ASYMPTOTIC STUDY OF ESTIMATORS IN SOME DISCRETE AND CONTINUOUS TIME MODELS

Arup Bose

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements
for the award of the degree of
Doctor of Philosophy

CALCUTTA

1986

ACKNOWLEDGEMENT

This work was done under the supervision of Professor G.J. Babu. Mere words cannot express my sense of gratitude for the encouragement, help and guidance he has given me. I had the liberty to draw upon his invaluable time at my will. His unfailing promptness at every stage of the work induced me to produce better than my best.

During the period 1981-82, while Professor G.J. Babu was on leave, I was very fortunate to have Professor J.K. Ghosh as my guide. The materials of Chapters 4 and 5 were developed under his guidance. I am extremely grateful to him.

My teacher and friend Dr. R.L. Karandikar has played an important role in my academic development. I am indebted to him beyond limits.

Professor B.V. Rao always lent ear to all my queries. It is a pleasure to record my thanks to him.

Teachers, friends and colleagues at the Stat-Math. Division have been instrumental in maintaining an excellent academic atmosphere, and this made for a great experience.

Throughout all these years, my friend Mr. Rabindranath Mukhopadhyay shared all my extra-academic problems and endeavoured to reduce my burdens. This is an opportunity to acknowledge it.

Mr. Dilip Kumar Bardhan did an excellent job of typing with a personal touch. Mr. Mukta Lal Khanra was as efficient as ever with the cyclostyling. They deserve my sincere appreciation.

Calcutta, May 1986 Arup Bose

CONTENTS

*****

INTRODUCTION AND SUMMARY

This thesis deals with an asymptotic study of estimators in some discrete time and continuous time models.

The first part deals with time series data modelled by moving average and autoregressive processes. Higher order asymptotics beyond asymptotic normality, of the usual estimators have been dealt by Phillips (1977,1978), Durbin (1980), Ochi (1983), Tanaka (1983), Fujikoshi and Ochi (1984). $o(n^{-1/2})$ or $O(n^{-1})$ expansions are obtained by these authors under the assumption of normality of errors. The reason for this was the lack of suitable expansions for normalized sums of dependent random vectors. Recently Gotze and Hipp (1983) have been able to solve this problem. They have obtained Edgeworth expansion for normalized sums of weakly dependent random vectors under fairly general conditions.

We start by showing that these recent results of Gotze and Hipp (1983) enable us to derive general asymptotic expansions for the distribution of autocovariances from certain linear processes of the form $Y_t = \sum_{r=0}^{\infty} \delta_r \varepsilon_{t-r}$ where (ε_i) are i.i.d. with sufficiently high order moments. The main conditions imposed are exponential decay of the sequence (δ_r) and Cramer's condition on $(\varepsilon_1, \varepsilon_1^2)$. These results are applied to moving average and autoregressive processes (of any order) under stability conditions. Edgeworth expansions of any order (depending upon the existence of moments of ε_1) for the distribution of the usual estimators follow from the main result. The Berry-Esseen bounds in all these situations hold as a corollary. These are the contents of Chapter 1.

Having proved that the normal approximation is of the order $O(n^{-1/2})$, it is natural to enquire whether there is a better approximation. This idea stems from the recent concept of bootstrap.

The bootstrap is basically a non-parametric procedure and was introduced by Efron (1979, 1982). Since then, it has also been used in parametric situations. Its performance on simulated data, both in parametric and non-parametric situations is quite encouraging. See e.g. Efron (1979,1982,1985), Bickel and Freedman (1983), Freedman and Peters (1984a,b,c) etc.

Simultaneously, various authors have tried to provide theoretical justification as to why this method performs well. The main works in this direction are by Bickel and Freedman (1980, 1981), Singh (1981), Beran (1982) and Babu and Singh (1984). These results deal with accuracy of the bootstrap approximation in various senses (e.g. asymptotic normality, Edgeworth expansions etc.) mainly for sample mean type statistics (or their functionals), quantiles etc. in the i.i.d. situation, where the basic asymptotic distribution theory is normal. For nice functionals, the bootstrap approximation out-performs the normal approximation.

It was anticipated by various authors that the bootstrap would work (in the sense of yielding the same asymptotic distribution as for the original statistic) even in dependent situations provided the resampling takes care of the dependence properly. Freedman (1984) confirms this by showing that it does work for two stage least squares estimates in linear autoregressions with possible exogeneous variables orthogonal to errors.

In Chapter 2 we show that the bootstrap not only works, but, as in the i.i.d. case, works well in autoregressions which satisfy the usual stability condition. In fact, the bootstrapped distribution of the least squares estimates approximates the original distribution with an error $o(n^{-1/2})$. The idea is to develop a bootstrap Edgeworth expansion parallel to the original Edgeworth expansion. (For this no extra conditions are needed). The leading term of these two expansions agree and the difference is $o(n^{-1/2})$. Thus the bootstrap approximation beats the normal approximation almost surely.

In Chapter 3, we prove similar results for the moment estimators in moving average models. The usual invertibility condition is assumed along with moment conditions and Cramer's condition for the errors.

The second part deals with asymptotics of estimation for continuous time data modelled by stochastic differential equations. Here the observable quantity is X_t. $(0 \leq t \leq T)$ which satisfies the stochastic differential equation

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

where $W(t)$ is a standard Brownian motion. The problem is to estimate θ and study its asymptotic properties as $T \to \infty$.

The particular (linear) case, $f(\theta,x) = \sum_{i=1}^{k} a_i(\theta)b_i(x)$, has been dealt by several authors, e.g. Taraskin (1971, 1974), LeBreton (1977), Brown and Hewitt (1975), Kulinic (1975), Lee and Kozin (1977) etc. These authors show that the m.l.e. is weakly/strongly consistent, asymptotically normal and efficient under the main assumption of existence of a stationary ergodic distribution.

The m.l.e. in the non-linear case has been considered by Kutoyants (1977), Prakasa Rao and Rubin (1981) and others. Prakasa Rao and Rubin (1981) proved the strong consistency and asymptotic normality of the m.l.e. when the parameter space is $[-1,1]$, under the assumption of stationarity and ergodicity.

In Chapter 4, we treat the non-linear case with the parameter space as the unit ball of $\mathbb{R}^d$. Strong consistency and asymptotic normality of the m.l.e. are proved. We do not assume stationarity and ergodicity. However, the conditions we impose are easier to verify under the above assumptions. The main technique is to use Kolmogorov type inequalities from the theory of diffusion processes. These are used to get probabilistic bounds for supremum of certain processes, and are of independent interest.

The spirit of Chapter 5 is in a sense Bayesian. Suppose there is a prior probability on the parameter space. A classical theorem for posteriors loosely stated says that the posterior density given the i.i.d. observations (X_i), $i \leq n$, converges in L^1 to a normal density under a fixed θ_0. See LeCam (1955, 1958). This result was extended to discrete time Markov processes by Borwanker, Kallianpur and Prakasa Rao (1971). In the context of diffusion processes, Prakasa Rao (1981) proved the result in a special linear case and extended it to diffusion fields in Prakasa Rao (1983). With the help of techniques developed in

Chapter 4 and a formula for ordinary differentiation under stochastic integrals due to Karandikar (1983), we obtain this theorem for non-linear diffusions. As consequences, the Bayes estimators and m.l.e. are $T^{1/2}$ equivalent a.s. and hence the asymptotic behaviour of the Bayes estimators are same as those of the m.l.e. shown in Chapter 4.

In Chapter 6, we have made a small attempt to study the rate of convergence of the m.l.e. to normality. Mishra and Prakasa Rao (1985) studied the case $f(\theta,x) = -\theta b(x)$, $\theta > 0$. Their results yield the rate $O(T^{-1/5})$ when applied to the case $b(x) = x$. But note that the model $dX(t) = -\theta X(t)dt + dW(t)$ is the continuous time analogue of the first order autoregressive process with i.i.d. $N(0,1)$ errors. Given the results of Chapter 1, better rates are naturally expected for this model. We obtain the Berry-Esseen bound of the order $O(T^{-1/2})$ for the normalized m.l.e. The proof involves a suitable change of measure and an application of Ito's formula. The general non-linear case seems to be hard and perhaps new techniques are needed to deal with this situation.

PART I

CHAPTER 1

EDGEWORTH EXPANSIONS FOR AUTOCOVARIANCES FROM LINEAR PROCESES WITH APPLICATIONS

1.1 Introduction

A wide class of time series models specify the observations (Y_t) in the form of a linear process $Y_t = \sum_{r=0}^{\infty} \delta_r \varepsilon_{t-r}$, where (ε_i) is an i.i.d. sequence. The usual ARMA (p,q) models belong to this class. The autocovariances $n^{-1} \sum_{t=1}^{n} Y_t Y_{t-i}$, $i = 0,1,\ldots,p$ (or their slightly perturbed forms) play an important role in time series analysis.

Under mild conditions, these autocovariances have an asymptotic normal distribution. Blume and Wittwer (1981) derived convergence rates for individual autocovariances. Kersten (1984) generalized this to vector of autocovariances under decay conditions on (δ_i) and moment conditions on (ε_i). The ARMA (p,q) models satisfy these conditions. The rates obtained are $O(n^{-1/2+\varepsilon})$ for some $\varepsilon > 0$.

We assume stronger conditions (exponential decay of the constants (δ_r) and Cramer's condition on $(\varepsilon_1,\varepsilon_1^2)$) to obtain higher order approximations (Edgeworth expansions) for the distribution of the autocovariances. This is done by applying recent results of Gotze and Hipp (1983). The ARMA (p,q) models satisfy our conditions.

The Berry-Esseen bound holds as a corollary. However, the Berry-Esseen bound is expected to hold under much weaker conditions. For example the Cramer's condition seems superfluous. But this was not our primary aim.

So this possibility was not explored. We were interested in exploiting the expansions.

As applications of these expansions, we obtain Edgeworth expansions for the distribution of the least squares estimators in autoregressive processes and the moment estimators in moving average processes. These results remove the rather strict condition of normality of errors assumed by several authors to derive Edgeworth expansions of the maximum likelihood estimate for the autoregressive parameters. See Remark 1.3.2(5).

The validity of these expansions also give rise to the possibility of bootstrapping the distribution of certain estimators in autoregressive and moving average processes with high accuracy. These are the subject matter of the Chapters 2 and 3.

This chapter is a revised form of Bose (1985c).

1.2 The main results :

Let (Y_t) be a linear process such that $Y_t = \sum_{r=0}^{\infty} \delta_r \varepsilon_{t-r}$ where we assume the following conditions on (δ_r) and (ε_i) :

(C1) (ε_i) is an i.i.d. sequence such that

$$E\varepsilon_t = 0,\ E\varepsilon_t^2 = 1,\ E\varepsilon_t^{2(s+1)} < \infty \quad \text{for some } s \geq 3.$$

(C2) $\forall$ large m, $\sum_{r=m}^{\infty} |\delta_r| \leq c \exp(-\alpha m),\ \alpha > 0.$

(C3) $(\varepsilon_1, \varepsilon_1^2)$ satisfies the Cramer's condition, i.e.,

$$\exists\, \delta > 0,\ d > 0 \ni \forall\ \|t\| \geq d,\ |E\exp(it'(\varepsilon_1, \varepsilon_1^2))| \leq 1 - \delta .$$

Define the variables $X_{it} = Y_t Y_{t-i} - \sigma_i$, $i = 0,1,\ldots,p$ and let

$$X_t' = (X_{ot},\ldots,X_{pt}) \qquad \ldots(1.2.1)$$

Here σ_i is the theoretical autocovariance of order i, i.e.

$$\sigma_i = EY_t Y_{t-i} = \sigma_{-i} .$$

Assume further that

(C4) $\Sigma = \lim_{n \to \infty} D(n^{-1/2} \sum_{t=1}^{n} X_t)$ is positive definite.

Note that under our conditions (C1) and (C2), $\Sigma = ((\sigma_{ij}))$ exists and is given by (Anderson (1971), pp.478)

$$\sigma_{ij} = \sum_{r=-\infty}^{\infty} (\sigma_{r+i}\sigma_{r+j} + \sigma_{r-i}\sigma_{r+j}) + \sigma_i\sigma_j(E\varepsilon_t^4 - 1) \qquad \ldots(1.2.2)$$

The earliest works on asymptotic expansion for dependent variables are those of Statulevicius (1969, 1970), Gotze and Hipp (1978) and Durbin (1980). For our purposes, the recent work of Gotze and Hipp (1983) is most covenient. We take the help of the following results of Gotze and Hipp (1983).

Let (X_t) be $\mathbb{R}^k$ valued random variables on $(\Omega, \mathfrak{F}, P)$. Introduce the following conditions.

Let there be σ-fields $\mathcal{D}_j$ (write $\sigma(\prod_{j=a}^{b} \mathcal{D}_j) = \mathcal{D}_a^b$) and $\alpha > 0$ such that

C(1.2.1) $EX_t = 0 \;\forall t$.

C(1.2.2) $E \| X_t \|^{s+1} \leq \beta_{s+1} < \infty \;\; \forall\, t$ for some $s \geq 3$.

C(1.2.3) $\quad \exists\ Y_{nm} \in \mathcal{D}_{n-m}^{n+m} \ni E\| X_n - Y_{nm} \| \leq c.\exp(-\alpha m)$.

C(1.2.4) $\quad \forall\ A \in \mathcal{D}_{-\infty}^{n},\ B \in \mathcal{D}_{n+m}^{\infty},$

$$|P(A \cap B) - P(A)P(B)| \leq c.\exp(-\alpha m).$$

C(1.2.5) $\quad \exists\ d, \delta > 0 \ni \forall\ \| t \| \geq d,$

$$E|E \exp(it' \sum_{j=n-m}^{n+m} X_j) / \mathcal{D}_j,\ j \neq n| < 1 - \delta < 1.$$

C(1.2.6) $\quad \forall\ A \in \mathcal{D}_{n-p}^{n+p},\ \forall\ n,p,m,$

$$E|P(A/\mathcal{D}_j,\ j \neq n) - P(A/\mathcal{D}_j,\ 0 < |j-n| \leq m+p)| \leq c.\exp(-\alpha m).$$

C(1.2.7) $\quad \lim_{n \to \infty} D(n^{-1/2} \sum_{t=1}^{n} X_t)$ exists and is positive definite.

Define the integer $s_o \leq s$ by

$$s_o = \begin{cases} s & \text{if } s \text{ is even} \\ s-1 & \text{if } s \text{ is odd.} \end{cases}$$

Let $\psi_{n,s}$ be the usual function associated with Edgeworth expansions. Let φ_Σ be the normal density with mean 0 and dispersion matrix Σ, where Σ is as defined below.

Define $S_n = n^{-1/2} \sum_{t=1}^{n} X_t$, $\Sigma = \lim_{n \to \infty} D(n^{-1/2} \sum_{t=1}^{n} X_t)$.

The following results are due to Gotze and Hipp (1983).

<u>Theorem 1.2.1.</u> Let $f : \mathbb{R}^k \longrightarrow \mathbb{R}$ denote a measurable function such that $|f(x)| \leq M(1 + \| x \|^{s_o})$ for every $x \in \mathbb{R}^k$. Assume that C(1.2.1) - C(1.2.7) hold. Then there exists a positive constant δ not depending

on f and M, and for arbitrary $k > 0$ there exists a positive constant C depending on M but not on f such that

$$|Ef(S_n) - \int f d\psi_{n,s}| \leq C\omega(f,n^{-k}) + o(n^{-(s-2+\delta)/2})$$

where $\omega(f,n^{-k}) = \int \sup(|f(x+y) - f(x)| : \|y\| \leq n^{-k})\varphi_\Sigma(x)dx$.

The term $o(.)$ depends on f through M only.

Corollary 1.2.2 : Under assumptions C(1.2.1) - C(1.2.7) we have uniformly for convex measurable $C \subseteq \mathbb{R}^k$,

$$P(S_n \in C) = \psi_{n,s}(C) + o(n^{-(s-2)/2}).$$

For non-negative integral k-vectors $\alpha = (\alpha_1,\dots,\alpha_k)$ define

$$D^\alpha = \frac{\partial^{\alpha_1}}{\partial^{\alpha_1}_{x_1}} \cdots \frac{\partial^{\alpha_k}}{\partial^{\alpha_k}_{x_k}}.$$

Theorem 1.2.3 : Let $f : \mathbb{R}^k \to \mathbb{R}$ denote an infinitely differentiable function such that $|f(x)| \leq M(1 + \|x\|^{s_0})$ for every $x \in \mathbb{R}^k$ and $|D^\alpha f(x)| \leq M_\alpha(1 + \|x\|^{p_\alpha})$ for every non-negative integral k-vector with positive constants M_α, P_α. Assume C(1.2.1) - C(1.2.4) and C(1.2.7) hold. Then

$$Ef(S_n) - \int f d\psi_{n,s} = o(n^{-(s-2)/2}).$$

Theorem 1.2.4 : Under conditions C(1.2.1) - C(1.2.4) and C(1.2.7),

$$E(1+\|S_n\|^{s_0})I(\|S_n\| > ((s-2)\log n)^{1/2}) = o(^{-(s-2)/2}).$$

We shall also need the following lemma to prove our main theorem.

<u>Lemma 1.2.5</u>. Let $A' = (A_i\ ,\ i = 0,1,\ldots,p)$, $B' = (\sigma_i,\ i = 0,1,\ldots,p)$

where $A_i = \sum_{\substack{s=-\infty \\ s\neq 0}}^{\infty} (\sigma_{i+s} + \sigma_{i-s})\varepsilon_s\ ,\ i = 0,1,\ldots,p.$

Then under (C4), there exist $\eta, \varepsilon > 0$ such that for all t with $\| t \| = 1$, $P(\| t'A, t'B \| \geq \eta) > \varepsilon > 0$.

<u>Proof</u>. First note that the dispersion matrix of A is given by

$$D(A) = 2\Sigma + C_o N \quad \text{where} \qquad \ldots(1.2.3)$$

$$C_o = 2(1 - E\ \varepsilon_t^4) < 0$$

$N = ((\sigma_i \sigma_j))$ and Σ is as defined in (1.2.2).

From the compactness of the unit ball, it suffices to show that there is such a choice of η and ε for every fixed t. Choose such a t and write $t = t_1 + t_2$ where $t_1 \perp t_2$, $t_1 \varepsilon$ sp $((\sigma_i,\ i = 0,1,\ldots,p))$. Fix $\alpha > 0$ (to be chosen).

<u>Case 1</u>. $\| t_1 \| \geq \alpha$. In this case,

$$\| t'A, t'B \| \geq |t_1' B| \geq \alpha\left(\sum_{i=0}^{p} \sigma_i^2\right)^{1/2} > 0.$$

<u>Case 2</u>. $\| t_2 \| \geq 1 - \alpha$. In this case, $\| t'A,\ t'B \| \geq |t'A|$ and

$$V(t'A) = t'D(A)t = t'(2\Sigma + C_o N)t.$$

Note that $t'Nt = t_1' N t_1 \leq (\sum_{i=0}^{p} \sigma_i^2)\alpha$.

Let λ_1 be the smallest eigen value of $\Sigma(\lambda_1 > 0)$. Choose α such that $|C_0|(\sum_{i=0}^{p} \sigma_i^2)\alpha < \lambda_1$. Thus $V(t'A) > \lambda_1 > 0$. This proves the lemma.

We now state our main theorem.

<u>Theorem 1.2.6</u> : Under conditions (C1) - (C4), Theorems (1.2.1), (1.2.3), (1.2.4) and Corollary (1.2.2) hold for X_t defined by (1.2.1).

<u>Remark 1.2.7</u> : Obviously, corresponding to the theorems of Gotze and Hipp, parts of the above theorem hold even when appropriate conditions from (C1) - (C4) are dropped.

<u>Proof of Theorem 1.2.6</u> : We have to verify that conditions C(1.2.1) - C(1.2.7) hold for the process (X_t). Take $\mathcal{D}_j = \sigma(\varepsilon_j)$ = the σ-field generated by ε_j. Conditions C(1.2.1), C(1.2.4) and C(1.2.6) hold trivially. C(1.2.2) follows from (C1). C(1.2.7) is assumption (C3). It remains to check C(1.2.3) and C(1.2.5).

Let $Y_{nm} = (Y_{onm}, \ldots, Y_{pnm})$ where

$$Y_{inm} = (\sum_{r=0}^{m} \delta_r \varepsilon_{n-r})(\sum_{r=0}^{m+i} \delta_r \varepsilon_{n-i-r}) - \sigma_i,\ i = 0,1,\ldots,m.$$

Hence
$$X_{in} - Y_{inm} = (\sum_{r=0}^{\infty} \delta_r \varepsilon_{n-r})(\sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r} - \sum_{r=0}^{m+i} \delta_r \varepsilon_{n-i-r})$$
$$+ (\sum_{r=0}^{m+i} \delta_r \varepsilon_{n-i-r})(\sum_{r=0}^{\infty} \delta_r \varepsilon_{n-r} - \sum_{r=0}^{m} \delta_r \varepsilon_{n-r})$$

$$E|Y_{inm} - X_{in}| \leq \left[E(\sum_{r=0}^{\infty} \delta_r \varepsilon_{n-r})^2\right]^{1/2} \left[E(\sum_{r=m+i+1}^{\infty} \delta_r \varepsilon_{n-i-r})^2\right]^{1/2}$$
$$+ \left[E(\sum_{r=0}^{m+i} \delta_r \varepsilon_{n-i-r})^2\right]^{1/2} \left[E(\sum_{r=m+1}^{\infty} \delta_r \varepsilon_{n-r})^2\right]^{1/2}$$
$$\leq C(\sum_{r=m+1}^{\infty} \delta_r^2)^{1/2}.$$

Condition (C2) ensures that C(1.2.3) is satisfied.

It remains to check C(1.2.5). In what follows, β shall denote a random variable such that β and ε_n are independent.

For $0 \leq i \leq p$, $\sum_{j=n-m}^{n+m} X_{ij} = \sum_{j=n-m}^{n+m} Y_j Y_{j-i} + \beta$

$$= \beta + \varepsilon_n A_{inm} + \varepsilon_n^2 B_{inm} \quad \text{where,}$$

$$A_{inm} = \sum_{j=n+i}^{n+m} \left(\delta_{j-n} \sum_{r \neq j-i-n} \delta_r \varepsilon_{j-r-i} + \delta_{j-i-n} \sum_{r \neq j-n} \delta_r \varepsilon_{j-r}\right)$$

$$B_{inm} = \sum_{j=n+i}^{n+m} \delta_{j-n} \delta_{j-i-n} .$$

Note that A_{inm} and ε_n are independent $\forall n$ and $B_{inm} \to \sigma_i$ as $m \to \infty$. Let (ε_i^*) be i.i.d., (ε_j^*) and (ε_j) independent and $\varepsilon_j^* \stackrel{\mathcal{L}}{=} \varepsilon_j$. Define A_{inm}^*, A_{nm}^* as A_{inm}, A_{nm} with ε_i's replaced by ε_i^*'s. $(A_{inm}^*,\ i = 0,1,\ldots,p) \Rightarrow (A_i,\ i = 0,1,\ldots,p)$ where

$$A_i = \sum_{j=0}^{i-1} \delta_j Y_{j-i}^* + \sum_{j=i}^{\infty} (\delta_j Y_{j-i}^* + \delta_{j-i} Y_j^*) - 2\varepsilon_0^* \sigma_i$$

which on simplification yields,

$$A_i = \sum_{\substack{s=-\infty \\ s \neq 0}}^{\infty} (\sigma_{i+s} + \sigma_{i-s}) \varepsilon_s^* , \quad i = 0,1,\ldots,p .$$

Thus $E\left|E \exp(it' \sum_{j=n-m}^{n+m} X_j) / \mathcal{D}_j ,\ j \neq n\right|$

$$= E\left|E \exp(i\varepsilon_n t' A_{nm}^* + i\varepsilon_n^2 t' B_{nm}) / \mathcal{D}_j^* ,\ j \neq n\right|$$

$$\leq (1-\delta) P(\| t' A_{nm}^*,\ t' B_{nm} \| \geq d) + P(\| t' A_{nm}^*,\ t' B_{nm} \| \leq d) .$$

Hence C(1.2.5) will follow if $\exists$ $d, d_1, \varepsilon > 0$ such that $\| t \| \geq d_1$ and all large n,m implies,

$P(\| t'A^*_{nm}, t'B_{nm} \| \geq d) > \varepsilon > 0$. This follows if there exists $\eta, \varepsilon > 0$ such that for all $\| t \| = 1$,

$$P(\| t'A_{nm}, t'B_{nm} \| \geq \eta) > \varepsilon \text{ for all large } n,m. \qquad \ldots(1.2.4)$$

Suppose (1.2.4) is not true. Then there exists sequences (η_m), (δ_m) tending to zero and (t_m) such that $\| t_m \| = 1$ and $P(\| t_m'A_{nm}, t_m'B_{nm} \| \geq \eta_m) \leq \delta_m$. Without loss of generality, assume that (t_m) converges (otherwise take a subsequence) to some t_o. Note that A_{nm} and B_{nm} are not dependent on n and $A_{nm} \Rightarrow A$ and $B_{nm} \rightarrow B$. The above inequality shows that $\| t_o'A, t_o'B \| = 0$ which is a contradiction to Lemma 1.2.5. Thus C(1.2.5) is verified and the theorem follows.

Remarks 1.2.7

(1) The autocorrelations are smooth functions of the autocovariances. Edgeworth expansions for autocovariances thus yield expansions for the autocorrelations too. A justification is not hard to give. This is available in Bhattacharya and Ghosh (1978) (their Lemma 2.1). Deducing here would just be repetition, so we chose to omit it.

(2) Note that as a corollary to Theorem 1.2.6, the Berry-Esseen bound holds for $S_n = n^{-1/2} \sum_{t=1}^{n} X_t$. However, the Cramer's condition is conjectured to be superfluous. Kersten (1984) obtained rates $O(n^{-1/2+\varepsilon})$, $\varepsilon > 0$ under conditions $E|\varepsilon_1|^{2k} < \infty$ for some $k \geq 2$ and

$$\max\left(\sum_{i=r-U}^{\infty} \delta_i^4, \sum_{j=r-U}^{\infty} \delta_j^2\right) \leq K_1 r^{-2k} \text{ for } r \geq U+1.$$

It seems plausible that a slightly stronger form of this decay condition shall yield the Berry-Esseen bound. However, this was not our aim and the idea was not pursued.

