and the functions $(t, x) \mapsto f(t, x)$, $(t, x) \mapsto G(t, x)$ are defined and measurable on $[0, T] \times R^d$ and satisfy the following properties:

There exists a constant $K > 0$ such that

(a) (Lipschitz condition) For all $t \in [0, T]$, $x, y \in R^d$

$$|f(t, x) - f(t, y)| + \|G(t, x) - G(t, y)\| \leq K|x - y|$$

(b) (Growth restriction) for all $t \in [0, T]$, $x \in R^d$

$$|f(t, x)|^2 + \|G(t, x)\|^2 \leq K^2(1 + |x|^2)$$

Then equation (1.9) has on $[0, T]$ a unique R^d -valued solution $X = \{X_t; t \in [0, T]\}$, continuous with probability-1 and such that $X_0 = x_0$. In fact X is a Markov process with transition probabilities given by

$$P(s, x, t, B) = \Pr(X_t \in B | X_s = x) = \Pr(X_t(s, x) \in B),$$

where $X_t(s, x)$ is the unique solution of

$$dX_t = f(t, X_t)dt + G(t, X_t)dW_t, \quad X_s = x, \quad 0 \leq s \leq t \leq T.$$

If in addition the functions $f(t, x)$ and $G(t, x)$ are continuous in t , then X is a d -dimensional diffusion with drift vector $f(t, x)$ and diffusion matrix $D(t, x) = G(t, x)G(t, x)^*$.

The proof of Theorem 3 can be found in Arnold [3], Liptser and Shiriyayev [27], etc.. The Theorem remains valid if the Lipschitz condition is replaced by:

(a¹)(Local Lipschitz condition) There exists an $N > 0$ and a constant K_N such that for all $t \in [0, T]$, $|x| \leq N$ and $|y| \leq N$ we have

$$|f(t, x) - f(t, y)| + \|G(t, x) - G(t, y)\| \leq K_N |x - y|. \quad (1.11)$$

3: Itô's formula and applications

The following Theorem is fundamental in the analysis of SDE's.

Theorem 4: Let $u: (t, x) \mapsto u(t, x)$ denote a continuous function defined on $[0, T] \times \mathbb{R}^d$ with values in $\mathbb{R}^p$ and with the continuous partial derivatives

$$\begin{aligned} \frac{\partial}{\partial t} u(t, x) &= u_t \\ \frac{\partial}{\partial x_k} u(t, x) &= u_{x_k}, \quad x = (x_1, \dots, x_d)^*, \quad k = 1, \dots, d \\ \frac{\partial^2}{\partial x_k \partial x_l} u(t, x) &= u_{x_k x_l}, \quad k, l = 1, \dots, d \end{aligned}$$

If the d -dimensional stochastic process X is defined on $[0, T]$ by the equation

$$dX_t = f(t)dt + G(t)dW_t,$$

where W is an n -dimensional Wiener process, then the p -dimensional process

$$Y_t = u(t, X_t)$$

defined on $[0, T]$ with initial value $Y_0 = u(0, X_0)$ satisfies

$$dY_t = \left\{ u_t(t, X_t) + u_x(t, X_t)f(t) + \frac{1}{2} \sum_{k=1}^d \sum_{l=1}^d u_{x_k x_l}(t, X_t) [G_t G_t^*]_{kl} \right\} dt + u_x(t, X_t) G_t dW_t,$$

where $u_x = (u_{x_1}, \dots, u_{x_d})$ is a $p \times d$ matrix and $u_{x_k x_l}$ is a p -dimensional column vector.

For proof of the above result, see for example Arnold [3].

One application of Itô's formula is to establish estimates for the moments of the solution of (1.9).

Theorem 5: Assume that the coefficients of (1.9) satisfy conditions (a) and (b) (or (a¹) and (b)) of Theorem 3 and that $E|x_0|^{2m} < \infty$ where m is a positive integer. Then there exists constants C and D depending only on m, K and T such that

$$E|X_t|^{2m} \leq (1 + E|x_0|^{2m}) e^{Ct} \quad (1.12)$$

$$E|X_t - x_0|^{2m} \leq D (1 + E|x_0|^{2m}) t^m e^{Ct}, \quad (1.13)$$

where $t \in [0, T]$, $C = 2m(2m + 1)K^2$.

Proof: In applying Theorem 4 the function u is chosen to be

$$\begin{aligned} u(t, x) &= |x|^{2m} \\ &= \sum_{k=1}^d x_k^{2m} \end{aligned}$$

Details of the proof can be found in Arnold [3]. □

An immediate consequence of Theorem 5 is that if $T < \infty$, then

$$E|X_t|^{2m} < \infty, \quad \text{for all } t \in [0, T]$$

whenever $E|x_0|^{2m} < \infty$. This together with the method of proof of Theorem 5 implies the following:

$$E|X_t - X_s|^{2m} \leq \tilde{C}(t - s)^m \quad 0 \leq s \leq t \leq T \quad (1.14)$$

where the constant $\tilde{C}$ depends only on m, K and T . We shall apply the inequality (1.14) in Chapter 3. However we need to have the constant $\tilde{C}$ which does not depend on T . The assumption that X is stationary in the wide sense is sufficient for such requirement.

Theorem 6: In addition to the hypothesis of Theorem 5 if the solution X of (1.9) is stationary, then for $s, t \in [0, T]$ such that $|t - s| < 1$ we have

$$E|X_t - X_s|^{2m} \leq C(t - s)^m \quad (1.15)$$

where the constant C depends only on m, d and K .

Proof: The growth restriction condition and Hölder's inequality implies

$$|f(t, x)|^{2m} + \|G(t, x)\|^{2m} \leq 2^{m-1} K^m (1 + d^{m-1} |x|^{2m}) \quad (1.16)$$

where the left hand side of (1.16) is as defined on page (v). By (1.10)

$$X_t = x_0 + \int_0^t f(u, X_u) du + \int_0^t G(u, X_u) dW_u,$$

so that

$$\begin{aligned} |X_t - X_s|^{2m} &\leq 2^{2m-1} \left\{ \left| \int_s^t f(u, X_u) du \right|^{2m} + \left| \int_s^t G(u, X_u) dW_u \right|^{2m} \right\} \\ &\leq 2^{2m-1} \left\{ (t-s)^{2m-1} \int_s^t |f(u, X_u)|^{2m} du + \left| \int_s^t G(u, X_u) dW_u \right|^{2m} \right\} \end{aligned}$$

by Hölder's inequality. Taking expectations on both sides we obtain

$$E|X_t - X_s|^{2m} \leq 2^{2m-1} \left\{ (t-s)^{2m-1} \int_s^t E|f(u, X_u)|^{2m} du + E \left| \int_s^t G(u, X_u) dW_u \right|^{2m} \right\}$$

The fact that $|t-s| < 1$ together with (1.8), (1.9) and stationarity implies

$$\begin{aligned} E|X_t - X_s|^{2m} &\leq 2^{2m-1} \left\{ (t-s)^{2m-1} \int_s^t 2^{m-1} K^m (1 + d^{m-1} E|X_u|^{2m}) du \right. \\ &\quad \left. + [m(2m-1)]^{m-1} (t-s)^{m-1} \int_s^t E\|G(u, X_u)\|^{2m} du \right\} \\ &\leq 2^{2m-1} \left\{ (t-s)^{2m-1} 2^{m-1} K^m (1 + d^{m-1} E|X_0|^{2m}) (t-s) \right. \\ &\quad \left. + [m(2m-1)]^{m-1} (t-s)^{m-1} 2^{m-1} K^m (1 + d^{m-1} E|X_0|^{2m}) (t-s) \right\} \\ &\leq 2^{3m-2} K^m (1 + d^{m-1} E|X_0|^{2m}) (1 + [m(2m-1)]^{m-1}) (t-s)^m \\ &= C(d, m, K)(t-s)^m \quad \square \end{aligned}$$

Remark: The case $|t-s| < 1$ in Theorem 6 is of interest to Chapter 3. However, if

$|t-s| \geq 1$, it is easy to see from the proof above that

$$E|X_t - X_s|^{2m} \leq C(t-s)^{2m} \quad C = C(d, m, K)$$

4: Absolute continuity of measures induced by solutions of SDE's

In this section we present a basic Theorem that will enable us to calculate likelihood functions in Chapter 2. Details of the formulae given below will be omitted. They can be found in Liptser and Shirayev [27], Chapter 7.

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, $F = (\mathcal{F}_t; t \in [0, T])$ be a non-decreasing family of sub- σ -algebras of $\mathcal{F}$ and $W = \{W_t; t \in [0, T]\}$ be an n -dimensional Wiener process. Denote by $(C_T = C([0, T], R^d), \mathcal{B}_T)$ the measurable space of continuous functions with values in R^d . $\mathcal{B}_T$ is the σ -algebra on $C[0, T]$ generated by $\{x_t^{-1}(B); t \in [0, T], x_t \in C[0, T], B \text{ a Borel set in } R^d\}$

Let $X = \{X_t; t \in [0, T]\}$ and $Z = \{Z_t; t \in [0, T]\}$ be processes on R^d having the differentials

$$\begin{aligned} dX_t &= a_t(X)dt + G_t(X)dW_t & X_0 &= Z_0 \\ dZ_t &= b_t(Z)dt + G_t(Z)dW_t & \sum_{k=1}^d |Z_k(0)| &< \infty \text{ a.s.} \end{aligned} \quad (1.17)$$

where $a_t(x)$ and $b_t(x)$ are R^d -valued functions and $G_t(x)$ a $d \times n$ -matrix-valued function. Denote by μ_X and μ_Z the measures on $(C_T, \mathcal{B}_T)$ generated by processes X and Z respectively. That is

$$\mu_X(B) = \Pr(X \in B)$$

$$\mu_Z(B) = \Pr(Z \in B) \quad B \in \mathcal{B}_T$$

Suppose that the system of algebraic equations

$$G_t(x)\alpha_t(x) = a_t(x) - b_t(x)$$

has a bounded solution $\alpha_t(x)$ for each $t \in [0, T]$, $x \in C_T$.

Theorem 7: Suppose the functionals $b_t(x)$ and $G_t(x)$ satisfy conditions (a) and (b) of Theorem 3. If $\mu_X = a.s.$

$$\int_0^T a_t^*(x) [G_t(x)G_t^*(x)]^- a_t(x) + b_t^*(x) [G_t(x)G_t^*(x)]^- b_t(x) dt < \infty \quad (1.19)$$

where A^- denotes the pseudo-inverse of the matrix A . Then $\mu_X \ll \mu_Z$. If in addition (1.19) is satisfied $\mu_Z = a.s.$ then $\mu_X \sim \mu_Z$ and

$$\begin{aligned} \frac{d\mu_X}{d\mu_Z}(t, X) = \exp \left\{ \int_0^t [a_s(X) - b_s(X)]^* [G_s(X)G_s^*(X)]^- dX_s \right. \\ \left. - \frac{1}{2} \int_0^t [a_s(X) - b_s(X)]^* [G_s(X)G_s^*(X)]^- [a_s(X) + b_s(X)] ds \right\} \end{aligned} \quad (1.20)$$

5: Other Preliminaries

The following result will enable us to solve some simple integral equations and also obtain some L^2 -bounds of solutions of SDE's.

Theorem 8: (Gronwall's inequality) Let $\alpha(t)$ and $\beta(t)$ be Lebesgue integrable functions on $[a, b]$ and suppose γ is a constant such that

$$\alpha(t) \leq \beta(t) + \gamma \int_a^t \alpha(s) ds \quad \text{for } t \in [a, b] \quad (1.21)$$

Then

$$\alpha(t) \leq \beta(t) + \gamma \int_a^t e^{\gamma(t-s)} \beta(s) ds \quad (1.22)$$

Moreover, if the inequality in hypothesis (1.21) is replaced by an equality, then the conclusion (1.22) also follows with an equality.

Proof: Let $A(t) = \int_a^t \alpha(s) ds$. The function $g(t) = A(t)e^{-\gamma t}$ is absolutely continuous on $[a, b]$ and

$$g'(t) = \frac{d}{dt} g(t) = \alpha(t)e^{-\gamma t} - \gamma A(t)e^{-\gamma t} \quad (a.e.)$$

Since

$$g'(t) \leq \beta(t)e^{-\gamma t} \quad (\text{a.e.}) \text{ on } [a, b],$$

$$g(t) - g(a) = \int_a^t g'(s) ds \leq \int_a^t \beta(s)e^{-\gamma s} ds, \quad t \in [a, b].$$

Thus

$$A(t) \leq e^{-\gamma t} \int_a^t \beta(s)e^{-\gamma s} ds,$$

and hence

$$\alpha(t) \leq \beta(t) + \gamma A(t) \leq \beta(t) + \gamma \int_a^t \beta(s)e^{-\gamma(s-a)} ds$$

The proof of the second part of the Theorem is obtained by replacing inequalities by equalities everywhere in the above arguments. $\square$

Finally, we present a powerful bound for the stochastic integral defined by (1.6)

Theorem 9: Let $G \in H_2[0, T]$. Then

$$Z(t) = \exp \left\{ \int_0^t G_s dW_s - \frac{1}{2} \int_0^t G_s^2 ds \right\} \quad (1.23)$$

is a supermartingale and

$$\Pr \left\{ \sup_{0 \leq t \leq T} \int_0^t G_s dW_s - \frac{\alpha}{2} \int_0^t G_s^2 ds > \beta \right\} \leq e^{-\alpha\beta}. \quad (1.24)$$

Proof: See McKean [31], page 25.

Chapter 2

MAXIMUM LIKELIHOOD ESTIMATION

1: Introduction

The theory of maximum likelihood (ML) estimation for continuous-time processes has been studied extensively. The aim of this chapter is to study those estimation problems which deal with stochastic differential equation (SDE) models. Specifically, we shall investigate the asymptotic properties of parameter estimates from the following model

$$dX_t = f(\theta, t, X)dt + G^{1/2}(t, X)dW_t \quad (2.1)$$

where $f(\theta, t, x)$ and $G(t, x)$ are known functions with values in R^N and $R^{m \times N}$ respectively. $\theta = (\theta_1, \dots, \theta_d)$ is a vector of unknown parameters to be estimated. $W = (W_t, t \geq 0)$ is an m -dimensional Wiener process with independent components. The solution $X = (X_t, t \geq 0)$ is a continuous stochastic process in R^N .

In section 2 we give a brief account of the general ML estimation theory for continuous time stochastic processes. This is based mainly on the work of Feigin [14]. Many studies on SDE models assume that the drift coefficient $f(\theta, t, x)$ is linear in the parameter θ (cf Brown and Hewitt [9], [10], Guttorp and Kulperger [18], Sørensen [39] etc.), that is

$$f(\theta, t, x) = \sum_{k=1}^d \theta_k f_k(t, x) \quad t \geq 0, \quad x \in R^N \quad (2.2)$$

and $f_k(t, x)$, $k = 1, \dots, d$ have values in R^N . From now on, SDE-LP (SDE's-LP) shall stand for *Stochastic Differential Equation(s) with Linear Parametrization*. This linearity assumption simplifies the asymptotic analysis considerably due to the fact that the SDE's-LP, form a class called *conditional exponential family* introduced by Feigin [14]. The ML

estimation theory for SDE's-LP is treated in its greatest generality by Taraskin [41]. This is discussed in section 3.

In sections 4,5 and 6 we study the asymptotic properties of ML estimators for homogeneous interacting systems of N SDE's-LP. This has applications to the mean field limit of interacting diffusions (cf. Dawson [12], Léonard [26]). In addition the results obtained in those sections also apply to the case in which we observe N replications of the process on the interval $[0, T]$. Modelling the drift coefficient as in (2.2) may not always be desirable (cf. Borkar and Bagchi [7]). In section 7 we shall study systems which are more general than SDE's-LP. Finally in section 8, we briefly mention the work of Le Breton [8]. As far as we know, this contribution by Le Breton is the first attempt ever to study ML estimators derived from discrete sampling.

2: General Martingale Approach

The aim of this section is to introduce classes of stochastic processes called *conditional exponential families* and demonstrate their significance in the ML estimation theory.

Preliminaries

Let $(\Omega, \mathcal{F}, P_\theta; \theta \in \Theta)$ be a class of probability spaces in which the index θ is a parametrization of the family of possible probability measures, and $\Theta \subset R^d$. We assume that, for all $\theta_1, \theta_2 \in \Theta$, P_{θ_1} is equivalent to P_{θ_2} and $(\Omega, \mathcal{F}, P_{\theta_1})$ is complete. The process $X = (X_t, t \geq 0)$ on $(\Omega, \mathcal{F})$ evolves according to one of the probability laws P_θ , let us say P_{θ^0} , that is θ^0 is the true parameter. Let $F = \{\mathcal{F}_t; t \geq 0\}$ be a non-decreasing complete family of sub- σ -algebras of $\mathcal{F}$ generated by X . We consider processes X which have right continuous sample paths almost surely, so that F is also right continuous.

We write P_θ^t for the restriction of P_θ to $\mathcal{F}_t$ and define the likelihood function

$$L_t(\theta) = \frac{dP_\theta^t}{dP_{\theta^0}^t} \quad \text{for all } t \geq 0$$

as a Radon-Nikodym derivative. Note that if $L_t(\theta)$ is calculated with respect to $P_{\theta_1}^t$ instead of $P_{\theta^0}^t$, where $\theta_1 \neq \theta^0$, then the likelihood would change only by a multiplicative factor independent of θ . Hence as far as inference on θ is concerned, $L_t(\theta)$ is independent of the dominating measure. Also, by the Radon-Nikodym Theorem, $L_t(\theta)$ is $\mathcal{F}_t$ -measurable and so, almost surely, can be expressed as a function of the sample path $\{X_s; 0 \leq s \leq t\}$. This is a necessary requirement if we wish to base our estimation on the observations of X on $[0, t]$.

Now let

$$\begin{aligned} l_t(\theta) &= \log L_t(\theta) \\ U_t(\theta) &= \frac{\partial}{\partial \theta} l_t(\theta) \end{aligned} \quad (2.3)$$

where for $\theta \in R^d$, the RHS of (2.3) denotes $\left(\frac{\partial}{\partial \theta_1} l_t(\theta), \dots, \frac{\partial}{\partial \theta_d} l_t(\theta) \right)$. Under mild regularity conditions $\{U_t(\theta), \mathcal{F}_t, t \geq 0\}$ is a zero mean martingale for each $\theta \in \Theta$ (see [14] for details). This is the key tool with which asymptotic properties of the ML estimators can be determined.

Conditional Exponential Families

Suppose that $\{U_t(\theta), \mathcal{F}_t; t \geq 0\}$ is a real zero mean square integrable martingale. Then by Liptser and Shirayev [27], there exist an increasing process $\{I_t(\theta); t \geq 0\}$, such that

$$E(|U_t(\theta) - U_s(\theta)|^2 | \mathcal{F}_s) = E(I_t(\theta) - I_s(\theta) | \mathcal{F}_s) \quad \text{a.s.} \quad (2.4)$$

for all $0 \leq s \leq t$. The family of probability measures $\{P_\theta; \theta \in \Theta\}$ on $(\Omega, \mathcal{F})$ is said to be a conditional exponential family if the following factorization holds

$$U_t(\theta) = I_t(\theta)(\hat{\theta}_t - \theta) \quad t \geq 0, \theta \in \Theta \quad (2.5)$$

where the ML estimate $\hat{\theta}_t$ is defined to be the solution of

$$U_t(\theta) = 0 \quad (2.6)$$

We assume that (2.6) has a unique solution almost surely, for each t and that this solution maximizes the likelihood. The advantage of the factorization (2.5) is that we can now study the behaviour of $\hat{\theta}_t$ via the martingale $\{U_t(\theta), \mathcal{F}_t; t \geq 0\}$.

