# ASYMPTOTIC THEORY OF ESTIMATION WHEN THE LIMIT OF THE LOG - LIKELIHOOD RATIOS IS MIXED NORMAL

P. Jeganathan



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

DOCTOR OF PHILOSOPHY

CALCUTTA

1980

Dedicated to

Professor B. Ramachandran

# A C K N O W L E D G E M E N T S

The present work was written under the supervision of Professor J. K. Ghosh.

In July 1979 - after having spent two and a half years at the Delhi campus of Indian Statistical Institute - I resumed work under the supervision of Professor J. K. Ghosh. I had just been through a critical period in my career - and it is not too much to say that without Professor Ghosh's kind help and encouragement I could never have recovered the strength and confidence to complete this work. He has been involved in every stage of preparation of the present work, and the present version owes very much to his careful reading and suggestions for improvements.

It is difficult to describe in this brief space how for the past two years Professor L. LeCam's pain-staking help and encouragement have contributed to the progress of this work. I will content myself by saying that the major part of this work could not have been done without his assistance.

As for Professor B. Ramachandran, this is not the place to go into the nature and extent of his association with this work. I am ever grateful to him.

It is a pleasure to express my gratitude to Professor K. R. Parthasarathy, Professor J. Sethuraman, and to Professor Divaker Sarma of Indian Institute of Technology, Kanpur, for their encouragement and for the interest they took in the progress of my research work.

I am thankful to several members of the Division of Theoretical Statistics and Mathematics for their kind help during the difficult time of my research work.

I have no intention of suppressing the positive aspects of the role played by Dr. B. L. S. Prakasa Rao - my supervisor during my two and a half years at Delhi campus - in my research work. It is a fact that I benefited from him during the initial stages of his supervision. More specifically, the problem treated in Chapter 8 was suggested by him in a specific form. He also read and corrected an unpublished manuscript written at the end of 1977. In 1977 and the early part of 1978 I felt he was genuinely interested in guiding me. During this period he spent a lot of time on me.

Finally, I wish to thank Mr. D. K. Bardhan for typing and Mr. Mukta L. Bag and Mr. Harish C. Prasad for cyclostyling of this work.

# C O N T E N T

## INTRODUCTION

In one of his fundamental papers LeCam (1960) introduced what is now called locally asymptotically normal (LAN) families of distributions and obtained several basic results regarding the asymptotic theory of estimation and testing. Roughly speaking, a sequence of families is said to satisfy the LAN condition if the corresponding sequence of appropriately normalised log-likelihood function is locally approximated with probability tending to one by the sum of two expressions, the first one being a sequence of random linear functions of the normalised parameter and the second one being a non-random quadratic form of the normalised parameter, and the sequence of random vectors involved in the linear term of the approximation converges weakly to the normal distribution with mean vector zero and the covariance matrix being the matrix involved in the quadratic form of the approximation. Actually LeCam (1960) considered a more general approximation in the sense that he allowed the above mentioned second term of the approximation to be any non-random function of the normalised parameter and the limit of the random vectors of the first term to be any arbitrary distribution and then he showed that if one further assumes the contiguity condition, which is impossible to avoid in a large part of statistical theory, the given families satisfy the LAN condition. This is indeed a remarkable result since it implies that if one could approximate by linearly indexed expential families one could also approximate by normal families.

An important thing to observe regarding the basic and " restricted" assumptions of LeCam (1960) is that a large part of asymptotic theory depends only on the approximating form of the likelihood function, and any specific property such as i.i.d or any other special form of dependence is irrelevent. Thus, to mention only a small fraction of the results of LeCam (1960), LeCam presented, under the LAN condition, a far-reaching geralisation of Wald's (1943) asymptotic theory of testing and showed that this testing problem can be simply treated as if it were regarding the normal distribution.

Based on LeCam (1960), more importantly based on the above mentioned observation, Hájek (1970, 1971 and 1972) further obtained several basic results regarding the asymptotic theory of estimation.

Though the LAN condition covers a large part of statistical theory associated with asymptotic normality, there are problems in which the assertions cannot be made in terms of asymptotic normality. Therefore, LeCam (1972 and 1974b) further developed his theory and obtained quite general and more forceful results in a more general framework which amount to the following. If one is interested in the asymptotic properties such as local asymptotic minimaxity and admissibility for the given sequence of families, it is just enough to obtain the results for the limit of the given sequence of families and then the corresponding limiting statements for the given sequence of families can be simply concluded from his results, even when the limits of the families are remote from the usual normal families. Thus, for example, the results of Hájek (1972) regarding local asymptotic minimaxity and admissibility for the LAN families can be viewed as particular cases of LeCam's more general results.

In recent times, there occur situations, e.g. in Galton-Watson branching processes and pure-birth process as has been discussed in e.g. Keiding (1974), Basawa and Scott (1976), Heyde and Feigin (1975), Heyde (1978) and Bhat (1978), in which LAN condition is not satisfied, but it can be seen that a quite similar and more general condition, which may be called locally asymptotically <u>mixed</u> normal (LAMN) condition, is satisfied. Roughly speaking, a sequence of families may be said to satisfy the LAMN condition if the corresponding sequence of appropriately normalised log-likelihood function is locally approximated, with probability tending to one, by the sum of two expressions, the first one being a sequence of random linear functions of the normalised parameter and the second one being a sequence of random quadratic forms of the normalised parameter, the sequence of random matrices involved in the quadratic forms being convergent weakly to an almost surely positive definite random matrix and the random vectors involved in the linear terms being convergent weakly to an appropriate mixed normal distribution.

In the first six chapters, which may be called the local part, of the present work, we present a detailed study of the LAMN families of distributions, and obtain some basic results which are the consequences of the LAMN condition. Our aim in these chapters is to extend, and if possible to strengthen, some of the basic results of LeCam and Hájek. We have also presented some results which have not been stated before even for the LAN case, even though they are implicit in the arguments of various works of LeCam and Hájek, at least for the LAN case.

It may be further noted here that we obtain our results directly without using the general results presented in LeCam (1972 and 1974b).

Before going into the details of the global part (Chapters 7 and 8) of the present work, we first give a brief summary of the first six chapters.

In Chapter 1, we introduce the precise definitions of LAMN families. We present several basic and preliminary results that will be frequently used in the chapters that follow. All the results of this chapter are almost essentially either contained or implicit in LeCam (1960) and Chapter 12 of LeCam (1974a) though the arguments of LeCam are intended for the LAN case. Using the results of Section 3 of this chapter, a method of constructing a specific sequence of estimators is presented in Section 4 and it is noted that this sequence together with a sequence of estimates of the random matrices of the LAMN condition form a sequence of asymptotically sufficient estimators in the sense of LeCam (1960); the specific method of construction given here is the same as the one given in LeCam (1960 and 1974a). The same method of construction for the LAMN case was earlier given by Davies (1979) under a further restriction on the sequence of random matrices of the LAMN condition. In Chapter 7 a detailed study of the traditional estimation procedures will be made and since these procedures yeild estimators that are also asymptotically sufficient in the sense of this section, we preferred to include the discussions of this tiny section 4.

In Chapter 2, a 'differentiability in quadratic mean' type regularity condition is introduced and it is shown that under this condition the LAMN condition is satisfied. Recent results on martingale central limit theorems are used in deriving the asymptotic mixed normality of the log-likelihood function. The results of this chapter are originally due to LeCam (1970) for the i.i.d. case. LAN condition for dependent observations has been studied, among several others, by Roussas (1972 and 1979) and for the independent but not necessarily identical case has been studied by Phillippou and Roussas (1973) and Ibragimov and Khasminskii (1975).

Chapter 3 presents two results concerning the invariance of the possible limits of distributions. The first one is a related, but different, version and the second one is a strengthened form of an invariance result given in LeCam (1979). These invariance results of LeCam are simple but their power can be seen from the applications given in Chapters 4,5 and 6 of the present work.

In Chapter 4, a certain kind of asymptotic differentiability condition, analogous to the one assumed in LeCam (1960), is introduced and it is shown that under this condition, the limit distribution, when it exists, of the log-likelihood function is a mixed normal for almost all points of the parameter space. This result extends the corresponding result of LeCam (1960, Theorem 4.1) Secondly we show that, without assuming the existence of the limit distribution, the log-likelihood function converges in a certain weak topology (introduced in Ch.3) to a mixed normal distribution. Though the convergence stated in this result is weaker than the one stated in the first result, the first result actually follows from this result and it appears that this result is more important than the first result. Using a statement of this second result it is further noted that in the special case considered by LeCam (1960, Theorem 4.1), LeCam's conclusion holds even when the existence of the limit distribution is not assumed. Thirdly we show that under a specific form of the asymptotic differentiability condition, asymptotic mixed normality is equivalent to the contiguity condition and a certain kind of invariance condition on the sequence of random matrices involved in the approximation. This third result was independently obtained by Davies (1979) also.

It is important to note that in this Chapter 4, asymptotic mixed normality occur through an argument which has nothing to do with " martingale differences". For a better explanation of why the Gaussian family is so pervasive, one should consult LeCam (1974a, Ch.11 and 1979, Ch.8).

In Chapter 5 we first show that when the given sequence of estimators satisfies a certain kind of invariance restriction, the limit distribution of any convergent sub-sequence of estimators can be <u>conditionally</u> decomposed as a convolution. This result extends and strengthens the convolution result of Hájek (1970). We would like to mention that the convolution result for the LAN case was also essentially obtained by Inagaki (1970) under restrictive assumptions. Secondly we show that, without assuming the invariance restriction, the limit distribution of any subsequence that is convergent in the weak topology introduced in Ch.3 can be decomposed conditionally as a convolution for almost all [Lebegue] points of the parameter space. This result extends and strengthens the corresponding result for the LAN case mentioned in LeCam (1973). Applying our conditional convolution results we deduce several results concerning the asymptotic lower bounds for risk functions; one of these results clarifies some of the statements made earlier by Heyde (1978).

We continue the study of asymptotic properties of risk functions in Chapter 6. We first present a general local result concerning the asymptotic lower bound for risk functions and then we characterise the estimators which attain this lower bound. An important thing to be noted here is that these two results do not depend on the dimensionality restriction of the parameter space. A more familiar result under the usual invariance restriction is presented as a simple corollary of our general result. The next two results extend the local asymptotic minimax and admissibility results presented in Hájek (1972) and LeCam (1972 and 1974a) for the LAN case. Next we present a general result concerning a certain kind of posterior approximation. Using this general result we deduce a result concerning the global asymptotic lower bound for risk functions; using this same general result we then characterise the estimators which attain the lower bound. The results concerning the global asymptotic lower bound and the corresponding characterisation occur explicitely for the LAN case in Strasser (1978).

Chapters 7 and 8 form the global part of the present work.

The main purpose of Chapter 7 is, under suitable global assumptions, to see what are the minimum possible local regularity conditions needed under which the sequences of maximum likelihood estimators, maximum probability estimators and a certain class of Bayes estimators can be locally approximated by the sequence of random vectors involved in the linear terms of the approximation of the LAMN condition and satisfy the asymptotic sufficieny criteria of Ch.1. A result concerning the posterior approximation at the true value of the parameter is also presented. Our arguments depend only on the approximating form of the log-likelihood ratios, and they do not in any way depend on any particular nature of the sample space. For example, given that a sequence of maximum probability estimators is constistant at a certain rate, the only additional condition we assume to show that this sequence satisfy the above mentioned requirements is the LAMN condition. We would like to point out that it is not the aim of this chapter to give less stringent regularity conditions than some of the possibly stringent conditions usually found in the literature. Our aim is just to clarify some of the local arguments usually found in the literature.

In Chapter 8 we try to extend the results of Ibragimov and Khasminskii (1972 and 1973) concerning the convergence of moments of statistical estimators that are considered in Ch.7 of the present work. Ibragimov and Khasminskii obtained their results when the observations are i.i.d. and when the parameter space is a subset of the real line, and some of their arguments depend in a crucial manner on the dimensionality restriction of the parameter space. We prove the results for the LAMN case and for the multi-dimensional parameter space. A result concerning the weak convergence of the sequence of likelihood ratio random processes to a a mixed Gaussian shift process is also presented. For a better introduction see the introduction presented in Chapter 8.

After the completion of the present work we received a copy of Ph.D. thesis from Swensen (1980, September), where he has independently obtained results which are closely related to some of the results of the present work. More specifically, the results of the first chapter of his thesis are related to the results of the chapter 2 of the present work. His thesis further contains the local asymptotic minimax and admissibility results of Chapter 6 of the present work; his proofs consist of first proving the minimax and admissibility results for the limit of the LAMN families and then using the general results of LeCam (1972 and 1974) to get the corresponding limiting statements for the sequence of LAMN families, whereas our proofs are directly based on certain approximation results for the LAMN case.

It may be noted that we have not treated the asymptotic testing problem. Swensen's thesis contains some important results regarding the asymptotic testing problem for the LAMN case. For an earlier important treatment on testing for the general case, under some specific assumptions, see Basawa and Scott (1977) and Feigin (1978). Basawa and Koul (1979) suggests, for the LAMN case, some test statistics, similar to the ones used for the LAN case, for multidimensional case, but the asymptotic optimality is not discussed.

CHAPTER 1

# LOCALLY ASYMPTOTICALLY MIXED NORMAL EXPERIMENT

## 1. INTRODUCTION

In this chapter we introduce the notion of a locally asymptotically mixed normal (LAMN) experiment and obtain some basic and preliminary results which will be frequently used in the chapters that follow; all the results of this chapter are almost essentially contained in LeCam (1960) and Chapter 12 of LeCam (1974a), though the arguments of LeCam are intended for a locally asymptotically normal (LAN) experiment.

Roughly speaking, an LAMN experiment means a sequence of appropriately normalised log-likelihood ratios is approximated with probability tending to one by the sum of two expressions, the first one being a sequence of random linear functions of the normalised parameter and the second one being a sequence of random quadratic forms of the normalised parameter, the sequence of matrices involved in the quadratic forms being convergent weakly to an a.s. positive definite (p.d.) random matrix and the sequence of random vectors involved in the linear terms being convergent weakly to a mixed normal distribution;

in the special case when the matrices of the quadratic terms are identical to a constant matrix we say that the experiment satisfies the LAN condition.

A sequence of estimators will be called <u>asymptotically centering sequence</u> (ACS) <u>of estimators</u> if it can be substituted in the sequence of linear terms of the approximation of the log-likelihood ratios.

In Section 2 we introduce some notations and precise definitions of an LAMN experiment and ACS estimators.

In Section 3 we first present a result on contiguity and then show that the sequence of random vectors and matrices of the LAMN experiment satisfies a certain invariance condition. We next show that one can select modified versions of the random vectors and matrices of the LAMN experiment in such a way that these versions satisfy certain regularity properties; these regularity properties will play a crucial role in Section 4 of the present chapter and also in the chapters that follow. We then present an exponential approximation result (Lemma 6). This result says that, locally, one can approximate the LAMN experiment, in the $L_1$-norm, by an another experiment which is a slightly perturbed mixed normal shift experiment with a slightly deformed likelihood ratios. In particular this will imply that the sequence of random vectors and matrices of the LAMN experiment forms a sequence of locally asymptotically sufficient statistics. As will be seen in the subsequent chapters, this ~~theorem~~ lemma will also serve as a powerful tool in extending certain results of Hájek and LeCam and also in clarifying certain local arguments usually associated with Bayes estimators.

In section 4, we present two results which are due to LeCam in the LAN case, may be described as follows. If one is given an LAMN experiment, then take a preliminary estimate, i.e., an estimate which takes values in a " small vicinity" of the true value, look at the logarithms of likelihood ratios and fit a quadratic to them. Take for estimate the point that maximises this quadratic. This sequence of estimators turns out to be a sequence of ACS estimators. The second result says that any sequence of ACS estimators together with certain estimates of the random matrices form a sequence of statistics which is " approximately " sufficient.

## 2. NOTATIONS AND DEFINITIONS

Let $\underline{\underline{E}}_n = \{\underline{\underline{X}}_n, \underline{\underline{A}}_n, P_{\theta,n} ; \theta \in \underline{\textcircled{H}}\}$ , $n \geq 1$ , be a sequence of experiments ; throughout what follows it will be assumed, without any further mentioning, that $\underline{\underline{H}}$ is an open subset of $R^k$.

We use the following notations. If P and Q are probability measures on a measurable space $(\underline{\underline{X}}, \underline{\underline{A}})$, then dP/dQ denotes the Radon-Nykodym derivative of the Q-continuous part of P with respect to Q. If p and q are densities of P and Q with respect to some σ-finite measure λ, then

$$\| P - Q \| = \int |p-q| d\lambda$$

is the $L_1$-norm. If Y is a random vector its distribution will be denoted by $\mathcal{L}(Y)$ or by $\mathcal{L}(Y|P)$ when $Y: (\underline{\underline{X}}, \underline{\underline{A}}) \longrightarrow (R^q, \underline{\underline{B}}^q)$, $q \geq 1$, $\underline{\underline{B}}^q$ being the σ-field of Borel subsets of $R^q$. For a vecto $h \varepsilon R^k$, $h'$ denotes the transpose of h and $|h|$ denotes the euclidea norm; for a square matrix D, $\| D \|$ denotes the norm defined by the the square root of the sum of squares of its elements. '$\Longrightarrow$' denotes the convergence in distribution. Log $\frac{dP_{s,n}}{dP_{\theta,n}}$, $\theta, s \varepsilon (\underline{\underline{H}})$, $n \geq 1$, will be denoted by $\Lambda_n(s,\theta)$.

We now introduce the following set of definitions.

<u>Definition 1</u>. The sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$ if the following two conditions are satisfied.

(A.1) There exists a sequence $\{W_n(\theta_o)\}$ of $\underline{\underline{A}}_n$-measurable k-vector and a sequence $\{T_n(\theta_o)\}$ of $\underline{\underline{A}}_n$-measurable $k \times k$ symmetric matrice such that $P_{\theta_o,n}[T_n(\theta_o) \text{ is p.d.}] = 1$ for every $n \geq 1$ and the difference

$$\Lambda_n(\theta_o+\delta_n h, \theta_o) - [h' T_n^{1/2}(\theta_o) W_n(\theta_o) - \frac{1}{2} h' T_n(\theta_o) h]$$

converges to zero in $P_{\theta_o,n}$-probability for every $h \varepsilon R^k$, where $\{\delta_n\}$ is a sequence of p.d. matrices such that $\|\delta_n\| \longrightarrow 0$ as $n \longrightarrow \infty$.

(A.2) There exists an almost surely (a.s.) p.d. random matrix $T(\theta_o)$ such that

$$\mathcal{L}(W_n(\theta_o), T_n(\theta_o) | P_{\theta_o,n}) \implies \mathcal{L}(W, T(\theta_o))$$

where $W$ is a copy of the standard k-variate normal distribution independent of $T(\theta_o)$.

Definition 2. Let $W_n(\theta_o), T_n(\theta_o), \delta_n, n \geq 1$, be as in Definition 1. Assume that (A.2) of Definition 1 is satisfied. Then the sequence $\{\underline{E}_n\}$ satisfies the LAMN condition in the strict sense at $\theta=\theta_o \varepsilon (\underline{H})$ if the difference

$$\Lambda_n(\theta_o+\delta_n h_n, \theta_o) - [h_n' T_n^{1/2}(\theta_o) W_n(\theta_o) - \frac{1}{2} h_n' T_n(\theta_o) h_n]$$

converges to zero in $P_{\theta_o,n}$- probability for every bounded sequence $\{h_n\}$ of elements of $R^k$.

Definition 3. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the LAMN-condition at $\theta=\theta_o\varepsilon(\underline{\underline{H}})$. Then the sequence $V_n:(\underline{\underline{X}}_n,\underline{\underline{A}}_n)\longrightarrow(R^k,\underline{\underline{B}}^k)$, $n\geq 1$, of estimators is said to be a sequence of ACS estimators at $\theta=\theta_o\varepsilon(\underline{\underline{H}})$ if the difference

$$\delta_n^{-1}(V_n-\theta_o)-T_n^{-1/2}(\theta_o)W_n(\theta_o)$$

converges to zero in $P_{\theta_o,n}$-probability.

## 3. SOME PRELIMINARY RESULTS

Lemma 1. Suppose that the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta=\theta_o\varepsilon(\underline{\underline{H}})$. Then the sequences $\{P_{\theta_o,n}\}$ and $\{P_{\theta_o+\delta_n h,n}\}$ are contiguous for every $h\varepsilon R^k$. In case the sequence $\{\underline{\underline{E}}_n\}$ satisfies the strictly LAMN condition at $\theta=\theta_o\varepsilon(\underline{\underline{H}})$, the sequences $\{P_{\theta_o,n}\}$ and $\{P_{\theta_o+\delta_n h_n,n}\}$ are contiguous for every bounded sequence $\{h_n\}$ of elements of $R^k$.

Proof. Using the independence of $W$ and $T(\theta_o)$ it follows that

$$E[\exp(h'T^{1/2}(\theta_o)W-\tfrac{1}{2}h'T(\theta_o)h)]=1$$

for every $h\varepsilon R^k$. Hence the result follows from the statement (5) of Theorem 2.1 of LeCam (1960), since there exists a subsequence $\{h_m\}\subseteq\{h_n\}$ such that $h_m\longrightarrow h$ as $m\longrightarrow\infty$ for some $h\varepsilon R^k$.

Lemma 2. Suppose that the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o \varepsilon(\underline{\underline{H}})$. Then for every $h \varepsilon R^k$

$$\mathscr{L}(T_n(\theta_o), T_n^{-1/2}(\theta_o) W_o(\theta_o) \mid P_{\theta_o + \delta_n h, n}) \Longrightarrow \mathscr{L}(T(\theta_o), T^{-1/2}(\theta_o) W + h).$$

Proof. For simplicity assume that $\dim(\underline{\underline{H}}) = 1$. According to the statement (6) of Theorem 2.1 of LeCam (1960) it follows that, for every $u, v, h \varepsilon R$,

$$E\{\exp[iuT_n^{-1/2}(\theta_o) W_n(\theta_o) + ivT_n(\theta_o)] \mid P_{\theta_o + \delta_n h, n}\}$$

$$\longrightarrow E\{\exp[iuT^{-1/2}(\theta_o) W + ivT(\theta_o) + hT^{1/2}(\theta_o) W - \tfrac{h^2}{2} T(\theta_o)]\}$$

$$= E\{\exp(ivT(\theta_o) + iuh) E^T[\exp(iu(T^{-1/2}(\theta_o) W - h) + hT^{1/2}(\theta_o) W - \tfrac{h^2}{2} T(\theta_o)]\}$$

where $E^T$ denotes the conditional expectation given $T(\theta_o)$. Using the independence of $W$ and $T(\theta_o)$ it follows that

$$E^T\{\exp[iu(T^{-1/2}(\theta_o) W - h) + hT^{1/2}(\theta_o) W - \tfrac{h^2}{2} T(\theta_o)]\} = E^T[\exp(iuT^{-1/2}(\theta_o) W)].$$

Hence we see that for every $u, v, h \varepsilon R$

$$E[\exp(iuT_n^{-1/2}(\theta_o) W_n + ivT_n(\theta_o)) \mid P_{\theta_o + \delta_n h, n}]$$

$$\longrightarrow E[\exp(iu(T^{-1/2}(\theta_o) W + h) + ivT(\theta_o))].$$

This gives the required result.

As a consequence of the above lemma we obtain the followi[ng]

Corollary. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta=\theta_o\varepsilon(\underline{H})$. Let $\{V_n\}$ be a sequence of ACS estimator[s]. Then for every $h\varepsilon R^k$

$$\mathcal{L}(T_n(\theta_o),\delta_n^{-1}(V_n-\theta_o-\delta_n h)\,|\,P_{\theta_o+\delta_n h,n}) \Longrightarrow \mathcal{L}(T(\theta_o),T^{-1/2}(\theta_o)W).$$

Proof. Since the sequences $\{P_{\theta_o,n}\}$ and $\{P_{\theta_o+\delta_n h,n}\}$ are contiguous for every $h\varepsilon R^k$, the difference

$$\delta_n^{-1}(V_n-\theta_o)-T_n^{-1/2}(\theta_o)W_n(\theta_o)$$

converges to zero in $P_{\theta_o+\delta_n h,n}$ - probability for every $h\varepsilon R^k$. Hence the result follows from Lemma 2.

Lemma 3. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the condition that the quantity

$$\| P_{\theta+\delta_n h_n,n}-P_{\theta+\delta_n h_n^*,n}\| \longrightarrow 0 \quad \text{as} \quad n\longrightarrow\infty \qquad (*)$$

for every $\theta\varepsilon(\underline{H})$ whenever the bounded sequences $\{h_n\}$ and $\{h_n^*\}$ of $R^k$ are such that $|h_n-h_n^*|\longrightarrow 0$ as $n\longrightarrow\infty$, where $\{\delta_n\}$ is a sequence of p.d. matrices such that $\|\delta_n\|\longrightarrow 0$ as $n\longrightarrow\infty$. Then there exists a construction of another sequence of experiment[s] $\underline{\underline{E}}_n^*=\{\underline{\underline{X}}_n,\underline{\underline{A}}_n,P_{\theta,n}^*;\theta\varepsilon(\underline{H})\}$, $n\geq 1$, with the following properties.

(i) The functions $\theta\longrightarrow P_{\theta,n}^*(A)$, $A\varepsilon\underline{\underline{A}}_n$, $n\geq 1$, are Borel measurable

(ii) the functions

$$(\underline{\underline{X}}_n \times (\underline{H}) \times R^k) \longrightarrow \log \frac{P^*_{\theta+\delta_n h, n}}{P^*_{\theta, n}}, \; n \geq 1,$$

are jointly measurable,

(iii) for every $\alpha > 0$ and $\theta \varepsilon (\underline{H})$

$$\sup_{|h| \leq \alpha} \| P_{\theta+\delta_n h, n} - P^*_{\theta+\delta_n h, n} \| \longrightarrow 0, \text{ and}$$

(iv) when the sequence $\{\underline{\underline{E}}_n\}$ further satisfies the condition that the sequences $\{P_{\theta,n}\}$ and $\{P_{\theta+\delta_n h_n, \theta}\}$ are contiguous for every $\theta \varepsilon (\underline{H})$ and for every bounded sequence $\{h_n\}$ of $R^k$, the difference

$$\log \frac{P_{\theta+\delta_n h_n, h}}{P^*_{\theta,n}} - \log \frac{P^*_{\theta+\delta_n h_n, n}}{P^*_{\theta,n}}$$

converges to zero in $P_{\theta,n}$-probability for every $\theta \varepsilon (\underline{H})$ and for every bounded sequence $\{h_n\}$ of $R^k$.

<u>Proof</u>. See LeCam (1974a, Ch. 12, pp. 153-155).

<u>Remark</u>. It is easy to see that any sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfying the strictly LAMN-condition at all $\theta \varepsilon (\underline{H})$ satisfies the condition (*) of the above lemma 3.

<u>Lemma 4</u>. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at all $\theta \varepsilon (\underline{H})$, then a sequence $\{T^*_n(\theta)\}$ of random $k \times k$ symmetric matrices can be constructed in such a way that

(i) the difference $T_n(\theta) - T^*_n(\theta)$ converges to zero in $P_{\theta,n}$-probability for every $\theta \varepsilon (\underline{H})$,



(ii) the difference $T^*_n(\theta) - T_n(\theta+\delta_n h)$ converges to zero in

$P_{\theta,n}$ - probability for every $\theta \varepsilon (\underline{H})$ and $h \varepsilon R^k$, and

(iii) $P_{\theta,n}[T_n^*(\theta)$ is p.d.$] = 1$ for every $n > 1$ and $\theta \varepsilon (\underline{H})$.

(iv) In case the sequence $\{\underline{\underline{E}}_n\}$ satisfies the strictly LAMN condition at all $\theta \varepsilon (\underline{H})$, the sequence $\{T_n^*(\theta)\}$ can be constructed in such a way that it further satisfies the condition that the matrices $T_n^*(\theta)$, $n \geq 1$, are $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$ - measurable and that the difference $T_n^*(\theta) - T_n^*(\theta + \delta_n h_n)$ converges to zero in $P_{\theta,n}$ - probability for every $\theta \varepsilon (\underline{H})$ and for every bounded sequence $\{h_n\}$ of $R^k$.

<u>Proof</u>. For simplicity assume that $\dim(\underline{H}) = 1$. First note that

$$T_n(\theta) = -4\left[H_n(\theta,1) - 2H_n(\theta,\tfrac{1}{2})\right]$$

where we set

$$H_n(\theta,h) = hT_n^{1/2}(\theta)W_n(\theta) - \frac{h^2}{2}T_n(\theta) .$$

Now set

$$T_n^*(\theta) = -4\left[\Lambda_n(\theta+\delta_n,\theta) - 2\,\Lambda_n(\theta+\tfrac{1}{2}\delta_n,\theta)\right]$$

where

$$\Lambda_n(\theta,h) = \log \frac{P_{\theta+\delta_n h,\theta}}{P_{\theta,n}} .$$

In view of contiguity we can assume without loss of generality that $P_{\theta+\delta_n h,n} \approx P_{\theta,n}$ for every $n \geq 1, \theta \varepsilon (\underline{H})$ and $h \varepsilon R^k$, where the symbol $\approx$ denotes mutual absolute continuity. Since the difference $\Lambda_n(\theta+\delta_n h,\theta) - H_n(\theta,h)$ converges to zero in $P_{\theta,n}$ - probability for every $\theta \varepsilon (\underline{H})$ and $h \varepsilon R^k$ the statement (i) follows. Now, in view of mutual absolute continuity, we have

$$\Lambda_n(\theta+\delta_n(h+s),\theta+\delta_n h) = \Lambda_n(\theta+\delta_n(h+s),\theta) - \Lambda_n(\theta+\delta_n h,\theta)$$

for every $n \geq 1$, $\theta \varepsilon (\underline{H})$ and $s,h \varepsilon R$, and so

$$T_n^*(\theta+\delta_n h) = -4\left[\Lambda_n(\theta+\delta_n(h+1),\theta) - 2\Lambda_n(\theta+\delta_n(h+\tfrac{1}{2}),\theta) + \Lambda_n[\theta+\delta_n h,\theta)\right].$$

Also observe that

$$T_n(\theta) = -4\left[H_n(\theta,h+1) - 2H_n(\theta,h+\tfrac{1}{2}) + H_n(\theta,h)\right].$$

Hence the statement (ii) follows. The $T_n^*(\theta)$'s constructed above need not be positive but this can be easily remedied, e.g. if we define

$$T_n^{**}(\theta) = T_n^*(\theta) \quad \text{if} \quad T_n^*(\theta) > 0$$
$$= 1 \qquad \text{otherwise,}$$

then it is easy to see that the sequence $\{T_n^{**}(\theta)\}$ satisfies the statements (i) and (ii) using the facts that the sequence $\{T_n^*(\theta)\}$ satisfies the statements (i) and (ii) and that, for every $\theta \varepsilon (\underline{H})$, the limit $T(\theta)$ is positive almost surely.

In case the sequence $\{\underline{\underline{E}}_n\}$ satisfies the strictly LAMN condition at all $\theta \varepsilon (\underline{H})$, we set

$$T_n^*(\theta) = -4\left[\Lambda_n^*(\theta+\delta_n,\theta) - 2\Lambda_n^*(\theta+\tfrac{1}{2}\delta_n,\theta)\right]$$

where

$$\Lambda_n^*(\theta+\delta_n h,\theta) = \log \frac{dP^*_{\theta+\delta_n h,n}}{dP^*_{\theta,n}} .$$

(The family $\{P^*_{\theta,n};\ \theta \varepsilon (\underline{H})\}$ is the one constructed in Lemma 3; see the remark following this lemma 3.) The joint measurability of $T_n^*(\theta)$ follows from the joint measurability of $\Lambda_n^*(\theta+\delta_n h,\theta)$. The remaining arguments needed to complete the proof of the statement (iv) are, in view of the statement (iv) of Lemma 3, identical to the arguments of the proof of the statements (i) - (iii) of the present lemma.

Lemma 5. Suppose that the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at all $\theta\varepsilon(\underline{H})$. Then a sequence $\{\Delta_n(\theta)\}$ of random k-vectors can be constructed in such a way that

(i) the difference $\Delta_n(\theta) - T_n^{-1/2}(\theta)W_n(\theta)$ converges to zero in $P_{\theta,n}$ - probability for every $\theta\varepsilon(\underline{H})$,

(ii) the difference $\Delta_n(\theta+\delta_n h) - [\Delta_n(\theta) - h]$ converges to zero in $P_{\theta,n}$ - probability for every $\theta\varepsilon(\underline{H})$ and $h\varepsilon R^k$.

(iii) In case the sequence $\{\underline{\underline{E}}_n\}$ satisfies the strictly LAMN condition at all $\theta\varepsilon(\underline{H})$, the sequence $\{\Delta_n(\theta)\}$ can be constructed in such a way that it further satisfies the condition that the vectors $\Delta_n(\theta), n\geq 1$, are $\underline{\underline{A}}_n\times\underline{\underline{B}}^k$ measurable and that the difference $\Delta_n(\theta+\delta_n h_n) - [\Delta_n(\theta) - h_n]$ converges to zero in $P_{\theta,n}$ - probability for every bounded sequence $\{h_n\}$ of $R^k$ and for every $\theta\varepsilon(\underline{H})$.

Proof. Let $\{u_j\}$ ; $|u_j|\leq 1$ be a basis of $R^k$. Construct $\Delta_n(\theta)$ by the relation

$$\Lambda_n(\theta+\delta_n u_j,\theta) = u_j' T_n^*(\theta)\Delta_n(\theta) - \tfrac{1}{2}u_j' T_n^*(\theta)u_j .$$

The proof of the statement (i) is immediate from the statements (i) and (iii) of Lemma 4. To prove the statement (ii) consider

$$\Lambda_n(\theta+\delta_n(u_j+h),\theta+\delta_n h)$$
$$= \Lambda_n(\theta+\delta_n(u_j+h),\theta) - \Lambda_n(\theta+\delta_n h,\theta)$$

and so

$$u_j' T_n^*(\theta+\delta_n h)\,\Delta_n(\theta+\delta_n h)$$
$$= \Lambda_n(\theta+\delta_n(u_j+h),\theta) - \Lambda_n(\theta+\delta_n h,\theta) + \frac{1}{2}\, u_j' T_n^*(\theta+\delta_n h)\, u_j$$

and this is approximated by

$$u_j' \left[ T_n^*(\theta)(\Delta_n(\theta) - h)\right] + \frac{1}{2}\, u_j' \left[ T_n^*(\theta+\delta_n h) - T_n^*(\theta)\right] u_j .$$

Hence in view of the statements (ii) and (iii) of Lemma 4, the statement (ii) of the present lemma follows.

