SOME CONTRIBUTIONS TO THE ASYMPTOTIC THEORY OF ESTIMATION IN NON-REGULAR CASE

TAPAS SAMANTA

INDIAN STATISTICAL INSTITUTE
CALCUTTA
1988

SOME CONTRIBUTIONS TO THE ASYMPTOTIC THEORY OF ESTIMATION IN NON-REGULAR CASE

Tapas Samanta

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements
for the award of the degree of
DOCTOR OF PHILOSOPHY
Calcutta
1988

ACKNOWLEDGEMENT

I express my sincere gratitude and thanks to Sri Kamal Kumar Roy, under whose supervision and careful guidance I have carried out the present research work. He took a lot of pains in going through the details of the progress at each stage of this work.

Next, I want to record my indebtedness to Professor J.K. Ghosh who introduced me to this area ; the initial direction about the basic approach to these problems was received from him.

Dr. Tapas Kumar Chandra was always eager and enthusiastic about the progress of my work. He used to encourage me at different stages of the work ; I am grateful to him. Special mention must also be made about my indebtedness to an anonymous referee of Sankhyā for his valuable comments and suggestions towards possible extension of my earlier works.

I am thankful to my teachers and co-fellows of the Stat-Math. Division, who have always expressed their interest in the progress of my work. I am particularly grateful to Professor B.V. Rao for clearing many of my doubts at the early stage of this work.

I am grateful to the Indian Statistical Institute for providing me with all facilities necessary for carrying out my work. I thank Subrata Kumar Roy for his cooperation. Babulal Seal and Rabindranath Mukhopadhyay provided valuable service towards proof reading of the typed script. I am thankful to Samir Kumar Chakraborty for excellent typing ; his neat dealing of this painful task deserves special mention. Thanks are also due to Muktalal Khanra who provided the final touch.

Tapas Samanta

CONTENTS

		page
	LIST OF NOTATION	
	INTRODUCTION	(i)-(ix)
CHAPTER 1	LOCAL ASYMPTOTIC MINIMAX ESTIMATION UNDER AN ASYMPTOTIC EXPANSION OF LIKELIHOOD RATIO	1-24
1.1	Introduction	1
1.2	Convergence of experiments assuming asymptotic expansions	2
1.3	Lower bound for asymptotic risk and an efficient estimator	6
1.4	A convolution theorem	17
1.5	Statement of results under condition (A2)	22
CHAPTER 2	EXAMPLES OF NON-REGULAR CASES	25-46
2.1	Introduction	25
2.2	Independent and identically distributed observations	25
2.3	Regression type model	30
2.4	Asymptotic properties of maximum probability estimators	33
2.5	Asymptotic properties of Bayes estimators for regression type model	35
CHAPTER 3	THE ASYMPTOTIC BEHAVIOUR OF POSTERIOR DISTRIBUTIONS AND BAYES ESTIMATORS IN NON-REGULAR CASES	47-70
3.1	Introduction	47
3.2	Limit of posterior distributions in non-regular cases	48

CONTENTS

		page
	LIST OF NOTATION	
	INTRODUCTION	(i)-(ix)
CHAPTER 1	LOCAL ASYMPTOTIC MINIMAX ESTIMATION UNDER AN ASYMPTOTIC EXPANSION OF LIKELIHOOD RATIO	1-24
1.1	Introduction	1
1.2	Convergence of experiments assuming asymptotic expansions	2
1.3	Lower bound for asymptotic risk and an efficient estimator	6
1.4	A convolution theorem	17
1.5	Statement of results under condition (A2)	22
CHAPTER 2	EXAMPLES OF NON-REGULAR CASES	25-46
2.1	Introduction	25
2.2	Independent and identically distributed observations	25
2.3	Regression type model	30
2.4	Asymptotic properties of maximum probability estimators	33
2.5	Asymptotic properties of Bayes estimators for regression type model	35
CHAPTER 3	THE ASYMPTOTIC BEHAVIOUR OF POSTERIOR DISTRIBUTIONS AND BAYES ESTIMATORS IN NON-REGULAR CASES	47-70
3.1	Introduction	47
3.2	Limit of posterior distributions in non-regular cases	48

LIST OF NOTATION

$\mathbb{R}$	the real line
$\mathbb{R}^+$	positive half of the real line $([0,\infty))$
$\mathcal{B}$	Borel σ-field on $\mathbb{R}$
$(\underline{\underline{X}}, \underline{\underline{A}})$ or $(S, \underline{\underline{S}})$	measurable space
$\{(\underline{\underline{X}}, \underline{\underline{A}}), P_\theta, \theta \in \Theta\}$	statistical experiment where $(\underline{\underline{X}}, \underline{\underline{A}})$ is a measurable space and $\{P_\theta, \theta \in \Theta\}$ is a collection of probability measures
$\mathcal{L}\{X\}$	distribution of a random element X
$\mathcal{L}\{X \mid P\}$	distribution of a random element X with respect to the measure P
$X_n \xrightarrow{P_n} X$	convergence of X_n in P_n- probability to X
$X_n \xrightarrow{\mathcal{L}} X$	convergence of X_n in distribution to X
$\Rightarrow$	weak convergence of probability measures
$\underline{\lim}$	limit inferior
$\overline{\lim}$	limit superior
$X \sim P$	the random element X follows distribution P
$\frac{dQ}{dP}$	Radon Nykodym derivative of the absolutely continuous component of Q with respect to P
$P * Q$	convolution of measures P and Q
F_X	distribution function of a random variable X

$E_\theta X$	expectation of X when θ is the value of the parameter
a.s. P	almost surely with respect to P
a.e. P	almost everywhere with respect to P
$X_n = O_p(a_n)$	$\{a_n^{-1} X_n\}$, $n \geq 1$ is bounded in probability
$L_1(\mu)$	space of all functions absolutely integrable with respect to measure μ
$\Vert \cdot \Vert$	norm in appropriate Banach space
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$\varphi(v,t)$	density of $N(0,v)$
1_A	indicator function of the set A
A^c	complement of A
$x \in A$	x belongs to A
$x \notin A$	x does not belong to A
///	end of a proof

INTRODUCTION

The introduction of the concept of asymptotic efficiency of statistical estimators in connection with proposing and developing the method of maximum likelihood by R.A. Fisher (Fisher (1922, 1925)) is really the starting point of the asymptotic theory of estimation. Historically, however, Laplace (1774) and Gauss (1809) had made two different studies earlier than Fisher both connected with asymptotic theory of estimation. Fisher considered only consistent asymptotically normal estimators and measured the asymptotic performance of an estimator by its asymptotic variance. Thus, a consistent asymptotically normal estimator with least possible asymptotic variance was defined to be an efficient estimator. Fisher also claimed to have proved that under certain regularity conditions the maximum likelihood estimator (MLE) is efficient in the above sense. In the thirties and forties several authors (Dugue (1936a, 1936b, 1937), Wilks (1938), Neyman (1949) and others) attempted to obtain a rigorous proof of the efficiency of the MLE and there was a general belief that there exists an efficient estimator in the general case which may be obtained by the method of maximum likelihood. This belief existed until J.L. Hodges produced in 1951 the " revolutionary " examples of " super efficient " estimators (one can see, for example, Ghosh (1985) or Le Cam (1953) where it first appeared). Hodges' example shows that in the usual regular cases there exist asymptotically normal estimators whose asymptotic variances are always less than or equal to that of the MLE and are strictly less than that of the MLE at particular values of the parameter and at these particular values the asymptotic variance may even be made equal to zero. Thus, the MLE is not efficient in the above sense and indeed, within the class of all asymptotically normal estimators no estimator with minimal asymptotic variance exists.

However, the ideas of Fisher and the existence of " super efficient " estimators greatly influenced the development of the theory of efficient estimation and a modern approach to the theory of asymptotic efficient estimation emerged in the fundamental paper of Le Cam (1953). The theory was further developed in the works of Le Cam (1960, 1964, 1972), Hajek (1970, 1972), Wolfowitz (1965), Millar (1983) and others. This approach which reached more or less its final form in the papers of Hajek (1972) and Le Cam (1972) considers all estimators in stead of restricting to the class of asymptotically normal estimators only, but the efficiency or the performance of these estimators is measured in a slightly different way. Millar (1983) presents a very clear exposition of this theory extending some of the basic results of Le Cam (1972). A lucid account of this development of Hajek - Le Cam theory of efficient estimation is available in Ghosh (1985).