(3) Extension to vector-valued processes involves no new ideas and involves only more notations and algebra.

(4) The ARMA (p,q) models satisfy our conditions by calculations of Kersten (1984, page 528).

1.3. Applications

(A) Moving average process

Let (Y_t) be a MA(2) process,

$Y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}$ where (ε_i) satisfies (C1) and (C3).

The moment estimators (slightly perturbed) are given by

$$\alpha_{2n} = n^{-1} \sum_{k=1}^{n} Y_k Y_{k-2}$$

$$\alpha_{1n} = n^{-1} \sum_{k=1}^{n} Y_k Y_{k-1} / (1 + n^{-1} \sum_{k=1}^{n} Y_k Y_{k-2}).$$

Note that $\sigma_0 = 1 + \alpha_1^2 + \alpha_2^2$

$$\sigma_1 = \alpha_1 (1 + \alpha_2)$$

$$\sigma_2 = \alpha_2 \text{ and } \sigma_j = 0 \ \forall \ |j| \geq 3.$$

To prove that Σ is positive definite, suffices to show (see (1.2.3)),

$$a_1(\sigma_{s+1} + \sigma_{1-s}) + a_2(\sigma_{2+s} + \sigma_{2-s}) = 0 \quad \forall s \neq 0$$

implies $a_1 = a_2 = 0$.

Taking $s = 1,2,3$ in the above expression,

$$a_1(\alpha_2 + 1 + \alpha_1^2 + \alpha_2^2) + a_2\alpha_1(1 + \alpha_2) = 0$$

$$a_1\alpha_1(1 + \alpha_2) + a_2(1 + \alpha_1^2 + \alpha_2^2) = 0$$

$$a_1\alpha_2 + a_2\alpha_1(1 + \alpha_2) = 0,$$

which clearly implies $a_1 = a_2 = 0$.

Thus we have asymptotic expansions for the random vector

$$n^{-1/2}\Big(\sum_{k=1}^{n}(Y_kY_{k-1} - \sigma_1), \sum_{k=1}^{n}(Y_kY_{k-2} - \sigma_2)\Big),$$ which in turn yields

asymptotic expansion for the joint distribution of $n^{1/2}(\alpha_{1n} - \alpha_1, \alpha_{2n} - \alpha_2)$ by Lemma 2.1 of Bhattacharya and Ghosh (1978).

Remark 1.3.1

The same principle applies to higher order models, but solving for estimates of $\alpha_1, \ldots, \alpha_p$ from the moment equations become cumbersome with increase in p.

(B) Autoregressive process :

Let $Y_t = \theta Y_{t-1} + \varepsilon_t$ where (ε_i) satisfies (C1) and (C3) and $|\theta| < 1$. The least squares estimate of θ is given by

$$\theta_n = \sum_{t=1}^{n} Y_tY_{t-1} \Big/ \sum_{t=1}^{n} Y_{t-1}^2,$$

which is a smooth function of autocovariances.

To prove that Σ is positive definite with $p = 1$, as in the previous example, it suffices to show that

$$a_1(\sigma_s + \sigma_{-s}) + a_2(\sigma_{1+s} + \sigma_{1-s}) = 0 \quad \forall\, s \neq 0 \text{ implies}$$

that $a_1 = a_2 = 0$.

Taking $s = 0$ and 1 in the above expression,

$$2\sigma_o a_1 + 2\sigma_1 a_2 = 0$$

$$2\sigma_1 a_1 + a_2(\sigma_2 + \sigma_o) = 0$$

Since $Y_t = \sum_{k=0}^{\infty} \theta^k \varepsilon_{t-k}$, it easily follows that

$$\sigma_i = \theta^i(1 - \theta^2)^{-1}, \quad i = 0,1,2.$$

Hence the above equations reduce to

$$a_1 + \theta a_2 = 0$$

$$2\theta a_1 + (\theta + \theta^2)a_2 = 0$$

The matrix $\begin{pmatrix} 1 & \theta \\ 2\theta & \theta+\theta^2 \end{pmatrix}$ is non-singular. Thus $a_1 = a_2 = 0$.

Hence we have asymptotic expansion for the random vector

$$n^{-1/2}\left(\sum_{t=1}^{n} (Y_{t-1}^2 - \frac{1}{1-\theta^2}),\ \sum_{t=1}^{n} (Y_t Y_{t-1} - \frac{\theta}{1-\theta^2})\right) \text{ which in turn}$$

yields asymptotic expansion for normalized θ_n, i.e. $(\frac{n}{1-\theta^2})^{1/2}(\theta_n - \theta)$.

Remark 1.3.2

(1) This result is stronger than $o(n^{-1/2})$ expansion in Bose (1985a).

(2) As a consequence of the above results we also have the Berry-Esseen theorem for θ_n, viz.

$$\sup_{x \in \mathbb{R}} \left| P\left(\left(\frac{n}{1-\theta^2}\right)^{1/2}(\theta_n - \theta) \le x\right) - \Phi(x) \right| \le Cn^{-1/2}.$$

(3) Note that Theorem 1.2.6 remains valid for $X_t = Y_{t-1}\varepsilon_t$ if we assume the following weaker conditions:

(C1)' $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$, $E|\varepsilon_t|^{s+1} < \infty$ for some $s \ge 3$.

(C2)' ε_1 satisfies Cramer's condition.

Hence under (C1)' and (C2)' we have Edgeworth expansions for the distribution of $(\sum_{t=1}^{n} Y_{t-1}^2)n^{-1/2}(\theta_n - \theta)$, since

$$\theta_n - \theta = \sum_{t=1}^{n} Y_{t-1}\varepsilon_t \Big/ \left(\sum_{t=1}^{n} Y_{t-1}^2\right).$$

(4) The same principle applies to l.s.e. from higher order autoregressions. We shall elaborate on this in Chapter 2.

(5) Quite a few authors have devoted themselves to obtaining Edgeworth expansions for the m.l.e. in time series models. See e.g. Phillips (1977-78), Durbin (1980), Ochi (1983), Tanaka (1983), Fujikoshi and Ochi (1984). All these works assume the restrictive condition of normality of errors. As our results show this is unnecessary, if we consider l.s.e. or moment estimators in place of m.l.e. in the absence of distribution assumptions.

(6) The polynomials involved in the expansions are not hard to get and involve only cumbersome algebra. Also note that application of Theorem 1.2.6 also yields approximations for the moments of the estimators considered.

(7) As mentioned in the introduction of this chapter, the approach of this chapter shall be used in the next two chapters to study the bootstrap principle in these two dependent models, viz. autoregressive and moving average models.

CHAPTER 2

BOOTSTRAP IN AUTOREGRESSIONS

2.1 Introduction : In this chapter, we shall be engaged in studying the probabilistic aspects of bootstrap approximation for distribution of parameter estimates in autoregressive models. A few words about the bootstrap is pertinent here. We shall not go into a formal description of the bootstrap procedure, since by now it is quite well known. Instead, we shall provide an indication of the literature available (which, by no means is claimed to be exhaustive).

The bootstrap was introduced by Efron (1979, 1982). Since then, there has been a fast growing literature on the topic. These move in two directions, one complimenting the other. Empirical evidence has suggested that the bootstrap performs usually very well. Relevant references for these are Efron (1979, 1982), Bickel and Freedman (1983), Daggett and Freedman (1984), Freedman and Peters (1984a,b,c). Simultaneously, there have been attempts to provide theoretical justification as to why this method performs well. These results provide an insight into the working of the bootstrap procedure. We would like to mention the papers by Bickel and Freedman (1980, 1981), Singh (1981), Beran (1982) and Babu and Singh (1984). These results deal with accuracy of bootstrap approximation in various senses (e.g. asymptotic normality, Edgeworth expansions etc.), mainly for sample mean type statistics (or their functionals), quantiles etc. in the i.i.d. situation, where the basic asymptotic distribution theory is normal. A recent paper by Abramovitch and

Singh (1985) deals with modifications and bootstrap approximations of statistics admitting Edgeworth expansion. Babu (1984) has showed that a modification of the bootstrap " works " (in the sense of yielding the same asymptotic distribution as for the original statistic) when the basic asymptotic distribution is Chi-square. Athreya, Ghosh, Low and Sen (1984) derived law of large numbers and asymptotic normality of bootstrapped U-statistics. Most of the above research has concentrated on univariate situations, whereas the potential of the bootstrap in multivariate situations is clear. We refer to the works of Babu and Singh (1983) and Beran and Srivastava (1985). An example of Babu (1984) shows the importance of existence of moments for the bootstrap to work. Athreya (1984a,b,c) in a series of papers has explored in detail the limiting bootstrap distribution under weak moment conditions. These show in particular that, in general the limiting distribution is not free of the sample sequence of observations. Further study in this area seems to be desirable (e.g. whether proper censoring makes the bootstrap work or whether it works for self normalizing statistics). The bootstrap has been also studied in sample survey problems, by Bickel and Freedman (1984) and Rao and Wu (1985). Recently, focus has been on bootstrap in regression problems. We refer the reader to Freedman (1981), Shorack (1982), Delaney and Chatterjee (1984), Wu (1986) and the references contained therein. One of the reasons for the growing interest in bootstrap procedure is the ever increasing computational facilities. There have appeared in quick succession review articles e.g. by Beran (1984), Efron and Tibshirani (1985) and " popular " articles e.g. by Efron and Gong (1983) and Peters and Freedman (1984) which indicate the popularity of the procedure. Among the many possibilities of application of the bootstrap is its use to obtain better confidence intervals and better approximations of quantiles than given by the normal theory. We refer to Efron (1984) and Babu and Bose (1986). An article by Schenker (1985) shows that the use of the bootstrap confidence interval might be disastrous if applied indiscriminately. To round off this brief indication of the available literature we mention the paper by Rubin (1981), which introduces a Bayesian analogue of the bootstrap ; this notion does not seem to have been pursued.

The bootstrap cannot in general work for dependent processes ; Singh (1981) provides an example. However, it was anticipated that the bootstrap would still work if the dependence is taken care of while resampling. Freedman (1984) confirms this by showing that it does work for certain linear dynamic models (e.g. for two stage least squares estimates in linear autoregressions with possible exogeneous variables orthogonal to errors). To the author's knowledge this is the only theoretical work available for bootstrap in dependent models. However, there has been empirical work for dependent models ; the references are contained in those already cited. While writing this thesis, the author has come across a paper by Chatterjee (1985) which deals with empirical work in ARMA models.

Our aim in this chapter is to study the bootstrap in autoregressions. Resampling is possible which takes care of the dependence. In

Chapter 1, we showed that Edgeworth expansions can be obtained for the distribution of the normalized least squares estimate. The main idea here is to develop an analogous expansion (of order $o(n^{-1/2})$) for the bootstrap distribution. The leading terms of these expansions are same and the difference is $o(n^{-1/2})$ a.s. Thus the bootstrap beats the normal approximation.

In Section 2, we obtain $o(n^{-1/2})$ expansion for the distribution of l.s.e. in first order autoregressions via a route different from that in Chapter 1, since this is more convenient while dealing with the bootstrap approximation - which, is dealt in Section 3. In Section 4, we outline how results of first order autoregressions can be extended to higher order autoregressions. This treatment shows the difficulty of obtaining general Edgeworth expansions for l.s.e. in higher order autoregressions. However, such an expansion is valid for a certain randomly normed version of l.s.e. (see Remark 2.4.2) and, in any case expansion of order $o(n^{-1/2})$ remains valid with usual norming. We also remark about extensions of the results to vector-valued case and to the case where the i.i.d. errors need not have known mean and variance.

This chapter is a revised form of Bose (1985a).

2.2 One term Edgeworth expansion of the l.s.e.

Let (Y_t) be a first order autoregressive process satisfying

$$Y_t = \theta Y_{t-1} + \varepsilon_t,\ t = 0, \pm 1, \pm 2, \ldots \qquad \ldots(2.2.1)$$

$$|\theta| < 1,\ (\varepsilon_t)\ \text{i.i.d.} \sim F_0,\ E\,\varepsilon_t = 0,\ V(\varepsilon_t) = 1\,.$$

We introduce the following assumptions :

(A1) $E\,\varepsilon_t^{2(s+1)} < \infty$ for some $s \geq 3$.

(A2) $(\varepsilon_1, \varepsilon_1^2)$ satisfies Cramer's condition, i.e. for every $d > 0$, there exists $\delta > 0$ such that $\sup_{\|t\| > d} |E\exp(it'(\varepsilon_1, \varepsilon_1^2))| \leq \exp(-\delta)$.

Given the observations $Y_0, Y_1, \ldots, Y_n$, the l.s.e. θ_n of θ satisfies the equation

$$\theta_n - \theta = \Big(\sum_{t=1}^{n} Y_{t-1}\varepsilon_t\Big)\Big(\sum_{t=1}^{n} Y_{t-1}^2\Big)^{-1}. \qquad \ldots(2.2.2)$$

Also note that the following equation holds

$$(1-\theta^2)\sum_{t=1}^{n} Y_{t-1}^2 = Y_0^2 - Y_n^2 + 2\theta \sum_{t=1}^{n} Y_{t-1}\varepsilon_t + \sum_{t=1}^{n} \varepsilon_t^2 \qquad \ldots(2.2.3)$$

Let $X_t' = (X_{1t}, X_{2t})$ where $X_{1t} = Y_{t-1}\varepsilon_t$, $X_{2t} = \varepsilon_t^2 - 1$.

In Chapter 1, we derived asymptotic expansions for the distribution of normalized θ_n. However, to deal with the bootstrap approximation, it is convenient to work with X_t instead of autocovariances.

$$\text{Define } s_0 = \begin{cases} s & \text{if } s \text{ is even} \\ s-1 & \text{if } s \text{ is odd} \end{cases}$$

To obtain an expansion for (X_t), we use Theorem 1.2.1 of Chapter 1 and verify the conditions C(1.2.1) - C(1.2.7) of Chapter 1.

We take $\mathcal{D}_j = \sigma(\varepsilon_j) = \sigma$-field generated by ε_j. Under assumption (A1), C(1.2.1) and C(1.2.2) hold. C(1.2.3) is easily verified. C(1.2.4) and C(1.2.6) hold autometically. Clearly

$$\begin{aligned} \operatorname{Cov}(X_{1t}, X_{1t'}) &= 0 \quad \text{if} \quad t \neq t' \\ &= (1-\theta^2)^{-1} \quad \text{if} \quad t = t' \\ \operatorname{Cov}(X_{2t}, X_{2t'}) &= 0 \quad \text{if} \quad t \neq t' \\ &= V(\varepsilon_1^2) \quad \text{if} \quad t = t' \quad \text{and} \\ \operatorname{Cov}(X_{1t}, X_{2t'}) &= 0. \end{aligned}$$

Thus $\Sigma = \lim_{n \to \infty} D(n^{-1/2} \sum_{t=1}^{n} X_t)$

$$= \begin{pmatrix} (1-\theta^2)^{-1} & 0 \\ 0 & V(\varepsilon_1^2) \end{pmatrix} \quad \text{which verifies C(1.2.7).}$$

We show now that under (A2), C(1.2.5) holds.

$$E \mid E \exp(it' \sum_{j=n-m}^{n+m} X_j)/\varepsilon_j ,\, j \neq n \mid$$

$$= E \mid E \exp(it_1 \sum_{j=n}^{n+m} Y_{j-1}\varepsilon_j + it_2\varepsilon_n^2)/\varepsilon_j , \quad j \neq n \mid .$$

But $$\sum_{j=n}^{n+m} Y_{j-1}\varepsilon_j = \sum_{j=n}^{n+m} \left(\sum_{t=0}^{\infty} \theta^t \varepsilon_{j-1-t}\right)\varepsilon_t$$

$$= \varepsilon_n \sum_{t=0}^{\infty} \theta^t \varepsilon_{n-1-t} + \varepsilon_n \sum_{j=n+1}^{n+m} \theta^{j-1-n}\varepsilon_j$$

+ terms involving ε_j, $j \neq n$.

Let $A_n = \sum_{t=0}^{\infty} \theta^t \varepsilon_{n-1-t}$ and $B_{nm} = \sum_{j=n+1}^{n+m} \theta^{j-1-n}\varepsilon_j$.

Note that A_n and B_{nm} are independent.

Using independence of (ε_i), the above expectation

$$= E \mid E \exp(it_1\varepsilon_n(A_n + B_{nm}) + it_2\varepsilon_n^2)/\varepsilon_j,\ j \neq n \mid \qquad \ldots(2.2.4)$$

$$\leq \exp(-\delta)P(\| t_{nm} \| \geq d) + P(\| t_{nm} \| \leq d)$$

where $t'_{nm} = (t_1(A_n + B_{nm}), t_2)$.

Note that $A_n + B_{nm} \xrightarrow{\mathcal{L}} Z_1 + Z_2$ where Z_i are iid $Z_1 \stackrel{\mathcal{L}}{=} \sum_{t=0}^{\infty} \theta^t \varepsilon_t$.

Choose λ such that $0 < \lambda < 1$ and $P(|Z_1 + Z_2| > \lambda) > 0$.

Then we have

$$P(\| t_{nm} \| \geq d) = P(t_1^2(A_n + B_{nm})^2 + t_2^2 \geq d^2)$$

$$\geq P((A_n + B_{nm})^2 \geq \lambda^2) \quad \text{if} \quad \| t \|^2 \geq d_1^2 = d^2/\lambda^2$$

$$\longrightarrow P(|Z_1 + Z_2| \geq \lambda) > 0.$$

Thus $\exists\ \varepsilon > 0$, and m_o such that

$P(\| t_{nm} \| \geq d) > \varepsilon \ \forall\ n,m \geq m_o$. This verifies C(1.2.5).

If $S_n = (n^{-1/2} \sum_{t=1}^{n} Y_{t-1}\varepsilon_t , \; n^{-1/2} \sum_{t=1}^{n} (\varepsilon_t^2 - 1))$, then we have the following theorem as an immediate consequence of the above discussion. (See Theorem 1.2.1).

<u>Theorem 2.2.1</u> : Assume (A1) and (A2) hold. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ denote a measurable function such that $|f(x)| \leq M(1+ \|x\|^{s_0}) \; \forall \; x \in \mathbb{R}^2$. Then there exists a positive constant δ not depending on f and M, and for arbitrary $k > 0$, there exists a positive constant C depending on M but not on f such that

$$| E \; f(S_n) - \int f d\psi_{ns} | \leq C . w(f, n^{-k}) + o(n^{-(s-2+\delta)/2})$$

where $w(f, n^{-k}) = \int \sup(|f(x+y) - f(x)| \; : \; \|y\| \leq n^{-k}) \varphi_\Sigma(x) dx$. φ_Σ is the normal density with zero mean and coveriance matrix Σ. ψ_{ns} is the usual function associated with Edgeworth expansions. The term $o(n^{-(s-2+\delta)/2})$ depends on f through M only.

We immediately have the following corollary.

<u>Corollary 2.2.2</u> : Assume (A1) and (A2). Then uniformly for convex measurable $C \subset \mathbb{R}^2$,

$$P(S_n \in C) = \psi_{n,s}(C) + o(n^{-(s-2)/2}).$$

Theorem 2.2.1 is not directly applicable to $(1-\theta^2)^{-1} n^{1/2} (\theta_n - \theta)$. Nevertheless, we have the following theorem which is enough to study the accuracy of bootstrap approximation.

<u>Theorem 2.2.3</u> : Assume (A1) and (A2). Then there exists a polynomial p in one variable which is a continuous function of θ and moments of

$(X_{j-1}\varepsilon_j, \varepsilon_j^2 - 1)$ of order less or equal to 3 such that

$$P\left(\left(\frac{n}{1-\theta^2}\right)^{1/2}(\theta_n - \theta) \le x\right) = \int_{-\infty}^{x} (1 + n^{-1/2}p(y))\varphi(y)dy + o(n^{-1/2}).$$

<u>Proof</u> :

$$\left(\frac{n}{1-\theta^2}\right)^{1/2}(\theta_n - \theta) = \frac{X_1(1-\theta^2)^{-1/2}}{1+n^{-1/2}(2\theta X_1 + X_2) + A_n}$$

where $X_1 = n^{-1/2} \sum_{t=1}^{n} Y_{t-1}\varepsilon_t$

$$X_2 = n^{-1/2} \sum_{t=1}^{n} (\varepsilon_t^2 - 1), \quad A_n = n^{-1}(Y_o^2 - Y_n^2).$$

Let $B_1 = \left\{ |X_1| \ge b \,.\, \log n \right\}$

$$B_2 = \left\{ |X_2| \ge b \log n \right\}$$

$$B_3 = \left\{ n^{3/4}|A_n| \ge b \log n \right\}$$

By Theorem 2.2.1, for sufficiently large b,

$$P(B_1),\ P(B_2) = o(n^{-1/2}) \text{ and obviously } P(B_3) = o(n^{-1/2}).$$

On $B_1^c \cap B_2^c \cap B_3^c$, we have,

$$\begin{aligned}\left(\frac{n}{1-\theta^2}\right)^{1/2}(\theta_n - \theta) &= X_1(1-\theta^2)^{-1/2}(1 - n^{-1/2}(2\theta X_1 + X_2)) + o(n^{-1/2}) \\ &= (\ell' X + n^{-1/2} X' A X)(\ell' \Sigma \ell)^{-1/2} + o(n^{-1/2})\end{aligned} \qquad \ldots(2.2.5)$$

where $\ell' = (1,0)$, $X' = (X_1, X_2)$, and $A = \begin{pmatrix} -2\theta & -1/2 \\ -1/2 & 0 \end{pmatrix}$.

Now the result follows from Theorem 2.2.1 and the following lemma due to Babu and Singh (1984) (which is essentially a version of Lemma 2.1 of Bhattacharya and Ghosh (1978)).

Lemma 2.2.4 : Let $\lambda = (\lambda_1,\ldots,\lambda_k)$ be a vector, $L = (L_{ij})$ be a $k \times k$ matrix and Q be a polynomial in k variables. Let $M \geq \max(|v_{ij}|, |u_{ij}|, |\lambda_i|, |L_{ij}|, |a_\lambda|)$ where $(v_{ij}) = V$, $(u_{ij}) = V^{-1}$ and a_λ are coefficients of Q. Let $|\lambda_1| > \lambda_0 > 0$ and $b_n = (\lambda_1 n^{1/2})^{-1}$. Then there exists a polynomial p in one variable, whose coefficients are continuous functions of λ_i, L_{ij}, v_{ij}, u_{ij} and a_λ such that

$$\int_A (1+n^{-1/2}Q(Z)\varphi_V(Z)dZ = \int_{-\infty}^{u} (1+b_n p(y))\varphi(y)dy + o(n^{-1/2})$$

where $A = \left\{ Z : \lambda.Z + n^{-1/2}Z'LZ < u(\lambda' V\lambda)^{1/2} \right\}$. The $o(.)$ term depends on M and λ_0.

2.3 The bootstrap approximation

In this section we assume (A1) and (A2). As a consequence, $D(\varepsilon_1, \varepsilon_1^2)$ is positive definite.

Given the l.s.e. θ_n, we " recover" the errors by

$$\hat{\varepsilon}_i = Y_i - \theta_n Y_{i-1}, \quad i = 1,\ldots,n.$$

Let $G_n(.)$ denote the empirical distribution function of $\hat{\varepsilon}_1,\ldots,\hat{\varepsilon}_n$ (i.e. the distribution which puts mass $1/n$ at each $\hat{\varepsilon}_i$, $i = 1,\ldots,n$). Let $F_n^*(x) = G_n(x - \bar{\varepsilon}_n)$ where $\bar{\varepsilon}_n = n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i$.

Let (ε_i^*), $i = 0, \pm 1, \pm 2, \ldots$ be iid with $F_n^*(\cdot)$. (Strictly speaking we should write ε_{in}^*, but we shall drop the suffix n to ease notations). Given (ε_i^*), $i = 0, \pm 1, \pm 2, \ldots$ generate Y_i^* by

$$Y_i^* = \theta_n Y_{i-1}^* + \varepsilon_i^*, \quad i = 0, \pm 1, \pm 2, \ldots.$$

Pretend that θ_n is unknown and obtain the l.s.e. of θ_n from the pseudo-data Y_i^*, $i = 0, 1, \ldots, n$ by

$$\theta_n^* = \Big(\sum_{i=1}^n Y_{i-1}^* Y_i^*\Big)\Big(\sum_{i=1}^n Y_{i-1}^{*2}\Big)^{-1}.$$

Hence

$$(1-\theta_n^2)^{-1/2}\, n^{1/2}(\theta_n^* - \theta_n) = (1-\theta_n^2)^{-1/2}\, n^{-1/2} \sum_{i=1}^n Y_{i-1}^* \varepsilon_i^* / n^{-1} \sum_{i=1}^n Y_{i-1}^{*2},$$

the bootstrap equivalent of $(1-\theta^2)^{-1/2}\, n^{1/2}(\theta_n - \theta)$.

In this section we make the convention that the presence of $(*)$ denotes that we are dealing with the bootstrap quantity and hence expectation etc. are taken under (ε_i^*) iid F_n^* given $Y_o, Y_1, \ldots, Y_n$.

To state and prove the main theorem we need the following lemmas. The paper of Gotze and Hipp (1983) shall be simply referred as (GH) in the sequel.

Let $H_n^*(\cdot)$ denote the characteristic function of $n^{-1/2} \sum_{j=1}^n Z_j^*$, where Z_j^* is a certain truncation (as in (GH)) of

$X_j^* = (Y_{j-1}^* \varepsilon_j^*,\ \varepsilon_j^{*2} - \sigma_n^{*2})$. Here $\sigma_n^{*2} = E^*(\varepsilon_n^{*2})$. We omit the explicit definition of Z_j^* since this shall not be used in subsequent calculations.