Example 1: Let the real process $X = (X_t; t \geq 0)$ be a solution of

$$dX_t = \theta X_t dt + dW_t, \quad t \geq 0, \quad \theta \in R^1 \quad (2.7)$$

where $W = (W_t; t \geq 0)$ is a Wiener process. Let P_θ and P_{θ_0} be the measures induced by X and W on $C_T(R^1)$ respectively. Then by Theorem(1.7)

$$\begin{aligned} L_t(\theta) &= \frac{dP_\theta^t}{dP_{\theta_0}^t} \\ &= \exp \left\{ \theta \int_0^t X_s dX_s - \frac{\theta^2}{2} \int_0^t X_s^2 ds \right\} \end{aligned}$$

and

$$U_t(\theta) = \int_0^t X_s dX_s - \theta \int_0^t X_s^2 ds$$

so that the ML estimator $\hat{\theta}_t$ is given by

$$\hat{\theta}_t = \frac{\int_0^t X_s dX_s}{\int_0^t X_s^2 ds}$$

The above implies

$$\begin{aligned} \hat{\theta}_t \int_0^t X_s^2 ds &= \int_0^t X_s dX_s \\ &= \theta \int_0^t X_s^2 ds + \int_0^t X_s dW_s \quad \text{by (2.7)} \end{aligned}$$

hence

$$\int_0^t X_s dW_s = \int_0^t X_s^2 ds \cdot (\hat{\theta}_t - \theta)$$

which yields the factorization (2.5), since in this case $I_t(\theta) = \int_0^t X_s^2 ds$. By Liptser and Shiryayev ([28] page 210)

$$\frac{\int_0^t X_s dW_s}{\int_0^t X_s^2 ds} \xrightarrow{p} 0 \quad \text{as } t \rightarrow \infty$$

Therefore $\hat{\theta}_t$ is strongly consistent. Moreover, from Theorem 1 below we have

$$t^{-1/2} \int_0^t X_s dW_s \xrightarrow{D} N(0, \sigma^2)$$

if

$$t^{-1} \int_0^t X_s^2 ds \xrightarrow{p} \sigma^2$$

and

$$t^{-1} \int_0^t E(X_s^2) ds \rightarrow \sigma^2$$

More examples are given in Feigin [14]. In the next two sections we will show that (2.5) holds in general for SDE's-LP and use the fact to prove some asymptotic properties of the ML estimator.

3: Stochastic Differential Equations with Linear Parametrization

Throughout this section we shall consider a special case of (2.1) in which $f(\theta, t, x)$ is as defined in (2.2) and $G(t, x)$ invertible for each $(t, x) \in [0, T] \times R^N$, $T \geq 0$, namely,

$$dX_t = \sum_{k=1}^d \theta_k f_k(t, X_t) dt + G^{1/2}(t, X_t) dW_t, \quad (2.8)$$

where $f_k(t, x)$; $k = 1, \dots, d$ are linearly independent for each $(t, x) \in [0, T] \times R^N$. Consider the equation.

$$dZ_t = G^{1/2}(t, Z_t) dW_t, \quad Z_0 = X_0 \quad (2.9)$$

Let P_θ^T and P_0^T be the probability measures on $\mathcal{F}_T$ generated by the solutions of (2.8) and (2.9) respectively. If

$$\sum_{k=1}^d \sum_{l=1}^d \theta_k \theta_l \int_0^T f_k^*(s) G_s^{-1} f_l(s) ds < \infty \quad (2.10)$$

as $T \rightarrow \infty$ and $n^{-1/2}T \rightarrow 0$.

Asymptotic efficiency

Let μ_θ^T be the measure induced by the process $\{X_t, 0 \leq t \leq T\}$ satisfying (3.1) and μ_W^T be the Wiener measure on $C([0, T], R)$. By Theorem 1.7, μ_θ^T is equivalent to μ_W^T and the Radon-Nykodim derivative is

$$\begin{aligned} L_T(\theta, X) &= \frac{d\mu_\theta^T}{d\mu_W^T} \\ &= \exp \left(\int_0^T f(\theta, X_t) dX_t - \frac{1}{2} \int_0^T f^2(\theta, X_t) dt \right) \end{aligned}$$

Prakasa Rao [35] defines efficiency as follows:

Definition 2 An estimator θ_T of θ^0 is said to be asymptotically efficient if there exist non-random functions $\alpha(\theta)$ and $\beta(\theta)$ such that

$$T^{-1/2} \frac{d}{d\theta} \log L_T(\theta, X) - \alpha(\theta) - \beta(\theta) T^{-1/2} (\theta_T - \theta) \rightarrow 0$$

as $T \rightarrow \infty$.

If we assume that $\log L_T(\theta, X)$ is differentiable with respect to θ and $\frac{d}{d\theta} \log L_T(\theta, X)$ can be obtained by interchanging the integral and the derivative operations (cf. Prakasa Rao [35]), then

$$\begin{aligned} \frac{d}{d\theta} \log L_T(\theta, X) &= \int_0^T f^{(1)}(\theta, X_t) dX_t - \int_0^T f(\theta, X_t) f^{(1)}(\theta, X_t) dt \\ &= \int_0^T f^{(1)}(\theta, X_t) dW_t. \end{aligned}$$

Theorem 1 The estimator $\bar{\theta}_{n,T}$ is asymptotically efficient as $T \rightarrow \infty$ and $n^{-1/2}T \rightarrow 0$.

Proof: see Prakasa Rao [35].

Remark: When applying definition (3.2) to the estimator $\bar{\theta}_{n,T}$, we use the likelihood $L_T(\theta, X)$ as an approximation. More appropriately, one should use the likelihood $L_{n,T}(\theta, X)$

$k = 1, \dots, d$. In matrix notation, (2.13) becomes

$$U_T(\theta^0) = I_T(\theta^0) (\hat{\theta}_T - \theta^0) \quad (2.14)$$

where the k -th element of the vector $U_T(\theta^0)$ is as the LHS of (2.13). As we have seen in the previous chapter (Theorem 1.2), $\{U_t(\theta^0), \mathcal{F}_t, t \geq 0\}$ is a zero mean square integrable martingale if

$$\sum_{k=1}^d \int_0^T E [f_k^*(s) G_s^{-1} f_l(s)] ds < \infty \quad (2.15)$$

Now suppose (2.15) holds for every $0 \leq T < \infty$, then we have the following

Theorem 1: (Taraskin [41]) *If the conditions*

$$C1 \quad \frac{1}{T} \int_0^T f_k^*(s) G_s^{-1} f_l(s) ds \rightarrow c_{kl}$$

$$C2 \quad \frac{1}{T} \int_0^T E [f_k^*(s) G_s^{-1} f_l(s)] ds \rightarrow c_{kl}$$

hold as $T \rightarrow \infty$, where c_{kl} are finite, then the distribution of

$$T^{-1/2} U_T(\theta^0) \quad (2.16)$$

converges to the normal law with zero mean and covariance matrix $I = (c_{kl})_{k,l=1, \dots, d}$

To prove the above Theorem, Taraskin [41] uses a time change technique applied to the martingale $\{U_t(\theta^0), \mathcal{F}_t; t \geq 0\}$. Now combining (2.14) and Theorem 1, it follows easily that $\hat{\theta}_T$ is a consistent estimate and that $\sqrt{T} (\hat{\theta}_T - \theta^0)$ is asymptotically normal with mean zero and covariance matrix I^{-1} .

Suppose that the drift coefficient and the diffusion operator in (2.8) depend only on x . That is $f_k(t, x) = f_k(x)$; $k = 1, \dots, d$ and $G(t, x) = G(x)$. Then by Theorem 1.3, the solution of such an equation will be a homogeneous Markov diffusion process. This case

shall be studied in detail in the next section. Further, let us suppose that it is recurrent and for each $\theta \in \Theta$ has the limiting distribution $G_\theta(x)$. Then the ergodic theorem implies that conditions C1 and C2 of Theorem 1 are satisfied and hence the asymptotic normality of (2.16) follows (see [41] for details).

Now let us suppose that $G(t, x) = G(t)$ and $f_k(t, x) = f_k(t)$, $k = 1, \dots, d$ (cf. Stoyanov [40]). The equation (2.14) takes the form.

$$U_T(\theta^0) = J_T(\theta^0) (\hat{\theta}_T - \theta^0)$$

where $J_T(\theta^0) = E [I_T(\theta^0)] = I_T(\theta^0)$

The above also implies that $\hat{\theta}_T$ is unbiased since

$$E[\hat{\theta}_T - \theta^0] = J_T^{-1}(\theta^0) E[U_T(\theta^0)] = 0$$

Theorem 2: The vector $\sqrt{T} (\hat{\theta}_T - \theta^0)$ has normal distribution with zero mean and covariance matrix $I_T^{-1}(\theta^0)$.

Proof: See [41].

Example 2: The Logistic Model

This is a special case of (2.1) given by the following equation:

$$dX_t = X_t(\mu - \lambda X_t)dt + \sigma X_t dW_t \quad (2.17)$$

σ is assumed to be known since the formula

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} (X_{kT2^{-n}} - X_{(k-1)T2^{-n}})^2 = \sigma^2 \int_0^T X_t^2 dt \quad a.s. \quad (2.18)$$

implies that if the process X can be observed continuously over $[0, T]$, then σ^2 can be calculated exactly. Equation (2.14) becomes

$$\begin{pmatrix} \int_0^T dW_s \\ -\int_0^T X_s dW_s \end{pmatrix} = \begin{bmatrix} T & -\int_0^T X_s ds \\ -\int_0^T X_s ds & \int_0^T X_s^2 ds \end{bmatrix} \begin{pmatrix} \mu - \mu_0 \\ \lambda - \lambda_0 \end{pmatrix} \quad (2.19)$$

where $\theta^0 = (\mu_0, \lambda_0)$ is the true parameter.

Guttorp and Kulperger [18] have proved that if $\phi = \mu_0 - \frac{\sigma^2}{2} > 0$ then

$$\frac{\sigma^2}{T} I_T(\theta^0) \rightarrow \begin{bmatrix} 1 & -\phi/\lambda_0 \\ -\phi/\lambda_0 & \mu_0\phi/\lambda_0^2 \end{bmatrix} = I(\theta^0)$$

4: Estimation for Large Systems of Equations

The asymptotic theory of the previous sections assumed that the length of time of observation T can be taken to be arbitrarily large. There are cases however, where this hypothesis may not be appropriate. When the system under consideration consists of a large number (N) of interacting subsystems, a natural question is whether the estimators derived from sampling on a fixed interval $[0, T]$ have desirable properties in the limit $N \rightarrow \infty$. We will show in the next two sections that the answer to this question is affirmative. The same is also true in the case where a large number of independent replications on $[0, T]$ is available.

We consider a model given by a system of Itô SDE's

$$dX_j(t) = \sum_{k=1}^d \theta_k f_{j,k}(X(t)) dt + g_j(X(t)) dW_j(t) \quad j = 1, \dots, N \quad (2.20)$$

where $X(t) = (X_1(t), \dots, X_N(t))^*$ and $W(t) = (W_1(t), \dots, W_N(t))^*$ is a Wiener process with independent components. The functions $f_{j,k}(x)$, $g_j(x)$ $j = 1, \dots, N$, $k = 1, \dots, d$ are such that the conditions (a) and (b) of Theorem 1.3 are satisfied. Hence the unique solution (up to equivalence) X is a homogeneous Markov diffusion process (see Theorem 1.3). We assume that we can observe a continuous trajectory $\{X(t), t \in [0, T]\}$. Then we study the properties of the ML estimator $\hat{\theta}_N$ of θ as $N \rightarrow \infty$. We suppress the time parameter T in the notation $\hat{\theta}_N$ because for the rest of the chapter, T is assumed to be fixed.

According to (2.12) the likelihood equations are

$$\begin{aligned} \sum_{j=1}^N \sum_{k=1}^d \hat{\theta}_{N,k} \int_0^t f_{j,k}(X(s)) g_j^{-2}(X(s)) f_{j,l}(X(s)) ds = \\ = \sum_{j=1}^N \int_0^t f_{j,l}(X(s)) g_j^{-2}(X(s)) dX_j(s) \end{aligned} \quad (2.21)$$

$l = 0, \dots, d$. Now let $b_{j,k}(s) = g_j^{-1}(X(s)) f_{j,k}(X(s))$. Then (2.21) combined with (2.20) yields

$$\sum_{j=1}^N \sum_{k=1}^d (\hat{\theta}_{N,k} - \theta_k^0) \int_0^t b_{j,k}(s) b_{j,l}(s) ds = \sum_{j=1}^N \int_0^t b_{j,l}(s) dW_j(s) \quad (2.22)$$

$l = 1, \dots, d$, and $\theta^0 = (\theta_1^0, \dots, \theta_d^0)$ is assumed to be the true parameter. We can write the above in the form (2.14) as follows:

$$I_N(t)(\hat{\theta}_N - \theta^0) = M_N(t) \quad t \in [0, T] \quad (2.23)$$

where

$$\begin{aligned} M_N(t) &= \begin{pmatrix} N^{-1} \sum_{j=1}^N \int_0^t b_{j,1}(s) dW_j(s) \\ \vdots \\ N^{-1} \sum_{j=1}^N \int_0^t b_{j,d}(s) dW_j(s) \end{pmatrix} \\ I_N(t) &= \begin{pmatrix} N^{-1} \sum_{j=1}^N \int_0^t b_{j,k}(s) b_{j,l}(s) ds \end{pmatrix}_{k,l=1, \dots, d} \end{aligned}$$

We are now in a position to prove an analogue of Theorem 1.

Theorem 3: Suppose that $\{b_{j,k}(s), s \in [0, T]\}$, $j = 1, \dots, N$; $k = 1, \dots, d$, are measurable and adapted processes satisfying:

$$N^{-1} \sum_{j=1}^N \int_0^t b_{j,k}(s) b_{j,l}(s) ds \xrightarrow{a.s.} c_{kl}(t) \quad (2.24)$$

$k, l = 1, \dots, d$, where $c_{kl}(t)$ are finite and continuous non-random functions of $t \in [0, T]$.

If also the limiting matrix $I(t) = (c_{kl}(t))_{k,l=1, \dots, d}$ is positive definite, $\lambda^* I(t) \lambda$ is increasing for all $\lambda \in R^d$ and $I(0) = 0$ then $\hat{\theta}_N$ is a consistent estimator of θ and $N^{1/2}(\hat{\theta}_N - \theta^0)$ is asymptotically normal with mean zero and covariance matrix $I^{-1}(t)$.

Remark: In the exchangeable case the hypothesis (2.24) follows from the McKean-Vlasov law of large numbers (cf. Léonard [26]). In particular (2.24) will be satisfied when $f_{j,k}(X) = f_k(X_j)$ and $g_j(X) = g(X_j)$. This corresponds to the case of independent replicated sampling on $[0, T]$

To prove Theorem 3, we need a simplified version of Rebolledo's central limit Theorem [36]. This is given as Theorem 4 below.

Theorem 4: Let M_N , $N \in \mathbb{Z}_+$ be a sequence of local square integrable martingales with $M_N(0) = 0$. Suppose the following holds.

$$\sum_{s \leq t} E \{ |\Delta M_N(s)|^2 \chi(|\Delta M_N(s)| > \epsilon) \} \rightarrow 0,$$

for all $t \in [0, T]$, $\epsilon > 0$; and

$$\langle M_N \rangle(t) \xrightarrow{p.p.} c(t)$$

for all $t \in [0, T]$, where $c(t)$ is a continuous increasing function with $c(0) = 0$. Then $M_N \Rightarrow M$ a continuous Gaussian martingale with mean zero and covariance function $K(s, t) = c(s \wedge t)$, $s, t \in [0, T]$.

Remarks: The notation $\Rightarrow$ above means weak convergence in the space $D[0, T]$ equipped with the Skorohod topology. $\langle M \rangle(t)$ is the integrable increasing process in the Doob-Meyer decomposition of M^2 . Finally, $\Delta M(s) = M(s) - M(s-)$ is the jump of M at the point s . It should be noted that Theorem 3 holds under much weaker hypothesis than the one given above. The following corollary is a simple application of Theorem 4 to a multidimensional case.

Corollary: If $M_N = (M_{N1}, \dots, M_{Nd})^*$ is a sequence of local square integrable mar-

tingales such that $M_N(0) = 0$, $N \in \mathbb{Z}_+$ and

$$\sum_{s \leq t} E \{ |\Delta M_{N_j}(s)|^2 \chi(|\Delta M_{N_j}(s)| > \epsilon) \} \rightarrow 0 \quad j = 1, \dots, d \quad (2.25)$$

$$\langle M_{N_k}, M_{N_l} \rangle(t) \rightarrow c_{kl}(t) \quad k, l = 1, \dots, d$$

as $N \rightarrow \infty$ where for each $k, l = 1, \dots, d$, $c_{kl}(t)$ is finite and continuous. Suppose also that the matrix $I(t) = (c_{kl}(t))_{k,l=1, \dots, d}$ is positive definite for $t > 0$, $I(0) = 0$ and $\lambda^* I(t) \lambda$ is increasing for all $\lambda \in \mathbb{R}^d$, then for each $t \in [0, T]$

$$M_N(t) \xrightarrow{D} M(t)$$

where $M(\cdot)$ is an $\mathbb{R}^d$ -valued mean zero continuous Gaussian martingale with $\text{Cov}(M(t)) = I(t)$.