For ease of exposition, let us consider the case where we have independent and identically distributed (i.i.d.) observations $X_1, X_2, \dots, X_n$ with a common distribution P_θ , θ being a real parameter. Let $K_n (\uparrow \infty)$ be the normalizing factor for the given family of distributions. A formal definition of a normalizing factor is given, for example, in Weiss and Wolfowitz (1974, p.13). Roughly speaking, " This means that the best that any estimator $T_n = T_n(X_1, \dots, X_n)$ can estimate θ is to within $O_p(K_n^{-1})$"

(Weiss and Wolfowitz (1974)). For example, in the regular cases $K_n = \sqrt{n}$. A natural measure of the asymptotic performance of a sequence of estimator T_n is given by

$$\lim_{n \to \infty} E_\theta L\left[K_n(T_n - \theta)\right] \tag{1}$$

where L is an appropriate loss function. In stead of (1), Hajek (1972) considered the local asymptotic risk (a smooth version of (1)) :

$$R(\theta , \{T_n\}) = \lim_{\delta \to 0} \ \varliminf_{n \to \infty} \ \sup_{|\theta' - \theta| < \delta} E_{\theta'} L\left[K_n(T_n - \theta')\right] \tag{2}$$

as a measure of the performance of $\{T_n\}$ at θ and obtained (under regularity conditions) a lower bound to this measure in the class of all estimators. Following Fabian and Hannan (1982) we consider a variant of the measure (2) :

$$\rho(\theta , \{T_n\}) = \lim_{A \to \infty} \ \varliminf_{n \to \infty} \ \sup_{|\theta' - \theta| \leq AK_n^{-1}} E_{\theta'} L\left[K_n(T_n - \theta')\right] \tag{3}$$

Thus an estimator T_n for which the local asymptotic risk $\rho(\theta , \{T_n\})$ (with $\varliminf$ replaced by lim) is equal to the local asymptotic minimax risk

$$\rho(\theta) = \lim_{A \to \infty} \ \lim_{n \to \infty} \ \inf_{T_n} \ \sup_{|\theta' - \theta| \leq AK_n^{-1}} E_{\theta'} L\left[K_n(T_n - \theta')\right]$$

may be considered as an efficient estimator.

Hajek's lower bound for $R(\theta, \{T_n\})$ is given by

$$R(\theta, \{T_n\}) \geq E\, L(X) \tag{4}$$

where X is a random variable with distribution $N(0, I^{-1}(\theta))$, $I(\theta)$ being Fisher's information at θ . Following the arguments in his proof one can indeed prove the sharper inequality

$$\rho(\theta, \{T_n\}) \geq \rho(\theta) \geq E\, L(X) \tag{5}$$

In this connection one can see the remarks following the proofs of (4) given in Hajek (1972) and Ibragimov and Hasminskii (1981). Fabian and Hannan (1982) had shown that the bound in (4) is not sharp and may not be attainable and therefore it seems natural to consider the measure $\rho(\theta, \{T_n\})$ in stead of $R(\theta, \{T_n\})$.

Hajek's inequality (4) was proved under the assumption of asymptotic normality of the log likelihood ratio. According to Hajek, " The statistical essence of regularity consists in the possibility of replacing the family of distributions by a normal family in a local asymptotic sense." This notion of regularity known as local asymptotic normality (LAN) was developed in the papers of Le Cam (1953, 1956, 1960). A family of distributions $\{P_\theta\}$ (or rather $\{P_\theta^n\}$, $n \geq 1$, P_θ^n being the n fold product of P_θ for i.i.d. case) is said to satisfy the LAN condition at some particular value θ if it admits the following local asymptotic expansion of the likelihood ratio :

$$\frac{d P^n_{\theta + u n^{-1/2}}}{d P^n_\theta} = \exp\left\{ u \, \Delta_n(\theta) - \frac{1}{2} u^2 I(\theta) + \varepsilon_n(u, \theta) \right\} , \quad u \in \mathbb{R}$$

where $I(\theta)$ is a positive finite number and Δ_n, ε_n are random variables such that

$$\mathcal{L}\left\{ \Delta_n(\theta) \mid P^n_\theta \right\} \Rightarrow N(0, I(\theta))$$

and

$$\varepsilon_n \xrightarrow{P^n_\theta} 0 .$$

It is important to note that in all these investigations the asymptotic properties of the likelihood ratio (in the neighbourhood of the true parameter point) play a very crucial role.

In 1972, Lucien Le Cam developed the concept of limiting experiments. The definition of limits of experiments essentially uses the notion of standard measures and comparison of experiments introduced by Blackwell (1951). The idea is to approximate a statistical experiment $E = \{P_\lambda, \lambda \varepsilon \Lambda\}$ by a simpler, known and mathematically tractable experiment $F = \{Q_\lambda, \lambda \varepsilon \Lambda\}$ so that we can solve the problem in F and use this solution to solve the problem in E. We can use this notion of limiting experiment to obtain a lower bound to the local asymptotic minimax risk. A nice account of this approach is given in Millar (1983). Theorem III.1.1 in Millar (1983) which is referred to as Hajek-Le Cam asymptotic minimax theorem states that if we have a sequence of experiments E^n converging to some experiment E, then the limit of the minimax risk for experiment E^n is greater than or equal to the minimax risk of the limiting experiment (a formal statement of this result is given in Section 1.3). Now the quantity

$$\inf_{T_n} \sup_{|\theta' - \theta| \leq A K_n^{-1}} E_{\theta'} L\left[K_n(T_n - \theta')\right]$$

can be expressed as the minimax risk for the experiment $E_n = \left\{P^n_{\theta + \lambda K_n^{-1}} : |\lambda| \leq A\right\}$, $n \geq 1$. If now one can show that this sequence of experiments converge to some simple known experiment $E_A = \left\{Q_{\lambda,\theta} : |\lambda| \leq A\right\}$, then the Hajek-Le Cam asymptotic minimax theorem gives a lower bound to the local asymptotic minimax risk $\rho(\theta)$ which is obtained by computing the minimax risk for the experiment $E = \left\{Q_{\lambda,\theta} : \lambda \varepsilon \mathbb{R}\right\}$ (one must verify that the limit of the minimax risk for E_A, as $A \longrightarrow \infty$, is the minimax risk for E. Millar (1983, p. 147-148) gives an argument).

In the regular cases, the limiting experiment is the Gaussian shift experiment $\left\{ N(\lambda, I^{-1}(\theta)) : \lambda \in \mathbb{R} \right\}$. One can easily compute the minimax risk for this experiment by a well known Bayesian argument and obtain the inequality (5).

Having obtained the lower bound to the local asymptotic risk one can now give sufficient regularity conditions under which the MLE and the Bayes estimators are asymptotically efficient for a natural class of loss functions in the sense that the lower bound is attained by these estimators (see, for example, Ibragimov and Hasminskii (1981)).

In the non-regular cases, however, it is well known that the method of maximum likelihood does not yield " efficient " estimators. Weiss and Wolfowitz (1974) studied a family of non-regular cases and suggested an estimator called maximum probability estimator. It was shown that although the MLE is not efficient for these non-regular examples, the maximum probability estimators which are equivalent to the MLE in the regular cases, continue to be efficient (in the sense of Weiss and Wolfowitz) in the non-regular cases.

In the present work we study a class of non-regular cases which include the Weiss-Wolfowitz examples and study the problem of efficient estimation in the Hajek-Le Cam approach indicated above. The starting point of this work is the remark made in Ghosh (1985) about the results of Weiss and Wolfowitz (1974), where he suggests that it is worth studying the non-regular cases using the Hajek-Le Cam-Millar approach. Ibragimov and Hasminskii (1981) also studied non-regular cases quite extensively but their methods are different from ours.

Let $\{ P^n_\theta , \theta \in \Theta \}$, $n \geq 1$, be a sequence of statistical experiments with a real parameter θ. It is noted that in many non-regular cases, the likelihood ratio $\frac{d P^n_{\theta + \lambda K_n^{-1}}}{d P^n_\theta}$ has certain local asymptotic expansion at all $\theta \in \Theta$. In Chapter 1 we obtain our results assuming such an asymptotic expansion of the likelihood ratio. It is shown that the sequence of experiments $E^n = \{ P^n_{\theta + \lambda K_n^{-1}} : \lambda \in \Lambda \}$ where Λ is some appropriate interval in $\mathbb{R}$, converges to an " exponential shift experiment ". We consider a wide class of loss functions and compute the minimax risk in the limiting experiment which gives us a lower bound to the local asymptotic minimax risk $\rho(\theta)$ by Hajek - Le Cam asymptotic minimax theorem. We then suggest an estimator which is shown to be efficient under certain assumptions. We also obtain a convolution theorem characterizing the class of possible limiting distributions of " regular " estimators. It states that the limiting distribution of any sequence of regular estimators can be expressed as the convolution of two probability distributions — one is the limiting distribution of the suggested estimator, the other being some probability measure depending on the choice of the regular estimator. An alternative proof of the result that the local asymptotic risk of the suggested estimator is minimum in the class of regular estimators follows as a corollary of the convolution theorem.