<u>Lemma 2.3.1</u> : $\forall \; \| t \| \leq C.n^{\varepsilon^0}$ we have

$$| D^{\alpha}(H_n^*(t) - \hat{\psi}_{n3}^*(t)) | \leq C.(1+\beta_{4n}^*)(1+ \| t \|^{6+|\alpha|})\exp(-C \| t \|^2) n^{-1/2-\varepsilon^0}$$

for some $\varepsilon^0 < 1/2$, and C depends on the bounds of β_{4n}^* (= 4th moment of X_j^*). $\hat{\psi}_{ns}^*(t)$ denotes the Fourier transform of $\psi_{n,s}^*$, the signed measure associated with the Edgeworth expansion of X_j^*. D^{α} is the usual differential operator with $|\alpha| \leq s+3$.

This lemma is proved in (GH) and hence we skip the proof.

Let $I_1 = \left\{ t : Cn^{\varepsilon^0} \leq \| t \| \leq C_1 n^{1/2} \right\}$

$I_2 = \left\{ t : C_1 n^{1/2} \leq \| t \| \leq \varepsilon^{-1} n^{1/2} \right\}$ where C_1 is to be chosen later and $0 < \varepsilon < 1$ is fixed.

<u>Lemma 2.3.2</u> : Under (A1), (A2), we have, for almost every sequence $Y_0, Y_1, \ldots$ and $|\alpha| \leq 6$,

$$\int_{C_1 n^{1/2} \leq \|t\| \leq \varepsilon^{-1} n^{1/2}} | D^{\alpha} H_n^*(t) | dt = o(n^{-1/2}).$$

<u>Proof</u> : A careful look at the proof of Lemma 3.43 of GH shows that it suffices to show that

$$E^* | E^* A_p^* / \mathcal{D}_j^*, \; j \neq j_p | < 1 \quad \text{uniformly in } t \in I_2 \text{ and } p = 1, \ldots, \lambda$$

where $A_p^* = \exp(itn^{-1/2} \sum_{j=j_p-m}^{j_p+m} Z_j^*)$. For definition of λ and j_p, one can consult GH. We omit these definitions since they are not used explicitly in our calculation.

But note that the effect of truncation is negligible and it suffices to deal with

$$\delta_{nm} = E^{*}|E^{*}\exp(itn^{-1/2}\sum_{j=j_p-m}^{j_p+m} X_j^{*})/\varepsilon_j^{*},\ j \neq j_p|.$$

If F_{nm}^{*} is the distribution of $\sum_{t=0}^{\infty} \theta_n^t \varepsilon_{n-1-t}^{*} + \sum_{j=m+1}^{n+m} \theta_n^{j-1-n} \varepsilon_j^{*}$

then writing $t_1' = t_1 n^{-1/2}, t_2' = t_2 n^{-1/2}$, we have

$$\delta_{nm} = \int \left| \int \exp(it_1'xy + it_2'x^2)dF_n^{*}(x) \right| dF_{nm}^{*}(y) \quad \text{(see (2.2.4))}$$

Note that a.s., $F_{nm}^{*} \Rightarrow F$ where F is the distribution of $Z_1 + Z_2$, Z_1, Z_2 iid $Z_1 \overset{\mathcal{L}}{=} \sum_{t=0}^{\infty} \theta^t \varepsilon_t$ and by Levy's theorem F is continuous.

Further $F_n^{*} \Rightarrow F_0$ a.s.

Noting that the convergence of F_{nm}^{*} to F is uniform we have

$$\delta_{nm} \longrightarrow \delta = \int \left| \int \exp(it_1'xy + it_2'x^2)dF_0(x) \right| dF(y) \quad \text{a.s.}$$

uniformly on compact sets of (t_1', t_2') i.e. uniformly over $t \in I_2$ and by Cramer's condition $\delta < 1$. This proves the lemma.

<u>Lemma 2.3.3</u> : Assume (A1), (A2) hold. For sufficiently small C_1, we have for almost every sequence $Y_0, Y_1, \ldots,$

$$\int_{t \in I_1} |D^{\alpha} H_n^{*}(t)| \, dt = o(n^{-1/2}).$$

<u>Proof</u> : As in Lemma 2.3.2, it is sufficient to deal with original variables instead of truncations. As before, we proceed as in Lemma 2.3.2

following GH but use a different estimate for $E^*|E^*A_p^*/\mathcal{D}_j^*,\ j \neq j_p|$ (see Lemma 2.3.2 for definition of A_p^*). We have to deal with

$$\delta_{nm}^* = E^*|E^*\exp(it_1 n^{-1/2}\ \varepsilon_n^*(A_n^* + B_{nm}^*) + it_2 n^{-1/2}\ \varepsilon_n^{*2})/\varepsilon_j^*,\ j \neq n|$$

where $A_n^* = \sum_{t=0}^{\infty} \theta_n^t\, \varepsilon_{n-1-t}^*$,

$$B_{nm}^* = \sum_{j=n+1}^{n+m} \theta_n^{j-1-n}\ \varepsilon_j^*$$

$$\delta_{nm}^* \leq P^*(|A_n^* + B_{nm}^*| \geq b)$$

$$+ E^*\Big[|E^*\exp(it_1 n^{-1/2}\varepsilon_n^*(A_n^* + B_{nm}^*) + it_2 n^{-1/2}\ \varepsilon_n^{*2})/\varepsilon_j^*\ j \neq n|$$

$$I(|A_n^* + B_{nm}^*| \leq b)\Big]$$

For large b, the first term is $< 1/2$.

In the second term, the inner expectation

$$= 1 - \frac{t_n'}{2n} D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n + \frac{\gamma}{6}\ \frac{\|t_n^*\|^3}{n^{3/2}} E^* \|\varepsilon_n^*, \varepsilon_n^{*2}\|^3$$

where $t_n' = (t_1(A_n^* + B_{nm}^*),\ t_2)$ and $|\gamma| \leq 1$.

In the above expression the last term is

$$\leq \frac{1}{6}\ b^3\ \frac{\|t\|^3}{n^{3/2}} \mu_{3n}^* \quad \text{where} \quad \mu_{3n}^* = E^* \|\varepsilon_n^*, \varepsilon_n^{*2}\|^3$$

$$\leq \beta\ \frac{b^3 C_1}{6}\ \frac{\|t\|^2}{n} \quad \text{a.s. noting that} \quad \mu_{3n}^* \xrightarrow{\text{a.s.}} E\,\|\varepsilon_1,\ \varepsilon_1^2\|^3$$

$$\leq \alpha\ \frac{\|t\|^2}{n},$$ where α is as small as we please by taking ε_1 sufficiently small.

Again note that $D(\varepsilon_n^*, \varepsilon_n^{*2}) \xrightarrow{a.s.} D(\varepsilon_1, \varepsilon_1^2)$,which is positive definite, and on $\{|A_n^* + B_{nm}^*| \leq b\}$ we have,

$$\left|\frac{t_n'}{2n} D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n\right| \leq |D(\varepsilon_n^*, \varepsilon_n^{*2})| \frac{b^2}{2n} \|t\|^2 \leq 1 \quad \text{if } C_1 \text{ is small}$$

and hence second term of δ_{nm}^* is bounded by

$$\delta_{2nm} = \alpha \frac{\|t\|^2}{n} + E^*\left(1 - \frac{1}{2n} t_n' D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n\right) I(|A_n^* + B_{nm}^*| \leq b)$$

$$\leq \alpha \frac{\|t\|^2}{n} + E^*\left(1 - \frac{1}{2n} I(|A_n^* + B_{nm}^*|| \leq b) \lambda_1(\Sigma_n^*) \| t_n \|^2\right)$$

where $\lambda_1(\Sigma_n^*)$ = smallest eigen value of $D(\varepsilon_n^*, \varepsilon_n^{*2})$ and hence $\lambda_1(\Sigma_n^*) \xrightarrow{a.s.} \lambda_1(D(\varepsilon_1, \varepsilon_1^2)) > 0$.

If b is large such that $P^*\{|A_n^* + B_{nm}^*| \leq b\} > \frac{1}{2}$, we have that

$$\left[E^*(A_n^* + B_{nm}^*)^2 I(|A_n^* + B_{nm}^*| \leq b) - E^*(A_n^* + B_{nm}^*)^2\right]$$

$$\leq \left[E^*(A_n^* + B_{nm}^*)^4 P^*(|A_n^* + B_{nm}^*| > b)\right]^{1/2}$$

$< \eta$, η sufficiently small by choosing b large enough.

(Note that $E^*(A_n^* + B_{nm}^*)^4 \xrightarrow{a.s.} E(Z_1 + Z_2)^4$, where Z_1 and Z_2 are iid, $Z_1 \stackrel{\mathcal{L}}{=} \sum_{t=0}^{\infty} \theta^t \varepsilon_t$).

And thus

$$\delta_{2nm} \leq \alpha \frac{\|t\|^2}{n} + E^*\left(1 - \frac{\lambda_1(\Sigma_n)}{2n} I(|A_n^* + B_{nm}^*| \leq b)(t_2^2 + t_1^2(A_n^* + B_{nm}^*)^2)\right)$$

This term in turn is dominated by

$$\alpha \frac{\| t \|^2}{n} + 1 - \gamma \frac{\| t \|^2}{n} \quad (\text{where } \gamma > \alpha)$$

$$\leq \exp(-\delta \frac{\| t \|^2}{n}).$$

This shows that $\delta^*_{nm} \leq \rho + \exp(-\delta \frac{\| t \|^2}{n})$ where $\rho < \frac{1}{2}$. A look at the proof of Lemma 3.43 of (GH) shows that this proves the lemma.

Our final lemma is stated in Babu and Singh (1984) and is a modified version of a lemma in Sweeting (1977). This lemma shall be used in Chapter 3 also.

<u>Lemma 2.3.4</u> : Let P and K be probability measures and Q be a signed measure on $\mathbb{R}^k$. Let f be a measurable function such that $M_s(f) < \infty$ for some $s \geq 2$. Further let $\alpha = K(x : \|x\| \leq 1) > \frac{1}{2}$ and $\beta = \int \|x\|^{s+2} K(dx) < \infty$. Then for any $0 < \varepsilon < 1$,

$$\left| \int f d(P-Q) \right| \leq (2\alpha-1)^{-1} \left[B(1-\alpha)/\alpha \right]^{-1+\varepsilon^{-1/4}} + \beta \varepsilon B$$

$$+ B \int (1 + \|x\|^s) | K^*_\varepsilon (P-Q) | (dx)$$

$$+ \sup_{\|x\| < \varepsilon^{1/4}} \int w(f, 2\varepsilon, x-y) |Q| dy$$

where $K_\varepsilon(dx) = K(\varepsilon^{-1} dx)$ and $B = 9^s M_s(f) \int (1 + \|x\|^s)(P + |Q|) dx$,

$$M_s(f) = \sup_x (1 + \|x\|^s)^{-1} |f(x)|$$

Further we have for any $0 < \|x\| < 1$, $0 < \delta < 1$,

$$\int w(f,\delta,x-y)\varphi(y)dy \leq 3 \int w(f,\delta,y)\varphi(y)dy + c_o M_s(f) \| x \|^{-k-s+1} \exp(-\tfrac{1}{8} \| x \|^{-2}).$$

We are now in a position to state the theorems on bootstrap.

Let $\quad S_n^* = n^{-1/2}\left(\sum_{j=1}^{n} Y_{j-1}^* \varepsilon_j^*, \ \sum_{j=1}^{n} (\varepsilon_j^{*2} - \sigma_n^{*2})\right).$

Theorem 2.3.5 : Assume (A1), (A2).

Let $\quad f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be measurable and $|f(x)| \leq M(1 + \|x\|^{s_o})$.

Then we have

$$| E^* f(S_n^*) - \int f(x)(1 + n^{-1/2} a(x,n)) d\Phi_{\Sigma_n^*}(x) | \leq M\delta_n + C\bar{w}(f, \delta_n \Sigma_n^*)$$

where $a(x,n)$ denotes a polynomial (of the same form as p) in x whose coefficients are continuous function of the moments of $Y_{j-1}^* \varepsilon_j^*$ and ε_j^{*2} of order 3 or less, $\delta_n = o(n^{-1/2})$ and is independent of f, $\Sigma_n^* = D^*(Y_{j-1}^* \varepsilon_j^*, \varepsilon_j^{*2})$.

Proof : The proof follows using Lemmas 2.3.1 - 2.3.4. We omit the details since they would be repetition of arguments given in Babu and Singh (1984) for the i.i.d. case.

Theorem 2.3.6 : Under assumptions (A1) , (A2) we have, for a.e. (Y_i),

$$P^*((1-\theta_n^2)^{-1/2} n^{1/2} (\theta_n^* - \theta_n) \leq x) = \int_{-\infty}^{x} (1 + n^{-1/2} p(y)) \varphi(y) dy + o(n^{-1/2})$$

$$= P\left(\left(\frac{n}{1-\theta^2}\right)^{1/2} (\theta_n - \theta) \leq x\right) + o(n^{-1/2})$$

where $p(\cdot)$ is a polynomial with coefficients continuous functions of θ.

<u>Proof</u> : Note that

$$(1-\theta_n^2)^{-1/2} n^{1/2} (\theta_n^* - \theta_n) = \frac{X_1^*(1-\theta_n^2)^{-1/2}}{1 + n^{-1/2}(2\theta X_1^* + X_2^*) + A_n^*}$$

where $X_1^* = n^{-1/2} \sum_{t=1}^{n} Y_{t-1}^* \varepsilon_t^*$

$$X_2^* = n^{-1/2} \sum_{t=1}^{n} (\varepsilon_t^{*2} - \sigma_n^{*2})$$

$$A_n^* = n^{-1}(Y_0^{*2} - Y_n^{*2}).$$

As in Theorem 2.2.3 define B_1^*, B_2^*, B_3^* analogous to B_1, B_2, B_3. By Theorem 2.3.5, $P^*(B_1^*)$, $P^*(B_2^*)$, $P^*(B_3^*)$ are $o(n^{-1/2})$ and on $B_1^{*c} \cap B_2^{*c} \cap B_3^{*c}$ we have as before

$$(1-\theta_n^2)^{-1/2} n^{1/2} (\theta_n^* - \theta_n) = (\lambda' X^* + n^{-1/2} X^{*\prime} A^* X)(\lambda' \Sigma_n^* \lambda)^{-1/2} + o(n^{-1/2})$$

where $\lambda' = (1,0)$, $X^{*\prime} = (X_1^*, X_2^*)$, $A^* = \begin{pmatrix} -2\theta_n, & -1/2 \\ -1/2, & 0 \end{pmatrix}$ and

$$\Sigma_n^* = \begin{pmatrix} (1-\theta_n^2)^{-1} & 0 \\ 0 & V(\varepsilon_1^{*2}) \end{pmatrix}$$

The theorem now follows from Lemma 2.2.4, and Theorem 2.3.5. (Note that the moments of $(X_{j-1}^* \varepsilon_j^*, \varepsilon_j^{*2} - \sigma_n^{*2})$ under E^* converge almost surely to those of $(X_{j-1}\varepsilon_j, \varepsilon_j^2 - 1)$ by ergodic theorem, and under our assumptions $\theta_n \rightarrow \theta$ a.s.).

Remark 2.3.7

The assumption of stationarity of (Y_t) was made since the calculations (e.g. of Σ) in this case is simpler. The results hold even if this assumption is dropped. This is fairly obvious, since the asymptotic structure does not change and the results of GH go through.

In the next section we move over to higher order autoregressions. We shall see that analogue of Theorem 2.3.6 is valid. However the distribution of the l.s.e. does not seem to admit an Edgeworth expansion of order s (see Remark 2.4.1).

2.4 Edgeworth expansions and bootstrap in higher order autoregressions

Let Y_t be a stationary autoregressive process satisfying

$$Y_t = \sum_{i=1}^{p} \theta_i Y_{t-i} + \varepsilon_t \qquad \ldots(2.4.1)$$

where (ε_t) satisfies assumptions (A1), (A2). We also assume that

(A3) Roots of $\sum_{j=0}^{p} \theta_j z^{p-j} = 0$ lies within the unit circle. Here $\theta_0 = 1$.

Under our conditions, $Y_t = \sum_{r=0}^{\infty} \delta_r \varepsilon_{t-r}$

where $\exists\ c, \alpha > 0$ and $N_0 \ni \forall\ N \geq N_0$, $\sum_{r=N}^{\infty} |\delta_r| \leq c \exp(-\alpha N)$.

Given $(Y_{1-p}, \ldots, Y_n)$, the least squares equations are

$$S_n \begin{bmatrix} \theta_{1n} \\ \vdots \\ \theta_{pn} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{n} Y_t Y_{t-1} \\ \vdots \\ \sum_{t=1}^{n} Y_t Y_{t-p} \end{bmatrix}$$

where $S_n = \begin{bmatrix} \sum_{t=1}^{n} Y_{t-1}^2 & \sum_{t=1}^{n} Y_{t-1}Y_{t-2} & \cdots & \sum_{t=1}^{n} Y_{t-1}Y_{t-p} \\ & \sum_{t=1}^{n} Y_{t-2}^2 & & \\ & & \ddots & \\ & & & \sum_{t=1}^{n} Y_{t-p}^2 \end{bmatrix}$

Using equation (2.4.1), this can be written as

$$S_n \begin{bmatrix} \theta_{1n} - \theta_1 \\ \vdots \\ \theta_{pn} - \theta_p \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{n} Y_{t-1}\varepsilon_t \\ \vdots \\ \sum_{t=1}^{n} Y_{t-p}\varepsilon_t \end{bmatrix}$$

Let $\sigma_i = \mathrm{Cov}(Y_o, Y_i)$, $i = 1, \ldots, (p-1)$.

It is well known that

$$\Sigma = \begin{bmatrix} \sigma_o & \sigma_1 & \cdots & \sigma_{p-1} \\ & \sigma_o & & \sigma_{p-2} \\ & & \ddots & \\ & & & \sigma_o \end{bmatrix}$$ is positive definite.

Define $X_{it} = Y_{t-i}\varepsilon_t, \quad i = 1,\ldots,p$

$$X_{p+1,t} = \varepsilon_t^2 - 1$$

$$X_t' = (X_{1t},\ldots,X_{p+1,t}), \quad \mathcal{D}_j = \sigma(\varepsilon_j).$$

It can be easily checked that

$$\lim_{n\to\infty} D(n^{-1/2} \sum_{t=1}^{n} X_t) = \begin{pmatrix} \Sigma & 0 \\ 0 & V(\varepsilon_1^2) \end{pmatrix}$$

which is positive definite.

As before, to obtain asymptotic expansion for the distribution of $n^{-1/2} \sum_{t=1}^{n} X_t$, we use Theorem 1.2.1 of Chapter 1.

Conditions $C(1.2.1) - C(1.2.4)$ and $C(1.2.6)$ are immediate and it remains to check $C(1.2.5)$. Below β shall stand for random variables which are independent of ε_n.

Let $t' = (t_1,\ldots,t_p,\ t_{p+1}) \in \mathbb{R}^{p+1}$

Then $t' \sum_{j=n-m}^{n+m} X_j = t_{p+1}\varepsilon_n^2 + \sum_{j=n}^{n+m} \sum_{i=1}^{p} t_i Y_{j-i}\varepsilon_j + \beta$

Observe that
$$\begin{aligned} Y_{j-i}\varepsilon_j &= \varepsilon_n \sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r} + \beta \quad \text{if} \quad j = n \\ &= \varepsilon_n \delta_{j-i-n} \varepsilon_j + \beta \quad \text{if} \quad j-i \geq n \\ &= \beta \qquad \text{if} \quad j-i < n \quad \text{but} \quad j > n. \end{aligned}$$

Hence $\sum_{j=n}^{n+m} \sum_{i=1}^{p} t_i Y_{j-i} \varepsilon_j$

$$= \varepsilon_n \sum_{i=1}^{p} t_i \sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r}$$

$$+ \varepsilon_n \sum_{i=1}^{p} t_i \sum_{j=n+i}^{n+m} \delta_{j-i-n} \varepsilon_j$$

$$= \varepsilon_n \left[\sum_{i=1}^{p} t_i \sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r} + \sum_{s=0}^{m-i} \delta_r \varepsilon_{s+i+n} \right]$$

Let $A_{in} = \sum_{r=0}^{\infty} \delta_r \varepsilon_{n-i-r}$

$$B_{inm} = \sum_{r=0}^{m-i} \delta_r \varepsilon_{r+i+n}, \quad i = 1,\ldots,p.$$

Note that $\forall i$, $i = 1,\ldots,p$, A_{in} and B_{inm} are independent and

$$Z'_{nm} = (A_{in} + B_{inm},\ i = 1,\ldots,p) \xrightarrow{\mathcal{L}} (Z_{i1} + Z_{i2},\ i = 1,\ldots,p)$$

where Z_{i1} and Z_{i2} are independent for every i and

$$\text{Cov}(Z_{i1}, Z_{j1}) = \text{Cov}(Z_{i2}, Z_{j2}) = \sigma_{i-j}.$$

Thus the limiting dispersion matrix of Z_{nm} is positive definite.

$$E \mid E \exp(it' \sum_{j=n-m}^{n+m} X_j)/\mathcal{D}_j,\ j \neq n \mid$$

$$= E \mid E \exp(i(t_1,\ldots,t_p)\varepsilon_n Z_{nm} + it_{p+1}\varepsilon_n^2)/\varepsilon_j,\ j \neq n \mid$$

$$\leq \exp(-\delta)P(\| t_{nm} \| \geq d) + P(\| t_{nm} \| \leq d) \qquad \ldots(2.4.2)$$

where $t_{nm} = ((t_1,\ldots,t_p)Z_{nm},\ t_{p+1})$.

Suppose that $\| t \|^2 = \sum_{i=1}^{p} t_i^2 + t_{p+1}^2 \geq d_1^2 = d^2/\lambda^2$,

where $0 < \lambda < 1$ is to be chosen.

Then $P(\| t_{nm} \| \geq d) \geq P((a'Z_{nm})^2 \geq \lambda^2)$ where $\| a \| = 1$.

Let $b_1, \ldots, b_{p+1}$ be $(p+1)$ points in $\mathbb{R}^{p+1}$ and let $r > 0$ be such that

$$P(Z_{j1} + Z_{j2},\ j = 1,\ldots,p) \in B(b_i, r)) > 0 \quad \forall i = 1,\ldots,p \text{ and}$$

$B(b_i, r)$ $i = 1,\ldots,p$ are such that not all $(b_1,\ldots,b_{p+1})$ lie in a given hyperplane of dimension $(p-1)$. This is possible since the dispersion matrix of $(Z_{j1} + Z_{j2},\ j = 1,\ldots,p)$ is positive definite. Choose λ sufficiently small. Then for any a with $\| a \| = 1$, $\left\{ x : (a'x)^2 < \lambda^2 \right\}$ does not intersect at least one of the balls.

Hence $\quad P((a'Z_{nm})^2 \geq \lambda^2) \geq \min\limits_{i=1,\ldots,(p+1)} (Z_{nm} \in B(b_i, r))$

and lim inf of the right hand side is positive. Thus there exists n_o, m_o large enough and $\varepsilon > 0$ such that

$$P(\| t_{nm} \|^2 \geq d^2) > \varepsilon > 0 \quad \forall \| t \|^2 \geq d_1^2,\ n \geq n_o,\ m \geq m_o.$$

Condition C(1.2.5) follows now from (2.4.2) and the above fact.

Hence we have asymptotic expansion for the distribution of $n^{-1/2} \sum_{t=1}^{n} X_t$.

<u>Remark 2.4.1</u> :

The above arguments also show that if conditions (A1) and

(A3) hold and ε_1 satisfies Cramer's condition, then the distribution of $n^{-1/2}\, S_n(\underset{\sim}{\theta}_n - \underset{\sim}{\theta})$ admits an Edgeworth expansion. Here $\underset{\sim}{\theta}_n' = (\theta_{1n}, \ldots, \theta_{pn})$.

To get an asymptotic expansion for $\underset{\sim}{\theta}_n'$ with usual normalization we proceed as follows.

Define
$$\tilde{Y}_t = \begin{bmatrix} Y_t \\ \vdots \\ Y_{t-p+1} \end{bmatrix}_{p \times 1}, \quad \tilde{Z}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{p \times 1}$$

$$B = \begin{pmatrix} \theta_1 \ldots \theta_{p-1} & \theta_p \\ I_{p-1} & 0 \end{pmatrix}$$

It is easy to see that (2.4.1) is equivalent to $\tilde{Y}_t = B\tilde{Y}_{t-1} + \tilde{Z}_t$ which gives

$$\tilde{Y}_t\tilde{Y}_t' = B\tilde{Y}_{t-1}\tilde{Y}_{t-1}'B' + 2B\tilde{Y}_{t-1}\tilde{Z}_t' + \tilde{Z}_t\tilde{Z}_t' \qquad \ldots(2.4.3)$$

Let $A_n = n^{-1}\sum_{t=1}^{n} \tilde{Y}_t\tilde{Y}_t'$, $B_n = n^{-1}\sum_{t=1}^{n} \tilde{Y}_{t-1}\tilde{Y}_{t-1}'$.

Clearly $A_n = B_n + O_p(n^{-1})$. Hence from (2.4.3),

$$B_n - BB_nB' = n^{-1}\sum_{t=1}^{n} \tilde{Z}_t\tilde{Z}_t' + 2Bn^{-1}\sum_{t=1}^{n} \tilde{Y}_{t-1}\tilde{Z}_t' + O_p(n^{-1}) \qquad \ldots(2.4.4)$$

Note that $\Sigma = E\,\tilde{Y}_t\tilde{Y}_t'$ satisfies the equation $\Sigma = B\Sigma B' + I^*$

where
$$I^* = \begin{bmatrix} 1 & 0 \ldots 0 \\ 0 & 0 \ldots 0 \\ \cdots & \cdots \\ \cdots & \cdots \\ 0 & 0 \ldots 0 \end{bmatrix}_{p \times p}$$

Equation (2.4.4) yields,

$$(B_n - \Sigma) + \theta(B_n - \Sigma)\theta' = n^{-1} \sum_{t=1}^{n} (\tilde{Z}_t \tilde{Z}_t' - I^*)$$

$$+ 2\theta n^{-1} \sum_{t=1}^{n} \tilde{Y}_{t-1} \tilde{Z}_t' + O_p(n^{-1})$$

$B_n - \Sigma$ can be solved uniquely from this equation and

$$B_n - \Sigma = G_1 \left[n^{-1} \sum_{t=1}^{n} (\tilde{Z}_t \tilde{Z}_t' - I^*) + 2\theta n^{-1} \sum_{t=1}^{n} \tilde{Y}_{t-1} \tilde{Z}_t' \right] G_2 + O_p(n^{-1})$$

where G_1, G_2 are independent of n but depend on $\theta_1, \ldots, \theta_p$.