Lemma 1: Let $\lambda_1, \dots, \lambda_d$ and $x_1, \dots, x_d$ be real numbers and define

$$x = \sum_{k=1}^d \lambda_k x_k.$$

If
$$\sum_{k=1}^d |\lambda_k| \leq 1$$

Then for any $\epsilon > 0$

$$x^2 \chi(|x| > \epsilon) \leq \sum_{k=1}^d x_k^2 \chi(|x_k| > \epsilon)$$

Proof: Clearly

$$|x| \leq \sum_{k=1}^d |\lambda_k| |x_k| \leq \max_{1 \leq k \leq d} |x_k|$$

and $x^2 \chi(|x| > \epsilon)$ is non-decreasing in $|x|$. Hence

$$x^2 \chi(|x| > \epsilon) \leq \max_k |x_k|^2 \chi(\max_k |x_k| > \epsilon) \leq \sum_{k=1}^d x_k^2 \chi(|x_k| > \epsilon) \quad \square$$

Proof: (of the corollary) We use the Cramer-Wold method (Billingsley [5]) Let

$$\lambda = (\lambda_1, \dots, \lambda_d)^* \in R^d \quad \text{such that} \quad \sum_{k=1}^d |\lambda_k| < 1$$

and let

$$M_N^\lambda = \sum_{k=1}^d \lambda_k M_{Nk}$$

Then the condition of Theorem 4 for M_N^λ follows from (2.25) and Lemma 1. M_N^λ is a local square integrable martingale with

$$\langle M_N^\lambda \rangle(t) = \sum_{k=1}^d \sum_{l=1}^d \lambda_k \lambda_l \langle M_{Nk}, M_{Nl} \rangle(t) \xrightarrow{p.p.} \sum_{k=1}^d \sum_{l=1}^d \lambda_k \lambda_l c_{kl}(t) = \lambda^* I(t) \lambda$$

By Theorem 3, $M_N^\lambda \Rightarrow M^\lambda$ a continuous gaussian martingale with mean zero and covariance function $\lambda^* I(s \wedge t) \lambda$, $s, t \in [0, T]$. In particular,

$$M_N^\lambda(t) \xrightarrow{p.p.} M^\lambda(t)$$

and the proof is complete. □

We are now in a position to prove Theorem 3

Proof: From (2.23) we have

$$I_N(t)(\hat{\theta}_N - \theta) = M_N(t) \quad (2.26)$$

The hypothesis of the Theorem implies

$$I_N(t) \xrightarrow{p.p.} I(t) \quad \text{for all } t \in [0, T] \quad (2.27)$$

$\{I(t), t \in (0, T)\}$ are non-singular matrices. Let $B^{d \times d}$ be the Banach algebra of $d \times d$ matrices and $C \subset B^{d \times d}$ the subset containing all invertible elements of $B^{d \times d}$. Since C is

open and $\{I(t), t \in [0, T]\} \subset C$, there exists a number $n_0 \geq 1$ such that $I_N(t) \in C$ for all $N \geq n_0, t \in [0, T]$. Therefore we have

$$\hat{\theta}_N - \theta^0 = [I_N(t)]^{-1} M_N(t) \quad (2.28)$$

for all $N \geq n_0, t \in [0, T]$. The continuity of the mapping $x \mapsto x^{-1}$ and (2.27) implies that

$$[I_N(t)]^{-1} \xrightarrow{s.s.} I(t)^{-1} \quad \text{as } N \rightarrow \infty$$

Now consider the k -th component of $M_N(t)$.

$$M_{Nk}(t) = N^{-1} \sum_{j=1}^N \int_0^t b_{jk}(s) dW_j(s) \quad (2.29)$$

For each $N \in \mathbb{Z}_+$, $\{M_{Nk}(t), \mathcal{F}_t; t \in [0, T]\}$ is a zero mean continuous square integrable martingale with

$$\langle M_{Nk}(t) \rangle(t) = N^{-2} \sum_{j=1}^N \int_0^t b_{jk}^2(s) ds \quad (2.30)$$

Theorem 1.9 yields

$$\Pr \left(\sup_{0 \leq t \leq T} \left\{ M_{Nk}(t) - \frac{\alpha}{2} \langle M_{Nk}(t) \rangle(t) \right\} > \beta \right) < e^{-\alpha\beta}$$

which by symmetry and (2.30) reduces to

$$\Pr \left(\sup_{0 \leq t \leq T} |M_{Nk}(t)| > \beta + \frac{\alpha}{2N^2} \sum_{j=1}^N \int_0^T b_{jk}^2(s) ds \right) < 2e^{-\alpha\beta}$$

If we choose $\alpha = N^a, \beta = N^{-b}$ where $0 < b < a < 1$ we obtain

$$\Pr \left(\sup_{0 \leq t \leq T} |M_{Nk}(t)| > \frac{1}{N^b} + \frac{N^{-1}}{2N^{1-a}} \sum_{j=1}^N \int_0^T b_{jk}^2(s) ds \right) < 2e^{-N^{a-b}} \quad (2.31)$$

Since $\sum_{N=1}^{\infty} e^{-N^{a-b}} < \infty$ and $N^{-1} \sum_{j=1}^N \int_0^T b_{jk}^2(s) ds \xrightarrow{s.s.} c_{kk}(T)$, the Borel-Cantelli lemma implies that for each $k = 1, \dots, d$

$$\sup_{0 \leq t \leq T} |M_{Nk}(t)| \xrightarrow{s.s.} 0 \quad \text{as } N \rightarrow \infty \quad (2.32)$$

This completes the proof of consistency.

To prove asymptotic normality, we again use identity (2.23) in the form

$$I_N(t)N^{1/2}(\hat{\theta}_N - \theta^0) = N^{1/2}M_N(t) \quad (2.33)$$

By (2.27) and the fact that $I(t)$, $t \in [0, T]$ are non-random, it will be sufficient to show that

$$N^{1/2}M_N(t) \xrightarrow{D} N_d(0, I(t))$$

The continuity of the process $\{N^{1/2}M_N(t), t \in [0, T]\}$ implies that the Lindeberg condition (2.25) is satisfied. Furthermore, we have

$$\langle N^{1/2}M_{Nk}, N^{1/2}M_{Nl} \rangle(t) = N^{-1} \sum_{j=1}^N \int_0^t b_{j,k}(s)b_{j,l}(s)ds = c_{kl}(t)$$

for all $k, l = 1, \dots, d$. Hence the desired result follows from the corollary

5: Exchangeable Interacting Diffusions

An interesting example of the system (2.20) is the case where the subsystems are interacting and exchangeable. We shall consider a simple mean field model described by a system of Itô SDE's

$$dX_j(t) = \alpha X_j(t)dt - \beta(X_j(t) - \bar{X}_N(t))dt + dW_j(t), \quad X_j(0) = x_j(0), \quad (2.34)$$

$j = 1, \dots, N$; where $\bar{X}_N(t) = N^{-1} \sum_{j=1}^N X_j(t)$, $\beta \neq \alpha$ and $\alpha \neq 0$. The middle term on the RHS of (2.34) can be viewed as an interaction among the subsystems which create a tendency for the subsystems to relax toward the center of gravity of the ensemble. Mean-field-type models have applications in physics, biology, economics, etc. (cf. Dawson [12], Léonard [26] and references therein). A special case where $\beta = 0$ corresponds to sampling

independent replications of the same process on $[0, T]$. We shall study the asymptotic behaviour of the ML estimator $\hat{\theta}_N$ of $\theta = (\alpha, \beta)$ arising from model (2.34)

In terms of the notation introduced in the previous section

$$f_{j1}(x) = x_j, \quad f_{j2}(x) = -(x_j - \bar{x}) \quad x = (x_1, \dots, x_N)^*$$

$$b_{jk}(s) = f_{jk}(X(s)) \quad \text{since} \quad g_j(x) = 1$$

$$j = 1, \dots, N, \quad k = 1, 2$$

$$N^{-1} \sum_{j=1}^N b_{j1}^2(s) = N^{-1} \sum_{j=1}^N X_j^2(s)$$

$$N^{-1} \sum_{j=1}^N b_{j2}^2(s) = N^{-1} \sum_{j=1}^N (X_j(s) - \bar{X}_N(s))^2$$

$$\begin{aligned} N^{-1} \sum_{j=1}^N b_{j1}(s)b_{j2}(s) &= N^{-1} \sum_{j=1}^N X_j(s)(X_j(s) - \bar{X}_N(s)) \\ &= -N^{-1} \sum_{j=1}^N (X_j(s) - \bar{X}_N(s))^2 \end{aligned}$$

Hence the observed information matrix is

$$\begin{aligned} I_N(t) &= N^{-1} \begin{bmatrix} \sum_{j=1}^N \int_0^t X_j^2(s) ds & - \sum_{j=1}^N \int_0^t (X_j(s) - \bar{X}_N(s))^2 ds \\ - \sum_{j=1}^N \int_0^t (X_j(s) - \bar{X}_N(s))^2 ds & \sum_{j=1}^N \int_0^t (X_j(s) - \bar{X}_N(s))^2 ds \end{bmatrix} \\ &= \begin{pmatrix} A_N(t) & -B_N(t) \\ -B_N(t) & B_N(t) \end{pmatrix} \end{aligned}$$

Then from (2.21) the ML estimator of $\theta = (\alpha, \beta)^*$ is given by

$$\hat{\theta}_N = \frac{N^{-1}}{\Delta_N(t)} \begin{pmatrix} B_N(t) & B_N(t) \\ B_N(t) & A_N(t) \end{pmatrix} \begin{bmatrix} \sum_{j=1}^N \int_0^T X_j(t) dX_j(t) \\ - \sum_{j=1}^N \int_0^T (X_j(t) - \bar{X}_N(t)) dX_j(t) \end{bmatrix}, \quad (2.35)$$

where $\Delta_N(t) = \det[I_N(t)]$. The following Theorem is the main result of this section

Theorem 5: Suppose model (2.34) satisfies

$$\bar{x}_N(0) = N^{-1} \sum_{j=1}^N x_j(0) \xrightarrow{a.s.} \mu_0 \quad \text{and} \quad N^{-1} \sum_{j=1}^N x_j^2(0) \xrightarrow{a.s.} \sigma_0^2 + \mu_0^2 \quad \text{as } N \rightarrow \infty$$

for some real constants μ_0 and $\sigma_0^2 > 0$. Then $\hat{\theta}_N$ is a strongly consistent estimator of θ and $N^{1/2}(\hat{\theta}_N - \theta)$ is asymptotically normal with mean zero and covariance matrix $I^{-1}(t)$ where

$$I(t) = \begin{pmatrix} A(t) & -B(t) \\ -B(t) & B(t) \end{pmatrix}$$

and

$$A(t) = \frac{x_0^2}{2\alpha} (e^{2\alpha t} - 1) + B(t)$$

$$B(t) = \frac{1}{\gamma^2} (e^{\gamma t} - 1) - \frac{t}{\gamma}, \quad \gamma = 2(\alpha - \beta)$$

Proof: From (2.34) we have

$$\bar{X}_N(t) = \bar{x}_N(0) + \alpha \int_0^t \bar{X}_N(s) ds + \bar{W}_N(t)$$

where

$$\bar{X}_N(t) = N^{-1} \sum_{j=1}^N X_j(t)$$

$$\bar{W}_N(t) = N^{-1} \sum_{j=1}^N W_j(t)$$

The process $\{\bar{W}_N(t), t \in [0, T]\}$ is equivalent to the process $\{N^{-1/2}W(t), t \in [0, T]\}$ where W is a standard Wiener process on R . From above, Theorem 1.8 and integration by parts we obtain

$$\begin{aligned} \bar{X}_N(t) &= \bar{x}_N(0) + \alpha \int_0^t \bar{X}_N(s) ds + N^{-1/2}W(t) \\ &= \bar{x}_N(0)e^{\alpha t} + N^{-1/2}e^{\alpha t} \int_0^t e^{-\alpha s} dW(s) \end{aligned} \quad (2.36)$$

Also

$$\begin{aligned} &\Pr \left(\sup_{0 \leq t \leq T} \left| N^{-1/2}e^{\alpha t} \int_0^t e^{-\alpha s} dW(s) \right| > \epsilon \right) \\ &\leq \Pr \left(N^{-1/2}K(T) \sup_{0 \leq t \leq T} \left| \int_0^t e^{-\alpha s} dW(s) \right| > \epsilon \right) \end{aligned}$$

where $K(T) = \max \{e^{\alpha T}, 1\}$

$$\begin{aligned} &\leq \left[\frac{K(T)}{N^{1/2}\epsilon} \right]^4 E \left\{ \int_0^T e^{-\alpha s} dW(s) \right\}^4 \\ &\leq \frac{K^4(T)}{N^2\epsilon^4} \cdot \frac{3T}{2\alpha} (1 - e^{-4\alpha T}) \quad \text{by Theorem (1.2)} \end{aligned}$$

Hence it follows by the Borel-Cantelli Lemma and (2.36) that

$$\sup_{0 \leq t \leq T} |\bar{X}_N(t) - \bar{x}_N(0)e^{\alpha t}| \xrightarrow{a.s.} 0.$$

The inequality $|a^2 - b^2| \leq (|a - b| + 2|b|)|a - b|$ for real numbers a, b implies that

$$\begin{aligned} \sup_{0 \leq t \leq T} |\bar{X}_N^2(t) - \bar{x}_N^2(0)e^{2\alpha t}| &\leq \sup_{0 \leq t \leq T} |\bar{X}_N(t) - \bar{x}_N(0)e^{\alpha t}| \times \\ &\times \left\{ \sup_{0 \leq t \leq T} |\bar{X}_N(t) - \bar{x}_N(0)e^{\alpha t}| + 2 \sup_{0 \leq t \leq T} \bar{x}_N(0)e^{\alpha t} \right\} \end{aligned}$$

Hence

$$\sup_{0 \leq t \leq T} |\bar{X}_N^2(t) - \bar{x}_N^2(0)e^{2\alpha t}| \xrightarrow{a.s.} 0 \quad (2.37)$$

This in turn implies that

$$\sup_{0 \leq t \leq T} \left| \int_0^t \bar{X}_N^2(s) ds - \int_0^t \bar{x}_N^2(0)e^{2\alpha s} ds \right| \xrightarrow{a.s.} 0 \quad (2.38)$$

Itô's formula applied to (2.34) gives

$$\begin{aligned} X_j^2(t) &= x_j^2(0) + t + 2(\alpha - \beta) \int_0^t X_j^2(s) ds + 2\beta \int_0^t \bar{X}_N(s) X_j(s) ds + \\ &\quad + 2 \int_0^t X_j(s) dW_j(s) \end{aligned} \quad (2.39)$$

Therefore

$$\begin{aligned} N^{-1} \sum_{j=1}^N X_j^2(t) &= N^{-1} \sum_{j=1}^N x_j^2(0) + t + 2(\alpha - \beta) N^{-1} \sum_{j=1}^N \int_0^t X_j^2(s) ds + \\ &\quad + 2\beta \int_0^t \bar{X}_N(s) ds + 2N^{-1} \sum_{j=1}^N \int_0^t X_j(s) dW_j(s), \end{aligned}$$

that is

$$Y_N(t) = a_N(t) + V_N(t) + M_N(t) + \gamma \int_0^t Y_N(s) ds \quad (2.40)$$

where

$$a_N(t) = N^{-1} \sum_{j=1}^N x_j^2(0) + t$$

$$V_N(t) = 2\beta \int_0^t \bar{X}_N^2(s) ds$$

$$M_N(t) = 2N^{-1} \sum_{j=1}^N \int_0^t X_j(s) dW_j(s)$$

$$Y_N(t) = N^{-1} \sum_{j=1}^N X_j^2(t)$$

and

$$\gamma = 2(\alpha - \beta)$$

By Theorem 1.8, the solution of (2.40) is given by

$$\begin{aligned} Y_N(t) &= a_N(t) + V_N(t) + M_N(t) + \gamma \int_0^t e^{\gamma(t-s)} (a_N(s) + V_N(s) + M_N(s)) ds \quad (2.41) \\ &= a_N(t) + \gamma \int_0^t e^{\gamma(t-s)} a_N(s) ds + V_N(t) + M_N(t) + \\ &\quad + \gamma \int_0^t e^{\gamma(t-s)} V_N(s) ds + \gamma \int_0^t e^{\gamma(t-s)} M_N(s) ds \end{aligned}$$

By (2.38),

$$V_N(t) \xrightarrow{p} \frac{\mu_0^2 \beta}{\alpha} (e^{2\alpha t} - 1)$$

uniformly in $t \in [0, T]$. Thus

$$\begin{aligned} \gamma \int_0^t e^{\gamma(t-s)} V_N(s) ds &\xrightarrow{p} \gamma \int_0^t e^{\gamma(t-s)} \frac{\mu_0^2 \beta}{\alpha} (e^{2\alpha s} - 1) ds \\ &= \frac{\gamma \mu_0^2}{2\alpha} e^{2\alpha t} - \mu_0^2 e^{\gamma t} + \frac{\mu_0^2 \beta}{\alpha} \end{aligned}$$

uniformly in $t \in [0, T]$. Hence if we can show that $\sup_{0 \leq t \leq T} M_N(t) \xrightarrow{p} 0$ and consequently

$$\sup_{0 \leq t \leq T} \left| \gamma \int_0^t e^{\gamma(t-s)} M_N(s) ds \right| \xrightarrow{p} 0$$

it would follow that

$$\begin{aligned} Y_N(t) &\xrightarrow{a.s.} \sigma_0^2 + \mu_0^2 + t + \gamma \int_0^t e^{\gamma(t-s)} (\sigma_0^2 + \mu_0^2 + s) ds + \frac{\mu_0^2 \beta}{\alpha} (e^{2\alpha t} - 1) \\ &\quad + \frac{\gamma \mu_0^2}{2\alpha} e^{2\alpha t} + \frac{\mu_0^2 \beta}{\alpha} - \mu_0^2 e^{\gamma t} \\ &= \sigma_0^2 e^{2\alpha t} + \frac{e^{\gamma t} - 1}{\gamma} = Y(t) \end{aligned}$$

uniformly in $t \in [0, T]$. We now show that $M_N(t)$ converges to zero almost surely, uniformly in t . In view of the inequality (2.31)

$$\Pr \left(\sup_{0 \leq t \leq T} |M_N(t)| > \frac{1}{N^b} + \frac{2}{N^{1-a}} \int_0^T Y_N(t) dt \right) \leq 2e^{-N^{a-b}} \quad (2.42)$$

where $0 < b < a < 1$. Moreover (2.41) yields

$$Y_N(t) \leq e^{\gamma t} \sup_{0 \leq t \leq T} \{a_N(t) + V_N(t) + M_N(t)\}.$$

Hence

$$\int_0^T Y_N(t) dt \leq C(T) \sup_{0 \leq t \leq T} \{a_N(t) + V_N(t) + M_N(t)\}, \quad (2.43)$$

where $C(T) = \int_0^T e^{\gamma t} dt$. Combining (2.42) and (2.43) yields

$$\begin{aligned} &\Pr \left(\left\{ 1 - \frac{2C(T)}{N^{1-a}} \right\} \sup_{0 \leq t \leq T} |M_N(t)| > \frac{1}{N^b} + \frac{2C(T)}{N^{1-a}} \sup_{0 \leq t \leq T} \{a_N(t) + V_N(t)\} \right) \\ &\leq \Pr \left(\sup_{0 \leq t \leq T} |M_N(t)| > \frac{1}{N^b} + \frac{2}{N^{1-a}} C(T) \sup_{0 \leq t \leq T} \{a_N(t) + V_N(t) + M_N(t)\} \right) \\ &\leq e^{-N^{a-b}}. \end{aligned} \quad (2.44)$$

By hypothesis $a_N(t)$ converges a.s.. Also we have shown that $V_N(t)$ converges a.s. as $N \rightarrow \infty$. We can therefore conclude from (2.44) and the Borel-Cantelli lemma that

$$\sup_{0 \leq t \leq T} |M_N(t)| \xrightarrow{a.s.} 0 \quad \text{as } N \rightarrow \infty.$$

Now each term in (2.41) converges and hence

$$\sup_{0 \leq t \leq T} |Y_N(t) - Y(t)| \xrightarrow{a.s.} 0 \quad (2.45)$$

Finally, we show that $A_N(t) \xrightarrow{a.s.} A(t)$ and $B_N(t) \xrightarrow{a.s.} B(t)$. By (2.45) and simple calculations we obtain

$$\begin{aligned} A_N(t) &= N^{-1} \sum_{j=1}^N \int_0^t X_j^2(s) ds \\ &= \int_0^t Y_N(s) ds \xrightarrow{a.s.} \int_0^t Y(s) ds \\ &= \frac{\sigma_0^2}{2\alpha} (e^{2\alpha t} - 1) + \frac{1}{\gamma^2} (e^{\gamma t} - 1) - \frac{t}{\gamma} \\ &= A(t) \end{aligned}$$

uniformly in $t \in [0, T]$, and

$$\begin{aligned} B_N(t) &= N^{-1} \sum_{j=1}^N \int_0^t (X_j(s) - \bar{X}_N(s))^2 ds \\ &= \int_0^t Y_N(s) ds - \int_0^t \bar{X}_N^2(s) ds \\ &\xrightarrow{a.s.} \frac{1}{\gamma^2} (e^{\gamma t} - 1) - \frac{t}{\gamma} \end{aligned}$$

uniformly in $t \in [0, T]$.