Chapter 2 deals with specific non-regular cases. In this chapter we apply the results of Chapter 1 for two important classes of non-regular examples. We first consider the case where the observations are independent and identically distributed with density whose support is an interval depending on θ. As a second example we consider a regression type model where the observations are independent but not identically distributed. We solve the problem of efficient estimation in these cases using the results obtained in Chapter 1. We also study the asymptotic properties of the maximum probability estimators and a class of Bayes estimators. This chapter and Chapter 1 are based on Samanta (1986a).

In Chapter 3 we prove the approximate Bayes property of the estimator suggested in Chapter 1. We consider only the i.i.d. case. It is well known that in the regular cases, for a wide variety of priors, the posterior tends to a normal distribution. This was first observed by Laplace (1774) and more recently by Bernstein (1917) and von Mises (1931). Using this result one can show that the MLE is asymptotically equivalent to the Bayes estimators for any prior satisfying some mild condition (see, for example, Bickel and Yahav (1969), Chao (1970) or Borwanker et al. (1971)). In Chapter 3 we prove an analogue of the Bernstein-von Mises theorem in non-regular case. The limiting posterior distribution is, however, not normal. This result is then used to study the asymptotic behaviour of the Bayes estimators and it is shown that the Bayes estimators are asymptotically equivalent to the estimator suggested in Chapter 1. We also use this result to obtain a lower bound to the local asymptotic minimax risk. It is noted that the proofs of the results of Bickel and Yahav (1969), Chao (1970) and Borwanker et al. (1971) on asymptotic normality of posterior are based on an assumption which is not satisfied even in the simplest regular cases. We show that we can obtain their results under a much weaker assumption. This chapter is based on Samanta (1986b).

In Chapters 1-3, we considered the case where there is only one unknown real parameter θ with respect to which the problem is non-regular. In Chapter 4 we consider the case in which there is an additional unknown parameter, say, φ . This type of problems were studied by Smith (1985), Cheng and Iles (1987) and others but these authors were concerned mainly with the problem of obtaining the asymptotic distribution of the maximum likelihood estimators or its alternatives. We here study the problem of efficient estimation from the Hajek-Le Cam-Millar point of view. It is assumed that the usual regularity conditions are satisfied with respect to the additional parameter φ. For simplicity, we consider only the case in which φ is a real parameter. An important result in this situation is that the problem of estimation of θ and φ , when considered together, are asymptotically independent and the limiting experiment is a product of a regular one and a non-reguler one.

CHAPTER 1

LOCAL ASYMPTOTIC MINIMAX ESTIMATION UNDER AN ASYMPTOTIC EXPANSION OF LIKELIHOOD RATIO

1.1 INTRODUCTION

Let $\{f_n(\cdot,\theta)\}$ $(n \geq 1)$ be a family of densities depending on a parameter θ taking values in Ⓗ, where Ⓗ is an open subset of the real line $\mathbb{R}$. Our problem is to estimate θ efficiently. Let $\{T_n\}$ be a sequence of estimators of θ. We consider the local asymptotic (maximum) risk

$$\rho(\theta, \{T_n\}) = \lim_{A \to \infty} \underline{\lim}_{n \to \infty} \sup_{|\theta' - \theta| \leq AK_n^{-1}} E_{\theta'} L(K_n(T_n - \theta'))$$

as a measure of the asymptotic performance of the estimator $\{T_n\}$ at θ, where L is an appropriate loss function and $K_n(\uparrow \infty)$ is the normalizing factor (see, for example, Weiss and Wolfowitz (1974)) for the given family of distributions. Thus, an estimator T_n for which the local asymptotic risk $\rho(\theta, \{T_n\})$ (with $\underline{\lim}$ replaced by lim) is equal to the local asymptotic minimax (LAM) risk

$$\rho(\theta) = \lim_{A \to \infty} \underline{\lim}_{n \to \infty} \inf_{T_n} \sup_{|\theta' - \theta| \leq AK_n^{-1}} E_{\theta'} L(K_n(T_n - \theta'))$$

may be considered as an efficient estimator. In this chapter we consider the problem of efficient estimation for a class of non-regular cases admitting certain local asymptotic expansion of the likelihood ratio. In Section 1.2 we use the results of Millar (1983) to get a sequence of experiments converging to some exponential shift experiment and then in Section 1.3 obtain a lower bound to the local asymptotic minimax risk using the Hajek-Le Cam asymptotic minimax theorem (Millar (1983)). In

Section 1.3 we also suggest an estimator which is shown to be efficient under certain assumptions. A convolution theorem, which gives the decomposition of the limiting distribution of a sequence of estimators, is proved in Section 1.4 using the notion of limiting experiments.

1.2 CONVERGENCE OF EXPERIMENTS ASSUMING ASYMPTOTIC EXPANSIONS

Let $\{(\underline{X}^n, \underline{A}^n), P^n_\theta ; \theta \in Ⓗ\}$, $n \geq 1$, be a sequence of statistical experiments, where Ⓗ is an open subset of $\mathbb{R}$. Let $\dfrac{dP^n_{\theta_2}}{dP^n_{\theta_1}}$ denote the derivative of the absolutely continuous component of $P^n_{\theta_2}$ with respect to $P^n_{\theta_1}$. Fix $\theta_0 \in Ⓗ$. We assume that either of the following two conditions holds a.e. $P^n_{\theta_0}$.

Condition (A1). For any $\lambda \geq 0$ and some sequence $K_n \uparrow \infty$,

$$\frac{dP^n_{\theta_0 + \lambda K_n^{-1}}}{dP^n_{\theta_0}} = \begin{cases} \exp\{\lambda \Delta_n(\theta_0) + \varepsilon_n(\lambda,\theta_0)\}, & \text{if } K_n(Z_n - \theta_0) > \lambda, \\ 0, & \text{if } K_n(Z_n - \theta_0) < \lambda, \end{cases} \tag{1.1}$$

where $\Delta_n(\theta_0)$ converges in $P^n_{\theta_0}$ - probability to $c(\theta_0)$ for some $c(\theta_0) > 0$, ε_n converges in $P^n_{\theta_0}$ - probability to zero, and Z_n is a random variable which does not depend on θ_0 and for which

$$\lim_{n \to \infty} P^n_{\theta_0} (K_n(Z_n - \theta_0) > t) = e^{-tc(\theta_0)} \quad \text{for all } t \geq 0.$$

Condition (A2). For any $\lambda \leq 0$ and some sequence $K_n \uparrow \infty$,

$$\frac{dP^n_{\theta_0 + \lambda K_n^{-1}}}{dP^n_{\theta_0}} = \begin{cases} \exp\{\lambda \Delta^*_n(\theta_0) + \varepsilon^*_n(\lambda,\theta_0)\}, & \text{if } K_n(Z^*_n - \theta_0) < \lambda, \\ 0, & \text{if } K_n(Z^*_n - \theta_0) > \lambda, \end{cases} \tag{1.2}$$

where $\Delta_n^*(\theta_o)$ converges in $P_{\theta_o}^n$ - probability to $c^*(\theta_o)$ for some $c^*(\theta_o) < 0$, ε_n^* converges in $P_{\theta_o}^n$ - probability to zero, and Z_n^* is a random variable which does not depend on θ_o and satisfies

$$\lim_{n \to \infty} P_{\theta_o}^n (K_n(Z_n^* - \theta_o) < t) = e^{-tc^*(\theta_o)} \quad \text{for all } t \leq 0 .$$

We define experiments

$$E^n = \left\{ P_{\theta_o + \lambda K_n^{-1}}^n : \lambda \geq 0 \right\} \quad \text{and} \quad E^{*n} = \left\{ P_{\theta_o + \lambda K_n^{-1}}^n : \lambda \leq 0 \right\} , \ n \geq 1.$$

We want to study the convergence of these sequences of experiments in the sense defined as follows (see, for example, Millar (1983)).