In view of the statements (iii) and (iv) of Lemma 4, the statement (iii) of the present lemma is similarly proved by constructing $\Delta_n(\theta)$ by the relation

$$\Lambda_n^*(\theta+\delta_n u_j,\theta) = u_j' T_n^*(\theta)\,\Delta_n(\theta) - \frac{1}{2}\, u_j' T_n^*(\theta)\, u_j ,$$

where $\Lambda_n^*(\theta+\delta_n h,\theta)$ is as defined in the proof of Lemma 4.

<u>Lemma 6</u>. Assume that the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN - condition at $\theta = \theta_0 \varepsilon (\underline{\underline{H}})$. Then there exist

(i) an increasing sequence $\{k_n\}$ tending to infinity as $n \longrightarrow \infty$,

(ii) functions $C_n : \underline{\underline{H}} \times R^{k} \to R$ such that

$$\sup_{|h| \leq \alpha} |C_n(\theta_o,h) - 1| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

for every $\alpha > 0$, such that the measures $Q_n(\theta_o,h)|\underline{\underline{A}}_n$, $Q_n(\theta_o,h) \ll P_{\theta_o,n}$, defined by

$$\frac{dQ_n(\theta_o,h)}{dP_{\theta_o,n}} = C_n(\theta_o,h) \exp\left[h'T_n^{1/2}(\theta_o)W_n^*(\theta_o) - \frac{1}{2} h'T_n(\theta_o)h\right]$$

with $W_n^*(\theta_o) = W_n(\theta_o)I(|T_n^{1/2}(\theta_o)W_n(\theta_o)| \leq k_n)$, are probability measures and satisfy

$$\| P_{\theta_o+\delta_n h,n} - Q_n(\theta_o,h) \| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

for every $h \varepsilon R^k$,

(iii) in case the sequence $\{\underline{\underline{E}}_n\}$ satisfies the strictly LAMN condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$ we have

$$\| P_{\theta_o+\delta_n h_n,n} - Q_n(\theta_o,h_n) \| \longrightarrow 0$$

as $n \longrightarrow \infty$ for every bounded sequence $\{h_n\}$ of $R^k$.

<u>Proof</u>. Define, for $\alpha > 0$,

$$W_n^\alpha(\theta_o) = W_n(\theta_o)I(|T_n^{1/2}(\theta_o)W_n(\theta_o)| < \alpha)$$

and

$$W^\alpha(\theta_o) = WI(|T^{1/2}(\theta_o)W| < \alpha) .$$

There is a dense set of values of $\alpha$ for which

$$\mathcal{L}(T_n(\theta_o), W_n^\alpha(\theta_o)|P_{\theta_o,n}) \Longrightarrow \mathcal{L}(T(\theta_o), W^\alpha(\theta_o)).$$

For any such $\alpha$, we have

$$\sup_{|h| \leq a} |E[\exp(h'T_n^{1/2}(\theta_o)W_n^a(\theta_o) - \frac{1}{2} h'T_n(\theta_o)h)]$$
$$- E[\exp(h'T^{1/2}(\theta_o)W^a(\theta_o) - \frac{1}{2} h'T(\theta_o)h)] |$$
$$\longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty ,$$

since the family of functions, defined on the space of k-vectors and $k \times k$ p.d. matrices,

$$\left\{(x,D) \longrightarrow \exp(h'D^{1/2}x - \frac{1}{2} h'Dh) : |h| \leq a \right\}$$

is uniformly bounded and equicontinuous whenever the domain of $x$ is bounded. Hence by a standard diagonal argument one can choose an increasing sequence $\{k_n\}$ tending to infinity such that

$$\sup_{|h| \leq k_n} |E[\exp(h'T_n^{1/2}(\theta_o)W_n^*(\theta_o) - \frac{1}{2} h'T_n(\theta_o)h)]$$
$$- E[\exp(h'T^{1/2}(\theta_o)W^{k_n}(\theta_o) - \frac{1}{2} h'T(\theta_o)h)] |$$
$$\longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

where $W_n^*(\theta_o) = W_n^{k_n}(\theta_o)$ and, hence

$$\sup_{|h| \leq b} |E[\exp(h'T_n^{1/2}(\theta_o)W_n^*(\theta_o) - \frac{1}{2} h'T_n(\theta_o)h)]$$
$$- E[\exp(h'T^{1/2}(\theta_o)W^{k_n}(\theta_o) - \frac{1}{2} h'T(\theta_o)h)] |$$
$$\longrightarrow 0 \qquad (1.1)$$

as $n \longrightarrow \infty$ for every $b > 0$.

We now show that, for every $b > 0$,

$$\sup_{|h| \leq b} |E[\exp(h'T^{1/2}(\theta_o)W^{k_n}(\theta_o) - \frac{1}{2} h'T(\theta_o)h)] - 1 | \longrightarrow 0 \qquad (1.2)$$

as $n \longrightarrow \infty$. Now note that

$$\sup_{|h|\leq b} |E[\exp(h'T^{1/2}(\theta_o)W^{k_n}(\theta_o) - \tfrac{1}{2} h'T(\theta_o)h)] - 1|$$

$$\leq E[\sup_{|h|\leq b} |E^T[\exp(h'T^{1/2}(\theta_o)W_n^{k_n}(\theta_o) - \tfrac{1}{2}h'T(\theta_o)h] - 1|]$$

and

$$\exp(h'T^{1/2}(\theta_o)W^{k_n}(\theta_o) - \tfrac{1}{2} h'T(\theta_o)h)$$

$$\leq \exp(h'T^{1/2}(\theta_o)W - \tfrac{1}{2} h'T(\theta_o)h) + 1$$

for every $n \geq 1$ and $h \in R^k$ and hence, using the independence of $W$ and $T(\theta_o)$,

$$\sup_{|h|\leq b} E^T[\exp(h'T^{1/2}(\theta_o)W^{k_n}(\theta_o) - \tfrac{1}{2} h'T(\theta_o)h)] \leq 2$$

for every $b > 0$ and $n \geq 1$. Hence (1.2) will follow if we show that, for each fixed $T(\theta_o)$,

$$\sup_{|h|\leq b} |E^T[\exp(h'T^{1/2}(\theta_o)W^{k_n}(\theta_o) - \tfrac{1}{2} h'T(\theta_o)h)] - 1| \longrightarrow 0$$

This is quite easy to see. From (1.1) and (1.2) we now have, for every $b > 0$,

$$\sup_{|h|\leq b} |E[\exp(h'T_n^{1/2}(\theta_o)W_n^*(\theta_o) - \tfrac{1}{2} h'T_n(\theta_o)h)] - 1| \longrightarrow 0 \quad (1.3)$$

as $n \longrightarrow \infty$. Set

$$C_n(\theta_o,h) = 1/E[\exp(h'T_n^{1/2}(\theta_o)W_n^*(\theta_o) - \tfrac{1}{2} h'T_n(\theta_o)h)] .$$

From (1.3) it follows that

$$\sup_{|h|\leq b} |C_n(\theta_o,h) - 1| \longrightarrow 0 \text{ as } n \longrightarrow \infty \quad (1.4)$$

for every $b > 0$. To complete the proof of the statements (i) and (ii) it remains to show that

$$\| P_{\theta_o+\delta_n h,n} - Q_n(\theta_o,h) \| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

for every $h \in R^k$. Without loss of generality we can assume that $P_{\theta_o+\delta_n h,n} \approx P_{\theta_o,n}$ for every $n \geq 1$, $h \in R^k$. In view of (1.4) and since $|W_n^*(\theta_o) - W_n(\theta_o)| \longrightarrow 0$ in $P_{\theta_o,n}$-probability, we see that the difference

$$Z_{n,\theta_o}(h) - Z'_{n,\theta_o}(h) \tag{1.5}$$

converges to zero in $P_{\theta_o,n}$-probability for every $h \in R^k$, where we set

$$Z_{n,\theta_o}(h) = \frac{dP_{\theta_o+\delta_n h,n}}{dP_{\theta_o,n}} \quad \text{and} \quad Z'_{n,\theta_o}(h) = \frac{dQ_n(\theta_o,h)}{dP_{\theta_o,n}} .$$

Further, in view of contiguity,

$$\lim_{\alpha \longrightarrow \infty} \limsup_{n \longrightarrow \infty} \int_{\{|Z_{n,\theta_o}(h)| > \alpha\}} Z_{n,\theta_o}(h)\, dP_{\theta_o}$$

$$= \lim_{\alpha \longrightarrow \infty} \limsup_{n \longrightarrow \infty} P_{\theta_o+\delta_n h,n} \left[ |Z_{n,\theta_o}(h)| > \alpha \right] = 0 . \tag{1.6}$$

Similarly we see that, since the sequences $\{Q_n(\theta_o,h)\}$ and $\{P_{\theta_o,n}\}$ are contiguous,

$$\lim_{\alpha \longrightarrow \infty} \limsup_{n \longrightarrow \infty} \int_{\{|Z'_{n,\theta_o}(h)| > \alpha\}} Z'_{n,\theta_o}(h)\, dP_{\theta_o,n}$$

$$= \lim_{\alpha \to \infty} \limsup_{n \to \infty} Q_n(\theta_o, h)[\,|Z'_{n,\theta_o}(h)| > \alpha\,) = 0. \quad (1.7)$$

Combining (1.5), (1.6) and (1.7) we see that

$$\int |Z_{n,\theta_o}(h) - Z'_{n,\theta_o}(h)|\,dP_{\theta_o,n} \longrightarrow 0$$

as $n \longrightarrow \infty$ for every $h \in R^k$. This completes the proof of the statements (i) and (ii). Statement (iii) is similarly proved.

Remark. In the statement (iii) of the above lemma, if we consider the jointly measurable versions, constructed in Lemmas 3 and 4, of $W_n(\theta), T_n(\theta), n \geq 1$, the quentities $C_n(\theta,h), n \geq 1$, will be $\underline{\underline{B}}^k \times \underline{\underline{B}}^k$ measurable and the likelihood ratios $\frac{dQ_n(\theta,h)}{dP_{\theta,n}}$ will be $\underline{\underline{A}}_n \times \underline{\underline{B}}^k \times \underline{\underline{B}}^k$ measurable.

## 4. ASYPTOTIC SUFFICIENCY

In this section we first present an extremely simple method of construction, using a preliminary estimate, of a sequence of ACS estimators; the specific method of construction given here is same as the one presented in LeCam (1960) and LeCam (1974a). Secondly we present a result concerning the asymptotic sufficiency of any sequence of ACS estimators; an important thing to be noted here is that a sequence of ACS estimators alone is not asymptotically sufficient, but it together with a sequence of estimates of the random matrices of the LAMN condition form a sequence of asymptotically sufficient estimators.

<u>Lemma 7</u>. Let $\{\hat{\theta}_n\}$ be a sequence of estimators such that, for each $\theta \varepsilon (\underline{\overline{H}})$, the sequence $\{\delta_n^{-1}(\hat{\theta}_n - \theta)\}$ is relatively compact for the sequence $\{P_{\theta,n}\}$. Then there are functions $\theta_n^*$ of $\hat{\theta}_n$ such that, for each $\theta \varepsilon (\underline{\overline{H}})$,

(i) the sequence $\{\delta_n^{-1}(\theta_n^* - \theta)\}$ is relatively compact for the sequence $\{P_{\theta,n}\}$, and

(ii) for each $b \varepsilon (0,\infty)$ there is a number $K(b)$ such that the number of possible values of $\theta_n^*$ contained in $\delta_n D_b$, where $D_b = \{h \varepsilon R^k : |h| \leq b\}$, never exceeds $K(b)$.

<u>Proof</u>. This lemma is the Lemma 4 of LeCam (1974, Ch. 12).

Practically, this lemma says that one computes $\delta_n^{-1} \hat{\theta}_n$ only upto a to a certain number of decimals.

<u>Theorem 1</u>. Let $\{\hat{\theta}_n\}$ be a sequence of estimators such that, for each $\theta \varepsilon (\underline{\overline{H}})$, the sequence $\{\delta_n^{-1}(\hat{\theta}_n - \theta)\}$ is relatively compact for the sequence $\{P_{\theta,n}\}$. Let $\theta_n^*$, $n \geq 1$, be the functions of $\hat{\theta}_n$ satisfying the requirements of lemma 7. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition in the strict sense for all $\theta \varepsilon (\underline{\overline{H}})$. Then the sequence $\{V_n\}$ of estimators, constructed by $V_n = \theta_n^* + \delta_n \triangle_n(\theta_n^*)$, $n \geq 1$, is a sequence of ACS estimators for all $\theta \varepsilon (\underline{\overline{H}})$, where the sequence $\{\triangle_n(\theta)\}$ of $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$-measurable k-vectors is the one constructed in the statement (iii) of Lemma 5.

<u>Proof</u>. In view of the statement (i) of Lemma 5, it is enough to show that the difference $\delta_n^{-1}(V_n - \theta) - \triangle_n(\theta)$ converges to zero in $P_{\theta,n}$ probability for every $\theta \varepsilon (\underline{\overline{H}})$. Now

$$\delta_n^{-1}(V_n - \theta) = \delta_n^{-1}(\theta_n^* - \theta) + \Delta_n(\theta^*).$$

In view of the statement (iii) of Lemma 5 and since the sequence $\{\theta_n^*\}$ satisfies the requirements of Lemma 7 it follows that the difference

$$\Delta_n(\theta_n^*) - [\Delta_n(\theta) - \delta_n^{-1}(\theta_n^* - \theta)]$$

converges to zero in $P_{\theta,n}$-probability for every $\theta \varepsilon (\underline{H})$. This completes the proof.

<u>Theorem 2</u>. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition in the strict sense for every $\theta \varepsilon (\underline{H})$. Let $\{V_n\}$ be a sequence of ACS estimators for every $\theta \varepsilon (\underline{H})$. Let $V_n^*$, $n \geq 1$, be the functions of $V_n$ satisfying the requirements of lemma 7. Let $\underline{\underline{F}}_n$ be the $\sigma$-field generated by the statistics $(V_n, T_n^*(V_n^*))$ where the sequence $\{T_n^*(\theta)\}$ of $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$-measurable p.d. matrices is the one constructed in the statement (iv) lemma 4. Then there is a sequence $\underline{\underline{E}}^* = \{\underline{\underline{X}}_n, \underline{\underline{A}}_n, P_{\theta,n}^*; \theta \varepsilon (\underline{H})\}$, $n \geq 1$, of experiments such that

(i) for each $\underline{\underline{E}}_n^*$ the $\sigma$-field $\underline{\underline{F}}_n$ is sufficient,

and (ii) for every $\theta \varepsilon (\underline{H})$ and $\alpha > 0$

$$\sup_{|h| \leq \alpha} \| P_{\theta+\delta_n h, n} - P_{\theta+\delta_n h, n}^* \| \longrightarrow 0.$$

<u>Proof</u>. In view of Lemma 6, the proof is identical to the proof of the corresponding result for the LAN case given in LeCam (1960) and (1974a).

CHAPTER 2

# DIFFERENTIABILITY IN QUADRATIC MEAN TYPE REGULARITY CONDITION AND THE LAMN EXPERIMENT

## 1. INTRODUCTION

In this chapter a differentiability in quadratic mean type regularity condition is introduced and it is shown that the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition under this condition. Recent results on martingale central limit theorems due to McLeish (1974), Hall (1977) and Aldous and Eagleson (1978) will be used to prove the weak convergence of log-likelihood ratios to a mixed normal distribution.

The results of the present chapter were originally obtained for the i.i.d. case by LeCam (1970). LAN - condition for the dependent observations has been recently studied by Roussas (1978) and for the independent but not necessarily identical case has been studied by Phillippou and Roussas (1973) and Ibraginov and Khasminskii (1975).

The results of the present chapter constitute a major part of sections 2 and 5 of Jeganathan (1979a); during the final stage of preparation of the present work we received a copy of Ph.D. thesis (1980, September), the results of the first chapter of which are related to the results of the present chapter.

In Section 2 we introduce some notations and the regularity conditions; the main result of this chapter is also stated in this section. In Section 3 the proof of the main result is presented through a series of lemmas; the proof is based mostly on the ideas of LeCam (1970), Roussas (1972 and 1979) and Ibragimov and Khasminskii (1975). In Section 3 we discuss, following the arguments of Hájek (1972) and LeCam (1970), some easily verifiable regularity conditions implying the more direct differentiability in quadratic mean type condition of Section 2.

## 2. NOTATIONS, DEFINITIONS AND THE MAIN RESULT

Let $(X_1, X_2, \ldots, X_n)$, $n \geq 1$, be a sequence of random vectors defined on a probability space $(\underline{\underline{X}}, \underline{\underline{A}}, P_\theta)$ where the k-dimensional parameter $\theta \varepsilon \underline{\underline{H}}$, an open subset of $R^k$, $k \geq 1$. Let $\underline{\underline{A}}_n = \sigma(X_1, \ldots, X_n)$ be the $\sigma$-field induced by the random vector $(X_1, \ldots, X_n)$ and $P_{\theta,n}$ be the restriction of $P_\theta$ to $\underline{\underline{A}}_n$. Let $\theta_o \varepsilon \underline{\underline{H}}$ be the 'true' value of the parameter. We further assume that, for $j \geq 2$, a regular conditional probability measure of $X_j$ given $(X_1, \ldots, X_{j-1})$ is absolutely continuous with respect to a $\sigma$-finite measure $\mu_j$ with a corresponding density $f_j(X_j | X_1, \ldots, X_{j-1}; \theta)$, and the probability measure of $X_1$ is absolutely continuous with respect to a $\sigma$-finite measure $\mu_1$ with a corresponding density $f_1(X_1; \theta)$. For the sake of simplicity we set $f_j(X_j | X_1, \ldots, X_{j-1}; \theta) = f_j(\theta)$, $j \geq 2$, and $f_1(X_1; \theta) = f_1(\theta)$.

Let

$$\Lambda_n(\theta,\theta_o) = \log \frac{\prod_{j=1}^{n} f_j(\theta)}{\prod_{j=1}^{n} f_j(\theta_o)}$$

$$= \sum_{j=1}^{n} \log \frac{f_j(\theta)}{f_j(\theta_o)} .$$

We now introduce the following set of assumptions. Notations of Ch. 1 are assumed in this chapter.

(A.1) There are p.d. matrices $\delta_n$, $n \geq 1$, depending neither on $\theta$ nor on the observations, and random vectors $\xi_j(\theta_o), j \geq 1$, such that for every $h \varepsilon R^k$

$$\sum_{j=1}^{n} E\left\{ \int [\xi_{nj}(\theta_o,h) - \frac{1}{2} h' \delta_n \xi_j(\theta_o)]^2 \, d\mu_j \right\} \longrightarrow 0$$

as $n \longrightarrow \infty$, where we set

$$\xi_{nj}(\theta_o,h) = f_j^{1/2}(\theta_o+\delta_n h) - f_j^{1/2}(\theta_o) .$$

Define

$$\eta_j(\theta_o) = \xi_j(\theta_o)/f_j^{1/2}(\theta_o) \quad \text{if} \quad f_j(\theta_o) \neq 0$$

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

(A.2) $E[\eta_j(\theta_o) | \underline{\underline{A}}_{j-1}] = 0$ for every $j \geq 1$.

(A.3) There exists a measurable function $T(\theta_o)$ mapping $\underline{\underline{X}}$ to the set $k \times k$ symmetric matrices such that $P_{\theta_o}(T(\theta_o)$ is p.d.$) = 1$ and the difference

$$\delta_n \sum_{j=1}^{n} E[\eta_j(\theta_o)\eta_j'(\theta_o) | \underline{\underline{A}}_{j-1}]\delta_n - T(\theta_o)$$

converges to zero in $P_{\theta_o}$ probability.

(A.4) For every $\varepsilon > 0$ and $h \in R^k$

$$\sum_{j=1}^{n} E[|h'\delta_n\eta_j(\theta_o)|^2 \, I(|h'\delta_n\eta_j(\theta_o)| > \varepsilon)] \longrightarrow 0 .$$

(A.5) For every $h \in R^k$, there exists a constant $K > 0$ such that

$$\sup_{n \geq 1} \sum_{j=1}^{n} E[|h'\delta_n\eta_j(\theta_o)|^2] \leq K .$$

Following is the main theorem of this chapter.

<u>Theorem 1</u>. Suppose that the assumptions (A.1) - (A.5) are satisfied. Then the sequence $\underline{\underline{E}}_n = \{\underline{\underline{X}}, \underline{\underline{A}}_n, P_{\theta,n}; \theta \in (\underline{\underline{H}})\}$, $n \geq 1$, of experiments satisfy the LAMN condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$ with

$$T_n(\theta_o) = \delta_n \sum_{j=1}^{n} E[\eta_j(\theta_o)\eta_j'(\theta_o) | \underline{\underline{A}}_{j-1}]\delta_n$$

and

$$W_n(\theta_o) = T_n^{-1/2}(\theta_o)\delta_n \sum_{j=1}^{n} \eta_j(\theta_o) .$$

<u>Remarks</u>. (1). Suppose that the assumption (A.1) is strengthened as follows : for every bounded sequence sequence $\{h_n\}$ of elements $R^k$

$$\sum_{j=1}^{n} E\left\{\int [\xi_{nj}(\theta_o, h_n) - \frac{1}{2} h_n'\delta_n\xi_j(\theta_o)]^2 \, d\mu_j\right\} \longrightarrow 0$$

as $n \longrightarrow \infty$.

Then this assumption together with the assumptions (A.2) - (A.5) implies that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the LAMN - condition in the strict sense at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$.

The proof of this statement is identical to the proof of the above theorem.

(2) Assumption (A.2) is imposed in order to invoke the central limit theorems for martingales. It is possible to relax the assumption (A.2) slightly if one uses the central limit theorems for 'near martingales' as considered by Hall (1977).

(3) It is possible to deduce the assumption (A.2) from (A.1) in some special cases; see e.g. LeCam (1970), Roussas (1972 and 1979). We were not able to deduce (A.2) from (A.1) in the general case.

(4) For each $n \geq 1$, the quantities $\sum_{j=1}^{n} E[\eta_j(\theta_o)\eta_j'(\theta_o)]$ and $\sum_{j=1}^{n} E[\eta_j(\theta_o)\eta_j'(\theta_o) | \underline{\underline{A}}_{j-1}]$ are generally called respectively the Fisher information matrix and the conditional Fishes information matrix.

## 3. PROOF OF THE THEOREM

The proof will be presented through a series of lemmas. We start with the following lemma, the proof of which is essentially contained in McLeish (1974, 3.15) (see also lemma (3.1) of Basawa and Scott (1977)).

<u>Lemma 1</u>. Suppose the assumptions (A.2) - (A.4) are satisfied. Then, for every $t \in R^k$, the difference

$$\sum_{j=1}^{n} |t'\delta_n \eta_j(\theta_o)|^2 - \sum_{j=1}^{n} E[|t'\delta_n \eta_j(\theta_o)|^2 | \underline{\underline{A}}_{j-1}]$$

converges to zero in $P_{\theta_o}$ probability.

Lemma 2. Suppose the assumptions (A.2) - (A.4) are satisfied. Then

$$\mathcal{L}(T_n(\theta_o), W_n(\theta_o) \mid P_{\theta_o,n}) \Longrightarrow \mathcal{L}(T(\theta_o), W \mid P_{\theta_o})$$

where $T_n(\theta_o)$ and $W_n(\theta_o)$, $n \geq 1$, are as defined in the theorem, and $W$ is a copy of $N(0,I)$ independent of $T(\theta_o)$.

Proof. Since, for every $\varepsilon > 0$ and $t \varepsilon R^k$,

$$E[\max_{j \leq n} |t'\delta_n \eta_j(\theta_o)|^2]$$

$$\leq \varepsilon^2 + \sum_{j=1}^{n} E[|t'\delta_n \eta_j(\theta_o)|^2 I(|t'\delta_n \eta_j(\theta_o)| > \varepsilon)]$$

we have by (A.4)

$$E[\max_{j \leq n} |t'\delta_n \eta_j(\theta_o)|^2] \longrightarrow 0 .$$

Now lemma 1 and (A.3) implies that for every $t \varepsilon R^k$ the quantity $\sum_{j=1}^{n} |t'\delta_n \eta_j(\theta_o)|^2$ converges in $P_{\theta_o}$ probability to $t'T(\theta_o)t$. Hence by Theorem 2 and the remarks preceding this theorem of Aldous and Eagleson (1978), we have

$$\delta_n \sum_{j=1}^{n} \eta_j(\theta_o) \Longrightarrow T^{1/2}(\theta_o)W \qquad (\text{stably}) \quad (P_{\theta_o})$$

where $W$ is a copy of $N(0,I)$, independent of $T(\theta_o)$. In particular,

$$\mathcal{L}(\delta_n \sum_{j=1}^{n} \eta_j(\theta_o), T(\theta_o) \mid P_{\theta_o}) \Longrightarrow \mathcal{L}(T^{1/2}(\theta_o)W, T(\theta_o)) .$$

In view of (A.3) we then have

$$\mathcal{L}(\delta_n \sum_{j=1}^{n} \eta_j(\theta_o), T_n(\theta_o) | P_{\theta_o,n}) \Longrightarrow \mathcal{L}(T^{1/2}(\theta_o)W, T(\theta_o))$$

This completes the proof by noting that

$$\delta_n \sum_{j=1}^{n} \eta_j(\theta_o) = T_n^{1/2}(\theta_o) W_n(\theta_o) .$$

The following lemma is an easy generalisation of Lemma 5 of LeCam (1974b).

Lemma 3. Suppose the assumption (A.1) is satisfied. Then, letting $Z_j$ for the indicator of the set $\{f_j(\theta_o) = 0\}$,

$$\sum_{j=1}^{n} E[\int Z_j f_j(\theta_o+\delta_n h) d\mu_j] \longrightarrow 0 \qquad (2.1)$$

and

$$\sum_{j=1}^{n} E[\int Z_j |h' \delta_n \xi_j(\theta_o)|^2 d\mu_j] \longrightarrow 0 \qquad (2.2)$$

as $n \longrightarrow \infty$ for every $h \varepsilon R^k$.

Proof. Fix $h \varepsilon R^k$. Let $Z_{1j}$ be the indicator of the set $\{f_j(\theta_o) = 0, h' \delta_n \xi_j(\theta_o) < 0\}$ and $Z_{2j}$ be the indicator of the set $\{f_j(\theta_o) = 0, h' \delta_n \xi_j(\theta_o) \geq 0\}$ so that $Z_j = Z_{1j} + Z_{2j}$. We then have

$$\sum_{j=1}^{n} E \left\{ \int Z_{1j} [\xi_{nj}(\theta_o, h) - \frac{1}{2} h' \delta_n \xi_j(\theta_o)]^2 d\mu_j \right\}$$

$$\geq \sum_{j=1}^{n} E[\int Z_{1j} f_j(\theta_o+\delta_n h) d\mu_j] + \frac{1}{4} \sum_{j=1}^{n} E[\int Z_{1j} |h' \delta_n \xi_j(\theta_o)|^2 d\mu_j].$$

By (A.1), the l.h.s. of the above expression tends to zero. Since both the terms of the r.h.s. of the above expression are positive, we have

$$\sum_{j=1}^{n} E\left[\int Z_{1j} |h' \delta_n \dot{\xi}_j(\theta_o)|^2 d\mu_j\right] \longrightarrow 0 \tag{2.3}$$

Now let $t < 0$. Then

$$\sum_{j=1}^{n} E\left\{\int Z_{2j} \left[\xi_{nj}(\theta_o, th) - \frac{t}{2} h' \delta_n \dot{\xi}_j(\theta_o)\right]^2 d\mu_j\right\}$$

$$\geq \sum_{j=1}^{n} E\left[\int Z_{2j} f_j(\theta_o + \delta_n th) d\mu_j\right]$$

$$+ \frac{t^2}{2} \sum_{j=1}^{n} E\left[\int Z_{2j} |h' \delta_n \dot{\xi}_j(\theta_o)|^2 d\mu_j\right]$$

By (A.1) this implies that

$$\sum_{j=1}^{n} E\left[\int Z_{2j} |h' \delta_n \dot{\xi}_j(\theta_o)|^2 d\mu_j\right] \longrightarrow 0 . \tag{2.4}$$

Combining (2.3) and (2.4) we have

$$\sum_{j=1}^{n} E\left[\int Z_j |h' \delta_n \dot{\xi}_j(\theta_o)|^2 d\mu_j\right] \longrightarrow 0.$$

This proves (2.2). To prove (2.1) consider the inequality

$$\sum_{j=1}^{n} E\left[\int Z_j \left|\xi_{nj}^2(\theta_o, h) - \frac{1}{4}|h' \delta_n \dot{\xi}_j(\theta_o)|^2\right| d\mu_j\right]$$

$$\leq 2 \sum_{j=1}^{n} E\left\{\int Z_j \left[\xi_{nj}(\theta_o, h) - \frac{1}{2} h' \delta_n \dot{\xi}_j(\theta_o)\right]^2 d\mu_j\right\}$$

$$+ \frac{1}{4} \sum_{j=1}^{n} E\left[\ Z_j |h' \delta_n \dot{\xi}_j(\theta_o)|^2 d\mu_j\right] ; \tag{2.5}$$

here we have used the inequality

$$|c^2 - d^2| \leq (1+\alpha)|c-d|^2 + d^2/\alpha\ , \quad \alpha > 0 \ \text{ and } \ c, d \in R . \tag{2.6}$$

The first term of the r.h.s. of (2.5) tends to zero by (A.1) while the second term tends to zero by (2.2). This completes the proof of the lemma.

To simplify the notation we set, in what follows,

$$\eta_{nj}(\theta_o,h) = [f_j^{1/2}(\theta_o+\delta_n h)/f_j^{1/2}(\theta_o)] - 1 \quad \text{if} \quad f_j(\theta_o) \neq 0$$

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

Lemma 4. Suppose the assumptions (A.1) and (A.5) are satisfied. Then

$$\sum_{j=1}^{n} \int |\xi_{nj}^2(\theta_o,h) - \tfrac{1}{4}|h'\delta_n\xi_j(\theta_o)|^2|\,d\mu_j \longrightarrow 0 \qquad (2.7)$$

in $P_{\theta_o}$ probability, and

$$\sum_{j=1}^{n} E[\,|\eta_{nj}^2(\theta_o,h) - \tfrac{1}{4}|h'\delta_n\eta_j(\theta_o)|^2|\,] \longrightarrow 0 \qquad (2.8)$$

Proof. Using the inequality (2.6) we have

$$\sum_{j=1}^{n} E[\,\int|\xi_{nj}^2(\theta_o,h) - \tfrac{1}{4}|h'\delta_n\xi_j(\theta_o)|^2|\,d\mu_j\,]$$

$$\leq (1+\alpha) \sum_{j=1}^{n} E\left\{\int[\xi_{nj}(\theta_o,h) - \tfrac{1}{2}\,h'\delta_n\xi_j(\theta_o)]^2\,d\mu_j\right.$$

$$+ \frac{1}{4\alpha} \sum_{j=1}^{n} E[\,\int|h'\delta_n\xi_j(\theta_o)|^2 d\mu_j\,], \quad \alpha > 0. \qquad (2.9)$$

For each fixed $\alpha > 0$, the first term of the r.h.s of (2.9) tends to zero as $n \longrightarrow \infty$ by (A.1). Now consider

$$\sum_{j=1}^{n} E[\,\int|h'\delta_n\xi_j(\theta_o)|^2 d\mu_j\,] = \sum_{j=1}^{n} E[\,|h'\delta_n\eta_j(\theta_o)|^2\,]$$

$$+ \sum_{j=1}^{n} E[\,\int Z_j\,|h'\delta_n\xi_j(\theta_o)|^2 d\mu_j\,]$$

where $Z_j$ is the indicator of the set $\{f_j(\theta_o) = 0\}$. Hence we see that the second term of the r.h.s. of (2.9) tends to zero, in view of (A.5) and (2.2), by first letting [illegible] and then $\alpha \longrightarrow \infty$. Thus r.h.s of (2.9) tends to zero by first letting $n \longrightarrow \infty$ and then $\alpha \longrightarrow \infty$. This proves (2.7). (2.8) also follows from this since

$$\sum_{j=1}^{n} E\left[\,|\eta_{nj}^{2}(\theta_o,h) - \tfrac{1}{4}|h'\delta_n\eta_j(\theta_o)|^2\,\right]$$

$$= \sum_{j=1}^{n} E\left[\int(1-Z_j)\,|\xi_{nj}^{2}(\theta_o,h) - \tfrac{1}{4}|h'\delta_n\xi_j(\theta_o)|^2|\,d\mu_j\right].$$

Hence the proof of the lemma is complete.

Lemma 5. Suppose that the assumptions (A.1) - (A.5) are satisfied. Then the difference

$$\sum_{j=1}^{n} \eta_{nj}^{2}(\theta_o,h) - \frac{1}{4}\,h'T(\theta_o)h$$

converges to zero in $P_{\theta_o}$ probability.

Proof. By assumption (A.3) and Lemma 1 we see that the difference

$$\sum_{j=1}^{n} |h'\delta_n\eta_j(\theta_o)|^2 - h'T(\theta_o)h$$

converges to zero in $P_{\theta_o}$ probability. Hence it is enough to show that the difference

$$\sum_{j=1}^{n} \eta_{nj}^{2}(\theta_o,h) - \frac{1}{4}\sum_{j=1}^{n} |h'\delta_n\eta_j(\theta_o)|^2$$

converges to zero in $P_{\theta_o}$ probability. This follows from (2.8) by applying Chebyshev's inequality.

Lemma 6. Suppose the assumptions (A.1) - (A.5) are satisfied. Then, for every $h \in R^k$,

$$\max_{j \le n} |\eta_{nj}(\theta_o,h)| \longrightarrow 0 \text{ in } P_{\theta_o} \text{ probability} \tag{2.11}$$

and

$$\sum_{j=1}^{n} |\eta_{nj}(\theta_o,h)|^3 \longrightarrow 0 \text{ in } P_{\theta_o} \text{ probability} \tag{2.12}$$

Proof. For every $\varepsilon > 0$, consider

$$P[\max_{j \le n} |\eta_{nj}(\theta_o,h)| > \varepsilon] \le \sum_{j=1}^{n} P[|\eta_{nj}(\theta_o,h)| > \varepsilon]$$

$$\le \sum_{j=1}^{n} P[|\eta_{nj}(\theta_o,h) - \frac{1}{2} h' \delta_n \eta_j(\theta_o)| > \varepsilon/2]$$

$$+ \sum_{j=1}^{n} P[|\frac{1}{2} h' \delta_n \eta_j(\theta_o)| > \varepsilon/2].$$

Now (2.10) implies by applying Chebychev's inequality, that

$$\sum_{j=1}^{n} P[|\eta_{nj}(\theta_o,h) - \frac{1}{2} h' \delta_n \eta_j(\theta_o)| > \varepsilon/2] \longrightarrow 0 \tag{2.13}$$

It is easily seen that the assumption (A.4) implies

$$\sum_{j=1}^{n} P[|\frac{1}{2} h' \delta_n \eta_j(\theta_o)| > \varepsilon/2] \longrightarrow 0. \tag{2.14}$$

Combining (2.13) and (2.14) we see that (2.11) is proved. To prove (2.12) consider

$$\sum_{j=1}^{n} |\eta_{nj}(\theta_o,h)|^3 \le \max_{j \le n} |\eta_{nj}(\theta_o,h)| \sum_{j=1}^{n} \eta_{nj}^2(\theta_o,h).$$

Hence (2.12) follows by applying Lemma 5 and (2.11). This completes the proof of the lemma.