Finally we verify that the hypothesis of Theorem 3 is satisfied by the matrices $I(t)$, $t \in [0, T]$. Take any $\lambda = (\lambda_1, \lambda_2) \in R^2$, then

$$\begin{aligned} \lambda^* I(t) \lambda &= (\lambda_1, \lambda_2) \begin{pmatrix} A(t) & -B(t) \\ -B(t) & B(t) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \\ &= \lambda_1^2 (A(t) - B(t)) + (\lambda_2 - \lambda_1)^2 B(t) \\ &> 0 \quad \text{for all } t > 0 \end{aligned}$$

Hence $\lambda^* I(t) \lambda$ is positive definite for all $t \in (0, 1]$. Furthermore it is easy to see that $\lambda^* I(t) \lambda$ is an increasing continuous function of t with $\lambda^* I(0) \lambda = 0$. This completes the proof of Theorem 5. $\square$

6: Testing of hypothesis

One possible application of Theorem 5 could be to test the hypothesis of no interaction

arising from model (2.42). Consider

$$H_0: \beta = 0$$

$$H_A: \beta \neq 0$$

The null hypothesis H_0 asserts that the processes $\{X_j(t), t \in [0, T]; j = 1, \dots, N\}$ are independent. By Theorem 5

$$N^{1/2}(\hat{\beta}_N - \beta) \stackrel{\text{appr}}{\sim} N\left(0, \frac{A(t)}{(A(t) - B(t))B(t)}\right)$$

Hence under H_0

$$N^{1/2}\hat{\beta}_N \sqrt{\frac{(A_N(t) - B_N(t))B_N(t)}{A_N(t)}} \stackrel{\text{appr}}{\sim} N(0, 1)$$

And the null hypothesis is rejected if

$$\left| N^{1/2}\hat{\beta}_N \sqrt{\frac{(A_N(t) - B_N(t))B_N(t)}{A_N(t)}} \right| > z_{\epsilon/2}$$

where ϵ is the chosen size of the test.

7: Non-Linear Stochastic Differential Equations

One important contribution to the non-linear theory of ML estimation is that of Borkar and Bagchi [7]. They considered the following equation:

$$dX_t = f(\theta, t, X)dt + dW_t, \quad X_0 = 0 \quad \text{a.s.} \quad (2.46)$$

To obtain the asymptotic results similar to those in the linear case, additional assumptions are required. They assumed that $\theta \in K \subset R^d$ where K is compact. Also they assumed that there exists a non-negative measurable function $c: C[0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ satisfying:

$$|f(\theta_1, t, x) - f(\theta_2, t, x)| \leq c(t, x)\|\theta_1 - \theta_2\| \quad (2.47)$$

where $\|\cdot\|$ is the Euclidean norm in R^d . Note that (2.47) holds for stochastic systems described by (2.8). The ML estimator $\hat{\theta}_T$ in this case is a solution of

$$\frac{\partial}{\partial \theta} l_T(\theta) = 0 \quad (2.48)$$

where

$$l_T(\theta) = \log L_T(\theta, X)$$

and the likelihood function $L_T(\theta, X)$ is given by Theorem 1.7 as

$$L_T(\theta, X) = \exp \left\{ \int_0^T f(\theta, s, X) dX_s - \frac{1}{2} \int_0^T f^2(\theta, s, X) ds \right\} \quad (2.49)$$

A key result in [7] is lemma 4.1 which we state as Theorem 6 for the sake of completeness

Theorem 6:

$$\lim_{t \rightarrow \infty} \sup_{\theta \in K} \frac{1}{t} \int_0^t h(X, s, \theta) dW_s = 0$$

with probability-1, where

$$h(\theta, s, X) = f(\theta, s, X) - f(\theta^0, s, X)$$

θ^0 being the true value of θ

Now consistency of the ML estimator $\hat{\theta}_T$ is implied by Theorem 6 by the following simple argument. Since $\hat{\theta}_T$ maximizes $l_T(\theta)$ (see (2.2)) we have, with probability-1

$$l_T(\hat{\theta}_T) \geq l_T(\theta) \quad \text{for all } \theta \in K$$

in particular

$$l_T(\hat{\theta}_T) \geq l_T(\theta^0) = 0$$

that is,
$$\frac{1}{T} \left\{ \int_0^T h(\theta, s, X) dW_s - \frac{1}{2} \int_0^T h^2(\theta, s, X) ds \right\}_{\theta = \hat{\theta}_T} \geq 0 \quad \text{a.s.}$$

Hence

$$\frac{1}{T} \int_0^T h(\theta, s, X) dW_s \Big|_{\theta=\hat{\theta}_T} \geq \frac{1}{2T} \int_0^T h^2(\hat{\theta}_T, s, X) ds \geq 0 \quad (2.50)$$

Theorem 6 and (2.50) implies that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T h^2(\hat{\theta}_T, s, X) ds = 0 \quad \text{a.s.} \quad (2.51)$$

Obviously (2.51) holds if $\hat{\theta}_T$ is replaced by θ^0 . Denote by

$$D = \left\{ \theta \in K; \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T h^2(\theta, s, X) ds = 0 \right\}$$

It easily follows from assumption (2.47) (essentially) that the family

$$\left\{ \frac{1}{T} \int_0^T h^2(\theta, s, X) ds, t \geq 0 \right\}$$

is equicontinuous in θ . Hence if we further assume that $D = \{\theta^0\}$, then (2.51) and the above implies that $\hat{\theta}_T \rightarrow \theta^0$ a.s.. For a detailed argument see [7]

8: Estimation from Continuous-Discrete Data

In practice, even if continuous data is available, in order to compute the ML estimate $\hat{\theta}_T$ from (2.12) the integrals

$$\int_0^T f_k^*(t) G_t^{-1} dX_t \quad \text{and} \quad \int_0^T f_k^*(t) G_t^{-1} f_l(t) dt \quad (2.52)$$

have to be approximated by appropriate finite sums of the form

$$\sum_{i=1}^n f_k^*(t_{i-1}) G_{t_{i-1}}^{-1} (X_{t_i} - \bar{X}_{t_{i-1}}) \quad (2.53)$$

and

$$\sum_{i=1}^n f_k^*(t_{i-1}) G_{t_{i-1}}^{-1} f_l(t_{i-1}) (t_i - t_{i-1})$$

where $\mathcal{W}_n = t_0, t_1, \dots, t_n$ defines a partition of $[0, T]$. The approximation (2.53) of an Itô's stochastic integral may need some adjustment (cf. Wong and Zakai [42], Sjögren [38]). Then a natural question is whether the resulting estimator $\hat{\theta}_{n,T}$ is the best one that could be obtained in view of the continuous-discrete data. Le Breton [8] studied this problem using the following SDE model

$$dX_t = \theta X_t dt + dW_t, \quad X_0 = 0, \quad (2.54)$$

where $W = (W_t; t \geq 0)$ is the standard Wiener process in R^N , θ is an unknown $N \times N$ matrix. Let $\hat{\theta}_T$ denote the ML estimate based on continuous sampling on $[0, T]$ of the process described by (2.54) and $\hat{\theta}_{n,T}$ be the corresponding approximation obtained by replacing stochastic and ordinary integrals by Reimann sums as in (2.53). In this case

$$\hat{\theta}_{n,T} = \left(\sum_{i=1}^n (X_{t_i} - X_{t_{i-1}}) X_{t_{i-1}}^* \right) \left(\sum_{i=1}^n (t_i - t_{i-1}) X_{t_{i-1}} X_{t_{i-1}}^* \right)^{-1} \quad (2.55)$$

Model (2.54) allows computation of the exact likelihood function based on observations at $t_0, t_1, \dots, t_n$. Let $\tilde{\theta}_{n,T}$ be the resulting ML estimate. Le Breton [8] proved that if $|\mathcal{W}_n| = \max_{1 \leq i \leq n} |t_i - t_{i-1}|$ goes to zero as $n \rightarrow \infty$, then $\lim_{n \rightarrow \infty} \hat{\theta}_{n,T} = \lim_{n \rightarrow \infty} \tilde{\theta}_{n,T} = \hat{\theta}_T$ in probability. He also showed that $|\mathcal{W}_n|^{-1/2}(\hat{\theta}_{n,T} - \hat{\theta}_T)$ and $|\mathcal{W}_n|^{-1/2}(\tilde{\theta}_{n,T} - \hat{\theta}_T)$ are bounded in probability. The latter statement implies that, if only continuous-discrete data is available, then $\hat{\theta}_{n,T}$ is *almost the best estimate* up to the order $|\mathcal{W}_n|^{1/2}$ and is easily computable. On the other hand, the likelihood equations for $\tilde{\theta}_{n,T}$ do not offer a closed form solution (see [8]).

One interesting question that remains unanswered by Le Breton's work is what happens if both $n \rightarrow \infty$ and $T \rightarrow \infty$. Does $\hat{\theta}_{n,T}$ converge to θ^0 the true parameter? We shall study this and other related problems in the next chapter in a more general setting.

Chapter 3

LEAST SQUARES METHODS

1: Introduction

In this chapter we study the problem of estimation for the following stochastic differential equation model

$$dX_t = f(\theta^0, X_t)dt + dW_t, \quad X_0 = x_0, \quad (3.1)$$

when only discrete data is available. A basic motivation for this problem is the fact that, despite the mathematical elegance of the maximum likelihood theory based on continuous sampling, in practice only a finite number of observations is available. This problem has not received much attention in the literature. In section 2, we give a brief review of the available results. In section 3, we extend Dorogovcev's [13] result by proving the strong consistency of the least squares estimator. The methods discussed in sections 2 and 3 are based on the assumption that the underlying process is stationary. In section 4, we introduce an estimator which has nice asymptotic properties even in the non-stationary case. This estimator is based on the new idea of approximating a solution of (3.1) by a solution of a deterministic equation.

2: Estimation from Stationary Processes

We shall assume throughout this section and in section 3.3 that the process $X = (X_t, t \geq 0)$ is stationary ergodic (cf. Glhman and Skorohod [16]). Let $t_0, \dots, t_n = T$ be the time points at which the process X is observed. Denote by $\Delta_k t = t_k - t_{k-1}$, $k =$

1, 2, ..., n. the time between consecutive observations. In this chapter unless otherwise stated, we will assume that $\Delta_k t = T/n$.

Definition 1. Let $\theta^0 \in \Theta$ denote the true parameter in equation (3.1) and

$$Q_{n,T}(\theta) = \sum_{k=1}^n \frac{|X_{t_k} - X_{t_{k-1}} - f(\theta, X_{t_{k-1}}) \Delta_k t|^2}{\Delta_k t}$$

Then $\tilde{\theta}_{n,T}$ is the least squares (LS) estimator of θ^0 if

$$Q_{n,T}(\tilde{\theta}_{n,T}) = \min_{\theta \in \Theta} Q_{n,T}(\theta)$$

where Θ is a subspace of a normed linear space

Consistency

As far as we know, the first person to study this type of estimator is Dorogovcev [13]. He proved (the proof appears to have a technical error on page 79) that $\tilde{\theta}_{n,T}$ is consistent. The method of proof was based on analysing the inequality

$$\begin{aligned} 2 \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}} - f(\theta^0, X_{t_{k-1}}) \Delta_k t) \cdot (f(\tilde{\theta}_{n,T}, X_{t_{k-1}}) - f(\theta^0, X_{t_{k-1}})) \\ \geq \sum_{k=1}^n |f(\tilde{\theta}_{n,T}, X_{t_{k-1}}) - f(\theta^0, X_{t_{k-1}})|^2 \Delta_k t \end{aligned}$$

where θ^0 is assumed to be the true value of the parameter. The above inequality follows easily from the definition of $\tilde{\theta}_{n,T}$. Among the regularity conditions in that paper was the assumption that for any $\delta > 0$

$$\Pr(F^2(\theta, X_0, \delta) = 0) = 0,$$

and for any $c > 0$ there exists a $\delta > 0$ such that

$$E[F^2(\theta, X_0, \delta)] > c$$

where

$$F^2(\theta, x, \delta) = \inf_{\theta_1, |\theta_1 - \theta| \geq \delta} |f(\theta_1, x) - f(\theta, x)|^2$$

and the parameter space Θ is a closed interval of the real line. The function $F^2(\cdot, \cdot)$ is not well defined since the set $\{\theta_1, |\theta_1 - \theta| \geq \delta\}$ could be empty for large δ . In order for the assumption to hold even when the above set is empty, we suggest to define $F^2(\theta, x, \delta) = \infty$ in this case.

The above assumption says that the function $f(\cdot, x)$ must distinguish different values of θ . This is a minimal requirement if we want an estimator based on continuous-discrete data to be consistent. In section 3.3 we shall use a similar assumption in proving strong consistency of the least squares estimator.

Prakasa Rao and Rubin's paper [3] provided a motivation to the study of the limiting behaviour of $\hat{\theta}_{n,T}$ via the function $Q_{n,T}(\theta) - Q_{n,T}(\theta^0)$. They proved that

$$Q_{n,T}(\theta) - Q_{n,T}(\theta^0) \rightarrow R_T(\theta) \quad \text{as } n \rightarrow \infty$$

in probability when T is held fixed, where $R_T(\theta)$ is proportional to the log-likelihood function. They also proved that the maximum likelihood estimator $\hat{\theta}_T$, which is obtained by minimizing $R_T(\theta)$, is strongly consistent, that is, as $T \rightarrow \infty$ (cf. Borkar and Bagchi [7] or section 1.7). The above results led us to suspect that $\hat{\theta}_{n,T}$ converges to θ^0 if we let both $n \rightarrow \infty$ and $T \rightarrow \infty$. In section 3.3 we confirm our suspicion by proving that $\hat{\theta}_{n,T}$ is strongly consistent.

Asymptotic Normality

Prakasa Rao [35] studied the asymptotic distribution of $T^{1/2}(\hat{\theta}_{n,T} - \theta^0)$. In this case $\hat{\theta}_{n,T}$ is defined to be a solution of $\frac{d}{d\theta} Q_{n,T}(\theta) = 0$ or

$$\sum_{k=1}^n (X_{t_k} - X_{t_{k-1}} - f(\theta, X_{t_{k-1}}) \Delta_k t) \frac{d}{d\theta} f(\theta, X_{t_{k-1}}) = 0 \quad (3.2)$$

Applying Taylor's Theorem on the above equation yields

$$T^{1/2}(\tilde{\theta}_{n,T} - \theta^0) = \frac{T^{-1/2} \sum_{k=1}^n U_k f_k^{(1)}}{n^{-1} \sum_{k=1}^n f_k^{(1)2} - T^{-1} \sum_{k=1}^n U_k f_k^{(2)} + o_p(1)} + T^{-1} \sum_{k=1}^n U_k v'_k$$

where for $k = 1, \dots, n$

$$\begin{aligned} \Delta_k X &= X_{t_k} - X_{t_{k-1}}, \\ f_k^{(1)} &= \frac{d}{d\theta} f(\theta, X_{t_k}, \cdot) \Big|_{\theta=\theta^0}, \\ f_k &= f(\theta^0, X_{t_k}, \cdot) \end{aligned} \quad (3.3)$$

If $f_k^{(2)}(\theta)$ denotes the second derivative of $f(\cdot, X_{t_k}, \cdot)$ evaluated at some θ , then

$$v'_k = f_k^{(2)}(\tilde{\theta}_{n,T}) - f_k^{(2)}(\theta^0)$$

and

$$U_k = \Delta_k X - f_k \Delta_k t \quad k = 1, 2, \dots, n$$

where $\tilde{\theta}_{n,T}$ lies between θ^0 and $\tilde{\theta}_{n,T}$. In summary the proof verified the following

$$\begin{aligned} T^{-1/2} \sum_{k=1}^n U_k f_k^{(1)} &\rightarrow N(0, \sigma^2 I(\theta^0)) \\ n^{-1} \sum_{k=1}^n (f_k^{(1)})^2 &\rightarrow I(\theta^0) \\ T^{-1} \sum_{k=1}^n U_k f_k^{(2)} &\rightarrow 0 \\ T^{-1} \sum_{k=1}^n U_k v'_k &= o_p(1) \end{aligned}$$

as $T \rightarrow \infty$ and $n^{-1/2}T \rightarrow 0$, where

$$I(\theta^0) = E \left[f^{(1)}(\theta^0, X_0) \right]^2$$

Then it follows by standard arguments that

$$T^{1/2}(\tilde{\theta}_{n,T} - \theta^0) \rightarrow N \left(0, \frac{\sigma^2}{I(\theta^0)} \right)$$

as $T \rightarrow \infty$ and $n^{-1/2}T \rightarrow 0$.

Asymptotic efficiency

Let μ_θ^T be the measure induced by the process $\{X_t, 0 \leq t \leq T\}$ satisfying (3.1) and μ_W^T be the Wiener measure on $C([0, T], R)$. By Theorem 1.7, μ_θ^T is equivalent to μ_W^T and the Radon-Nykodim derivative is

$$\begin{aligned} L_T(\theta, X) &= \frac{d\mu_\theta^T}{d\mu_W^T} \\ &= \exp \left(\int_0^T f(\theta, X_t) dX_t - \frac{1}{2} \int_0^T f^2(\theta, X_t) dt \right) \end{aligned}$$

Prakasa Rao [35] defines efficiency as follows:

Definition 2 An estimator θ_T of θ^0 is said to be asymptotically efficient if there exist non-random functions $\alpha(\theta)$ and $\beta(\theta)$ such that

$$T^{-1/2} \frac{d}{d\theta} \log L_T(\theta, X) - \alpha(\theta) - \beta(\theta) T^{-1/2} (\theta_T - \theta) \rightarrow 0$$

as $T \rightarrow \infty$.

If we assume that $\log L_T(\theta, X)$ is differentiable with respect to θ and $\frac{d}{d\theta} \log L_T(\theta, X)$ can be obtained by interchanging the integral and the derivative operations (cf. Prakasa Rao [35]), then

$$\begin{aligned} \frac{d}{d\theta} \log L_T(\theta, X) &= \int_0^T f^{(1)}(\theta, X_t) dX_t - \int_0^T f(\theta, X_t) f^{(1)}(\theta, X_t) dt \\ &= \int_0^T f^{(1)}(\theta, X_t) dW_t. \end{aligned}$$

Theorem 1 The estimator $\bar{\theta}_{n,T}$ is asymptotically efficient as $T \rightarrow \infty$ and $n^{-1/2}T \rightarrow 0$.

Proof: see Prakasa Rao [35].

Remark: When applying definition (3.2) to the estimator $\bar{\theta}_{n,T}$, we use the likelihood $L_T(\theta, X)$ as an approximation. More appropriately, one should use the likelihood $L_{n,T}(\theta, X)$

based on measures $\mu_n^{n,T}$ generated by $\{X_{t_k}; t_k \in [0, T], k = 1, 2, \dots, n\}$. However, $L_{n,T}(\theta, X)$ is cumbersome and asymptotically equivalent to $L_T(\theta, X)$

3: Strong Consistency of the LS Estimator

In this section we prove the strong consistency of the least squares estimator $\hat{\theta}_{n,T}$. First we partition the sum of squares $Q_{n,T}(\theta)$ into three components

$$Q_{n,T}(\theta) = Q_{n,T}(\theta^0) - 2 \sum_{k=1}^n u_k h(\theta, X_{t_{k-1}}) + g_n^2(\theta) \quad (3.4)$$

$$\text{where } u_k = X_{t_k} - X_{t_{k-1}} - f(\theta^0, X_{t_{k-1}}) \Delta_k t$$

$$h(\theta, x) = f(\theta, x) - f(\theta^0, x)$$

$$\text{and } g_n^2(\theta) = \sum_{k=1}^n |h(\theta, X_{t_{k-1}})|^2 \Delta_k t$$

Note that minimizing the function $Q_{n,T}$ is equivalent to minimizing the function

$$\begin{aligned} D_{n,T}(\theta) &= T^{-1} \{Q_{n,T}(\theta) - Q_{n,T}(\theta^0)\} \\ &= T^{-1} g_n^2(\theta) - 2T^{-1} \sum_{k=1}^n u_k h(\theta, X_{t_{k-1}}) \end{aligned} \quad (3.5)$$

We will assume throughout that the following conditions hold

A1: The function $f = f(\theta, x)$ is Lipschitz in both variables, that is, there exists continuous functions $K(\cdot)$ and $C(\cdot)$ such that

$$|f(\theta, x) - f(\theta, y)| \leq |K(\theta)| |x - y|$$

$$|f(\theta, x) - f(\phi, x)| \leq |C(x)| |\theta - \phi|$$

for all $\theta, \phi \in \Theta$, $x, y \in R^N$ where

$$\sup_{\theta \in \Theta} |K(\theta)| = K < \infty$$

$$E|C(X_0)|^m = C_m < \infty \text{ for some } m > 16$$

The growth condition

$$|f(\theta, x)|^2 \leq L(1 + |x|^2)$$

for some constant L , is also satisfied.