Definition. Let $E^n = \left\{ (S^n, \underline{\underline{S}}^n), Q_\lambda^n , \lambda \varepsilon \Lambda \right\}$, $n \geq 1$ and $E = \left\{ (S, \underline{\underline{S}}), Q_\lambda , \lambda \varepsilon \Lambda \right\}$ be experiments with parameter set Λ . Then E^n converges to E if for every finite subset $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ of Λ ,

$$\mathcal{L}\left\{ \left(\frac{dQ_{\lambda_1}^n}{d\mu^n}, \ldots, \frac{dQ_{\lambda_k}^n}{d\mu^n} \right) \Big| \mu^n \right\} \Rightarrow \mathcal{L}\left\{ \left(\frac{dQ_{\lambda_1}}{d\mu}, \ldots, \frac{dQ_{\lambda_k}}{d\mu} \right) \Big| \mu \right\} ,$$

where $\mu^n = \sum_{i=1}^k Q_{\lambda_i}^n$, $\mu = \sum_{i=1}^k Q_{\lambda_i}$.

The following proposition provides a simple method for checking convergence of experiments.

Proposition (Millar). Let $E^n = \left\{ Q_\lambda^n \right\}$, $E = \left\{ Q_\lambda \right\}$ be experiments with parameter set Λ. Suppose there exists $\lambda_o \varepsilon \Lambda$ such that for each $\lambda \varepsilon \Lambda$, Q_λ is absolutely continuous with respect to Q_{λ_o} and Q_λ^n is contiguous to $Q_{\lambda_o}^n$. Then E^n converges to E if for any finite subset $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ of Λ ,

$$\mathcal{L}\left\{\left(\frac{dQ^n_{\lambda_1}}{dQ^n_{\lambda_0}}, \ldots, \frac{dQ^n_{\lambda_k}}{dQ^n_{\lambda_0}}\right) \mid Q^n_{\lambda_0}\right\} \Rightarrow \mathcal{L}\left\{\left(\frac{dQ_{\lambda_1}}{dQ_{\lambda_0}}, \ldots, \frac{dQ_{\lambda_k}}{dQ_{\lambda_0}}\right) \mid Q_{\lambda_0}\right\} .$$

Let Q_{λ,θ_0} $(\lambda \geq 0)$ denote a probability on $\mathbb{R}$ with density

$$q_{\lambda,\theta_0}(x) = \begin{cases} c(\theta_0)\, e^{-c(\theta_0)(x-\lambda)}, & \text{for } x > \lambda , \\ 0 , & \text{for } x \leq \lambda , \end{cases}$$

and Q^*_{λ,θ_0} $(\lambda \leq 0)$ denote a probability on $\mathbb{R}$ with density

$$q^*_{\lambda,\theta_0}(x) = \begin{cases} -c^*(\theta_0) e^{-c^*(\theta_0)(x-\lambda)}, & \text{for } x < \lambda , \\ 0, & \text{for } x \geq \lambda . \end{cases}$$

Then we have the following result :

Theorem 1. (i) Under condition (A1), the sequence of experiments E^n converges to $E = \{Q_\lambda : \lambda \geq 0\}$.

(ii) Under condition (A2), the sequence of experiments E^{*n} converges to $E^* = \{Q^*_\lambda : \lambda \leq 0\}$.

(We write just Q_λ and Q^*_λ in place of Q_{λ,θ_0} and Q^*_{λ,θ_0}).

Proof. We will give the proof for case (i) only. The proof of case (ii) is exactly similar.

Set $Q^n_\lambda = P^n_{\theta_0 + \lambda K_n^{-1}}$.

It is given that for all $\lambda \geq 0$,

$$\frac{dQ^n_\lambda}{dQ^n_0} = \begin{cases} \exp(Y_n) & \text{on } B_n , \\ 0 , & \text{otherwise} , \end{cases}$$

where $Y_n \xrightarrow{Q_o^n} \lambda c(\theta_o)$ and $Q_o^n(B_n) \longrightarrow \exp(-\lambda c(\theta_o))$ as $n \longrightarrow \infty$.
This gives us

$$\mathcal{L}\left\{ \frac{dQ_\lambda^n}{dQ_o^n} \mid Q_o^n \right\} \Rightarrow \mathcal{L}\left\{ \frac{dQ_\lambda}{dQ_o} \mid Q_o \right\} .$$

Since $E_{Q_o}\left(\frac{dQ_\lambda}{dQ_o}\right) = 1$, by a result on contiguity (referred to as LeCam's 1st lemma in Hajek and Sidak (1967)) it follows that Q_λ^n is contiguous to Q_o^n for all $\lambda \geq 0$.

Further, using the asymptotic expansion (1.1) again we can prove that for $0 \leq \lambda_1 < \lambda_2 < \dots < \lambda_k$,

$$\mathcal{L}\left\{\left(\frac{dQ_{\lambda_1}^n}{dQ_o^n}, \frac{dQ_{\lambda_2}^n}{dQ_o^n}, \dots, \frac{dQ_{\lambda_k}^n}{dQ_o^n}\right) \mid Q_o^n\right\} \Rightarrow \mathcal{L}\left\{\left(\frac{dQ_{\lambda_1}}{dQ_o}, \frac{dQ_{\lambda_2}}{dQ_o}, \dots, \frac{dQ_{\lambda_k}}{dQ_o}\right) \mid Q_o\right\} .$$

Hence by the above proposition of Millar the theorem is proved. ///

Remark 1.1 Contiguity plays an important role in the proof of the above theorem. Millar's results cannot be applied if $P^n_{\theta_o + \lambda K_n^{-1}}$ is not contiguous to $P^n_{\theta_o}$ and it is usually very difficult to solve the problem if contiguity does not hold. In the proof of the above theorem we have seen that condition (A1) implies contiguity. Now suppose (1.1) holds for all $\lambda \geq 0$, where $\Delta_n(\theta_o)$ and ε_n are as in condition (A1) and Z_n is a sequence of random variables such that $\mathcal{L}\{K_n(Z_n - \theta_o) \mid P^n_{\theta_o}\}$ converges weakly to some arbitrary distribution. Then to have contiguity we must have

$$\lim_{n \to \infty} P^n_{\dot\theta_o}(K_n(Z_n - \dot\theta_o) > t) = e^{-tc(\dot\theta_o)} \quad \text{for all} \quad t \geq 0 .$$

This follows from a result on contiguity.

Remark 1.2. If condition (A1) is replaced by the following stronger condition :

$(A1)^*$ For any real u and v such that $u < v$,

$$\frac{dP^n_{\dot\theta_o + vK_n^{-1}}}{dP^n_{\dot\theta_o + uK_n^{-1}}} = \begin{cases} \exp\left\{ (v-u)\Delta_n(\dot\theta_o) + \varepsilon_n \right\}, & \text{if } K_n(Z_n - \dot\theta_o) > v , \\ 0 , & \text{otherwise} , \end{cases}$$

where $\Delta_n(\dot\theta_o)$ and ε_n are as in (A1), but the convergence is with respect to $P^n_{\dot\theta_o + uK_n^{-1}}$ and Z_n is such that

$$P^n_{\dot\theta_o + uK_n^{-1}}\left[K_n(Z_n - \theta_o) > v\right] \longrightarrow e^{-(v-u)c(\dot\theta_o)} ,$$

then proceeding as above the sequence of experiments $\left\{P^n_{\dot\theta_o + \lambda K_n^{-1}}, \lambda \in \mathbb{R}\right\}$ may be shown to converge to the experiment $\left\{Q_\lambda, \lambda \in \mathbb{R}\right\}$.

From now onwards, we will consider only the case where condition (A1) is satisfied. The treatment for the case where condition (A2) holds is similar with obvious modifications.

1.3 LOWER BOUND FOR ASYMPTOTIC RISK AND AN EFFICIENT ESTIMATOR

In this section we obtain a lower bound to the local asymptotic minimax risk using Theorem 1 and the Hajek-Le Cam asymptotic minimax theorem (stated below) and suggest an estimator for which the local asymptotic risk is equal to this lower bound.

For completeness, we state below the decision theoretic set up of the Hajek-Le Cam asymptotic minimax theorem.

Suppose we have an experiment $E = \{(S, \underline{\underline{S}}), P_\theta : \theta \in \Theta\}$ and a decision space D. We assume D to be a separable metric space and let $\underline{\underline{D}}$ be the Borel sigma field on D. A procedure b is a Markov kernel of $(S, \underline{\underline{S}})/(D, \underline{\underline{D}})$, i.e.,

for each $x \in S$, $b(x,.)$ is a probability on $(D, \underline{\underline{D}})$

and for each $A \in \underline{\underline{D}}$, $b(.,A)$ is $\underline{\underline{S}}$- measurable.

Such procedures are also known as randomized decision rules in Statistical Decision Theory.