<u>Lemma 7</u>. Suppose the assumptions (A.1) - (A.5) are satisfied. Then the quantity

$$2\sum_{j=1}^{n}\eta_{nj}(\theta_o,h) - h'\delta_n \sum_{j=1}^{n}\eta_j(\theta_o) + \frac{1}{4}h'T_n(\theta_o)h$$

converges to zero in $P_{\theta_o}$ probability for every $h \varepsilon R^k$.

Proof. Consider the identity

$$E[\eta_{nj}(\theta_o,h)|\underline{\underline{A}}_{j-1}] = \int f_j^{1/2}(\theta_o+\delta_n h)f^{1/2}(\theta_o)d\mu_j - 1$$

$$= -\frac{1}{2}\int \xi_{nj}^2(\theta_o,h)d\mu_j .$$

By (2.7), (2.2) and Lemma 1, we see from this identify that the difference

$$2\sum_{j=1}^{n}E[\eta_{nj}(\theta_o,h)|\underline{\underline{A}}_{j-1}] + \frac{1}{4}h'T_n(\theta_o)h$$

converges to zero in $P_{\theta_o,n}$ probability for every $h \varepsilon R^k$.

Hence, since $E[\eta_j(\theta_o)|\underline{\underline{A}}_{j-1}] = 0$, $j \geq 1$, it is enough to show that the quantity

$$\sum_{j=1}^{n}[Y_j - E(Y_j|\underline{\underline{A}}_{j-1})]$$

converges to zero in $P_{\theta_o,n}$ probability, where we set

$$Y_j = 2[\eta_{nj}(\theta_o,h) - \frac{1}{2}h'\delta_n\eta_j(\theta_o)].$$

Since the summands are martingale differences, we have

$$E\left\{\sum_{j=1}^{n}[Y_j - E(Y_j|\underline{\underline{A}}_{j-1})]\right\}^2$$

$$= \sum_{j=1}^{n}E[Y_j - E(Y_j|\underline{\underline{A}}_{j-1})]^2 \leq \sum_{j=1}^{n}E[|Y_j|^2]$$

$$\longrightarrow 0.$$

since

$$\sum_{j=1}^{n} E\left\{\left[\eta_{nj}(\theta_o,h) - \frac{1}{2} h' \delta_n \eta_j(\theta_o)\right]^2\right\}$$

$$= \sum_{j=1}^{n} E\left\{\int (1-Z_j)\left[\xi_{nj}(\theta_o,h) - \frac{1}{2} h' \delta_n \xi_j(\theta_o)\right]^2 d\mu_j\right\}$$

$$\longrightarrow 0 \text{ by } (A.1).$$

This completes the proof of the lemma by applying Chebyshev's inequality.

Proof of Theorem 1. In view of (2.11) and the Taylor's expansion we have, with $P_{\theta_o,n}$ probability tending to one, the equality

$$\Lambda_n(\theta_o+\delta_n h,\theta_o) = 2 \sum_{j=1}^{n} \log (1+\eta_{nj}(\theta_o,h))$$

$$= 2 \sum_{j=1}^{n} \eta_{nj}(\theta_o,h) - \sum_{j=1}^{n} \eta_{nj}^2(\theta_o,h) + \sum_{j=1}^{n} \alpha_{nj}|\eta_{nj}(\theta_o h)|^3$$

where $|\alpha_{nj}| \leq 1$. By (2.12) we then see that the quantity

$$\Lambda_n(\theta_o+\delta_n h,\theta_o) - 2 \sum_{j=1}^{n} \eta_{nj}(\theta_o,h) + \sum_{j=1}^{n} \eta_{nj}^2(\theta_o,h)$$

converges to zero in $P_{\theta_o}$-probability for every $h \in R^k$. Hence the result follows from Lemmas 2,5 and 7.

## 4. DISCUSSIONS ON THE ASSUMPTION (A.1) OF SECTION 2

The arguments of this section are based on LeCam (1970) and Hájek (1972).

Consider the following set of assumptions.

(A.6) The functions $f_j(X_j|X_1,\ldots,X_{j-1};\theta) = f_j(\theta) : (\underline{\overline{H}}) \longrightarrow R$ are absolutely continuous in $\theta$ for all $(X_1,\ldots,X_j)$, $j \geq 1$.

(A.7) For every $\theta\varepsilon(\underline{\overline{H}})$ the $\theta$ derivative $f_j'(\theta) = (\partial/\partial\theta)f_j(\theta)$ exists for $\mu_1\times\ldots\times\mu_j$ almost all $(X_1,\ldots,X_j)$, $j \geq 1$.

Define for every $\theta\varepsilon(\underline{\overline{H}})$ and $j \geq 1$

$$\begin{aligned}\xi_j(\theta) &= f_j'(\theta)f^{-1/2}(\theta) \quad \text{if the derivative exists and } f_j(\theta) > 0\\ &= 0 \quad \text{otherwise.}\end{aligned}$$

Suppose that we have selected the sequence $\{\delta_n\}$ ; one way of selection is to define

$$\delta_n'\delta_n = \left\{\sum_{j=1}^{n} E_\emptyset\left[\int\xi_j(\emptyset)\xi_j'(\emptyset)\,d\mu_j\right]\right\}^{-1} \text{ for some fixed } \emptyset\varepsilon(\underline{\overline{H}}).$$

(A.8) For every $h\varepsilon R^k$ and $\theta\varepsilon(\underline{\overline{H}})$

$$E\left[\int|h'\delta_n\xi_j(\theta)|^2 d\mu_j\right] < \infty, \quad 1 \leq j \leq n < \infty.$$

(A.9) For every $h\varepsilon R^k$ and for every $\theta\varepsilon(\underline{\overline{H}})$

$$\sup_{a\leq t\leq b}\sum_{j=1}^{n} E\left\{\int|h'\delta_n[\xi_j(\theta+t\delta_n h) - \xi_j(\theta)]|^2 d\mu_j\right\} \longrightarrow 0.$$

<u>Proposition</u>. Suppose the assumptions (A.6) - (A.9) are satisfied. Then the assumption (A.1) is satisfied for every $\theta\varepsilon(\underline{\overline{H}})$.

<u>Proof</u>. An application of the inequality (2.6) and (A.8) shows that (A.9) in particular entails, for every $h\varepsilon R^k$ and $P_{\theta,j-1}\times\mu_j$ almost all $(X_1,\ldots,X_j)$,

$$\int_a^b |h'\delta_n\xi_j(\theta+t\delta_n h)|^2 dt < \infty, \quad 1 \leq j \leq n < \infty.$$

Hence according to Lemma (A.1) of Hájek (1972, p.189), for every $\theta \varepsilon (\underline{H})$ and $h \varepsilon R^k$ the functions $t \longrightarrow f_j^{1/2}(\theta+t\delta_n h)$, $1 \leq j \leq n \leq \infty$ are absolutely continuous in the interval $(a,b)$ for $P_{\theta,j-1} \times \mu_j$ almost all $(X_1,\ldots,X_j)$. Hence we can write for $P_{\theta,j-1} \times \mu_j$ almost all $(X_1,\ldots,X_j)$, for all $h \varepsilon R^k$ and $\theta \varepsilon (\underline{H})$

$$f_j^{1/2}(\theta+t_2\delta_n h) - f_j^{1/2}(\theta+t_1\delta_n h) = \frac{1}{2} \int_{t_1}^{t_2} h'\delta_n \xi_j(\theta+t\delta_n h)\,dt$$

for every $t_1$ and $t_2$ such that $a < t_1 < t_2 < b$. Hence

$$\sum_{j=1}^{n} E\left[\int \left|f_j^{1/2}(\theta+\delta_n h) - f_j^{1/2}(\theta) - \frac{1}{2} h'\xi_j(\theta)\right|^2 d\mu_j\right]$$

$$= \frac{1}{4} \sum_{j=1}^{n} E\left\{\int \left|\int_0^1 h'\delta_n\left[\xi_j(\theta+t\delta_n h) - \xi_j(\theta)\right] dt\right|^2 d\mu_j\right\}$$

$$\leq \frac{1}{4} \sum_{j=1}^{n} E\left\{\int_0^1 dt \int \left|h'\delta_n\left[\xi_j(\theta+t\delta_n h) - \xi_j(\theta)\right]\right|^2 d\mu_j\right\}$$

$$\leq \frac{1}{4} \sup_{0 \leq t \leq 1} E\left\{\int \left|h'\delta_n\left[\xi_j(\theta+t\delta_n h) - \xi_j(\theta_o)\right]\right|^2 d\mu_j\right\}$$

$$\longrightarrow 0 \text{ by (A.9).}$$

The proof is complete.

Remark. In connection with the above result it should be mentioned here that LeCam (1974b) has given some results, based on Lusin's (N) - condition (cf. Hewitt and Stromberg (1965, p.288)) instead of absolute continuity, which are applicable to more general situations; LeCam's arguments are restricted to the i.i.d case but the above discussion shows that his arguments are applicable to the general case also.

CHAPTER 3

# SOME RESULTS CONCERNING THE INVARIANCE OF THE POSSIBLE LIMITS OF DISTRIBUTIONS

## 1. INTRODUCTION

The purpose of this chapter is to present two results concerning certain types of invariance of limit distributions. A very detailed and deep discussions on the invariance of the possible limits of experiments and distributions can be found in LeCam (1974, Ch.11 and 1979, Ch.8). The first result of the present chapter is a related, but different, version, and the second one is a strengthened form, of an invariance result given in LeCam (1979, Ch.8). These invariance results of LeCam look so simple and innocent, but their power appears to be remarkably surprising, as is seen from the applications given in Chapters 4,5 and 6. One can also see that the basic ideas of these invariance results are implicit in the proof of Theorem (4.1) of LeCam (1960). Another paper where appropriate rescaling procedure has been employed and then the idea of invariance in some sense was used is Bahadur (1964).

In Section 2 we first introduce some basic terminologies needed to state the results and then state our results. Proofs of the results are presented in Section 3.

## 2. STATEMENTS OF THE RESULTS

Let $G(\underline{\underline{B}}^q)$ be the space of all sub-stochastic measures on $\underline{\underline{B}}^q$, $\underline{\underline{B}}^q$ being the Borel σ-field on $R^q, q \geq 1$. Let $(\underline{\underline{S}}, \underline{\underline{F}}, v)$ be a σ-finite measure space. Consider the (sub-stochastic) kernels $P : \underline{\underline{S}} \longrightarrow G(\underline{\underline{B}}^q)$. Let $C_{oo}$ be the space of continuous functions vanishing outside compacts. Define $C_{oo}(R^q) \underline{\otimes} L_1(v)$ topology of the set of all kernels to be the smallest topology such that all functions

$$P \longrightarrow \iint f(x)P(t)(dx)g(t)v(dt)$$

$f \varepsilon C_{oo}$, $g \varepsilon L_1(v)$, are continuous; this topology was introduced in LeCam (1973). It is known that the set of all kernels endowed with this topology is meterizable and compact; a proof can be found in LeCam (1979, Ch. 8).

We now state the results of this chapter. In what follows $\mu^k | \underline{\underline{B}}^k$ denotes the Lebesgue measure.

Theorem 1. Let $(\underline{H})$ be a measurable subset of $R^k$. Let $\{F_n\}$ be a sequence of kernels $F_n(\theta,h) = F_{\theta+\delta_n h,n} : (\underline{H}) \times R^k \longrightarrow G(\underline{\underline{B}}^q)$, where $\{\delta_n\}$ is a sequence of p.d. matrices such that $\|\delta_n\| \longrightarrow 0$. Then the following two statements hold.

(i) The sequence $\theta \longrightarrow F_n(\theta,h), h\varepsilon R^k, n \geq 1$, is $C_{oo}(R^q) \underline{\otimes} L_1(\mu^k)$ convergent to a kernel $F(\theta,h)$ if and only if the sequence $\theta \longrightarrow F_n(\theta,0)$, $n \geq 1$, is $C_{oo}(R^q) \underline{\otimes} L_1(\mu^k)$ convergent to the kernel $F(\theta,0) = F(\theta)$.

(ii) The kernel $F(\theta,h)$ satisfies the invariance condition

$$\int_{(\bar{H})}\int_{R^q} f(x)F(\theta)(dx)g(\theta)\mu^k(d\theta) = \int_{(\bar{H})}\int_{R^q} f(x)F(\theta,h)(dx)g(\theta)\mu^k(d\theta)$$

for every $f \varepsilon C_{oo}, g \varepsilon L_1(\mu^k)$ and $h \varepsilon R^k$.

Theorem 2. Assume that $(\bar{H})$ and the sequences $\{F_n\}$ and $\{\delta_n\}$ are as in the above Theorem 1. Then the following two statements hold.

(i) The sequence $(\theta,h) \longrightarrow F_n(\theta,h+u)$, $n \geq 1, u \varepsilon R^k$, is $C_{oo}(R^q) \otimes L_1(\mu^k \otimes \mu^k)$ convergent to a kernel $F(\theta,h+u)$ if and only if the sequence $\theta \longrightarrow F_n(\theta,0), n \geq 1$, is $C_{oo}(R^q) \otimes L_1(\mu^k)$ convergent to a kernel $K(\theta)$.

(ii) Let $F(\theta,h+u)$ and $K(\theta)$ be as above. Then

$$\int_{(\bar{H})}\int_{R^k}\int_{R^q} f(x)F(\theta,h+u)(dx)g(h)\mu^k(dh)m(\theta)\mu^k(d\theta)$$

$$= \int_{R^k} g(h)\mu^k(dh) \int_{(\bar{H})}\int_{R^q} f(x)K(\theta)(dx)m(\theta)\mu^k(d\theta)$$

for every $g,m \varepsilon L_1(\mu^k), f \varepsilon C_{oo}$ and $u \varepsilon R^k$.

The following invariance result presented in LeCam (1979, Ch.8) and Strasser (1978) follows from the above Theorem 2.

Theorem 3. Assume that $(\bar{H})$ and the sequences $\{F_n\}$ and $\{\delta_n\}$ are as in Theorem 1. Then the following two statements hold.

(i) The sequence $(\theta,h) \longrightarrow F_n(\theta,h+u), n \geq 1, u \varepsilon R^k$, is $C_{oo}(R^q) \otimes L_1(\mu^k \otimes \mu^k)$ convergent to a kernel $F(\theta,h+u)$ if and only if the sequence $(\theta,h) \longrightarrow F_n(\theta,h), n \geq 1$, is $C_{oo}(R^q) \underline{\otimes} L_1(\mu^k \underline{\otimes} \mu^k)$ convergent to the kernel $F(\theta,h)$.

(ii) The kernel $F(\theta,h+u)$ satisfies the invariance condition

$$\int_{\underline{H}} \int_{R^k} \int_{R^q} f(x)F(\theta,h+u)(dx)g(h)\mu^k(dh)n(\theta)\mu^k(d\theta)$$

$$= \int_{\underline{H}} \int_{R^k} \int_{R^q} f(x)F(\theta,h)(dx)g(h)\mu^k(dh)n(\theta)\mu^k(d\theta)$$

for every $g,n \in L_1(\mu^k), f \in C_{oo}$ and $u \in R^k$.

## 3. PROOFS OF THE RESULTS

Proof of the Theorem 1. To prove the result it is enough to consider a subsequence $\{r\} \subseteq \{n\}$ such that both the sequences $\theta \longrightarrow F_r(\theta,h), r \geq 1, h \in R^k$, and $\theta \longrightarrow F_r(\theta,0), r \geq 1$, are $C_{oo}(R^q) \underline{\otimes} L_1(\mu^k)$ convergent to some kernels $F(\theta,h)$ and $F(\theta)$ respectively. For simplicity assume that $\underline{H} = R$ and $q = k = 1$. Now note that

$$\int_{t_1}^{t_2} \int_R f(x)F_r(\theta,h)(dx)\mu(d\theta) = \int_{t_1+\delta_n h}^{t_2+\delta_n h} \int_R f(x)F_r(\theta,0)(dx)\mu(d\theta)$$

for every $t_1, t_2, h \in R, t_1 < t_2$ and $f \in C_{oo}$. Hence the difference

$$\int_{t_1}^{t_2} \int_R f(x)F_r(\theta,h)(dx)\mu(d\theta) - \int_{t_1}^{t_2} \int_R f(x)F_r(\theta,0)(dx)\mu(d\theta)$$

is absolutely bounded by

$$C\mu([t_1,t_2] \triangle ([t_1,t_2] + \delta_r h)) \quad (\text{for some} \quad C > 0)$$

$$\leq C\, 2\, \delta_r |h| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty .$$

Hence we have

$$\int_{t_1}^{t_2} \int_R f(x) F(\theta,h)(dx)\mu(d\theta) = \int_{t_1}^{t_2} \int_R f(x) F(\theta)(dx)\mu(d\theta)$$

for every $t_1, t_2 \varepsilon R$ and $f \varepsilon C_{oo}$.

This implies

$$\int_R f(x) F(\theta,h)(dx) = \int_R f(x) F(\theta)(dx) \quad \text{a.s [Lebesgue]}$$

for every $h \varepsilon R$ and $f \varepsilon C_{oo}$. This proves the result.

Proof of the Theorem 2. To prove the result it is enough to consider a subsequence $\{r\} \subseteq \{n\}$ such that the sequence $(\theta,h) \longrightarrow F_r(\theta,h+u), r \geq 1, u \varepsilon R^k$, is $C_{oo}(R^q) \otimes L_1(\mu^k \otimes \mu^k)$ convergent to a kernel $F(\theta,h+u)$, and the sequence $\theta \longrightarrow F_r(\theta,0), r \geq 1$, is $C_{oo}(R^q) \otimes L_1(\mu^k)$ convergent to a kernel $K(\theta)$. For simplicity assume that $(\underline{\overline{H}}) = R^k, k = q = 1$. As in the proof of Theorem 1, the proof will follow if we show that the difference

$$\int_{t_1}^{t_2} \int_R \int_R f(x) F_r(\theta,h+u)(dx) g(h)\mu(dh)\mu(d\theta)$$

$$- \int_R g(h)\mu(dh) \int_{t_1}^{t_2} \int_R f(x) F_r(\theta,0)(dx)\mu(d\theta)$$

converges to zero for every $t_1, t_2, u \varepsilon R, t_1 < t_2, g \varepsilon L_1(\mu)$ and $f \varepsilon C_{oo}$. To show this it is enough to show that the difference

$$\int_{t_1}^{t_2} \int_R f(x) F_r(\theta, t)(dx)\mu(d\theta) - \int_{t_1}^{t_2} \int_R f(x) F_r(\theta, 0)(dx)\mu(d\theta)$$

converges to zero for every $t_1, t_2, t \varepsilon R, t_1 < t_2$ and $f \varepsilon C_{oo}$. This is proved in the proof of Theorem 1. Hence the proof is complete.

CHAPTER 4

# ASYMPTOTIC DIFFERENTIABILITY OF THE LOG-LIKELIHOOD RATIOS AND THE LAMN EXPERIMENTS

## 1. INTRODUCTION

LeCam (1960) has shown that a certain kind of asymptotic differentiability of the log-likelihood ratios together with the contiguity condition implies that the limit distribution of the suitably normalised log-likelihood function is normal. A remarkable thing to be noted here is that the asymptotic normality occur through an argument which has nothing to do with sums of independent random variables or martingale differences. It is the purpose of this chapter to extend and strengthen this and other related results of LeCam to a situation where the limit of the experiments turns out to be a mixed normal experiment.

More specifically, in his definition of asymptotic differentiability, LeCam assumed that the sequence of normalised log-likelihood ratios is approximated, with probability tending to one, by the sum of two expressions, the first one being a sequence of random linear functions of the normalised parameter and the second one being a non-random function of the normalised parameter. In this chapter we assume that this second expression is also a sequence of random functions of the normalised parameter and then we first establish (Theorem 1) that the limit distribution, when it exists, is a mixed normal for almost all points of the parameter space. Secondly we establish (Theorem 2), without assuming the existence of the limit distribution, the log-likelihood ratios converge in the weak topology introduced in Ch. 3 to a mixed normal distribution; though the convergence stated here is very much weaker than the convergence stated in Theorem 1, Theorem 1 actually follows from this result and it appears that this result is more important than Theorem 1. In the special case when the second expression mentioned above is assumed to be a non-random function of the normalised parameter, it is possible to obtain the convergence stated in Theorem 1 under the assumptions of Theorem 2 (see the remark following this theorem 2) ; thus the conclusion of Theorem 4.1 of LeCam (1960) holds even when the existance of the limit distribution is not assumed. Thirdly we establish (Theorem 3) that, when the second expression mentioned above is a sequence of random quadratic forms of the normalised parameter and when the sequence of random matrices of this quadratic forms satisfies a certain invariance condition, the limit distribution is a mixed normal for all points of the parameter space. It may be mentioned here that the contiguity condition plays a crucial role in establishing all these results.

Our approximation of the log-likelihood ratios, stated in section 2, is slightly weaker than the one assumed in LeCam (1960), and therefore we will have to further assume that the random quantities involved in the approximation are jointly measurable in the observations and the parameter, and that the given sequence of family of probability measures are measurable in a certain sense. In Section 4 it is shown that these measurability restrictions can be removed when the approximation of the log-likelihood ratios is analogous to the one assumed in LeCam (1960).

Assumptions and the main results are stated in Section 2 and the proofs of the main results are presented in Section 3.

This chapter is the revised version of Jeganathan (1979d); after the completion of this ~~work~~ paper we came to know of a related work Davies (1979) which contains a version of the third result (Theorem 3) of the present paper.

## 2. ASSUMPTIONS AND THE MAIN RESULTS

Let $\underline{\underline{E}}_n = \{\underline{\underline{X}}_n, \underline{\underline{A}}_n, P_{\theta,n}; \theta \varepsilon \underline{\textcircled{H}}\}, n \geq 1$, be a sequence of experiments; through out this chapter it will be assumed, without further mentioning, that $\underline{\textcircled{H}}$ is an open subset of $R^k, k \geq 1$.

Notations of Ch.1 are assumed in this chapter.

<u>Definition</u>. A sequence of experiments $\underline{\underline{E}}_n = \{\underline{\underline{X}}_n, \underline{\underline{A}}_n, P_{\theta,n}; \theta\varepsilon \underline{\textcircled{H}}\}$, $n \geq 1$, will be called asymptotically differentiable on $\underline{\textcircled{H}}$ if the following six assumptions are satisfied.

(A.1) The functions $\theta \longrightarrow P_{\theta,n}(A), A \varepsilon \underline{\underline{A}}, n \geq 1$, are Borel measurable.

(A.2) There exists $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$ -measurable functions $W_n(\cdot) : \underline{\underline{X}}_n \times \underline{\textcircled{H}} \longrightarrow R^k$ and $\Lambda_n(h,\cdot) : \underline{\underline{X}}_n \times \underline{\textcircled{H}} \longrightarrow R$, $n \geq 1$, $h \varepsilon R^k$, such that the difference

$$\frac{dP_{\theta+\delta_n h,n}}{dP_{\theta,n}} - \exp[h'W_n(\theta) - A_n(h,\theta)]$$

converges to zero in $P_{\theta,n}$ - probability for every $h \varepsilon R^k$ and $\theta \varepsilon (\underline{\underline{H}})$, where $\{\delta_n\}$ is a sequence of positive definite (p.d.) matrices such that $\|\delta_n\| \longrightarrow 0$ as $n \longrightarrow \infty$.

(A.3) The sequences $\{P_{\theta+\delta_n h,n}\}$ and $\{P_{\theta,n}\}$ are contiguous for every $h \varepsilon R^k$ and $\theta \varepsilon (\underline{\underline{H}})$.

(A.4) For every $\theta \varepsilon (\underline{\underline{H}})$, there exist a random function $h \longrightarrow A(h,\theta)$ and a random vector $W(\theta)$ defined on some probability space $(\underline{\underline{Y}}, \underline{\underline{F}}, \lambda_\theta)$ such that for every finite $\{h_i ; i = 1,2,\ldots,n\}$

$$\mathcal{L}(W_n(\theta), A_n(h_i,\theta), i = 1,2,\ldots,m | P_{\theta,n})$$

$$\Longrightarrow \mathcal{L}(W(\theta), A(h_i,\theta), i = 1,2,\ldots,m | \lambda_\theta).$$

(A.5) For every $\theta \varepsilon (\underline{\underline{H}})$, there exists a set $N_\theta \varepsilon \underline{\underline{F}}$ of $\lambda_\theta$ - measure zero such that the functions $h \longrightarrow A(h,\theta)$ are continuous for all points outside the set $N_\theta$.

(A.6) For every $s, h \varepsilon R^k$ and $\theta \varepsilon (\underline{\underline{H}})$, the difference

$$A_n(h, \theta+\delta_n s) - A_n(h,\theta)$$

converges to zero in $P_{\theta,n}$ - probability.

<u>Theorem 1</u>. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the conditions (A.1) - (A.6). Then there are $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$ -

measurable k-vectors $\gamma_n(\theta)$ and positive semi-definite (p.s.d.) $k \times k$ matrices $T_n(\theta)$, $n \geq 1$, $\underline{\underline{F}} \times \underline{\underline{B}}^k$ -measurable k-vector $\gamma(\theta)$ and a p.s.d. $k \times k$ matrix $T(\theta)$, and a Lebesgue null set $N \subseteq (\underline{H})$ such that for every $\theta \varepsilon (\underline{H}) - N$

(i) the difference

$$A_n(h,\theta) - [h'\gamma_n(\theta) + \frac{1}{2} h' T_n(\theta) h]$$

converges to zero in $P_{\theta,n}$ - probability, and

(ii) $$\mathcal{L}(W_n(\theta), \gamma_n(\theta), T_n(\theta) | P_{\theta,n})$$

$$\Longrightarrow \mathcal{L}(T^{1/2}(\theta) Z + \gamma(\theta), \gamma(\theta), T(\theta) | \lambda_\theta)$$

where $Z$ is a copy of the standard k-variate normal distribution independent of both $\gamma(\theta)$ and $T(\theta)$.

<u>Corollary</u>. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the conditions (A.1) - (A.6). Further assume that $\mathcal{L}(W(\theta), A(h_i,\theta), i = 1,2,\ldots,m | \lambda_\theta)$ is a continuous function of $\theta$ for every finite $\{h_i ; i = 1,2,\ldots,m\}$. Then the statements (i) and (ii) of Theorem 1 hold for every $\theta \varepsilon (\underline{H})$.

<u>Theorem 2</u>. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies, in addition to the conditions (A.1)-(A.3), (A.5) and (A.6), the condition

(A.4') for every finite $\{h_i ; i = 1,2,\ldots,m\}$ and $\theta \varepsilon (\underline{H})$

$$\mathcal{L}(A_n(h_i,\theta) ; i = 1,2,\ldots,m | P_{\theta,n})$$

$$\Longrightarrow \mathcal{L}(A(h_i,\theta) ; i = 1,2,\ldots,m | \lambda_\theta).$$

Let the random functions $\gamma_n(\theta), T_n(\theta), n \geq 1, \gamma(\theta)$ and $T(\theta)$ be as in Theorem 1. Then

(i) there exists a Lebesgue null set $N$ ($\subseteq$ ($\underline{H}$) such that for every $\theta \varepsilon (\underline{H})$, the difference

$$A_n(h,\theta) - [h'\gamma_n(\theta) + \frac{1}{2} h' T_n(\theta) h]$$

converges to zero in $P_{\theta,n}$ probability, and

(ii) the sequence $\mathcal{L}(W_n(\theta), \gamma(\theta), T_n(\theta) | P_{\theta,n})$ is $C_{oo}(R^{2k+k^2}) \underline{\otimes} L_1(\mu^k)$ convergent to the stochastic kernel $\mathcal{L}(T^{1/2}(\theta)Z + \gamma(\theta), \gamma(\theta), T(\theta) | \lambda_\theta)$.

Note that in Theorem 2, the existence of the limit distribution of the sequence $\{W_n(\theta)\}$ is not assumed.

<u>Remark</u>. In the special case when $A_n(h,\theta) = A(h,\theta)$ for every $n \geq 1$ where the function $h \longrightarrow A(h,\theta)$ is non random, it is easy to see directly from (i) of Theorem 2 that there exists a Lebesgue null set $N$ ($\subseteq$ ($\underline{H}$) such that for every $\theta \varepsilon (\underline{H}) - N$ $\mathcal{L}(W_n(\theta) | P_{\theta,n})$ converges weakly to the k-variate normal distribution with mean vector $\gamma(\theta)$ and covariance matrix $T(\theta)$.

<u>Theorem 3</u>. Assume that, for $\theta_o \varepsilon (\underline{H})$,

(i) the assumption (A.1) of Def. 1 of Ch. 1 is statisfied,

(ii) there exists an almost surely p.s.d. matrix $T(\theta_o)$ such that

$$\mathcal{L}(T_n(\theta_o) | P_{\theta,n}) \Longrightarrow \mathcal{L}(T(\theta_o)).$$

Then the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o \in (\underline{H})$ if and only if

(iii) the sequences $\{P_{\theta_o,n}\}$ and $\{P_{\theta_o+\delta_n h,n}\}$, $n \geq 1$, are contiguous for every $h \in R^k$, and

(iv) $\mathcal{L}(T_n(\theta_o) | P_{\theta_o+\delta_n h,n}) \Rightarrow \mathcal{L}(T(\theta_o))$ for every $h \in R^k$.

## 3. PROOFS OF THE RESULTS

Note that the theorem 1 actually follows from Theorem 2. However we will present the proof of Theorem 1 only for the following reasons. Firstly the proofs of both the theorems are essentially identical. Secondly the arguments in the proof of Theorem 1 are more transparent and notationally less cumbersome.

We will assume in what follows, for the sake of simplicity only, that $\dim(\underline{H}) = 1$. The proof of the next lemma is based on the proposition 1 of LeCam (1974, Ch. 11).

<u>Lemma 1</u>. Suppose that the assumptions of Theorem 1 are satisfied. Then there exist random variables $\gamma(\theta)$ and $T(\theta)$, and a Lebesgue null set $N \subseteq (\underline{H})$ such that

$$A(h,\theta) = h\gamma(\theta) + \frac{1}{2} h^2 T(\theta) \quad \text{a.s.}$$

for every $h \in R$ and $\theta \in (\underline{H}) - N$.

<u>Proof</u>. Denote, for $s, h \in R$ and $\theta \in (\underline{H})$,

$$Z_{n,\theta}(h|s) = \frac{dP_{\theta+\delta_n h,h}}{dP_{\theta+\delta_n s,n}} .$$

Further, the vector whose elements are $A_n(s,\theta), A_n(h,\theta)$, $A_n(\frac{s+h}{2}, \theta), A_n(s+u,\theta), A_n(h+u,\theta)$ and $A_n(\frac{s+h}{2} + u, \theta)$ will be denoted by $V_n(s,h,u,\theta)$, and the vector whose elements are $A(s,\theta), A(h,\theta), A(\frac{s+h}{2}), \theta), A(s+u,\theta), A(h+u,\theta)$ and $A(\frac{h+s}{2} + u, \theta)$ will be denoted by $V(s,h,u,\theta)$. In view of the statement (6) of Theorem 2.1 of LeCam (1960), it follows from the given conditions that

$$\mathcal{L}(V_n(s,h,u,\theta), Z_{\theta,n}(h+w|s+w) | P_{\theta+\delta_n(s+w),n})$$
$$\Longrightarrow \mathcal{L}(V(s,h,u,\theta), \frac{dG_{\theta,h+w}}{dG_{\theta,s+w}} | G_{\theta,s+w}) \qquad (4.1)$$

for every $\theta\varepsilon(\underline{H})$ and $s,h,u,w\varepsilon R$, where

$$dG_{\theta,h} = \exp[hW(\theta) - A(h,\theta)]\, d\lambda_\theta, \; h\varepsilon R.$$

(Note that in view of the condiguity condition $G_{\theta,h}$ is a probability measure for every $\theta\varepsilon(\underline{H})$ and $h\varepsilon R$.) Further, the condition (A.6) and the invariance theorem 1 of Ch.3 implies that both the sequences

$$\mathcal{L}(V_n(s,h,u,\theta), \; Z_{\theta,n}(h|s) | P_{\theta+\delta_n s,n})$$

and

$$\mathcal{L}(V_n(s,h,u,\theta), \; Z_{\theta,n}(h+w|s+w) | P_{\theta+\delta_n(s+w),n})$$

are $C_{oo}(R^7) \otimes L_1(\mu)$ convergent to a same kernel for every $s,h,u,w\varepsilon R$. Hence, this fact together with (4.1) implies that, for every $s,h,u,w \varepsilon R$, there exists a Lebesgue null set $N(s,h,u,w)$ possibly depending on $(s,h,u,w)$ such that

$$\mathcal{L}\left(V(s,h,u,\theta), \frac{dG_{\theta,h+w}}{dG_{\theta,s+w}} \Big| G_{\theta,s+w}\right)$$

$$= \mathcal{L}\left(V(w,h,u,\theta), \frac{dG_{\theta,h}}{dG_{\theta,s}} \Big| G_{\theta,s}\right) \tag{4.2}$$

whenever $\theta \varepsilon (\underline{H}) - N(s,h,u,w)$. Let $D$ be the set of all points in $R^4$ with rational co-ordinates, and let $N = \bigcup_{(s,h,u,w)\varepsilon D} N(s,h,u,w)$. Then it follows that, whenever $\theta\varepsilon(\underline{H}) - N$, the equality in the expression (4.2) holds for every $(s,h,u,w)\varepsilon D$. In particular it easily follows that, for every $(s,h,u,w)\varepsilon D$ and $\theta\varepsilon(\underline{H}) - N$,

$$E^V\left\{\sqrt{\frac{dG_{\theta,h+u}}{d\lambda_\theta}\ \frac{dG_{\theta,s+u}}{d\lambda_\theta}}\right\} = E^V\left\{\sqrt{\frac{dG_{\theta,h}}{d\lambda_\theta}\ \frac{dG_{\theta,s}}{d\lambda_\theta}}\right\} \text{ a.s} \tag{4.3}$$

and

$$E^V\left(\frac{dG_{\theta,w}}{d\lambda_\theta}\right) = 1 \quad \text{a.s.} \tag{4.4}$$

where $E^V$ denotes the conditional expectation given $V(s,h,u,\theta)$ with the underlying probability space being $(\underline{Y}, \underline{F}, \lambda_\theta)$. In what follows assume that $\theta \varepsilon (\underline{H}) - N$ is fixed. Since we have, as is easily checked using (4.4),

$$\log E^V\left\{\sqrt{\frac{dG_{\theta,h+u}}{d\lambda_\theta}\ \frac{dG_{\theta,s+u}}{d\lambda_\theta}}\right\}$$

$$= -\frac{1}{2}\left[A(s+u,\theta) + A(h+u,\theta) - 2A\left(\frac{s+h}{2} + u,\theta\right)\right]$$

for every rationals s,h,u$\varepsilon$R, it follows from (4.3) that for every rationals s,h,u,$\varepsilon$R

$$A(s+u,\theta) + A(h+u,\theta) - 2A(\frac{s+h}{2} + u,\theta)$$

$$= A(s,\theta) + A(h,\theta) - 2A(\frac{s+h}{2}, \theta) \quad \text{a.s.}$$

Hence in view of the condition (A.5) there exists a set of $\lambda_\theta$-measure zero such that outside this null set

$$A(s+u,\theta) + A(h+u,\theta) - 2A(\frac{s+h}{2} + u,\theta)$$

$$= A(s,\theta) + A(h,\theta) - 2A(\frac{s+h}{2}, \theta)$$

for every s,h,u$\varepsilon$R, i.e., the random function $h \longrightarrow A(h,\theta)$ has constant second differences outside a set of $\lambda_\theta$-measure zero. This implies that there exist random variables $\gamma(\theta)$ and $T(\theta)$ such that, (note that $A(0,\theta) = 0$ a.s.),

$$A(h,\theta) = h\gamma(\theta) + \frac{1}{2} h^2 T(\theta) \quad \text{a.s.}$$

for every $\theta\varepsilon(\underline{H}) - N$. This completes the proof of the lemma.