Let $$V_n = \begin{bmatrix} \sum_{t=1}^{n} X_{1t} \\ \vdots \\ \sum_{t=1}^{n} X_{pt} \end{bmatrix}$$

Clearly $$n^{1/2}(\underset{\sim}{\theta}_n - \underset{\sim}{\theta}) = \left[\frac{\sum \tilde{Y}_{t-1} \tilde{Y}_{t-1}'}{n} \right]^{-1} n^{-1/2} V_n$$

$$= (B_n - \Sigma + \Sigma)^{-1} n^{-1/2} V_n$$

$$= \Sigma^{-1} \left[I + (B_n - \Sigma)\theta \Sigma^{-1} \right]^{-1} n^{-1/2} V_n \quad \text{which is equal to}$$

$$\Sigma^{-1} \left[I + G_1 (n^{-1/2} \sum_{t=1}^{n} (\tilde{Z}_t \tilde{Z}_t' - I^*) + 2\theta n^{-1} \sum_{t=1}^{n} \tilde{Y}_{t-1} \tilde{Z}_t') G_2 \Sigma^{-1} \right]^{-1} n^{-1/2} V_n .$$

Let $$B_i = \left\{ n^{-1/2} \left| \sum_{t=1}^{n} X_{it} \right| \geq c \log n \right\}, \; i = 1, \ldots, (p+1).$$

Since we have asymptotic expansions for the distribution of $n^{-1/2} \sum_{t=1}^{n} X_{it}$, it follows that for a large c, $P(B_i) = o(n^{-1/2})$ $\forall\, i = 1,\ldots,(p+1)$.

On $\bigcap_{i=1}^{p+1} B_i^c$, similar to equation (2.2.5),

$$n^{1/2}(\underset{\sim}{\theta}_n - \underset{\sim}{\theta}) = \Sigma^{-1} n^{-1/2} \begin{bmatrix} \sum_{t=1}^{n} X_{1t} \\ \vdots \\ \sum_{t=1}^{n} X_{pt} \end{bmatrix} + n^{-1/2} \begin{bmatrix} X' L_1 X \\ \vdots \\ X' L_p X \end{bmatrix} + o_p(n^{-1/2})$$

$$X = n^{-1/2} \sum_{t=1}^{n} X_t \qquad \ldots(2.4.5)$$

Thus by multidimensional version of Lemma 2.1 of Bhattacharya and Ghosh (1978) or of Lemma 2.2.4, quoted from Babu and Singh (1984) we have the following theorem.

<u>Theorem 2.4.2</u> : Under assumptions (A1) - (A3),

$$\sup_{\underset{\sim}{x}} \left| P(n^{1/2}\Sigma^{1/2}(\underset{\sim}{\theta}_n - \underset{\sim}{\theta}) \leq \underset{\sim}{x}) - \int_{-\infty}^{\underset{\sim}{x}} (1+n^{-1/2}p(\underset{\sim}{y}))\varphi(\underset{\sim}{y})d\underset{\sim}{y} \right| = o(n^{-1/2})$$

where p is a polynomial whose coefficients are continuous functions of moments of $Y_{j-i}\varepsilon_j$, $i = 1,\ldots,p$ and $\varepsilon_j^2 - 1$, of order three or less.

The distribution can be bootstrapped as in the case $p = 1$.

Here $\quad \hat{\varepsilon}_t = Y_t - \sum_{i=1}^{p} \theta_{in} Y_{t-i} \quad t = 1,\ldots,n.$

Compute the empirical distribution function as before and proceed exactly as before. We have the following multidimensional version of Theorem 2.3.6.

<u>Theorem 2.4.3</u> : Under assumptions (A1) - (A3), for a.e (Y_i),

$$\sup_x \left| P^*(n^{1/2}\Sigma_n^*(\underset{\sim}{\theta}_n^* - \underset{\sim}{\theta}_n) \leq x) - P(n^{1/2}\Sigma(\underset{\sim}{\theta}_n - \underset{\sim}{\theta}) \leq x) \right| = o(n^{-1/2}).$$

<u>Proof</u> : As in the case $p=1$, the bootstrapped $n^{-1/2}\sum_{t=1}^{n} X_t^*$ admits an Edgeworth expansion. Representation analogous to (2.4.5) holds for the bootstrapped $n^{1/2}(\underset{\sim}{\theta}_n - \underset{\sim}{\theta})$, yielding the bootstrap analogue of Theorem 2.4.2. As before, by ergodic theorem the bootstrapped moments of $Y_{j-i}^*\varepsilon_j^*$ converges to those of $Y_{j-i}\varepsilon_j$. Thus an application of Theorem 2.4.2 completes the proof.

<u>Remark 2.4.4</u>

The assumptions $E\,\varepsilon_t = 0$ $E\,\varepsilon_t^2 = 1$ may seem to be too restrictive. Actually these restrictions were imposed to keep the proofs simpler. We sketch below how the case $E\,\varepsilon_t = \mu$, $E\,\varepsilon_t^2 = \sigma^2$ can be tackled. We illustrate the case $p = 1$ only.

The model in this case is,

$Y_t = \theta Y_{t-1} + \varepsilon_t + \mu$ where (ε_t) satisfies (A1) - (A3) but $E\,\varepsilon_t^2 = \sigma^2 > 0$ and μ and σ^2 are unknown.

Under assumptions (A1) - (A3), Edgeworth expansion is valid for the distribution of

$$n^{-1/2}\left(\sum_{t=1}^{n}(Y_t - \alpha_1),\ \sum_{t=1}^{n}(Y_tY_{t-1} - \alpha_2),\ \sum_{t=1}^{n}(Y_t^2 - \alpha_3)\right) \qquad \ldots(2.4.6)$$

where $\alpha_1 = EY_t$, $\alpha_2 = EY_tY_{t-1}$ and $\alpha_3 = EY_t^2$.

Estimates θ_n and μ_n of θ and μ are obtained by solving

$$\sum_{t=1}^{n} (Y_t - \theta_n Y_{t-1} - \mu_n) = 0$$

and $$\sum_{t=1}^{n} Y_{t-1}(Y_t - \theta_n Y_{t-1} - \mu_n) = 0 .$$

An estimate σ_n^2 of σ^2 is given by

$$\sigma_n^2 = n^{-1} \sum_{t=1}^{n} (Y_t - \theta_n Y_{t-1} - \mu_n)^2 .$$

Thus, the estimates θ_n, μ_n and σ_n are all smooth functions of $\sum_{t=1}^{n} Y_t$, $\sum_{t=1}^{n} Y_t Y_{t-1}$ and $\sum_{t=1}^{n} Y_t^2$ except for terms which can be neglected.

Thus for a suitable normalizing factor β, the distribution of $n^{1/2}\beta(\theta_n - \theta)$ admits an Edgeworth expansion upto $o(n^{-1/2})$, with the leading term as $\Phi(x)$, and the coefficients involved in the polynomial in the second term are smooth functions of θ, μ, σ^2 and of moments of Y_t, $Y_t Y_{t-1}$ and Y_t^2 of order less or equal to three. β can be explicitly calculated and depends on θ, μ and moments of ε_1.

The empirical distribution is computed by putting mass $1/n$ at each $\hat{\varepsilon}_i = Y_i - \theta_n Y_{i-1} - \mu_n$, $i = 1, \ldots, n$. Proceeding as in the case $\mu = 0$, $\sigma^2 = 1$, an asymptotic expansion is valid for the bootstrapped version of (2.4.6), which yields an expansion of order $o(n^{-1/2})$ for the distribution of $n^{1/2}\beta_n(\theta_n^* - \theta_n)$ where β_n is the variance-normalizing factor, the bootstrap equivalent of β. The leading term of this expansion is also $\Phi(x)$ and the polynomial involved in the second term is of the same form as that in the expansion of $n^{1/2}\beta(\theta_n - \theta)$. By erodic theorem, the empirical moments of Y_t, Y_tY_{t-1} and Y_t^2 converge to the true moments a.s.. θ_n, μ_n and σ_n are all strongly consistent estimators of θ, μ_n and σ respectively. Thus the difference between the two Edgeworth expansions is $o(n^{-1/2})$ a.s.

Remark 2.4.5

The results of this chapter are of course valid for vector-valued autoregressive processes. The proofs are exactly along the same lines as that of the scalar case, with added complexity in notations.

CHAPTER 3

BOOTSTRAP IN MOVING AVERAGE MODELS

3.1 Introduction

In Chapter 2, we gave a terse review of the available literature on bootstrap. We have seen that it works very well for distribution of the least squares estimate in autoregressions.

In this chapter, we shall study the probabilistic aspects of the bootstrap procedure for moving average models and show that the bootstrap approximation beats the normal approximation.

Let (Y_t) be a moving average process satisfying $Y_t = \varepsilon_t + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}$, $t \geq 1$, where (ε_i) are i.i.d. Other conditions on (ε_i) shall be introduced at proper places. $\alpha_1,\ldots,\alpha_p$ are unknown real parameters.

In the absence of distribution assumption on (ε_i), $\alpha_1,\ldots,\alpha_p$ can be estimated by the method of moments. This procedure, though in principle can be applied for any p, becomes increasingly difficult with increase in the value of p. In Chapter 1, we showed that these moment estimators admit Edgeworth expansions under suitable conditions. This together with apparent simplicity of the structure of the process indicates a possibility of bootstrapping the distribution of the estimates with high accuracy.

The structure of the process enables us to resample the errors. Then pseudo data can be generated. We show that the distribution of the moment estimators can be bootstrapped with accuracy $o(n^{-1/2})$ under invertibility condition. The bootstrap fails when the process is not invertible.

We first show that asymptotic expansion can be obtained for bootstrapped distribution of the autocovariances. The moment estimators are smooth functions of these autocovariances. This helps to conclude via a lemma of Bhattacharya and Ghosh (1978) (henceforth referred to as BG) that the bootstrap approximations are accurate upto the order $o(n^{-1/2})$.

The proofs are provided for the case $p = 1$ and sketched for $p = 2$. The general case is not different except for complexities in calculations.

This chapter is a revised version of Bose (1985b).

3.2 Preliminaries

Let (Y_t) be a process satisfying $Y_t = \varepsilon_t + \alpha\varepsilon_{t-1}$ where we assume that

(A1) (ε_t) are i.i.d. $\sim F_o$, $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$, $E\varepsilon_t^{2(s+1)} < \infty$ for some $s \geq 3$.

(A2) $(\varepsilon_1, \varepsilon_1^2)$ satisfies Cramer's condition.

The moment estimate of α, given the observations (Y_i), $0 \leq i \leq n$, is given by

$$\alpha_n = n^{-1} \sum_{t=1}^{n} Y_t Y_{t-1}.$$

Define $\tilde{\varepsilon}_i = \sum_{j=0}^{i-1} (-1)^j \alpha^j Y_{i-j}$, $i = 2, \ldots, n$ and $\tilde{\varepsilon}_1 = Y_1$. Using the structure of the process,

$$\tilde{\varepsilon}_i = \varepsilon_i - (-\alpha)^i \varepsilon_o . \qquad \ldots(3.2.1)$$

This shows that $\tilde{\varepsilon}_i$ and ε_i are close enough for large i only if $|\alpha| < 1$; which in turn shows that resampling of errors shall be proper only in this situation. (For $p > 1$, this condition should be replaced by the invertibility condition (see Hannan (1970)).

Under our conditions, α_n is a strongly consistent estimate of α. Hence motivated by (3.2.1) we define the pseudo errors $\hat{\varepsilon}_{in}$ as

$$\hat{\varepsilon}_{in} = \sum_{j=0}^{i-1} (-1)^j \alpha_n^j Y_{i-j} , \quad i = 2,\ldots,n, \ \hat{\varepsilon}_{1n} = Y_1 .$$

For ease of notations we shall often drop the subscript n.

Let G_n denote the empirical distribution function which puts mass n^{-1} at each $\hat{\varepsilon}_{in}$, $i = 1,\ldots,n$. Define $\hat{F}_n(x) = G_n(x - \bar{\varepsilon}_n)$ where $\bar{\varepsilon}_n = n^{-1} \sum_{i=1}^{n} \varepsilon_{in}$. It is expected that $\hat{F}_n$ shall be close to F_o with increasing n. Take an i.i.d. sample (ε^*_{in}) from $\hat{F}_n$ and define $Y^*_i = \varepsilon^*_i + \alpha_n \varepsilon^*_{i-1}$. Pretend that α_n is unknown and obtain its moment estimate by $\alpha^*_n = n^{-1} \sum_{t=1}^{n} Y^*_t Y^*_{t-1}$. Then it is reasonable to expect that for almost every sequence $Y_o, Y_1, \ldots$, the distribution of α^*_n (given $Y_o, Y_1, \ldots, Y_n$) mimics the distribution of α_n with some accuracy. This statement is made precise in the next section.

3.3 The main results

To prove our main results, we shall require the following lemmas.

Let $\tilde{F}_n$ denote the empirical distribution function of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$.

Lemma 3.3.1 : Under (A1), we have,

(a) $\quad n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_i^k \xrightarrow{\text{a.s.}} E_{F_o}(\varepsilon_1^k) \quad \forall\, k \leq 2(s+1).$

(b) $\quad n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_{in}^k \xrightarrow{\text{a.s.}} E_{F_o}(\varepsilon_1^k) \quad \forall\, k \leq 2(s+1).$

(c) $\quad \tilde{F}_n \Rightarrow F_o \quad$ a.s.

(d) $\quad \hat{F}_n \Rightarrow F_o \quad$ a.s.

Proof : (a) By SLLN, it suffices to show that $n^{-1} \sum_{i=1}^{n} (\tilde{\varepsilon}_i^k - \varepsilon_i^k) \xrightarrow{\text{a.s.}} 0$. But $n^{-1}(\varepsilon_1^k - \varepsilon_1^k) \xrightarrow{\text{a.s.}} 0$ trivially and

$$\left| n^{-1} \sum_{i=2}^{n} (\tilde{\varepsilon}_i^k - \varepsilon_i^k) \right| \leq n^{-1} \sum_{j=0}^{k-1} {}^kC_j \, |\varepsilon_o|^{k-j} \sum_{i=2}^{n} |\varepsilon_i^j| \, |\alpha|^i .$$

Thus it suffices to show that $n^{-1} \sum_{i=1}^{n} |\varepsilon_i|^j |\alpha|^i \to 0$ a.s. We quote the part of Theorem 2.18 of Hall and Heyde (1980) which ensures the above convergence.

Proposition : Let $(S_n = \sum_{i=1}^{n} X_i, \mathcal{F}_n, n \geq 1)$ be a zero mean martingale and $(U_n, n \geq 1)$ a non-decreasing sequence of positive constants. Then

$$\lim_n U_n^{-1} S_n = 0 \text{ a.s. on the set}$$

$$\left\{ \lim_{n\to\infty} U_n = \infty, \ \sum_{i=1}^{\infty} U_i^{-2} E(|X_i|^p / \mathcal{F}_{i-1}) < \infty \right\}, \ p \geq 2 .$$

In the above proposition, take $U_n = n$, $X_i = |\alpha|^i(|\varepsilon_i|^j - E_{F_o}|\varepsilon_i^j|)$, $\mathcal{F}_n = \sigma(X_1,\dots,X_n)$. Then

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} |\alpha|^i(|\varepsilon_i|^j - E_{F_o}|\varepsilon_i|^j) \to 0 \quad \text{a.s.}$$

But $n^{-1}\sum_{i=1}^{n}|\alpha|^i \to 0$. This yields the required convergence and (a) is proved.

(b) By (a), it suffices to show that $n^{-1}\sum_{i=1}^{n}(\tilde{\varepsilon}_i^k - \hat{\varepsilon}_i^k) \xrightarrow{a.s.} 0$.

Note that $\alpha_n \xrightarrow{a.s.} \alpha$ and $|\alpha| < 1$. Hence for all large n,

$$|\alpha| + |\alpha_n - \alpha| \le \beta < 1. \qquad \ldots(3.3.1)$$

Also note that $\forall\, j \ge 1$,

$$\begin{aligned} |\alpha_n^j - \alpha^j| &= |(\alpha_n - \alpha + \alpha)^j - \alpha^j| \\ &< Cj\,|\alpha_n - \alpha|\,\beta^{j-1} \\ &< C\,|\alpha_n - \alpha|\,\delta^j \quad \text{for some } \delta < 1. \qquad \ldots(3.3.2) \end{aligned}$$

$$\left|n^{-1}\sum_{i=1}^{n}(\tilde{\varepsilon}_i^k - \hat{\varepsilon}_i^k)\right|$$

$$= \left|n^{-1}\sum_{i=2}^{n}\left[\Big(\sum_{j=0}^{i-1}(-1)^j\alpha^jY_{i-j}\Big)^k - \Big(\sum_{j=0}^{i-1}(-1)^j\alpha_n^jY_{i-j}\Big)^k\right]\right| \text{ which is dominated by,}$$

$$n^{-1}\sum_{i=2}^{n}\Big(\sum_{j=0}^{i-1}|\alpha_n^j - \alpha^j|\,|Y_{i-j}|\Big)2^{k-1}\Big(\Big|\sum_{t=0}^{i-1}(-1)^t\alpha_n^tY_{i-j}\Big|^{k-1} + \Big|\sum_{t=0}^{i-1}(-1)^t\alpha^tY_{i-t}\Big|^{k-1}\Big)$$

$$\le C.n^{-1}|\alpha_n - \alpha|\sum_{i=2}^{n}\Big(\sum_{j=0}^{i-1}\delta^j|Y_{i-j}|\Big)^k \quad \text{(by (3.3.1) and (3.3.2)).}$$

Thus it is sufficient to show that $n^{-1} \sum_{i=2}^{n} (\sum_{j=0}^{i-1} \delta^j |Y_{i-j}|)^k$ is bounded by a constant a.s. Since $Y_i = \varepsilon_i + \alpha \varepsilon_{i-1}$, it suffices to show that

$$n^{-1} \sum_{i=2}^{n} (\sum_{j=0}^{i-1} \delta^j |\varepsilon_{i-j}|)^k \text{ is bounded a.s.}$$

Define $$Z_i = \sum_{j=0}^{\infty} \delta^j |\varepsilon_{i-j}| \quad \text{(is well defined)} \qquad \ldots(3.3.3)$$

Then the sequence (Z_i) is a stationary autoregressive process of order 1 and hence ergodic (see Hannan (1970), p.204). Thus

$$n^{-1} \sum_{i=1}^{n} Z_i^k \xrightarrow{a.s.} E Z_1^k < \infty . \quad \text{But} \quad (\sum_{j=0}^{i-1} \delta^j |\varepsilon_{i-j}|)^k \le Z_i^k .$$

This proves (b).

(c) Since (ε_i) are i.i.d. F_o this follows readily from (3.2.1).

(d) Suffices to show that $\tilde{F}_n(x) - G_n(x) \xrightarrow{a.s.} 0$ at every continuity point x of F_o. Fix such an x and $\varepsilon > 0$.

$$|\tilde{F}_n(x) - G_n(x)| \le n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_{in} \le x, \tilde{\varepsilon}_i > x)$$
$$+ n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_{in} > x, \tilde{\varepsilon}_i \le x).$$

We handle the second term. The first term can be similarly taken care of.

$$n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_{in} > x, \tilde{\varepsilon}_i \le x)$$
$$\le n^{-1} \sum_{i=1}^{n} I(|\hat{\varepsilon}_{in} - \tilde{\varepsilon}_i| > \varepsilon)$$
$$+ n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_{in} > x, \tilde{\varepsilon}_i \le x, |\hat{\varepsilon}_{in} - \tilde{\varepsilon}_i| \le \varepsilon).$$

The second term in the above expression is bounded by

$$n^{-1} \sum_{i=1}^{n} I(\tilde{\varepsilon}_i \leq x, \tilde{\varepsilon}_i > x-\varepsilon) \longrightarrow F_o(x) - F_o(x-\varepsilon).$$

Since ε is arbitrary and x is a continuity point of F_o, this term is 0.

$$|\hat{\varepsilon}_{in} - \varepsilon_i| \leq \sum_{j=0}^{i-1} |\alpha_n^j - \alpha^j||Y_{i-j}|$$

$$\leq C.|\alpha_n - \alpha| \sum_{j=0}^{i-1} \delta^j |\varepsilon_{i-j}|$$

$$\leq C.|\alpha_n - \alpha| Z_i \quad (\text{see } (3.3.1) - (3.3.3)).$$

This shows that $n^{-1} \sum_{i=1}^{n} I(|\hat{\varepsilon}_{in} - \tilde{\varepsilon}_i| > \varepsilon)$ also tends to zero a.s.

Hence (d) is proved and the proof of the lemma is complete.

In what follows we make the convention that the presence of (*) indicates that we are dealing with the bootstrapped quantity and hence expectation etc. are taken w.r.t. (ε_i^*) i.i.d. $\hat{F}_n$ given $Y_o, Y_1, \ldots, Y_n$.

Define $\quad X_j = Y_j Y_{j-1} - \alpha, \; j \geq 1$

Z_j = the truncation of X_j (as in GH). As in Chapter 1, we omit the details of truncation.

$$H_n(t) = \text{the characteristic function of } n^{-1/2} \sum_{j=1}^{n} Z_j.$$

We have the following lemmas. The proofs are only sketched and the details can be filled in from GH.

<u>Lemma 3.3.2</u> : $\forall\, |t| \leq C.n^{\varepsilon_o}$, we have

$$|D^{\alpha}(H_n^*(t) - \hat{\psi}_{n,s}^*(t)| \leq C(1+\beta_{s+1,n}^*)(1+|t|^{3(s-1)+|\alpha|})\exp(-C|t|^2)n^{-(s-2+\varepsilon_o)/2}$$

for some $\varepsilon_o < 1/2$ and C dependes on the bounds of $\beta^*_{s+1,n} = (s+1)\underline{th}$ moment of X^*_j. D^α is the usual differential operator, $\hat{\psi}_{n,s}(t)$ the Fourier transform of $\psi_{n,s}$, the usual function associated with Edgeworth expansions and $|\alpha| \leq s+2$.

The proof is exactly as the proof of Lemma 3.3 of GH and we omit it.

Let $I_1 = \left\{ t : Cn^{\varepsilon_o} \leq |t| \leq C_1 n^{1/2} \right\}$

$I_2 = \left\{ t : C_1 n^{1/2} \leq |t| \leq \varepsilon^{-1} n^{1/2} \right\}$ where C_1 is to be chosen and $0 < \varepsilon < 1$ is fixed.

Lemma 3.3.3 : Under (A1) and (A2), we have for almost every sequence $Y_o, Y_1, \ldots$

$$\int_{t \varepsilon I_2} |D^\alpha H^*_n(t)| dt = o(n^{-(s-2)/2}).$$

Proof : A careful look at the proof of Lemma 3.43 of GH shows that it suffices to show that $E^*|E^* A^*_p / \mathcal{D}^*_j, j \neq j_p| < 1$ uniformly in $t \varepsilon I_2$ and $p = 1, 2, \ldots, \ell$

where $A^*_p = \exp(itn^{-1/2} \sum_{j=j_p-m}^{j_p+m} Z^*_j)$, $\mathcal{D}^*_j = \sigma(\varepsilon^*_j)$ and for definition of j_p and m see GH. As in Chapter 1, we omit the details of the definitions. But note that the effect of truncation is negligible and it suffices to deal with

$$\delta^*_{nm} = E^*|E^* \exp(itn^{-1/2} \sum_{j=j_p-m}^{j_p+m} X^*_j) / \varepsilon^*_j, \quad j \neq j_p| .$$

Note that $\sum_{j=j_p-m}^{j_p+m} X_j^* = \varepsilon_{j_p}^*(Y_{j_p-1}^* + \alpha_n Y_{j_p+2}^* + \varepsilon_{j_p+1}^* + \alpha_n^2 \varepsilon_{j_p-1}^*) + \alpha_n \varepsilon_{j_p}^{*2} + \beta$

where $\beta \perp \varepsilon_{j_p}^*$.

Let K_n^* denote the distribution function of

$$Y_{j_p-1}^* + \alpha_n Y_{j_p+2}^* + \varepsilon_{j_p+1}^* + \alpha_n^2 \varepsilon_{j_p-1}^*.$$

Then $\delta_{nm}^* = \int \left| \int \exp(itn^{-1/2}xy + itn^{-1/2}\alpha_n x^2)dF_n^*(x) \right| dK_n^*(y)$.

As t varies in I_2, $(tn^{-1/2}, tn^{-1/2}\alpha_n)$ varies in a compact set bounded away from zero. Let D denote such a set in $\mathbb{R}^2$.

$$\delta_{nm}^* \le \sup_{(\alpha_1,\alpha_2)\in D} \int \left| \int \exp(i\alpha_1 xy + i\alpha_2 x^2)dF_n^*(x) \right| dK_n^*(y).$$

Let $b_1, b_2 > 0$ (to be chosen). Then

$$\delta_{nm}^* \le K_n^*(b_1 \le |Y| \le b_2)I_{1n} + K_n^*(|Y| < b_1) + K_n^*(|Y| > b_2)$$

where $I_{1n} \le \sup_{b_1 \le |y| \le b_2} \sup_{(\alpha_1,\alpha_2)\in D} \left| \int \exp(i\alpha_1 xy + i\alpha_2 x^2)dF_n^*(x) \right|$.