Let us now consider a loss function $L(\theta, d)$ on $\Theta \times D$. We assume that L is nonnegative and for each θ, $L(\theta, d)$ is a lower semi-continuous function of d. The risk function of a procedure b is then given by

$$\rho(\cdot b, \theta) = \int_S \int_D L(\theta, y)\ b(x,dy).P_\theta(dx) .$$

In order to compactify the collection of all procedures, Le Cam (1955) introduced the notion of generalized procedures. We consider the Banach space M of all finite signed measures on $(S, \underline{\underline{S}})$, with the total variation norm. Let V_o be the collection of all finite linear combinations of the form $\Sigma a_i \mu_i$, where a_i's are real and for each i, $\mu_i \in M$ is absolutely continuous with respect to some P_θ, $\theta \in \Theta$. We then define $V = V(E)$ to be the closure of V_o in M. Let C(D) denote the Banach space of all bounded continuous real valued functions on D, with supremum norm.

Definition. A generalized procedure is defined to be a bilinear form on $V \times C(D)$ such that

(i) b is positive, i.e., $b(\mu,c) \geq 0$ if $\mu \geq 0$, $c \geq 0$

(ii) $|b(\mu,c)| \leq \|\mu\| \cdot \|c\|$

(iii) $b(\mu,1) = \|\mu\|$ if $\mu \geq 0$.

Any Markov kernel procedure $b(x, dy)$ is also a generalized procedure if we define

$$b(\mu, c) = \iint c(y)\, b(x, dy)\, \mu(dx) .$$

The risk function of a generalized procedure b is defined as

$$\rho(b,\theta) = \sup b(P_\theta, c)$$

where the supremum is over all $c \in C(D)$ such that $c(y) \leq L(\theta, y)$.

The collection of all generalized procedure is now compact with respect to the topology of pointwise convergence.

In general it is not true that all generalized procedures are given by Markov kernels. However, for many important statistical experiments it is true. Consider, for example, the Gaussian shift experiment $\{N(\lambda,1), \lambda \in \mathbb{R}\} = G$, say, with decision space $D = \mathbb{R}$. It is a well known result that in this case all generalized procedures are given by Markov kernels (Millar (1983, page 131)). Now this result can be used to show that the same is true also for the limiting experiment $E = \{Q_\lambda, \lambda \geq 0\}$ defined in Section 1.2. We first note that there exist probabilities μ_1, μ_2 on $(\mathbb{R}, \mathcal{B})$ such that $V(E)$ and $V(G)$ are isometrically isomorphic to $L_1(\mu_1)$ and $L_2(\mu_2)$ respectively (see Millar

(1983, page 81)). More specifically, we can choose μ_1 to be Q_0 and μ_2 to be $N(0,1)$. Now using the fact that μ_2 is symmetric about zero and the restriction of μ_2 on $\mathbb{R}^+$ is equivalent to μ_1 it can be shown that all generalized procedures on $L_1(\mu_1) \times C(\mathbb{R})$ are given by Markov kernels.

We now state the Hajek-Le Cam asymptotic minimax theorem as given in Millar (1983, Ch. III). Suppose we have experiments $E^n = \{(S^n, \underline{\underline{S}}^n), Q_\lambda^n, \lambda \varepsilon \Lambda\}$, $n \geq 1$ and $E = \{(S, \underline{\underline{S}}), Q_\lambda, \lambda \varepsilon \Lambda\}$. Let D be a fixed decision space and L a loss function on $\Lambda \times D$. Let $\rho_n(b,\lambda)$ and $\rho(b,\lambda)$ be the risk functions of a procedure b in the decision theoretic structure (E^n, D, L) and (E, D, L) respectively. Then we have

<u>Theorem (Hajek-Le Cam asymptotic minimax theorem)</u>. If E^n converges to E, then

$$\underline{\lim}_{n \to \infty} \inf_b \sup_\lambda \rho_n(b,\lambda) \geq \inf_b \sup_\lambda \rho(b,\lambda)$$

where the infimum in either side is over all generalized procedures for the corresponding experiment.

We now consider the problem of estimating θ when condition (A1) holds for all $\theta_0 \varepsilon (\overline{H})$.

<u>Definition</u>. A loss function of the form $L(\theta, a) = L(\theta - a)$ is said to be subconvex if L satisfies the following conditions:

(i) $L(x) \geq 0$ for all x

(ii) $L(x) = L(|x|)$ for all x

(iii) $\{x : L(x) \leq c\}$ is closed and convex for all $c > 0$.

All the loss functions considered in this paper will be assumed to be subconvex.

Lemma 1. Under assumption (A1), for any subconvex loss function L,

$$\lim_{A \to \infty} \varliminf_{n \to \infty} \inf_{T_n} \sup_{|\theta - \theta_0| \leq AK_n^{-1}} E_\theta L\left[K_n(T_n - \theta)\right]$$

$$\geq \inf_{\delta} \sup_{0 \leq \lambda < \infty} \rho(\delta, \lambda) \tag{1.3}$$

where the infimum in left hand side is over all estimators T_n of θ, the infimum in right hand side is over all randomized (Markov kernel) procedures for the experiment E with decision space as $\mathbb{R}$ and parameter space as $[0, \infty)$, and $\rho(\delta,\lambda)$ is the risk of the procedure δ at λ with loss function L.

Proof. The proof is similar to that of Theorem VII.2.6 of Millar (1983). For any $A > 0$, the sequence of experiments $E_A^n = \{Q_\lambda^n, 0 \leq \lambda \leq A\}$ converges to $E_A = \{Q_\lambda, 0 \leq \lambda \leq A\}$. Hence by the Hajek-Le Cam asymptotic minimax theorem and a change of variable we have

$$\varliminf_{n \to \infty} \inf_{T_n} \sup_{|\theta - \theta_0| \leq AK_n^{-1}} E_\theta L\left[K_n(T_n - \theta)\right]$$

$$= \varliminf_{n \to \infty} \inf_{T_n} \sup_{|\lambda| \leq A} E_{\theta_0 + \lambda K_n^{-1}} L\left[K_n(T_n - \theta_0) - \lambda\right]$$

$$\geq \inf_{b} \sup_{0 \leq \lambda \leq A} \rho(b,\lambda)$$

where the infimum is over all generalized procedures for the limiting experiment.

Let us now denote by $\mathcal{M}_A(\mathcal{M})$ the set of all probability measures μ on $[0,A]$ ($[0,\infty)$) with finite support. For any procedure b, we define $\rho(b,\mu) = \int \rho(b,\lambda) d\,\mu(\lambda)$. Then using an ordinary minimax theorem (Theorem III.1.3 in Millar (1983)) we have

$$\lim_{A \to \infty} \inf_{b} \sup_{0 \le \lambda \le A} \rho(b,\lambda)$$

$$= \lim_{A \to \infty} \inf_{b} \sup_{\mu \varepsilon \mathcal{M}_A} \rho(b,\mu)$$

$$= \lim_{A \to \infty} \sup_{\mu \varepsilon \mathcal{M}_A} \inf_{b} \rho(b,\mu)$$

$$= \sup_{\mu \varepsilon \mathcal{M}} \inf_{b} \rho(b,\mu) \quad [\text{since } \mathcal{M}_A \uparrow \mathcal{M} \text{ as } A \to \infty]$$

$$= \inf_{b} \sup_{\mu \varepsilon \mathcal{M}} \rho(b,\mu)$$

$$= \inf_{b} \sup_{0 \le \lambda \le \infty} \rho(b,\lambda) .$$

The result now follows because for the experiment E, all generalized procedures are given by Markov kernels. ///

We will now compute the minimax risk given in the right hand side of (1.3). We will use a well-known technique of finding minimax risk.

We assume that

C(i) $E_{Q_0} L(X - a) = \int L(x - a) d\, Q_0(x)$ exists and is finite for some a and there exists $b = b(\theta_0)$ such that

$$E_{Q_0} L(X - b(\theta_0)) = \inf_a E_{Q_0} L(X - a) = R_{\theta_0} \text{ , say.}$$

C(ii) For every $\varepsilon > 0$, there exists $N > 0$ such that for all $a \in \mathbb{R}$,

$$\int_0^N L(x - a) d\, Q_0(x) \ge R_{\theta_0} - \varepsilon .$$

C(iii) $b(\theta)$ is a continuous function of θ.

Lemma 2. For any subconvex loss function satisfying conditions C(i) and C(ii), we have

$$\inf_{\delta} \sup_{0\leq\lambda<\infty} \rho(\delta,\lambda) = \int_0^{\infty} L(x - b(\theta_0))c(\theta_0)e^{-c(\theta_0)x}\, dx$$

where the minimax risk in the left hand side is as described in Lemma 1.