<u>Proof of Theorem 1.</u> First note that for almost all $(\underline{H}), T(\theta)$ is the second difference of the random function $h \longrightarrow A(h,\theta)$ at $h=0$ and that $\gamma(\theta)$ can be expressed in terms of the first and second differences of the function $h \longrightarrow A(h,\theta)$ at $h = 0$. Hence it follows from the relation (4.4) and Lemma 1 together with a simple continuity argument that there exists a Lebesgue null set $N \subseteq (\underline{H})$ such that

$$E^{(\gamma(\theta),T(\theta))}[\exp hW(\theta)] = \exp[h\gamma(\theta) + \frac{1}{2} h^2 T(\theta)] \quad \text{a.s.}$$

for every $h \in R$ and $\theta \in (\underline{H}) - N$. Hence it follows that $W(\theta)$ is distributed as $T^{1/2}(\theta)Z + \gamma(\theta)$ where $Z$ is a copy of the standard normal distribution independent of both $\gamma(\theta)$ and $T(\theta)$. Now let $T_n(\theta)$ be the second difference of $h \longrightarrow A_n(h,\theta)$ at $h = 0$, $\theta \in (\underline{H}), n \geq 1$, and define $\gamma_n(\theta)$ similarly. This $T_n(\theta)$ need not be non-negative, but this can be easily remedied since $T(\theta)$ is non-negative. This completes the proof of Theorem 1.

Proof of Theorem 3. The proof of the necessary part follows from Lemmas 1 and 2 of Ch. 1. To prove the sufficiency part, first note that, in view of the statement (4) of Theorem 2.1 of LeCam (1960), the sequence $\left\{T_n^{1/2}(\theta_o)W_n(\theta_o), T_n(\theta_o)\right\}$ is relatively compact for the sequence $\left\{P_{\theta_o,n}\right\}$. Hence for every sub-sequence there exists a further sub-sequence $\{m\} \subseteq \{n\}$ and a random vector $(T', W)$ such that

$$\mathcal{L}(T_n(\theta_o)\ ,\ T_m^{1/2}(\theta_o)W_m(\theta_o)\,|\,P_{\theta_o,m}) \Longrightarrow \mathcal{L}(T', W).$$

In view of the statement (6) of Theorem (2.1) of LeCam (1960) we then have from the statement (iii), that

$$\mathcal{L}(T') = \mathcal{L}(T'\,|\,R_h) \qquad (4.5)$$

for every $h \in R$, where the probability measure $R_h$ is defined by

$$R_h = \exp(hW - \frac{h^2}{2}T')d\,\mathcal{L}(T', W).$$

In particular (4.5) implies that

$$E^{T'}\left[\exp(hW)\right] = \exp(\tfrac{1}{2}h^2T')$$

and hence

$$E[\exp(itT' + iuW)] = E[\exp(itT' - \tfrac{1}{2} u^2 T')]$$

$$= E[\exp(itT(\theta_0) - \tfrac{1}{2} u^2 T(\theta_0))]$$

This proves the sufficiency part.

SECTION 4.

The purpose of this section is to show that, under a condition which is slightly stronger than (A.2), the joint measurability of $W_n(\theta)$ and $A_n(h,\theta), n \geq 1$, can be removed in both the theorems 1 and 2 and further that the condition (A.1) can be removed in Theorem 1. Consider the following condition:

(A.2') There exist $\underline{\underline{A}}_n$ - measurable functions $W_n(\theta), A_n(h,\theta), n \geq 1$, $h \in R^k, \theta \in (\underline{H})$ such that the difference

$$\frac{dP_{\theta+\delta_n h_n, n}}{dP_{\theta,n}} - \exp[h_n' W_n(\theta) - A_n(h_n,\theta)]$$

converges to zero in $P_{\theta,n}$ probability for every bounded sequence $\{h_n\}$ of $R^k$ and that the difference

$$A_n(h_n,\theta) - A_n(h_n^*,\theta)$$

converges to zero for every sequence $\{h_n^*\}$ of $R^k$ satisfying $|h_n - h_n^*| \longrightarrow 0$, where the sequence $\{\delta_n\}$ of p.d. matrices is such that $\|\delta_n\| \longrightarrow 0$.

<u>Lemma 2</u>. Suppose that the conditions (A.2') and (A.3) are

satisfied. Then the quantity

$$\| P_{\theta+\delta_n h_n, n} - P_{\theta+\delta_n h_n^*, n} \| \longrightarrow 0 \qquad (4.6)$$

for every bounded sequences $\{h_n\}$ and $\{h_n^*\}$ of $R^k$ satisfying $|h_n - h_n^*| \longrightarrow 0$.

<u>Proof</u>. It is easy to see that the given conditions entails that the sequences $\{P_{\theta+\delta_n h_n, n}\}$, $\{P_{\theta+\delta_n h_n^*, n}\}$ and $\{P_{\theta,n}\}$ are contiguous, and hence we can assume without loss of generality that $P_{\theta+\delta_n h_n, n} \approx P_{\theta,n} \approx P_{\theta+\delta_n h_n^*, n}$ for every $\theta \varepsilon (\underline{H})$ and $n \geq 1$, where the symbol $\approx$ denotes the mutual absolute continuity.

Set, for every $h \varepsilon R^k$, $\theta \varepsilon (\underline{H})$ and $n \geq 1$,

$$Z_{\theta,n}(h) = \frac{dP_{\theta+\delta_n h, n}}{dP_{\theta,n}} .$$

Now note that the difference

$$Z_{\theta,n}(h_n) - Z_{\theta,n}(h_n^*) \qquad (4.7)$$

converges to zero in $P_{\theta,n}$ probability for every $\theta \varepsilon (\underline{H})$.

Next note that

$$\int_{\{|Z_{\theta,n}(h_n)| > \alpha\}} Z_{\theta,n}(h_n) dP_{\theta,n} = P_{\theta+\delta_n h_n, n}\left[ |Z_{\theta,n}(h_n)| > \alpha \right]$$

and hence, in view of contiguity,

$$\lim_{\alpha \longrightarrow \infty} \limsup_{n \longrightarrow \infty} \int_{\{|Z_{\theta,n}(h_n)| > \alpha\}} Z_{\theta,n}(h_n) dP_{\theta,n} = 0. \qquad (4.8)$$

Similarly

$$\lim_{\alpha \to \infty} \limsup_{n \to \infty} \int_{\{|Z_{\theta,n}(h_n^*)| > \alpha\}} Z_{\theta,n}(h_n^*)\, dP_{\theta,n} = 0 \qquad (4.9)$$

Now combining (4.7), (4.8) and (4.9) we have

$$\int |Z_{\theta,n}(h_n) - Z_{\theta,n}(h_n^*)|\, dP_{\theta,n} \longrightarrow 0 .$$

Hence the proof of the lemma is complete.

<u>Proposition</u>. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the conditions (A.2'), (A.3), (A.5) and (A.6) and the condition (A.4') of Theorem 2. Then there exist $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$ - measurable functions $W_n^*(\theta) : \underline{\underline{X}}_n \times (\underline{\underline{H}}) \longrightarrow R^k$ and $\underline{\underline{A}}_n \times \underline{\underline{B}}^k \times \underline{\underline{B}}^k$ - measurable functions $A_n^*(h,\theta) : \underline{\underline{X}}_n \times R^k \times (\underline{\underline{H}}) \longrightarrow R, n \geq 1$, such that

(i) the condition (A.2') is satisfied with the functions $W_n(\theta)$ and $A_n(h,\theta)$ are replaced by $W_n^*(\theta)$ and $A_n^*(h,\theta)$ respectively,

(ii) for every $h \in R^k$ and $\theta \in (\underline{\underline{H}})$, the differences

$$A_n^*(h,\theta) - [A_n(h,\theta) - \sum_{i=1}^{k} h_i A_n(\varepsilon_i,\theta)]$$

and

$$h' W_n^*(\theta) - [h' W_n(\theta) - \sum_{i=1}^{k} h_i A_n(\varepsilon_i,\theta)]$$

converges to zero in $P_{\theta,n}$ - probability where $\{\varepsilon_i ; i = 1,2,\ldots,k\}$ is a basis of $R^k$ and $h_i$'s are such that $h = \sum_{i=1}^{k} h_i \varepsilon_i$,

(iii) for every $\varepsilon > 0$, $h \varepsilon R^k$ and $g \varepsilon L_1(\mu^k)$,

$$\int_{(\underline{H})} \int_{\underline{\underline{X}}_n} I[\,|A_n^*(s,\theta+\delta_n h) - A_n^*(s,\theta)| > \varepsilon] \, dP_{\theta+\delta_n h,n} \, g(\theta)\, d\theta \longrightarrow 0 \qquad (4.10)$$

and (iv) the condition (A.5) is satisfied for the corresponding limit of the sequence $\{A_n^*(h,\theta)\}$.

Remark. Note that the above condition (4.10) is weaker than the condition (A.6), but what we have really used in the proof of Theorem 1 is the above condition (4.10). It is possible to show, under the condition (A.4) which is stronger than the condition (A.4') of Theorem 2, that there exists a lebesgue null set $N \subseteq (\underline{H})$ such that the condition (A.6) is satisfied for the sequence $\{A_n^*(h,\theta)\}$ whenever $\theta \varepsilon (\underline{H}) - N$; the proof of this statement will not be presented here though the arguments of the proof seem to be somewhat non-trivial.

Proof. Let $\underline{\underline{E}}_n^* = \{\underline{\underline{X}}_n, \underline{\underline{A}}_n, P_{\theta,n}^*; \theta \varepsilon (\underline{H})\}$ be a sequence of experiments satisfying the statements (i) - (iv) of Lemma 3 of Ch.1; a construction of such a sequence exists since, in view of Lemma 2, the condition(*) of Lemma 3 of Ch.1 is satisfied. Let $\{\varepsilon_i ; i = 1,2,\ldots,k\}$ be a basis of $R^k$. Define $W_n^*(\theta)$ by

$$h' W_n^*(\theta) = \sum_{i=1}^{k} h_i \wedge_n^*(\theta+\delta_n \varepsilon_i, \theta)$$

where $h_i$'s are such that $h = \sum_{i=1}^{k} h_i \varepsilon_i$, $h \varepsilon R^k$, and

$\Lambda_n^*(\theta+\delta_n h,\theta) = \log \dfrac{dP^*_{\theta+\delta_n h,\theta}}{dP^*_{\theta,n}}$. Now define $A_n^*(h,\theta)$ by

$$A_n^*(h,\theta) = h'W_n^*(\theta) - \Lambda_n^*(\theta+\delta_n h,\theta) .$$

It is clear from the statements (ii) and (iv) of Lemma 3 of Ch.1, that the condition (A.2') is satisfied with the functions $W_n(\theta)$ and $A_n(h,\theta)$ are replaced by $W_n^*(\theta)$ and $A_n^*(h,\theta)$ respectively and that the functions $W_n^*(\theta)$ and $A_n^*(h,\theta)$ are jointly measurable. Now note that

$$\begin{aligned} A_n^*(h,\theta) &= h'W_n^*(\theta) - \Lambda_n^*(\theta+\delta_n h,\theta) \\ &= \sum_{i=1}^{k} h_i \Lambda_n^*(\theta+\delta_n \varepsilon_i,\theta) - \Lambda_n^*(\theta+\delta_n h,\theta) \end{aligned}$$

and, in view of the statement (iv) of Lemma 3 of Ch.1, this can be approximated with $P_{\theta,n}$- probability tending to one by $A_n(h,\theta) - \sum_{i=1}^{k} h_i A_n(\varepsilon_i,\theta)$. Hence the statements (ii) and (iv) follows. Now note that, since for every $s\varepsilon R^k$ and $\theta\varepsilon(\underline{H})$

$$\Lambda_n^*(\theta+\delta_n s,\theta) - \exp[s'W_n(\theta) - A_n(s,\theta)]$$

converges to zero in $P_{\theta,n}$- probability, the invariance theorem 1 of Ch.3 and the condition (A.6) implies that, for every $\varepsilon > 0$, $h,s\varepsilon R^k$ and $g\varepsilon L_1(\mu^k)$

$$\int_{(\underline{H})}\int_{\underline{X}_n} I[\,|\quad{}^*_n(\theta+\delta_n(s+h),\theta+\delta_n h) - [s'W_n(\theta+\delta_n h) - A_n(s,\theta)]\,| > \varepsilon] \, dP_{\theta+\delta_n h,n}\, g(\theta)\, d\theta \longrightarrow 0 \qquad (4.11)$$

Now, writing $s = \sum_{i=1}^{k} s_i \varepsilon_i$,

$$A_n^*(s,\theta+\delta_n h) = \sum_{i=1}^{k} s_i \wedge_n^*(\theta+\delta_n(\varepsilon_i+h),\theta+\delta_n h)$$

$$- \wedge_n^*(\theta+\delta_n(s+h),\theta+\delta_n h)$$

and hence the statement (iii) follows by using (4.11) and the statement (ii). This completes the proof of the proposition.

CHAPTER 5

# CONDITIONAL DECOMPOSITION OF THE LIMIT DISTRIBUTION AND SOME APPLICATIONS

## 1. INTRODUCTION

When the sequence of experiments satisfies the LAN condition, Hájek (1970) has established a basic result that the limit distribution, when it is invariant in the limit in some sense, of a sequence of estimators can be decomposed as a convolution for all points of the parameter space. Independently, this result was also essentially obtained by Inagaki (1970) under restrictive assumptions. LeCam (1972) has extended Hájek's convolution result to a much more general sequence of experiments than that of LAN experiments. Furthermore, LeCam (1973), while discussing certain results concerning the possible invariance of the limits of experiments, pointed out that Hájek's convolution result can be obtained, without assuming the above mentioned invariance restriction, for almost all points of the parameter space ; it is important to note here that the usual examples (see, e.g. LeCam (1953)) show that the invariance restriction cannot be relaxed if one tries to establish such a convolution result for all points of the parameter space.

In this chapter we first show that, when the sequence of experiments satisfies the LAMN condition, a conditional convolution result holds for any convergent subsequence of estimators under a certain type of invariance restriction; this result extends and strengthens the convolution result of Hájek (1970). Secondl we show that, without assuming the invariance restriction, the limit distribution of any convergent subsequence of estimators can be conditionally decomposed as a convolution for almost all points of the parameter space.

The ideas of our proof are based on Bickel's simple short proof of Hájek's convolution theorem (see Roussas (1972) for the published version of Bickel's proof.)

We would like to mention that it is possible to deduce Hájek's convolution result, for the LAN case, without the invariance restriction, for almost all points of the parameter space from Corollary 2 of Strasser (1978). However, Strasser requires, in addition to the LAN condition, a strong restriction concerning the existence of ACS estimators. (This restriction can be easily removed if one uses Lemma 5 of Ch.1 in the arguments of his proof). Furthermore Strasser (1978) uses the invariance theorem 3 of Ch.3 to get the conclusion of his result, but it appears that one has to use the stronger result, Theorem 2 of Ch.3.

Results are stated in Section 2, and the proofs of the results are presented in Section 3. By applying our conditional convolution results we deduce several results concerning the asymptotic lower bounds for risk functions; one of these results clarifies some of the statements made earlier by Heyde (1978). This chapter is a revised version of Jeganathan (1979c and Sec.3 of 1979a).

## 2. STATEMENTS OF THE MAIN RESULTS

In addition to the notations of Ch.1, $\mu^k|\mathbb{B}^k$ denotes the Lebesgue measure and $C_{oo}(R^q), q \geq 1$, denotes the set of all continuous functions vanishing outside compacts.

Let $\{r\} \subseteq \{n\}$ be a subsequence and $H_{\theta_o}$ be a (substochastic) measure such that

$$\mathcal{L}(T_r(\theta_o), \delta_r^{-1}(V_r-\theta_o) | P_{\theta_o,r}) \Longrightarrow H_{\theta_o} .$$

Let $\mathcal{L}_{\theta_o}$ be the law of $T(\theta_o)$ and let $\bar{R}^k$ be one-point compactification of $R^k$. Define

$$\bar{H}_{\theta_o}(B \times \{\infty\}) = \mathcal{L}_{\theta_o}(B) - H_{\theta_o}(B \times R^k)$$

and

$$\bar{H}_{\theta_o}(B \times A) = H_{\theta_o}(B \times A)$$

for every Borel sets $B \subseteq R^{k^2}$ and $A \subseteq R^k$. Then $\bar{H}_{\theta_o}$ is a probability measure defined on $R^{k^2} \times \bar{R}^k$ induced by $H_{\theta_o}$. Let $\bar{\mathcal{L}}_{T(\theta_o)}$ be a regular conditional probability measure (on $\bar{R}^k$) such that

$$\bar{H}_{\theta_o}(C) = \int I(C)\, \bar{\mathcal{L}}_t(dx)\, \mathcal{L}_{\theta_o}(dt)$$

for every Borel set $C \subseteq R^{k^2} \times \bar{R}^k$.

Theoren 1. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the LAMN-condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$. Let $\{V_n\}$ be a sequence of estimators such that the difference

$$\mathcal{L}[f(T_n(\theta_o), \delta_n^{-1}(V_n - \theta_o - \delta_n h)) | P_{\theta_o + \delta_n h, n}]$$
$$- \mathcal{L}[f(T_n(\theta_o), \delta_n^{-1}(V_n - \theta_o)) | P_{\theta_o, n}]$$

converges to zero for every $h \varepsilon R^k$ and $f \varepsilon C_{oo}(R^{k^2+k})$.

Let the regular conditional probability measure $\bar{\mathcal{L}}_{T(\theta_o)}$ be as above and let $\mathcal{L}_{T(\theta_o)}$ be the restriction of $\bar{\mathcal{L}}_{T(\theta_o)}$ to $R^k$. Then there exists a (sub-strochustic) kernel $K_{T(\theta_o)}$ such that

$$\mathcal{L}_{T(\theta_o)} = K_{T(\theta_o)} * N(0, T^{-1}(\theta_o)).$$

The following familier version, in which the existonce of the limit distribution is assumed, is immediate from the above theoren 1.

Corollary 1. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o$. Let $\{V_n\}$ be a sequence of estimators such that, for every $h \varepsilon R^k$,

$$\mathcal{L}(T_n(\theta_o), \delta_n^{-1}(V_n - \theta_o - \delta_n h) | P_{\theta_o} + \delta_n h, n) \Longrightarrow \mathcal{L}(T(\theta_o), V(\theta_o))$$

for some random k-vector $V(\theta_o)$. Let $\mathcal{L}_{T(\theta_o)}$ be a regular conditional probability measure of $V(\theta_o)$ given $T(\theta_o)$. Then there exists a stochastic kernel $K_{T(\theta_o)}$ such that

$$\mathcal{L}_{T(\theta_o)} = K_{T(\theta_o)} * N(0, T^{-1}(\theta_o)).$$

Now assume that the assumptions of Theorem 2 below are satisfied and hence we can further assume, without loss of generality, that the functions $V_n(\theta), T_n(\theta)$ are jointly measurable (see Lemmas 4 and 5 of Chapter 1).

Let $\{r\} \subseteq \{n\}$ be a subsequence and $H_\theta$ be a (sub-stochastic) kernel such that the sequence $\mathcal{L}(T_n(\theta), \delta_n^{-1}(V_n-\theta) | P_{\theta,n})$ is $C_{oo}(R^{k^2+k}) \otimes L_1(\mu^k)$ convergent to $H_\theta$. Define, as above, the stochastic kernel $\overline{H}_\theta$ on $R^{k^2} \times \overline{R}^k$ induced by $H_\theta$. Let $\overline{\mathcal{L}}_{T(\theta)}$ be a stochastic kernel (on $\overline{R}^k$) such that

$$\overline{H}_\theta(C) = \int I(C)\, \overline{\mathcal{L}}_t(dx)\, \mathcal{L}_\theta(dt)$$

for every $\theta \in (\underline{H})$ and for every Borel set $C \subseteq R^{k^2} \times \overline{R}^k$.

<u>Theorem 2</u>. Suppose that the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN-condition for $\mu^k$-almost all $\theta \in (\underline{H})$. Further assume that the functions $\theta \longrightarrow P_{\theta,n}(A), A \in \underline{\underline{A}}_n, n \geq 1$ are $\underline{\underline{B}}^k$-measurable and that the random functions $\Lambda_n(\theta+\delta_n h, \theta), h \in R^k$, are $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$-measurable. Let the kernel $\overline{\mathcal{L}}_{T(\theta)}$ be as above and let $\mathcal{L}_{T(\theta)}$ be the restriction of $\overline{\mathcal{L}}_{T(\theta)}$ to $R^k$. Then there exists a Lebesgue null set $N \subseteq (\underline{H})$ and a (sub-stochastic) kernel $K_{T(\theta)}$ such that

$$\mathcal{L}_{T(\theta)} = K_{T(\theta)} * N(0, T^{-1}(\theta))$$

for every $\theta \in (\underline{H}) - N$.

<u>Corollary 2</u>. Suppose that the assumptions of the above Theorem 2 are satisfied. Let $\{V_n\}$ be a sequence of estimators such that for every $\theta \in (\underline{H})$.

$$\mathcal{L}(T_n(\theta), \delta_n^{-1}(V_n-\theta) \mid P_{\theta,n}) \Longrightarrow \mathcal{L}(T(\theta), V(\theta))$$

for some random k-vector $V(\theta)$. For every $\theta \varepsilon (\underline{H})$, let $\mathcal{L}_{T(\theta)}$ be a regular conditional probability measure of $V(\theta)$ given $T(\theta)$. Then there exists a Lebesgue null set $N \subseteq (\underline{H})$ and a stochastic kernel $K_{T(\theta)}$ such that

$$\mathcal{L}_{T(\theta)} = K_{T(\theta)} * N(0, T(\theta))$$

for every $\theta \varepsilon (\underline{H}) - N$.

## 3. PROOFS OF THE MAIN RESULTS

Before going into the details of the proof of Theorem 2 let us observe that when the measurability condition of Theorem 2 is satisfied, the random functions $T_n^*(\theta), n \geq 1$, constructed, under the LAMN condition, in Lemma 4 will be $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$-measurable.

Proof of Theorem 2. Let $\{r\} \subseteq \{n\}$ be a subsequence and $H_\theta$ be a kernel such that the sequence $\{\mathcal{L}(T_r(\theta), \delta_r^{-1}(V_r-\theta) \mid P_{\theta,r})\}$ is $C_{oo}(R^{k^2+k}) \otimes L_1(\mu^k)$ convergent to $H_\theta$. Then according to the invariance theorem 1 of Ch.3 and the statements (i) and (ii) of of Lemma 4/Ch.1 together with the contiguity condition, the sequence $\{\mathcal{L}(T_r(\theta), \delta_r^{-1}(V_r-\theta-\delta_r h) \mid P_{\theta+\delta_r h, r})\}$ is also $C_{oo}(R^{k^2+k}) \otimes L_1(\mu^k)$ convergent to the same kernel $H_\theta$. For simplicity assume that $(\underline{H}) = R$. Let $f(u,z,x,y) = (e^{iux}-1)(e^{izy}-1)/ixiy; u,z,x,y \varepsilon R$. Note that $f(u,z,x,y) \longrightarrow 0$ as $|(x,y)| \longrightarrow \infty$ for every $u, z \varepsilon R$. Hence we have, setting $\theta_n = \theta + \delta_n h$,

$$\int_R \int_{\underline{\underline{X}}_r} f(u,z,T_r(\theta),\delta_r^{-1}(V_r-\theta_r))\,dP_{\theta_r,r}\,g(\theta)\,d\theta$$

$$\longrightarrow \int_R \int_{R^2} f(u,z,t,v)H_\theta(dt,dv)g(\theta)d\theta \qquad (5.1)$$

for every $g(\theta) \varepsilon L_1(\mu)$ and $u,z\varepsilon R$.

Now, there exists a subsequence $\{m\} \subseteq \{r\}$ and a kernel $Q_\theta$ such that the sequence $\{\mathcal{L}(T_m(\theta),W_m(\theta),\delta_m^{-1}(V_m-\theta)|P_{\theta,m})\}$ is $C_{oo}(R^3) \otimes L_1(\mu)$ convergent to $Q_\theta$. Without loss of generality assume that $\{m\} = \{r\}$. In view of contiguity, we can further assume, without loss of generality, that $P_{\theta_n,n} << P_{\theta,n}$ for every $n \geq 1, h \varepsilon R$ and $\mu$-almost all $\theta \varepsilon (\underline{H})$. Hence the l.h.s. of (5.1) can be written as

$$\int_R \int_{\underline{\underline{X}}_n} f(u,z,T_r(\theta),\delta_r^{-1}(V_r-\theta_r)) \frac{dP_{\theta_r,r}}{dP_{\theta,r}}\,dP_{\theta,r}\,g(\theta)\,d\theta$$

and it is not difficult to see from the LAMN condition that this converges to

$$\int_R \int_{R^3} f(u,z,t,(v-h))\exp(ht^{1/2}w-\tfrac{h^2}{2}t)Q_\theta(dt,dw,dv)g(\theta)d\theta$$

for every $g(\theta) \varepsilon L_1(\mu)$ and $u,z,h \varepsilon R$. Hence we see that for every $(u,z,h) \varepsilon R^3$ there exists a Lebesgue null set $N(u,z,h)$, possibly depending on $(u,z,h)$, such that

$$\int_{R^2} f(u,z,t,v)H_\theta(dt,dv) = \int_{R^3} f(u,z,t,(v-h))\exp(ht^{1/2}w-\tfrac{h^2}{2}t)Q_\theta(dt,dw,dv) \qquad (5.2)$$

for every $\theta \varepsilon (\underline{H}) - N(u,z,h)$.

Define

$$\overline{Q}_\theta(A \times B \times \{\infty\}) = F_\theta(A \times B) - Q_\theta(A \times B \times R^k)$$

and

$$\overline{Q}_\theta(A \times B \times C) = Q_\theta(A \times B \times C)$$

for every Borel sets $A \subseteq R^{k^2}$, $B \subseteq R^k$ and $C \subseteq \overline{R}^k$, where $F_\theta$ is the law of $(T(\theta), W)$. Then $\overline{Q}_\theta$ is a probability measure on $R^{k^2+k} \times \overline{R}^k$ induced by $Q_\theta$. Let the probability measure $\overline{H}_\theta$ be as defined in Section 2. Note that $\overline{H}_\theta$ and $F_\theta$ are marginals of $\overline{Q}_\theta$ for $\mu$-almost all $\theta \varepsilon (\underline{H})$; let this exceptional set be $N'$. Now the equality (5.2) can be written as

$$\int_{R^2} f(u,z,t,v)\overline{H}_\theta(dt,dv) = \int_{R^3} f(u,z,t,(v-h))\exp(ht^{1/2}w - \frac{h^2}{2}t)\ \overline{Q}_\theta(dt,dw,dv) \tag{5.3}$$

for every $\theta \varepsilon (\underline{H}) - N(u,z,h)$.

Let $D$ be the set of all rationals in $R^3$ and let $N'' = \bigcup_{(u,z,h) \varepsilon D} N(u,z,h)$. Let $N = N' \cup N''$. Then whenever $\theta \varepsilon (\underline{H}) - N$ the equality (5.3) holds for every $(u,z,h) \varepsilon D$. Now for every $(u,z,h) \varepsilon R^3$ there exists a sequence $(u_m, z_m, h_m) \varepsilon D$ such that $(u_m, z_m, h_m) \longrightarrow (u,z,h)$ as $m \longrightarrow \infty$. Clearly

$$\int_{R^2} f(u_m,z_m,t,v)\overline{H}_\theta(dt,dv) \longrightarrow \int_{R^2} f(u,z,t,v)\overline{H}_\theta(dt,dw) .$$

We now show that

$$\int_{R^3} f(u_m,z_m,t,(v-h_m))\exp(h_m t^{1/2} w - \frac{h_m^2}{2} t)\overline{Q}_\theta(dt,dw,dv)$$

converges for every $\theta \in (\bar{\underline{H}}) - N$ to

$$\int_{R^3} f(u,z,t,(v-h)\exp(ht^{1/2}w - \frac{h^2}{2}t)\bar{Q}_\theta(dt,dw,dv) .$$

Since

$$f(u_m,z_m,t,(v-h_m))\exp(h_m t^{1/2}w - \frac{h_m^2}{2}t)|$$

$$\leq C\exp(h_m t^{1/2}w - \frac{h_m^2}{2}t) \text{ (for some } C > 0)$$

it is enough to show that, for every $\theta \in (\bar{\underline{H}}) - N$, $\exp(h_m t^{1/2}w - \frac{h_m^2}{2}t)$ converges in the first mean to $\exp(ht^{1/2}x - \frac{h^2}{2}t)$ as $m \longrightarrow \infty$. This easily follows via Scheffe's theorem and from the facts that

$$\exp(h_m t^{1/2}w - \frac{h_m^2}{2}t) \longrightarrow \exp(ht^{1/2}w - \frac{h^2}{2}t)$$

as $m \longrightarrow \infty$ and

$$\int_{R^2} \exp(h_m t^{1/2}w - \frac{h_m^2}{2}t)\bar{Q}_\theta(dt,dw,\bar{R}) = 1$$

$$= \int_{R^2} \exp(ht^{1/2}w - \frac{h^2}{2}t)\bar{Q}_\theta(dt,dw,\bar{R})$$

for every $\theta \in (\bar{\underline{H}}) - N$.

Thus we see from (5.3) that whenever $\theta \in (\bar{\underline{H}}) - N$

$$\int_{R^2} f(u,z,t,v)\bar{H}_\theta(dt,dv) = \lim_{m \longrightarrow \infty} \int_{R^2} f(u_m,z_m,t,v)\bar{H}_\theta(dt,dv)$$

$$= \lim_{m \longrightarrow \infty} \int_{R^3} f(u_m,z_m,t,(v-h_m))\exp(h_m t^{1/2}w - \frac{h_m^2}{2}t)\bar{Q}_\theta(dt,dw,dv)$$

$$= \int_{R^3} f(u,z,t,(v-h))\exp(ht^{1/2}w - \frac{h^2}{2}t)\bar{Q}_\theta(dt,dw,dv)$$

for every $(u,z,h) \in R^3$. Now this implies (cf. Loéve (1963), p. 189) that whenever $\theta \in (\underline{H}) - N$

$$\int_{R^2} \exp(iut+izv)\bar{H}_\theta(dt,dv)$$
$$= \int_{R^3} \exp(iut+iz(v-h))\exp(ht^{1/2}w - \frac{h^2}{2}t)\bar{Q}_\theta(dt,dw,dv) \tag{5.4}$$

for every $(u,z,h) \in R^3$. Let $\bar{\mathscr{L}}_{T(\theta)}$ be the regular conditional probability measure as defined in Section 2. Then (5.4) can be written as, for every $\theta \in (\underline{H}) - N$ and $(u,z,h) \in R^3$,

$$\int_R \text{ex}(iut) [\int_R \exp(izv) \bar{\mathscr{L}}_t(dv)] \mathscr{L}_\theta(dt)$$
$$= \int_R \exp(iut) [\int_{R^2} \exp(iz(v-h))\exp(ht^{1/2}w - \frac{h^2}{2}t) \bar{\mathscr{L}}'_t(dw,dv)] \mathscr{L}_\theta(dt) \tag{5.5}$$

for some regular conditional probability measure $\bar{\mathscr{L}}'_{T(\theta)}$, where $\mathscr{L}_\theta$ is the law of $T(\theta)$. Now a simple continuity argument shows that (5.5) entails

$$\int_R \exp(izv)\bar{\mathscr{L}}_{T(\theta)}(dv)$$
$$= \int_{R^2} \exp(iz(v-h))\exp(hT^{1/2}(\theta)w - \frac{h^2}{2}T(\theta)) \bar{\mathscr{L}}'_{T(\theta)}(dw,dv) . \tag{5.6}$$

for every $\theta \in (\underline{H}) - N$ and $u, h \in R$. In what follows assume that $\theta \in (\underline{H}) - N$ and $T(\theta)$ are fixed. It can be shown that the r.h.s. of (5.6) is analytic in $h$. Hence we have

$$\int_R \exp(izv)\,\bar{\mathcal{L}}_{T(\theta)}(dv)$$
$$= \int_{R^2} \exp(izv+zh)\exp(ihT^{1/2}(\theta)w+\frac{h^2}{2}T(\theta))\,\bar{\mathcal{L}}'_{T(\theta)}(dw,dv)$$

for every $z,h \in R$. Setting $h=-T^{-1}(\theta)z$ in this equality we have

$$\int_R \exp(izv)\,\bar{\mathcal{L}}_{T(\theta)}(dv)$$
$$= \exp(-\tfrac{1}{2}T^{-1}(\theta)z^2)\int_{R^2}\exp[iz(v-T^{-1/2}(\theta)w)]\,\bar{\mathcal{L}}'_{T(\theta)}(dw,dv)$$

for every $z$. This proves the result.