Note that by Lemma 3.3.1, $K_n^* \Rightarrow K$ a.s. where K is the distribution function of $Y_{j-1} + \alpha Y_{j+2} + \varepsilon_{j+1} + \alpha^2 \varepsilon_{j-1}$, which is non-degenerate. Thus b_1 and b_2 can be chosen such that for large n,

$$K_n^*(|Y| < b_1) + K_n^*(|Y| > b_2) < \alpha_0 < 1.$$

Note that $F_n^* \Rightarrow F_0$ a.s. and we have Cramer's condition for $(\varepsilon_1, \varepsilon_1^2) \sim F_0$. Using the fact that the convergence of a sequence of characteristic functions to its limit is uniform over compact sets, we have $I_{1n} < 1 - \delta < 1$

for large n. Thus $\delta^*_{nm} \leq (1-\delta) + \delta\alpha_o < 1$.

<u>Lemma 3.3.4</u> : Under (A1) and (A2), for sufficiently small C_1, we have for almost every sequence $Y_o, Y_1, \ldots$.

$$\int_{t \in I_1} |D^{\alpha} H_n^*(t)| dt = o(n^{-(s-2)/2}).$$

<u>Proof</u> : As in Lemma 3.3.3 it is sufficient to deal with the original variables instead of truncations. As before we proceed as in Lemma 3.3.3 following GH but use a different estimate for $E^*|E^* A_p^*/\varepsilon_j^*, j \neq j_p|$. We have to deal with

$$\delta^*_{nm} = E^*|E^* \exp(itn^{-1/2}(\varepsilon_n^* A_n^* + \alpha_n \varepsilon_n^{*2}))/\mathcal{D}_j^*, j \neq n|$$

where $A_n^* = Y_{n-1}^* + \alpha_n Y_{n+2}^* + \varepsilon_{n+1}^* + \alpha_n^2 \varepsilon_{n-1}^*$.

$$\delta^*_{nm} \leq P^*(|A_n^*| \geq b) + E^*|E^* \exp(itn^{-1/2}(\varepsilon_n^* A_n^* + \alpha_n \varepsilon_n^{*2}))/\mathcal{D}_j^*, j \neq n| I(|A_n^*| \leq b).$$

For large b, the first term is small.

In the second term, the inner expectation equals

$$1 - \frac{t_n'}{2n} D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n + \frac{\gamma}{6} \frac{\| t_n^* \|^3}{n^{3/2}} E^* \| \varepsilon_n^*, \varepsilon_n^{*2} \|^3$$

where $t_n' = (tA_n^*, t\alpha_n)$, $|\gamma| \leq 1$.

The last term is bounded by

$$\frac{C}{6} b^3 \frac{|t|^3}{n^{3/2}} \mu_{3n}^* \quad \text{where} \quad \mu_{3n}^* = E^* \| \varepsilon_n^*, \varepsilon_n^{*2} \|^3 .$$

$$\leq \beta \frac{b^3 C_1 |t|^2}{6n} \text{ a.s., since } \mu_{3n}^* \xrightarrow{\text{a.s.}} E \| \varepsilon_1, \varepsilon_1^2 \|^3 \quad \text{by Lemma 3.3.1.}$$

By taking C_1 sufficiently small the above term is dominated by $\alpha \frac{|t|^2}{n}$ where α is as small as we please.

Let $\lambda_{1n}(\lambda_1)$ be the smallest eigen values of $D(\varepsilon_n^*, \varepsilon_n^{*2})(D(\varepsilon_1, \varepsilon_1^2))$.
By Lemma 3.3.1, $D(\varepsilon_n^*, \varepsilon_n^{*2}) \xrightarrow{a.s.} D(\varepsilon_1, \varepsilon_1^2)$ which is positive definite.

$$\text{On } \left\{ |A_n^*| \leq b \right\}, \; \left|\frac{t_n'}{2n} D(\varepsilon_n^*, \varepsilon_n^{*2})\right| \leq C \, \| D(\varepsilon_n^*, \varepsilon_n^{*2}) \| \frac{b^2}{2n} |t|^2$$

$$\leq 1 \quad \text{if } C_1 \text{ is small.}$$

$$\text{For large } b, \; \left| E^* A_n^{*2} I(|A_n^*| \leq b) - E^* A_n^{*2} \right|$$

$$\leq \left[E^*(A_n^{*4}) P^*(|A_n^*| > b) \right]^{1/2}$$

$< \eta$, η sufficiently small by choosing b large enough.
(By Lemma 3.3.1, $E^*(A_n^{*4}) \xrightarrow{a.s.} E_{F_0}(A^4) < \infty$).

Thus the second term of δ_{nm}^* is bounded by

$$\delta_{2n} = \frac{\alpha |t|^2}{n} + E^*(1 - \frac{1}{2n} t_n' D(\varepsilon_n^*, \varepsilon_n^{*2}) t_n) I(|A_n^*| \leq b)$$

$$\leq \alpha \frac{|t|^2}{n} + E^*(1 - \frac{1}{2n} I(|A_n^*| \leq b) \lambda_{1n} |t_n|^2)$$

$$\leq \alpha \frac{|t|^2}{n} + 1 - \frac{\gamma |t|^2}{n} \quad \text{where } \gamma > \alpha$$

$$\leq \exp(-\frac{\delta |t|^2}{n}), \; \delta > 0.$$

This shows that $\delta_{nm}^* \leq \rho + \exp(-\frac{\delta |t|^2}{n})$ where $0 < \rho < \frac{1}{2}$. A look at the proof of Lemma 3.43 of GH shows that this proves the lemma.

We introduce a few more notations. For a real valued measurable function f on $\mathbb{R}$.

Let
$$M_r(f) = \sup_x (1+|x|)^{-r}|f(x)|$$
$$\bar{\omega}(f,\varepsilon,\sigma^2) = \int \omega(f,\varepsilon,x)\varphi_{\sigma^2}(x)dx$$

where $\omega(f,\varepsilon,x) = \sup_{|y-x|\leq\varepsilon} |f(y)-f(x)|$ and φ_{σ^2} is the normal density with mean 0 and variance σ^2.

$$S_n = n^{-1/2} \sum_{k=1}^{n} (Y_k Y_{k-1} - \alpha) = n^{1/2}(\alpha_n - \alpha)$$

$$S_n^* = n^{-1/2} \sum_{k=1}^{n} (Y_k^* Y_{k-1}^* - \alpha_n) = n^{1/2}(\alpha_n^* - \alpha_n)$$

$$s_o = \begin{cases} s & \text{if } s \text{ is even} \\ s-1 & \text{if } s \text{ is odd.} \end{cases}$$

The following results follow from developments of Chapter 1.

<u>Theorem 3.3.5</u> : Let $f : \mathbb{R} \to \mathbb{R}$ be measurable such that $M_{s_o}(f) < \infty$. Assume (A1) and (A2). Then there exists a positive δ not depending on f and for arbitrary $k > 0$ there exists a positive constant depending on f only through M_{s_o} such that

$$\left| Ef(S_n) - \int f d\psi_{n,s} \right| \leq c\bar{\omega}(f,n^{-k},\sigma^2) + o(n^{-(s-2+\delta)/2}).$$

The term $o(\cdot)$ depends on f through M_{s_o} only and σ^2 = limiting variance of S_n. $\psi_{n,s}$ is the usual function associated with Edgeworth expansions.

<u>Corollary 3.3.6</u> : Under (A1) and (A2), uniformly for convex measurable $C \subseteq \mathbb{R}$,

$$P(S_n \in C) = \psi_{n,s}(C) + o(n^{-(s-2)/2}).$$

We are now in a position to study the accuracy of bootstrap approximation.

Theorem 3.3.7 : Assume (A1), (A2) and $|\alpha| < 1$. Suppose $f : \mathbb{R} \to \mathbb{R}$ is such that $M_{s_0}(f) < \infty$. Let $\sigma_n^{*2} = E^*(S_n^{*2})$. For a.e. $Y_0, Y_1, \ldots$ and uniformly over $x \in \mathbb{R}$, almost surely,

(a) $\quad \left| E^* f(S_n^*) - \int f d\psi_{n,3}^* \right| \leq c.\bar{\omega}(f, n^{-k}, \sigma_n^{*2}) + o(n^{-1/2}).$

(b) $\quad P^*(\sigma_n^{*-1} S_n^* \leq x) = \psi_{n,3}^*(\sigma_n^* x) + o(n^{-1/2}).$

(c) $\quad P^*(\sigma_n^{*-1} S_n^* \leq x) = \int_{-\infty}^{x} d\psi_{n,3}^*(\sigma_n^* y) + o(n^{-1/2}) = P(\sigma^{-1} S_n \leq x) + o(n^{-1/2}).$

Proof : (a) follows from Lemmas 3.3.2 - 3.3.4 and Lemma 2.3.4 of Chapter 2. (b) is a consequence of (a). (c) follows by combining (b) and Corollary 3.3.6 and noting that the difference of moments involved in $\psi_{n,3}$ and $\psi_{n,3}^*$ converge to zero by Lemma 3.3.1. To have a more detailed idea of the proof, the reader may refer to the proof of an analogous result in the i.i.d. case by Babu and Singh (1984).

For $p > 1$, the recovery of errors is slightly more complicated and is done as follows.

Under our conditions,

$$\varepsilon_t = \sum_{r=0}^{\infty} c_r Y_{t-r}$$

where $c_0 = 1$, $c_1 = -\alpha_1$, $c_2 = -\alpha_1 c_1 - \alpha_2, \ldots, c_{q-1} = -\alpha_1 c_{q-2} - \alpha_2 c_{q-3} - \cdots,$

$$c_j = -(\alpha_1 c_{j-1} + \cdots + \alpha_q c_{j-q}),\ j = q, q+1, \ldots .$$

Thus, the pseudo errors are defined as

$$\hat{\varepsilon}_{in} = \sum_{r=0}^{i-1} \hat{c}_r Y_{i-r}, \quad i = 1,\dots,n, \quad \text{where } \hat{c}_r\text{'s are estimates}$$

of c_r's, obtained by replacing α_i's by their moment estimates. Lemma 3.3.1 - 3.3.4 can be extended to cover this situation.

The following theorem can now be easily proved by extending previous arguments.

<u>Theorem 3.3.8</u> : Let H be a function from $\mathbb{R}^p \longrightarrow \mathbb{R}$ which is thrice continuously differentiable in a neighbourhood of 0. Let λ denote the vector of first order partial derivatives of H at 0. Assume $\lambda \neq 0$.

Let $T(F) = n^{1/2}\left[H(n^{-1}\sum_{k=1}^{n}(Y_kY_{k-i}-\beta_i),\ i = 1,\dots,p)-H(0)\right],\ \sigma^2 = \lambda'\Sigma\lambda$

$$T(F_n^*) = n^{1/2}\left[H(n^{-1}\sum_{k=1}^{n}(Y_k^*Y_{k-1}^*-\beta_{in}),\ i = 1,\dots,p))-H(0)\right],\ \sigma_n^{*2} = \lambda'\Sigma_n^*\lambda$$

where $\Sigma = \lim D(n^{-1/2}\sum_{k=1}^{n} Y_kY_{k-i},\ i = 1,\dots,p),\quad \beta_i = EY_kY_{k-i}$

$$\Sigma_n^* = D^*(n^{-1/2}\sum_{k=1}^{n} Y_k^*Y_{k-i}^*,\ i = 1,\dots,p),\quad \beta_{in} = n^{-1}\sum_{t=1}^{n} Y_tY_{t-i}.$$

Then $\sup_x |P(\sigma^{-1}T(F) \leq x) - P^*(\sigma_n^{*-1}T(F_n^*) \leq x)| = o(n^{-1/2})$. a.s.

<u>Proof</u> : Note that Theorems 3.3.5 and 3.3.7(a) remain valid for

$(n^{-1/2}\sum_{k=1}^{n}(Y_kY_{k-i}-\beta_i),\ i=1,\dots,p)$ and $(n^{-1/2}\sum_{t=1}^{n}(Y_t^*Y_{t-i}^*-\beta_{in}),\ i=1,\dots,p)$.

Thus arguments analogous to Theorem 3 and Corollary 2 of BS yields the theorem. The arguments are exactly as those in BS, so we omit the details.

Remark 3.3.9.

The above result, with proper modifications is true with vector-valued H. This is because Theorem 3 and Corollary 2 of BS are true for such functions. The estimates of $\alpha_1,\ldots,\alpha_p$ are smooth functions of $n^{-1}\sum_{j=1}^{n} Y_j Y_{j-i}$. Hence, once we have obtained Theorem 3.3.8, we should be able to utilize it to prove results for the parameter estimates too.

We have the following theorem on accuracy of bootstrap approximation.

Theorem 3.3.10 : Let ψ_n denote the distribution function of $\Sigma^{-1/2}\, n^{1/2}(\alpha_{1n}-\alpha_1,\ldots,\alpha_{pn}-\alpha_p)$, where Σ is the limiting variance-covariance matrix of $n^{1/2}(\alpha_{1n}-\alpha_1,\ldots,\alpha_{pn}-\alpha_p)$ and ψ_n^* denote the corresponding bootstrapped distribution function. Then under assumptions (A1), (A2) and the invertibility condition, for a.e. $Y_0, Y_1, \ldots,$

$$\sup_{x\in\mathbb{R}^p} |\psi_n(x)-\psi_n^*(x)| = o(n^{-1/2}) \quad \text{a.s.}$$

Proof : The case $p = 1$ is Theorem 3.3.8.

For $p = 2$, the moment equations are

$$\alpha_{2n} = n^{-1}\sum_{t=1}^{n} Y_t Y_{t-2}$$

$$\alpha_{1n}(1+\alpha_{2n}) = n^{-1}\sum_{t=1}^{n} Y_t Y_{t-1}\,.$$

Thus $\alpha_{2n} - \alpha_2 = n^{-1} \sum_{t=1}^{n} (Y_t Y_{t-2} - \beta_2) = \bar{Z}_{2n}$ say

$$\alpha_{1n}(1+\alpha_{2n}) - \alpha_1(1+\alpha_2) = n^{-1} \sum_{t=1}^{n} (Y_t Y_{t-1} - \beta_1) = \bar{Z}_{1n} \quad \text{say}$$

Hence $\alpha_{1n} = \left[\bar{Z}_{1n} + \alpha_1(1+\alpha_2) \right] / (1+\alpha_2+\bar{Z}_{2n})$

or $$\alpha_{1n} - \alpha_1 = \frac{\bar{Z}_{1n} + \alpha_1(1+\alpha_2)}{(1+\alpha_2+\bar{Z}_{2n})} - \alpha_1 .$$

Thus $$(\alpha_{1n} - \alpha_1, \alpha_{2n} - \alpha_2) = \left(\frac{\bar{Z}_{1n} + \alpha_1(1+\alpha_2)}{1+\alpha_2+\bar{Z}_{2n}} - \alpha_1, \bar{Z}_{2n} \right) .$$

Now the result follows from the multidimensional version of Theorem 3.3.8.

Remark 3.3.11.

The idea of proof for general p is clear from what we have shown. However, solving for the estimates $\alpha_{1n}, \ldots, \alpha_{pn}$ becomes increasingly difficult with increase in p.

Remark 3.3.12.

As in Chapter 2, the assumptions $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$ can be weakened. We illustrate how this can be achieved, for the case $p = 1$.

The model is $Y_t = \mu + \varepsilon_t + \alpha\varepsilon_{t-1}$ where (ε_i) satisfies (A1), (A2) but $E\varepsilon_t^2 = \sigma^2$ (unknown), $|\alpha| < 1$.

The estimates $\mu_n, \alpha_n, \sigma_n^2$ of μ, α and σ^2 are obtained by solving the equations,

$$n^{-1} \sum_{t=1}^{n} Y_t = \mu_n$$

$$n^{-1} \sum_{t=1}^{n} Y_t Y_{t-1} = \mu_n^2 + \alpha_n \sigma_n^2$$

$$n^{-1} \sum_{t=1}^{n} Y_t^2 = \mu_n^2 + \sigma_n^2(1 + \alpha_n^2).$$

Thus the key quantities here are

$$n^{-1/2}\left(\sum_{t=1}^{n} (Y_t - \beta_1),\ \sum_{t=1}^{n} (Y_t Y_{t-1} - \beta_2),\ \sum_{t=1}^{n} (Y_t^2 - \beta_3)\right)$$

where $\beta_1 = EY_t$, $\beta_2 = EY_t Y_{t-1}$ and $\beta_3 = EY_t^2$.

Edgeworth expansions for this vector and its bootstrapped version can be established by modifying arguments already given for the case $\mu = 0$, $\sigma^2 = 1$. μ_n, α_n and σ_n^2 are smooth functions of these. Hence result analogous to Theorem 3.3.10 can be established for these estimates.

Remark 3.3.13.

Note that Theorem 3.3.10 is specially relevant in testing problems since we have shown that approximate tests can be carried out by the bootstrapped distribution. They beat the tests based on normal approximation.

Remark 3.3.14.

The models we have dealt in Chapter 2 and the present chapter have a linear structure. It would be interesting to derive analogous results in more complicated models. This problem at the moment seems quite hard.

PART II

CHAPTER 4

ASYMPTOTIC THEORY OF MAXIMUM LIKELIHOOD ESTIMATION IN DIFFUSION PROCESSES

4.1 Introduction

In this and the next two chapters, we move over to continuous time models. To be more specific, we consider a diffusion model of the form,

$$dX(t) = f(\theta, X(t))dt + dW(t), \ t \geq 0 \qquad \ldots(4.1.1)$$

The problem is to estimate θ and study the asymptotic properties of the estimator.

The situation where $f(\theta,x) = \sum_{i=1}^{k} a_i(\theta)b_i(x)$ is known as the linear case. This case has been dealt by several authors, e.g. Taraskin (1971,1974), LeBreton (1977), Brown and Hewitt (1975), Kulinic (1975), and Lee and Kozin (1977). Work has concentrated on showing that the m.l.e. is strongly/weakly consistent, asymptotically normal and efficient. These authors assume the existence of a stationary ergodic distribution.

The m.l.e. in the non-linear case has been considered by Kutoyants (1977), Prakasa Rao and Rubin (1981) and others. Here too, the main assumption has been the existence of stationary, ergodic distribution. Prakasa Rao and Rubin (1981) proved the strong consistency and asymptotic normality of the m.l.e. when the parameter space is the interval $[-1,1]$. They use Fourier analytic methods, which are of independent interest.

These methods can be generalized to cover the multiparameter case but the proofs will become quite lengthy. The approach of Kutoyants (1977) is slightly different and imposes slightly different conditions.

Borkar and Bagchi (1981) deal with the characterization of the limit set of the m.l.e. dropping the assumption of stationarity and ergodicity.

Lanska (1979) deals with minimum contrast estimation, which uses a differentiation under the stochastic integral. However, no justification is provided.

Other works dealing with the above parametric model are discussed in Chapters 5 and 6.

For works in related problems, e.g. non-parametric estimation of f, maximum probability estimation and the practically important least squares estimation when $X(t)$ is observed at finitely many points, we refer the reader to the article by Prakasa Rao (1985a).

We treat the non-linear case, with the parameter space being the unit ball of $\mathbb{R}^d$. We show that under certain conditions, the m.l.e. is strongly consistent and asymptotically normal. We do not assume stationarity or ergodicity. However, in practice, the conditions we impose are easier to verify under the above assumptions.

The main technique is to use Kolmogorov type inequalities from the theory of diffusion processes. These are used to get probabilistic bounds for supremum of certain processes, and, are of independent interest.

This chapter is a revised form of Bose (1983a).

4.2. Notations and Assumptions

Let $f(\theta,x)$ be a real-valued function on $\Omega \times \mathbb{R}$ where $\Omega = \{\theta \in \mathbb{R}^d : |\theta| \leq 1\}$, $d \geq 1$ (finite) and $\theta_0 \in \Omega^0$ (the interior of Ω) is the unknown true value. $\mathbb{R}$ denotes the set of real numbers. The observation $X(t)$, $t \geq 0$ evolves as the unique strong solution of the stochastic differential equation

$$dX(t) = f(\theta_0, X(t))dt + dW(t), \; t \geq 0 \tag{4.2.1}$$

$X(0) = X$ (given) with $E(X^2) < \infty$. Here $(W(t), t \geq 0)$ is a standard Wiener process.

Sufficient conditions for the existence of unique strong solutions of stochastic differential equations can be found in McKean (1969) or Friedman (1975). See Remark 4.2.1.

We introduce the following assumptions.

A1 (i) $f(\theta,x)$ is continuous on $\Omega \times \mathbb{R}$.

(ii) $\sup_\theta |f(\theta,x)| \leq J_1(x) \; \forall x \in \mathbb{R}$.

(iii) $|f(\theta,x) - f(\varphi,x)| \leq J_2(x)|\theta - \varphi|^{\beta_1}$, $0 < \beta_1 \leq 1$.

(iv) $\sup_\theta |f(\theta,x) - f(\theta,y)| \leq L|x-y|^{\alpha_1}$, $0 \leq \alpha_1 \leq 1$.

A2 (i) The partial derivatives $f_\theta^{(i)}$ of f w.r.t. θ_i (where $\theta' = (\theta_1,\ldots,\theta_d)$) exists $\forall i = 1,2,\ldots,d$.

Denote by $f_\theta^{(i)}(\theta^*,x)$ the derivative w.r.t. θ_i evaluated at θ^*.

(ii) $\sup_{\theta \in V_{\theta_0}} \sum_{i=1}^{d} | f_\theta^{(i)}(\theta,x) | \leq J_3(x)$ in a neighbourhood V_{θ_0} of θ_0.

(iii) $| f_\theta^{(i)}(\theta,x) - f_\theta^{(i)}(\varphi,x) | \leq J_4(x) | \theta - \varphi |^{\beta_2}$, $0 \leq \beta_2 \leq 1$.

A3 (i) There exists $\alpha_0 > 0$ with $\beta_1(d+\alpha_0) > d$ and

$$\sup_{T \geq 1} T^{-1} \int_0^T EJ_2^{d+\alpha_0}(X(t)dt < J_2 < \infty .$$

(ii) $\limsup_{T \to \infty} T^{-1} \int_0^T J_1(X(t))J_2(X(t))dt \leq M < \infty$ a.s.

(iii) There exists $\alpha_0 > 0$ with $\beta_2(d+\alpha_0) > d$ and

$$\sup_{T \geq 1} T^{-1} \int_0^T EJ_4^{d+\alpha_0}(X(t))dt < \infty .$$

(iv) $\limsup_{T \to \infty} T^{-1} \int_0^T J_3(X(t))J_4(X(t))dt \leq M < \infty$ a.s.

A4 (i) $\lim_{T \to \infty} T^{-1} I_T(\theta) = I(\theta)$ a.s. and $I(\theta) > 0 \ \forall \ \theta \neq \theta_0$. (For definition of $I_T(\theta)$, see below).

(ii) $\lim_{T \to \infty} T^{-1} \int_0^T (\nabla f_\theta(\theta,X(t))(\nabla f_\theta(\theta,X(t))' dt = J(\theta)$ a.s. and $J(\theta)$ is positive definite.

<u>Remark 4.2.1</u>

Strong solution of (4.2.1) exists if A1(ii) is satisfied with $J_1(x) = c(1+|x|)$ and A1(iv) with $\alpha_1 = 1$.

For any $T \geq 0$, $\mathcal{C}[0,T]$ will denote the space of all continuous functions on $[0,T]$ equipped with the supremum norm topology.

Under condition A1(i), the solution $X(t)$ is a continuous function of t. Let μ_θ^T denote the measure generated by $X(t)$ on $\mathcal{C}[0,T]$ when θ is the true value.

Sufficient conditions for the mutual absolute continuity of $\mu_\theta^T, \theta \in \Omega$ can be obtained directly from results of the theory of diffusion processes. See e.g. Liptser and Shiryayev (1977). In particular, $\mu_\theta^T << \mu_{\theta_0}^T$ under assumption A1(i), and the Radon-Nikodym derivative (likelihood function) is given by

$$\frac{d\mu_\theta^T}{d\mu_{\theta_0}^T}(X(t) ; 0 \leq t \leq T) = \exp\left(\int_0^T v(\theta,X(s))dW(s) - \frac{1}{2}\int_0^T v^2(\theta,X(s))ds\right)$$

$$= \exp(Z_T^*(\theta) - I_T(\theta)/2)$$

where $v(\theta,x) = f(\theta,x) - f(\theta_0,x)$

$$Z_T^*(\theta) = \int_0^T v(\theta,X(s))dW(s)$$

$$I_T(\theta) = \int_0^T v^2(\theta,X(s))ds$$

Define $R_T(\theta) = I_T(\theta) - 2Z_T^*(\theta)$.

4.3 Auxiliary Results

Before we state and prove the main results, we shall derive some auxiliary results. These will serve as important tools in subsequent developments. They have found applications in other areas too. See Remark 4.5.4(3).

The following lemma is stated and proved in Stroock (1982, page 7).

<u>Lemma 4.3.1</u> : Let p and ψ be strictly increasing continuous functions on $[0,\infty]$ such that $p(0) = \psi(0) = 0$ and $\psi(\infty) = \infty$. Also suppose that L is a normed linear space and that $f : \mathbb{R}^d \to L$ is strongly continuous on $B(a,\rho) = \{x \in \mathbb{R}^d : |x-a| < \rho\}$. Then

$$\int_{B(a,\rho)} \int_{B(a,\rho)} \psi\left[\frac{\|f(x)-f(y)\|}{p(|x-y|)}\right] dxdy \leq B \text{ implies that}$$

$$\|f(x)-f(y)\| \leq 8 \int_0^{|x-y|} \psi^{-1}\left[\frac{4^{d+2}B}{\beta^2 u^{2d}}\right] p(du), x,y \in B(a,\rho)$$

where $\beta = \inf\limits_{x \in B(a,\rho)} \inf\limits_{1<\rho^* \leq 2} \dfrac{|B(x,\rho^*) \cap B(a,1)|}{\rho^d}$

and $|A|$ = Lebesgue measure of A.

The following corollary is immediate.

<u>Corollary 4.3.2</u> : In particular if $\psi(x) = x^r$, $p(x) = x^{\delta/r}$ $r > 0$, $\delta > 2d$ in Lemma 4.3.1, then

$$\|f(x)-f(y)\| \leq c(r,\delta,d)\, |x-y|^{(\delta-2d)/r}\, B^{1/r}.$$

<u>Lemma 4.3.3</u> : Suppose $\{Y(\theta) : \theta \in \mathbb{R}^d\}$ is a class of random variables taking values in a normed linear space L. Let $\alpha > 0$.