A2: The parameter space Θ is a compact subspace of a normed linear space.

A3: The process $X = (X_t; t \geq 0)$ is ergodic stationary and therefore the following law of large numbers holds

$$T^{-1} \int_0^T |C(X_s)|^2 ds \rightarrow E|C(X_0)|^2$$

a.s. as $T \rightarrow \infty$, where C is defined in A1.

A4: Let θ^0 be the true parameter value in (3.1). Then

$$E|f(\theta, X_0) - f(\theta^0, X_0)|^2 = 0 \quad \text{iff} \quad \theta = \theta^0$$

The following is the main result of this section.

Theorem 2 Suppose that assumptions A1-A4 and the hypothesis of Lemma 3.2 hold. Then $\hat{\theta}_{n,T} \rightarrow \theta^0$ a.s. as $T \rightarrow \infty$ and $n \rightarrow \infty$ if T^2/n is bounded from above.

Key to the proof of this Theorem is the following result due to Frydman [15].

Theorem 3 If the functions L_n satisfy three conditions:

C1: $L_n(\theta) \rightarrow L(\theta)$ with probability 1 uniformly in $\theta \in \Theta$ as $n \rightarrow \infty$

C2: The limiting deterministic function L is such that

$$L(\theta^0) \leq L(\theta) \quad \text{for all} \quad \theta \in \Theta$$

C3: $L(\theta) = L(\theta^0)$ iff $\theta = \theta^0$.

then $\theta_n \rightarrow \theta^0$ a.s. as $n \rightarrow \infty$ where

$$L_n(\theta_n) = \min_{\theta \in \Theta} L_n(\theta) /$$

Proof: see Theorem 1 in Frydman [35]

Proof: (of Theorem 2) From (3.5)

$$D_{n,T}(\theta) = T^{-1} g_n^2(\theta) - 2T^{-1} \sum_{k=1}^n u_k h(\theta, X_{t_{k-1}})$$

Lemma 3.1 and Lemma 3.2 below implies

$$T^{-1} \sum_{k=1}^n u_k h(\theta, X_{t_{k-1}}) \rightarrow 0 \quad \text{a.s. uniformly in } \theta \in \Theta.$$

By Lemma 3.3

$$T^{-1} g_n^2(\theta) \rightarrow E |f(\theta, X_0) - f(\theta^0, X_0)|^2 \quad \text{a.s.}$$

uniformly in θ . Thus condition C1 of Theorem 3 is satisfied as applied to $D_{n,T}$. The limiting function $D(\theta) = E |f(\theta, X_0) - f(\theta^0, X_0)|^2$ satisfies conditions C2 and C3 by assumption A4. □

We now prove Lemmas 3.1, 3.2 and 3.3. The second term on the RHS of (3.5) can be split further into

$$\begin{aligned} T^{-1} \sum_{k=1}^n u_k h(\theta, X_{t_{k-1}}) &= T^{-1} \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}} - f(\theta^0, X_{t_{k-1}}) \Delta_k t) h(\theta, X_{t_{k-1}}) \\ &= T^{-1} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} [f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})] h(\theta, X_{t_{k-1}}) ds + T^{-1} \sum_{k=1}^n h(\theta, X_{t_{k-1}}) \Delta_k W \end{aligned} \quad (3.6)$$

In the next two Lemmas, we will show that

$$\sup_{\theta \in \Theta} \left| T^{-1} \sum_{k=1}^n u_k h(\theta, X_{t_{k-1}}) \right| \rightarrow 0 \quad \text{a.s.}$$

as $n \rightarrow \infty$

Lemma 3.1 *With probability 1, we have*

$$\sup_{\theta \in \Theta} \left| T^{-1} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} [f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})] h(\theta, X_{t_{k-1}}) ds \right| \rightarrow 0$$

if $T \rightarrow \infty$ and T^2/n is bounded from above.

Proof: For $m > 0$, we have

$$\begin{aligned} E \left\{ \sup_{\theta \in \Theta} \left| T^{-1} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} [f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})] h(\theta, X_{t_{k-1}}) ds \right|^{2m} \right\} \\ = E \left\{ \sup_{\theta \in \Theta} \left| T^{-1} \int_0^T G_n(s) ds \right|^{2m} \right\} \end{aligned} \quad (3.7)$$

where $G_n(s) = [f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})] h(\theta, X_{t_{k-1}})$ if $t_{k-1} \leq s < t_k$. Hölder's inequality implies that (3.7) is bounded by

$$\begin{aligned} &\leq T^{-2m} E \left(\sup_{\theta \in \Theta} T^{2m-1} \int_0^T |G_n(s)|^{2m} ds \right) \\ &= T^{-2m} E \left(\sup_{\theta \in \Theta} T^{2m-1} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} |f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})|^{2m} |h(\theta, X_{t_{k-1}})|^{2m} ds \right) \\ &\leq T^{-1} D_m \sum_{k=1}^n \int_{t_{k-1}}^{t_k} E \left(|f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})|^{2m} |C(X_{t_{k-1}})|^{2m} \right) ds \end{aligned}$$

by A1 where $D_m = \sup_{\theta \in \Theta} |\theta - \theta^0|^{2m} < \infty$. Schwarz's inequality yields:

$$\begin{aligned} &\leq T^{-1} D_m \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \left(E |f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})|^{4m} \right)^{1/2} \left(E |C(X_{t_{k-1}})|^{4m} \right)^{1/2} ds \\ &\leq T^{-1} D_m K(\theta^0) \left(E |C(X_0)|^{4m} \right)^{1/2} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \left(E |X_s - X_{t_{k-1}}|^{4m} \right)^{1/2} ds \end{aligned}$$

by condition A1. Theorem 1.6 yields

$$\leq T^{-1} D_m K(\theta^0) \left(E |C(X_0)|^{4m} \right)^{1/2} B_{2m} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (s - t_{k-1})^m ds$$

$$\begin{aligned}
&= D_m K(\theta^0) B_{2m} \left(E |C(X_0)|^{4m} \right)^{1/2} T^{-1} \sum_{k=1}^n \frac{(\Delta_k t)^{m+1}}{m+1} \\
&\leq \frac{D_m K(\theta^0) B_{2m}}{m+1} \left(E |C(X_0)|^{4m} \right)^{1/2} (T^2/n)^{m/2} n^{-m/2}, \quad m > 4
\end{aligned}$$

Without loss of generality, let $T = T_n \rightarrow \infty$ as $n \rightarrow \infty$. Then Chebyshev's inequality and the above implies that for any $\epsilon > 0$

$$\sum_{n=1}^{\infty} \Pr \left\{ \sup_{\theta \in \Theta} \left| T_n^{-1} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} [f(\theta, X_s) - f(\theta^0, X_{t_{k-1}})] h(\theta, X_{t_{k-1}}) ds \right| > \epsilon \right\} \leq L \sum_{n=1}^{\infty} n^{-m/2} < \infty, \quad (3.8)$$

for some constant L independent of n and T . The Borel-Cantelli lemma and (3.8) yields the desired result. $\square$

Lemma 3.2 Suppose that for each x , $h(\cdot, x)$ lives in a Hilbert space $\mathcal{H}$ having the property that $\mathcal{H}$ is continuously imbedded into $C(\Theta)$. Assume that $\{e_m\}$ is a complete orthonormal basis for $\mathcal{H}$ and that

$$h(\theta, x) = \sum_m a_m(x) e_m(\theta) \quad (3.9)$$

with $|a_m(x)| \leq c_m |x|$ and $\sum_m m^{1+\gamma} c_m^4 < \infty$

for some $\gamma > 0$. Then

$$\sup_{\theta \in \Theta} \left| T^{-1} \sum_{k=1}^n h(\theta, X_{t_{k-1}}) \Delta_k W \right| \rightarrow 0 \quad \text{as } \underline{\hspace{1cm}}$$

if $T = t_n = O(n^\epsilon)$ for some $\epsilon > 0$ and $n^{-1}T \rightarrow 0$.

Example 1 $W^{m,2}(R^d) = \{f \in L^2(R^d); D^\alpha f \in L^2(R^d) \text{ for } 0 \leq |\alpha| \leq m\}$

where for $f = f(x_1, \dots, x_d)$ and $\alpha = (\alpha_1, \dots, \alpha_d) \in Z_+^d$

$$D^\alpha f = \prod_{j=1}^d \frac{\partial^{\alpha_j} f}{\partial x_j^{\alpha_j}} \quad \text{and} \quad |\alpha| = \sum_{j=1}^d \alpha_j,$$

$W^{m,2}(R^d)$ is a separable Hilbert space with inner product

$$\langle f, g \rangle_m = \sum_{0 \leq |\alpha| \leq m} \langle D^\alpha f, D^\alpha g \rangle$$

where $\langle f, g \rangle$ is the inner product in $L^2(R^d)$ (see Adams [2], Theorem 3.5 page 47). Therefore $W^{m,2}(R^d)$ has a complete orthonormal system. Moreover, by the Sobolev imbedding Theorem (cf. Adams [2], page 97), there exists a continuous imbedding $W^{m,2}(R^d) \hookrightarrow C(R^d)$ as long as $m > d/2$.

Remark: The hypothesis of Lemma 3.2 is by no means necessary Borkar and Bagchi [7] have proved that

$$\sup_{\theta \in \Theta} \frac{1}{t} \int_0^t h(\theta, X_s) dW_s \rightarrow 0 \quad \text{a.s.}$$

as $t \rightarrow \infty$, by assuming only A1-A4. However, their proof is long and technical

Proof:(of Lemma 3.2) Since $\mathcal{H} \hookrightarrow C(\Theta)$, it is enough to show that

$$\left\| T^{-1} \sum_{k=1}^n h(\theta, X_{t_{k-1}}) \Delta_k W \right\|_{\mathcal{H}} \rightarrow 0$$

as $T \rightarrow \infty$, $n \rightarrow \infty$. By the representation (3.9)

$$\left\| T^{-1} \sum_{k=1}^n h(\theta, X_{t_{k-1}}) \Delta_k W \right\|_{\mathcal{H}}^2 = \sum_m \left(T^{-1} \sum_{k=1}^n a_m(X_{t_{k-1}}) \Delta_k W \right)^2 \quad (3.10)$$

Let

$$A_{mn}(s) = \sum_{k=1}^n a_m(X_{t_{k-1}}) \chi_{[t_{k-1}, t_k]}(s)$$

where $\chi_{[t_{k-1}, t_k]}$, $k = 1, 2, \dots, n$ are indicator functions. Then

$$T^{-1} \sum_{k=1}^n a_m(X_{t_{k-1}}) \Delta_k W = T^{-1} \int_0^T A_{mn}(s) dW_s \quad (3.11)$$

and the integral on the RHS of (3.11) is well defined since A_{mn} are non-anticipating. By the exponential inequality, we have

$$\Pr \left\{ \int_0^T A_{mn}(s) dW_s - \frac{\alpha}{2} \int_0^T A_{mn}^2(s) ds > \beta \right\} < e^{-\alpha\beta}$$

for any $\alpha, \beta > 0$. Thus

$$\Pr \left\{ \frac{1}{T} \int_0^T A_{mn}(s) dW_s > \frac{\beta}{T} + \frac{\alpha}{2T} \int_0^T A_{mn}^2(s) ds \right\} < e^{-\alpha\beta}$$

and

$$\Pr \left\{ \left| \frac{1}{T} \int_0^T A_{mn}(s) dW_s \right| > \frac{\beta}{T} + \frac{\alpha}{2T} \sum_{k=1}^n a_m^2(X_{t_{k-1}}) \Delta_k t \right\} < 2e^{-\alpha\beta}$$

Since

$$T^{-1} \sum_{k=1}^n a_m^2(X_{t_{k-1}}) \Delta_k t \leq c_m^2 T^{-1} \sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t$$

and by A3

$$T^{-1} \sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t \rightarrow E[X_0^2] > 0$$

there exist a random variable V such that

$$T^{-1} \sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t < V \quad \text{a.s. for all } T > 0, n \geq 1,$$

where $P\{V < \infty\} = 1$. Denote

$$Z_{mn} = \frac{1}{t_n} \int_0^{t_n} A_{mn}(s) dW_s.$$

We have used t_n instead of T in order to emphasize the dependence of T on n . Choose

$$\alpha = \frac{m^a}{t_n^\delta}, \quad \beta = \frac{t_n^\gamma}{m^b}$$

where $\delta < \gamma < 1$, and $\frac{1}{2} < b < a < \frac{1+\gamma}{2}$. Then

$$\Pr \left\{ |Z_{mn}| > \frac{1}{t_n^{1-\gamma} m^b} + \frac{m^a c_m^2 V}{2t_n^\delta} \right\} < 2e^{-m^{a-\delta} t_n^{\gamma-\delta}}$$

The above implies that

$$\begin{aligned}
& \Pr \left\{ \sum_m Z_{mn}^2 > \sum_m \left(\frac{1}{t_n^{1-\gamma} m^b} + \frac{m^a c_m^2 V}{2t_n^\delta} \right)^2 \right\} \\
& \leq \sum_m \Pr \left\{ Z_{mn}^2 > \left(\frac{1}{t_n^{1-\gamma} m^b} + \frac{m^a c_m^2 V}{2t_n^\delta} \right)^2 \right\} \\
& = \sum_m \Pr \left\{ |Z_{mn}| > \frac{1}{t_n^{1-\gamma} m^b} + \frac{m^a c_m^2 V}{2t_n^\delta} \right\} \\
& \leq 2 \sum_m \exp(-m^{a-b} t_n^{\gamma-\delta}) \\
& \leq 2e^{-t_n^{1-\delta}} \sum_m e^{-m^{a-b}}
\end{aligned}$$

Hence

$$\begin{aligned}
& \sum_n \Pr \left\{ \sum_m Z_{mn}^2 > \sum_m \left(\frac{1}{t_n^{1-\gamma} m^b} + \frac{m^a c_m^2 V}{2t_n^\delta} \right)^2 \right\} \\
& \leq 2 \left(\sum_n e^{-t_n^{1-\delta}} \right) \left(\sum_m e^{-m^{a-b}} \right) < \infty
\end{aligned}$$

since $\gamma - \delta > 0$ and $a - b > 0$. The above implies

$$\sum_n \Pr \left\{ \sum_m Z_{mn}^2 > \frac{2}{t_n^{2(1-\gamma)}} \sum_m m^{-2b} + \frac{V^2}{t_n^{2\delta}} \sum_m m^{2a} c_m^4 \right\} < \infty$$

The hypothesis of the Theorem and Borel-Cantelli Lemma implies that

$$\sum_m \left(t_n^{-1} \sum_{k=1}^n a_m(X_{t_{k-1}}) \Delta_k W \right)^2 \rightarrow 0 \quad \text{a.s.}$$

as $n \rightarrow \infty$. This completes the proof of the Lemma. □

Lemma 3.3 Under the assumptions A1-A4 and if T^2/n is bounded, we have

$$T^{-1} g_n^2(\theta) \rightarrow E \left| h(\theta, X_0) \right|^2 \quad \text{a.s. as } n \rightarrow \infty$$

uniformly in θ .

Proof: By the law of large numbers

$$T^{-1} \int_0^T |h(\theta, X_s)|^2 ds \rightarrow E|h(\theta, X_0)|^2 \quad (3.12)$$

a.s. as $T \rightarrow \infty$ for each $\theta \in \Theta$. A1 implies

$$\begin{aligned} T^{-1} \int_0^T |h(\theta, X_s)|^2 ds &\leq T^{-1} |\theta - \theta^0|^2 \int_0^T |C(X_s)|^2 ds \\ &\leq \sup_{\theta \in \Theta} |\theta - \theta^0|^2 T^{-1} \int_0^T |C(X_s)|^2 ds \\ &\leq B \end{aligned}$$

a.s. for some random variable B by A1, A2 and A3. It also follows easily from A1-A4 that

$$\begin{aligned} \left| T^{-1} \int_0^T |h(\theta_1, X_s)|^2 ds - T^{-1} \int_0^T |h(\theta_2, X_s)|^2 ds \right| \\ \leq M |\theta_1 - \theta_2| \quad a.s. \end{aligned} \quad (3.13)$$

for some random variable M and $\theta_1, \theta_2 \in \Theta$. Thus the family of functions

$$\left\{ T^{-1} \int_0^T |h(\cdot, X_s)|^2 ds, T \geq 0 \right\}$$

is equicontinuous. Hence by the Arzelà-Ascoli Theorem, the convergence (3.12) is uniform

Now it is enough to show that

$$T^{-1} \int_0^T |h(\theta, X_s)|^2 ds - T^{-1} g_n^2(\theta) \rightarrow 0 \quad (3.14)$$

a.s. uniformly in θ .

$$\begin{aligned} &E \left\{ \sup_{\theta \in \Theta} \left| \int_0^T |h(\theta, X_s)|^2 ds - g_n^2(\theta) \right|^{2m} \right\} \\ &= E \left\{ \sup_{\theta \in \Theta} \left| \int_0^T |h(\theta, X_s)|^2 ds - \sum_{k=1}^n |h(\theta, X_{t_{k-1}})|^2 \Delta_k t \right|^{2m} \right\} \end{aligned}$$

$$= E \left\{ \sup_{\theta \in \Theta} \left| \sum_{k=1}^n \int_{t_{k-1}}^{t_k} [h(\theta, X_s) - h(\theta, X_{t_{k-1}})] \cdot [h(\theta, X_s) + h(\theta, X_{t_{k-1}})] ds \right|^{2m} \right\}$$

Hölder's inequality implies

$$\begin{aligned} &\leq T^{2m-1} E \left\{ \sup_{\theta \in \Theta} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} |h(\theta, X_s) - h(\theta, X_{t_{k-1}})|^{2m} |h(\theta, X_s) + h(\theta, X_{t_{k-1}})|^{2m} ds \right\} \\ &\leq T^{2m-1} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} E \left\{ \sup_{\theta \in \Theta} |h(\theta, X_s) - h(\theta, X_{t_{k-1}})|^{2m} \times \right. \\ &\quad \times \left. \sup_{\theta \in \Theta} |h(\theta, X_s) + h(\theta, X_{t_{k-1}})|^{2m} \right\} ds \\ &\leq T^{2m-1} K^{2m} 2^{2m} D_m \sum_{k=1}^n \int_{t_{k-1}}^{t_k} E \left\{ |X_s - X_{t_{k-1}}|^{2m} \right. \\ &\quad \left. (|C(X_s)|^{2m} + |C(X_{t_{k-1}})|^{2m}) \right\} \end{aligned}$$

where $D_m = \sup_{\theta \in \Theta} |\theta - \theta^0|^{2m} < \infty$

$$\begin{aligned} &\leq T^{2m-1} K^{2m} 2^{2m+1} D_m \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (E|X_s - X_{t_{k-1}}|^{4m})^{1/2} \\ &\quad (E|C(X_s)|^{4m} + E|C(X_{t_{k-1}})|^{4m})^{1/2} ds \end{aligned}$$

Theorem 1.6 and stationarity yields

$$\begin{aligned} &\leq T^{2m-1} K^{2m} 2^{2m+1} (E|C(X_0)|^{4m})^{1/2} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (s - t_{k-1})^m ds \\ &\leq L_m T^{2m-1} n (T/n)^{m+1} \end{aligned}$$

where $L_m = K^{2m} 2^{2m+1} (E|C(X_0)|^{4m})^{1/2}$. Hence if $m > 4$,

$$\begin{aligned} E \left\{ \sup_{\theta \in \Theta} \left| T^{-1} \int_0^T |h(\theta, X_s)|^2 ds - T^{-1} g_n^2(\theta) \right|^{2m} \right\} &\leq L_m (T/n)^m \\ &\leq L_m (T^2/n)^{m/2} n^{-m/2} \end{aligned}$$

again the Borel-Cantelli argument yields the required result. □

4: Non-Stationary Processes

While the least squares method of the previous section yields consistent and asymptotically normal estimates, it suffers from severe bias when T/n is not small. We shall investigate this aspect more closely in the next chapter. Moreover, the techniques that were used to study the asymptotic properties fail when the process under consideration is assumed to be non-stationary. In this section we present an estimation procedure that attempts to overcome these problems. The main idea in this method is the use of a piecewise deterministic approximation to the solution of (3.1). The assumption of equal spacing of observations will not be required in this section.