Proof. We shall exhibit a sequence τ_M of prior distributions on $[0, \infty)$ and show that

$$\lim_{M\to\infty} \inf_{\delta} r(\delta,\tau_M) \geq R_{\theta_0} \tag{1.4}$$

where the infimum in the left hand side is over all randomized (Markov kernel) procedures and $r(\delta,\tau_M)$ is the Bayes risk of δ with respect to the prior τ_M.

We choose τ_M as the uniform distribution over the interval $(0,M)$. Let $\varepsilon > 0$ and N be such that $\int_0^N L(x - a)d\, Q_0(x) \geq R_{\theta_0} - \varepsilon$ for all a. Proceeding as in Ferguson (1967, Section 4.5, p.172) we can prove that for any $M > N$ and any nonrandomized decision rule δ,

$$r(\delta,\tau_M) \geq (R_{\theta_0} - \varepsilon)\, \frac{M - N}{M} .$$

Therefore, for any $M > N$, $r(\delta,\tau_M) \geq (R_{\theta_0} - \varepsilon)\, \frac{M - N}{M}$ for all " randomized " procedures δ which are probabilities over the space of non-randomized decision rules. This proves (1.4) using a result on equivalence of two methods of randomization (see, for example, Wald and Wolfowitz (1951)). Since $X - b(\theta_0)$ is an equalizer rule with constant risk R_{θ_0}, the lemma is proved. ///

Now, from Lemma 1 and Lemma 2 we get the following result:

<u>Theorem 2</u>. Under assumption (A1), for any subconvex loss function L satisfying C(i) and C(ii),

$$\lim_{A\to\infty} \underline{\lim}_{n\to\infty} \inf_{T_n} \sup_{|\theta-\theta_0|\le AK_n^{-1}} E_\theta L\left[K_n(T_n-\theta)\right]$$
$$\ge \int_0^\infty L(x-b(\theta_0))\, c(\theta_0)\, e^{-c(\theta_0)x}\, dx .$$

<u>Remark</u>. To prove Theorem 2, we need not assume that Z_n (in Condition (A1)) is independent of θ_0. Indeed, we may replace the set $\{K_n(Z_n-\theta_0)>\lambda\}$ by $\{\tau_n>\lambda\}$, where τ_n is a random variable such that

$$\lim_{n\to\infty} P^n_{\theta_0}(\tau_n>t) = e^{-tc(\theta_0)} \qquad \text{for all } t\ge 0 .$$

Our problem is now to search for an estimator $\hat{\theta}_n$ for which

$$\left.\begin{aligned} &\lim_{A\to\infty} \underline{\lim}_{n\to\infty} \inf_{T_n} \sup_{|\theta-\theta_0|\le AK_n^{-1}} E_\theta L\left[K_n(T_n-\theta)\right] \\ &= \lim_{A\to\infty} \lim_{n\to\infty} \sup_{|\theta-\theta_0|\le AK_n^{-1}} E_\theta L\left[K_n(\hat{\theta}_n-\theta)\right] \quad \text{for all } \theta_0 \in (H) \end{aligned}\right\} \quad (1.5)$$

<u>Definition</u>. An estimator $\hat{\theta}_n$ for which (1.5) holds is said to be a locally asymptotically minimax (LAM) estimator of θ.

It follows from Theorem 2 that an estimator $\hat{\theta}_n$ for which

$$\lim_{A\to\infty} \lim_{n\to\infty} \sup_{|\theta-\theta_0|\le AK_n^{-1}} E_\theta L\left[K_n(\hat{\theta}_n-\theta)\right]$$
$$= \int_0^\infty L(x-b(\theta_0))\, c(\theta_0) e^{-c(\theta_0)x}\, dx$$

is a locally asymptotically minimax estimator.

Let us now consider the case for which condition (A1) is satisfied for all $\theta_0 \in \textcircled{H}$. Condition (A1) ensures the existence of a sequence of statistics Z_n for which $K_n(Z_n - \theta_0)$ converges in distribution (under $P^n_{\theta_0}$) to a random variable X with distribution Q_0.

Definition. A sequence of estimators T_n is said to be regular at $\theta_0 \in \textcircled{H}$ if for some probability distribution G,

$$\mathcal{L}\left\{K_n(T_n - \theta_0 - \lambda K_n^{-1}) \mid P^n_{\theta_0 + \lambda K_n^{-1}}\right\} \Rightarrow G \quad \text{as} \quad n \to \infty$$

uniformly in $\{|\lambda| \leq c\}$ for any $c > 0$.

Theorem 3. Suppose condition (A1) holds for all $\theta_0 \in \textcircled{H}$ and the sequence of statistics Z_n is regular at all values θ_0 in $\textcircled{H}$.

Set $\hat{\theta}_n = Z_n - K_n^{-1}\, b(Z_n)$.

Then the following results hold:

(i) For any bounded subconvex loss function satisfying conditions C(i), C(iii) (condition C(ii) is satisfied for bounded loss function), $\hat{\theta}_n$ is LAM.

(ii) Suppose that for some $r > 0$,

$$\lim_{A \to \infty} \lim_{n \to \infty} \sup_{|\theta - \theta_0| \leq AK_n^{-1}} E_\theta |K_n(\hat{\theta}_n - \theta)|^r < \infty \tag{1.6}$$

for all $\theta_0 \in \textcircled{H}$. Then for any subconvex loss function L satisfying conditions C(i), C(ii) and C(iii) for which

$$L(u) \leq B(1 + |u|^s) \quad \text{for all} \quad u \in \mathbb{R}$$

for some $B > 0$ and $0 < s < r$, we have

$$\lim_{A \to \infty} \lim_{n \to \infty} \sup_{|\theta - \theta_0| \leq AK_n^{-1}} E_\theta L\left[K_n(\hat{\theta}_n - \theta)\right]$$

$$= \int L(x - b(\theta_0))dQ_{0,\theta_0} \quad \text{for all } \theta_0 \in (H)$$

and hence $\hat{\theta}_n$ is LAM.

Proof. Fix $A > 0$. Under the conditions of the theorem for any $\theta_0 \in (H)$ and for any sequence $\{\theta_n\}$ satisfying $|K_n(\theta_n - \theta_0)| \leq A$,

$$\mathcal{L}\left\{K_n(Z_n - \theta_n) | P^n_{\theta_n}\right\} \Rightarrow Q_{0,\theta_0} .$$

Since $b(\theta)$ is continuous in θ, $b(Z_n)$ converges in $P^n_{\theta_n}$ - probability to $b(\theta_0)$. Thus,

$$\mathcal{L}\left\{K_n(\hat{\theta}_n - \theta_n) | P^n_{\theta_n}\right\} \Rightarrow \mathcal{L}\{X - b(\theta_0)\}$$

where X is a random variable with distribution Q_{0,θ_0}.

We shall now prove that

$$\mathcal{L}\left\{L\left[K_n(\hat{\theta}_n - \theta_n)\right] \mid P^n_{\theta_n}\right\} \Rightarrow \mathcal{L}\left\{L(X - b(\theta_0))\right\} . \tag{1.7}$$

Take any $t \geq 0$. $B_t = \{x : L(x) \leq t\}$ is closed convex subset of $\mathbb{R}$. Since the Lebesgue measure of the boundary of any convex set is zero, B_t is a continuity set with respect to the distribution of $X - b(\theta_0)$ and we have

$$\lim_{n \to \infty} P^n_{\theta_n}\left(L\left[K_n(\hat{\theta}_n - \theta_n)\right] \leq t\right)$$

$$= \lim_{n \to \infty} P^n_{\theta_n}\left[K_n(\hat{\theta}_n - \theta_n) \in B_t\right]$$

$$= Q_{0,\theta_0}(\{x : (x - b(\theta_0)) \varepsilon B_t\})$$
$$= Q_{0,\theta_0}(\{x : L(x-b(\theta_0)) \leq t\}) .$$

Hence (1.7) is proved.

