CHAPTER 2
MAXIMUM LIKELIHOOD ESTIMATION FOR DEPENDENT OBSERVATIONS

2.1 INTRODUCTION

The method of ML is one of the classical methods of estimation. In the case of iid it is well known that under certain regularity conditions the ML estimate is the solution of the likelihood equation and is consistent, asymptotically normal, asymptotically efficient and asymptotically sufficient. Under certain conditions on conditional probability densities, in the case of dependent observations Bhat (1974) has proved that, the ML estimate is consistent, asymptotically efficient and asymptotically normal.

Prasad and Prakasa Rao (1976) have obtained the results relating to strong consistency, asymptotic normality and first-order efficiency of the ML estimate in the case of dependent random variables under a set of verifiable conditions. The conditions are expressed in terms of the probability density of individual observations, conditioned upon all past observations. As an application of these results Prasad and Prakasa Rao (1976) have proved the Bernstein Von-mises theorem, regarding the convergence of posterior density to normal.

In this Chapter, we have obtained certain results pertaining to consistency and asymptotic normality of the MLE for
dependent observations. Also, we have obtained some results on MLE when the transient parameters are present in the observations. In Section 2.2 we have proved the consistency of the ML estimate based on the regularity conditions made by Prasad and Prakasa Rao (1976). Also, we have obtained some conditions for weak consistency. Results pertaining to asymptotic normality of the MLE for dependent observations have been obtained in Section 2.3. Also, we have proved the consistency and asymptotic normality of MLE for dependent observations when transient parameters are present provided with some examples. In the case of estimating a parameter on the basis of observations which are iid or come from a stationary ergodic Markov chain (MC), it is known that the ML estimate enjoys certain optimality properties. In Section 2.4 we have obtained an analogous result for the much more complex case where the sample is from a general stochastic process. Also, we have shown that, the MLE produces the best asymptotic probability of concentration in symmetric intervals.

2.2 CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE

Under certain regularity conditions Billingsley (1961) has shown that the properties of consistency, asymptotic efficiency and asymptotic normality of the ML estimate continue to hold for ergodic Markov processes. He used Levy’s CLT for martingales to establish the results. Billingsley has pointed out
this theorem can be used to extend the properties of ML estimate for processes more general than the Markov process. This extension was simultaneously done by Silvey(1961), Bar-shalom(1971) and Bhat(1974).

We let, \( x^n = (x_1, x_2, ..., x_n) \) be a sequence of real-valued, possibly dependent random variables, with joint probability density function with respect to a \( \sigma \)-finite measure \( \mu^n \), given by,

\[
p(x_1, x_2, ..., x_n, \theta) = p(x^n, \theta)
\]

(2.2.1)
depending on a real parameter, \( \theta \in \Theta \), where \( \Theta \) is an open subset of the real line. Let \( \theta_0 \) be the true but unknown value of \( \theta \). Let \( \hat{\theta}(x^n) \) be the Borel measurable function used as an estimator of \( \theta \) when \( x^n \) is observed. The conditional probability density function of \( x_k \) given \( x^{k-1} \) is

\[
p_{k,\theta} = p(x_k, \theta | x^{k-1}) = p(x_k, \theta) / p(x^{k-1}, \theta)
\]

(2.2.2)
so that

\[
p(x^n, \theta) = \prod_{k=1}^{n} p_{k,\theta}
\]

which will be assumed to be continuous in \( \theta \). Let \( E_{k-1,\theta} \) and \( \sigma^2_{k-1,\theta_0} \) denote the conditional expectation and variance respectively, given \( x^{k-1} \) under \( p_{k,\theta_0} \).

Bar-Shalom(1971) and Bhat(1974), Prasad and Prakasa Rao(1976) have presented the conditions in terms of the probability density of individual observations, conditioned upon all past observations. Loeve(1963) results relating to strong law
of large numbers and CLT for dependent random variables have been used to obtain the strong consistency and asymptotic normality of the ML estimate respectively. As an application of these results by presenting the conditions Prasad and Prakasa Rao (1976) obtained the Bernstein Von - Mises theorem for dependent random variables, when the random variables are iid. This result was proved by Le Cam (1970), while obtaining some asymptotic properties of Bayes' estimates, they obtained the generalised version of this result for Markov processes.

The following are the regularity conditions made by Prasad and Prakasa Rao (1976).

C2-1. The parameter space $\Theta$ is an open interval of the real line.

C2-2. $\frac{\delta^i h_k(\theta)}{\delta \theta^i}, i = 1, 2$ exist for almost all $x^k$ and for all $\theta \in \Theta$ and continuous in $\theta$. For almost all $x^{k-1}$ and for all $\theta \in \Theta$.

C2-3. $E_{\theta} \left\{ \frac{\delta h_k(\theta)}{\delta \theta} | x^{k-1} \right\} = 0$ and $E_{\theta} \left\{ \frac{\delta^2 h_k(\theta)}{\delta \theta^2} | x^{k-1} \right\} = \sigma_k^2$

where $\sigma_k^2(\theta)$ is independent of $x_1, x_2, \ldots, x_{k-1}$.
C2-4. \[ \lim_{n \to \infty} n^{-1} \sum_{k=1}^{n} \sigma_k^2(\theta) = i(\theta) , \]

where \( i(\theta) \) is finite for all \( \theta \in \Theta \). Here the assumption is

\( i(\theta) \) is continuous and non-zero for \( \theta \in \Theta \).

C2-5. There exists a neighbourhood \( V(\theta) \), such that, for every \( \theta' \in V(\theta) \)

\[ \left| \frac{\delta^2 h_k(\theta')}{\delta \theta^2} - \frac{\delta^2 h_k(\theta)}{\delta \theta^2} \right| \leq |\theta' - \theta| G(x_k;\theta) \]

where \( G(x_k;\theta) \geq 0 \) is a function not depending on \( x_k \) and

\[ E_{\theta} G(x_k;\theta) \leq M, \] for all \( k \geq 1 \).

C2-6. For all \( \theta \in \Theta \),

\[ \begin{align*}
(1) & \quad \sum_{k=1}^{\infty} k^{-2} \text{Var}_\theta \left[ \frac{\delta h_k(\theta)}{\delta \theta} \right] < \infty \\
(II) & \quad \sum_{k=1}^{\infty} k^{-2} \text{Var}_\theta \left[ \frac{\delta^2 h_k(\theta)}{\delta \theta^2} \right] < \infty \quad \text{and} \\
(iii) & \quad \sum_{k=1}^{\infty} k^{-2} \text{Var}_\theta G(x_k;\theta) < \infty \\
\end{align*} \]

The following condition is weaker than C2-6.