Suppose a) $\forall\, w$, $\theta \to Y(\theta,w)$ is continuous on $\overline{B(a,\rho)}$

b) $E\,\|Y(\theta) - Y(\varphi)\|^r \leq C\,|\theta - \varphi|^{d+\alpha} \quad \forall\, \theta,\varphi \in \overline{B(a,\rho)}$.

Then $\forall\, \delta \in (2d, 2d+\alpha)$ and $\lambda > 0$, with C as above,

$$P\left(\sup_{\theta,\varphi \in B(a,\rho)} \frac{\|Y(\theta) - Y(\varphi)\|}{|\theta-\varphi|^{\beta}} \geq c(r,\delta,d)\lambda^{1/r}\right) \leq \frac{CA}{\lambda}$$

where $\beta = (\delta - 2d)/r$, $A = \int_{B(a,\rho)}\int_{B(a,\rho)} |\theta-\varphi|^{d+\alpha-\delta}\, d\theta d\varphi$

Proof : By (b),

$$E\left[\int_{B(a,\rho)}\int_{B(a,\rho)} \left(\frac{\|Y(\theta)-Y(\varphi)\|}{|\theta-\varphi|^{\delta/r}}\right)^r d\theta d\varphi\right] \leq CA.$$

Hence $P\left(\int_{B(a,\rho)}\int_{B(a,\rho)} \left(\frac{\|Y(\theta)-Y(\varphi)\|}{|\theta-\varphi|^{\delta/r}}\right)^r d\theta d\varphi \geq \lambda\right) \leq CA/\lambda$

and whenever $\int_{B(a,\rho)}\int_{B(a,\rho)} \left(\frac{\|Y(\theta)-Y(\varphi)\|}{|\theta-\varphi|^{\delta/r}}\right)^r d\theta d\varphi \leq \lambda$ we have by Corollary 4.3.2,

$$\|Y(\theta) - Y(\varphi)\| \leq c(r,\delta,d)\,|\theta-\varphi|^{\beta}\,\lambda^{1/r}, \quad \forall\, \theta,\varphi \in B(a,\rho).$$

This proves the lemma.

The following corollary is immediate.

Corollary 4.3.4 : Suppose in additions to conditions of Lemma 4.3.3, there exists a $\theta_0 \in B(a,\rho)$ such that $Y(\theta_0) = 0$. Then

$$P\left(\sup_{\theta \in \overline{B(a,\rho)}} \|Y(\theta)\| \geq (2\rho)^{\beta} c(r,\delta,d)\lambda^{1/r}\right) \leq CA/\lambda.$$

Lemma 4.3.5 : Let g be a function on $\Omega \times \mathbb{R}$ such that

$$|g(\theta,x) - g(\varphi,x)| \leq J(x)\,|\theta-\varphi|^{\beta}, \ \forall\, x \in \mathbb{R}, \ \forall\, \theta, \varphi \in \Omega.$$

Suppose $Y_T(\theta) = \int_0^T g(\theta, X(t))dW(t)$, $T \geq 0$, is well defined as a stochastic integral. Let $\alpha_0 \geq 0$ and $d+\alpha_0 \geq 2$. Then

$$E\left(\sup_{0 \leq t \leq T} |Y_t(\theta) - Y_t(\varphi)|^{d+\alpha_0}\right) \leq c_{d+\alpha_0} |\theta - \varphi|^{\beta_1(d+\alpha_0)} T^{\frac{d+\alpha_0}{2} - 1} \int_0^T E(J^{d+\alpha_0}(X(t))dt$$

Proof : Let $\Delta_T = E\left(\sup_{0 \leq t \leq T} |Y_t(\theta) - Y_t(\varphi)|^{d+\alpha_0}\right)$. By applying Burkholder's inequality (see Stroock and Varadhan (1979) page 116) and Holder's inequality in succession,

$$\begin{aligned} \Delta_T &\leq c_{d+\alpha_0} E\left(\int_0^T |g(\theta, X(t)) - g(\varphi, X(t))|^2 dt\right)^{\frac{d+\alpha_0}{2}} \\ &\leq c_{d+\alpha_0} T^{\frac{d+\alpha_0}{2} - 1} \int_0^T E(J^{d+\alpha_0}(X(t))dt. \end{aligned}$$

This proves the lemma.

Finally, we state the following lemma on the existence of continuous version of a process. For a proof see Stroock (1982, page 9).

Lemma 4.3.6 : Let $X(\theta)$, $\theta \in \mathbb{R}^d$, be a family of Banach space valued random variables with the properties that for some $\alpha > 0$ and $p \geq d + \alpha$,

$$E\left[\|X(\theta) - X(\varphi)\|^p\right] \leq c\,|\theta - \varphi|^{d+\alpha}, \ \theta, \varphi \in \mathbb{R}^d.$$

Then there is a family of random variables $\tilde{X}(\theta)$, $\theta \in \mathbb{R}^d$ such that $\tilde{X}(\theta) = X(\theta)$ a.s. for each $\theta \in \mathbb{R}^d$ and $\theta \to \tilde{X}(\theta)$ is a.s. strongly continuous.

The next section is devoted to proving strong consistency of the m.l.e.

4.4 Strong consistency of the m.l.e.

Let θ_T denote the m.l.e. when $X(t)$ is observed over the time period $0 \leq t \leq T$. We shall assume the existence and measurability of θ_T. For sufficient conditions and questions regarding these see Lanska (1979).

We first prove two lemmas.

Lemma 4.4.1 : Assume A1 (iii) and A3 (i) hold. Then

(1) $$P\Big(\sup_{\theta} \sup_{0 \leq t \leq T} \| Z_t^*(\theta) \| \geq c_1 \lambda^{1/(d+\alpha_0)}\Big) \leq c_2 \lambda^{-1} T^{(d+\alpha_0)/2}$$

(2) For any $\delta > 1/(d+\alpha_0)$, there exists $H > 0$ such that

$$\limsup_{T \to \infty} \sup_{\theta} \frac{|Z_T^*(\theta)|}{T^{1/2}(\log T)^{\delta}} \leq H \quad \text{a.s.}$$

Proof : (1) Take $g(\theta,x) = v(\theta,x)$ in Lemma 4.3.5 to get

$$E\Big(\sup_{0 \leq t \leq T} |Z_t^*(\theta) - Z_t^*(\varphi)|^{d+\alpha_0}\Big) \leq c_{d+\alpha_0} J_2 T^{(d+\alpha_0)/2} |\theta - \varphi|^{\beta_1(d+\alpha_0)}.$$

By Lemma 4.3.6, there is a continuous version of $\theta \to Z_t^*(\theta)$, $0 \leq t \leq T$. The conditions of Corollary 4.3.4 are valid. These together yield the result.

(2) Prakasa Rao and Rubin (1981) have proved this in the situation where $d = 1$ and $X(t)$ has a stationary ergodic initial distribution. We follow their proof.

Define $A_n = \left\{ \sup_{2^{n-1} \le t \le 2^n} \sup_{\theta} |Z_t^*(\theta)| \ge H' 2^{n/2} n^{\delta} \right\}$, $n \ge 1$.

By (1), $P(A_n) \le c_2' \dfrac{(2^n)^{(d+\alpha_0)/2}}{(2^{n/2} n^{\delta})^{d+\alpha_0}}$.

By the choice of δ, $\sum_{n=1}^{\infty} P(A_n) < \infty$. An application of the Borel-Cantelli lemma completes the proof.

Lemma 4.4.2 : Assume that A1 (ii),(iii), A3 (ii) and A4 (i) hold.

Then $\inf_{|\theta - \theta_0| \ge \delta} \dfrac{I_T(\theta)}{T} \xrightarrow{\text{a.s.}} \lambda(\delta)$ as $T \to \infty$ for some $\lambda(\delta) > 0$.

Proof : *The proof for the stationary ergodic case and* $d = 1$ is given by Prakasa Rao and Rubin (1981). We provide simple modification of their proof.

$$I_T(\theta) - I_T(\varphi) = \int_0^T (f(\theta,X(t)) - f(\varphi,X(t))) \left\{ f(\varphi,X(t)) + f(\theta,X(t)) - 2f(\theta_0,X(t)) \right\} dt .$$

Hence $|I_T(\theta) - I_T(\varphi)| \le cT |\theta - \varphi|^{\beta_1}$ a.s.

Thus it follows that $\dfrac{I_T(\theta)}{T} \to I(\theta)$ uniformly in $\theta \in \Omega$ as $T \to \infty$.

But $I_T(\theta_0) = 0$ and $\lim_{T \to \infty} T^{-1} I_T(\theta) > 0$ a.s. for $\theta \ne \theta_0$.

This proves the lemma.

We are now in a position to prove strong consistency of m.l.e.

Theorem 4.4.3 : Assume conditions of Lemma 4.4.1 and 4.4.2.

Then $\theta_T \to \theta_0$ a.s. as $T \to \infty$.

Proof : The key results used in proving this theorem are Lemma 4.4.1(2) and Lemma 4.4.2. The arguments are easy and are available in Prakasa Rao and Rubin (1981). Hence we omit the details.

In the next section we prove asymptotic normality of θ_T under enough assumptions.

4.5 Asymptotic normality of the m.l.e.

We shall need two lemmas to prove the asymptotic normality of m.l.e. The first of these provides an approximation for $I_T(\theta)$.

$$\text{For any } \theta^* \in \Omega^0, \text{ let } \nabla f_\theta(\theta^*,x) = \begin{bmatrix} f_\theta^{(1)}(\theta^*,x) \\ \vdots \\ f_\theta^{(d)}(\theta^*,x) \end{bmatrix}$$

Lemma 4.5.1 : Assume A2 (i) - (iii) and A3 (iv). Let $\psi = T^{1/2}(\theta - \theta_0)$.

For any $A_T > 0$, $\sup_{|\psi| \leq A_T} \left| I_T(\theta) - T^{-1}\int_0^T (\psi'\nabla f_\theta(\theta_0,X(t)))^2 dt \right| \leq MA_T^{2+\beta_2} T^{-\beta_2/2}$ a.s.

Proof : $$I_T(\theta) = \int_0^T v^2(\theta,X(t))dt$$

$$= \int_0^T \left[(\theta-\theta_0)'\nabla f_\theta(\theta_0,X(t))\right]^2 dt$$

$$+ \int_0^T \left\{\left[(\theta-\theta_0)'\nabla f_\theta(\theta^*,X(t))\right]^2 - \left[(\theta-\theta_0)'\nabla f_\theta(\theta_0,X(t))\right]^2\right\}dt$$

where $|\theta^* - \theta_0| \leq |\theta - \theta_0|$.

The integrand in the second expression

$$= (\theta-\theta_0)'(\nabla f_\theta(\theta^*,x) + \nabla f_\theta(\theta_0,x))(\theta-\theta_0)'(\nabla f_\theta(\theta^*,x) - \nabla f_\theta(\theta_0,x))$$

Hence using A2 (ii) and (iii),

$$\left| I_T(\theta) - T^{-1} \int_0^T \left[\psi'\nabla f_\theta(\theta_0,X(t))\right]^2 dt \right| \leq c|\theta-\theta_0|^{2+\beta_2} \int_0^T J_3(X(t))J_4(X(t))dt.$$

The lemma follows immediately from the above expression.

Lemma 4.5.2 : Let $\psi = T^{1/2}(\theta-\theta_0)$ and

$$v_T(\psi,x) = f(\theta_0+\psi T^{-1/2},x) - f(\theta_0,x) - \psi' T^{-1/2}\nabla f_\theta(\theta_0,x).$$

Assume that A2 (iii) and A3 (iii) hold. Then for any $\delta \in (2d, 2d+\alpha_0)$,

and $$B_T = \int_{B(0,A_T)}\int_{B(0,A_T)} |\theta-\varphi|^{d+\alpha_0-\delta}\, d\theta d\varphi ,$$

$$P\left(\sup_{|\psi|\leq A_T} \left|\int_0^T v_T(\psi,X(t))dW(t)\right| \geq c_1(\delta)\lambda^{1/(d+\alpha_0)}\right) \leq \frac{cB_T}{\lambda}(A_T T^{-1/2})^{\beta_2(d+\alpha_0)}$$

Proof : First note that $\int_0^T v_T(\psi,X(t))dW(t)$ is well-defined.

$v_T(\psi,x) - v_T(\psi_1,x) = (\psi-\psi_1)'\nabla v_T(\psi^*,x)$ where ψ^* lies between ψ and ψ_1.

But $$v_T^{(i)}(\psi,x) = T^{-1/2}f_\theta^{(i)}(\theta_0+\psi T^{-1/2},x) - T^{-1/2}f_\theta^{(i)}(\theta_0,x)$$

which yields $|v_T^{(i)}(\psi,x)| \leq T^{-1/2}J_4(x)(A_T T^{-1/2})^{\beta_2}$.

Now the result follows exactly as in Lemma 4.4.1(1).

We are now in a position to state and give a quick proof of the asymptotic normality of m.l.e.

Theorem 4.5.3 : Assume A1 - A4. Then

$$T^{1/2}(\theta_T - \theta_0) \xrightarrow{\mathcal{L}} N(0, J^{-1}(\theta_0)).$$

Proof : Since $T^{-1} \int_0^T \nabla f_\theta(\theta_0, X(t))(\nabla f_\theta(\theta_0, X(t)))' dt \rightarrow J(\theta_0)$ a.s.,

we have by the central limit theorem for stochastic integrals due to Kutoyants (1975); (for a proof see Basawa and Prakasa Rao (1980, page 405)),

$$T^{-1/2} \int_0^T \nabla f_\theta(\theta_0, X(t)) dW(t) \xrightarrow{\mathcal{L}} N(0, J(\theta_0)) \quad \text{as} \quad T \rightarrow \infty.$$

Since θ_T is consistent by Theorem 4.4.3, $\theta_T \in V_{\theta_0}$ with probability one as $T \rightarrow \infty$. Choose $A_T = \log T$ in Lemma 4.5.1 and 4.5.2. We then have that the asymptotic distribution of $T^{1/2}(\theta_T - \theta_0)$ which minimizes $R_T(\theta)$ is same as $\hat{\psi}$ where $\hat{\psi}$ is that which minimizes $\psi' J(\theta_0) Z - 2\psi' Z$ and Z is normal with mean zero and variance-covariance matrix $J(\theta_0)$.

But $\hat{\psi} = J^{-1}(\theta_0) Z \sim N(0, J^{-1}(\theta_0))$.

Hence it follows that $T^{1/2}(\theta_T - \theta_0) \xrightarrow{\mathcal{L}} N(0, J^{-1}(\theta_0))$.

Remarks 4.5.4 (1) Prakasa Rao and Rubin (1981) prove the above results for the case $d = 1$, assuming stationarity and ergodicity of $X(t)$, $t \geq 0$. They use Fourier analytic methods; which can be extended to the case $d > 1$. However the proofs will become quite lengthy.

(2) The case of vector valued observations poses no extra difficulties except for complexities in notations.

(3) The convergence of the quantities $T^{-1} \int_0^T EJ_i^{d+\alpha_0}(X(t))dt$ etc. is most easily verified under stationarity and ergodicity. The general set up broadens the scope of the results.

(4) The techniques developed in Section 3 has uses elsewhere. It is specially useful in dependent models and has been exploited by Prakasa Rao (1985c).

However the drawback of the technique is the existence of sufficiently high order moments.

(5) The techniques of this chapter in conjunction with available limit theorems for martingales can be easily adapted to prove analogous results for the least squares estimates in discrete time models of the form

$X(n) = f(\theta, X(n-1)) + \varepsilon_n$, $n \geq 1$ where (ε_n) are i.i.d error variables and $X(0) = X$ (given).

CHAPTER 5

ASYMPTOTIC BEHAVIOUR OF POSTERIORS AND BAYES ESTIMATORS IN DIFFUSION PROCESSES

5.1 Introduction

A classical theorem for posteriors in the set up of i.i.d. observations, loosely stated, says the following :

Let (X_i) be i.i.d. observations from a parametric family of probabilities $\left\{P_\theta : \theta \in \Omega\right\}$. Let $\Lambda(\cdot)$ be a prior probability distribution on Ω. Then, under sufficient regularity conditions, the posterior distribution of θ (suitably normalized) given the observations (X_i), $i \leq n$ is asymptotically normal under a fixed θ_0.

This was proved in the general set up of i.i.d. observations by LeCam (1955, 1958), generalizing the results of Laplace, Bernstein, von-Mises, Kolmogorov who obtained them for particular cases. Borwanker, Kallianpur and Prakasa Rao (1971) have extended this result to Markov processes.

For diffusions with $f(\theta,t,x) = a(t,x) + \theta b(t,x)$, analogous result was proved by Prakasa Rao (1981) and then extended to diffusion fields in Prakasa Rao (1983).

In this chapter we show that this theorem holds for non-linear diffusions too. As a consequence of the main theorem, we obtain that the Bayes estimators for smooth loss functions and smooth priors, are asymptotically normal.

The basic line of argument is similar to Borwanker et.al. (1971). However, the intermediate results needed are obtained by applying techniques developed in Chapter 4. We also need a formula for interchanging the order of stochastic integration and ordinary differentiation due to Karandikar (1983).

This chapter is a revised form of Bose (1983b).

5.2 Notations and Assumptions

The model is as in the previous chapter. Here we introduce some more assumptions on f apart from those given in the previous chapter.

A5 (i) The partial derivatives $f_\theta^{(i,j)}(\theta,x) = \dfrac{\partial^2}{\partial\theta_i \partial\theta_j} f(\theta,x)$ exists $\forall\ i,j = 1,\ldots,d$ and are continuous.

(ii) $\sup\limits_\theta |f_\theta^{(i,j)}(\theta,x)| \leq J_5(x) \quad \forall\ i,j = 1,2,\ldots,d,\ x \in \mathbb{R}$.

(iii) $|f_\theta^{(i,j)}(\theta,x) - f_\theta^{(i,j)}(\varphi,x)| \leq J_6(x)\,|\theta-\varphi|^{\beta_3},\ 0 \leq \beta_3 \leq 1$.

To keep conformity with notations of the previous chapter introduce the following assumption.

A3 (v) There exists $\alpha_0 > 0$ with $\beta_3(d+\alpha_0) > d$ and

$$\sup_{T\geq 1} T^{-1} \int_0^T E\, J_6^{d+\alpha_0}(X(t))dt < J_6 < \infty .$$

(vi) $\sup\limits_{T\geq 1} T^{-1} E \int_0^T f''^2(\theta_0,X(t))dt < \infty$, f'' denotes second derivatives.

(vii) $\sup\limits_{T\geq 1} T^{-1} \int_0^T J_2(X(t))J_5(X(t))dt \leq M < \infty$ a.s.

In Chapter 4 we have seen that under assumptions A1 (ii), (iii), (iv), A3 (i), (ii) and A4 (i) the m.l.e. θ_T is strongly consistent.

Suppose now that Λ is a prior probability on $(\Omega, \underline{\underline{B}})$, where $\underline{\underline{B}}$ is the σ-algebra of Borel subsets of Ω. Assume that Λ has a density $\lambda(\cdot)$ w.r.t. the Lebesgue measure and the density is continuous and positive in an open neighbourhood of θ_0.

The posterior density of θ given $(X(s) : 0 \leq s \leq T)$ is

$$p(\theta/X(s) : 0 \leq s \leq T) = \frac{d\mu_\theta^T}{d\mu_{\theta_0}^T}(X(s) : 0 \leq s \leq T)/\alpha$$

where $\alpha = \int\limits_{\Omega} \frac{d\mu_\theta^T}{d\mu_{\theta_0}^T}(X(s) : 0 \leq s \leq T)\lambda(\theta)d\theta$

Let $t = T^{1/2}(\theta - \theta_T)$. Then the posterior density of $T^{1/2}(\theta - \theta_T)$ is

$$p^*(t/X(s) : 0 \leq s \leq T) = T^{-1/2}p(\theta_T + tT^{-1/2}/X(s) : 0 \leq s \leq T).$$

Let $\gamma_T(t) = \dfrac{d\mu^T_{\theta_T + tT^{-1/2}}}{d\mu^T_{\theta_0}}(X(s) : 0 \leq s \leq T) \Big/ \dfrac{d\mu^T_{\theta_T}}{d\mu^T_{\theta_0}}(X(s) : 0 \leq s \leq T)$

$$C_T = \int_{-\infty}^{\infty} \gamma_T(t)\lambda(\theta_T + tT^{-1/2})dt$$

$$V(\theta,x) = f''(\theta,x) - f''(\theta_0,x),$$

$$X(t,\theta) = \int_0^t V(\theta,X(s))dW(s).$$

For the rest of the chapter, we shall assume that $d = 1$. With proper modifications, every argument in this special case goes through for higher dimensions. We will continue writing d in general.

5.3 The main results

We first quote a result from Karandikar (1983) which serves as an important tool.

Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space, $(\mathfrak{F}_t)_{t \geq 0}$ be an increasing family of sub σ-fields of $\mathfrak{F}$ such that $\mathfrak{F}_0$ contains all the P-null sets. All the processes are $(\mathfrak{F}_t)$ adapted. Let (W_t) be an $(\mathfrak{F}_t)$ Brownian motion. Let $\mathcal{L}_2$ be the collection of all progressively measurable processes f on $[0,1]$ such that

$$E \int_0^1 f^2(t,\omega)dt < \infty .$$

Theorem 5.3.1 : Let $\{f(\theta,.,.) : \theta \in \mathbb{R}\} \subseteq \mathcal{L}_2$ be such that for all t,ω, $\frac{d}{d\theta} f(\theta,t,\omega) = f'(\theta,t,\omega)$ exists and $\{f'(\theta,.,.) : \theta \in \mathbb{R}\} \subseteq \mathcal{L}_2$. Further assume that the following condition holds.

There exists constants c,β_1,β_2 ; $0 < c < \infty$, $0 < \beta_1,\beta_2 \leq 1$ such that

$$|f(\theta_1,t,\omega) - f(\theta_2,t,\omega)| \leq c\,|\theta_1 - \theta_2|^{\beta_1}$$

$$|f'(\theta_1,t,\omega) - f'(\theta_2,t,\omega)| \leq c\,|\theta_1 - \theta_2|^{\beta_2} .$$

Then there exists a version $X(\theta,t,.)$ of $\int_0^t f(\theta,u,.)dW(u)$ such that for all ω,t; $\theta \to X(\theta,t,\omega)$ is differentiable in θ and

$$\frac{d}{d\theta} X(\theta,t,\cdot) = \int_0^t f'(\theta,u,\cdot)dW(u).$$

<u>Remark 5.3.2</u>

Under the assumptions A1,A2,A3 and A5, all the stochastic integrals occurring henceforth can be defined pathwise. See Karandikar (1981). Further, by a slight modification of the above result, it is possible to differentiate (w.r.t. θ_i's) within the stochastic integral, pathwise outside a fixed null set. This set is dropped out of consideration henceforth.

<u>Remark 5.3.3</u>

The m.l.e. θ_T satisfies the equation

$$\frac{\partial}{\partial\theta} \log L_T(\theta) \Big|_{\theta=\theta_T} = 0 .$$

Thus by Remark 5.3.2,

$$\left[\int_0^T f'(\theta,X(s))dW(s)\right]_{\theta=\theta_T} = \int_0^T f'(\theta_T,X(s)) \left[f(\theta_T,X(s)) - f(\theta_0,X(s))\right] ds \qquad \ldots(5.3.1)$$

Note that $\int_0^T f'(\theta_T,X(s))dW(s)$ is not defined as an Ito integral since $f'(\theta_T,X(s))$ is not non-anticipative. However, we will use this notation for the expression $\left[\int_0^T f'(\theta,X(s))dW(s)\right]_{\theta=\theta_T}$. The same comment holds for all other similar expressions appearing in the sequel and also for $Y_T(t)$ defined earlier.

We now state and prove a series of lemmas which will lead us to the main theorem.

Lemma 5.3.4 : Under the assumption A3 (vi),

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f''(\theta_0, X(s))\,dW(s) = 0 \quad \text{a.s.}$$

Proof : $g(t) = \int_0^t f''(\theta_0, X(s))\,dW(s)$ $t \geq 0$, is a martingale. Hence, by the martingale inequality, for every $T > 0$,

$$P\Big(\sup_{1 \leq t \leq T} |g(t)| > \lambda\Big) \leq \lambda^{-2} E\Big(\int_0^T f''(\theta_0, X(s))\,dW(s)\Big)^2$$

$$= \lambda^{-2} E \int_0^T f''^2(\theta_0, X(s))\,ds\,.$$

Let $A_n = \Big\{ \sup_{2^{n-1} \leq t \leq 2^n} |t^{-1} g(t)| \geq 2^{-n/4} \Big\}$

Then $P(A_n) \leq P\Big(\sup_{2^{n-1} \leq t \leq 2^n} |g(t)| \geq 2^{n-1} 2^{-n/4}\Big)$

$$\leq P\Big(\sup_{0 \leq t \leq 2^n} |g(t)| \geq 2^{n-1} 2^{-n/4}\Big)$$

$$\leq E \int_0^{2^n} f''^2(\theta_0, X(s))\,ds \Big/ (2^{n-1} 2^{-n/4})^2$$

Hence $\sum_{n=1}^{\infty} P(A_n) < \infty$, and the lemma follows from Borel Cantelli Lemma.

Lemma 5.3.5 : Under assumptions A3 (v),(vi) and A5 (iii),

(i) $P\Big(\sup_{\theta} \sup_{0 \leq t \leq T} |X(t,\theta)| \geq c_1 \lambda^{1/(d+\alpha_0)}\Big) \leq c_2 \lambda^{-1} T^{(d+\alpha_0)/2}$

where c_1, c_2 are positive constants independent of T.

(ii) $\forall\, \gamma > 1/(d+\alpha_0)$, there exists an H such that

$$\lim_{T \to \infty} \sup_{\theta} \frac{|X(T,\theta)|}{T^{1/2}(\log T)^{\gamma}} \leq H \quad \text{a.s.}$$

(iii) $$\lim_{T \to \infty} \sup_{\theta} \frac{|X(T,\theta)|}{T} = 0 \quad \text{a.s.}$$

(iv) $$\lim_{T \to \infty} \sup_{\theta} \frac{1}{T} \int_0^T f''(\theta, X(s))\,dW(s) = 0 \quad \text{a.s.}$$

Proof : Proofs of (i) and (ii) are exactly same as Lemma 4.4.1 (1),(2) of the previous chapter. (iii) is immediate from (ii). (iv) follows from (iii) and Lemma 5.3.4.