Proof of Theorem 1. The proof is essentially contained in the above proof of Theorem 2.

## 4. SOME APPLICATIONS

Let $L$ be the class of all loss functions $\ell: R^k \longrightarrow R$ of the form $\ell(0)=0$, $\ell(y)=\ell(|y|)$ and $\ell(y) \leq \ell(z)$ if $|y| \leq |z|$. Let $\lambda|\underline{\underline{B}}^k$ be a $\sigma$-finite measure such that $\lambda \ll \mu^k$. Let $L_1^+(\lambda)$ be the class of all positive integrable functions in $(R^k,\underline{\underline{B}}^k,\lambda)$. The distribution of $T^{-1/2}(\theta)W$ will be denoted by $\phi_\theta$.

Proposition 1. Assume that the assumptions of Theorem 2 are satisfied. Let $\{V_n\}$ be a sequence of estimators. Then

$$\liminf_{n\to\infty} \int_{(\underline{\underline{H}})}\int_{\underline{\underline{X}}_n} \ell(\delta_n^{-1}(V_n-\theta))\,dP_{\theta,n}\,b(\theta)\lambda(d\theta)$$
$$\geq \int_{(\underline{\underline{H}})}\int_{R^k} \ell(x)b(\theta)\phi_\theta(dx)\lambda(d\theta)$$

for every $b \in L_1^+(\lambda)$ and $\ell \in L$, provided the r.h.s. of the above inequality is finite.

Proof. The proof is an easy consequence of Theorem 2.

Proposition 2. Assume that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the LAMN condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$. Let $\{V_n\}$ be a sequence of estimators satisfying the invariance condition of Theorem 1. Then

$$\liminf_{n \longrightarrow \infty} E_{\theta_o} [\ell(\delta_n^{-1}(V_n - \theta_o))] \geq E[\ell(T^{-1/2}(\theta_o)W)]$$

for every $\ell \in L$, provided the r.h.s. of this inequality is finite.

Proof. The proof is an easy consequence of Theorem 1.

The proof of the statement (i) of the following proposition 3 is immediate from Corollary 1; the proof of the statement (ii) is also a consequence of Corollary 1, and the proof of which is quite similar to the proof of the corresponding result under the LAN-condition, see, e.g. Roussas (1972, pp. 141-147).

Proposition 3. Assume that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$. Let $\{V_n\}$ be a sequence of estimators satisfying the invariance condition of Corollary 1. Then

(i) $E[\ell(V(\theta_o))] \geq E[\ell(T^{-1/2}(\theta_o)W)]$

provided the r.h.s. of this inequality is finite, and

(ii) for every $q \in R^k$ and $t_1, t_2 > 0$,

$$P[-t_1 < q' V(\theta_o) < t_2] \leq P[-t_1 < q' T^{-1/2}(\theta_o) W < t_2]$$

provided, for every $q \in R^k$,

$$P[q' V(\theta_o) \geq 0 | T(\theta_o)] \geq \tfrac{1}{2} \text{ and } P[q' V(\theta_o) \leq 0 | T(\theta_o)] \geq \tfrac{1}{2}.$$

Remark 1. A result ~~similar to Proposition 2~~ was earlier obtained under some specific assumptions and with special reference to maximum likelihood estimators by Heyde (1978). See also Basawa and Scott (1979).

Note that in Proposition 1 we have imposed the invariance restriction on the sequence $\{T_n(\theta_o), \delta_n^{-1}(V_n - \theta_o)\}$; it is enough to impose the restriction on the sequence $\{\delta_n^{-1}(V_n - \theta_o)\}$, see Corollary 1 of the next chapter (Ch.6).

Remark 2. Proposition 1 occur explicitely in the form given here in Strasser (1978) for the LAN case; this result seems to be implicit for the LAN case in, for example, LeCam (1973) also since Theorem 2 was essentially mentioned in this paper for the LAN case.

Remark 3. Note that the familiar result (LeCam (1953) and Bahadur (1964)) concerning Fisher's bound for asymptotic variances for almost all points of the parameter space can be easily deduced from Corollary 2. A more general result of Pfanzagl (1970, Theorem 2) can also be deduced from the Corollary 2 when it is specialised to the LAN case; an analogous result for the LAMN case can also be deduced under appropriate conditions. Extension of Theorem 1 of Pfanzagl (1970) to the LAMN case can be found in Bhat and Prasad (1978).

CHAPTER 6

# SOME ASYMPTOTIC PROPERTIES OF RISK FUNCTIONS

## 1. INTRODUCTION

In a fundamental paper LeCam (1953) has obtained some basic results concerning the asymptotic properties of risk functions. Extending and improving the results of LeCam, Hájek (1972) obtained when $\dim(\underline{H}) = 1$, certain asymptotic lower bound (local asymptotic minimax result) for a certain class of risk functions of estimators under the LAN condition; under the same LAN condition Strasser (1978) has recently obtained certain global asymptotic properties of risk functions of estimators. Furthermore, Hájek (1972) and Strasser (1978) characterise those estimators which attain these lower bounds. In two important papers LeCam (1972, 1979b) obtained certain extremely general results, in a general framework, concerning local asymptotic minimaxity and admissibility and showed that the aboved mentioned results of Hájek may be viewed as special cases of these general results; another important feature of LeCam (1972) is that the results were stated in an approximation framework, i.e., in terms of certain distances. The main purpose of this chapter is to obtain analogous lower bounds, to characterise those estimators which attain these lower bounds and to present some further results concerning the asymptotic properties of risk functions and posterior approximation for the more general LAMN case.

More specifically, in Theorem 1 we present a general result concerning the asymptotic lower bound for risk functions; a more familiar result, Corollary 1, follows from this result under the usual invariance restriction. Sequences of estimators which attain the lower bound of Theorem 1 are characterised in Theorem 2. An important feature of these two results is that they do not depend on the dimensionality restriction of the parameter space.

In Theorem 3 we present a result, for the LAMN case, which is an extension of the local asymptotic minimax results of Hájek (1972) and LeCam (1972 and 1974b); this result turns out to be an immediate consequence of Theorem 1. In Theorem 4 we present, when $\dim(\underline{H}) = 1$, an extension of the uniqueness result of Hájek (1972); it may be noted here that this unqueness result does not hold when $\dim(\underline{H}) > 2$ for the reasons explained in LeCam (1972). It may be further noted here that the uniqueness result of Theorem 2 is fairly weaker than the uniqueness result of Theorem 4, as is easily seen by considering James - Stein type estimators.

In Theorems 5 and 6 we present analogous global asymptotic properties of risk functions; actually, we deduce these results from a general result (Proposition 1) concerning a certain kind of posterior approximation.

In connection with Theorems 3 and 4 of this chapter, the following remarks should be made. As we have already remarked, LeCam (1972, 1974b) has obtained some very deep and general results which amount to say the following. It one is interested in proving asymptotic properties such as local asymptotic minimaxity and admissibility of experiments, it is just enough to prove the statements for the limit of the experiments and then the corresponding limiting statements for the sequence of experiments can be concluded from his results. Thus once we have proved minimax and admissibility results for the limit of the LAMN experiment, the conclusions of Theorems 3 and 4 are the consequences of LeCam's results, since LeCam's results are not restricted to any particular form of the limit of the experiments. In the present chapter, as we have remarked, the proof of Theorem 3 turns out to be an immediate consequence of Theorem 1. The main reason for presenting a rather complete proof of Theorem 4 is the following. Once the powerful Lemma 6 of Ch.1 is given, it turns out that the proof of the unique admissibility result for the limit of the LAMN experiments and the proof of the local asymptotic admissibility result for the LAMN experiments are almost identical.

In connection with Theorem 6, we would like to remark that Strasser (1978), in proving this result for the LAN case, has assumed a strong restriction concerning the existence of ACS estimators. This restriction is removed in the present chapter.

This chapter is a revised version of Jeganathan (1980); a referee of this paper has remarked that some of the results of this paper might be already known to some of the workers working in this field, e.g. R.B. Davies. During the final stage of preparation of the present work we received a copy of Ph.D. thesis from Swensen (1980, September), where he has independently obtained our Theorems 3 and 4. His proof consist of first proving minimax and admissibility results for the limit of the LAMN experiment and then using the above mentioned results of LeCam to get the desired conclusion, whereas our proofs are based directly on Lemma 6 of Ch.1.

In section 2 we present the results and in Section 3 we present some preliminary lemmas. In Section 4 we present the proofs of the results. It may be noted that Theorem 5 of this chapter already occurs in Ch.5. See the remarks at the end of Ch.5.

## 2. STATEMENTS OF THE RESULTS

In addition to the notations of Ch.1, $\mu^k | \underline{B}^k$ denotes the Lebesgue measure.

Let L be the class of all loss functions $\ell : R^k \longrightarrow [0,1]$ of the form $\ell(0)$, $\ell(y) = \ell(|y|)$ and $\ell(y) = \ell(|y|)$ if $|y| \leq |z|$. We point out that for the purpose of simplicity only we consider bounded loss functions, and the results can be shown to hold true for a certain more general class of unbounded loss functions.

In what follows $E_{\theta+\delta_n h}$ denotes the expectation with respect to the measure $P_{\theta+\delta_n h,n}$, $n \geq 1, h \in R^k$. We set $D_\alpha = \{h \in R^k : |h| \leq \alpha\}$, $\alpha > 0$.

<u>Theorem 1</u>. Assume that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the LAMN condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$. Further assume that the functions $h \longrightarrow P_{\theta_o+\delta_n h,n}(A), A \in \underline{\underline{A}}_n, n \geq 1$, are Borel measurable. Then for every sequence $\{V_n\}$ of estimators and for every $\ell \in L$

$$\lim_{\alpha \longrightarrow \infty} \liminf_{n \longrightarrow \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} E_{\theta_o+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta_o-\delta_n h))]\, dh$$

$$\geq E[\ell(T^{-1/2}(\theta_o)W)] . \qquad (6.1)$$

<u>Corollary 1</u>. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o \varepsilon (\underline{\underline{H}})$. Let $\{V_n\}$ be a sequence of estimators such that the difference

$$\mathcal{L}(f(\delta_n^{-1}(V_n-\theta_o-\delta_n h))|P_{\theta_o+\delta_n h,n}) - \mathcal{L}(f(\delta_n^{-1}(V_n-\theta_o))|P_{\theta_o,n})$$

converges to zero for every $h \varepsilon R^k$ and for every continuous functions $f : R^k \longrightarrow R$ vanishing outside compacts. Then

$$\liminf_{n \longrightarrow \infty} E_{\theta_o}[\ell(\delta_n^{-1}(V_n-\theta_o))] \geq E[\ell(T^{-1/2}(\theta_o)W)] .$$

<u>Theorem 2</u>. Suppose that the assumptions of Theorem 1 are satisfied. Further suppose that for a sequence $\{V_n\}$ of estimators and for a non-constant $\ell \in L$

$$\lim_{\alpha \to \infty} \lim_{n \to \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} E_{\theta_o+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta_o-\delta_n h))]\, dh$$

$$= E[\ell(T^{-1/2}(\theta_o)W)]. \quad (6.2)$$

Then for every $\varepsilon > 0$

$$\lim_{\alpha \to \infty} \lim_{n \to \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} P_{\theta_o+\delta_n h, n}[|\delta_n^{-1}(V_n-\theta_o) - T_n^{-1/2}(\theta_o)W_n(\theta_o)|$$

$$> \varepsilon]\, dh = 0. \quad (6.3)$$

Theorem 3. Suppose that the sequence $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o \varepsilon (\underline{H})$. Then for every sequence $\{V_n\}$ of estimators and for every $\ell \varepsilon L$

$$\lim_{\alpha \to \infty} \liminf_{n \to \infty} \sup_{|h| \leq \alpha} E_{\theta_o+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta_o-\delta_n h))]$$

$$\geq E[\ell(T^{-1/2}(\theta_o)] .$$

Theorem 4. Suppose that $\dim(\underline{H}) = 1$ and that the assumption of Theorem 3 is satisfied. Further assume that $E(T^{-1/2}(\theta_o)) < \infty$. Let $\{V_n\}$ be a sequence of estimators such that for a non-constant $\ell \varepsilon L$ and for every $h \varepsilon R^k$

$$\limsup_{n \to \infty} E_{\theta_o+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta_o-\delta_n h))] \leq E[\ell(T^{-1/2}(\theta_o)W)]. \quad (6.5)$$

Then the difference

$$\delta_n^{-1}(V_n-\theta_o) - T_n^{-1/2}(\theta_o)W_n(\theta_o)$$

converges to zero in $P_{\theta_o,n}$ - probability.

Proposition 1. Let $\lambda | \underline{\underline{B}}^k$ be a measure such that $\lambda << \mu^k$ and $\lambda((\underline{\underline{H}})) < \infty$. Assume that the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfies the LAMN condition for $\mu^k$-almost all $\theta \varepsilon (\underline{\underline{H}})$. Further assume that the functions $(\theta, h) \rightarrow P_{\theta+\delta_n h, n}(A), A \varepsilon \underline{\underline{A}}_n$, $n \geq 1$, are $\underline{\underline{B}}^k$-measurable and that the functions $\wedge_n(\theta+\delta_n h, \theta), h \varepsilon R^k, n \geq 1$, and $W_n(\theta), T_n(\theta)$, $n \geq 1$, are $\underline{\underline{A}} \times B^k$-measurable. Set

$$S_n^*(\theta,h) = \frac{|\det T_n(\theta)|^{1/2}}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2}(h - T_n^{-1/2}(\theta) W_n(\theta))' T_n(\theta)(h - T_n^{-1/2}(\theta) W_n(\theta))\right]$$

Let $\underline{\underline{H}}$ be a class of uniformly bounded Borel measurable functions of $R^k$. Then the difference

$$\int_{(\underline{\underline{H}})} E_\theta\left[f(\delta_n^{-1}(V_n - \theta))\right] \lambda(d\theta)$$

$$- \int_{(\underline{\underline{H}})} \int_{\underline{\underline{X}}_n} \left\{ \int_{R^k} f(\delta_n^{-1}(V_n - \theta) - h) S_n^*(\theta,h)\, dh \right\} dP_{\theta,n} \lambda(d\theta)$$

converges to zero uniformly for all $f \varepsilon \underline{\underline{H}}$ and for all sequence $\{V_n\}$ of estimators.

Theorem 5. Let the measure $\lambda$ be as in Proposition 1. Assume that the assumptions of Proposition 1 are satisfied. Then for every sequence $\{V_n\}$ of estimators and for every $\ell \varepsilon L$

$$\liminf_{n \to \infty} \int_{(\underline{\underline{H}})} E_\theta\left[\ell(\delta_n^{-1}(V_n - \theta))\right] \lambda(d\theta) \geq \int_{(\underline{\underline{H}})} E\left[\ell(T^{-1/2}(\theta) W)\right] \lambda(d\theta). \tag{6.6}$$

Theorem 6. Let the measure $\lambda$ be as in Proposition 1. Assume that the assumptions of Proposition 1 are satisfied. Further assume that for a sequence $\{V_n\}$ of estimators and for a non-constant $\ell \varepsilon L$

$$\lim_{n \to \infty} \int_{(\underline{H})} E_\theta [\ell(\delta_n^{-1}(V_n - \theta))] \lambda(d\theta) = \int_{(\underline{H})} E[\ell(T^{-1/2}(\theta)W)] \lambda(d\theta). \tag{6.7}$$

Then for every $\varepsilon > 0$

$$\lim_{n \to \infty} \int_{(\underline{H})} P_{\theta,n}[|\delta_n^{-1}(V_n - \theta) - T_n^{-1/2}(\theta)W_n(\theta)| > \varepsilon] \lambda(d\theta) = 0.$$

3. SOME PRELIMINARY RESULTS

Throughout what follows we set

$$S_n(\theta,h) = \exp[h' T_n^{1/2}(\theta) W_n^*(\theta) - \frac{1}{2} h' T_n(\theta) h],$$

$$S_n^*(\theta,h) = \frac{|\det T_n(\theta)|^{1/2}}{(2\pi)^{k/2}} \exp(-\frac{1}{2}[(h - T_n^{-1/2}(\theta)W_n^*(\theta))' T_n(\theta) (h - T_n^{-1/2}(\theta)W_n^*(\theta))]$$

and

$$S(\theta,h) = \exp[(h' T^{1/2}(\theta) W - \frac{1}{2} h' T(\theta) h]$$

where the sequence $\{W_n^*(\theta)\}$ is the one constructed in Lemma 6 of Ch. 1. Further let $\underline{\underline{C}}$ be the class of all sequences of $\underline{\underline{A}}_n$-measurable k-vectors and let $\underline{\underline{H}}$ be a class of uniformly bounded Borel measurable functions of $R^k$. Without further mentioning, we will also use the sequence $\{Q_n(\theta_o,h)\}$ of probability measures that was constructed in Lemma 6 of Ch. 1.

<u>Lemma 1</u>. Suppose that the assumptions of Theorem 1 are satisfied. Then for every $\alpha > 0$, the difference

$$\frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} E_{\theta_o+\delta_n h}\left[f(Z_n-h)\right] dh$$

$$- \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \left\{ \frac{\int_{D_\alpha} f(Z_n-u) S_n^*(\theta_o,u)\,du}{\int_{D_\alpha} S_n^*(\theta_o,u)\,du} \right\} dQ_n(\theta_o,h)\,dh \qquad (6.8)$$

converges to zero uniformly for all $f \varepsilon \underline{\underline{H}}$ and $Z_n \varepsilon \underline{\underline{C}}$.

Proof. First note that the difference between the r.h.s. of the above difference (6.8) and the quantity

$$\frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \left\{ \frac{\int_{D_\alpha} f(Z_n-u) S_n^*(\theta_o u)\,du}{\int_{D_\alpha} S_n^*(\theta_o,u)\,du} \right\} S_n(\theta_o,h)\,dP_{\theta_o,n}dh$$

$$= \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} f(Z_n-h) S_n(\theta_o,h)\,dP_{\theta_o,n}dh$$

converges to zero uniformly for all $f \varepsilon \underline{\underline{H}}$ and $z_n \varepsilon \underline{\underline{C}}$ by the statment (i) of Lemma 5 of Ch.1. Hence the result follows again from Lemma 6 of Ch.1.

Lemma 2. Suppose that the assumptions of Theorem 1 are satisfied. Then the difference

$$\frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} E_{\theta_o+\delta_n h}\left[f(Z_n-h)\right] dh$$

$$- \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \left\{ \int_{R^k} f(Z_n-u) S_n^*(\theta_o,u)\,du \right\} dQ_n(\theta_o,h)\,dh \qquad (6.9)$$

converges to zero uniformly for all $f \in \underline{\underline{H}}$ and $Z_n \in \underline{\underline{C}}$ by first letting $n \to \infty$ and then letting $\alpha \to \infty$.

Proof. In view of Lemma 1 it is enough to show that the difference between the r.h.s. of (6.8) and the r.h.s. of (6.9) converges to zero uniformly for all $f \in \underline{\underline{H}}$ and $Z_n \in \underline{\underline{C}}$ by first letting $n \to \infty$ and then $\alpha \to \infty$. It is easily checked that this difference is absolutely bounded by

$$\frac{2}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \left\{ \int_{D_\alpha^c} S_n^*(\theta_o, u)\,du \right\} dQ_n(\theta_o, h)\,dh \,, \qquad (6.10)$$

where $D_\alpha^c$ denotes the complement of the set $D_\alpha$. Let, for each fixed $T(\theta_o)$, $N_{T(\theta_o)}$ denotes the k-variate normal distribution with mean vector $0 \in R^k$ and co-variance matrix $T^{-1}(\theta_o)$. Then first letting $n \to \infty$ and using the statement (6) of Theorem 2.1 of LeCam (1960) it is easily seen that (6.10) converges to

$$\frac{2}{\mu^k(D_\alpha)} \int_{D_\alpha} \int N_{T(\theta_o)}(D_\alpha^c - T^{-1/2}(\theta_o)W) S(\theta_o, h)\, d\mathcal{L}(W, T(\theta_o))\,dh$$

$$= E_T \left\{ \frac{2}{\mu^k(D_\alpha)} \int_{D_\alpha} \int N_{T(\theta_o)}(D_\alpha^c - T^{-1/2}(\theta_o)W) S(\theta_o, h)\, d\mathcal{L}(W)\,dh \right\}$$

(using the independence of $T(\theta_o)$ and $W$)

$$= E_T \left\{ \frac{2}{\mu^k(D_\alpha)} \int_{D_\alpha} N_{T(\theta_o)} * N_{T(\theta_o)}(D_\alpha^c - h)\,dh \right\}$$

where $E_T$ denotes the expectation w.r.t. the law of $T(\theta_o)$. The lemma now follows from the following lemma whose ideas are contained in Hájek (1970), and stated and proved separately in Strasser (1978, Lemma 5).

Lemma 3. Let $P|\underline{\underline{B}}^k$ be a probability measure. Then

$$\lim_{\alpha \longrightarrow \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} P(D_\alpha^c - h)\,dh = 0 .$$

The following simple lemma will be used in the proof of Theorem 4 ; in the present context it serves the purpose of Lemmas 3.1, 3.2 and 3.3 of Hájek (1972).

Lemma 4. Let $\pi(h) = \frac{1}{\sigma(2\pi)^{1/2}} \exp(\frac{-h^2}{2\sigma^2})$, $\sigma > 0$. Assume that $\dim(\underline{\underline{H}}) = 1$ and that the assumptions of Theorem 1 are satisfied. Then the difference

$$\int_R E_{\theta_o+\delta_n h}[\ell(Z_n-h)]\,\pi(h)\,dh$$
$$- \int_R \int_{\underline{\underline{X}}_n} \int_R \ell(Z_n-t)\,\psi_\sigma(t,W_n^*,T_n)\,dt\,dP_{\theta_o+\delta_n h,n}\,\pi(h)\,dh \qquad (6.11)$$

tends to zero as $n \longrightarrow \infty$ for every sequence $\{Z_n\}$ of $\underline{\underline{A}}_n$-measurable k-vectors and for $\ell \in L$, where

$$\psi_\sigma(h,w,t) = \frac{(1+t\sigma^2)^{1/2}}{\sigma(2\pi)^{1/2}} \exp[-\tfrac{1}{2}(\frac{1+t\sigma^2}{\sigma^2})(h - \frac{t^{1/2}\sigma^2 w}{1+t\sigma^2})^2] .$$

Proof. First note that the difference between the $\ell.h.s$ of the above difference (6.11) and the quantity

$$\int_{|h| \leq \alpha} E_{\theta_o+\delta_n h}[\ell(Z_n-h)]\,\pi(h)\,dh, \alpha > 0 , \qquad (6.12)$$

is absolutely bounded by

$$\int_{|h|>\alpha} \pi(h)\,dh . \tag{6.13}$$

Further, since

$$\int_{|h|\le\alpha} \int_{\underline{\underline{X}}_n} \ell(Z_n-h)S_n(\theta_o,h)\,dP_{\theta_o,n}\,\pi(h)\,dh$$

$$= \int_{|h|\le\alpha} \int_{\underline{\underline{X}}_n} \left\{ \frac{\int_{|t|\le\alpha} \ell(Z_n-t)S_n^*(\theta_o,t)\pi(t)\,dt}{\int_{|t|\le\alpha} S_n^*(\theta_o,t)\pi(t)\,dt} \right\} S_n(\theta_o,h)\,\pi(h)\,dh ,$$

the difference between (6.12) and the quantity

$$\int_{|h|\le\alpha} \int_{\underline{\underline{X}}_n} \left\{ \frac{\int_{|t|\le\alpha} \ell(Z_n-t)S_n^*(\theta_o,t)\pi(t)\,dt}{\int_{|t|\le\alpha} S_n^*(\theta_o,t)\pi(t)\,dt} \right\} dP_{\theta_o+\delta_n h,n}\,\pi(h)\,dh \tag{6.14}$$

converges to zero as $n \longrightarrow \infty$ for every $\alpha > 0$ by Lemma 6 of Ch.1. Now the difference between (6.14) and the quantity

$$\int_R \int_{\underline{\underline{X}}_n} \left\{ \frac{\int_{|t|\le\alpha} \ell(Z_n-t)S_n^*(\theta_o,t)\pi(t)\,dt}{\int_{|t|\le\alpha} S_n^*(\theta_o,t)\pi(t)\,dt} \right\} dP_{\theta_o+\delta_n h,n}\,\pi(h)\,dh \tag{6.15}$$

is absolutely bounded by (6.13). Moreover it is easy to see that the difference between (6.15) and the r.h.s. of the difference (6.11) is absolutely bounded by

$$2\int_R\int_{\underline{\underline{X}}_n}\left\{\frac{\int_{|t|>\alpha} S_n^*(\theta_o,t)\,\pi(t)\,dt}{\int_R S_n^*(\theta_o,t)\,\pi(t)\,dt}\right\} dP_{\theta_o+\delta_n h,n}\,\pi(h)\,dh. \tag{6.16}$$

(Note that

$$\frac{S_n^*(\theta_o,h)\pi(h)}{\int S_n^*(\theta_o,h)\,\pi(h)\,dh} = \psi_\sigma(h, W_n^*, T_n).)$$

Obviously (6.13) tends to zero as $\alpha \longrightarrow \infty$. Using the statement (6) of Theorem 2.1 of LeCam (1960) it is easy to see that, for every $\alpha > 0$, (6.16) converges to a limit as $n \longrightarrow \infty$, and it is clear that this limit tends to zero as $\alpha \longrightarrow \infty$. Hence the proof of the lemma is complete.

## 4. PROOFS OF THE RESULTS

Proof of Theorem 1. We have

$$\lim_{\alpha\to\infty}\ \liminf_{n\to\infty} \frac{1}{\mu^k(D_\alpha)}\int_{D_\alpha} E_{\theta_o+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta_o)-h)]\,dh$$

$$= \lim_{\alpha\to\infty}\ \liminf_{n\to\infty} \frac{1}{\mu^k(D_\alpha)}\int_{D_\alpha}\int_{\underline{\underline{X}}_n}\left\{\int_{R^k}\ell(\delta_n^{-1}(V_n-\theta_o)-u)S_n^*(\theta_o,u)\,du\right\} dQ_n(\theta_o,h)\,dh \quad \text{(by Lemma 2)}$$

$$\geq \lim_{\alpha\to\infty}\ \liminf_{n\to\infty} \frac{1}{\mu^k(D_\alpha)}\int_{D_\alpha}\int_{\underline{\underline{X}}_n}\left\{\int_{R^k}\ell(T_n^{-1/2}(\theta_o)W_n^*(\theta_o)-u)S_n^*(\theta_o,u)\,du\right\} dQ_n(\theta_o,h)\,dh \quad \text{(by Anderson (1955))}$$

$$= \lim_{\alpha \to \infty} \liminf_{n \to \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} E_{\theta_o + \delta_n h} [\ell(T_n^{-1/2}(\theta_o) W_n^*(\theta_o) - h)] \, dh$$

(by Lemma 2)

$$= \lim_{\alpha \to \infty} \liminf_{n \to \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int P_{\theta_o + \delta_n h, n} [I(|T_n^{-1/2}(\theta_o) W_n^*(\theta_o) - h| > a] \, d\ell(a) \, dh$$

$$\geq \lim_{\alpha \to \infty} \int_{D_\alpha} \int \ell(T^{-1/2}(\theta_o) W - h) S(\theta_o, h) \, d\mathcal{L}(W, T(\theta_o)) \, dh \; .$$

Using independence of $W$ and $T(\theta_o)$ it is easily seen that

$$\int \ell(T^{-1/2}(\theta_o) W - h) S(\theta_o, h) \, d\mathcal{L}(W, T(\theta_o)) = \int \ell(T^{-1/2}(\theta_o) W) \, d\mathcal{L}(W, T(\theta_o)).$$

Hence the proof is complete.

<u>Proof of Theorem 2</u>. Let $\lambda_{n1} < \lambda_{n2} < \ldots < \lambda_{nk}$ be the eigen values of $T_n(\theta_o)$. An application of a result of Anderson (1955) shows that whenever $|\delta_n^{-1}(V_n - \theta_o) - T_n^{-1/2}(\theta_o) W_n^*(\theta_o)| > \varepsilon > 0, 0 < \delta \leq \lambda_{n1}$, $\lambda_{nk} \leq M$ and $\ell \varepsilon L$ is non-constant there exists a continuous function $\eta(\varepsilon, \lambda_{n1}, \ldots, \lambda_{nk})$ of $(\lambda_{n1}, \ldots, \lambda_{nk})$ such that $\eta(\varepsilon, \lambda_{n1}, \ldots, \lambda_{nk}) > 0$ and that the difference

$$\int_{R^k} \ell(\delta_n^{-1}(V_n - \theta_o) - T_n^{-1/2}(\theta_o) W_n^*(\theta_o) - u) \exp(-\tfrac{1}{2} u' T_n(\theta_o) u) \, du$$

$$- \int_{R^k} \ell(u) \exp(-\tfrac{1}{2} u' T_n(\theta_o) u) \, du$$

is greater than or equal to $\eta(\varepsilon, \lambda_{n1}, \ldots, \lambda_{nk})$. Let $\eta'(\varepsilon, \delta, M) = \inf \{ \eta(\varepsilon, \lambda_{n1}, \ldots, \lambda_{nk}) : \delta \leq \lambda_{n1} < \ldots < \lambda_{nk} \leq M \}$. In view of continuity $\eta'(\varepsilon, \delta, M) > 0$. Let $A_n = \{ |\delta_n^{-1}(V_n - \theta_o) - T_n^{-1/2}(\theta_o) W_n^*(\theta_o)| > \varepsilon, \delta \leq \lambda_{n1}, \lambda_{nk} \leq M \}$. Now it follows from the above arguments that

$$\frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \left\{ \int_{R^k} \ell(\delta_n^{-1}(V_n-\theta_o)-u) S_n^*(\theta_o,u)\,du \right\} dQ_n(\theta_o,h)\,dh$$

$$\geq \frac{\eta'(\varepsilon,\delta,M)}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} I(A_n)\,dQ_n(\theta_o,h)\,dh$$

$$+ \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \left\{ \int_{R^k} \ell(T_n^{-1/2}(\theta_o) W_n^*(\theta_o)-u) S_n^*(\theta_o,u)\,du \right\}$$

$$dQ_n(\theta_o,h)\,dh\,. \qquad (6.17)$$

In view of the given condition (6.2) and Lemma 2 we have

$$\lim_{\alpha\to\infty} \lim_{n\to\infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \int_{R^k} \ell(\delta_n^{-1}(V_n-\theta_o)-u) S_n^*(\theta_o,u)\,du \quad dQ_n(\theta_o,h)\,dh$$

$$= E[\ell(T^{-1/2}(\theta_o)W)]. \qquad (6.18)$$

Furthermore, in view of the arguments of the proof of Theorem 1 we see that

$$\lim_{\alpha\to\infty} \liminf_{n\to\infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \int_{R^k} \ell(T_n^{-1/2}(\theta_o) W_n^*(\theta_o)-u) S_n^*(\theta_o,u)\,du \; dQ_n(\theta_o,h)\,dh$$

$$\geq E[\ell(T^{-1/2}(\theta_o)W)]. \qquad (6.19)$$

From (6.17), (6.18) and (6.19) it now follows that

$$\lim_{\alpha\to\infty} \lim_{n\to\infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} I(A_n)\,dQ_n(\theta_o,h)\,dh = 0\,. \qquad (6.20)$$

Further, in view of the invariance relation (see Lemma 2 of Ch. 1)

$$\mathcal{L}(T_n(\theta_o) \mid P_{\theta_o+\delta_n h,n}) \Longrightarrow \mathcal{L}(T(\theta_o))$$

and since $T(\theta_o)$ is p.s. almost surely we see that for every $\varepsilon > 0$ there exist positive constants $\delta$ and $M$ such that

$$\lim_{\alpha \to \infty} \limsup_{n \to \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} [I(\delta > \lambda_{n1}) + I(\lambda_{nk} > M)] \, dQ_n(\theta_o, h) \, dh \leq \varepsilon \tag{6.21}$$

From (6.20) and (6.21) it follows that for every $\varepsilon > 0$

$$\lim_{\alpha \to \infty} \lim_{n \to \infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} I(|\delta_n^{-1}(V_n - \theta_o) - T_n^{-1/2}(\theta_o) W_n^*(\theta_o)| > \varepsilon) \, dQ_n(\theta_o, h) \, dh = 0 .$$

Hence the proof follows from Lemma 6 of Ch.1.

<u>Proof of Theorem 3.</u> Note that when the measurability condition of Theorem 1 is assumed, the proof is immediate from Theorem 1. To prove the general case, first note that the sequence $\{Q_n(\theta_o, h)\}$ satisfies the measurability condition of Theorem 1 and that Theorem 1 is valid when the sequence $\{P_{\theta_o + \delta_n h, n}\}$ is replaced by $\{Q_n(\theta_o, h)\}$. Partition $D_\alpha$ into blocks $C_{1m}, \ldots, C_{mm}$ such that $\sup_{1 \leq j \leq m} \mu^k(C_{jm}) \longrightarrow 0$ as $n \longrightarrow \infty$. Let $h_j$ be a fixed point in $C_{jm}$, $j = 1, \ldots, m$. It is enough to consider continuous $\ell \in L$. Then it is easy to see that the difference

$$\frac{1}{\mu^k(D_\alpha)} \sum_{j=1}^{m} \left[\int_{\underline{\underline{X}}_n} \ell(\delta_n^{-1}(V_n - \theta_o) - h_j) \, dQ_n(\theta_o, h_j)\right] \mu^k(C_{jm})$$
$$- \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} \int_{\underline{\underline{X}}_n} \ell(\delta_n^{-1}(V_n - \theta_o) - h) \, dQ_n(\theta_o, h) \, dh \tag{6.22}$$

converges to zero by first letting $n \longrightarrow \infty$ and then $n \longrightarrow \infty$. Further, it is clear that the difference between the l.h.s. of (6.22) and the quantity

$$\frac{1}{\mu^k(D_\alpha)} \sum_{j=1}^{m} E_{\theta_o+\delta_n h_j}[\ell(\delta_n^{-1}(V_n-\theta_o)-h_j)]\,\mu^k(C_{jm})$$

converges to zero as $n \longrightarrow \infty$ for every $m$ and $\alpha > 0$. Hence

$$\lim_{\alpha\to\infty} \liminf_{n\to\infty} \sup_{|h|\leq\alpha} E_{\theta_o+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta_o)-h)]\mu^k(C_{jm})$$

$$\geq \lim_{\alpha\to\infty} \lim_{m\to\infty} \liminf_{n\to\infty} \frac{1}{\mu^k(D_\alpha)} \sum_{j=1}^{m} E_{\theta_o+\delta_n h_j}[\ell(\delta_n^{-1}(V_n-\theta_o)-h_j)]\mu^k(C_{jm})$$

$$= \lim_{\alpha\to\infty} \liminf_{n\to\infty} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha}\int_{\underline{\underline{X}}_n} \ell(\delta_n^{-1}(V_n-\theta_o)-h)\,dQ_n(\theta_o,h)\,dh.$$

Hence the proof.