C2-6a. For all \( \theta \in \Theta \),

\[ \begin{align*}
(1) & \quad \lim_{k \to \infty} k^{-1} \text{Var}_\theta \left[ \frac{\delta h_k(\theta)}{\delta \theta} \right] < \infty \\
\end{align*} \]

30
(ii) \( \lim_{k \to \infty} k^{-1} \text{Var}_{\theta} \left( \frac{\delta^2 h_k(\theta)}{\delta^2 \theta} \right) < \infty \). and

(iii) \( \lim_{k \to \infty} k^{-1} \text{Var}_{\theta} G(x_k; \theta) < \infty \)

C2-7. For all \( \theta \in \Theta \).

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \text{E}_{\theta} \left[ \frac{\delta h_k(\theta)}{\delta \theta} \right]^{2+\delta} < \infty
\]

for some \( \delta > 0 \), \( \delta \) independent of \( \theta \).

C2-8. Let for each \( \theta \in \Theta \) and any \( \varepsilon > 0 \),

\[ R_k(\theta, \varepsilon) \equiv \sup \left\{ (h_k(\theta') - h_k(\theta)) : |\theta' - \theta| \geq \varepsilon, \theta' \in \Theta \right\} \]

Then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \text{E}_{\theta} \left[ R_k(\theta, \varepsilon) \right] \text{ is a finite negative quantity}
\]

and \( \sum_{k=1}^{\infty} k^{-2} \text{Var}_{\theta} \left[ R_k(\theta, \varepsilon) \right] < \infty \).

Let \( \theta_0 \) denote the true parameter and \( P_0 = P_{\theta_0} \). Let \( H \) be a non-negative measurable function satisfying the following condition.

C2-9. There exist a number \( \varepsilon, 0 < \varepsilon < i_0 \) (\( i_0 = i(\theta_0) \)) such that

\[
M(\theta_0) = \frac{i_0}{2\pi} \int_{-\infty}^{\infty} H(t) \exp\left(-\left(i_0 - \varepsilon\right) t^2/2\right) dt
\]

is finite, and for every \( h > 0 \) and every \( \delta > 0 \)

\[
\exp(-\delta n) \int H \left( n^{1/2} t \right) \lambda(\theta_n + t n^{1/2}) dt \to 0 \text{ a.s } P_0 \text{ as } n \to \infty \text{ where } \theta_n \text{ denotes a MLE.} \]
C2-10. $\Delta$ is a prior probability measure on $(\Theta, \mathcal{B})$, where $\mathcal{B}$ is the $\sigma$-field of Borel subsets of $\Theta$. $\Delta$ has a density $\lambda$ with respect to the Lebesgue measure. The prior density $\lambda$ is continuous and positive in an open neighbourhood of the true parameter $\theta_0$.

Based on the above regularity conditions Prasad and Prakasa Rao (1976) have obtained the following lemmas.

**LEMMA 2.1.1**

If the random variables $x_k$, $k = 1, 2, \ldots, n$ are such that

$$\sum_{k=1}^{n} b_n^{-2} V(x_k) < \infty \text{ with } b_n \to \infty,$$

then as $n \to \infty$,

$$b_n^{-1} \sum_{k=1}^{n} \left\{ x_k - \mathbb{E}(x_k | x_1, \ldots, x_{k-1}) \right\} \to 0 \text{ a.s.}$$

**LEMMA 2.1.2**

Let $(x_k, k \geq 1)$ be a sequence of random variables. If the following conditions are satisfied, then

$$n^{-1/2} \sum_{k=1}^{n} x_k \to N(0,1),$$

where $s_n^2 = n^{-1/2} \sum_{k=1}^{n} \mathbb{E}(x_k^2)$ and

1. $\lim_{n \to \infty} n^{1+\delta/2} \sum_{k=1}^{n} \mathbb{E} \left| x_k \right|^{2+\delta} = 0$, for some $\delta > 0$

2. $\mathbb{E}(x_k | x_1, \ldots, x_{k-1}) = 0$ and

3. $\mathbb{E}(x_k^2 | x_1, \ldots, x_{k-1}) = \sigma_k^2$

where $\sigma_k^2$ is independent of $x_1, \ldots, x_{k-1}$.
Under the above regularity conditions, they have obtained the consistency and asymptotically normal of the MLE by using the above lemmas. Based on these conditions, Bhat (1974) has proved the following theorem.

**THEOREM 2.2.1**

If for an $\epsilon > 0$

$\lim_{n \to \infty} \left\{ -n \sum_{k=1}^{n} E_{k-1,0} \log p_{k,0^+\epsilon} \right\}$

and $\lim_{n \to \infty} \left\{ -n \sum_{k=1}^{n} E_{k-1,0} \log p_{k,0^-\epsilon} \right\}$ are finite,

and (ii) $\sum_{k=1}^{\infty} \left\{ \sigma^2_{k-1,0^+} \frac{(\log p_{k,0^+\epsilon} / k)}{k} \right\} < \infty$

and $\sum_{k=1}^{\infty} \left\{ \sigma^2_{k-1,0^-} \frac{(\log p_{k,0^-\epsilon} / k)}{k} \right\} < \infty$

then the probability of the inequality

$\frac{1}{n} \left\{ n \sum_{k=1}^{n} \log p_{k,0^+\epsilon} - n \sum_{k=1}^{n} \log p_{k,0^-\epsilon} \right\} < 0$

tends to one as $n \to \infty$.

With a minor modification of conditions in Theorem 2.2.1, we now state and prove the following for the consistency of ML estimate.

**THEOREM 2.2.2**

Let $\log p_{k,0}$ be differentiable in an interval including the true value, for all $k$. Then the maximum-likelihood equation has a root with probability one as $n \to \infty$ which is consistent for $\theta$. 

33
PROOF:

As \( n \to \infty \), for almost all sample sequences \( l(x|\theta) \) will be greater at \( \theta_0 \) than at \( \theta_0 + \epsilon \). Since \( l(x|\theta) \) is continuous in \( (\theta_0 + \epsilon) \), there is a local maximum of \( l(x|\theta) \) within \( \theta_0 + \delta \). If \( l(x|\theta) \) is differentiable, its derivative must vanish at that point. Since \( \epsilon \) is arbitrary a root \( \hat{\theta} \) so located is consistent for \( \theta_0 \).