Lemma 5.3.6 : Assume the conditions (A1) - (A5) except A3 (iii). Then

(i) For each fixed t, $\lim_{T \to \infty} \log \gamma_T(t) = -\beta t^2/2$ a.s.

(ii) For every ε, $0 < \varepsilon < \beta$, there exists δ_0 and T_0 such that for $|t| \leq \delta_0 T^{1/2}$ and $T \geq T_0$,

$$\gamma_T(t) \leq \exp(-\tfrac{1}{2}t^2(\beta - \varepsilon)) \quad \text{a.s.}$$

(iii) For every $\delta > 0$, there exists a positive ε and T_0 such that for $T \geq T_0$,

$$\sup_{|t| > \delta T^{1/2}} \gamma_T(t) \leq \exp(-T\varepsilon/4) \quad \text{a.s.}$$ Here β denotes $J(\theta_0)$ (defined in Chapter 1), when $d = 1$.

Proof : $$\log \gamma_T(t) = \int_0^T \left[f(\theta_T + T^{-1/2}t, X(s)) - f(\theta_T, X(s))\right] dW(s)$$

$$-1/2 \int_0^T \left[f(\theta_T + T^{-1/2}t, X(s)) - f(\theta_o, X(s))\right]^2 ds$$

$$+ 1/2 \int_0^T \left[f(\theta_T, X(s)) - f(\theta_o, X(s))\right]^2 ds .$$

Applying mean-value theorem and then the likelihood equation (5.3.1), it easily follows that

$$\log \gamma_T(t) = I_1 + I_2 + I_3 + I_4$$

where $$I_1 = \frac{-t^2}{2T} \int_0^T f'^2(\theta_o, X(s)) ds$$

$$I_2 = \frac{t^2}{2T} \int_0^T \left[f'^2(\theta_o, X(s)) - f'^2(\theta_T^{**}, X(s))\right]^2 ds$$

$$I_3 = \frac{t^2}{2T} \int_0^T f''(\theta_T^*, X(s)) dW(s)$$

$$I_4 = \int_0^T \left[f(\theta_T, X(s)) - f(\theta_o, X(s))\right] \left[f(\theta_T + tT^{-1/2}, X(s)) - f(\theta_T, X(s)) - tT^{-1/2} f'(\theta_T, X(s))\right] ds$$

where $\max(|\theta_T^* - \theta_T|, |\theta_T^{**} - \theta_T|) \leq |t| T^{-1/2}$.

By assumption A4 (ii), $I_1 \rightarrow -\beta t^2/2$ a.s. as $T \rightarrow \infty$.

By assumption A2 (ii) and (iii),

$$|I_2| \leq t^2 T^{-1} |\theta - \theta_T^{**}|^{\beta_2} \int_0^T J_3(X(s)) J_4(X(s)) ds .$$

Using consistency of θ_T and assumption A4 (iv) it follows that $I_2 \rightarrow 0$ a.s. as $T \rightarrow \infty$.

By Lemma 5.3.5, $I_3 \rightarrow 0$ a.s. as $T \rightarrow \infty$.

$$|I_4| \leq c\,|\theta - \theta_T|^{\beta_1}\, t^2 T^{-1} \int_0^T J_2(X(s))J_5(X(s))ds .$$

Again, by using consistency of θ_T and assumption A3 (vii) it follows that $I_4 \rightarrow 0$ a.s. as $T \rightarrow \infty$. This proves (i).

(ii) Fix $\varepsilon_1 > 0$. Clearly there exists a T_1 such that $\forall\, T \geq T_1$,

$$-\frac{t^2}{2T}\int_0^T f'^2(\theta_0, X(s))ds \leq -\frac{1}{2}\, t^2(\beta - \varepsilon_1) \quad \text{a.s.} \qquad \ldots(5.3.2)$$

By Lemma 5.3.5 there exists a T_2 such that $\forall\, T \geq T_2$,

$$\sup_\theta \frac{1}{T}\left|\int_0^T f''(\theta, X(s))dW(s)\right| \leq \varepsilon_1/2 \quad \text{a.s.} \qquad \ldots(5.3.3)$$

$$|I_2| \leq \frac{t^2}{2T}\,|\theta_T^{**} - \theta_0|^{\beta_2} \int_0^T J_3(X(s))J_4(X(s))ds$$

$$\leq \frac{t^2}{2T}\,(|t|T^{-1/2} + |\theta_T - \theta_0|)^{\beta_2} \int_0^T J_3(X(s))J_4(X(s))ds .$$

Using A3 (iv) and choosing δ_0 suitably and using consistency of θ_T, it follows that there exist δ_0 and T_3 such that

$$|t|\,T^{-1/2} \leq \delta_0 \text{ and } T \geq T_3 \text{ implies } I_2 \leq \frac{t^2}{2T}\,\varepsilon_1 \quad \text{a.s.} \qquad \ldots(5.3.4)$$

Similarly using mean-value theorem and arguing as above, there exist

T_4 and δ_1 such that (here assumptions A1 (iii), A3 (vii) and A5 (ii) are used)

$$|t|\,T^{-1/2} \leq \delta_1 \quad \text{and} \quad T \geq T_4 \quad \text{implies} \quad I_4 \leq \frac{t^2}{2T}\,\varepsilon_1 \quad \text{a.s.} \qquad \ldots(5.3.5)$$

Combining the estimates (5.3.2) - (5.3.5), (ii) follows.

$$\text{(iii)} \quad \log \frac{\gamma_T(t)}{T} = \frac{1}{T}\int_0^T \left[f(\theta_T + tT^{-1/2}, X(s)) - f(\theta_T, X(s))\right] dW(s)$$

$$- \frac{1}{2T}\int_0^T \left[f(\theta_T + tT^{-1/2}, X(s)) - f(\theta_o, X(s))\right]^2 ds$$

$$+ \frac{1}{2T}\int_0^T \left[f(\theta_T, X(s)) - f(\theta_o, X(s))\right]^2 ds$$

$$= A_1(t,T) + A_2(t,T) + A_3(T), \quad \text{say.}$$

Note that A_3 does not involve t and by arguments given earlier $A_3(T) \longrightarrow 0$ a.s. as $T \longrightarrow \infty$ under A1 (iii) and A3 (i).

By Lemma 4.4.1(2) under A1 (iii) and A3 (i),

$$\sup_t |A_1(t,T)| \leq 2 \sup_\theta \frac{1}{T}\left|\int_0^T f(\theta, X(s))\,dW(s)\right| \longrightarrow 0 \quad \text{a.s.}$$

Finally, by strong consistency, there exists a T_o such that for all $T \geq T_o$, $|\theta_T - \theta_o| \leq \delta/2$ a.s.

Hence if $|t|\,T^{-1/2} \geq \delta$ and $T \geq T_o$, we have

$$|\theta_T + tT^{-1/2} - \theta_o| > \delta/2.$$

Thus $A_2 \leq -\frac{1}{2} \inf_{|\theta - \theta_0| \geq \delta/2} \frac{I_T(\theta)}{T} \rightarrow -\frac{1}{2}\lambda(\delta/2)$ a.s. by Lemma 4.4.2, under A1 (iii), (iv), A3 (ii) and A4 (i).

Combining these estimates of A_1, A_2 and A_3, (iii) is proved.

Let K be a non-negative measurable function such that

(K1) There exists a number ε, $0 < \varepsilon < \beta$, for which

$$\int_{-\infty}^{\infty} K(t)\exp(-(\beta - \varepsilon)t^2/2)dt < \infty$$

(K2) For every $h > 0$ and every $\delta > 0$,

$$e^{-T\delta} \int_{|t|>h} K(T^{1/2}t)\lambda(\theta_T + t)dt \rightarrow 0 \quad \text{a.s. as } T \rightarrow \infty .$$

<u>Lemma 5.3.7</u> : Under assumption (A1) - (A5) except A3 (iii),

(a) There exists a $\delta_0 > 0$ such that a.s.,

$$\lim_{T \rightarrow \infty} \int_{|t| \leq \delta_0 T^{1/2}} K(t)\left|\gamma_T(t)\lambda(\theta_T + tT^{-1/2}) - \lambda(\theta_0)\exp(-\beta t^2/2)\right| dt = 0.$$

(b) For every $\delta > 0$ a.s.,

$$\lim_{T \rightarrow \infty} \int_{|t| > \delta T^{1/2}} K(t)\left| \gamma_T(t)\lambda(\theta_T + tT^{-1/2}) - \lambda(\theta_0)\exp(-\beta t^2/2) \right| dt = 0.$$

<u>Proof</u> : We follow the lines of argument in Borwanker et. al. (1971).

(a) $$\int_{|t| \leq \delta_o T^{1/2}} K(t) \left| \gamma_T(t)\lambda(\theta_T + tT^{-1/2}) - \lambda(\theta_o)\exp(-\beta t^2/2) \right| dt$$

$$\leq \int_{|t| \leq \delta_o T^{1/2}} K(t)\lambda(\theta_o) \left| \gamma_T(t) - \exp(-\beta t^2/2) \right| dt$$

$$+ \int_{|t| \leq \delta_o T^{1/2}} K(t)\gamma_T(t) \left| \lambda(\theta_o) - \lambda(\theta_T + tT^{-1/2}) \right| dt .$$

Choose $0 < \varepsilon < \beta$ such that

$$\int K(t)\exp(-(\beta-\varepsilon)t^2/2)dt < \infty .$$

By Lemma 5.3.6 (ii), there exists a δ_1 and T_1 such that

$$\gamma_T(t) \leq \exp(-(\beta-\varepsilon)t^2/2) \quad \forall\ |t| \leq \delta_1 T^{1/2} \text{ and } T \geq T_1 \quad \text{a.s.}$$

Thus using Lemma 5.3.6 and dominated convergence theorem

$$\lim_{T \to \infty} \int_{|t| \leq \delta_1 T^{1/2}} K(t)\lambda(\theta_o) \left| \gamma_T(t) - \exp(-\beta t^2/2) \right| dt = 0 \quad \text{a.s.}$$

Since θ_T is consistent, for large T, $|\theta - \theta_T| \leq \delta_o$ a.s.

Thus $$\int_{|t| \leq \delta_o T^{1/2}} K(t)\gamma_T(t)|\lambda(\theta_o) - \lambda(\theta_T + tT^{-1/2})| dt$$

$$\leq \sup_{|\theta - \theta_o| \leq 2\delta_o} |\lambda(\theta) - \lambda(\theta_o)| \int_{|t| \leq \delta_o T^{1/2}} K(t)\exp(-(\beta-\varepsilon)t^2/2)dt \quad \text{a.s.}$$

(a) follows now by continuity of λ around θ_o and condition (K1).

(b) $$\int_{|t|>\delta T^{1/2}} K(t)\,|\,\gamma_T(t)\lambda(\theta_T + tT^{-1/2}) - \lambda(\theta_o)\exp(-\beta t^2/2)\,|\,dt$$

$$\leq \int_{|t|>\delta T^{1/2}} K(t)\gamma_T(t)\lambda(\theta_T + tT^{-1/2})dt$$

$$+ \int_{|t|>\delta T^{1/2}} K(t)\lambda(\theta_o)\exp(-\beta t^2/2)dt$$

By Lemma 5.3.6 (iii), for all large T, the first term is dominated by (for some $\varepsilon > 0$),

$$\exp(-T\varepsilon/4) \int_{|t|>\delta T^{1/2}} K(t)\lambda(\theta_T + tT^{-1/2})dt \quad \text{a.s.}$$

By assumption (K2), this converges to zero a.s. as $T \to \infty$.

Using (K1) and dominated convergence theorem, the second term is easily seen to converge to zero. This completes the proof of the lemma.

<u>Theorem 5.3.8</u> : Under assumptions (K1), (K2), (A1) - (A5), except A3 (iii),

$$\lim_{T\to\infty} \int_{-\infty}^{\infty} K(t)\,|\,p^*(t/X(s) : 0 \leq s \leq T) - (\beta/2\pi)^{1/2}\exp(-\beta t^2/2)|\,dt = 0 \quad \text{a.s.}$$

<u>Proof</u> : The proof follows easily from Lemma 5.3.7. We give it here for the sake of completeness.

Note that $K(t) \equiv 1$ satisfies (K1) and (K2). Thus

$C_T = \int \gamma_T(t)\lambda(\theta_T + tT^{-1/2})dt$ converges to

$\lambda(\theta_o)\int \exp(-\beta t^2/2)dt = \lambda(\theta_o)(2\pi\beta^{-1})^{1/2}$.

Hence

$$\int K(t)\,|\,p^*(t/X(s)\ :\ 0\leq s\leq T)-(\beta/2\pi)^{1/2}\exp(-\beta t^2/2)|\,dt$$

$$\leq \int K(t)\,|\,C_T^{-1}\lambda(\theta_T+tT^{-1/2})\gamma_T(t)-C_T^{-1}\lambda(\theta_0)\exp(-\beta t^2/2)\,|\,dt$$

$$+\int K(t)\,|\,C_T^{-1}\lambda(\theta_0)-(\beta/2\pi)^{1/2}|\,\exp(-\beta t^2/2)\,dt$$

Now the proof is obvious.

<u>Corollary 5.3.9</u> : If further $\int_{-\infty}^{\infty} |\theta|^m\lambda(\theta)d\theta < \infty$ for some m, then a.s.,

$$\lim_{T\to\infty}\int_{-\infty}^{\infty} |t|^m\,|\,p^*(t/X(s)\ :\ 0\leq s\leq T)-(\beta/2\pi)^{1/2}\exp(-\beta t^2/2)\,|\,dt=0$$

<u>Proof</u> : It can be easily checked that $K(t)=|t|^m$ satisfies the conditions of Theorem 5.3.8. (See Borwanker et. al. (1971)). This proves the corollary.

<u>Remark 5.3.10</u>

The case $m=0$ yields the analogue of the classical theorem on posteriors.

5.4 <u>Bayes estimation</u>

Suppose $\ell(\theta,\varphi)$ is a loss function defined on $\Omega\times\Omega$. Assume that $\ell(\theta,\varphi)=\ell(|\theta-\varphi|)\geq 0$ and $\ell(t)$ is nondecreasing. Suppose R is a non-negative function and K and G are functions such that

(B1) $R(T)\ell(tT^{-1/2})\leq G(t)$ for all $T\geq 0$.

(B2) $R(T)\ell(tT^{-1/2})\to K(t)$ uniformly on bounded intervals of t as $T\to\infty$.

(B3) $\int_{-\infty}^{\infty} K(t+m)\exp(-\beta t^2/2)dt$ has a strict minimum at $m = 0$.

(B4) G satisfies (K1) and (K2).

A Bayes estimate $\hat{\theta}_T$ based on $(X(s) : 0 \leq s \leq T)$ is that which minimizes

$$B_T(\psi) = \int_{\Omega} \ell(\theta,\psi)p(\theta / X(s) : 0 \leq s \leq T)d\theta$$

Assume that such an estimator exists.

<u>Theorem 5.4.1</u> : Under (A1) - (A5),(B1) - (B4) and (K1), (K2),

(i) $T^{1/2}(\hat{\theta}_T - \theta_T) \to 0$ a.s. as $T \to \infty$

(ii) $\lim_{T\to\infty} R(T)B_T(\theta_T) = \lim_{T\to\infty} R(T)B_T(\hat{\theta}_T)$

$$= (\beta/2\pi)^{1/2} \int_{-\infty}^{\infty} K(t)\exp(-\beta t^2/2)dt \quad \text{a.s.}$$

<u>Proof</u> : The proof uses Theorem 5.3.8 and the properties of m.l.e. θ_T proved in Chapter 4. The arguments are exactly as in Borwanker et. al. (1971). We shall give only a brief sketch. Below $a_T \equiv R(T)$.

$$\limsup_T a_T B_T(\hat{\theta}_T) \leq \limsup_T a_T B_T(\theta_T)$$

$$= \limsup_T \int a_T \ell(tT^{-1/2})p^*(t/X(s) : 0 \leq s \leq T)dt.$$

Writing $p^*(t)$ for $p^*(t/X(s) : 0 \leq s \leq T)$, the r.h.s. of the above inequality is bounded by

$$\limsup_T \int \left| a_T \lambda(tT^{-1/2}) - K(t) \right| \left| p^*(t) - \exp(-\beta t^2/2)(\beta/2\pi)^{1/2} \right| dt$$

$$+ \limsup_T (\beta/2\pi)^{1/2} \int \left| a_T \lambda(tT^{-1/2}) - K(t) \right| \exp(-\beta t^2/2) dt$$

$$+ \limsup_T \int K(t) p^*(t) dt.$$

The first term is dominated by

$$\limsup_T \int 2G(t) \left| p^*(t) - (\beta/2\pi)^{1/2} \exp(-\beta t^2/2) \right| dt$$

which by Theorem 5.3.8 is zero.

The second term is zero by dominated convergence theorem.

The last term is

$$(\beta/2\pi)^{1/2} \int K(t) \exp(-\beta t^2/2) dt = \alpha \quad \text{say.}$$

Thus $\limsup_T a_T B_T(\hat{\theta}_T) \leq \limsup_T a_T B_T(\theta_T) \leq \alpha$ ...(5.4.1)

Let $U_T = T^{1/2}(\theta_T - \hat{\theta}_T)$.

The next step is to show that $\sup_T |U_T| < \infty$ a.s.

Suppose this is not true. We assume w.l.g. that for every $M > 0$ there exist A_M such that $P_{\theta_0}(A_M) > 0$ and $U_{T_n}(x) > M \ \forall x \in A_M$ for a sequence $T_n \to \infty$. On A_M,

$$\begin{aligned} a_{T_n} B_{T_n}(\hat{\theta}_{T_n}) &= \int a_{T_n} \lambda \left[(t + U_{T_n}) T_n^{-1/2}\right] p^*_{T_n}(t) dt \\ &\geq \int_{|t| \leq M} a_{T_n} \lambda \left[(t + U_{T_n}) T_n^{-1/2}\right] p^*_{T_n}(t) dt \\ &\geq \int_{|t| \leq M} a_{T_n} \lambda \left[(t + M) T_n^{-1/2}\right] p^*_{T_n}(t) dt. \end{aligned}$$

This expression converges to

$$\int_{|t|\leq M} K(t+M)(\beta/2\pi)^{1/2}\exp(-\beta t^2/2)dt.$$

However $\lim_{M\to\infty} \int_{|t|\leq M} K(t+M)(\beta/2\pi)^{1/2}\exp(-\beta t^2/2)dt > \alpha$.

Hence for large M, on a set of positive probability ,

$$\liminf_n a_{T_n} B_{T_n}(\hat{\theta}_{T_n}) > \alpha > \limsup_n a_{T_n} B_{T_n}(\theta_{T_n}).$$

This contradicts the definition of $\hat{\theta}_T$. Thus $\sup_T |U_T| < \infty$ a.s.

Fix $\varepsilon > 0$. Let B_M be a set such that $P(B_M) > 1-\varepsilon$ and $|U_T(x)| \leq M$ $\forall x \in B_M$, $\forall T > 0$. Fix $x \in B_M$.

Take a sequence $T_n \to \infty$ such that $U_{T_n}(x) \to m$. Suppose if possible $m \neq 0$.

$$\begin{aligned}\liminf_n a_{T_n} B_{T_n}(\hat{\theta}_{T_n}) &\geq \liminf_n \int_{-T_0}^{T_0} a_{T_n} \lambda\left[(t+U_{T_n})T_n^{-1/2}\right] p^*_{T_n}(t)dt \\ &\geq \int_{-T_0}^{T_0} \liminf_n a_{T_n} \lambda\left[(t+U_{T_n})T_n^{-1/2}\right] p^*_{T_n}(t)dt \\ &= (\beta/2\pi)^{1/2} \int_{-T_0}^{T_0} K(t+m)\exp(-\beta t^2/2)dt. \qquad \text{...(5.4.2)}\end{aligned}$$

Thus $\liminf_n a_{T_n} B_{T_n}(\hat{\theta}_{T_n}) \geq (\beta/2\pi)^{1/2} \int_{-\infty}^{\infty} K(t+m)\exp(-\beta t^2/2)dt$

$$> \alpha \geq \limsup_n a_T B_T(\hat{\theta}_T)$$

This is impossible. Thus $m = 0$. Hence $T^{1/2}(\theta_T - \hat{\theta}_T) \rightarrow 0$ a.s. Further using relations (5.4.1) and (5.4.2),

$$\lim_T a_T B_T(\theta_T) = \lim_T a_T B_T(\hat{\theta}_T) = \alpha \quad \text{a.s.}$$

This proves the theorem completely.

Remark 5.4.2

Assume (A1) - (A5) and (B1) - (B4). Then as $T \rightarrow \infty$, $\hat{\theta}_T \rightarrow \theta_0$ a.s. and $T^{1/2}(\hat{\theta}_T - \theta_0) \rightarrow N(0,\beta^{-1})$.

This follows easily from previous results.

Remark 5.4.3

(1) Once the asymptotic normality of the Bayes estimator is established, it would be interesting to obtain the rate of convergence to normality. There has been some recent work in this direction by Mishra and Prakasa Rao (1985).

(2) It is also interesting to see whether law of iterated logarithm holds for such estimators. For some work in this direction we refer the reader to Prakasa Rao (1985b).

CHAPTER 6

BERRY-ESSEEN BOUND FOR THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE ORNSTEIN-UHLENBECK PROCESS

6.1 Introduction

In Chapter 4 we showed that the maximum likelihood estimator in certain diffusion processes has an asymptotic normal distribution. The next question which naturally arises is its rate of covergence to normality. In the general set up of Chapter 4, this problem seems to be hard.

However, there has been some work for a class of simpler processes. Suppose that $X(t)$ is governed by the stochastic differential equation

$$dX(t) = \theta a(X(t))dt + b(X(t))dW(t) \qquad \ldots(6.1.1)$$

$X(0) = 0$, $t \geq 0$, $\theta > 0$ and $W(t)$ is a standard Wiener process. The m.l.e. θ_T of θ, based on the observation $X(t)$, $0 \leq t \leq T$, satisfies the equation

$$\theta_T - \theta = \int_0^T \frac{a(X(t))}{b(X(t))}\, dW(t) \Big/ \int_0^T \frac{a^2(X(t))}{b^2(X(t))}\, dt .$$

Mishra and Prakasa Rao (1985) has studied the rate of convergence of the above estimator. Their main result is given below.

Let $I_T = \int_0^T \frac{a^2(X(t))}{b^2(X(t))}\, dt$. Introduce the following assumptions :

(A1) $0 < E(I_T) < \infty$.

(A2) There exist positive functions $Q(T)\uparrow\infty$, $\varepsilon(T)\downarrow 0$ such that $Q(T)\,\varepsilon^2(T)\rightarrow\infty$ and

$$\sup_{\theta\in\Omega} P_{\theta,T}\left(\left|\frac{I_T}{Q_T}-1\right|\geq\varepsilon(T)\right)=O(\varepsilon^{1/2}(T))$$

Then

$$\sup_{\theta\in\Omega}\ \sup_x \left|P_{\theta,T}(Q^{1/2}(T)(\theta_T-\theta)\leq x)-\Phi(x)\right|$$

$$\leq 2\varepsilon(T)^{1/2}+2P_{\theta,T}\left(\left|\frac{I_T}{Q_T}-1\right|\geq\varepsilon(T)\right)+\varepsilon(T) \quad \text{if (A1) holds}$$

$$= O(\varepsilon^{1/2}(T)) \quad \text{if both (A1) and (A2) hold.}$$

Note that this involves a growth condition on I_T, which incidentally is hard to check. However, this is the only approach known for obtaining rate of covergence to normality of martingales; the rate directly depends on the rate of convergence of the conditional variance to the limiting variance. See e.g. the developments in Hall and Heyde (1980).

For the case $a(x)=-x$, $b(x)=1$, the process $X(t)$ satisfies $dX(t)=-\theta X(t)dt+dW(t)$, $X(0)=0$, $\theta>0$ and is called the Ornstein-Uhlenbeck (O - U) process. The growth conditions hold for this process with $Q(T)=T^{1/2}$ and $\varepsilon(T)=T^{-2/5}$, yielding the rate $O(T^{-1/5})$ for m.l.e.

However, the O - U process being a natural continuous time analogue of the first order discrete autoregressive process with i.i.d $N(0,1)$ errors, one is led to believe that the above rate can be sharpened. (Recall the strong results of Chapter 1 on linear processes).

Mishra and Prakasa Rao (1985) use simple Markov inequalities to tackle the denominator. The numerator is embedded in a Brownian motion by Kunita-Watanabe theorem and then Lemma 3.2 of Hall and Heyde (1980) is invoked. These two together limit the rate obtainable to $T^{-1/5}$. For the O-U process, Burkholder's inequality can be used for the denominator (after applying Ito's formula) to yield the better rate $T^{-1/4+\varepsilon}$, $\varepsilon > 0$. However, as long as we use embedding technique, the rate cannot be better than $T^{-1/4}$.

We take an alternative approach. By extending an argument in Liptser and Shiryayev (1978) (henceforth referred as LS), we obtain the characteristic function of the numerator for suitable values of the argument. This allows us to use Esseen's lemma yielding the rate $O(T^{-1/2})$ for the numerator. The denominator is linked with the numerator via Ito's formula. This helps us to get the final result.

This result opens up the possibility of obtaining faster rates of convergence for general linear diffusions. It also shows that the embedding technique might not lead to the strongest possible results.

This chapter is a revised version of Bose (1986).

6.2 Preliminaries

Let $(X(t),\ t \geq 0)$ be a diffusion process satisfying the stochastic differential equation

$$dX(t) = -\theta X(t)dt + dW(t),\ t \geq 0,\ X(0) = 0.$$

Here $W(\cdot)$ is a standard Brownian motion and $\theta > 0$ is the unknown parameter. It is well known that the solution $X(t)$ is a continuous Gaussian process. In fact, $X(t) = \int_0^t e^{-\theta(t-s)} dW(s)$.