Now

$$\lim_{n \to \infty} \sup_{|\theta - \theta_0| \leq AK_n^{-1}} E_\theta L\left[K_n(\hat{\theta}_n - \theta)\right]$$

$$= \lim_{n \to \infty} E_{\theta_n} L\left[K_n(\hat{\theta}_n - \theta_n)\right]$$

for some sequence $\{\theta_n\}$, $n \geq 1$, satisfying $|K_n(\theta_n - \theta_0)| \leq A$. The proof of the 1st part of the theorem is now obvious. We will now prove the 2nd part of the theorem. We are given that

$$\lim_{n \to \infty} E_{\theta_n} |K_n(\hat{\theta}_n - \theta_n)|^r < \infty .$$

Since $L(u) \leq B(1 + |u|^s)$ for all $u \varepsilon \mathbb{R}$, there exists $\varepsilon > 0$ such that for some $B_1, B_2 > 0$,

$$\left[L(u)\right]^{1+\varepsilon} \leq B_1 + B_2 |u|^r \quad \text{for all } u \varepsilon \mathbb{R}.$$

Therefore we have

$$\lim_{n \to \infty} E_{\theta_n} \left[L\left(K_n(\hat{\theta}_n - \theta_n)\right)\right]^{1+\varepsilon} \leq B_1 + B_2 \lim_{n \to \infty} E_{\theta_n} |K_n(\hat{\theta}_n - \theta_n)|^r$$

$$< \infty .$$

This together with (1.7) proves the theorem. ///

Remark 3.1. Part (i) of Theorem 3 holds for any sequence of regular estimators $\hat{\theta}_n$ for which $K_n(\hat{\theta}_n - \theta_0)$ converges in distribution (under $P^n_{\theta_0}$) to $X - b(\theta_0)$ $(X \sim Q_0)$ for all $\theta_0 \varepsilon ⒽH$.

Remark 3.2. It is interesting to note that condition (A1) itself implies that for all $\lambda \geq 0$,

$$\mathcal{L}\left\{n(Z_n - \theta_0 - \lambda K_n^{-1}) \mid P^n_{\theta_0 + \lambda K_n^{-1}}\right\} \Rightarrow Q_0 . \tag{1.8}$$

This follows from the fact that $P^n_{\theta_0 + \lambda K_n^{-1}}$ is contiguous to $P^n_{\theta_n}$. Similarly condition $(A1)^*$ implies that (1.8) holds for all real λ. Moreover, under uniform versions of condition (A1) and $(A1)^*$, where the convergence of ε_n are uniform in λ and (u,v) belonging to compact sets, the convergence (1.8) may be shown to be uniform for all λ in compact sets.

Remark 3.3. The LAM estimator here depends on the loss function chosen whereas in the regular case it is possible to find LAM estimators not depending on the loss function.

1.4 A CONVOLUTION THEOREM

We shall now try to characterize the class of possible limiting distributions of appropriately normalized estimators. We shall consider only the class of estimators T_n (which includes all regular estimators) for which

$$\mathcal{L}\left\{K_n(T_n - \theta_0 - \lambda K_n^{-1}) \mid P^n_{\theta_0 + \lambda K_n^{-1}}\right\} \Rightarrow G \quad \text{for all } \lambda \geq 0 \tag{1.9}$$

where G is some probability distribution not depending on λ.

The convolution theorem was first proved by Hajek (1970) and Inagaki (1970) for regular cases. Consider, for example, the case when the observations are i.i.d. with a density $f(x, \theta)$, $\theta \in \mathbb{R}$. Let T_n

be any regular estimator of θ based on a sample of size n. Then under the usual assumptions of the regular case, the limiting distribution G_θ of $\sqrt{n}(T_n - \theta)$ is a convolution of $N(0, I^{-1}(\theta))$ and some other probability distribution μ which depends on the choice of the estimator T_n ($I(\theta)$ denotes the Fisher information). Since convolution " spreads out mass " G_θ is more spread out than $N(0, I^{-1}(\theta))$ and thus, an estimator T_n for which the measure μ is unit mass at $\{0\}$ (i.e., the limiting distribution G_θ is $N(0, I^{-1}(\theta))$) may be considered as an efficient estimator. The convolution theorem can also be used to obtain lower bound for asymptotic risk of estimators.

Hajek (1970) proved his result for the regular cases. A simpler proof, due to Bickel, is given in Roussas (1972). A more general result based on the notion of limiting experiment is proved in Millar (1983) using Kakutani's fixed point theorem. The proof was originally sketched in Le Cam (1972).

In the non-regular case where condition (A1) holds we have the following theorem.

Theorem 4. Suppose condition (A1) holds. Then for any estimator T_n satisfying (1.9), the limiting distribution G of $K_n(T_n - \theta_0)$ under $P^n_{\theta_0}$ is a convolution of Q_0 and some probability distribution μ depending on $\{T_n\}$:

$$G = Q_0 * \mu .$$

To prove Theorem 4, we shall use a slightly different version of the convolution theorem, than the one given in Millar (1983, Ch.III), which we state and prove below (For an alternative proof of Theorem 4 see Ibragimov and Hasminskii (1981, Ch.V)).

Theorem (Millar). Let $E^n = \{(S^n, \underline{\underline{S}}^n), Q^n_\lambda, \lambda \geq 0\}$, $n \geq 1$, $E = \{(\mathbb{R}, \mathcal{B}), Q_\lambda, \lambda \geq 0\}$ be statistical experiments ($\mathcal{B}$ denotes the Borel σ-field on $\mathbb{R}$). Assume E^n converges to E. Suppose R_n is a sequence of statistics on $(S^n, \underline{\underline{S}}^n)$ taking values in $\mathbb{R}$. Assume further

i) there is a family of probabilities $\{G_\lambda, \lambda \geq 0\}$ on $(\mathbb{R}, \mathcal{B})$ such that for each $\lambda \geq 0$,

$$\mathcal{L}\{R_n \mid Q^n_\lambda\} \Rightarrow G_\lambda .$$

ii) Q_o is concentrated on $\mathbb{R}^+$ and is absolutely continuous with respect to Lebesgue measure. Also the number 0 belongs to the support of Q_o.

iii) $Q_\lambda(A) = Q_o(A-\lambda)$, $G_\lambda(A) = G_o(A-\lambda)$ for all $\lambda \geq 0$ and all $A \in \mathcal{B}$.

Then there is a probability μ on $\mathbb{R}$ such that

$$G_o = Q_o * \mu .$$

Proof. Let $E^R = \{(\mathbb{R}, \mathcal{B}), G_\lambda, \lambda \geq 0\}$. Then by an argument given in Millar (1983, p.98) there exists a Markov kernel K of $(\mathbb{R}, \mathcal{B})/(\mathbb{R}, \mathcal{B})$ such that $G_\lambda = KQ_\lambda$ for all $\lambda \geq 0$. Let $\mathcal{K}_o$ be the collection of all Markov kernels K of $(\mathbb{R}, \mathcal{B})/(\mathbb{R}, \mathcal{B})$ such that

$$G_\lambda = KQ_\lambda \quad \text{for all } \lambda \geq 0.$$

For all $g \geq 0$, we define a map $K \longrightarrow gK$ as

$$gK(x,A) = K(x+g, A+g),\ A \in \mathcal{B},\ x \in \mathbb{R},\ K \in \mathcal{K}_o .$$

Then $\mathcal{K}_o$ is a compact convex subset of a linear topological space. Also, the family $\{g : g \geq 0\}$ is a commuting family of continuous linear mappings which map $\mathcal{K}_o$ into itself. Therefore, by Markov-Kakutani fixed point theorem (see Dunford Schwartz, Vol. I, p. 456) there exists $K \in \mathcal{K}_o$ such that

$$gK = K \quad \text{for all} \quad g \geq 0$$

i.e., for every $\mu \in V(E)$, every Borel set A and every $g \geq 0$,

$$\int gK(x,A)\,d\mu(x) = \int K(x,A)\,d\mu(x) \ .$$

Since $V(E) = L_1(\nu)$ for some probability ν with support $\mathbb{R}^+$ which is equivalent to the Lebesgue measure, this implies for every $g \geq 0$ and $A \in \mathcal{B}$,

$$K(x,A) = K(x+g,\ A+g) \quad \text{a.e.} \quad x \geq 0.$$

Therefore, by Fubini's theorem there exists a null set N such that

$$\text{for} \quad x \notin N,\ x \geq 0,\ \{K(x,A) = K(x+g,\ A+g) \quad \text{for all} \quad A \in \mathcal{B}\} \ \text{a.e.}\ g \geq 0.$$

We now choose a sequence $\alpha_n \downarrow 0$, $\alpha_n \in N^c$ for all n.

For all $n \geq 1$, there is a null set N_n such that for all $g \notin N_n$, $g \geq 0$,

$$K(\alpha_n + g,\ A + g) = K(\alpha_n,\ A) \quad \text{for all} \quad A \in \mathcal{B} \ .$$

Therefore, for all $x \notin N_n + \alpha_n$, $x \geq \alpha_n$,

$$K(x,\ A+x) = K(\alpha_n,\ A+\alpha_n) \quad \text{for all} \quad A \in \mathcal{B}.$$

Let $N_o = \bigcup_{n \geq 1} (N_n + \alpha_n)$.