We now proceed to prove MLE to be consistent based on Wald (1949) and Doob (1934), by adopting the asymptotic normality approach. Here, it may be remembered that the derivatives of the likelihood function as done in Cramer (1946) are not required.

Let the observations, \( x_0, x_1, \ldots, x_n \) have known conditional probability density functions \( f(x_t|\Theta_{t-1}; \theta) \) where \( \theta \in \Theta \), an open subset of \( \mathbb{R}^k \), and \( \Theta_{t-1} \) is the \( \sigma \)-field generated by \( x_0, x_1, \ldots, x_{t-1} \). The log-likelihood function, conditional on \( x_0 \) is given by

\[
\ell_n(\theta) = \sum_{t=1}^{n} \ln f(x_t|\Theta_{t-1}; \theta).
\]

A solution of the likelihood equation \( \ell'_n(\theta) = 0 \) will be denoted by \( \hat{\theta}_n \), and the true parameter by \( \theta_0 \). Based on Taylor series expansion up to two terms,

\[
\ell'_n(\theta) = \ell'_n(\theta_0) + \ell''_n(\theta_0, \theta)(\theta - \theta_0)
\]

(2.2.5)

Where \( \ell''_n(\theta_0, \theta) \) denoted the matrix of order \( n \times n \) of second derivatives of \( \ell_n(\theta) \) with respect to \( \theta \), with rows evaluated at possibly different points on the line segments between \( \theta_0 \) and \( \theta \).
We now proceed as follows:

For weak consistency a sufficient condition is given which requires the matrix \(-B_n^{-1/2} I_n'(\theta_0, \theta)\) to tends to infinity in some sense whenever \(\theta\) is near \(\theta_0\), \(B_n\) being the Fisher information matrix. This may be interpreted that, information on \(\theta_0\), contained in \(I_n''\), steadily accrues as \(n \to \infty\) and that, there is sufficient continuity of \(-I_n''(\theta_0, \theta)\) in \(\theta\) near \(\theta_0\) for this function to resemble \(B_n = E [-I_n''(\theta_0, \theta)]\). The asymptotic normality of \(\theta_n\) this condition is augmented by \(-B_n^{-1} I_n''(\theta_0, \hat{\theta}_n) \to I_k\) in probability by \(k \times k\) unit matrix, after that a martingale CLT is applied.

The standard regularity conditions are,

\[
E[I'_n(\theta_0)] = 0, \quad V[I'_n(\theta_0)] = E[-I''_n(\theta_0, \theta_0)] = B_n
\]

(2.2.6)

The information matrix \(B_n\) is assumed to be positive definite.

From (2.2.5), we have

\[
B_n^{-1/2} I_n'(\theta) = B_n^{-1/2} I_n'(\theta_0) + B_n^{-1/2} I_n''(\theta_0, \theta) (\theta - \theta_0)
\]

(2.2.7)

the existence is assumed of \(\Delta > 0\) and some sequence \(\{c_n\}\) tending to infinity such that, when \(|\theta - \theta_0| = \delta \leq \Delta\),

\[
P(-c_n^{-1/2} (\theta - \theta_0)^T B_n^{-1/2} I_n''(\theta_0, \theta)(\theta - \theta_0) \geq \delta^2) \to 1 \text{ as } n \to \infty
\]

(2.2.8)

To prove this, it may be noted that,

\[
c_n^{-1/2} B_n^{-1/2} I_n'(\theta_0) \to 0
\]

(2.2.9)

is an immediate consequence of
which follows from (2.2.6). We define,
\[ g_n = c_n^{-1/2} B_n^{-1/2} I_n'(\theta_0) + c_n^{-1/2} B_n^{-1/2} I_n''(\theta_0, \theta_0) (\theta - \theta) \]  
(2.2.11)

It can be seen that combining (2.2.8) and (2.2.9), we get
\[ P[\theta - \theta_0, \nabla g_n < 0] \to 1 \quad \text{as} \quad n \to \infty, \]
for \( |\theta - \theta_0| = \delta \), given \( 0 < \delta \leq \Delta \)  
(2.2.12)

using an equivalent of Browner's fixed point theorem as in Aitchison and Silvey (1958), with the probability tending to 1 as \( n \to \infty \), \( g_n \) has a zero at \( \hat{\theta}_n \), say with \( |\hat{\theta}_n - \theta_0| < \delta \). Thus, there exists a weakly consistent solution \( \hat{\theta}_n \) of the likelihood function. Also it may be possible to derive a bound for the mean-square error of \( \hat{\theta}_n \). In cases, where \( -I_n''(\theta, \theta) \) is positive definite throughout \( \Theta \), with probability 1, \( \hat{\theta}_n \) will be unique and will give a true maximum of the likelihood function.

2.3 ASYMPTOTIC NORMALITY OF THE ML ESTIMATE

Consider the distribution function of \( \hat{\theta}_n \) which satisfies
\[ (2.2.3), \]
\[ \hat{\theta}_n - \theta_0 = -I_n''(\theta_0, \hat{\theta}_n)^{-1} B_n B_n^{-1} I_n'(\theta_0) \]  
(2.3.1)

Here, we assume that \( I_n''(\theta_0, \hat{\theta}_n) \) is non-singular, then
\[ B_n^{-1} I_n''(\theta_0, \hat{\theta}_n) \to I_k \]  
(2.3.2)

In which, the case \( I_n''(\theta_0, \hat{\theta}_n) \) is non-singular with probability one and the problem is simplified to dealing with \( B_n^{-1} I_n'(\theta_0) \) by a
matrix version of the convergence theorem of Cramer (1946). By the
Cramer and Wold (1936), $Z_n = z^T I_n'(\theta_0)$ for arbitrary unit $z$.