Proof of Theorem 4. First note that the condition (6.5) and the conclusion of Theorem 4 hold for the sequence $\{P_{\theta_o+\delta_n h,n}\}$, $h \in R^k$, if and only if they hold for the sequence $\{Q_n(\theta_o,h)\}$. Further note that Lemma 4 is valid when the sequence $\{P_{\theta_o+\delta_n h,n}\}$ is replaced by $\{Q_n(\theta_o,h)\}$. Hence it is enough to prove the theorem as it stands with the additional assumption that the functions $h \longrightarrow P_{\theta_o+\delta_n h,n}(A), A \in \underline{\underline{A}}_n$, $n \geq 1$, are Borel measurable, since this measurability assumption is satisfied for the sequence $\{Q_n(\theta_o,h)\}$. We then have, (in what follows we suppress $\theta_o$ and write $W_n, T_n$ and $T$ instead of $W_n(\theta_o), T_n(\theta_o)$ and $T(\theta_o)$),

$$E[\ell(T^{-1/2}W)] \geq \limsup_{n\to\infty} \int_R E_{\theta_o+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta_o-\delta_n h))]\pi(h)\,dh$$

(by (6.5))

$$= \limsup_{n\to\infty} \int\int_{R\underline{\underline{X}}_n}\int_R \ell(\delta_n^{-1}(V_n-\theta_o-\delta_n t))\psi_\sigma(t,W_n T_n)\,dt\,dP_{\theta_o+\delta_n h,n}\,\pi(h)\,dh$$

(by Lemma 4)

$$\geq \liminf_{n\to\infty} \int_R \int_{\underline{X}_n} \int_R \ell\left(\frac{T_n^{1/2} W_n \sigma^2}{1 + T_n \sigma^2} - t\right) \psi_\sigma(t, W_n, T_n)\, dt\, dP_{\theta_o+\delta_n h, n}\, \pi(h)\, dh$$

$$\geq \int_R \int_{R^2} \int_R \ell\left(\frac{T^{1/2} W\sigma^2}{1+T\sigma^2} - t\right) \psi_\sigma(t, W, T)\, dt\, S(\theta_o, h)\, d\mathcal{L}(T, W)\, \pi(h)\, dh$$

$$= \int_R \int_{R^2} \int_R \frac{(1+T\sigma^2)^{1/2}}{\sigma(2\pi)^{1/2}} \ell(t) \exp\left(-\frac{(1+\sigma^2 T)}{2\sigma^2} t^2\right) dt\, S(\theta_o, h)\, d\mathcal{L}(T, W)\, \pi(h)\, dh$$

$$= E\left[\frac{(1+T\sigma^2)^{1/2}}{\sigma(2\pi)^{1/2}} \int_R \ell(t) \exp\left(-\frac{(1+\sigma^2 T)}{2\sigma^2} t^2\right) dt\right]$$

[since $\mathcal{L}(T) = \mathcal{L}(T | R_{\theta_o, h})$ where $dP_{\theta_o, h} = S(\theta_o, h)\, d\mathcal{L}(T, W)$]

$$> E\left[\frac{T^{1/2}}{(2\pi)^{1/2}} \int_R \ell(t) \exp\left(-\frac{(1+\sigma^2 T)}{2\sigma^2} t^2\right) dt\right]$$

$$\geq E[\ell(T^{-1/2} W)] - E\left\{\frac{T^{1/2}}{(2\pi)^{1/2}} \int \left[\exp\left(-\frac{Tt^2}{2}\right) - \exp\left(-\frac{(1+\sigma^2 T)}{2\sigma^2} t^2\right)\right] dt\right\}$$

$$= E[\ell(T^{-1/2} W)] - E(g(T, \sigma))$$

where we set $g(T,\sigma) = \left\{(1+T\sigma^2)^{1/2} \left[(1+T\sigma^2)^{1/2} + T^{1/2}\sigma\right]\right\}^{-1}$.

Let, for some $\varepsilon > 0$, $M > \delta > 0$,

$$A_n = \left\{ \left| \delta_n^{-1}(V_n - \theta_o) - \frac{T_n^{1/2} W_n \sigma^2}{1 + T_n \sigma^2} \right| > \varepsilon,\ \delta < T_n < M,\ |W_n| \leq M \right\}.$$

Now whenever $\sigma > a > 0$ and the event $A_n$ is true the inequality $\delta < \left(\frac{1+T_n\sigma^2}{\sigma^2}\right) < M + a^{-2}$ holds. Hence it is easily seen that there exists a positive constant K depending only on $\varepsilon, \delta, a$ and M such that

$$\int_R \int_{\underline{\underline{X}}_n} \int_R \ell(\delta_n^{-1}(V_n - \theta_o - \delta_n t)) \psi_\sigma(t, W_n, T_n) \, dt \, dP_{\theta_o + \delta_n h, n} \pi(h) \, dh$$

$$\geq K \int_R \int_{\underline{\underline{X}}_n} I(A_n) \, dP_{\theta_o + \delta_n h, n} \pi(h) \, dh$$

$$+ \int_R \int_{\underline{\underline{X}}_n} \int_R \ell\left(\frac{T_n^{1/2} W_n \sigma^2}{1 + T_n \sigma^2} - t\right) \psi_\sigma(t, W_n, T_n) \, dt \, dP_{\theta_o + \delta_n h, n} \pi(h) \, dh$$

for every $n \geq 1$ whenever $\sigma > a$. Hence it follows from the series of inequalities presented in the beginning of the proof, that

$$\limsup_{n \to \infty} \int_R \int_{\underline{\underline{X}}_n} I(A_n) P_{\theta_o + \delta_n h, n} \pi(h) \, dh \leq K^{-1} E(g(T, \sigma)) \tag{6.23}$$

whenever $\sigma > a$. Now note that the difference

$$\int_{|h| \leq \alpha} \int_{\underline{\underline{X}}_n} I(A_n) S_n(\theta_o, h) \, dP_{\theta_o, n} \pi(h) \, dh$$

$$- \int_{|h| \leq \alpha} \int_{\underline{\underline{X}}_n} I(A_n) \, dP_{\theta_o + \delta_n h, n} \pi(h) \, dh$$

tends to zero for every $\alpha > 0$ by Lemma 6 of Ch. 1. Further

$$\int_{|h| > \alpha} \int_{\underline{\underline{X}}_n} I(A_n) S_n(\theta_o, h) \, dP_{\theta_o, n} \pi(h) \, dh$$

$$\leq \int_{|h| > \alpha} \int_{\underline{\underline{X}}_n} I(\delta < T_n < M, |W_n| \leq M) S_n(\theta_o, h) \, dP_{\theta_o, n} \pi(h) \, dh$$

which clearly tends to zero by first letting $n \longrightarrow \infty$ and then $\alpha \longrightarrow \infty$. Also

$$\int_{|h|>\alpha} \int_{\underline{\underline{X}}_n} I(A_n)\, dP_{\theta_o+\delta_n h, n}\, \pi(h)\, dh \leq \int_{|h|>\alpha} \pi(h)\, dh \longrightarrow 0 \text{ as } \alpha \longrightarrow \infty .$$

Therefore from (6.23) we have

$$\limsup_{n\longrightarrow\infty} \int_R \int_{\underline{\underline{X}}_n} I(A_n) S_n(\theta_o, h)\, dP_{\theta_o, n}\, \pi(h)\, dh \leq K^{-1} E(g(T,\sigma)) \qquad (6.24)$$

whenever $\sigma > a$. Now note that

$$\int_R \int_{\underline{\underline{X}}_n} I(A_n) S_n(\theta_o, h)\, dP_{\theta_o, n}\, \pi(h)\, dh$$

$$= \int_{\underline{\underline{X}}_n} \frac{1}{(1+\sigma^2 T_n)}^{1/2} I(A_n) \exp\left(\frac{T_n \sigma^2 W_n^2}{2(1+T_n\sigma^2)}\right) dP_{\theta_o, n}$$

$$\geq \exp\left(-\frac{\delta M^2\sigma^2}{1+M\sigma^2}\right) (1+\sigma^2 M)^{-1/2} \int_{\underline{\underline{X}}_n} I(A_n)\, dP_{\theta_o, n} ,$$

since $\delta < T_n < M$ and $|W_n| \leq M$ whenever the event $A_n$ occurs. Hence from (6.24) we have

$$\limsup_{n\longrightarrow\infty} [\, |\delta_n^{-1}(V_n-\theta_o) - T_n^{-1/2} W_n| > 2\varepsilon, \delta < T_n < M, |W_n| < M \,]$$

$$\leq \limsup_{n\longrightarrow\infty} P_{\theta_o, n}(A_n) + \limsup_{n\longrightarrow\infty} P_{\theta_o, n}[\, |T_n^{-1/2} W_n - \frac{T_n^{-1/2} W_n \sigma^2}{1+T_n\sigma^2}| > \varepsilon \,]$$

$$\leq K^{-1} \sigma M \exp\left(\frac{\delta M^2 \sigma^2}{1+M\sigma^2}\right) E(g(T,\sigma))$$

$$+ \limsup_{n\longrightarrow\infty} P_{\theta_o, n}[\, |T_n^{-1/2} W_n - \frac{T_n^{1/2} W_n \sigma^2}{1+T_n\sigma^2}| > \varepsilon \,] \qquad (6.25)$$

whenever $\sigma > a$. Now note that $g(T,\sigma)$ is bounded by both $T^{-1}\sigma^{-2}$ and $T^{-1/2}\sigma^{-1}$. Hence applying the dominated convergence theorem we see that the last term of the inequality (6.25) tends to zero as $\sigma \longrightarrow \infty$ whenever $E(T^{-1/2}) < \infty$. Hence the proof of the theorem is complete by choosing $\delta$ and $M$ in such a way that

$$\limsup_{n \to \infty} \left[ P_{\theta_o,n}(T_n < \delta) + P_{\theta_o,n}(T_n > M) + P_{\theta_o,n}(|W_n| > M) \right] \leq \varepsilon .$$

Before going into the details of the proof of the proposition 1, let us first observe that, when the measurability condition of Proposition 1 is satisfied, the the random functions $T_n^*(\theta)$, $\Delta_n(\theta), n \geq 1$, constructed, under the LAMN condition, in Lemmas 4 and 5 will be $\underline{\underline{A}}_n \times \underline{\underline{B}}^k$-measurable; furthermore, it is easy to see that it is enough to prove the statement of Proposition 1 with the sequences $\{W_n(\theta)\}$ and $\{T_n(\theta)\}$ replaced by $\{\Delta_n(\theta)\}$ and $\{T_n^*(\theta)\}$. Therefore in what follows we will assume, without loss of generality, that the random functions $W_n(\theta)$, $T_n(\theta), n \geq 1$, satisfy the regularity properties of Lemmas 4 and 5 and that they are jointly measurable.

Let us also observe that Lemma 2 of the present chapter is valid when the sequence $\{W_n^*(\theta_o)\}$ is replaced by $\{W_n(\theta_o)\}$ since the difference $W_n(\theta_o) - W_n^*(\theta_o)$ converges to zero in $P_{\theta_o,n}$ probability, and that the functions $C_n(\theta,h)$ will be jointly measurable whenever $W_n(\theta)$ and $T_n(\theta)$ are jointly measurable.

The following well-known result of Lebesgue will be used repeatedly in the proof, for the sake of convenience we state it separately.

Lemma 5. For any function $f \in L_p(\mu^k), p \geq 1$, the function

$$\int |f(x+h) - f(x)|^p dx$$

is uniformly continuous in $h$.

Proof. See, e.g., Corollary 39.2 of Parthasarathy (1977).

Proof of Proposition 1. For simplicity assume that $(\underline{H}) = R^k$.
Now consider

$$\int_{R^k} E_{\theta+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta-\delta_n h))]\, g(\theta)\, d\theta$$

$$= \int_{R^k} E_\theta[\ell(\delta_n^{-1}(V_n-\theta))]\, g(\theta-\delta_n h)\, d\theta$$

where $g(\theta)$ is the density of $\lambda$ with respect to $\mu^k$. Hence it follows from Lemma 5 that the difference

$$\int_{R^k} \frac{1}{\mu^k(D_\alpha)} \int_{D_\alpha} E_{\theta+\delta_n h}[\ell(\delta_n^{-1}(V_n-\theta-\delta_n h))]\, dh g(\theta)\, d\theta$$

$$- \int_{R^k} E_\theta[\ell(\delta_n^{-1}(V_n-\theta))]\, g(\theta)\, d\theta \qquad (6.26)$$

converges to zero for every $\alpha > 0$. Now note that, for $u \in R^k$,

$$\int_{R^k} \ell(\delta_n^{-1}(V_n-\theta) - h) S_n^*(\theta,h)\, dh$$

$$= \int_{R^k} \ell(\delta_n^{-1}(V_n-\theta)-h-u) S_n^*(\theta,h+u)\, dh.$$

Hence in view of Lemmas 4 and 5 of Ch.1 it is easily seen that the difference

$$\int_{R^k} \ell(\delta_n^{-1}(V_n-\theta)-h)S_n^*(\theta,h)\,dh$$
$$- \int_{R^k} \ell(\delta_n^{-1}(V_n-0)-h-u)S_n^*(\theta+\delta_n u,h)\,dh$$

converges to zero in $P_{\theta,n}$ probability and hence, by contiguity, in $P_{\theta_o+\delta_n u,n}$ probability also for every $u \varepsilon R^k$. In particular the difference

$$\int_{\underline{\underline{X}}_n}\int_{R^k} \ell(\delta_n^{-1}(V_n-\theta)-h)S_n^*(\theta,h)\,dh\,dP_{\theta+\delta_n u,n}$$
$$- \int_{\underline{\underline{X}}_m}\int_{R^k} \ell(\delta_n^{-1}(V_n-\theta)-h-u)S_n^*(\theta+\delta_n u,h)\,dh\,dP_{\theta+\delta_n u,n}$$

converges to zero as $n \longrightarrow \infty$. Hence by an application of Lemma 5 it follows that the difference

$$\int_{R^k}\int_{\underline{\underline{X}}_n}\int_{R^k} \ell(\delta_n^{-1}(V_n-\theta)-h)S_n^*(\theta,h)\,dh\,dP_{\theta+\delta_n u,n}\,g(\theta)\,d\theta$$
$$- \int_{R^k}\int_{\underline{\underline{X}}_n}\int_{R^k} \ell(\delta_n^{-1}(V_n-\theta)-h)S_n^*(\theta,h)\,dh\,dP_{\theta,n}\,g(\theta)\,d\theta$$

converges to zero for every $u \varepsilon R^k$. This in turn implies, in view of Lemma 6 of Ch.1, that the difference between the r.h.s. of this expression and the quantity

$$\frac{1}{\mu^k(D_\alpha)}\int_{D_\alpha}\int_{R^k}\int_{\underline{\underline{X}}_n}\int_{R^k} \ell(\delta_n^{-1}(V_n-\theta)-h\ )S_n^*(\theta,h)\,dh\,dQ_n(\theta,u)\,g(\theta)\,d\theta\,du \tag{6.27}$$

converges to zero for every $\alpha > 0$. Hence the result follows since, by Lemma 2, the difference between the l.h.s. of (6.26) and (6.27) converges to zero by first letting $n \longrightarrow \infty$ and then $\alpha \longrightarrow \infty$.

<u>Proof of Theorems 5 and 6</u>. Using the arguments similar to the proof of Theorems 1 and 2, the proof easily follows from Proposition 1.

CHAPTER 7

# ASYMPTOTICALLY CENTERING NATURE OF THE SEQUENCES OF MAXIMUM LIKELIHOOD ESTIMATORS, MAXIMUM PROBABILITY ESTIMATORS AND BAYES ESTIMATORS

## 1. INTRODUCTION

In order that the sequences of maximum likelihood estimators (MLE), maximum probability estimators (MPE), and Bayes estimators to be the sequences of ACS estimators, it is first of all necessary that they should take values in a small vicinity of the true value of the parameter. It is the main purpose of this chapter to see what are the minimum possible local regularity conditions needed under which the sequences of MLEs, MPEs and Bayes estimators turn out to be the sequences of ACS estimators, given that these estimators take values in a small vicinity of the true value of the parameter. A result concerning the posterior approximation at the true value of the parameter is also presented. We would like to point out that it is not the aim of this chapter to give certain regularity conditions which are less stringent than some of the possibly stringent conditions usually found in the literature dealing with the dependent observations. Our aim is just to clarify some of the local arguments usually found in the literature and to show that a lot of local conditions can be removed. Particularly our arguments depend only on the approximating form of the log-likelihood ratios, and they do not in any way depend on the nature of the sample space such as that the observations are i.i.d or any other specific from of dependent nature. For example, given that a sequence of MPEs is consistent at a certain rate, the only additional condition we assume to show that this sequence is a sequence of ACS estimators is the LAMN condition.

Results are stated in Section 2 and the proofs of the results are given in Section 3.

This chapter constitute the section 4 of Jeganathan (1979a).

## 2. STATEMENTS OF THE RESULTS

In this chapter we assume the set-up and notations of Ch.2. We define

$$L_n(X_1, \ldots, X_n; \theta) = L_n(\theta) = \prod_{j=1}^{n} f_j(\theta)$$

and

$$\Lambda_n(\theta_o + \delta_n h, \theta_o) = \log \frac{L_n(\theta_o + \delta_n h)}{L_n(\theta_o)} .$$

In the discussions of MPE and Bayes estimators it will be assumed, without further mentioning, that the functions $L_n(X_1, X_2, \ldots, X_n; \theta), n \geq 1$, are jointly measurable in $(X_1, X_2, \ldots, X_n, \theta)$. We further assume that we have chosen a particular sequence $\{\delta_n\}$ of normalising matrices of the LAMN condition.

Definition 1. A maximum probability estimator $\bar{\theta}_n(a)$ with respect to the set $D_a = \{h \varepsilon R^k : |h| \leq a\}$, $a > 0$, is defined as that value of $d$ for which the integral

$$\int L_n(\theta)\, d\theta$$

over the set $\{d - \delta_n^{-1} D_a\}$, is maximum.

We assume that a measurable maximum probability estimator exists. A detailed discussion of MPE can be found in Weiss and Wolfowitz (1974).

Theorem 1. Suppose that (i) the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfy the LAMN condition at $\theta = \theta_o$, and (ii) the sequence $\{\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o)\}$ is relatively compact for $P_{\theta_o}$. Then the the sequence $\{\bar{\theta}_n(a)\}$, $a > 0$, is a sequence of ACS estimators at $\theta = \theta_o$.

Definition 2. A measurable function $\hat{\theta}_n = \theta_n(X_1, \ldots, X_n)$ is called a maximum likelihood estimator if

$$L_n(\hat{\theta}_n) \geq L_n(\theta)$$

for all $\theta \varepsilon (\underline{\underline{H}})$. We assume that a measurable MLE exists.

Theorem 2. Suppose that (i) the sequence $\{\underline{\underline{E}}_n\}$ of experiments satisfy the LAMN-condition at $\theta = \theta_o$, (ii) for every $\varepsilon > 0$ and $\alpha > 0$, setting $D_\alpha = \{h \varepsilon R^k : |h| \leq \alpha\}$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_{\theta_o}\Big[ \sup_{|h_2 - h_2| \leq \delta} \big| \Lambda_n(\theta_o + \delta_n h_2, \theta_o) - \Lambda_n(\theta_o + \delta_n h_1, \theta_o) \big| > \varepsilon,\ h_1, h_2 \varepsilon D_\alpha \Big] = 0,$$

and (iii) the sequence $\{\delta_n^{-1}(\hat{\theta}_n-\theta_o)\}$ is relatively compact for $P_{\theta_o}$. Then the sequence $\{\hat{\theta}_n\}$ is a sequence of ACS estimators at $\theta=\theta_o$.

We next consider Bayes estimators with respect to the loss functions $\ell_n(\theta,\phi) = |\delta_n^{-1}(\theta-\phi)|^a, a\geq 1, n\geq 1$. We assume, without further mentioning, that we are given a prior density $\pi(\theta)$ such that $\pi(\theta)$ is continuous, $\sup \pi(\theta) < \infty$ and $\pi(\theta) > 0$ for all $\theta \varepsilon \underline{\underline{H}}$. The above restrictions on $\ell_n(\theta,\phi), n\geq 1$, and $\pi(\theta)$ are assumed for simplicity only and can be relaxed to a great extent.

Important works dealing with the asymptotic dehaviour of Bayes estimators are, among others, LeCam (1953, 1958), Bickel and Yahav (1969), Borwanker, et.el (1971), Ibragimov and Khasminskii (1971), Levit (1974) and Prakasa Rao (1974); it may be mentioned here that the results of these papers and the present chapter are not entirely in the "Bayesian" spirit since the results are obtained at the true value of the parameter; some deep results in the 'Bayesian framework can be found in LeCam (1974a, Ch.13), and some further results can be found in Ch.6 of the present work. Recently, Ghosh, Sinha and Joshi (1980) have obtained the asymptotic expansions associated with posterior distributions in the Bayesian' framework.

Definition 3. We define a regular Bayes estimator $t_n = t_n(X_1, X_2, \ldots, X_n)$ as an estimator which minimises

$$B_n(\emptyset) = \int \ell_n(\theta,\emptyset) f_n(\theta|X_1,\ldots,X_n)\,d\theta$$

where

$$f_n(\theta|X_1,\ldots,X_n) = \frac{\Pi(\theta) L_n(X_1,\ldots,X_n;\theta)}{\int \Pi(\theta) L_n(X_1,\ldots,X_n;\theta)\,d\theta}$$

We assume that a regular measurable Bayes estimator exists.

Theorem 3. Suppose that (i) the sequence of experiments $\{E_n\}$ satisfies the LAMN condition at $\theta = \theta_o$ and (ii) for every $\varepsilon > 0$ and for some $a \geq 0$

$$\lim_{\alpha \to \infty} \limsup_{n \to \infty} P_{\theta_o}\left[\int_{|h|>\alpha} |h|^a \exp[\Lambda_n(\theta_o+\delta_n h,\theta_o)]\,dh > \varepsilon\right] = 0. \tag{7.1}$$

Then for every sequence $\{V_n\}$ of ACS estimators at $\theta = \theta_o$ we have that, for every $0 \leq a' \leq a$, the quantity

$$\int |h|^{a'} |f_n^*(V_n+\delta_n h) - J\exp(-\tfrac{1}{2} h' T(\theta_o) h)|\,dh$$

converges to zero in $P_{\theta_o}$ probability, where we set

$$f^*(V_n+\delta_n h) = \frac{(V_n+\delta_n h)\exp\Lambda_n(V_n+\delta_n h,\theta_o)}{\int (V_n+\delta_n h)\exp\Lambda_n(V_n+\delta_n h,\theta_o)\,dh}$$

and

$$J = \frac{|\det T(\theta_o)|^{1/2}}{(2\pi)^{k/2}}.$$

Theorem 4. Suppose that (i) the sequence of experiments $\{E_n\}$ satisfies the LAMN condition at $\theta = \theta_o$ and further suppose that (7.1) is satisfied for some $a \geq 1$ and for every $\varepsilon > 0$.

Then the sequence $\{t_n\}$ of Bayes estimators with respect to the loss functions $|\delta_n^{-1}(\theta - \emptyset)|^a$ is a sequence of ACS estimators at $\theta = \theta_o$ and

$$B_n(t_n) \longrightarrow J \int |h|^a \exp(-\tfrac{1}{2} h' T(\theta_o) h) dh$$

in $P_{\theta_o}$ probability.

PROOFS OF THE RESULTS

Proof of the theorem 1. First note that the difference

$$W_n(\theta_o) - W_n^*(\theta_o)$$

converges to zero in $P_{\theta_o}$ probability, where $\{W_n^*(\theta_o)\}$ is as defined in Lemma 6 of Ch. 1. Hence it is enough to show that, for every $\delta > 0$,

$$P_{\theta_o}[|\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o) - T_n^{-1/2}(\theta_o) W_n^*(\theta_o)| > \delta] \longrightarrow 0. \quad (7.2)$$

Select $\alpha$ sufficiently large such that, for a given $\varepsilon > 0$,

$$\limsup_{n \to \infty} P_{\theta_o}[|\delta_n(\bar{\theta}_n(a) - \theta_o)| > \alpha - a] < \varepsilon/2$$

and

$$\limsup_{n \to \infty} P_{\theta_o}[|T_n^{-1/2}(\theta_o) W_n^*(\theta_o)| > \alpha - a] < \varepsilon/2.$$

Hence (7.2) will follow if we show that for every given $\varepsilon > 0$ and $\delta > 0$, there exists an $n_o$ such that

$$P_{\theta_o}(A_n) < \varepsilon/2 \quad (7.3)$$

for all $n \geq n_o$ where we set

$$A_n = \Big\{ |\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o) - T_n^{-1/2}(\theta_o) W_n^*(\theta_o)| > \delta, |\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o)| \leq \alpha - a, |T_n^{-1/2}(\theta_o) W_n^*(\theta_o)| \leq \alpha - a \Big\}.$$

Now in view of Lemma 6 of Ch. 1 we have

$$\int \int_{D_\alpha} |\exp \wedge_n(\theta_o + \delta_n h, \theta_o) - \frac{dQ_n(\theta_o, h)}{dP_{\theta_o, n}}| dP_{\theta_o} dh \longrightarrow 0 \qquad (7.4)$$

since $\int |\exp \wedge_n(\theta_o + \delta_n h, \theta_o) - \frac{dQ_n(\theta_o, h)}{dP_{\theta_o, n}}| dP_{\theta_o} \leq 2$,

where the sequence $\{Q_n(\theta_o, h)\}$ of probability measures is as defined in Lemma 6 of Ch.1. Further, since $\sup_{|h| \leq \alpha} |C_n(\theta_o, h) - 1| \longrightarrow 0$ in Lemma 6 of Ch. 1 for every $\alpha > 0$, (7.4) implies that

$$E_{\theta_o} \Big[ \int_{D_\alpha} |\exp \wedge_n(\theta_o + \delta_n h, \theta_o) - S_n(\theta_o, h)| dh \Big] \longrightarrow 0 \qquad (7.5)$$

where

$$S_n(\theta_o, h) = \exp(h' T_n^{1/2}(\theta_o) W_n^*(\theta_o) - \frac{1}{2} h' T_n(\theta_o) h).$$

Now $|\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o)| \leq \alpha - a$ implies that, setting

$H_1 = \{\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o) - D_a\}$, $H_1 \subseteq D_\alpha$. Hence (7.5) implies that

$$\int_{A_n} \int_{H_1} |\exp \wedge_n(\theta_o + \delta_n h, \theta_o) - S_n(\theta_o, h)| dh dP_{\theta_o} \longrightarrow 0 . \qquad (7.6)$$

Similarly, setting $H_2 = \{T^{-1/2}(\theta_o) W_n^*(\theta_o) - D_a\}$,

$$\int_{A_n}\int_{H_2} |\exp \Lambda_n(\theta_o+\delta_n h,\theta_o) - S_n(\theta_o,h)|\, dh dP_{\theta_o} \longrightarrow 0 \,. \qquad (7.7)$$

Now suppose that (7.3) is not true. Then for every $n_o$ there exists a $\delta > 0$ such that

$$P_{\theta_o}(A_n) > \delta \quad \text{for some} \quad n \geq n_o \,.$$

It can be easily checked <u>when the event</u> $A_n$ <u>is true</u>, that

$$\eta + \int_{H_1} S_n(\theta_o,h)\,dh < \int_{H_2} S_n(\theta_o,h)\,dt$$

for some r.v. $\eta > 0$ a.s. Since $P(A_n) > \delta > 0$ this implies that

$$\eta' + \int_{A_n}\int_{H_1} S_n(\theta_o,h)\,dh\rho P_{\theta_o} < \int_{A_n}\int_{H_2} S_n(\theta_o,h)\,dh dP_{\theta_o}$$

for some $\eta' > 0$. In view of (7.6) and (7.7) this implies that for all sufficiently large $n_o$ there exists a constant $\eta' > 0$ such that

$$\eta' + \int_{A_n}\int_{H_1} \exp \Lambda_n(\theta_o+\delta_n h,\theta_o)\,dh dP_{\theta_o} < \int_{A_n}\int_{H_2} \exp \Lambda_n(\theta_o+\delta_n h)\,dh dP_{\theta_o}$$

for some $n > n_o$. On the other hand the definition of MPE entails that

$$\int_{A_n}\int_{H_1} \exp \Lambda_n(\theta_o+\delta_n h,\theta_o)\,dh dP_{\theta_o} \geq \int_{A_n}\int_{H_2} \exp \Lambda_n(\theta_o+\delta_n h,\theta_o)\,dh dP_{\theta_o}$$

for every $n$. Thus we have arrived at a contradiction. This completes the proof.

Before presenting the proof of Theorem 2 we first prove the following lemma.

Lemma 1. Suppose that the sequence of experiments $\{\underline{\underline{E}}_n\}$ satisfies the LAMN condition at $\theta = \theta_o$. Then

$$\sup_{|h| \leq \alpha} | \Lambda_n(\theta_o + \delta_n h, \theta_o) - R_n(h) | \longrightarrow 0 \text{ in } P_{\theta_o} \text{ probability}$$

for every $\alpha > 0$, where we set

$$R_n(h) = h' T^{1/2}(\theta_o) W_n(\theta_o) - \frac{1}{2} h' T(\theta_o) h) ,$$

if and only if for every $\varepsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_{\theta_o, n} \Big[ \sup_{|h_1 - h_2| < \delta} | \Lambda_n(\theta_o + \delta_n h_2, \theta_o) - \Lambda_n(\theta_o + \delta_n h_1, \theta_o) | > \varepsilon ; h_1, h_2 \varepsilon D_\alpha \Big] = 0 . \tag{7.8}$$

Proof. Using the fact the sequence $\{W_n(\theta_o), T(\theta_o)\}$ is relatively compact for $P_{\theta_o}$ it is easily seen that for every $\varepsilon > 0$ and $\alpha > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_{\theta_o} \Big[ \sup_{|h_2 - h_1| < \delta} |R_n(h_2) - R(h_1)| > \varepsilon ; h_1, h_2 \varepsilon D_\alpha \Big] = 0 . \tag{7.9}$$

Set $Y_n(h) = \Lambda_n(\theta_o + \delta_n h, \theta_o) - R_n(h)$. Then from (7.8) and (7.9) it easily follows that for every $\varepsilon > 0$ and $\alpha > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_{\theta_o} \Big[ \sup_{|h_2 - h_1| < \delta} |Y_n(h_2) - Y_n(h_1)| > \varepsilon ; h_1, h_2 \varepsilon D_\alpha \Big] = 0 . \tag{7.10}$$

It can be now easily concluded that (7.10) together with the LAMN condition at $\theta = \theta_o$ implies that for every $\varepsilon > 0$ and $\alpha > 0$

$$P_{\theta_o}\left[ \sup_{|h| \leq \alpha} |Y_n(h)| > \varepsilon \right] \longrightarrow 0 .$$

The other part of the proof follows easily by noting that

$$\Lambda_n(\theta_o + \delta_n h, \theta_o) = Y_n(h) + R_n(h) .$$

This comples the proof of the lemma.

<u>Proof of the Theorem 2</u>. Let $\lambda$ be the smallest eigen value of $T(\theta_o)$. Since $T(\theta_o)$ is positive definite with $P_{\theta_o}$ probability one, for any given $\varepsilon > 0$, there exists a $\gamma > 0$ such that

$$P_{\theta_o}[\lambda > \gamma] \geq 1 - \varepsilon .$$

Hence, without loss of generality we can assume in what follows that $\lambda > \gamma > 0$ always. Further, for any given $\varepsilon > 0$ and $\delta > 0$, there exists an $\alpha > 0$ such that

$$\limsup_{n \to \infty} \; P_{\theta_o}[|\hat{h}_n| > \alpha] + P_{\theta_o}[|T^{-1/2}(\theta_o) W_n(\theta_o)| > \alpha - \delta] \leq \varepsilon$$

where we set $\hat{h}_n = \delta_n^{-1}(\hat{\theta}_n - \theta_o)$. Hence it is enough to show that for every $\alpha > 0$, $\delta > 0$ and $\varepsilon > 0$, there exists an $n_o$ such that

$$P_{\theta_o}[|\hat{h}_n - T^{-1/2}(\theta_o) W_n(\theta_o)| > \delta, |\hat{h}_n| \leq \alpha, |T^{-1/2}(\theta_o) W_n(\theta_o)| \leq \alpha - \delta] \leq \varepsilon$$

for every $n \geq n_o$. Equivalently we shall prove that

$$P_{\theta_o}\Big[\sup_{h\varepsilon D_\alpha} \exp \Lambda_n(\theta_o+\delta_n h,\theta_o) \leq \sup_{h\varepsilon E} \exp \Lambda_n(\theta_o+\delta_n h,\theta_o),$$
$$|h_n| \leq \alpha, |T^{-1/2}(\theta_o)W_n(\theta_o)| \leq \alpha - a\Big] \leq \varepsilon \tag{7.11}$$

for every $n \geq n_o$, where we set

$$E = \left\{h\varepsilon D_\alpha : |h - T^{-1/2}(\theta_o)W_n(\theta_o)| > \delta\right\}.$$

Denote the event inside the bracket of (7.11) by $B_n$. Let

$$A_n = \left\{ \sup_{|h|\leq\alpha} |\exp \Lambda_n(\theta_o+\delta_n h,\theta_o) - \exp R_n(h)| \leq \eta\right\}, \tag{7.12}$$

$\eta > 0$, where $R_n(h)$ is as defined in Lemma 1. By lemma 1 there exists an $n_o$ such that

$$P_{\theta_o}[A_n^c] \leq \varepsilon \quad \text{for every} \quad n \geq n_o$$

where $A_n^c$ denotes the complement of the set $A_n$. Now

$$P_{\theta_o}[B_n] \leq P_{\theta_o}[B_n \cap A_n] + P_{\theta_o}[A_n^c].$$

We shall show that $P_{\theta_o}[A_n \cap B_n) = 0$ for all $n \geq n_o$ which will prove $P_{\theta_o}[B_n] \leq \varepsilon$ for all $n \geq n_o$. Suppose that the event $A_n \cap B_n$ is true. Then by (7.12)

$$\sup_{|h|\leq\alpha} \exp \Lambda_n(\theta_o+\delta_n h,\theta_o) \geq \sup_{|h|\leq\alpha} \exp R_n(h) - \eta \tag{7.13}$$

and

$$\sup_{h\varepsilon E} \exp \Lambda_n(\theta_o+\delta_n h,\theta_o) \leq \sup_{h\varepsilon E} \exp R_n(h) + \eta. \tag{7.14}$$

Also note that (since $|T^{-1/2}(\theta_o)W_n(\theta_o)| < \alpha - \delta$ when the event $B_n$ is true)

$$\sup_{|h|\leq \alpha} \exp R_n(h) = \exp\left(\frac{1}{2} W_n'(\theta_o) W_n(\theta_o)\right) \tag{7.15}$$

and

$$\sup_{h\varepsilon E} \exp R_n(h) \leq \exp\left[-\frac{\gamma\delta^2}{2} + \frac{1}{2} W_n'(\theta_o W_n(\theta_o)\right]. \tag{7.16}$$

Since $\gamma > 0$, there exists an $n_o$ such that

$$\exp \frac{1}{2} W_n'(\theta_o) W_n(\theta_o) - \eta_o > \exp\left[-\frac{\gamma\delta^2}{2} + \frac{1}{2} W_n'(\theta_o) W_n(\theta_o)\right] + \eta_o . \tag{7.17}$$

Thus we see from (7.13) - (7.17), that whenever $\eta \leq \eta_o$ and the event $A_n \cap B_n$ is true

$$\sup_{|h|\leq \alpha} \exp \Lambda_n(\theta_o + \delta_n h, \theta_o) > \sup_{h\varepsilon E} \exp \Lambda_n(\theta_o + \delta_n h, \theta_o)$$

for all $n \geq n_o$. But this is a contradiction since on $B_n$ we have, for every $n \geq 1$,

$$\sup_{|h|\leq \alpha} \exp \Lambda_n(\theta_o + \delta_n h, \theta_o) \leq \sup_{h\varepsilon E} \exp \Lambda_n(\theta_o + \delta_n h, \theta_o) .$$

Hence $P_{\theta_o}[A_n \cap B_n] = 0$ for all $n \geq n_o$. This proves the result.