Approximation of solution.

Consider a partitioning $0 \leq t_0 < t_1 < \dots < t_n = T$ of the interval $[0, T]$. On each subinterval $[t_{k-1}, t_k]$ define a function $U_k(\theta, t)$ as a solution of

$$dU_k(\theta, t) = f(\theta, U_k(\theta, t))dt, \quad t \in [t_{k-1}, t_k] \quad (3.15)$$

$$U_k(\theta, t_{k-1}) = X_{t_{k-1}}, \quad k = 1, 2, \dots, n$$

The process $U(\theta, \cdot) = \sum_{k=1}^n U_k(\theta, \cdot) \chi_{[t_{k-1}, t_k)}(\cdot)$ is a function of observations on X . In the following Lemma, we show that the mean square difference between X and U vanish asymptotically.

Lemma 4.1: Let $\Delta_n = \max_{1 \leq k \leq n} |t_k - t_{k-1}|$ and suppose that $\Delta_n \rightarrow 0$ as $n \rightarrow \infty$.

Then

$$n^{-1} \sum_{k=1}^n E |X_{t_k} - U_k(\theta, t_k)|^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

Proof:

From (2.1) and (2.2) we have

$$X_t = X_{t_{k-1}} + \int_{t_{k-1}}^t f(\theta, X_s) ds + W_t - W_{t_{k-1}}$$

$$U_k(\theta, t) = X_{t_{k-1}} + \int_{t_{k-1}}^t f(\theta, U_k(\theta, s)) ds$$

So that

$$E|X_t - U_k(\theta, t)|^2 \leq 2E \left| \int_{t_{k-1}}^t (f(\theta, X_s) - f(\theta, U_k(\theta, s))) ds \right|^2 + 2E|W_t - W_{t_{k-1}}|^2$$

$$\leq 2(t - t_{k-1}) \int_{t_{k-1}}^t E|f(\theta, X_s) - f(\theta, U_k(\theta, s))|^2 ds + 2(t - t_{k-1})$$

$$\leq 2(t - t_{k-1})K \int_{t_{k-1}}^t E|X_s - U_k(\theta, s)|^2 ds + 2(t - t_{k-1})$$

by the Lipschitz property of $f(\theta, \cdot)$. Now let

$$\varphi(t) = E|X_t - U_k(\theta, t)|^2, \quad t \in [t_{k-1}, t_k]$$

The above inequality reduces to

$$\varphi(t) \leq 2\Delta_k K \int_{t_{k-1}}^t \varphi(s) ds + 2(t - t_{k-1}), \quad \Delta_k t = t_k - t_{k-1}$$

Gronwall's inequality yields

$$\varphi(t) \leq 2K\Delta_k t \int_{t_{k-1}}^t e^{2K\Delta_k t(t-s)} 2(s - t_{k-1}) ds + 2(t - t_{k-1})$$

$$= (e^{2K\Delta_k t(t-t_{k-1})} - 1) / \Delta_k t$$

Hence

$$E|X_{t_k} - U_k(\theta, t_k)|^2 \leq \frac{(e^{2K(\Delta_k t)^2} - 1)}{\Delta_k t} \leq 2K\Delta_k t e^{2K(\Delta_k t)^2}$$

and

$$n^{-1} \sum_{k=1}^n E|X_{t_k} - U_k(\theta, t_k)|^2 \leq 2K\Delta_n e^{2K\Delta_n^2}$$

Above we have used the fact that $\frac{e^x - 1}{x} \leq e^x$ for $x > 0$. This completes the proof of the Lemma. □

Definition 3 Consider the following quadratic function

$$Q_n(\theta) = \sum_{k=1}^n |X_{t_k} - U_k(\theta, t_k)|^2 \quad (3.16)$$

The modified least squares estimate of θ is defined to be the quantity $\hat{\theta}_n \in \Theta$ which minimises $Q_n(\theta), \theta \in \Theta$

Example 2 Consider the linear stochastic differential equation

$$dX_t = \theta X_t dt + dW_t, \quad X_0 = x_0, \quad \theta \in R^{N \times N}, \quad X_t \in R^N \quad (3.17)$$

for each $t \in [0, T]$. By definition (3.15),

$$U_k(\theta, t) = e^{\theta(t-t_{k-1})} X_{t_{k-1}}, \quad t \in [t_{k-1}, t_k] \quad (3.18)$$

and (3.16) becomes $\sum_{k=1}^n |X_{t_k} - e^{\theta \Delta_k} X_{t_{k-1}}|^2$. We shall return to this example shortly

Now let

$$h_k(\theta) = U_k(\theta, t_k) - U_k(\theta^0, t_k)$$

$$g_n^2(\theta) = n^{-1} \sum_{k=1}^n |h_k(\theta)|^2$$

The following is the main result

Theorem 4: Suppose that given $\epsilon > 0$, there exists $\eta > 0$ such that

$$\lim_{n \rightarrow \infty} \Pr \left(\omega, \inf_{|\theta - \theta^0| > \epsilon} g_n^2(\omega, \theta) > \eta \right) = 1$$

Then $\hat{\theta}_n \rightarrow \theta^0$ in probability as $n \rightarrow \infty$ and $\Delta_n = \max_{1 \leq k \leq n} |t_k - t_{k-1}| \rightarrow 0$

Proof: Suppose that θ^0 is the true value in (3.1). By definition of $\hat{\theta}_n$,

$$\begin{aligned} Q(\theta^0) &\geq Q(\hat{\theta}_n) = \sum_{k=1}^n |X_{t_k} - U_k(\hat{\theta}_n, t_k)|^2 \\ &= \sum_{k=1}^n |X_{t_k} - U_k(\theta^0, t_k) + U_k(\theta^0, t_k) - U_k(\hat{\theta}_n, t_k)|^2 \\ &= Q(\theta^0) - 2 \sum_{k=1}^n (X_{t_k} - U_k(\theta^0, t_k)) \cdot (U_k(\hat{\theta}_n, t_k) - U_k(\theta^0, t_k)) \\ &\quad + \sum_{k=1}^n |U_k(\hat{\theta}_n, t_k) - U_k(\theta^0, t_k)|^2 \end{aligned}$$

Therefore

$$2 \sum_{k=1}^n (X_{t_k} - U_k(\theta^0, t_k)) \cdot (U_k(\bar{\theta}_n, t_k) - U_k(\theta^0, t_k)) \geq \sum_{k=1}^n |U_k(\bar{\theta}_n, t_k) - U_k(\theta^0, t_k)|^2$$

The above inequality yields

$$2n^{-1} \sum_{k=1}^n Z_k(\theta^0) \cdot h_k(\bar{\theta}_n) \geq g_n^2(\bar{\theta}_n)$$

where

$$Z_k(\theta) = X_{t_k} - U_k(\theta, t_k)$$

which implies (using Schwarz's inequality)

$$4n^{-1} \sum_{k=1}^n |Z_k(\theta^0)|^2 \geq g_n^2(\bar{\theta}_n). \quad (3.19)$$

Lemma 4.1 and (3.19) yields

$$E[g_n^2(\bar{\theta}_n)] \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.20)$$

Now

$$\begin{aligned} \Pr(g_n^2(\bar{\theta}_n) \leq \eta) &= \Pr(g_n^2(\bar{\theta}_n) \leq \eta, |\bar{\theta}_n - \theta^0| \leq \epsilon) + \Pr(g_n^2(\bar{\theta}_n) \leq \eta, |\bar{\theta}_n - \theta^0| > \epsilon) \\ &\leq \Pr(|\bar{\theta}_n - \theta^0| \leq \epsilon) + \Pr(A_{\epsilon, \eta}^c) \end{aligned}$$

because

$$\{\omega; g_n^2(\omega, \bar{\theta}_n(\omega)) \leq \eta, |\bar{\theta}_n(\omega) - \theta^0| > \epsilon\} \subset A_{\epsilon, \eta}^c$$

where

$$A_{\epsilon, \eta} = \left\{ \omega; \inf_{|\theta - \theta^0| > \epsilon} g_n^2(\omega, \theta) > \eta \right\}$$

Hence

$$\Pr(|\bar{\theta}_n - \theta^0| > \epsilon) \leq \Pr(g_n^2(\bar{\theta}_n) > \eta) + \Pr(A_{\epsilon, \eta}^c) \quad (3.21)$$

In view of (3.20), (3.21) and the hypothesis of the theorem we conclude that $\bar{\theta}_n \rightarrow \theta^0$ in probability as $n \rightarrow \infty$ and $\Delta_n \rightarrow 0$. □

Example 3 (continuation of example 2) We specialize to the case where $N = 1$, $\Theta = \mathbb{R}^1$, $\Delta_k t = T/n$ and $T = O(n^\alpha)$, $0 < \alpha < 10$ and show that the hypothesis of Theorem 4 is satisfied. From (3.18);

$$\begin{aligned} |U_k(\theta, t_k) - U_k(\theta^0, t_k)|^2 &= X_{t_{k-1}}^2 |e^{\theta \Delta_k t} - e^{\theta^0 \Delta_k t}|^2 \\ &= X_{t_{k-1}}^2 e^{2\theta^0 \Delta_k t} |e^{(\theta - \theta^0) \Delta_k t} - 1|^2 \\ &\geq X_{t_{k-1}}^2 (\Delta_k t)^2 e^{2\theta^0 \Delta_k t} |\theta - \theta^0|^2 \\ &\geq X_{t_{k-1}}^2 (\Delta_k t)^2 |\theta - \theta^0|^2 \end{aligned}$$

We have assumed w.l.o.g. that $\theta > \theta^0$. Hence

$$n^{-1} \sum_{k=1}^n |U_k(\theta, t_k) - U_k(\theta^0, t_k)|^2 > |\theta - \theta^0|^2 n^{-3} T^2 \sum_{k=1}^n X_{t_{k-1}}^2$$

By Lemma 5.4 below, the RHS $\rightarrow \infty$ as $n \rightarrow \infty$. Therefore in this case the hypothesis of Theorem 4 is satisfied.

5: The Ornstein-Uhlenbeck Process

We now specialize to example 3 and prove a strong consistency result. We shall consider the following linear SDE

$$dX_t = \theta X_t dt + dW_t, \quad X_0 = x_0 \quad (3.22)$$

where $\theta > 0$ (W_t ; $t \geq 0$) is a real standard Wiener process. From now on, unless otherwise stated, we will assume that $\Delta_k t = \Delta$ for all $k = 1, 2, \dots$ so that $T = \Delta n$. This assumption is not essential but simplifies the proofs significantly.

For $t > s$, we can solve (3.22) to obtain

$$X_t = X_s e^{\theta(t-s)} + e^{\theta t} \int_s^t e^{-\theta u} dW_u \quad (3.23)$$

The estimate of θ will be a number $\hat{\theta}_n$ which minimizes

$$Q(\theta) = \sum_{k=1}^n \frac{(X_{t_k} - X_{t_{k-1}} e^{\theta \Delta_k t})^2}{\Delta_k t} \quad (3.24)$$

The above expression is the same as (3.16) with each term in the summation weighted by multiplication by $1/\Delta_k t$. The motivation behind this modification is as follows. Let

$$Z_k = e^{\theta t_k} \int_{t_{k-1}}^{t_k} e^{-\theta u} dW_u, \quad k = 1, 2, \dots$$

Then $(Z_k; k = 1, 2, \dots)$ is a sequence of independent random variables with mean zero and variance $\text{var}(Z_k) = (e^{2\theta \Delta_k t} - 1)/2\theta$. If the intervals $\{\Delta_k t\}_k^\infty$ are small, then $(e^{2\theta \Delta_k t} - 1)/2\theta \approx \Delta_k t$. (3.23) and (3.24) implies $Q(\theta) = \sum_{k=1}^n \frac{Z_k^2}{\Delta_k t}$ which is proportional to the log-likelihood function.

Lemma 5.1 Suppose that for some real number $\alpha > -1$, $\Delta_k t = O(n^\alpha)$. Then for any real number β

$$n^{-\beta} \sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t \rightarrow \infty \quad \text{a.s. as } n \rightarrow \infty.$$

Proof: From (3.23) we obtain the form

$$X_{t_k} = X_{t_{k-1}} e^{\theta(t_k - t_{k-1})} + e^{\theta t_k} \int_{t_{k-1}}^{t_k} e^{-\theta s} dW_s, \quad k = 1, \dots, n \quad (3.25)$$

or in particular $X_t = X_0 e^{\theta t} + e^{\theta t} \int_0^t e^{-\theta s} dW_s$. Put

$$Y_t = \int_0^t e^{-\theta s} dW_s = e^{-\theta t} X_t - X_0 \quad (3.26)$$

Then $(Y_t, \mathcal{F}_t, t \in [0, T])$ is a zero mean gaussian martingale with

$$E|Y_t|^2 = \int_0^t e^{-2\theta s} ds = \frac{(1 - e^{-2\theta t})}{2\theta} \leq \frac{1}{2\theta}$$

where $\mathcal{F}_t = \sigma(W_s, X_0; s \leq t)$. By the martingale convergence Theorem, there exists an L^2 -random variable Y such that

$$Y_t \rightarrow Y \text{ a.s. as } t \rightarrow \infty. \quad (3.27)$$

Let $\beta_s = e^{-\theta s}$ $s \in [0, T]$ and λ a real number. The fact that

$$M_t(\lambda) = \exp \left(i\lambda \int_0^t \beta_s dW_s + \frac{\lambda^2}{2} \int_0^t \beta_s^2 ds \right)$$

is a martingale with $E(M_t(\lambda)) = 1$ imply that

$$E(e^{i\lambda Y_t}) = \exp \left(\frac{-\lambda^2}{2} \frac{1}{2\theta} (1 - e^{-2\theta t}) \right) \rightarrow \exp \left(\frac{-\lambda^2}{2} \frac{1}{2\theta} \right)$$

as $t \rightarrow \infty$. Hence $Y \sim N(0, \frac{1}{2\theta})$ and together with (3.26) and (3.27) we have

$$e^{-2\theta t} X_t^2 \rightarrow (Y + X_0)^2 \text{ a.s. as } t \rightarrow \infty \text{ and so } Pr((Y + X_0)^2 = 0) = 0$$

Now the Lemma follows from the inequality

$$\begin{aligned} n^{-\beta} \sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t &\geq n^{-\beta} X_{t_{n-1}}^2 \Delta_n t, \\ &= O(e^{2\theta t_{n-1}} n^{\alpha-\beta}) \end{aligned}$$

Lemma 5.2 (Toeplitz) Let $\{a_{nk}; k = 1, 2, \dots, \nu_n, n = 1, 2, \dots\}$ be an array of real numbers such that $\nu_n \rightarrow \infty$ as $n \rightarrow \infty$ and

- (i) $\sum_{k=1}^{\nu_n} a_{nk}$ converges absolutely as $n \rightarrow \infty$
- (ii) $\sum_{k=1}^{\nu_n} a_{nk} = a_n \rightarrow a$ as $n \rightarrow \infty$
- (iii) $a_{nk} \rightarrow 0$ as $n \rightarrow \infty$ for each $k = 1, 2, \dots$

Then $y_k \rightarrow y$ as $k \rightarrow \infty$ implies that

$$\sum_{k=1}^{\nu_n} a_{nk} y_k \rightarrow ay \text{ as } n \rightarrow \infty.$$

The proof of Toeplitz lemma is elementary and shall not be given here. We want to apply Lemma 5.2 to show that

$$e^{-2\theta^0 t_n} \sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t \rightarrow \eta \quad \text{a.s. as } n \rightarrow \infty \quad (3.28)$$

where $P(\eta > 0) = 1$.

Let

$$Y_k = e^{-2\theta^0 t_{k-1}} X_{t_{k-1}}^2 \quad (3.29)$$

$$a_{nk} = e^{-2\theta^0 (t_n - t_{k-1})} \Delta_k t \quad k = 1, 2, \dots, n, n = 1, 2, \dots$$

Assume that $\sum_{k=1}^n a_{nk} \rightarrow a$ as $n \rightarrow \infty$. This assumption holds for example, when $\Delta_k t = \Delta$ for all $k = 1, 2, \dots$. Since in this case

$$\begin{aligned} \sum_{k=1}^n a_{nk} &= e^{-2\theta^0 t_n} \sum_{k=1}^n e^{2\theta^0 t_{k-1}} \Delta_k t \\ &= \Delta e^{-2\theta^0 \Delta n} \sum_{k=1}^n e^{2\theta^0 \Delta (k-1)} \\ &= \frac{\Delta (e^{-2\theta^0 n} - 1)}{(1 - e^{2\theta^0 \Delta})} \end{aligned}$$

Hence

$$\sum_{k=1}^n a_{nk} \rightarrow \frac{\Delta}{(e^{2\theta^0 \Delta} - 1)} \quad \text{as } n \rightarrow \infty$$

The above implies that conditions (i) and (ii) of Lemma 5.2 are satisfied by the array $\{a_{nk}; k = 1, 2, \dots, n, n = 1, \dots\}$. Condition (iii) is immediate from the definition of the array. From the proof of Lemma 5.1, we have $Y_k \rightarrow (X_0 + Y)^2$ a.s. as $k \rightarrow \infty$ and $P([X_0 + Y]^2 > 0) = 1$. By Toeplitz lemma

$$\sum_{k=1}^n a_{nk} Y_k = e^{-2\theta^0 t_n} \sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t \rightarrow a(X_0 + Y)^2$$

as required in (3.28).

We shall need the following generalisation of the strong law of large numbers (SLLN)

Lemma 5.3 Let $\{Y_1, Y_2, \dots\}$ be a sequence of independent random variables with $E(Y_k) = 0$ and $\sum_{k=1}^{\infty} \frac{\text{var}(Y_k)}{a_k^2} < \infty$ where $a_k \rightarrow \infty$ as $k \rightarrow \infty$. Then

$$a_n^{-1} \sum_{k=1}^n Y_k \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty$$

The proof of the above Lemma is essentially the same as that of the classical SLLN (see Billingsley [6]). The following result follows from Lemmas (5.2) and (5.3)

Lemma 5.4 Suppose that the spacings $\{\Delta_k t\}_{k=1}^{\infty}$ are such that all conditions of Lemma 5.2 are satisfied with the array $\{a_{nk}; k = 1, 2, \dots, n, n = 1, 2, \dots\}$ as defined in (3.29). Then

$$U_n = \frac{\sum_{k=1}^n X_{t_{k-1}} Z_k}{\sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t} \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty$$

where $Z_k = e^{\theta^0 t_k} \int_{t_{k-1}}^{t_k} e^{-\theta^0 s} dW_s$.