Let,

$$
U_t = \frac{\delta}{\delta \theta} \ln f(x_t | \theta_{t-1} ; \theta_0),
$$

$$
V_t = \frac{\delta^2}{\delta \theta^2} \ln f(x_t | \theta_{t-1} ; \theta_0),
$$

so that, $I_n'(\theta_0) = \sum_{t=1}^{n} U_t$, $I_n''(\theta_0, \theta_0) = \sum_{t=1}^{n} V_t$

From (2.2.4),

$$
E [U_t | B_{t-1}] = 0 \text{ and } V [U_t | B_{t-1}] = E [-V_t | B_{t-1}]. \quad (2.3.3)
$$

Let $X_n = Z_n - Z_{n-1} = z^T U_n$, then $E [U_n | B_{n-1}] = 0$
so $Z_n$ is a martingale with respect to the $\theta_0$ distribution.

Also, $E [X_n^2 | B_{n-1}] = z^T E [U_n U_n^T | B_{t-1}] z = z^T E [-V_n | B_{n-1}] z \quad (2.3.4)$
so $s_n^2 = E[Z_n^2] = \sum_{t=1}^{n} E[X_t^2] = z^T E \left[ \sum_{t=1}^{n} V_t \right] z = z^T B_n z \quad (2.3.5)$

A substantial solution to the problem is contained in
the CLT for martingales given by Brown(1971) and Scott(1973). An
equivalent set of sufficient conditions are given below.

Considering the result due to Brown(1971),
that is, $Z_n s_n^{-1} \rightarrow N(0,1)$

if $s_n^{-2} \sum_{t=1}^{n} E \left[ X_t^2 | B_{t-1} \right] \rightarrow 1 \text{ in probability} \quad (2.3.6)$

and $s_n^{-2} \sum_{t=1}^{n} E \left[ X_t^2 I \{X_t \geq \varepsilon s_n \} \right] \rightarrow 0$, for every $\varepsilon > 0.$ \quad (2.3.7)
where $I\{A\}$ denotes the indicator function of the set $A$. The above equation (2.3.7) is the Lindeberg condition which ensures that the contribution of any individual $X_t$ is asymptotically negligible.

Thus, in large samples, we note that, $B_n^{-1} I_n'(\theta_0)$ is approximately distributed as normal with mean 0 and variance $B_n^{-1}$ and the same may be said of $\hat{\theta}_n - \theta_0$.

It is thus observed that, for weak consistency of $\hat{\theta}_n$, besides the routine regularity conditions, (2.2.6) is sufficient. (2.3.2), (2.3.6) and (2.3.7) describes the asymptotic normality.

Now, consider the case, MLE when transient parameters are present.

There are cases in which some parameters ($\theta$) in a model are consistently estimable be MLE and the rest ($\phi$) may not be estimable. In generally, in a model containing $k+r$ independent parameters only $k$ independent parametric functions may have the MLE consistency property. The log-likelihood function, conditional on $Y_0$, is now $l_n(\theta, \phi)$, where the parameter $\theta \in \Theta, \phi \in \Phi$, open subsets of $R^k, R^r$ respectively. The likelihood equations are

$$I_{(0)}(\theta, \phi) = 0, \quad \text{and} \quad I_{(\phi)}(\theta, \phi) = 0, \quad (2.3.8)$$

$I_{(0)}$, $I_{(\phi)}$ denoting the first derivative of $l_n(\theta, \phi)$ with respect to $\theta(\phi)$. Second derivative matrices are denoted by $I_{\theta\theta}$, $I_{\theta\phi}$, $I_{\phi\theta}$, $I_{\phi\phi}$ and the corresponding Fisher information submatrices by $B_\Theta$ etc. The full $(k+r) \times (k+r)$ information matrix is $B_n$. Conditions are found under which the $\theta$ likelihood equations, $I_{(\theta)}(\theta, \phi) = 0,$
gives a solution \( \hat{\theta}_n \) weakly consistent for \( \theta_0 \) true parameter. We require that \( \hat{\theta}_n \) is not too much disrupted by the lack of a consistent estimator for \( \phi_0 \), and that the property of \( I_\theta \) corresponding to (2.2.8) is not destroyed by a contribution from \( I_{\theta\phi} \). Further conditions are given under which \( \hat{\theta}_n \) is asymptotically normal. Therefore, a connection between the consistency of a MLE and the tendency to infinity of the information on that parameter. If the corresponding diagonal element of \( R_n \) does not tend to \( \infty \) as \( n \to \infty \) we will call the parameter "transient". In this case it follows from the Cramer-Rao bound that no transient parameter can possess an estimator consistent in the mean-square sense. This concept is refined by the examples.

We now proceed to prove the Consistency and asymptotic normality of \( \hat{\theta}_n \).

The Taylor expansion of the first derivative \( I_n' \) of the log likelihood function \( l_n \) with respect to \((\theta, \phi)\) about the population parameter \((\theta_0, \phi_0)\) is given by,

\[
I_n'(\theta, \phi) = I_n'(\theta_0, \phi_0) + I_n''(\theta_0, \theta, \phi_0, \phi) \left[ \begin{array}{c} \theta - \theta_0 \\ \phi - \phi_0 \end{array} \right]
\]

(2.3.9)

where \( I_n''(\theta_0, \theta, \phi_0, \phi) = \begin{bmatrix} I_{\theta\theta} & I_{\theta\phi} \\ I_{\phi\theta} & I_{\phi\phi} \end{bmatrix} \) is the \((k+r) \times (k+r)\) matrix of second derivatives of \( l_n \) with rows evaluated at points strictly between \((\theta_0, \phi_0)\) and \((\theta, \phi)\).
The $\theta$-likelihood equations are obtained by equating the first $k$ components of (2.3.9) to zero. A solution $\hat{\theta}_n$ will in general depend on $\phi$ and $\phi_0^*$. We seek a condition under which, for arbitrarily small $\delta > 0$, a $\hat{\theta}_n$ exists in probability with

$$|\hat{\theta}_n - \theta_0| \leq \delta, \quad \phi = \phi_0$$

be a sequence and irrespective of the unknown value of $\phi_0$. Then, multiplication by $B_{\theta\theta}^{-1/2}$ the $\theta$-likelihood equation becomes,

$$B_{\theta\theta}^{-1/2} I_{\theta}(\theta_0, \phi_0) + B_{\theta\theta}^{-1/2} I_{\theta\theta}(\theta_0, \theta, \phi_0, \phi)(\theta_0 - \theta)$$

$$+ B_{\theta\theta}^{-1/2} I_{\theta\phi}(\theta_0, \theta, \phi_0, \phi)(\phi_0 - \phi) = 0 \quad (2.3.10)$$