Let $C[0,T]$ = The space of real valued continuous function on $[0,T]$.

M_θ^T = The measure generated by $X(t)$, $0 \leq t \leq T$ on $C[0,T]$

M_W^T = The measure generated by $W(t)$, $0 \leq t \leq T$ on $C[0,T]$.

It is well known that $M_\theta^T << M_W^T$ and the Radon-Nikoydm derivative (likelihood function) can be explicitly computed (see e.g. Chapter 4). This in turn yields that the m.l.e. θ_T of θ based on the " observation " $X(t)$, $0 \leq t \leq T$ satisfies

$$\theta_T - \theta = \int_0^T X(t)dW(t)\left(\int_0^T X^2(t)dt\right)^{-1}.$$

Ito's formula (see Elliott (1982)) gives

$$2\int_0^T X(s)dW(s) - X^2(T) = 2\theta\int_0^T X^2(t)dt - T. \qquad \ldots(6.2.1)$$

This relation shall be used later.

C shall denote a generic constant (perhaps depending on θ, but not on anything else).

6.3 The main result

We begin with a few lemmas.

<u>Lemma 6.3.1</u> : For $Z_1, Z_2 \in \mathbb{C}^2$,

$$\text{Let } \varphi_T(Z_1,Z_2) = E \exp(Z_1 \int_0^T X^2(t)dt + Z_2 X^2(T)).$$

Then $\varphi_T(Z_1,Z_2)$ exists for $|Z_i| < \delta$, $i = 1,2$ for some $\delta > 0$ and

$$\varphi_T(Z_1,Z_2) = \exp(\tfrac{\theta T}{2}) \left[\frac{2\lambda}{(\lambda-\theta+2Z_2)\exp(-\lambda T) + (\lambda+\theta-2Z_2)\exp(\lambda T)}\right]^{1/2} \quad \ldots(6.3.1)$$

where $\lambda = (\theta^2 - 2Z_1)^{1/2}$ and we always choose the principal branch of the square root.

<u>Proof</u> : First assume that $Z_i = a_i \in \mathbb{R}$, $i = 1,2$ and a_i are sufficiently small.

Define $\lambda = (\theta^2 - 2a_1)^{1/2}$ and $dX_t^\lambda = \lambda X_t^\lambda dt + dW(t)$, $X_o^\lambda = 0$.

Also recall that $dX_t^\theta = -\theta X_t^\theta dt + dW(t)$.

Then we have (see LS)

$$\frac{dM_\theta^T}{dM_\lambda^T}(X^\lambda(.)) = \exp\left[(-\theta-\lambda)\int_0^T X_t^\lambda dX_t^\lambda - (\frac{\theta^2-\lambda^2}{2})\int_0^T (X_t^\lambda)^2 dt\right] \quad \ldots(6.3.2)$$

Note that

$$\varphi_T(a_1,a_2) = E_\theta \exp(a_1 \int_0^T (X_t^\theta)^2 dt + a_2 (X_T^\theta)^2).$$

If we change the measure to that generated by X_t^λ, then by (6.3.2),

$$\varphi_T(a_1,a_2) = E \exp\left[a_1 \int_0^T (X_t^\lambda)^2 dt + a_2 (X_t^\lambda)^2 - (\theta+\lambda)\int_0^T X_t^\lambda dX_t^\lambda - a_1 \int_0^T (X_t^\lambda)^2 dt\right]$$

$$= E \exp\left[a_2 (X_T^\lambda)^2 - (\theta+\lambda)\int_0^T X_t^\lambda dX_t^\lambda\right] \quad \ldots(6.3.3)$$

By Ito's formula,

$d(X_t^{\lambda})^2 = 2X_t^{\lambda}\, dX_t^{\lambda} + dt.$ Using this in (6.3.3),

$$\varphi_T(a_1,a_2) = E \exp\left[(X_t^{\lambda})^2 (a_2 - \frac{\theta+\lambda}{2}) + \frac{T}{2}(\theta+\lambda)\right]$$

Note that $X_t^{\lambda} \sim N(0, \frac{\exp(2\lambda T)-1}{2\lambda})$. Thus

$$\varphi_T(a_1,a_2) = \exp(\frac{T(\theta+\lambda)}{2})\left[\frac{2\lambda}{2\lambda + (\lambda+\theta-2a_2)(\exp(2\lambda T)-1)}\right]^{1/2}$$

which on simplification yields (6.3.1) for Z_i's real, around a neighbourhood of zero. Thus, there is no problem of existence of $\varphi_T(Z_1,Z_2)$ around zero in C^2 and since we have shown that the m.g.f. exists, $\varphi_T(Z_1,Z_2)$ is an analytic function. On the other hand (6.3.1) defines an analytic function in the relevant domain and agrees with $\varphi_T(Z_1,Z_2)$ for Z_1,Z_2 real. This finishes the proof.

<u>Lemma 6.3.2</u> : For $|t| \leq \varepsilon\, T^{1/2}$, where ε is sufficiently small,

$$\left|E \exp(it(\frac{2\theta}{T})^{1/2}\int_0^T X(s)dW(s)) - \exp(-t^2/2)\right| \leq C.\exp(-t^2/4)|t|^3\, T^{-1/2}.$$

<u>Proof</u> : By Ito's formula,

$$\int_0^T X(s)dW(s) = \theta\int_0^T X^2(t)dt - \frac{T}{2} + \frac{X^2(T)}{2}\ .$$

Hence $E \exp(it(\frac{2\theta}{T})^{1/2}\int_0^T X(s)dW(s)) = \varphi_T(Z_1,Z_2)\exp(\frac{-it}{2}(2\theta T)^{1/2})$ ...(6.3.4)

where $Z_1 = it\theta(\frac{2\theta}{T})^{1/2}$, $Z_2 = \frac{it}{2}(\frac{2\theta}{T})^{1/2}$.

Note that (Z_1, Z_2) satisfies the condition of Lemma 6.3.1 by choosing ε sufficiently small.

Clearly $\lambda - \theta = O(|t|T^{-1/2})$, $\lambda + \theta = 2\theta + O(|t|T^{-1/2})$ and

$$\lambda = \theta\beta_T(t) + O(|t|^3T^{-1/2})$$

where $\beta_T(t) = 1 - \frac{Z_1}{\theta^2} - \frac{Z_1^2}{2\theta^4}$.

Let $\alpha_T(t)$ denote any function which is $O(|t|T^{-1/2})$. Using these simple estimates,

$$\varphi_T(Z_1, Z_2) = \exp(\tfrac{\theta T}{2})\left[\frac{(1+\alpha_T(t))\exp(T\theta\beta_T(t) + O(|t|^3T^{-1/2}))}{\alpha_T(t) + (2+\alpha_T(t))\exp(2T\theta\beta_T(t) + O(|t|^3T^{-1/2})}\right]^{1/2}$$

Using this in (6.3.4), the required expectation equals

$$\left[\frac{1+\alpha_T(t)}{\alpha_T(t)\exp(-\frac{T\theta}{2}) + (1+\alpha_T(t))\exp(\psi_T(t))}\right]^{1/2}$$

where $\psi_T(t) = T\theta\beta_T(t) - \theta T + \frac{it}{2}(2\theta T)^{1/2} + O(|t|^3T^{-1/2})$

$$= t^2 + O(|t|^3T^{-1/2}).$$

Thus, the difference to be estimated, is, in absolute value

$$= \left|\exp(-t^2/2)(1+\alpha_T(t))\exp(O(|t|^3T^{-1/2})) - \exp(-t^2/2)\right|$$

$$\leq C.\exp(-t^2/2)|t|^3T^{-1/2}\exp(O(|t|^3T^{-1/2}))$$

$$\leq C.|t|^3T^{-1/2}\exp(-t^2/4) \text{ choosing } \varepsilon \text{ sufficiently small.}$$

This proves the lemma.

Let $\quad Y(T) = (\frac{2\theta}{T})^{1/2} \int_0^T X(s)dW(s).$

Lemma 6.3.2 and the well known Esseen's lemma immediately yields the following lemma.

<u>Lemma</u> 6.3.3 : $\sup_{x \in \mathbb{R}} \left|P(Y(T) \leq x) - \Phi(x)\right| \leq C.T^{-1/2}.$

We now state the main theorem.

<u>Theorem 6.3.4</u> : $\sup_{x \in \mathbb{R}} \left|P((\frac{T}{2\theta})^{1/2}(\theta_T - \theta) \leq x) - \Phi(x)\right| \leq C.T^{-1/2}$

and the bound is uniform over any fixed compact subset of values of θ.

<u>Proof</u> : Note that

$$(\frac{T}{2\theta})^{1/2}(\theta_T - \theta) = \frac{Y(T)}{2\theta T^{-1} \int_0^T X^2(t)dt}$$

(6.2.1) yields $\quad 2\theta T^{-1} \int_0^T X^2(t)dt = 1 - T^{-1/2}(\frac{2}{\theta})^{1/2} \frac{Y(T)}{T} - \frac{X^2(T)}{T}.$

Let $\quad B_1 = \left\{|Y(T)| \geq \delta \log T\right\}$

$\quad B_2 = \left\{T^{-1/4}X^2(T) \geq \delta \log T\right\}$ where δ is large.

By Lemma 6.3.3,

$$\left|P(B_1) - P(|N(0,1)| \geq \delta \log T)\right| \leq C.T^{-1/2}.$$

Using simple approximation for the tails of a normal distribution, $P(|N(0,1)| \geq \delta \log T) = O(T^{-1/2}).$

Thus $\quad P(B_1) = O(T^{-1/2}). \qquad \ldots(6.3.5)$

Note that $X(T) \sim N(0, \frac{1 - \exp(-2\theta T)}{2\theta})$. The variance being bounded in T, simple Markov inequality gives

$$P(B_2) = O(T^{-1/2}).$$

On $B_1^c \cap B_2^c$, $|(\frac{T}{2\theta})^{1/2}(\theta_T - \theta)| \leq C.\log T$ for some C.

Also by a simple binomial expansion of the denominator, on $B_1^c \cup B_2^c$,

$$(\frac{T}{2\theta})^{1/2}(\theta_T - \theta) = Y(T) + T^{-1/2}(\frac{2}{\theta})^{1/2} Y^2(T) + o(T^{-1/2})$$

$$= Y(T) + T^{-1/2} a Y^2(T) + o(T^{-1/2}) \quad \text{say.}$$

Now note that for any $|u| \leq C \log T$,

$$x + T^{-1/2} a x^2 \leq u \quad \text{iff}$$

$$(x + \frac{T^{1/2}}{2a})^2 \leq \frac{uT^{1/2}}{a} + \frac{T}{4a^2}$$

iff $a_2(T,u) \leq x \leq a_1(T,u)$ where

$$a_1(T,u) = (\frac{uT^{1/2}}{a} + \frac{T}{4a^2})^{1/2} - \frac{T^{1/2}}{2a}$$

$$= u + O(u^2/T)$$

$$= u + O(T^{-1}(\log T)^2)$$

and $\quad a_2(T,u) = -(\frac{uT^{1/2}}{a} + \frac{T}{4a^2})^{1/2} - \frac{T^{1/2}}{2a} \leq -C.T^{-1/2}$.

But as in (6.3.5), $P(Y(T) \leq -C.T^{1/2}) = O(T^{-1/2})$, and $\quad |P(Y(T) \leq x + O(T^{-1}(\log T)^2) - \Phi(x)|$

$$\leq C.T^{-1/2} + |\Phi(x) - \Phi(x + O(T^{-1}(\log T)^2))|$$

$\leq C.T^{-1/2}$ uniformly over x. This proves the theorem completely.

Remark 6.3.5

(1) Lemma 6.3.1 is derived by a " change of measure " technique. This technique has been used in some entirely different contexts very fruitfully. Two very good examples are the derivations of Wald's identity in sequential analysis, and of moderation deviation probabilities (Michel (1976)).

(2) One of the indispensable tools in the theory of diffusions is the Ito's formula, which is natural since it is a " change of variable formula". Here we have yet another novel application of this formula.

(3) We believe that under reasonable conditions, the rate $O(T^{-1/2})$ is attainable for general linear diffusions of the type (6.1.1). But apparently it has to wait for the development of entirely new techniques. The problem for non-linear diffusions treated in Chapter 4 seems even harder.

REFERENCES

Abramovitch, L. and Singh, K. (1985) : Edgeworth corrected pivotal statistics and the bootstrap. Ann. Statist., 13, 1, 116-132.

Anderson, T.W. (1971) : The statistical analysis of time series. John Wiley and Sons.

Athreya, K.B. (1984a) : Bootstrap of the mean in the infinite variance case I. Preprint, Iowa State University.

Athreya, K.B. (1984b) : Bootstrap of the mean in the infinite variance case II. Preprint, Iowa State University.

Athreya, K.B. (1984c) : Weak convergence of the bootstrap c.d.f. of the mean when F is in the domain of attraction of a stable law. Preprint, Iowa State University.

Athreya, K.B., Ghosh, M., Low, L.Y. and Sen, P.K. (1984) : Law of large numbers for bootstrap U-statistics. J. Stat. Planning and Inf., 9, 185-194.

Babu, G.J. (1984) : Bootstrapping statistics with linear combinations of Chi-squares as weak limit. Sankhyā Ser.A, 46,1, 85-93.

Babu, G.J. and Bose, A. (1986) : Bootstrap confidence intervals, with applications to autoregressive models. Preprint.

Babu, G.J. and Singh, K. (1983) : Inference on means using the bootstrap. Ann. Statist. 11,3,999-1003.

Babu, G.J. and Singh, Kesar (1984) : On one term Edgeworth correction by Efron's bootstrap. Sankhyā Ser.A, 40,2, 219-232.

Basawa, I.V. and Prakasa Rao, B.L.S. (1980) : Statistical Inference for Stochastic Processes. Academic Press, London.

Beran, R. (1982) : Estimated sampling distribution : the bootstrap and competitors. Ann. Statist., 10, 212-225.

Beran, R. (1984) : Bootstrap methods in statistics. Jber. d. Dt. Math-Verein. 86, 14-30.

Beran, R. and Srivastava, M.S. (1985) : Bootstrap tests and confidence regions for functions of a covariance matrix. Ann. Statist., 13,1, 95-115.

Bhattacharya, R.N. and Ghosh, J.K. (1978) : On the validity of the formal Edgeworth expansions. Ann. Statist., 6, 434-451.

Bickel, P.J. and Freedman, D. (1980) : On Edgeworth expansion for the bootstrap. Preprint.

Bickel, P.J. and Freedman, D. (1981) : Some asymptotic theory for the bootstrap. Ann. Statist., 9, 1196-1217.

Bickel, P.J. and Freedman, D. (1983) : Bootstrapping regression models with a large number of parameters. In the Lehmann Festschrift ed. P. Bickel, J.L. Hodges and K. Doksum. Wadsworth, Belmont California.

Bickel, P.J. and Freedman, D. (1984) : Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist., 12,2, 470-482.

Blume, W. and Wittwer, G. (1981) : On the asymptotic distribution of covariance estimates of stationary random sequences. Math. Oper. Statist. Ser. statistics, 12, 193-199.

Borkar, V. and Bagchi, A. (1981) : Parameter estimation in continuous - time stochastic processes. Memorandum NR. 331, Dept. of Applied Maths, Twente University of Technology, Enschede, Netherlands.

Borwanker, J.D., Kallianpur, G. and Prakasa Rao, B.L.S. (1971) : The Bernstein-von Mises theorem for Markov processes. Ann. Math. Statist., 42, 1241-1253.

Basu (Bose), A. (1983a) : Asymptotic theory of estimation in non-linear stochastic differential equations for the multiparameter case. Sankhyā Ser.A, 45,1, 56-65.

Bose (Basu), A. (1983b) : The Bernstein-von Mises theorem for a class of diffusion processes. Sankhyā Ser. A, 45,2, 150-160.

Bose, A. (1985a) : Expansions and one term Edgeworth correction by bootstrap in autoregressions. T/R No.10/85 Stat. Math. Div., ISI, Calcutta.

Bose, A. (1985b) : Bootstrap in moving average models. T/R No.17/85, Stat. Math. Div., ISI, Calcutta.

Bose, A. (1985c) : Higher order approximations for auto-covariances from linear processes with applications. T/R No.19/85, Stat. Math. Div., ISI, Calcutta.

Bose, A. (1986) : Berry-Esseen bound for the maximum likelihood estimator in the Ornstein-Uhlenbeck process. Sankhyā Ser.A, 48, Part 2. (To appear).

Brown, B.M. and Hewitt, J.T. (1975) : Asymptotic likelihood theory for diffusion processes. J. Appl. Prob., 12, 228-238.

Chatterjee, S. (1985) : Bootstrapping ARMA models : Some simulations. Preprint, College of Business Administration, Northeastern University.

Daggett, R. and Freedman, D.A. (1984) : Econometrics and the law : a case study in the proof of antitrust damages. Proceedings of the Berkeley Conference in honor of J. Neyman and J. Kiefer, ed. LeCam and Olshen, Vol.1, 123-172. Wadsworth, Inc., Belmont California.

Delaney, N.J. and Chatterjee, S. (1984) : Use of the bootstrap and cross-validation in ridge regression. J. of Bus. and Eco. Stat. (To appear).

Durbin, J. (1980) : Approximations for densities of sufficient estimators. Biometrika, 67, 311-333.

Efron, B. (1979) : Bootstrap methods : another look at the Jackknife. Ann. Statist., 7, 1-26.

Efron, B. (1982) : The Jackknife, the bootstrap and other resampling plans. CBMS-NSF Regional Conference Series in Applied Maths. Monograph 38, SIAM, Philadelphia.

Efron, B. (1985) : Bootstrap confidence intervals for a class of parametric problems. Biometrika, 72, 45-58.

Efron, B. and Gong, G. (1983) : A leisurely look at the bootstrap, the jackknife, and cross validation. The American Statistician, 37, 36-48.

Efron, B. and Tibshirani, R. (1985) : The bootstrap method for assessing statistical accuracy. Behaviormetrika, 17, 1-35.

Elliott, R.J. (1982) : Stochastic Calculus and Applications. Springer, New York.

Freedman, D. (1981) : Bootstrapping regression models. Ann. Statist., 9, 1218-1228.

Freedman, D. (1984) : On bootstrapping two stage least squares estimates in stationary linear models. Ann. Statist., 12, 827-842.

Freedman, D. and Peters, S. (1984a) : Bootstrapping a regression equation; some empirical results. J. Amer. Statist. Assoc., 79, 97-106.

Freedman, D. and Peters, S. (1984b) : Bootstrapping an econometric model; some empirical results. J. Bus. Econ. Statist., 2, 150-158.

Freedman, D.A. and Peters, S.C. (1984c) : Using the Bootstrap to Evaluate Forecasting Equations. Tech. Report No.20, Univ. of California, Berkeley, Dept. of Statistics. To appear in J. of Forecasting.

Friedman, A. (1975) : Stochastic Differential Equations, Vol.I. Academic Press, New York.

Fujikoshi, Y. and Ochi, Y. (1984) : Asymptotic properties of the maximum likelihood estimate in the first order autoregressive process. Ann. Inst. Statist. Math., Vol.36, Part A, 119-128.

Gotze, F. and Hipp, C. (1978) : Asymptotic expansions in the central limit theorem under moment conditions. Zeit. Wahr. ver. Gebiete, 42, 67-87.

Gotze, F. and Hipp, C. (1983) : Asymptotic expansions for sums of weakly dependent random vectors. Zeit. Wahr. ver. Gebiete, 64, 211-239.

Hall, P. and Heyde, C.C. (1980) : Martingale limit theory and its application. Academic Press, New York.

Hannan, E.J. (1970) : Multiple time series. John Wiley and Sons.

Karandikar, R.L. (1981) : Pathwise solutions of stochastic differential equations. Sankhyā, Ser.A, 43,2, 121-132.

Karandikar, R.L. (1983) : Interchanging the order of stochastic integration and ordinary differentiation. Sankhyā Ser. A, 45,1, 120-124.

Kersten, N. (1984) : On the multidimensional asymptotic distribution of covariance estimates of linear stationary processes. Math. Oper. Statist. ser. statistics, 15,4, 525-631.

Kulinic, G.L. (1975) : On an estimation of the drift parameter of a stochastic diffusion equation. Theory Prob. Appl., 20, 384-387.

Kutoyants, Yu. A. (1975) : On a hypothesis testing problem and asymptotic normality of stochastic integrals. Theory Prob. Appl., 20, 376-384.

Kutoyants, Yu. A. (1977) : Estimation of the trend parameter of a diffusion process in the smooth case. Theory Prob. Appl., 22, 399-406.

Lanska, V. (1979) : Minimum contrast estimation in diffusion processes. J. Appl. Prob., 16, 65-75.

LeBreton, A. (1977) : Parameter estimation in a vector linear stochastic differential equation. Transactions of the 7th Prague Conference on Information theory, Stat. Decision Functions and Random Processes. Vol.A, 353-366.

LeCam, L. (1955) : On the asymptotic theory of estimation and testing of hypotheses. Proc. Third Berkeley Symp. Math. Statist. Prob. Univ. of California Press.

LeCam, L. (1958) : Les proprietes asymptotiques des solutions de Bayes. Publ. Inst. Statist. Univ. Paris, 7, 18-35.

Lee, T.S. and Kozin, F. (1977) : Almost sure asymptotic likelihood theory for diffusion processes. J. Appl. Prob., 14, 527-537.

Liptser, R.S. and Shiryayev, A.N. (1977) : Statistics of Random Processes I General Theory. Springer-Verlag, Berlin.

Liptser, R.S. and Shiryayev, A.N. (1978) : Statistics of Random Processes II Applications. Springer-Verlag, Berlin.

McKean, H.P. (1969) : Stochastic Integrals. Academic Press, New York.

Michel, R. (1976) : Non-uniform central limit bounds with applications to probabilities of deviations. Ann. Prob., 4, 102-106.

Mishra, M.N. and Prakasa Rao, B.L.S. (1985) : On the Berry-Esseen bound for maximum likelihood estimator for linear homogeneous diffusion processes. Sankhyā Ser.A, 47,3, 392-398.

Mishra, M.N. and Prakasa Rao, B.L.S. (1985) : Rate of convergence in the Bernstein-von Mises theorem for a class of diffusion processes. Preprint.

Ochi, Y. (1983) : Asymptotic expansion for the distribution of an estimator in the first-order autoregressive process. J. Time Ser. Anal., 4,1, 57-67.

Peters, S.C. and Freedman, D.A. (1984) : *Some notes on the bootstrap in regression problems.* J. of Bus. and Eco. Stat., 2,4, 406-409.

Phillips, P.C.B. (1977) : Approximations to some finite sample distributions associated with a first-order stochastic difference equation. Econometrica, 45,2, 463-484.

Phillips, P.C.B. (1978) : Edgeworth and saddle point approximations in the first-order non-circular autoregression. Biometrika, 65,1, 91-98.

Prakasa Rao, B.L.S. (1981) : The Bernstein-von Mises theorem for a class of diffusion processes (in Russian). Theor. Random Processes, 9, 95-101.

Prakasa Rao, B.L.S. (1983) : On Bayes estimation for diffusion fields. In Statistics : *Applications and New Directions*, Proc. of the ISI Golden Jubilee Conference, ed. J.K. Ghosh and J. Roy, 504-511.

Prakasa Rao, B.L.S. (1985a) : Estimation of the drift for diffusion process. Statistics, 16,2, 263-275.

Prakasa Rao, B.L.S. (1985b) : Law of iterated logarithm for fluctuation of posterior distributions for a class of diffusion processes and a sequential test of power one. Preprint, Indian Statistical Institute, New Delhi.

Prakasa Rao, B.L.S. (1985c) : On the rate of convergence of the least squares estimator in non-linear regression model for multiparameter. Preprint, Indian Statistical Institute, New Delhi.

Prakasa Rao, B.L.S. and Rubin, H. (1981) : Asymptotic theory of estimation in non-linear stochastic differential equations. Sankhyā Ser.A, 43, 170-189.

Rao , J.N.K. and Wu, C.F.J. (1985) : Bootstrap inference for sample surveys. Proc. of the 1984 ASA meeting, Section on Survey Research Methods, 106-112.

Rubin, D.B. (1981) : The Bayesian bootstrap. Ann. Statist., 9,1, 130-134.

Schenker, N. (1985) : Qualms about bootstrap confidence intervals. J. Amer. Stat. Assoc., 80, 390, 360-361.

Shorack, G.R. (1982) : Bootstrapping robust regression. Comm. Statist., Theor. Meth., 11(9), 961-972.

Singh, K. (1981) : On asymptotic accuracy of Efron's bootstrap. Ann. Statist., 9, 1187-1195.

Statulevicius, V. (1969) : Limit theorems for sums of random variables that are connected in a Markov chain, I, II. Litovsk. Mat. Sb., 9, 345-362, 635-672.

Statulevicius, V. (1970) : Limit theorems for sums of random variables that are connected in a Markov chain III, Litovsk. Mat. Sb., 10, 161-169.

Stroock, D.W. (1982) : Lectures on Topics in Stochastic Differential Equations. TIFR lecture notes Bombay. Published by Springer-Verlag.

Stroock, D.W. and Varadhan, S.R.S. (1979) : Multidimensional Diffusion Processes. Springer-Verlag.

Sweeting, T.J. (1977) : Speeds of convergence for the multidimensional central limit theorems. Ann. Prob., 5, 28-41.

Tanaka, K. (1983) : Asymptotic expansion associated with the AR(1) model with unknown mean. Econometrica, 51,4, 1221-1231.

Taraskin, A.F. (1971) : Parameter estimation by method of maximum likelihood for a stationary process. In Mathematical Physics no.9, Naukova Dumka, Kiev, 123-131 (in Russian).

Taraskin, A.F. (1974) : On the asymptotic normality of vector-valued stochastic integrals and estimates of a multidimensional diffusion process. Theory Prob. Math. Stat., 2, 209-224.

Wu, C.F.J. (1986) : Jackknife, bootstrap and other resampling methods in regression analysis. To appear in Ann. Statist.