We next consider the proofs of the theorems 3 and 4; since the proofs are long we split the proofs into several lemmas. To simplify the notations we set $\eta_n(\theta_o) = T^{1/2}(\theta_o)\delta_n^{-1}(V_n - \theta_o)$ and $R_n^*(h) = \exp(h' T^{1/2}(\theta_o)\eta_n(\theta_o) - \frac{1}{2} h' T(\theta_o) h)$, where $\{V_n\}$ is a sequence of ACS estimators at $\theta = \theta_o$.

<u>Lemma 2</u>. For every $\varepsilon > 0$ and $a \geq 0$

$$\lim_{\alpha \to \infty} \limsup_{n \to \infty} P_{\theta_o}\left[\int_{|h|>\alpha} |h|^a R_n^*(h)\, dh > \varepsilon\right] = 0 .$$

__Proof.__ Consider

$$\int |h|^{a+1} R_n^*(h)dh = Q(a, \eta_n(\theta_o), T(\theta_o))$$

where $Q(a,.,.)$ is a continuous function. Since the sequence $\{\eta_n(\theta_o), T(\theta_o)\}$ is relatively compact for $P_{\theta_o}$ we see that, for any given $\varepsilon > 0$, there exists a constant $A > 0$ such that

$$P_{\theta_o}[Q(a, \eta_n(\theta_o), T(\theta_o)) \leq A] \geq 1 - \varepsilon$$

for every $n$. Let $\alpha_o = A/\varepsilon$. Since

$$\int_{|h| > \alpha} |h|^a R_n^*(h)dh \leq \alpha^{-1} \int |h|^{a+1} R_n^*(h)dh$$

we then have

$$P_{\theta_o}\left[\int_{|h| > \alpha} |h|^a R_n^*(h)dh \leq A/\alpha \leq \varepsilon\right] \geq 1 - \varepsilon$$

for every $\alpha \geq \alpha_o$ and for all $n \geq 1$. This proves the result.

__Lemma 3.__ Suppose the assumptions of Theorem 3 are satisfied. Then for every $0 \leq a' \leq a$, the quantity

$$\int |h|^{a'} |\pi(\theta_o + \delta_n^{-1}h) \exp \Lambda_n(\theta_o + \delta_n h, \theta_o) - \pi(\theta_o) R_n^*(h)| dh$$

tends to zero in $P_{\theta_o}$ probability.

__Proof.__ (7.5) implies that, for every $\alpha > 0$,

$$\int_{|h| \leq \alpha} |\exp \Lambda_n(\theta_o + \delta_n h, \theta_o) - S_n(\theta_o, h)| \longrightarrow 0 \qquad (7.18)$$

in $P_{\theta_o}$ probability. Since the sequence $\{V_n\}$ is a sequence of ACS estimators at $\theta = \theta_o$,

$$|T^{1/2}(\theta_o)\eta_n(\theta_o) - T_n^{1/2}(\theta_o)W_n(\theta_o)| \longrightarrow 0$$

in $P_{\theta_o}$ probability. Hence it is not difficult to see from (7.18) that

$$\int_{|h|\leq \alpha} |\exp \Lambda_n(\theta_o+\delta_n h,\theta_o) - R_n^*(h)|dh \longrightarrow 0$$

in $P_{\theta_o}$ probability for every $\alpha > 0$. Hence it follows that, since $\pi(\theta)$ is continuous at $\theta = \theta_o$,

$$\int_{|h|\leq \alpha} |h|^{a'} |\pi(\theta_o+\delta_n h)\exp \Lambda_n(\theta_o+\delta_n h,\theta_o) - \pi(\theta_o)R_n^*(h)|dh \longrightarrow 0 \qquad (7.19)$$

in $P_{\theta_o}$ probability for every $a' \geq 0$ and $\alpha > 0$. Now (7.1) implies that, since $\sup_{\theta\varepsilon(\overline{H})} \pi(\theta) < \infty$, for every $\varepsilon > 0$ and $0 \leq a' \leq a$

$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} P_{\theta_o}\left[\int_{|h|>\alpha} |h|^{a'}\pi(\theta_o+\delta_n h)\exp\Lambda_n(\theta_o+\delta_n h,\theta_o) > \varepsilon\right] = 0. \qquad (7.20)$$

Now (7.19) and (7.20) together with the Lemma 2 give the required result.

<u>Lemma 4</u>. Suppose that the assumptions of Theorem 3 are satisfied. Then for every $0 \leq a' \leq a$,

$$e_n(\theta_o,a') = \int |h|^{a'} |f_n^*(\theta_o+\delta_n h) - J\exp\left[-\tfrac{1}{2}(h-T^{-1/2}(\theta_o)\eta_n(\theta_o))'T(\theta_o)(h-T^{-1/2}(\theta_o)\eta_n(\theta_o))\right]| dh \longrightarrow 0 \qquad (7.21)$$

in $P_{\theta_o}$ probability, where we set

$$R_n^*(\theta_o+\delta_n h) = \frac{\pi(\theta_o+\delta_n h)\exp \Lambda_n(\theta_o+\delta_n h,\theta_o)}{\int \pi(\theta_o+\delta_n h)\exp \Lambda_n(\theta_o+\delta_n h,\theta_o)dh} .$$

Proof. Let

$$Y_n = \int \pi(\theta_o+\delta_n h)\exp \Lambda_n(\theta_o+\delta_n h,\theta_o)dh .$$

Setting $a'=0$ in Lemma 3, we see that

$$|Y_n - \pi(\theta_o)J^{-1}\exp(\tfrac{1}{2}\eta_n'(\theta_o)\eta_n(\theta_o))| \longrightarrow 0 \tag{7.22}$$

in $P_{\theta_o}$ probability. In particular we can assume that the sequence $\{Y_n^{-1}\}$ is well defined with $P_{\theta_o}$ probability tending to one. Now consider the inequality

$$e_n(\theta_o,a') \leq Y_n^{-1}\int |h|^{a'}|\pi(\theta_o+\delta_n h)\exp\Lambda_n(\theta_o+\delta_n h,\theta_o) - \pi(\theta_o)R_n^*(h)|dh$$
$$+ |Y_n^{-1}\pi(\theta_o) - J\exp[-\tfrac{1}{2}\eta_n'(\theta_o)\eta_n(\theta_o)]|\int |h|^{a'}R_n^*(h)dh$$
$$= I_1 + I_2 \text{ , say .}$$

First note that $I_2 \longrightarrow 0$ in $P_{\theta_o}$ probability by (7.22) since the proof of the lemma 2 shows that the sequence $\{\int |h|^{a'}R_n^*(h)dh\}$ is bounded in $P_{\theta_o}$ probability. Secondly $I_1 \longrightarrow 0$ in $P_{\theta_o}$ probability by Lemma 3 since (7.22) implies that the sequence $\{Y_n^{-1}\}$ is bounded in $P_{\theta_o}$ probability. This completes the proof of the lemma.

Proof of the theorem 3. Let

$$g_n = h - \delta_n^{-1}(V_n - \theta_o) .$$

Then

$$|g_n|^{a'} \leq c_{a'} |h|^{a'} + c_{a'} |\delta_n^{-1}(V_n - \theta_o)|^{a'}$$

where $c_{a'} = 1$ or $2^{a'-1}$ according as $a' \leq 1$ or $a' > 1$.

Using this inequality we have, noting $\eta_n(\theta_o) = T^{1/2}(\theta_o)\delta_n^{-1}(V_n - \theta_o)$,

$$\int |g|^{a'} |f_n(V_n + \delta_n g) - J \exp(-\tfrac{1}{2} g T(\theta_o) g)| dg$$

$$= \int |g_n|^{a'} |f_n(V_n + \delta_n g_n) - J \exp(-\tfrac{1}{2} g_n T(\theta_o) g_n)| dg_n$$

$$\leq c_{a'} e_n(\theta_o, a') + c_{a'} |\delta_n^{-1}(V_n - \theta_o|^{a'} e_n(\theta_o, 0) \longrightarrow 0$$

in $P_{\theta_o}$ probability by Lemma 4 where the quantities $e_n(\theta_o, a')$, $n \geq 1$, are as defined in (7.21). This completes the proof of the theorem.

<u>Proof of Theorem 4</u>. Since $T(\theta_o)$ is positive definite a.s. we will assume without loss of generality that both the smallest and largest eigen values of $T(\theta_o)$ are bounded both below and above by some positive constants. Now define

$$g(x) = |x|^a, \ a \geq 1,$$

and, for a given $\alpha > 0$,

$$g_\alpha(x) = |x|^a \quad \text{if} \quad |x|^a \leq \alpha$$

$$= \alpha \quad \text{otherwise}.$$

Then select $\alpha_o$ so large such that, for a given $\varepsilon > 0$,

$$J \int g_{\alpha_o}(h) \exp(-\tfrac{1}{2} h' T(\theta_o) h) dh \geq J \int g(h) \exp(-\tfrac{1}{2} h' T(\theta_o) h) dh - \varepsilon/2 .$$

According to Theorem 3 we have, setting $V_n = \delta_n^{-1} T^{-1/2}(\theta_o) W_n(\theta_o) + \theta_o$,

$$\int g(h)\,|f_n^*(V_n+\delta_n h) - J \exp(-\tfrac{1}{2} h' T(\theta_o) h)|\,dh \longrightarrow 0$$

in $P_{\theta_o}$-probability and, since $g_{\alpha_o}(x)$ is bounded by $\alpha_o$,

$$\int g_{\alpha_o}(h+u_n)\,|f_n^*(V_n+\delta_n h) - J \exp(-\tfrac{1}{2} h' T(\theta_o) h)|\,dh \longrightarrow 0$$

in $P_{\theta_o}$ probability, where we set $u_n = \delta_n^{-1}(t_n - V_n)$. Hence, for any given $\varepsilon > 0$, there exists an $n_o$ such that

$$P_{\theta_o}[A_{n\varepsilon}^{(1)} \cap A_{n\varepsilon}^{(2)}] > 1 - \varepsilon \qquad (7.22)$$

for all $n \geq n_o$, where we set

$$A_{1\varepsilon}^{(1)} = \left\{ \int g(h) f_n^*(V_n+\delta_n h)\,dh \leq J \int g(h) \exp(-\tfrac{1}{2} h' T(\theta_o) h)\,dh + \varepsilon \right\}$$

and

$$A_{2\varepsilon}^{(2)} = \left\{ \int g_{\alpha_o}(h+u_n) f_n^*(V_n^*+\delta_n h)\,dh \geq J \int g_{\alpha_o}(h+u_n) \exp(-\tfrac{1}{2} h' T(\theta_o) h) - \varepsilon \right\}.$$

First we shall prove that the sequence $\{t_n\}$ is a sequence of ACS estimators at $\theta = \theta_o$, i.e., we want to prove that, for every $\delta > 0$, $P_{\theta_o}[|u_n| > \delta] \longrightarrow 0$. In view of (7.22) it is enough to prove that for all sufficiently small $\varepsilon$, the event $B_{n\varepsilon} = A_{n\varepsilon}^{(1)} \cap A_{n\varepsilon}^{(2)} \cap \{|u_n| > \varepsilon\}$ is impossible for all $n \geq n_o$. In what follows suppose that the event $B_{n\varepsilon}$ is true. Using the defifition of Bayes estimators we then have on the set $A_{n\varepsilon}^{(1)}$, for every $n \geq n_o$,

$$\int g_{\alpha_o}(h+u_n)f^*(V_n+\delta_n h)\,dh \leq \int g(h+u_n)f_n^*(V_n+\delta_n h)\,dh$$
$$= B_n(t_n)$$
$$\leq B_n(V_n)$$
$$= \int g(h)f_n^*(V_n+\delta_n h)\,dh$$
$$\leq J\int g(h)\exp(-\tfrac{1}{2}h'T(\theta_o)h)\,dh + \varepsilon . \quad (7.23)$$

Now note that $g_{\alpha_o}(h)$ is a non-constant function and satisfies $g_{\alpha_o}(0) = 0$, $g_{\alpha_o}(|h|) = g_{\alpha_o}(h)$ and $g_{\alpha_o}(h_1) \leq g_{\alpha_o}(h_2)$ if $|h_1| \leq |h_2|$. Hence we have, for some $\eta > 0$,

$$J\int g_{\alpha_o}(h+u_n)\exp(-\tfrac{1}{2}h'T(\theta_o)h)\,dh$$
$$> J\int g_{\alpha_o}(h)\exp(-\tfrac{1}{2}h'T(\theta_o)h)\,dh + \eta$$

whenever $|u_n| > \delta > 0$. Thus on the set $A_{n\varepsilon}^{(2)} \cap \{|u_n| > \delta\}$ we have for every $n \geq n_o$

$$\int g_{\alpha_o}(h+u_n)f_n^*(V_n+\delta_n h)\,dh \geq J\int g_{\alpha_o}(h+u_n)\exp(-\tfrac{1}{2}h'T(\theta_o)h)\,dh - \varepsilon/2$$
$$> J\int g_{\alpha_o}(h)\exp(-\tfrac{1}{2}h'T(\theta_o)h)\,dh - \varepsilon/2 + \eta$$
$$\geq J\int g(h)\exp(-\tfrac{1}{2}h'T(\theta_o)h) + \varepsilon \quad (7.24)$$

for all $0 < \varepsilon \leq \eta$. From (7.23) and (7.24) we thus see that the event $B_{n\varepsilon}$ is impossible for all $n \geq n_o$ and $\varepsilon \leq \eta$. This proves that the sequence $\{t_n\}$ is a sequence of ACS estimators.

Now it follows from the previous arguments that

$$J \int g_{\alpha_o}(h+u_n)\exp(-\tfrac{1}{2}\, h' T(\theta_o)h)\,dh - \varepsilon/2$$

$$\leq \int g_{\alpha_o}(h+u_n) f_n^*(V_n+\delta_n h)\,dh$$

$$\leq \int g(h+u_n) f_n^*(V_n+\delta_n h)\,dh$$

$$\leq J \int g(h)\exp(-\tfrac{1}{2}\, h' T(\theta_o)h)\,dh + \varepsilon \qquad (7.25)$$

with $P_{\theta_o}$-probability tending to one. Since $P_{\theta_o}[\,|u_n| > \varepsilon] \longrightarrow 0$ for every $\varepsilon > 0$, it easily follows that the quantity

$$J \int g_{\alpha_o}(h+u_n)\exp(-\tfrac{1}{2}\, h' T(\theta_o)h)\,dh$$

converges in $P_{\theta_o}$ probability to

$$J \int g_{\alpha_o}(h)\exp(-\tfrac{1}{2}\, h' T(\theta_o)h)\,dh$$

$$\geq J \int g(h)\exp(-\tfrac{1}{2}\, h' T(\theta_o)h)\,dh - \varepsilon/2 . \qquad (7.26)$$

Combining (7.25) and (7.26) we see that

$$B_n(t_n) = \int g(h+u_n) f_n^*(V_n+\delta_n h)\,dh$$

converges in $P_{\theta_o}$ probability to

$$J \int g(h)\exp(-\tfrac{1}{2}\, h' T(\theta_o)h)\,dh .$$

This proves the theorem completely.

CHAPTER 8

# CONVERGENCE OF MOMENTS OF STATISTICAL ESTIMATORS

## 1. INTRODUCTION

In an important paper Ibragimov and Khasminskii (1972 and 1973) (henceforth this paper will be referred in short by I.K.) considered the asymptotic behaviour of maximum likelihood estimators (MLE) and a certain class of Bayes estimators when the observations are i.i.d and when the parameter space is a subset of the real line. Among other things, they proved that the moments of any order of the above mentioned estimators converge to the corresponding moments of a normal distribution. Their investigations are based on a general method which amounts to treat the likelihood function as a random function of the parameter; we would like to mention here that this general method of investigating MLE was first developed by Prakasa Rao (1968). A similar method of investigating MLE was also developed by LeCam (1970). This approach offered some fresh insights into the problems and as a result I.K. were able to prove powerful results under quite general assumptions. However, so far as the practical purposes are considered, the situation considered by I.K. was not quite general in the sense that the parameter space was assumed to be a subset of the real line. Though the methods of analysis of I.K.

were simple, some of the arguments depend in a very crucial manner on the dimension of the parameter space. For example, they invoke Kolmogorov's sufficient condition for the continuity of a random process to get some estimates of the continuity modulus of the realisations of the likelihood function (see Prokhorov (1956, p.180)). However, it does not seem to be possible to extend this idea to multidimension. In fact LeCam (1970) mentions that some of the arguments about continuity of sample paths for random processes do not extend directly to more than one-dimension. At the same time there is no reason to suppose that the results on the convergence of moments of statistical estimators would depend on the dimensionality restriction of the parameter space. This suggests that one can obtain the same type of estimates of the continuity modulus by some other methods whose arguments would not depend on the dimension of the parameter space. It is also important to know how far the results on the convergence of moments can be extended to the situation where the observations are not necessarily i.i.d. Thus our aim in this chapter is to prove that the moments of any order of MLE, maximum probability estimators (MPE) and a certain class of Bayes estimators converge to the corresponding moments of a mixed normal distribution when the observations are not necessarily i.i.d, including the LAMN case, and the parameter space is a subset of $R^k$, $k \geq 1$; as a by product we also present a weak convergence result for the likelihood ratio random processes.

Mejority of the ideas of this chapter are either inspired by or adapted from I.K. though we have substantially simplified the proofs. Also we have avoided using weak convergence results for random processes, which I.K. use freely in their paper. We would like to point out that the weak convergence results for the likelihood ratio process have been extended by Inagaki and Ogata (1975 and 1977) to the situations where the parameters space is of multidimension and the observations are from a strictly stationary markov process, with a number of interesting applications. When the parameter space is a subset of the real line, the results of I.K. on the convergence of moments have been extended to the idependent not necessarily identically distributed case by Ibragimov and Khasminskii (1975) and to a certain class of markov chains by Levit (1974).

In Section 2 we introduce the assumptions. One of the assumptions((A.10)) of this section is very direct. The reason is that we are not able to impose satisfactory conditions on the densities implying this assumption in the general case. However, it is possible to verify this assumption directely in some problems (for example, for a certain class of mixed Gaussian processes). In the situations where the observations have a certain 'mild' form of dependence it is possible to impose conditions on the densities implying this assumption. These things are done in Section 5. In section 3 we obtain some preliminary results on the behaviour of the likelihood function which are used in Section 4 where the results on the converyence of moments of MLE, MPE and a certain class of Bayes estimators are obtained. In Section 3 we also show, as a by-product of our assumptions, that the likelihood functions belong to a certain comple separable metric space for all sufficiently large sample size with probability one and the corresponding sequence of induced probability measures on this metric space converge weakly to the probability measure induced by a mixed Gaussian shift process.

This chapter is a revised version of Jeganathan (1979b).

## 2. ASSUMPTIONS

Notations and the set up of Ch.2 are assumed in this chapter. In what follows, unless otherwise specified, all the probability concepts and expectations are with respect to $P_{\theta_o}$ where $\theta_o \varepsilon \underline{(H)}$ denotes the " true " value of the parameter.

(A.1) For all $(X_1, \ldots, X_j)$ and for every $j \geq 1$ the functions $\theta \longrightarrow f_j(\theta), j \geq 1$, are absolutely continuous in $\theta$.

(A.2) For $\mu_1 \times \ldots \times \mu_j$ almost all $(X_1, \ldots, X_j)$ and for every $j \geq 1$, the functions $\theta \longrightarrow \log f_j(\theta)$ are differentiable in $\theta$.

Remark. Note that implicit in (A.2) is the assumption that, for $\mu_1 \times \ldots \times \mu_j$ almost all $(X_1, \ldots, X_j)$ and for every $j \geq 1$, $\log f_j(\theta)$ is finite for all $\theta$.

Set $\eta_j(\theta) = (\partial / \partial \theta) \log f_j(\theta)$ if the derivative exists,

$= 0$ otherwise.

Suppose that we have selected a suitable sequence $\{\delta_n\}$ of normalising matrices; one way of selection is to set

$$\delta_n'\delta_n = \left[ \sum_{j=1}^{n} E_{\phi} \left[ \eta_j'(\phi)\eta_j(\phi) \right] \right]^{-1}$$

for some fixed $\phi \varepsilon (\underline{H})$, where $E_{\phi}$ denotes the expectation with respect to $P_{\phi}$. Further we set

$$\xi_j(\theta) = \eta_j(\theta) f_j^{1/2}(\theta).$$

(A.3) For every $h \varepsilon R^k$

$$E\left[ \int |h'\delta_n \xi_j(\theta_o)|^2 d\mu_j \right] < \infty, \quad 1 \leq j \leq n < \infty.$$

(A.4) For every $h \varepsilon R^k$, for some $a < 0$ and $b > 1$

$$\sup_{a \leq t \leq b} \sum_{j=1}^{n} E\left\{ \int |h'\delta_n[\xi_j(\theta_o + \delta_n h) - \xi_j(\theta_o)]|^2 d\mu_j \right\} \longrightarrow 0.$$

(A.5) $E[\eta_j(\theta_o) | A_{j-1}] = 0$ for every $j \geq 1$.

(A.6) For every $\varepsilon > 0$ and $h \varepsilon R^k$

$$E[|h'\delta_n \eta_j(\theta_o)|^2 I(|h'\delta_n \eta_j(\theta_o)| > \varepsilon)] \longrightarrow 0.$$

(A.7) There exists an a.s. positive definite random matrix $T(\theta_o)$ such that the difference

$$\delta_n \sum_{j=1}^{n} [\eta_j'(\theta_o)\eta_j(\theta_o) | \underline{A}_{j-1}] \delta_n - T(\theta_o)$$

converges to zero in probability.

(A.8) $\sup_{n \geq 1} \| \sum_{j=1}^{n} E[\delta_n \eta_j'(\theta_o)\eta_j(\theta_o)\delta_n] \| \leq K$ for some $K > 0$.

(A.9) $$E\left\{\sup_{a \le |h| \le a+1} \left|\delta_n \sum_{j=1}^{n} [\eta_j(\theta_o+\delta_n h) - \eta_j(\theta_o)]\right|\right\} \le C a^p$$

for some constants $C > 0$ and $p > 0$ and for all sufficiently large $n$.

(A.10) To any positive $N$ there exists an $n_o$ and a constant $C_N$ depending only on $N$ such that for every $n \ge n_o$

$$P\left[\prod_{j=1}^{n} [f_j(\theta_o+\delta_n h)/f_j(\theta_o)] \ge \frac{1}{|h|^N}\right] \le \frac{C_N}{|h|^N}$$

Remark. See Section 5 for a discussion of this condition (A.10).

The next assumption will be used only in proving the results for MPE and Bayes estimators.

(A.11) There exists an $\varepsilon_o > 0$ such that for all $0 < \varepsilon \le \varepsilon_o$ and for all sufficiently large $n$

$$E\left\{\sup_{|h| \le \varepsilon} \left|\delta_n \sum_{j=1}^{n} [\eta_j(\theta_n+\delta_n h) - \eta_j(\theta_o)]\right|\right\} \le C\varepsilon$$

where $C$ is some constant.

## 3. STUDY OF THE LIKELIHOOD FUNCTION

Throughout what follows we set

$$Z_n(h) = \prod_{j=1}^{n} (f_j(\theta_o+\delta_n h)/f_j(\theta_o)] .$$

Theorem 1. Suppose the assumptions (A.1) - (A.8) are satisfied. Then for every $h \varepsilon R^k$, the difference

$$Z_n(h) - \exp[h' T^{1/2}(\theta_o) W_n(\theta_o) - \tfrac{1}{2} h' T(\theta_o) h]$$

converges to zero in probability and

$$\mathcal{L}(W_n(\theta_o), T(\theta_o)) \Longrightarrow \mathcal{L}(Z, T(\theta_o))$$

where

$$W_n(\theta_o) = T^{-1/2}(\theta_o)\left[\delta_n \sum_{j=1}^{n} \eta_j(\theta_o)\right]$$

and $Z$ is a copy of the standard k-variate normal distribution independent of $T(\theta_o)$.

Proof. The proof follows from Theorem 1 and proposition 1 of Ch. 2.

The next lemma gives an estimate of the continuity modulus of the processes $h \longrightarrow \log Z_n(h)$, $n \geq 1$.

Lemma 1. Suppose the assumptions (A.2), (A.5), (A.8) and (A.9) are satisfied. Then for some constant $C > 0$

$$P\left[\sup_{|h_2-h_1| \leq d} |\log Z_n(h_2,\theta_o) - \log Z_n(h_1,\theta_o)| \geq d^{1/2};\ h_1, h_2 \in B_a\right] \leq C a^p d^{1/2}$$

where the set $B_a = \{h \in R^k; a \leq |h| \leq a+1\}$ and $p$ is the positive constant occuring in the condition (A.9).

Proof. Consider, for $\mu_1 \times \mu_2 \times \ldots \times \mu_n$ almost all $(X_1, X_2, \ldots, X_n)$

$$\log Z_n(h_2) - \log Z_n(h_1) = (h_2-h_1)'\delta_n \sum_{j=1}^{n} \eta_j(\theta_n^*)$$

$$= (h_2-h_1)'\delta_n \sum_{j=1}^{n} \eta_j(\theta_o) + (h_2-h_1)'\delta_n \sum_{j=1}^{n} \left[\eta_j(\theta_n^*) - \eta_j(\theta_o)\right]$$

where

$$|\theta_n^* - (\theta_o + \delta_n h_1)| \leq |\delta_n(h_2 - h_1)| .$$

Now

$$P\Big[\sup_{|h_2-h_1|\leq d} |(h_2-h_1)'\delta_n \sum_{j=1}^{n} \eta_j(\theta_o)| \geq d^{1/2}/2\Big]$$

$$\leq 4d\, E\Big[\,|\, \phi_n \sum_{j=1}^{n} \eta_j(\theta_o)|^2\, \|\delta_n \phi_n^{-1}\|^2\Big] \qquad (8.1)$$

where $\phi_n'\phi_n = \left[\sum_{j=1}^{n} E[\eta_n(\theta_o)\eta_j'(\theta_o)]\right]^{-1}$. It can be easily seen, using the fact that

$$E[\eta_j(\theta_o)|\underline{\underline{A}}_{j-1}] = 0 \quad \text{for all} \quad j \geq 1$$

and

$$E\left\{\phi_n \sum_{j=1}^{n} [\eta_j(\theta_o)\eta_j'(\theta_o)]\, \phi_n\right\} = I \quad \text{(unit matrix)}$$

for all $n$, that

$$E\Big[\,|\phi_n \sum_{j=1}^{n} \eta_j(\theta_o)|^2\Big] = k. \qquad (8.2)$$

Hence from (8.1), (8.2) and (A.8) we see that for some constant $C > 0$

$$P\Big[\sup_{|h_2-h_1|\leq d} |(h_2-h_1)'\delta_n \sum_{j=1}^{n} \eta_j(\theta_o)| \geq d^{1/2}/2\Big] \leq d\,4\,C \qquad (8.3)$$

for all sufficiently large $n$. Next by (A.9) we have

$$P\left\{\sup_{|h_2-h_1|\leq d} |(h_2-h_1)'\delta_n \sum_{j=1}^{n} [\eta_j(\theta_n^*) - \eta_j(\theta_o)] \geq d^{1/2}/2\,;\, h_1,h_2 \in B_a\right\}$$

$$\leq 2d^{1/2} C a^p \qquad (8.4)$$

for some constants $C > 0$ and $p > 0$. Hence the result follows from (8.3) and (8.4).

The results of the next theorem will be instrumental in the next section.

Theorem 2. Suppose that the assumptions of Lemma 1 and (A.10) are satisfied. Then to any positive $N$ there exist an $n_o$ and a constant $C_N$ depending only on $N$ such that for $n \geq n_o$

$$P\Big[ \sup_{|h| \geq a} Z_n(h) \geq \frac{1}{a^N} \Big] \leq \frac{C_N}{a^N}, \quad a \geq 2 \tag{8.5}$$

and

$$P\Big[ \sup_{a \leq |h| \leq a+1} Z_n(h) \geq \frac{1}{a^N} \Big] \leq \frac{C_N}{a^N}, \quad a \geq 2. \tag{8.6}$$

Proof. By virtue of the inequality

$$P\Big[ \sup_{|h| \geq a} Z_n(h) \geq 2 \sum_{k=0}^{\infty} \frac{1}{(a+h)^N} \Big]$$

$$\leq \sum_{k=0}^{\infty} P\Big[ \sup_{a+k \leq |h| \leq a+k+1} Z_n(h) > \frac{1}{(a+k)^N} \Big]$$

relation (8.5) is a consequence of (8.6), whose derivation we shall now consider. We partition the set $\{h : a \leq |h| \leq a+1\}$ into cubes of sides of length $a^{-5N}$. Then totally we will have $a^{5kN}$ number of cubes. Denote the $i$ th cube and its center by $D_i$ and $t_i$ respectively. Then, for $a \geq 2$,

$$P\Big[ \sup_{a \leq |h| \leq a+1} Z_n(h) > \frac{1}{a^N} \Big] \leq P\Big[ \sup_{i} \sup_{h \in D_i} Z_n(h) > \frac{1}{a^N} \Big]$$

$$\leq P\Big[ \sup_{i} Z_n(t_i) > \frac{1}{a^{5kN+N}} \Big]$$

$$+ P\Big[ \sup_{i} \sup_{h \in D_i} |Z_n(h) - Z_n(t_i)| > \frac{1}{a^{5N}} \Big].$$

Now using the fact that, for every $x > 0$ and $y > 0$, $x < \frac{\delta}{2}^{1/2}$ and $|\log x - \log y| \leq \delta^{1/2}$ implies $|x - y| \leq \delta$ whenever $\delta^{1/2} \leq \log 2$, we have, for $a \geq 2$,

$$P[\sup_i \sup_{h \varepsilon D_i} |Z_n(h) - Z_n(t_i)| > \frac{1}{a^{5N}}]$$

$$\leq P[\sup_i Z_n(t_i) > \frac{1}{2a^{\frac{5}{2}N}}] + P[\sup_i \sup_{h\varepsilon D_i} |\log Z_n(h) - \log Z_n(t_i)| > \frac{1}{a^{\frac{5}{2}N}}].$$

Further, for $a \geq 2$,

$$P[\sup_i Z_n(t_i) \geq \frac{1}{2a^{\frac{5}{2}N}}] \leq P[\sup_i Z_n(t_i) \geq \frac{1}{a^{5kN+N}}]$$

$$\leq \sum_{i=1}^{a^{5kN}} P[Z_n(t_i) > \frac{1}{a^{5kN+N}}] \leq \frac{C_N}{a^N} \quad \text{(by (A.10))}.$$

Hence using Lemma 1 we have, for $a \geq 2$,

$$P[\sup_{a \leq |h| \leq a+1} Z_n(h) > \frac{1}{a^N}] \leq \frac{2C_N}{a^N} + \frac{C_N}{a^{\frac{5}{2}N-p}} \leq \frac{C_N}{a^N} \quad \text{whenever } n > \frac{2}{3}p$$

(for some $N > 0$).

Hence the result follows since if (8.6) holds for some $N_o$, then it will hold for $N \leq N_o$.

Let $C_o$ be the space of functions which are continuous on $\overline{R^k}$ the one point compactification of $R^k$, and for which $\lim_{|x| \to \infty} f(x) = 0$ endowed with the usual uniform metric. Suppose that $\Omega_n = \{h : \theta_o + \delta_n h \varepsilon (\underline{H})\} = R^k$ for all sufficiently large $n$. Then the above theorem in particular implies that $h \longrightarrow Z_n(h) \in C_0$ a.s. for all sufficiently large $n$. When $\bigcirc_n \neq R^k$ for all sufficien ly large $n$ w make the following modification. First define

$$Z_n'(h) = \begin{cases} Z_n(h) & \text{if } \theta_0+\delta_n h \in (\underline{H}) \\ 0 & \text{otherwise} \end{cases}$$

and then define

$$\overline{Z}_n(h) = \begin{cases} Z_n'(h) & \text{if } \theta_0+\delta_n h \in (\underline{H}) \\ 0 & \text{if } d(\theta_0+\delta_n h, (\overline{\underline{H}})) \geq \frac{1}{n} \\ \text{continuous in such a way that } \overline{Z}_n(h) \leq Z_n'(h), & \text{otherwise}, \end{cases}$$

where $(\overline{\underline{H}})$ is the closure of $(\underline{H})$ and $d(x,A)$ means the distance between $x$ and the set $A$ in the usual sense. Clearly, the conclusions of Theorems 1 and 2 are valid for the sequence of processes $h \longrightarrow \overline{Z}_n(h)$, $n \geq 1$. Also it is seen that, by Theorem 2 and Lemma 1, the sequence of processes $\overline{Z}_n(h)$ is uniformly equi-continuous in probability, under the assumptions of Theorem 2. In other words, for every $\varepsilon > 0$

$$\lim_{\delta \longrightarrow 0} \limsup_{n \longrightarrow \infty} P\Big[ \sup_{|h_2-h_1| \leq \delta} |\overline{Z}_n(h_2) - \overline{Z}_n(h_1)| > \varepsilon ; h_1, h_2 \in \overline{R}^k \Big] = 0 .$$

We thus have the following theorem by invoking appropriate theorems in Prakasa Rao (1975) or Straf (1972).