A sufficient criterion for what is sought is that, one can find $\Delta > 0$, sequence $\{c_n\}$ with $c_n \to \infty$, sequence $\{\phi_n\}$ such that,

when $|\theta - \theta_0| = \delta \leq \Delta$, where $\delta > 0$. Therefore,

$$P \left[ B_{\theta\theta}^{-1/2} (\theta - \theta_0)^T I_{\theta\theta}^{-1/2} \left\{ I_{\theta\theta}(\theta_0, \theta, \phi_0, \phi)(\theta - \theta_0) + I_{\theta\phi}(\theta_0, \theta, \phi_0, \phi)(\phi - \phi_0) \right\} \right] \leq \delta^2 \to 1$$

as $n \to \infty$. \quad (2.3.11)

For asymptotic normality of $\hat{\theta}_n$, From (2.3.10) we have,

$$\hat{\theta}_n - \theta_0 = - I_{\theta\theta}(\theta_0, \phi_0^*, \phi_0^*)^{-1} B_{\theta\theta} B_{\theta\theta}^{-1} I_{\theta}(\theta_0, \phi_0)$$

$$+ I_{\theta\theta}(\theta_0, \theta, \phi_0, \phi)(\phi_0 - \phi_0) = 0$$

(2.3.12)

If

$$- B_{\theta\theta}^{-1} I_{\theta\theta}(\theta_0, \theta, \phi_0, \phi_0) \xrightarrow{p} I_k \quad (2.3.13)$$

$$B_{\theta\theta}^{-1} I_{\theta\phi}(\theta_0, \theta, \phi_0, \phi)(\phi_0 - \phi_0) \xrightarrow{p} 0 \quad (2.3.14)$$

40
it only remains to find conditions under which $B^{-1}_{\theta\theta} I_0(\theta_0, \phi_0)$ is asymptotically normal.

We now provide the following examples to describe the transience.

**EXAMPLE 2.3.1**

In this example, we consider the regression model where the $Y_i$ are independent $N(\theta_0 + \phi_0 x_i)$.

$$B_{\theta\theta} = -1_{\theta\theta}(\theta, \phi) = n,$$
$$B_{\theta\phi} = -1_{\theta\phi} = \sum_{i=1}^{n} x_i$$
and $B_{\phi\phi} = -1_{\phi\phi}$

and $(\hat{\theta}_n, \hat{\phi}_n)^T = B^{-1}_n(\Sigma_t^t, \Sigma x_t y_t)^T$, so $(\hat{\theta}_n, \hat{\phi}_n)$ are joint distribution of $N(\theta_0, \phi_0)$.

CASE (i) : $x = t^\alpha$ for given $\alpha > 1/2$.

Since $\sum x_i^2 = \Sigma t^{-2\alpha} \rightarrow \infty$, $\phi$ is transient. For consistency we check the condition given in (2.3.11), we have to verify that, for some $c_n \rightarrow \infty$ the criterion expression

$$E_n = c_n^{-1/2} (\theta - \theta_0) n^{-1/2} \{ -n(\theta - \theta_0) - \Sigma t^{-\alpha}(\phi_n - \phi_0) \} < -\delta^2$$
in probability. Where $\delta = |\theta - \theta_0|$

But, $E_n = -\left(\frac{n}{c_n}\right)^{1/2} \delta^2 \left\{ 1 + \frac{\phi_n - \phi_0}{\theta - \theta_0} \frac{1}{n} \Sigma t^{-\alpha} \right\} \rightarrow -2\delta^2$

as $n \rightarrow \infty$ by choosing $\phi_n = 0$ and $c_n = n/4$.

In fact since $E\left[|\hat{\theta}_n - \theta_0|^2\right] = \Sigma t^{-2\alpha} / d_n$

and $\hat{\phi}_n$ is normal with mean $\phi_0$ and variance $n d_n$, where $d_n = n \Sigma t^{-2\alpha} - (\Sigma t^{-\alpha})^2$, it is easily verified
directly that \( \hat{\theta}_n \) is mean-square consistent and \( \hat{\phi}_n \) is not even weakly consistent.

CASE (ii) : Suppose \( x_t = t^{-\alpha} \) for given \( 0 < \alpha < 1/2 \). Although the effect of \( \phi \) on the data tends to zero as \( t \to \infty \), \( \phi \) is not transient. However, the criterion (2.3.11) may still be used to show that \( \hat{\theta}_n \) is consistent. The general pattern is that in cases where it is difficult to show that certain parameters are consistent it may be easier to demonstrate this property for a subset of them, regarding the rest as \( \phi \)'s and using (2.3.11) this partial solution to the problem may be adequate for the application in hand.

CASE (iii) : Suppose that \( x_t = 1 + t^{-\alpha} \), for given \( \alpha > 1/2 \).

Neither \( \theta \) and \( \phi \) are consistent. However,

\[
V(\hat{\theta}_n) = \sum (1 + t^{-\alpha})^2/d_n \quad \text{and} \quad V(\hat{\phi}_n) = n/d_n
\]

Hence as \( n \to \infty \) both variances tends to zero. That is,

\[
\left[ \sum -2\alpha \right]^{-1} \quad \text{tends to} \quad 0. \quad \text{So neither MLE is consistent.}
\]

This may be as follows : For unit \( z \),

\[
s_n^2 = z^T B_n z = n(z_1 + z_2)^2 + 2z_2(z_1 + z_2) \sum t^{-\alpha} + z_2^2 \sum t^{-2\alpha}.
\]

In this direction we have \( z_1 + z_2 = 0 \), and \( s_n^2 \to \infty \), so transience exists in one dimension of the two-dimensional parameter space. This may be removed by re-parameterizing that \( (\theta, \phi) = h(\psi_1, \psi_2) \). Where

\[
h = (h_1, h_2), \quad \text{a vector function with 2x2 Jacobian matrix,}
\]

\[
(h')_{ij} = \delta h_j / \delta \psi_i. \quad \text{In terms of the new parameters the information matrix is} \quad B_n^\psi = h'B_n h'^T \quad \text{and, the inspection of 2x2}
\]

42
element, it is found that $\psi_2$ is transient if $(\delta h_1/\delta \psi_1) + (\delta h_2/\delta \psi_2) = 0$. One such information is $\psi_1 = \theta + \phi$, $\psi_2 = \phi$, when the model $Y_i$'s independent of $N(\psi_1+\psi_2, t^{-\alpha}, 1)$ as in the first case. Although in the present case such a chance of parameters is obvious from the start the procedure demonstrates a general method. These results are reported in Muthukrishnan(1997a).