Remark: Note the resemblance between the above and the condition that

$$\frac{\int_0^t X_s dW_s}{\int_0^t X_s^2 ds} \rightarrow 0 \quad \text{a.s. as } t \rightarrow \infty$$

in the proof of consistency of the maximum likelihood estimator. In the latter case, the condition is both necessary and sufficient. See Liptser and Shirayayev [28], page 210.

Proof: By Schwarz's inequality

$$\begin{aligned} U_n^2 &\leq \frac{\sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t \cdot \sum_{k=1}^n \left(\frac{Z_k^2}{\Delta_k t} \right)}{\left(\sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t \right)^2} \\ &\leq \frac{\sum_{k=1}^n \left(\frac{Z_k^2}{\Delta_k t} \right)}{\sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t} \end{aligned}$$

From (3.28), we have $e^{-2\theta^0 t_n} \sum_{k=1}^n X_{t_k-1}^2 \Delta_k t \rightarrow \eta > 0$. Hence we only need to show that $e^{-2\theta^0 t_n} \sum_{k=1}^n \frac{Z_k^2}{\Delta_k t} \rightarrow 0$ a.s. $Z_k \sim N\left(0, (e^{2\theta^0 \Delta_k t} - 1)/2\theta^0\right)$. Therefore

$$e^{-2\theta^0 t_n} \sum_{k=1}^n \frac{Z_k^2}{\Delta_k t} = e^{-2\theta^0 t_n} \sum_{k=1}^n Y_k + e^{-2\theta^0 t_n} \sum_{k=1}^n m_k \quad (3.30)$$

where

$$m_k = \frac{e^{2\theta^0 \Delta_k t} - 1}{2\theta^0 \Delta_k t}, \quad Y_k = \frac{Z_k^2}{\Delta_k t} - m_k$$

The last term of (3.30)

$$e^{-2\theta^0 t_n} \sum_{k=1}^n m_k \leq e^{-2\theta^0 t_n} \sum_{k=1}^n e^{2\theta^0 \Delta_k t} \rightarrow 0 \text{ as } n \rightarrow \infty$$

if $\sum_{k=1}^{\infty} e^{-2\theta^0 t_k} e^{2\theta^0 \Delta_k t} < \infty$. But $\sum_{k=1}^{\infty} e^{-2\theta^0 t_{k-1}} < \infty$ as long as t_k does not grow too slowly. For example if $t_k = O(k^\gamma)$, $\gamma > 0$ then the above is satisfied. Now consider the first term on the RHS of (3.30). Let $a_n = e^{2\theta^0 t_n}$. Then

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\text{var} Y_k}{a_k^2} &= \sum_{k=1}^{\infty} e^{-4\theta^0 t_k} K \left(\frac{e^{2\theta^0 \Delta_k t} - 1}{2\theta^0 \Delta_k t} \right)^2 \\ &\leq K \sum_{k=1}^{\infty} e^{-4\theta^0 t_{k-1}} < \infty \end{aligned}$$

for some constant $K > 0$. By Lemma 5.3, we conclude that $a_n^{-1} \sum_{k=1}^n Y_k \rightarrow 0$ a.s. as $n \rightarrow \infty$. This completes the proof of the Lemma. $\square$

Theorem 5 If $\Delta_k t = \Delta$, for all $k = 1, 2, \dots$, Then the modified least squares estimator $\hat{\theta}_n$ is strongly consistent.

Proof: From (3.24)

$$Q_n(\theta) = \sum_{k=1}^n \frac{(X_{t_k} - X_{t_{k-1}} e^{\theta \Delta})^2}{\Delta}$$

setting $\frac{dQ_n(\theta)}{d\theta} = 0$ we get

$$\begin{aligned} \sum_{k=1}^n X_{t_{k-1}} (X_{t_k} - X_{t_{k-1}} e^{\theta \Delta}) &= 0 \\ (e^{\theta \Delta} - e^{\theta^0 \Delta}) \sum_{k=1}^n X_{t_{k-1}}^2 &= \sum_{k=1}^n X_{t_{k-1}} Z_k \end{aligned}$$

where $Z_k = e^{\theta^0 t_k} \int_{t_{k-1}}^{t_k} e^{-\theta^0 s} dW_s$, $k = 1, 2, \dots, n$ are iid random variables. Hence

$$\left| e^{\theta_n \Delta} - e^{\theta^0 \Delta} \right| = \frac{\sum_{k=1}^n X_{t_{k-1}} Z_k}{\sum_{k=1}^n X_{t_{k-1}}^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

by Lemma 5.4. []

Another consequence of Lemma 5.4 is the following result concerning the least squares estimator $\tilde{\theta}_{n,T}$.

Theorem 6 *If the conditions of Lemmas 5.1 and 5.2 are satisfied with $\Delta_k t = \Delta$, $k = 1, \dots, n$. Then*

$$\tilde{\theta}_{n,T} \xrightarrow{a.s.} \theta^0 \quad \text{as } n \rightarrow \infty, \Delta \rightarrow 0$$

Proof: By (3.22) and Definition 1 we have

$$\begin{aligned} \tilde{\theta}_{n,T} &= \frac{\sum_{k=1}^n X_{t_{k-1}} (X_{t_k} - X_{t_{k-1}})}{\sum_{k=1}^n X_{t_{k-1}}^2 \Delta t} \\ &= \frac{\sum_{k=1}^n X_{t_{k-1}} \left(X_{t_{k-1}} e^{\theta^0 \Delta} + e^{\theta^0 t_k} \int_{t_{k-1}}^{t_k} e^{-\theta^0 s} dW_s - X_{t_{k-1}} \right)}{\sum_{k=1}^n X_{t_{k-1}}^2 \Delta} \quad \text{by (3.23)} \\ &= \frac{e^{\theta^0 \Delta} - 1}{\Delta} + U_n \quad \text{using the notation of Lemma 5.4} \end{aligned}$$

Hence

$$\tilde{\theta}_{n,T} = \theta^0 + o(\Delta) + U_n$$

where $o(\Delta) \rightarrow 0$ as $\Delta \rightarrow 0$. The above and Lemma 5.4 yields the desired result []

Chapter 4

SIMULATION STUDIES

1: Introduction

In the preceding chapters, we have mainly studied pointwise convergence of the ML and LS estimators. Other statistical properties such as bias and mean-square-error (MSE) are generally difficult to study analytically (cf. Liptser and Shirayev [28]). We shall therefore use simulation techniques to get an insight on the moment behaviour of the estimators. We shall confine ourselves to two simple linear models, the mean-field model of section 2.5 and the Ornstein-Uhlenbeck process of section 3.5.

In section 3 we shall demonstrate that for any fixed interval of observation $[0, T]$, $\text{MSE}(\hat{\theta}_{N,T})$ is a decreasing function of N , where N is the number of equations in (2.34) and $\hat{\theta}_{N,T}$ is the ML estimator of θ^0 . The notation $\hat{\theta}_{N,T}$ is used in this chapter instead of $\hat{\theta}_N$ because we shall also investigate the dependence of $\hat{\theta}_N$ on T . We shall see that for a fixed N , $\text{MSE}(\hat{\theta}_{N,T})$ appears to behave like an exponential decay function of T . Also in section 3, we will investigate the behaviour of $\text{MSE}(\hat{\theta}_{n,T})$ for the same model, where $\hat{\theta}_{n,T}$ is the LS estimator defined in section 3.2. In this case too, the MSE is a decreasing function of N . In section 4 we compare the estimators $\hat{\theta}_{n,T}$ and the modified LS estimator $\tilde{\theta}_{n,T}$ arising from the Ornstein-Uhlenbeck process (see section 3.4). It is shown that $\tilde{\theta}_{n,T}$ improves on $\hat{\theta}_{n,T}$ both in terms of bias and MSE. For $\tilde{\theta}_{n,T}$ we also determine the optimal rate at which T should go to infinity.

2: Description of the Simulation Procedure

In order to describe the simulation schemes of the subsequent sections, we shall use the

following Itô stochastic differential equation

$$dX(t) = f(\theta, X(t))dt + dW(t), \quad X(0) = x_0 \quad (4.1)$$

where $(W(t), t \geq 0)$ is a Wiener process in R^N with independent components, $f(\theta, x)$ is a function $f: R^d \times R^N \rightarrow R^N$ and $X = (X(t), t \geq 0)$ is a solution of (4.1) in R^N .

Since it is impossible to simulate a continuous trajectory of X on $[0, T]$ we shall use its discrete approximation $\tilde{X}$. This estimate is obtained through the following procedure. The interval $[0, T]$ is divided into η subintervals $0 < t_1 < \dots < t_\eta$. We then use the Euler scheme

$$\tilde{X}_k = \tilde{X}_{k-1} + f(\theta^0, \tilde{X}_{k-1})\Delta_k t + \Delta_k W, \quad \tilde{X}_0 = x_0 \quad (4.2)$$

to compute η observations of $\tilde{X}$. In (4.2), $\Delta_k t = t_k - t_{k-1}$, $\Delta_k W = W_{t_k} - W_{t_{k-1}}$, $k = 1, \dots, \eta$. The increments of the Wiener process $\Delta_k W$ are simulated as independent Normal

R^N -valued random vectors with mean zero and covariance matrix $\text{diag}(\sqrt{\Delta_1 t}, \dots, \sqrt{\Delta_N t})$.

We shall only consider the case when $\Delta_k t = T/\eta$, $k = 1, \dots, \eta$. If we define

$$\tilde{X}^\eta(t) = \tilde{X}_{[\eta t / T]}$$

where $[x]$ = largest integer which is less or equal to x , then $\tilde{X}^\eta(t)$ converges to $X(t)$ almost surely and in mean square (cf. E. Pardoux and D. Talay [33]) as $\eta \rightarrow \infty$. The error of approximation of X can therefore be minimized by taking η large.

We can now approximate functions of sample paths $\phi(X) = \phi(X(t), 0 \leq t \leq T)$ by $\phi(\tilde{X}) = \phi(\tilde{X}^\eta(t), 0 \leq t \leq T)$. In particular for each simulated trajectory $\tilde{X}$ we can calculate the parameter estimate $\hat{\theta}_{N,T}$ as $\hat{\theta}_{N,T} = \hat{\theta}_{N,T}(\tilde{X})$. In order to estimate the moments of $\phi(X)$ we generate M independent replications of $\tilde{X}$. Let $\tilde{X}_{(j)}$ be the j -th replication. Then the p -th moment of $\phi(X)$ is estimated by

$$\mathcal{E}_{M,p}(\phi) = M^{-1} \sum_{j=1}^M |\phi(\tilde{X}_{(j)})|^p \quad (4.3)$$

By the law of large numbers

$$\ell_{M,p}(\phi) \rightarrow E|\phi(\tilde{X})|^p \quad \text{a.s. as } M \rightarrow \infty$$

In addition if η is large then $E|\phi(\tilde{X})|^p$ will be close to $E|\phi(X)|^p$. The preceding statement holds under some regularity conditions on ϕ (cf. E. Pardoux and D. Talay [33]).

3: The Mean Field Model

In this section we investigate numerically the behaviour of the means and MSE's of the ML estimator $\hat{\theta}_{N,T}$ and the LSE $\tilde{\theta}_{n,T}$ of $\theta = (\alpha, \beta)$ from model (2.34). We begin with $\hat{\theta}_{N,T}$ given by

$$\hat{\theta}_{N,T} = \frac{1}{\Delta_N(T)} \begin{pmatrix} B_N(T) & B_N(T) \\ B_N(T) & A_N(T) \end{pmatrix} \begin{bmatrix} \sum_{j=1}^N \int_0^T X_j(s) dX_j(s) \\ - \sum_{j=1}^N \int_0^T (X_j(s) - \bar{X}_N(s)) dX_j(s) \end{bmatrix} \quad (4.4)$$

where $\Delta_N(T) = \det[I_N(T)]$ (see page 28) and

$$\begin{aligned} A_N(T) &= N^{-1} \sum_{j=1}^N \int_0^T X_j^2(s) ds \\ B_N(T) &= N^{-1} \sum_{j=1}^N \int_0^T (X_j(s) - \bar{X}_N(s))^2 ds \end{aligned} \quad (4.5)$$

The ordinary integrals (4.5) are approximated by finite Riemann sums. For example

$$\begin{aligned} A_N(T) &= N^{-1} \sum_{j=1}^N \sum_{k=1}^{\eta} X_j^2(t_{k-1}) \Delta_k t \\ &= N^{-1} \cdot \frac{T}{N} \sum_{j=1}^N \sum_{k=1}^{\eta} X_j^2(t_{k-1}). \end{aligned} \quad (4.6)$$

The stochastic integrals on the RHS of (4.4) can be simplified by Itô's formula

$$\sum_{j=1}^N \int_0^T X_j(s) dX_j(s) = \sum_{j=1}^N \frac{1}{2} (X_j^2(T) - X_j^2(0) - T) \quad (4.7)$$

To calculate (4.7) we use the simulated trajectory $\tilde{X}$. Also

$$\sum_{j=1}^N \int_0^T (X_j(s) - \bar{X}_N(s)) dX_j(s) = \sum_{j=1}^N \int_0^T X_j(s) dX_j(s) - N \int_0^T \bar{X}_N(s) d\bar{X}_N(s) \quad (4.8)$$

The first term on the RHS of (4.8) is given by (4.7). From (2.34) we have

$$d\bar{X}_N(t) = \alpha \bar{X}_N(t) dt + N^{-1/2} dW(t), \quad \bar{X}_N(0) = x_N(0) \quad (4.9)$$

where $(W(t), t \geq 0)$ is a Wiener process on the real line. Hence (4.9) and Itô's formula yields

$$N \int_0^T \bar{X}_N(s) d\bar{X}_N(s) = \frac{1}{2} (N \bar{X}_N^2(T) - N \bar{X}_N^2(0) - T) \quad (4.10)$$

and (4.10) is approximated by using $\tilde{X}$ instead of $\bar{X}$.

In this experiment we have chosen $\theta^0 = (5.0, 2.0)$ that is $\alpha = 5.0$ and $\beta = 2.0$ as the true parameter values, $n = 1000$ and $M = 30$. We carried out the experiment for three different values of N , namely $N = 5, 7$ and 10 . For each N we repeated the experiment ten times using different values of T . We chose $T = 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6$. Following (4.3) the approximate mean and MSE are calculated as

$$\begin{aligned} \hat{E}(\hat{\theta}_{N,T}) &= \frac{1}{30} \sum_{j=1}^{30} \hat{\theta}_{N,T,j}(\tilde{X}) \\ \widehat{MSE}(\hat{\theta}_{N,T}) &= \frac{1}{30} \sum_{j=1}^{30} [\hat{\theta}_{N,T,j}(\tilde{X}) - \theta^{0,2}]^2 \end{aligned} \quad (4.11)$$

where $\hat{E}$ denotes the estimate of the expected value of and $\hat{\theta}_{N,T,j}(\tilde{X})$ denotes the estimator of θ^0 based on the j -th simulated replication of $\tilde{X}$. Below we present a graphical summary of our results.

Figure 1
MSE as a Function of T , for $N = 5, 7, 10$

Figure 1 illustrates the relative gain in efficiency of $\theta_{N,T}$ when N is large. For example if the desired maximum error is 0.05 when $N = 10$ then one needs to observe the process for only 70% of the time needed when $N = 5$. In addition to this, the curves in figure 1 provides an estimate of the function $f(N, T) = E(\theta_{N,T})$. For each N f appears to be exponential decay function of T .

Figure 2
Log(MSE) as a Function of T , for $N = 5, 7, 10$

Figure 2 led us to suspect that

$$\text{MSE}(\hat{\theta}_N, T) = a(N) \exp(-\gamma T)$$

112

where $a(\cdot)$ is a decreasing function and $\gamma > 0$.

We now turn to the LS estimator $\hat{\theta}_N, T$. This is given by definition 2.1 as

$$\hat{\theta}_N, T = \frac{1}{\Delta_N(T)} \begin{bmatrix} B_{N,T} & B_{N,T} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^N \sum_{j=1}^N A_i(t_{i,j}) (A_i(t_{i,j}) - A_j(t_{i,j})) \\ \sum_{i=1}^N \sum_{j=1}^N (X_i(t_{i,j}) - X_j(t_{i,j})) A_i(t_{i,j}) - A_j(t_{i,j}) \end{bmatrix}$$

where

$$\Delta_n(T) = (A_{n,T} - B_{n,T}) B_{n,T}$$

$$A_{n,T} = n^{-1} \sum_{k=1}^n \sum_{j=1}^N X_j^2(t_{k-1}) \Delta_k t \quad (4.13)$$

$$B_{n,T} = n^{-1} \sum_{k=1}^n \sum_{j=1}^N (X_j(t_{k-1}) - \bar{X}_N(t_{k-1}))^2 \Delta_k t$$

Note that $\hat{\theta}_{n,T}$ has the same form as $\hat{\theta}_{N,T}$ except that the stochastic integrals in (4.4) are replaced by the Itô form of the Riemann-Stieltjes sums

We used the following values $\eta = 1000$, $M = 30$, $N = 2, 5$, $n = 25, 64, 100, 400$ and $T = 0.2n^{1/2}$. The trajectory $\bar{X}$ is simulated over the interval $[0, 5.0]$. To calculate the mean and mean square errors we used formulas (4.11). The following tables summarises the results of the experiment

Table 1a
 $N = 2$

n	$\hat{E}(\hat{\alpha}_{n,T})$	$\hat{E}(\hat{\beta}_{n,T})$	$\widehat{MSE}(\hat{\alpha}_{n,T})$	$\widehat{MSE}(\hat{\beta}_{n,T})$	Total MSE
25	4.85205	2.27319	.427E+01	.101E+02	.144E+02
64	5.16679	2.25635	.307E-01	.163E+00	.194E+00
100	5.11740	2.10362	.149E-01	.489E-01	.638E-01
400	5.00074	1.97150	.146E-04	.457E-01	.457E-01

Table 1b
 $N = 5$

n	$\hat{E}(\hat{\alpha}_{n,T})$	$\hat{E}(\hat{\beta}_{n,T})$	$\widehat{MSE}(\hat{\alpha}_{n,T})$	$\widehat{MSE}(\hat{\beta}_{n,T})$	Total MSE
25	4.89929	1.79344	.496E+01	.507E+01	.100E+02
64	5.14619	2.08921	.276E+00	.253E+00	.529E+00
100	5.09977	2.06913	.290E-01	.225E-01	.515E-01
400	4.99950	2.00391	.509E-04	.209E-03	.260E-03

Both Tables 1a and 1b show that there is a steady improvement in bias and MSE as n and T increases. It is interesting to note that the performance of $\alpha_{n,T}$ is the same in both

tables while $\beta_{n,T}$ improves significantly in Table 1b. However the combined performance as given by the total MSE is better in Table 1b. Hence it appears that the LSE $\bar{\theta}_{n,T}$ also improves when N is large. It is also evident from both tables that the MSE and bias decrease as n increases.