Theorem 3. Suppose the assumptions (A.1) - (A.10) are satisfied. Then the distribution in $C_o$ generated by the process $\bar{Z}_n(h)$ converge as $n \longrightarrow \infty$ to the distribution generated by the process

$$R(h) = \exp(h' T^{1/2}(\theta_o)\, Z - \frac{1}{2} h' T(\theta_o) h)$$

where $Z$ is a copy of the standard normal distribution $N(0,I)$ independent of $T(\theta_o)$. In particular, if $f$ is a continuous functional on $C_o$, then for all $x \in R$

$$\lim_{n \longrightarrow \infty} P[f[Z_n(h)] < x] = P[f[R(h)] < x].$$

4. CONVERGENCE OF MOMENTS

(a) maximum likelihood estimators.

Theorem 4. Suppose the assumptions (A.1) - (A.10) are satisfied. Then for any $m > 0$

$$\lim_{n \longrightarrow \infty} E[|\delta_n^{-1}(\hat{\theta}_n - \theta_o)|^m] = E[|T^{-1/2}(\theta_o) Z|^m]$$

where $\hat{\theta}_n$ is an MLE as defined in Definition 2 of Ch.7, and $Z$ is a copy of the standard k-variate normal distribution independent of $T(\theta_o)$.

Before giving the proof of this theorem we first prove the following

Lemma 2. Suppose the asumptions of Theorem 2 are satisfied. Then for any given $N > 0$ there exists an $n_o$ and a constant $C_N$ depending only on $N$ such that

$$P[\,|\delta_n^{-1}(\hat{\theta}_n - \theta_o)| > x\,] \leq \frac{C_N}{x^N}$$

for all $n \geq n_o$ and for all sufficiently large $x > 0$.

Proof. Consider, for $x > 0$,

$$P[\,|\delta_n^{-1}(\hat{\theta}_n - \theta_o)| > x\,] \leq P[\,\sup_h Z_n(h) \leq \sup_{|h|>x} Z_n(h)\,]$$

$$\leq P[\,\sup_h Z_n(h) \leq \sup_{|h|>x} Z_n(h)\,;\ \sup_{|h|>x} Z_n(h) \leq \frac{1}{x^N}\,]$$

$$+ P[\,\sup_{|h|>x} Z_n(h) > \frac{1}{x^N}\,].$$

Now note that $\sup_h Z_n(h) \geq Z_n(0) = 1$. Hence for all $x$ such that $x^{-N} < 1$ we have by Theorem 2

$$P[\,|\delta_n^{-1}(\hat{\theta}_n - \theta_o)| > x\,] \leq P[\,\sup_{|h|>x} Z_n(h) > \frac{1}{x^N}\,] < \frac{C_n}{x^N}.$$

Hence the proof of the lemma is complete.

Proof of Theorem 4. Lemma 2 in particular entails that the sequence $\{\delta_n^{-1}(\hat{\theta}_n - \theta_o)\}$ is relatively compact. Hence in view of Lemma 1, and Theorem 2 of Ch.7 it follows first that

$$\mathscr{L}(\delta_n^{-1}(\hat{\theta}_n - \theta_o)) \Longrightarrow N(0, T^{-1}(\theta_o)).$$

Secondly, it is easily seen from Lemma 2 that the sequence $\{E[\,|\delta_n^{-1}(\hat{\theta}_n - \theta_o)|^{m+1}\}$, $m > 0$, is uniformly bounded for all sufficiently large $n$ and hence the sequence $\{|\delta_n^{-1}(\hat{\theta}_n - \theta_o)|^m\}$ is uniformly integrable. Now the proof can be easily concluded from these two facts.

(b) maximum probability estimators.

Theorem 5. Suppose the assumptions (A.1) - (A.11) are satisfied. Then for any $m > 0$,

$$\lim_{n \longrightarrow \infty} E[\,|\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o)|^m\,] = E[\,|T^{-1/2}(\theta_o)Z|^m\,]$$

where $\bar{\theta}_n(a)$ is an MPE as defined in Definition 1 of Ch.7.

Before giving the proof of this theorem we shall first present some preliminary lemmas.

Lemma 3. Suppose the assumptions (A.1), (A.2) and (A.11) are satisfied. Then there are positive constants $C$ and $\varepsilon_o$ such that for all $0 < \varepsilon \leq \varepsilon_o$

$$P[\int_{|u| \leq \varepsilon} Z_n(u)\, du \leq \varepsilon] \leq C\, \varepsilon^2 .$$

Proof. Because of the assummed continuity there exists an $u^*$ which may depend on the observations, $|u^*| \leq \varepsilon$, such that

$$P[\inf_{|u| \leq \varepsilon} Z_n(u) = Z_n(u^*)] = 1 .$$

Hence

$$P[\int_{|u| \leq \varepsilon} Z_n(u)du \leq \varepsilon] \leq P[Z_n(u^*) < \tfrac{1}{2}]$$

$$= P[\,|\log Z_n(u^*)| \geq |\log (\tfrac{1}{2})|\,]$$

$$\leq P[\sup_{|u| \leq \varepsilon} |\log Z_n(u)| \geq |\log (\tfrac{1}{2})|\,] .$$

In view of the condition (A.11), the last term of the above expression can be shown to be less than or equel to $C\,\varepsilon^2$ by following the arguments similar to the proof of Lemma 1. Hence the proof of the lemma is complete.

Lemma 4. Suppose that the assumptions of Theorem 2 are satisfied. Then for any given $N > 0$ and $a \geq 0$ there exists an $n_o$ and a constant $C_N$ depending only on $N$ such that for all $n \geq n_o$

$$P\left[ \int_{|h|>M} |h|^a Z_n(h)\,dh > M^{-N} \right] \leq \frac{C_N}{M^N}, \quad M \geq 1.$$

Proof. It is enough to prove the result for $N \geq N_o$ for some $N_o > 0$, since if it is true for $N_o$ then it will be true for $N \leq N_o$. Consider, for $M \geq 1$,

$$P\left[ \int_{|h|>M} |h|^a Z_n(h)\,dh > \frac{1}{M^N} \right] \leq P\left[ \int_{|h|>M} |h|^a Z_n(h)\,dh > \frac{1}{M^{N+a+1}} \right]$$

$$\leq P\left[ \int_{|h|>M} |h|^a Z_n(h)\,dh > \sum_{k=0}^{\infty} \frac{1}{(M+k)^{N+a+2}} \right] \quad (\text{for } N \geq 2)$$

$$\leq \sum_{k=0}^{\infty} P\left[ \int_{M+k \leq |h| \leq M+k+1} |h|^a Z_n(h)\,dh > \frac{1}{(M+k)^{N+a+2}} \right]$$

$$\leq \sum_{k=0}^{\infty} P\left[ \int_{M+k \leq |h| \leq M+k+1} Z_n(h) > \frac{1}{(M+k)^{N+2}} \right]$$

$$< \sum_{k=0}^{\infty} \frac{C_N}{(M+k)^{N+2}} \quad (\text{by Theorem 2})$$

for all $n \geq n_o$, where $n_o$ and $C_N$ are as in Theorem 2. This completes the proof of the lemma.

Lemma 5. Suppose the assumptions of Theorem 2 are satisfied. Then there exist an $n_o$ and a constant $C_N$ depending only on $N$ such that for all $n \geq n_o$

$$P\left[\, |\delta_n^{-1}(\bar\theta_n(a) - \theta_o)| > x \right] \le \frac{C_N}{x^N}\ , \quad a > 0\ ,$$

for all  ufficiently large  $> 0$.

<u>Proof</u>. Consider, setting $D_a = \left\{ h \in R^k : |h| \le a \right\}$

$$P\left[\, |\delta_n^{-1}(\bar\theta_n(a) - \theta_o)| > x \right]$$

$$\le P\left[ \sup_{|u| > x} \int_{u-D_a} Z_n(h)\,dh \ge \sup_{|u| < x} \int_{u-D_a} Z_n(h)\,dh \right]$$

$$\le P\left[ \int_{|h| \ge (x-a)} Z_n(h)\,dh \ge \int_{D_a} Z_n(h)\,dh \right]$$

$$\le P\left[ \int_{|h| \ge (x-a)} Z_n(h)\,dh \ge \int_{D_a} Z_n(h)\,dh\ ; \right.$$

$$\int_{|h| \ge (x-a)} Z_n(h)\,dh \le \frac{1}{(x-a)^N}\ ;$$

$$\left. \int_{D_a} Z_n(h)\,dh \ge \frac{2^{N/2}}{x^{N/2}} \right]$$

$$+ P\left[ \int_{|h| > (x-a)} Z_n(h)\,dh > \frac{1}{(x-a)^N} \right]$$

$$+ P\left[ \int_{D_a} Z_n(h)\,dh < \frac{2^{N/2}}{x^{N/2}} \right]$$

$$= I_1 + I_2 + I_3\ , \quad \text{say} .$$

Assume in what follows that $x > \max(2a, 2)$. Then we have $\frac{1}{(x-a)^N} < \frac{2^N}{x^N} < \frac{2^{N/2}}{x^{N/2}}$ and so $I_1 = 0$.

By Lemma 4

$$I_2 \le C_N \frac{2^N}{x^N} \quad (\text{for some } C_N > 0)$$

and by Lemma 3

$$I_3 \leq \frac{2^N}{x^N} .$$

Hence

$$P[\,|\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o)| > x\,] \leq \frac{C_N(1+2^N)}{x^N} \quad \text{(for some } C_N > 0) .$$

This completes the proof of the lemma.

Proof of Theorem 5. Lemma 5 in particular implies that the sequence $\{\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o)\}$ is relatively compact. Hence in view of Theorem 1 of the present chapter and Theorem 1 of Ch.7 it follows that

$$\mathcal{L}(\delta_n^{-1}(\bar{\theta}_n(a) - \theta_o)) \Longrightarrow N(0, T^{-1}(\theta_o)) .$$

Now proceeding as in the proof of Theorem 4, the proof of the theorem is completed.

(C) Bayes estimators.

Theorem 6. Suppose that the assumptions (A.1) - (A.11) are satisfied. Further assume that

(A.12) the largest eigen value of $T^{-1}(\theta_o)$ has finite moments of all orders.

Then for every $m > 0$

$$\lim_{n \to \infty} E[\,|\delta_n^{-1}(t_n - \theta_o)|^m\,] = E[\,|T^{-1/2}(\theta_o)Z|^m\,]$$

where $\{t_n\}$ is a sequence of Bayes estimators as defined in

Definition 3 of Ch. 7, with respect to the loss functions $|\delta_n^{-1}(\theta - \emptyset)|^a, a \geq 1$.

Before giving the proof of this theorem we present some preliminary lemmas. We shall suppose for brevity that $\pi(\theta) \equiv 1$ since the passage to the general case causes no difficulties.

Lemma 6. Let $a \geq 1$. Set

$$D_n(H) = \frac{Z_n(h)}{\int Z_n(h)\,dh} - \frac{|\det T(\theta_o)|^{1/2}}{(2\pi)^{k/2}} \exp\left[-\frac{1}{2}|T^{1/2}(\theta_o)(h-h_n^*)|^2\right]$$

where $h_n^* = \delta_n^{-1}(\hat{\theta}_n - \theta_o)$. Suppose that the assumptions of Theorem 6 are satisfied. Then for any $N > 0$ there exists an $n_o$ and a constant $C_N$ depending only on $N$ such that for all $n \geq n_o$

$$P\left[\int_{|h|>M} |h|^a |D_n(h)|\,dh > M^{-N}\right] \leq C_N M^{-N}, \quad M \geq 1.$$

Proof.
$$P\left[\sup_{|h|>M} \frac{Z_n(h)}{\int Z_n(h)\,dh} > \frac{1}{2M^N}\right] \leq P\left[\int_{|h|>M} Z_n(h) > M^{-\frac{3}{2}N}\right] + P\left[\int Z_n(h)\,dh < 2M^{-\frac{1}{2}N}\right] \leq C_N M^{-N}$$

by Theorem 2 and Lemma 3, for all sufficiently large $n$. Hence from the arguments of Lemma 4 it follows that for all sufficiently large $n$

$$P\left\{ \frac{\int_{|h|>M} |h|^a Z_n(h)\,dh}{\int Z_n(h)\,dh} > \frac{1}{2M^N} \right\} \le \frac{C_N}{M^N} \text{ (for some } C_N > 0\text{)} .$$

Denote the eigen values of $T(\theta_o)$ by $\lambda_1 < \ldots < \lambda_k$. Now $|h_n^*| \le M/2$ and $\frac{1}{\lambda_1} < M$ implies that

$$\frac{|\det T(\theta_o)|^{1/2}}{(2\pi)^{k/2}} \int_{|h|>M} |h|^a \exp\left[-\frac{1}{2}|T^{1/2}(\theta_o)(h-h_n^*)|^2\right] dh$$

$$\le \frac{K|\det T(\theta_o)|^{1/2}}{(2\pi)^{k/2}} \int_{|h-h_n^*|>\frac{M}{2}} |h-h_n^*|^a \exp\left(-\frac{1}{2}|T^{1/2}(\theta_o)(h-h_n^*)|^2\right] dh$$

(for some $k > 0$)

$$= \frac{k(\lambda_1 \ldots \lambda_n)^{1/2}}{(2\pi)^{k/2}} \int_{|h|>\frac{M}{2}} (\Sigma h_i^2)^{a/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} \lambda_i h_i^2\right) dh$$

$$\le \frac{(\lambda_1 \ldots \lambda_k)^{1/2}}{(2\pi)^{k/2}} \frac{K 2^{4N}}{M^{4N}} \int \left(\sum_{i=1}^{k} h_i^2\right)^{\frac{a}{2}+2N} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} \lambda_i h_i^2\right) dh$$

where $h_i$'s are the components of the vector $h$. In what follows we assume without loss of generality that a is an integer. Now using the fact that

$$\frac{(\lambda_1 \ldots \lambda_k)^{1/2}}{(2\pi)^{k/2}} \int \prod_{i=1}^{k} (\lambda_i h_i)^{\ell_i} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} \lambda_i h_i^2\right) dh < \infty$$

for every $\ell_i > 0$, $i = 1, 2, \ldots, k$, we see that the above integral is bounded by

$$\frac{K2^{4N}}{M^{4N}} \left( \sum_{i=1}^{k} \frac{1}{\lambda_i} \right)^{\frac{a}{2}+2N} \quad \text{(for some } K > 0)$$

$$\leq \frac{K2^{4N}}{M^{4N-\frac{a}{2}-N}} \leq \frac{K2^{N}}{M^{N}}$$

for some $K > 0$ and for all $N \geq 2$, since $\frac{1}{\lambda_1} < M$. Therefore for every $N > 0$

$$P\left[\frac{|\det T(\theta_o)|^{1/2}}{(2\pi)^{k/2}} \int_{|h|>M} |h|^a \exp\left(-\frac{1}{2}|T^{1/2}(\theta_o)(h-h_n^*)|^2\right) dh > \frac{1}{2M^N}\right]$$

$$\leq P\left[|h_n^*| > \frac{M}{2}\right] + P\left[\frac{1}{\lambda_1} > M\right] \leq \frac{C_N}{M^N} \quad \text{(for some } C_N)$$

for all sufficiently large $n$, by Theorem 4 and the given assumption (A.12).

Lemma 7. Let $a \geq 1$. Set

$$D_n^*(h) = f_n^*(\theta_n+\delta_n h) - \frac{|T(\theta_o)|^{1/2}}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2} h' T(\theta_o) h\right),$$

where

$$f_n^*(\theta_n+\delta_n h) = \frac{\prod_{j=1}^{n} f_j(\theta_n+\delta_n h)}{\int \prod_{j=1}^{n} f_j(\theta_n+\delta_n h)\, dh}.$$

Suppose the assumptions of Theorem 6 are satisfied. Then for any given $N > 0$, there exist an $n_o$ and constants $C_N^{(1)}$ and $C_N^{(2)}$ depending only on $N$ such that for all $n \geq n_o$ and $M \geq 1$

$$P\Big[\int_{|h|\geq M} |h|^a |D_n^*(h)|\,dh > C_N^{(1)}\ M^{-N}\Big] \leq C_N^{(2)}\ M^{-N}.$$

Proof. Let $g_n = h + h_n^*$. Then

$$|h| \leq d_a |g_n|^a + d_a |h_n^*|^a$$

where $d_a = 2^{a-1}$. Using this inequality we have

$$\int_{|h|>M} |h|^a |D_n^*(h)|\,dh \leq d_a \int_{|g-h_n^*|>M} |g|^a |D_n(g)|\,dg$$

$$+ d_a |h_n^*|^a \int_{|g-h_n^*|>M} |D_n(g)|\,dg\ ,$$

where $D_n(g)$ is as defined in Lemma 6. Consider

$$P\Big[\int_{|g-h_n^*|>M} |g|^a |D_n(g)|\,dg > 2^N M^{-N}\Big]$$

$$\leq P\Big[\int_{|g-h_n^*|>M} |g|^a |D_n(g)|\,dg > 2^N M^{-N};\ |h_n^*| \leq \tfrac{M}{2}\Big] + P\Big[|h_n^*| > \tfrac{M}{2}\Big]$$

$$\leq P\Big[\int_{|g|>\frac{M}{2}} |g|^a |D_n(g)|\,dg > 2^N M^{-N}\Big] + P\Big[|h_n^*| > \tfrac{M}{2}\Big]$$

$$\leq C_N^{(2)}\ M^{-N} \quad (\text{for some } C_N^{(2)} > 0) \qquad (8.7)$$

by the previous lemma 6 and Theorem 4. Similarly it can be shown that, for some $C_N^{(1)} > 0$ and $C_N^{(2)} > 0$,

$$P\Big[|h_n^*|^a \int_{|g-h_n^*|\geq M} |D_n(h)|\,dh > C_N^{(1)}\ M^{-N}\Big] \leq C_N^{(2)}\ M^{-N}. \qquad (8.8)$$

Hence the result follows from (8.7) and (8.8).

<u>Lemma 8</u>. Suppose that the assumptions of Theorem 6 are satisfied. Then for every $N > 0$, there exist an $n_o$ and a constant $C_N$ depending only on $N$ such that

$$P[\,|\delta_n^{-1}(t_n - \theta_o)| > M\,] \leq C_N\, M^{-N}$$

for all $n \geq 0$ and $M \geq 1$.

<u>Proof</u>. Since $t_n$ is a Bayes estimator with respect to the loss function $|\delta_n^{-1}(\theta - \emptyset)|^a$, $a \geq 1$, we have, putting $u_n = \delta_n^{-1}(t_n - \hat{\theta}_n)$,

$$\int |h|^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh \geq \int |h + u_n|^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh$$

$$\geq \int_{|h| \leq \frac{M}{4}} |h + u_n|^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh$$

Hence, since $|h| \leq M/4$ and $|u_n| > M$ implies that $|h + u_n| \geq |h| + M/2$,

$$P[\,|\delta_n^{-1}(t_n - \hat{\theta}_n)| > M\,]$$

$$\leq P\left\{ \int |h|^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh \geq \int_{|h| \leq \frac{M}{4}} \left(|h| + \frac{M}{2}\right)^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh \right\}$$

$$\leq P\left\{ \int |h|^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh \geq \int_{|h| \leq \frac{M}{4}} |h|^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh + K \int_{|h| \leq \frac{M}{4}} f_n^*(\hat{\theta}_n + \delta_n h)\,dh \right\}$$

(for some $K > 0$)

$$\leq P\left\{ \int_{|h| > \frac{M}{4}} |h|^a f_n^*(\hat{\theta}_n + \delta_n h)\,dh \geq K \int_{|h| \leq \frac{M}{4}} f_n^*(\hat{\theta}_n + \delta_n h)\,dh \right\}. \tag{8.9}$$

Now note that

$$\int f_n^*(\theta_n+\delta_n h)\,dh = 1 = \frac{|\det T(\theta_o)|^{1/2}}{(2\pi)^{1/2}}\int \exp[-\tfrac{1}{2}h'T(\theta_o)h]\,dh .$$

Hence by the arguments of the proof of the lemma 6 and by the previous lemma 7, for any given $N>0$, there exist constants $C_N^{(1)}$ and $C_N^{(2)}$ and an $n_o$ depending only on $N$ such that for all $n\geq n_o$

$$P\left[\int_{|h|>M/4}|h|^a f_n^*(\theta_n+\delta_n h)\,dh > C_N^{(1)}\ M^{-N}\right]\leq C_N^{(2)}\ M^{-N}$$

and

$$P\left\{\left|\int_{|h|\leq M/4} f_n^*(\theta_n+\delta_n h)\,dh - \frac{|\det T(\theta_o)|^{1/2}}{(2\pi)^{k/2}}\int_{|h|\leq M/4}\exp[-\tfrac{1}{2}h'T(\theta_o)h]\,dh\right|\right.$$
$$\left.\geq C_N^{(1)}\ M^{-N}\right\}\leq C_N^{(2)}\ M^{-N} .$$

Hence it is easily seen that the last term of the above inequality (8.9) is less than or equel to $2C_N^{(2)}\ M^{-N}$ for all $n\geq n_o$, that is,

$$P[\delta_n^{-1}(t_n-\hat{\theta}_n)|>M]\leq 2C_N^{(2)}\ M^{-N} \qquad (8.10)$$

for all $n\geq n_o$. Now

$$P[|\delta_n^{-1}(t_n-\theta_o)|>M]\leq P[|\delta_n^{-1}(t_n-\hat{\theta}_n)|>M/2]$$
$$+P[|\delta_n^{-1}(\hat{\theta}_n-\theta_o)|>M/2]$$
$$\leq 2^{N+1}C_N^{(2)}\ M^{-N}+2^N M^{-N}E[|\delta_n^{-1}(\hat{\theta}_n-\theta_o)|^N]$$

by (8.10). By Theorem 4, the proof of the lemma is complete.

Proof of Theorem 6. First note that the lemma 4 implies that for every $\varepsilon > 0$ and $a \geq 1$

$$\lim_{\alpha \to \infty} \limsup_{n \to \infty} P\left[ \int_{|h| > \alpha} |h|^a Z_n(h)\,dh \geq \varepsilon \right] = 0 .$$

Hence by Theorem 1 of the present chapter and Theorem 4 of Ch.7 it follows that

$$\mathcal{L}(\delta_n^{-1}(t_n - \theta_o)) \Longrightarrow N(0, T^{-1}(\theta_o)) .$$

Now proceeding as in the proof of Theorem 4, the proof of the theorem is completed.

## 5. DISCUSSIONS ON THE ASSUMPTION (A.10) AND SOME EXAMPLES

Consider a class of mixed Gaussian processes having the following form

$$Z_n(h) = \exp(h' U_n(\theta_o) - \tfrac{1}{2} h' T_n(\theta_o) h) \qquad (8.11)$$

where, for every $n \geq 1$, $U_n$ is a random k-vector and $T_n(\theta_o)$ is a p.d. random $k \times k$ matrix. We further assume that the moments of any order of the largest eigen value of the matrices $T_n^{-1}(\theta_o)$ are uniformly bounded for all large $n$. Let $\lambda_n$ be the smallest eigen value of $T_n(\theta_o)$. In order to verify (A.10) it is enough to show that for every $N > 0$ there exists an $n_o$ and $C_N$ depending only on $N$ such that

$$P\left[ Z_n^{1/2}(h) > |h|^{-N} ; \lambda_n^{-1} < |h| \right] \leq C_N |h|^{-N}$$

since we have assumed that for every $m > 0$, $\sup_{n \geq n_o} E(\lambda_n^{-m}) < \infty$ for some $n_o$. Now

$$P[Z_n^{1/2}(h) > |h|^{-N};\ \lambda_n^{-1} < |h|]$$

$$\leq |h|^N E\{I(\lambda_n^{-1} < |h|) Z_n^{1/2}(h)\}$$

$$\leq |h|^N E\{I(\lambda_n^{-1} < |h|) Z_n(h/2) \exp(-\frac{\lambda_n}{8}|h|^2)\}$$

$$\leq |h|^N E\{I(\lambda_n^{-1} < |h|) Z_n(h/2) K_N \lambda_n^{-2N} |h|^{-4N}\}$$

(for some $K_n > 0$)

$$\leq |h|^{3N} |h|^{-4N} K_N E[Z_n(h/2)]$$

$$\leq K_N |h|^{-N} \quad (\text{since } E[Z_n(h/2)] = 1).$$

Thus we see that the assumption (A.10) is satisfied in this case.

In some cases it is possible to verify that

$$\sum_{j=1}^{n} \inf_{\underline{X}_{j-1}} \int [f_j^{1/2}(\theta_o + \delta_n h) - f^{1/2}(\theta_o)]^2 \, d\mu_j \geq C|h|^2 \qquad (8.12)$$

for some $C > 0$ and for all sufficiently large $n$, where $\underline{X}_j = (X_1, X_2, \ldots, X_j)$. Then putting

$$a_j = \inf_{\underline{X}_{j-1}} \int [f_j^{1/2}(\theta_o + \delta_n h) - f^{1/2}(\theta_o)]^2 \, d\mu_j ,$$

$$E[Z_n^{1/2}(h)] \leq \prod_{j=1}^{n} (1 - \frac{a_j}{2}) .$$

Note that $|a_i| \leq 1$, since

$$0 \leq \sup_{\underline{X}_{j-1}} \int f_j^{1/2}(\theta_o + \delta_n h) f_j^{1/2}(\theta_o) \, d\mu_j \leq 1 .$$

Hence

$$E[Z_n^{1/2}(h)] \leq \exp[-\frac{1}{2}\sum_{j=1}^{n} a_j] \leq \exp[-C|h|^2]$$

for all sufficiently large $n$, when (8.12) holds.

A situation where (8.12) can be easily verified, with $\delta_n'\delta_n = n^{-1} I$, is the following example.

<u>Example</u>. Let $X_0, X_1, \ldots,$ be a sequence of Markov chain for which the state space consist of the numbers 0 and 1 ; the transition matrix is

| $X_0$ \ $X_1$ | 0 | 1 |
|---|---|---|
| 0 | $(1-p)+\pi p$ | $(1-\pi)p$ |
| 1 | $(1-\pi)(1-p)$ | $\pi+(1-\pi)p$ |

and the initial distribution is $f(1,p,\pi) = 1 - f(0,p,\pi) = p$ where $(p,\pi) \in \underline{\bar{H}} = (0,1)\times(0,1)$.

We finally point out that the assumptions (A.1) - (A.11) can be considerably simplified when the observations are i.i.d.

## REFERENCES


Aldous, D.J. and Eagleson, G.K. (1978). On mixing and stability of limit theorems. Ann. Probability, Vol.6, No.2, 325-331.

Anderson, T.W. (1955). The integral of a symmetric unimodel function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. 6, 170-176.

Bahadur (1964). On Fisher's lower bound for asymptotic variances. Ann. Math. Statist.

Bhat, B.R. (1978). Some problems of estimation for dependent observations. Tech. Report, University of Poona, India.

Bhat, B.R. and Prasad, M.S. (1978). W-efficiency of maximum likelihood estimates. Tech. Report, University of Poona.

Basawa, I.V. and Scott, D.J. (1977). Efficient tests for branching processes, Biometrica, 63 531-536.

__________ (1977). Efficient tests for stochastic processes, Sankhyā Ser. A, 39, 21-31.

__________ (1979). Efficient estimation for stochastic processes. J. Indian Statist. Assoc. To appear.

Basawa, I.V. and Koul, H.L. (1979). Asymptotic tests of composite hypothesis for non-ergodic type stochastic processes. Stoch. Proc. Appl. 9, 291-305.

Bickel, P. and Yahav, J. (1969). Some contributions to the asymptotic theory of Bayes solutions. Zeit. f. Wahrscheinlichkeitstheorie u. v. Gebiete, 11, 257-279.

Borwanker, J.D., Kallianpur, G. and Prakasa Rao, B.L.S. (1971). The Berstein-von Mises theorem for Markov processes. Ann. Math. Statist. 42, 1241-1253.

Davies, R.B. (1978). Asymptotic inference when the amount of information is random. Tech. Report, Wellington Univ., New Zealand.

Feigin, P.D. (1978). The efficiency criteria problem for stochastic processes, Stoch. Proc. Appl., 6, 115-127.

Ghosh, J.K., Joshi, S.W. and Sinha, B.K. (1980). On the asymptotic expansions for posterior distributions. Tech. Report, Indian Statistical Institute, Calcutta.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. Zeit. f. Wahrscheinlichkeitsthorie u.v. Gabiete, Vol. 14, 323-330.

__________ (1971). Limiting properties of likelihoods and inference. Foundations of Statistical Inference. Godambe and Sprott, editors, Holt, Rinehart and Winston, Toronto.

__________ (1972). Local asymptotic minimax and admissibility in estimation Proc. 6th Berkeley Symp. Vol. 7, 175-194.

Hall, P.G. (1977). Martingale invariance principles. Ann. Probability, 5, 875-887.

Hewitt, E. and Stromberg, K. (1965). Real and abstract analysis. Springer-Verlog.

Heyde, C.C. and Feigin, P.D. (1975). On efficiency and exponential families in stochastic process estimation. Statistical Distributions in Scientific Work. Vol. 1. Patil, G.P., Kotz, S. and Ord, J.K., editors, Reidel, Dordrecht, 227-240.

Heyde, C.C. (1978). On an optimal property of maximum likelihood estimator of a parameter from a stochastic process. Stoch. Proc. Appl. 8, 1-9.

Ibragimov, I.A. and Khasminskii, R.Z. (1972 and 1973). Asymptotic behaviour of some estimators in the smooth case, I and II. Theory Prob. Applications. 17, 445-462; 18, 76-91.

__________ (1975). Local asymptotic normality for non-identically distributed observations. Theory Prob. Applications. 20.

__________ (1975). Properties of maximum likelihood and Bayes estimators for non-identically distributed observations. Theory Prob. Applications. 20, 689-637.

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property. Ann. Inst. Statist. Math. 22, 1-13.

Inagaki, N. and Ogata, Y. (1975). The weak convergence of likelihood ratio random fields and its applications. Ann. Inst. Statist. Math., 27, 391-419.

Inagaki, N. and Ogata, Y. (1977). The weak convergence of the likelihood ratio random fields for Markov observations. Ann. Inst. Statist. Math. 29, 165-187.

Jeganathan, P. (1979a). On the asymptotic theory of statistical estimation when the limit of the log-likelihood ratios is mixed normal. Sankhyā (to appear).

__________ (1979b). On the convergence of moments of statistical estimators. (Submitted for publication)

__________ (1979c). On a decomposition of the limit distribution of a sequence of estimators. Sankhyā (to appear).

__________ (1979d). An extension of a result of L. LeCam concerning asymptotic normality. Sankhyā (to appear).

__________ (1980). Some asymptotic properties of risk functions in estimation when the limit of the experiments is a mixed normal. Sankhyā (to appear).

Keiding, N. (1975). Maximum likelihood estimation in the birth and death process. Ann. Statist. 3, 366-372.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood and Bayes estimates. Univ. Clif. Publ. Statist. 1, 277-330.

__________ (1958). Les Proprietes asymptotiques des solutions de Bayes. Pub. Inst. Statist. Univ. Paris, 8, 17-35.

__________ (1960). Locally asymptotically normal families of distributions. Univ. Calif. Publ. Statist. 3, 37-98.

LeCam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. Ann. Math.Statist. 41, 802-828.

__________ (1972). Limits of experiments. Proc. 6th Berkeley Symp. Vol.I, 245-61.

__________ (1973). Sur les contraints imposees par les passages la limite usuels en statistique. I.S.I. Congress, Vienne.

__________ (1974a). Notes on asymptotic methods in Statistical decision theory. Centre de Recherches Mathematiques, Universite de Montreal.

__________ (1974b). On some results of J. Hájek concerning asymptotic normality. (Pre-print).

__________ (1979). Asymptotic methods in Statistical decision theory. Manuscript for book in preparation.

Levit, V. Ya (1974). On the behaviour of generalised Bayes estimates for Markov observations. Theory Prob. Applications, 19, 382-346.

McLeish, D.L. (1974). Dependent central limit theorems and invariance principles. Ann. Probability, 2, 620-628.

Parthasarathy, K.R. (1977). Introduction to probability and measure. Macmillan Company of India Limited.

Pfanzagl, J. (1970). On the asymptotic efficiency of median unbiased estimates. Ann. Math. Stat. 41, 1550-1559.

Phillippou, A.N. and Roussas, G.G. (1974). Asymptotic distribution of the log-likelihood function in the independent not identically distributed case. Ann. Statist. 1, 545-571.

Prakasa Rao, B.L.S. (1968). Estimation of the location of the cusp of a continuous density. Ann. Math. Statist., 39, 76-87.

__________ (1974). Statistical inference for stochastic processes. Tech. Report. CRM 465, Centre de Recherches Mathematiques, Universite de Montreal.

__________ (1975). Tightness of probability measures generated by stochastic processes on metric spaces. Bull. Inst. Math. Academic. Sinica, 3, 353-367.

Prokorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. Theory Prob. Applications, 1, 157-214.

Roussas, G.G. (1972). Contiguity of probability measures. Cambridge University Press.

__________ (1979). Asymptotic distribution of the log-likelihood function for stochastic processes. Zeit. f. Wahrscheinlichkeitstheoric u.v. Gabiete. 47, 31-46.

Straf, M.L. (1970). Weak convergence of random processes with several parameters. Proc. 6th Berkeley Symp. Math. Statist. Prob. 2, 187-221.

Strasser, H.L. (1978). Global asymptotic properties of risk functions in estimation. Zeit. f. Wahrscheinlichkeitstheoric u.v. Gabiete. 45, 35-48.

Swensen, A.R. (1980). Asymptotic inference for a class of stochastic processes. Ph.D. thesis. Univ. Calif., Berkeley.

Wald, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc., 54, 426-482.

Weiss, L. and Wolfowitz, J. (1974). Maximum Probability Estimators and Related Topics. (Lecture Notes in Mathematics), Springer-Verlog.