We now proceed to furnish certain results on asymptotic properties of MLE for Stochastic processes.

2.4 ASYMPTOTIC PROPERTIES OF MLE FOR STOCHASTIC PROCESSES

In the case of estimating a parameter on the basis of observations which are independent and identically distributed or come from a stationary ergodic Markov chain, it is known that the MLE enjoys certain optimality properties. In this Section we have obtained an analogous result for the much more complex case where the sample is from a general stochastic process. Also we have to show that the MLE produces the best asymptotic probability of concentration in symmetric intervals.

Consider a sample $Y_1, Y_2, \ldots, Y_n$ of consecutive observations from some stochastic process whose distribution depends on a single parameter, $\theta$. Where $\theta \in \Theta$, $\Theta$ being an open interval. Let $L_n(\theta)$ be the likelihood function associated with $Y_1, Y_2, \ldots, Y_n$ and suppose the $L_n(\theta)$ is differentiable with respect to $\theta$ and

$$E_{\theta} \left[ \frac{d \log L_n(\theta)}{d \theta} \right]^2 < \infty \text{ for each } n.$$
Suppose in addition that if \( P(Y_1, Y_2, \ldots, Y_n) = L_n(\theta) \) is the joint probability density function of \( Y_1, Y_2, \ldots, Y_n \). We define \( F_k \) for the \( \sigma \)-field generated by \( Y_1, Y_2, \ldots, Y_k \), \( k \geq 1 \), take \( F_0 \) as the trivial \( \sigma \)-field and set \( L_0 = 1 \). Then we have,

\[
\frac{d \log L_n(\theta)}{d\theta} = \sum_{i=1}^{n} u_i(\theta)
\]

also,

\[
E_\theta[u_i(\theta)|F_{i-1}] = 0 \text{ almost surely.}
\]

so that \( \left\{ \frac{d \log L_n(\theta)}{d\theta}, F_n, n \geq 1 \right\} \) is a square integrable, that is a martingale. In addition, we have define,

\[
I_n(\theta) = \sum_{i=1}^{n} E_\theta[u_i^2(\theta)|F_{i-1}]
\]

and noted that, under the conditions imposed above,

\[
v_1(\theta) = \frac{d u_1(\theta)}{d\theta}
\]

which satisfies,

\[
E_\theta[u_1^2(\theta)|F_{i-1}] = -E_\theta[v_1(\theta)|F_{i-1}] \text{ almost surely.}
\]

Also, we have

\[
J_n(\theta) = \sum_{i=1}^{n} v_i(\theta)
\]

we noted that \( \{J_n(\theta) + I_n(\theta), F_n, n \geq 1 \} \) is a martingale. The quantity \( I_n(\theta) \) is a form of conditional information which reduces to the standard Fisher information in the case where the \( Y_i \)'s are independent random variables. The role of \( I_n(\theta) \) in the stochastic process estimation context is a vital one. The importance of \( I_n(\theta) \) is easily seen from the following expansion.
We suppose that, \( \theta \) is the true parameter value. Then, we can use Taylor's expansion to write for \( \theta' \in \Theta \),

\[
\begin{align*}
\frac{d \log L_n(\theta')}{d\theta'} &= \sum_{i=1}^{n} u_i(\theta') \\
&= \sum_{i=1}^{n} u_i(\theta) + (\theta' - \theta) \sum_{i=1}^{n} v_i(\theta^*) \\
&= \sum_{i=1}^{n} u_i(\theta) - (\theta' - \theta) I_n(\theta) \\
&\quad + (\theta' - \theta) (J_n(\theta^*) + I_n(\theta)) \tag{2.4.1}
\end{align*}
\]

where \( \theta^*_n = \theta + v(\theta' - \theta) \) with \( v = v(n, \theta) \) satisfying \( |v| < 1 \).

Further, since \( (I_n(\theta))^{-1} \sum_{i=1}^{n} u_i(\theta) \xrightarrow{a.s.} 0 \) a.s.

provided \( I_n(\theta) \xrightarrow{a.s.} \infty \) as \( n \to \infty \),

we see that the likelihood equation has a root \( \hat{\theta}_n \) which is strongly consistent for \( \theta \) if \( I_n(\theta) \xrightarrow{a.s.} \infty \) and

\[
\lim \sup_{n \to \infty} |I_n(\theta) + J_n(\theta^*)| < 1 \text{ a.s.}
\]

for \( |\theta^*_n - \theta| < \delta \) sufficiently small.

In order to utilize most effectively the MLE from large samples it is necessary for \( \sum_{i=1}^{n} u_i(\theta) \) to converge in distribution, when appropriately normalized, to some proper limit law. Furthermore, it can be arranged in many cases that this limit law is normal, which is most convenient for confidence interval purposes. The most effective general norming seems to be provided
by \( I_n^{1/2}(\theta) \) and \([I_n(\theta)]^{-1/2} \sum_{i=1}^{n} u_i(\theta)\) converges in distribution to normality under quite wide ranging circumstances. Suppose that,

\[
[I_n(\theta)]^{-1/2} \sum_{i=1}^{n} u_i(\theta) \xrightarrow{d} N(0,1).
\]

In addition to the above conditions, \( J_n(\theta^{**}) = -I_n(\theta) (1+O(1)) \) in probability as \( n \to \infty \) for \( \theta^{**} = \theta + \xi(\hat{\theta} - \theta) \) with any \( \xi \) satisfying \( |\xi| \leq 1 \), we have from (2.4.1) that

\[
I_n^{1/2}(\theta) (\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1).
\]