Another question regarding the LS estimator $\bar{\theta}_{n,T}$ is how it compares with the ML estimator $\hat{\theta}_{N,T}$. To study this problem, the following experiment was conducted. We fixed T and N at $T = 1.5$ and $N = 5$, where $[0, T]$ is the observation time interval and N is the number of components in system (2.34). We shall therefore drop the subscripts T and N in the notation $\bar{\theta}_{n,T}$ and $\hat{\theta}_{N,T}$ and denote the LSE and the MLE simply as $\bar{\theta}_n$ and $\hat{\theta}$ respectively.

We chose $\eta = 1000$, $M = 30$, $\theta^0 = (5.0, 2.0)^*$ and varied n as shown in Table 2. The purpose of this experiment was to see whether $\bar{\theta}_n$ converge to $\hat{\theta}$ as n increases and Δ_n goes to zero, where $\Delta_n = T/n$. In Table 2

$$\hat{P}(|\bar{\theta}_n - \hat{\theta}| > \epsilon) = M^{-1} \sum_{j=1}^M X_{(|\bar{\theta}_n - \hat{\theta}| > \epsilon)} \quad (4.14)$$

is the proportion of estimates $\bar{\theta}_{n,j}$, $j = 1, \dots, M$ which are at a distance of more than ϵ units from $\hat{\theta}$. We chose $\epsilon = 0.15$.

Table 2
The Convergence of LSE to MLE

n	$\hat{E}(\bar{\alpha}_{n,T})$	$\hat{E}(\bar{\alpha}_{n,T})$	$\hat{E} \bar{\theta}_n - \hat{\theta} ^2$	$\hat{P}(\bar{\theta}_n - \hat{\theta} > \epsilon)$
100	5.17173	2.11036	.0512365	1.000
125	5.13281	2.03634	.0338198	.733
200	5.07514	2.05500	.0172221	.167
250	5.05811	2.04241	.0129472	.133
500	5.01833	2.01396	.0072587	.067

The MLE in this case is $\hat{\theta} = (5.01834, 1.95269)^*$ with $MSE(\hat{\theta}) = 0.002307$ Table

2 convincingly suggests that $\hat{\theta}_n$ converges to θ both in probability and in mean square as $\Delta_n \rightarrow 0$ and $n \rightarrow \infty$. This confirms the results of Le Breton [8] (cf. section 2.8 page 36).

4: The Ornstein-Uhlenbeck Process

In this section we compare the LS estimator $\hat{\theta}_{n,T}$ defined in section 3.1 and the modified LSE $\tilde{\theta}_{n,T}$ of section 3.4. We shall simulate the following SDE

$$dX_t = \theta^0 X_t dt + dW_t, \quad X_0 = 0.5, \quad \theta^0 > 0 \quad (4.14)$$

We shall take advantage of the form (3.25) to obtain an estimate of X which is better than the Euler estimate. By (3.25) and (4.14)

$$X_{t_k} = X_{t_{k-1}} e^{\theta^0 \Delta_k t} + e^{\theta^0 t_k} \int_{t_{k-1}}^{t_k} e^{-\theta^0 s} dW_s, \quad k = 1, \dots, n \quad (4.15)$$

where $\Delta_k t = T/n$. Now let $U_k = e^{\theta^0 t_k} \int_{t_{k-1}}^{t_k} e^{-\theta^0 s} dW_s$, $k = 1, \dots, n$. Then $(U_k, k = 1, \dots, n)$ are independent Normal random variables with mean zero and variance $(e^{\theta^0 T/n} - 1)/2\theta^0$. Therefore to simulate X we first generate $(U_k, k = 1, \dots, n)$ using the IMSL subroutine GGNML and then use (4.15) to calculate $(X_{t_k}, k = 1, \dots, n)$. The estimators $\hat{\theta}_{n,T}$ and $\tilde{\theta}_{n,T}$ are given by

$$\begin{aligned} \hat{\theta}_{n,T} &= \frac{\sum_{k=1}^n X_{t_{k-1}} (X_{t_k} - X_{t_{k-1}})}{\sum_{k=1}^n X_{t_{k-1}}^2 \Delta_k t} \\ \tilde{\theta}_{n,T} &= \frac{n}{T} \log \left\{ \frac{|\sum_{k=1}^n X_{t_{k-1}} X_{t_k}|}{\sum_{k=1}^n X_{t_{k-1}}^2} \right\} \end{aligned} \quad (4.16)$$

We have chosen $\theta^0 = 2.0$, $T = n^{1/2}$, $n = 10, 50, 100, 250$, and 500 ; $M = 100$. Then the interval between observations is $\Delta_n = T/n = n^{-1/2}$. As in (4.11) the mean and MSE are estimated

by

$$\begin{aligned}\hat{E}(\bar{\theta}_{n,T}) &= \frac{1}{30} \sum_{l=1}^{30} \bar{\theta}_{n,T,l} \\ \widehat{MSE}(\bar{\theta}_{n,T}) &= \frac{1}{30} \sum_{l=1}^{30} (\bar{\theta}_{n,T,l} - \theta^0)^2\end{aligned}\quad (4.17)$$

where $\theta^0 = 2.0$ and $(\bar{\theta}_{n,T,l}, l = 1, \dots, 30)$ are independent estimates of θ^0 obtained via equation (4.16). The same calculations are carried out for $\bar{\theta}_{n,T}$. The following table gives a summary of the results.

Table 3
LSE vs Modified LSE

n	$\hat{E}(\bar{\theta}_{n,T})$	$\widehat{MSE}(\bar{\theta}_{n,T})$	$\hat{E}(\bar{\theta}_{n,T})$	$\widehat{MSE}(\bar{\theta}_{n,T})$
10	2.7301761	.60175E+00	1.9644609	.26501E-01
50	2.3115223	.97047E-01	2.0000112	.55669E-08
100	2.2140255	.45807E-01	1.9999982	.13150E-11
250	2.1319951	.17423E-01	1.9999982	.23557E-11
500	2.0921674	.84950E-02	1.9999976	.40709E-11

Note that for $\bar{\theta}_{n,T}$ to attain the same order of accuracy (MSE) as $\bar{\theta}_{n,T}$ based on 10 observations, one needs 250 observations. This means that $\bar{\theta}_{n,T}$ is at least 25 times more efficient than $\bar{\theta}_{n,T}$. However, $\bar{\theta}_{n,T}$ improves steadily as $n \rightarrow \infty$. Moreover, it seems as though $\widehat{MSE}(\bar{\theta}_{n,T})$ does not depend on n . This was confirmed by an experiment whereby T was fixed but large and n was varied as in Table 3. We found that the $\widehat{MSE}(\bar{\theta}_{n,T})$ essentially remained the same for all n .

Now we repeat the experiment that produced Table 3 but with $\theta^0 = 2.0$, so that the solution of (4.14) has a limiting stationary distribution. We obtain

Table 4
LSE vs Modified LSE when $\theta^0 = -2.0$

n	$\hat{E}(\tilde{\theta}_{n,T})$	$\widehat{MSE}(\tilde{\theta}_{n,T})$	$\hat{E}(\tilde{\theta}_{n,T})$	$\widehat{MSE}(\tilde{\theta}_{n,T})$
10	-2.0987299	.77663E+00	-4.3384028	.17480E+02
50	-1.9982407	.24597E+00	-2.3805977	.64996E+00
100	-1.9784782	.25562E+00	-2.2245271	.45129E+00
250	-2.1165625	.22225E+00	-2.2811474	.36043E+00
500	-2.1377542	.15993E+00	-2.2508217	.23628E+00

Table 4 indicate that in this case, $\tilde{\theta}_{n,T}$ performs slightly better than the modified LSE $\tilde{\theta}_{n,T}$. At $n = 250$ there is a 38% gain in MSE when $\tilde{\theta}_{n,T}$ is used instead of $\tilde{\theta}_{n,T}$; while at $n = 500$ the gain is only 32%. Here, gain in MSE is defined by

$$G(\tilde{\theta}_{n,T}, \tilde{\theta}_{n,T}) = \frac{MSE(\tilde{\theta}_{n,T}) - MSE(\tilde{\theta}_{n,T})}{MSE(\tilde{\theta}_{n,T})} \times 100 \quad (4.18)$$

In all the above experiments, we made an arbitrary choice of $T = n^{1/2}$. This choice may not be the best. Since the performance of $\tilde{\theta}_{n,T}$ depends very much on n and T , we would like to find an optimal spacing $\Delta_n = T/n$ in the case $\theta^0 = 2.0$. The series of experiments that are described below intend to arrive at a reasonable guess of the optimal rate at which $T \rightarrow \infty$.

We do a series of five experiments. In each experiment, we fix n ($n = 10, 50, 100, 250$, and 500) and vary the length of the time intervals Δ between observations. We then determine a Δ_* that minimises the $\widehat{MSE}(\tilde{\theta}_{n,T})$. In every experiment, we found that $\widehat{MSE}(\tilde{\theta}_{n,T})$ as a function of Δ , has a global minimum point at some Δ_* . This is illustrated by Figure 4, in the case $n = 10$. Now if we choose Δ_* as the length of the subintervals between observations, then $T = n\Delta_*$. We summarize the results in the following table and graph

Table 5
Optimal Spacing for Different Values of n

n	10	50	100	250	500
Δ	0.25	0.055	0.03	0.015	0.009
T	2.5	2.75	3.0	3.75	4.5

Figure 3
Optimal T as a Function of n

The above graph suggests that the choice $T = .73 \log(n)$ is optimal. We now repeat the first experiment of this section with this choice of T instead of $T = n^{1/2}$. We obtain

Table 6
 $\widehat{MSE}(\bar{\theta}_{n,T})$ vs n when $\theta^0 = 2.0$

n	$\hat{E}(\bar{\theta}_{n,T})$	$\widehat{MSE}(\bar{\theta}_{n,T})$
10	1.8890889	.14436E + 01
50	2.1107919	.12874E - 01
100	2.0638104	.44213E - 02
250	2.0291297	.11065E - 02
500	2.0187584	.35640E - 03

Comparing Table 3 and Table 6, we find that there is a remarkable improvement in the performance of $\bar{\theta}_{n,T}$ in Table 6.

Figure 4
 $\widehat{MSE}(\bar{\theta}_{n,T})$ vs Δ for $n = 10$

5: Directions for further Research

In this section we outline three interesting problems which are a by-product of our study.

For a system of N interacting particles described in Chapter 2, suppose the data consists of fewer than N particles. How sensitive is the ML estimator of θ^0 with respect to such deficiency of data. Further, if the MLE is non-robust, is it possible to develop alternative robust estimation procedures. In order to obtain an insight to this problem we again experimented with the mean field model as in section 4.3. In the first part of the experiment, we calculated the MLE $\hat{\theta}_{N,T} = (\hat{\alpha}_{N,T}, \hat{\beta}_{N,T})$ of $\theta^0 = (\alpha^0, \beta^0)$ using the complete data $D_1 = \{X_j(t), t \in [0, T], j = 1, \dots, N\}$. We chose $\theta^0 = (5, 2)$ and $N = 10$ and obtained the following results

Table 7a

$\hat{E}(\hat{\alpha}_{N,T})$	$\hat{E}(\hat{\beta}_{N,T})$	$\widehat{MSE}(\hat{\alpha}_{N,T})$	$\widehat{MSE}(\hat{\beta}_{N,T})$
5.018357	2.002762	.338E-3	271E-3

where the notation is the same as in section 4.3. In the second part of the experiment we simulated D_1 from model (2.34) but used the incomplete data $D_2 = \{X_j(t), t \in [0, T], j = 1, \dots, N-1\}$ to calculate $\hat{\theta}_{N,T}$. The result was

Table 7b

$\hat{E}(\hat{\alpha}_{N,T})$	$\hat{E}(\hat{\beta}_{N,T})$	$\widehat{MSE}(\hat{\alpha}_{N,T})$	$\widehat{MSE}(\hat{\beta}_{N,T})$
5.018357	1.136049	.338E-3	.944

Note the remarkable change in the $\widehat{MSE}(\hat{\beta}_{N,T})$ and $bias(\hat{\beta}_{N,T})$ from Table 7a to Table 7b. The stability of $\hat{\alpha}_{N,T}$ is probably due to the fact that $\hat{\alpha}_{N,T}$ depends on D_1 (or D_2) through $\bar{X}_N(t) = N^{-1} \sum_{j=1}^N X_j(t); t \in [0, T]$. These results suggests that in general

the MLE is non-robust with respect to incomplete observability. This problem calls for further theoretical studies and possibly development of some robust procedures

In the study of LS estimators we have for the most part assumed equal spacing of data. A more realistic type of data is one where observations occur at random times. For example observations may occur at the jump times of a poisson process. Estimators based on this type of data have been studied by Stoyanov [40] for the case of brownian motion with a drift. It would be nice to extend Stoyanov's results to more general models such as those considered in this study.

The last problem we wish to comment about concerns the modified least squares estimator $\hat{\theta}_{n,T}$ of section 3.4. The hypothesis of Theorem 4 which yield the proof of consistency of $\hat{\theta}_{n,T}$ has been verified in the case of non-stationary Ornstein-Uhlenbeck process. There have been suggestions that this was possible due to some specific features of the non-stationary Ornstein-Uhlenbeck process. It would therefore be worthwhile to look for more examples in order to demonstrate the generality of the modified least squares method.

REFERENCES

- [1] O.O. Aalen (1977), Weak Convergence of Stochastic Integrals Related to Counting Processes, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 38, 261-272.
- [2] R.A. Adams (1975), Sobolev Spaces, *Academic Press*.
- [3] A. Arnold (1974), Stochastic Differential Equations, Theory and Applications, *John Wiley and Sons, New York*.
- [4] I.V. Basawa and B.L.S. Prakasa Rao (1980), Statistical Inference for Stochastic Processes, *Academic Press, London*.
- [5] P. Billingsley (1968), Convergence of Probability Measures, *John Wiley & Sons Inc*.
- [6] P. Billingsley (1979), Probability and Measure, *John Wiley & Sons*.
- [7] V. Borkar and A. Bagchi (1982), Parameter Estimation in Continuous-Time Stochastic Processes, *Stochastics*, vol. 8, pp. 193-212.
- [8] A. Le Breton (1975), On Continuous and Discrete Sampling for Diffusion Type Processes, *Math. Prog. Studies* 5, 124-144.
- [9] B.M. Brown and J.I. Hewitt (1975a) Inference for The Diffusion Branching Processes, *J Appl. Prob* 12, 588-594.
- [10] B.M. Brown and J.I. Hewitt (1975b) Asymptotic Likelihood Theory for Diffusion Processes, *J Appl. Prob.*, 12, 228-238.
- [11] K.L. Chung and R.J. Williams (1983), Introduction to Stochastic Integration, *Birkhauser*.
- [12] D. Dawson (1983), Critical Dynamics and Fluctuations for a Mean-Field Model of Cooperative Behavior, *Journal of Statistical Physics*, Vol. 31, No. 1.
- [13] A.Ja. Dorogovcev (1976), The Consistency of an Estimate of a Parameter of a Stochastic Differential Equation, *Theory of Prob. and Math. Statist.*, no 10, 73-82.
- [14] P.D. Feigin (1976), Maximum Likelihood Estimation for Continuous-Time Stochastic Processes, *Adv. Appl. Prob.* 8, 712-736.
- [15] R. Frydman (1980), A Proof of The Consistency of Maximum Likelihood Estimators of Non-Linear Regression Models With Autocorrelated Errors, *Econometrica*, vol. 48, No. 4.

- [16] I.I. Gihman and A.V. Skorohod (1972), *Stochastic Differential Equations*, Springer-Verlag.
- [17] V.P. Godambe (1985), The Foundations of Finite Sample Estimation in Stochastic Processes, *Biometrika*, 72, 2, pp.419-428.
- [18] P. Guttorp and R. Kulperger (1984), Statistical Inference for Some Volterra Population Processes in a Random Environment, *The Canadian Journal of Statistics*, Vol.12, No.4, pp.289-302.
- [19] P. Hall and C.C. Heyde (1980), *Martingale Limit Theory and Applications*, Academic Press.
- [20] I.S. Helland (1982), Central Limit Theorems for Martingales with Discrete or Continuous Time, *Scand. J. Statist.*, 9, 79-94.
- [21] R.I. Jennrich (1969), Asymptotic Properties of Non-Linear Least Squares Estimators, *The Annals of Math. Statist.*, Vol.40, No.2, 633-643.
- [22] R.H. Jones (1986), Time Series Regression With Unequally Spaced Data, *J. Appl. Prob.*, Vol.23A
- [23] G. Kallianpur (1980), *Stochastic Filtering Theory*, Springer-Verlag
- [24] H. Kunita and S. Watanabe (1967), On Square Integrable Martingales, *Nagoya Mathematical Journal*, No.30, 209-245
- [25] V. Lánka (1979), Minimum Contrast Estimation in Diffusion Processes, *J. Appl. Prob.*, 16, 65-75.
- [26] C. Léonard (1984), Une Loi des Grands Nombres Pour des Systèmes de Diffusions Avec Interaction et à Coefficients Non Bornés, *Serie de Monographies du Laboratoire de Recherche en Statistique et Probabilités, Carleton University/Université d'Ottawa*, No.48.
- [27] R.S Liptser and A.N. Shirayev (1977), *Statistics of Random Processes I, General Theory*, Springer-Verlag, Berlin.
- [28] R.S Liptser and A.N. Shirayev (1977), *Statistics of Random Processes II, Applications*, Springer-Verlag, Berlin.
- [29] E. Malinvaud (1970), The Consistency of Non-Linear Regressions, *Annals of Math. Statist.*, 41, 953-969.
- [30] I.W. McKeague (1984), Estimation for Diffusion Processes Under Misspecified Models, *J. Appl. Prob.* 21, 511-520.

- [31] H.P. McKean Jr. (1969), *Stochastic Integrals*, Academic Press.
- [32] D.R. McNeil and S. Schach (1973), Central Limit Analogues for Markov Population Processes, *J. Royal Stat. Society, Sec. B* 35, 1-23
- [33] E. Pardoux and D. Talay (1985), Discretization and Simulations of Stochastic Differential Equations, *Acta Applicandae Mathematicae*, 3, 23-47
- [34] B.L.S. Prakasa Rao and H. Rubin (1981), Asymptotic Theory of Estimation in Non-Linear Stochastic Differential Equations, *Senkhyā*, Vol 43, Series A, Pt 2
- [35] B.L.S. Prakasa Rao (1983), Asymptotic Theory for Non-Linear Least Squares Estimator for Diffusion Processes, *Math. Operationsforsch. u. Statist., Ser. Statist.*, Berlin 14, 2
- [36] R. Rebolledo (1980), Central Limit Theorems for Local Martingales, *Z. W.* 51, 269-286
- [37] P.M. Robinson (1977), Estimation of a Time Series Model from Unequally Spaced Data, *Stoch. Proc. Appl.*, 6, 9-24
- [38] P. Sjögren (1981), Convergence of Riemann Sums for Stochastic Integrals, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 56, 181-193
- [39] M. Sørensen (1983), On Maximum Likelihood Estimation in Randomly stopped Diffusion-Type Processes, *International Statistical Review*, 51, pp. 93-110
- [40] J.M. Stoyanov (1982), Problems of Estimation in Continuous-Discrete Stochastic Models, *Proceedings of the Seventh Conference on Probability Theory August 29-September 4, Braşov, Romania*
- [41] A.F. Taraskin (1974), On The Asymptotic Normality of Vector-Valued Stochastic Integrals and Estimates of Drift Parameters of a Multidimensional Diffusion Process, *Theory of Prob. and Math. Statist.* No 2
- [42] E. Wong and M. Zakai (1965), On the Convergence of Ordinary Integrals to Stochastic Integrals, *Ann. Math. Statist.*, 36, 1580-1584

END

0/2.01.88

FIN