Then, if \( T_n = T_n(Y_1, \ldots, Y_n) \) is any consistent estimator of \( \theta \) for which \( I_n^{1/2}(\theta) (T_n - \theta) \xrightarrow{d} N(0, \beta^2(\theta)) \) where \( \beta(\theta) \) is bounded and continuous in \( \theta \). The MLE is optimal within this class since \( \beta(\theta) = 1 \) when \( T_n \) is the MLE. It is desirable if comparisons can be made between the MLE and those other consistent estimators for which \( I_n^{1/2}(\theta) (T_n - \theta) \) converges in distribution to a proper law or perhaps does not even converge. Such comparisons have been made in the case where the \( \{Y_i\} \) form a stationary ergodic Markov chain by using the concept of asymptotic efficiency in the Wolfowitz sense. Minor modifications of the standard theory as presented in Roussas(1972) leads to comparisons of the kind

\[
\lim_{n \to \infty} P_{\theta} \left\{ -c < (E_{\theta} I_n(\theta)^{1/2} (\hat{\theta}_n - \theta) < b \right\}
\]

\[
= \lim_{n \to \infty} \sup_{\theta} P_{\theta} \left\{ -c + w(\theta) < (E_{\theta} I_n(\theta)^{1/2} (T_n - \theta) < W(\theta) + b \right\}
\]

for arbitrarily positive \( b \) and \( c \) and certain \( W(\theta) \geq w(\theta) \). This
inequality holds under conditions that,
\[ I_n(\theta) (E_{\theta} I_n(\theta))^{-1} \overset{P}{\to} 1 \quad\text{as}\quad n \to \infty. \]

It is however, the cases where this last condition is not satisfied that pose the real interest and challenge in the treatment of stochastic process estimation. Our arguments in this section principally concerned with the commonly occurring circumstances under which
\[ I_n(\theta) (E_{\theta} I_n(\theta))^{-1} \overset{P}{\to} \gamma(\theta) \quad\text{as}\quad n \to \infty, \quad \gamma \text{ being a random variable in general. These difficulties can be avoided if we restrict consideration to symmetric concentration intervals. Then using of Weiss and Wolfowitz (1966) we proved that} \]
\[ \lim_{n \to \infty} P_{\theta} \left\{ -c < (E_{\theta} I_n(\theta))^{1/2} (\hat{\theta}_n - \theta) < b \right\} \]
\[ \geq \lim_{n \to \infty} \sup_{\theta} P_{\theta} \left\{ -c < (E_{\theta} I_n(\theta))^{1/2} (T_n - \theta) < c \right\} \]
for all \( c > 0. \)

We now state and prove the following theorem.

**THEOREM 2.4.1**

Let \( \theta_n = \theta + 2C (E_{\theta} I_n(\theta))^{-1/2}, \quad C > 0 \) and write \( P_{\theta}^{(n)} \) for the probability measure corresponding to the likelihood \( L_n(\theta). \) Suppose that \( T_n \) is any estimator of \( \theta \) which satisfies the condition that for any \( \theta \in \Theta, \)
\[ \lim_{n \to \infty} P_{\theta}^{(n)} \left\{ (E_{\theta} I_n(\theta))^{-1/2} (T_n - \theta) > -C \right\} \]
\[ - P_{\theta}^{(n)} (E_{\theta} I_n(\theta))^{-1/2} (T_n - \theta_n) - C) \right\} = 0 \]
If there is a MLE \( \hat{\theta}_n \) which is consistent for any \( \theta \in \Theta \), then
\[
\lim_{n \to \infty} P_{\theta}^{(n)} \left\{ -C < ((E_\theta I_n(\theta)))^{-1/2} (\hat{\theta}_n - \theta) < C \right\} \\
\geq \lim_{n \to \infty} \sup P_{\theta}^{(n)} \left\{ -C < ((E_\theta I_n(\theta)))^{-1/2} (T_n - \theta) < C \right\}
\]
Furthermore, if \( I_n^{1/2}(\theta) (T_n - \theta) \overset{d}{\longrightarrow} N(0, \gamma^2(\theta)) \), continuously in \( \theta \), where \( \gamma(\theta) \) is bounded, then \( \gamma^2(\theta) \geq 1 \).

**PROOF:**

Suppose that \( (x_1, \ldots, x_n) \) possesses a density \( P_n(x_1, \ldots, x_n) \) i.e., \( L_n(\theta) \), which is continuous in \( \theta \), with respect to a \( \sigma \)-finite measure \( \mu_n \). We now make the following two assumptions.

**C2-11.**

(i) \( I_n(\theta) \overset{a.s.}{\longrightarrow} \omega, \quad I_n(\theta)/E_\theta I_n(\theta) \overset{P}{\longrightarrow} \eta(\theta) \quad (> 0 \text{ a.s.}) \) for some random variable \( \eta \), and \( J_n(\theta)/I_n(\theta) \overset{P}{\longrightarrow} -1 \) as \( n \to \infty \), the convergences in probability being uniform in compacts of \( \theta \).

(ii) \( [I_n(\theta)]^{-1/2} d \log L_n(\theta)/d\theta \longrightarrow N(0,1) \), continuously in \( \theta \).

**C2-12.**

For \( \delta > 0 \), suppose \( |\hat{\theta}_n - \theta_0| \leq \delta/(E_\theta I_n(\theta_0))^{1/2} \). Then,

(i) \( E_{\hat{\theta}_n} I_n\theta_0 = E_{\theta_0} I_n\theta_0 (1 + o(1)) \) as \( n \to \infty \)

(ii) \( I_n(\theta) = I_n(\theta_0) (1 + o(1)) \text{ a.s. as } n \to \infty \), and

(iii) \( J_n(\theta) = J_n(\theta_0) + o(I_n(\theta_0)) \text{ a.s. as } n \to \infty \).

The proof the theorem is based on Weiss and Wolfowitz (1966).

From (2.4.1) we have,
\[
\sum_{i=1}^{n} u_i(\theta) + (\hat{\theta}_n - \theta)^* J_n(\hat{\theta}_n) = 0 \quad (2.4.2)
\]

where \( \hat{\theta}_n = \theta \gamma(n,\theta) (\theta_n - \theta) \) with \( |\gamma(n,\theta)| < 1 \).

By assumptions C2-11 and C2-12 gives,

\[
\lim_{n \to \infty} P_{\theta}^{(n)} (\sqrt{n} \gamma(\theta_n - \theta) < \beta) = \lim_{n \to \infty} P_{\theta}^{(n)} (\sqrt{n} \gamma(\theta_n - \theta) < \beta) = P (\eta^{1/2}(\theta) N(0,1) < \beta) \quad (2.4.3)
\]

for any \( \beta, -\infty < \beta < \infty \), where \( \eta \) and \( N \) are independent.

Let, \( \phi_n = \log \frac{L_n(\theta)}{L_n(\theta_n)} \) with \( \{L_n(\theta_n) / L_n(\theta) > 0\} \) and \( \phi_n \) is arbitrarily defined on \( \{L_n(\theta_n) / L_n(\theta) = 0\} \). Then by Taylor's expansion we have,

\[
\phi_n = (\hat{\theta}_n - \theta)^* \sum_{i=1}^{n} u_i(\theta) + \frac{1}{2} (\hat{\theta}_n - \theta)^2 J_n(\hat{\theta}_n)
\]

for \( \theta_n^* \in [\theta, \theta_n] \), and using (2.4.2), we have,

\[
\phi_n = -2 C (E_{\theta} I_n(\theta))^{-1/2} (\hat{\theta}_n - \theta)^* J_n(\hat{\theta}_n)
\]

\[
+ 2 C^2 (E_{\theta} I_n(\theta))^{-1} (\hat{\theta}_n - \theta) J_n(\hat{\theta}_n)
\]

thus by assumptions C2-11 and C2-12, we have,

\[
\{\phi_n < 0\} = \{\hat{\theta}_n < \theta + C(E_{\theta} I_n(\theta))^{-1/2} (1 + o_p(1))\}
\]

Here, \( o_p(1) \) represents the term which tends in probability to zero as \( n \to \infty \), while,

\[
\lim_{n \to \infty} P_{\theta}^{(n)} (\phi_n = 0) = \lim_{n \to \infty} P_{\theta}^{(n)} (\phi_n = 0) = 0
\]
The second part of the theorem, we note that the mixing form of convergence results together with assumption C2-11, we have,

$$
\lim_{n \to \infty} P^{(n)}_{\theta} (-C < (E_{\theta} I_{\theta}^{(n)}(\theta))^{1/2} (T_{n} - \theta) < C) = P(-C < \eta^{1/2}(\theta)N(0,1) < C)
$$

where $\eta$ and $N$ are independent. Hence obtained the results by (2.4.3).

Various structural requirements need to be imposed on the process $\{x_{i}\}$ in order to allow the various assumptions to be checked. The imposition of stationarity, for example, leads to useful simplification in the requirements. However, we shall here restrict attention to the case where the stochastic process $\{x_{i}\}$ is a time - homogeneous Markov process whose distribution belongs to a conditional exponential family.

The concept of a conditional exponential family is a generalization of that of the exponential family for the case of independent $X_{1}$. For $\{X_{1}\}$ belonging to a conditional exponential family,

$$
\frac{d \log L_{n}(\theta)}{d \theta} = I_{\theta}^{(n)}(\hat{\theta}_{n} - \theta) \tag{2.4.4}
$$

and $I_{\theta}^{(n)}(\theta) = \Phi(\theta) \sum_{i=1}^{n} H(X_{i-1})$ where $\Phi$ does not involve the $X_{i}$ and $H$ does not involve $\theta$. Furthermore, the result (2.4.4) holds if and only if the conditional probability density function of $X_{k}$ given $X_{k-1}$, $f(X_{k} | X_{k-1}, \theta)$ satisfies,
\[ \frac{d}{d\theta} \log f(x|y, \theta) = \Phi(\theta) H(y) \{ m(x,y) - \theta \} \]

where, \( E_{\theta} \{ m(X_1, X_{i-1}) | F_{i-1} \} = \theta \) a.s.

For the case of the conditional family we easily see that assumption C2-12 holds if \( \Phi(\theta) \) is continuous and differentiable.

As a particular example we shall consider the estimation of the mean \( \theta \) of the offspring distribution of a supercritical Galton-Watson branching process.

Here we have,
\[ 1 < \theta = E_{\theta}(X_1|X_0 = 1) \]
and we shall suppose that
\[ \sigma^2 = V_{\theta}(X_1|X_0 = 1) < \infty. \]

In this case it is known that the conditional exponential family consists of the family of power series distributions. These are distributions for which
\[ P_j = P(X_1 = j|X_0 = 1) = b_j \beta^j [f(\beta)]^{-1}, \quad j = 0, 1, 2, \ldots, \beta > 0. \]
where \( b_j \geq 0 \) and \( f(\beta) = \sum_{j=0}^{\infty} b_j \beta^j \). Then,
\[ \theta = \beta f'(\beta) [f(\beta)]^{-1}, \quad \sigma^2 = \left[ \frac{d}{d\theta} \log \beta \right]^{-1} \]
and
\[ d \left( \log f(x|y, \theta) \right) / \delta \theta = \sigma^{-2}(x-y\theta), \]
so that,
\[ d \left( \log L_n(\theta) \right) / \delta \theta = \sigma^{-2} Y_n \left( \frac{\hat{\theta}_n - \theta}{n} \right), \]
where \( Y_k = \sum_{j=0}^{k} X_j \) and \( \hat{\theta}_n = (Y_n - Y_0) Y_n^{-1} \).
For this family, $\theta = \theta(\beta)$ is known to be a non-negative monotone increasing function of $\beta$, so that the parametrization can be equally well expressed in terms of $\theta$.

Suppose $X_0 = 1$ and $P(X_0=0) = 0$ for definiteness and convenience. The latter gives $X_n \to \infty$ a.s. as $n \to \infty$.

If $P(X_1=0) > 0$, it is well known that $X_n \to \infty$ a.s. on the non-extinction set and the results which we shall describe hold conditionally on non-extinction. It follows from Heyde (1970) that,

$$Y_n (E_0 Y_n)^{-1} \to \eta(\theta) \text{ a.s. with } \eta \text{ non-degenerate and a.s. positive.}$$

Further,

$$I_n(\theta) = \sigma^{-2}(\theta) Y_{n-1}$$

and

$$J_n(\theta) = \frac{d\sigma^2(\theta)}{d \theta} = (Y_n - 1 - \theta Y_{n-1}) - \sigma^{-2}(\theta) Y_{n-1}^2,$$

and assumption C2-11 is satisfied since convergence in probability holds uniformly for $\theta \geq 1 + \epsilon$, any $\epsilon > 0$. Therefore, we have

$$\frac{1}{(E_0 Y_n(\theta))^{1/2}} \frac{d \log L_n(\theta)}{d \theta} \to N(0,1).$$

Therefore assumption C2-11 is satisfied. Also, we note that, $\sigma^2(\theta)$ is continuous in $\theta$ and differentiable so that assumption C2-12 holds. The application of this theorem in this context substantially clarifies Heyde's (1975) investigation of the class of estimators which are called asymptotically efficient for $\theta$. 

52