Then N_o is a null set and for any $x, y > 0$ such that $x \notin N_o$, $y \notin N_o$ we have

$$K(x,\ A+x) = K(y,\ A+y) \quad \text{for all} \quad A \in \mathcal{B} \ .$$

To see this choose $\alpha_n < x, y$ and note that $x, y \notin N_o$ implies $x, y \notin N_n + \alpha_n$ and hence $K(x,\ A+x) = K(\alpha_n,\ A+\alpha_n) = K(y,\ A+y)$.

Suppose the common value is $\mu(A)$.

$$\text{i.e., for all } x \notin N_o,\ K(x,A+x) = \mu(A) \text{ for all } A \in \mathcal{B} .$$

This implies

$$K(x,A) = K(x,(A-x)+x) = \mu(A-x) \quad \text{for all } A \in \mathcal{B}, \text{ for all } x \notin N_o$$

and therefore,

$$G_\lambda(A) = \int K(x,A)\, dQ_\lambda(x) = \int \mu(A-x)\, dQ_\lambda(x) .$$

Since μ is a probability this proves the theorem. ///

Proof of Theorem 4.

Consider $E^n = \left\{ Q_\lambda^n : \lambda \geq 0 \right\}$, $E = \left\{ Q_\lambda : \lambda \geq 0 \right\}$,

$$R_n = K_n(T_n - \theta_o) ,$$

where Q_λ is as defined earlier in this section

and $Q_\lambda^n = P^n_{\theta_o} + \lambda K_n^{-1}$ for $\lambda \geq 0$.

It is easy to see that all the conditions of the above theorem are satisfied and hence

$$G = Q_o * \mu$$

for some probability μ on $\mathbb{R}$. ///

Corollary. Under the conditions of Theorem 4, for any sequence of estimators T_n satisfying (1.9), we have

$$\varliminf_{n \to \infty} E_{\theta_o} L\left[K_n(T_n - \theta_o)\right] \geq \lim_{n \to \infty} E_{\theta_o} L\left[K_n(\hat{\theta}_n - \theta_o)\right]$$

where L is a loss function satisfying the conditions given in Theorem 3.

Proof. By convolution theorem

$$\mathcal{L}\left\{L\left[K_n(T_n-\theta_o)\right] \mid P^n_{\theta_o}\right\} \Rightarrow \mathcal{L}\{L(X+\xi)\},$$

where X and ξ are independent random variables and $X \sim Q_o$.
Using Fatou's lemma

$$\begin{aligned}\lim_{n\to\infty} E_{\theta_o} L\left[K_n(T_n-\theta_o)\right] &\geq E\, L(X+\xi)\\ &= \int E\, L(X+y)\,dF_\xi(y)\\ &\geq E\, L(X-b(\theta_o))\\ &= \lim_{n\to\infty} E_{\theta_o} L\left[K_n(\hat{\theta}_n-\theta_o)\right] . \quad ///\end{aligned}$$

1.5 STATEMENT OF RESULTS UNDER CONDITION (A2)

In Sections 1.3 and 1.4 we considered only the cases when condition (A1) holds. Proceeding in an exactly similar manner we can prove results for the case when condition (A2) is satisfied. In this section we only state these results.

Part (ii) of Theorem 1 states that the sequence of experiments $E^{*n} = \{P^n_{\theta_o+\lambda K_n^{-1}},\ \lambda \leq 0\}$ converges to the experiment $E^* = \{Q^*_\lambda,\ \lambda \leq 0\}$, where Q^*_λ is defined as in Section 1.2. This fact together with the Hajek-Le Cam asymptotic minimax theorem gives us a lower bound to the local asymptotic minimax risk.

We consider a subconvex loss function L satisfying the following conditions.

C^*(i) $E_{Q^*_o} L(X-a) = \int L(x-a)\,dQ^*_o(x)$ exists and is finite for some a and there exists $b = b(\theta_o)$ such that

$$E_{Q_0^*} L(X - b(\theta_0)) = \inf_a E_{Q_0^*} L(X - a) = R_{\dot{\theta}_0} \text{ , say}$$

$C^*(ii)$ For every $\varepsilon > 0$, there exists $N > 0$ such that for all $a \in \mathbb{R}$,

$$\int_{-N}^{0} L(x - a)\,dQ_0^*(x) \geq R_{\dot{\theta}_0} - \varepsilon$$

$C^*(iii)$ $b(\dot{\theta})$ is a continuous function of θ .

We now have the following theorems.

Theorem 2(a). Under assumption (A2), for any subconvex loss function L satisfying $C^*(i)$ and $C^*(ii)$,

$$\lim_{A \to \infty} \underline{\lim}_{n \to \infty} \inf_{T_n} \sup_{|\dot{\theta} - \dot{\theta}_0| \leq AK_n^{-1}} E_{\dot{\theta}} L\left[K_n(T_n - \dot{\theta})\right]$$
$$\geq -\int_{-\infty}^{0} L(x - b(\dot{\theta}_0))\; c^*(\dot{\theta}_0)\; e^{-c^*(\dot{\theta}_0)x}\, dx .$$

A locally asymptotically minimax estimator of $\dot{\theta}$ is suggested in the following theorem:

Theorem 3(a). Suppose condition (A2) holds for all $\dot{\theta}_0 \in \text{Ⓗ}$ and the sequence of statistics Z_n is regular at all $\dot{\theta}_0$ in Ⓗ .

Set $\hat{\theta}_n = Z_n^* - K_n^{-1}\, b(Z_n^*)$. Then both part (i) and part (ii) of Theorem 3 hold with Q_0 replaced by Q_0^* .

The convolution theorem can be stated as follows :

Theorem 4(a). Suppose condition (A2) holds. Let T_n be any estimator for which

$$\mathcal{L}\left\{K_n(T_n - \theta_0 - \lambda K_n^{-1}) \mid P^n_{\theta_0 + \lambda K_n^{-1}}\right\} \Rightarrow G \quad \text{for all } \lambda \leq 0$$

where G is some probability distribution not depending on λ. Then G is a convolution of Q_o^* and some probability distribution μ depending on $\{T_n\}$:

$$G = Q_o^* * \mu .$$

CHAPTER 2

EXAMPLES OF NON-REGULAR CASES

2.1 INTRODUCTION

In Chapter 1 we obtained our results for a class of non-regular cases admitting certain local asymptotic expansion of the likelihood ratio. In this chapter we apply the results of Chapter 1 for the estimation problem in two important classes of non-regular examples. In Section 2.2 we consider the case where the observations are independent and identically distributed with density whose support is an interval which is monotonic in θ. In Section 2.3 we consider a regression type model. We verify that in both these cases the local asymptotic expansion (A1) or (A2) of Chapter 1 is valid and hence the conclusions of all the theorems of Chapter 1 hold. In Sections 2.4 and 2.5 we study the asymptotic properties of the maximum probability estimator and a class of Bayes estimators.

2.2 INDEPENDENT AND IDENTICALLY DISTRIBUTED OBSERVATIONS

Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations, each X_i having density $f(x,\theta)$ on $\mathbb{R}$ with respect to the Lebesgue measure, where $\theta \in \textcircled{H}$ an open subset of $\mathbb{R}$. We assume that $f(x,\theta)$ is strictly positive for all x in a closed interval (bounded or unbounded) $S(\theta)$ depending on θ and is zero outside $S(\theta)$. Let $A_1(\theta)$, $A_2(\theta)$, $(A_1 < A_2)$ be the boundaries of $S(\theta)$. We consider the following cases :

Case I. The support $S(\theta)$ is nonincreasing in θ, i.e., $S(\theta_2) \subset S(\theta_1)$ whenever $\theta_2 > \theta_1$.

Case II. The support $S(\theta)$ is nondecreasing in θ, i.e., $S(\theta_2) \supset S(\theta_1)$ whenever $\theta_2 > \theta_1$.

We now make the following assumptions on the density $f(x,\theta)$ (Weiss and Wolfowitz (1974) have similar assumptions when they study properties of maximum probability estimators).

1. $A_1(\theta)$ and $A_2(\theta)$ are continuously differentiable functions of θ (if not infinity).

2. On the set $\{(x,\theta) : x \in S(\theta)\}$, $f(x,\theta)$ is jointly continuous in (x,θ)

3. The derivatives $\frac{\partial f(x,\theta)}{\partial \theta}$, $\frac{\partial^2 \log f(x,\theta)}{\partial \theta^2}$ exist for all (x,θ) in $\{(x,\theta) : A_1(\theta) < x < A_2(\theta)\}$.

4. For all $\theta_0 \in \textcircled{H}$, there exists a neighbourhood $N(\theta_0)$ of θ_0 and a constant $D(\theta_0) > 0$ such that for all $\theta \in N(\theta_0)$,

$$\left| \frac{\partial^2 \log f(x,\theta)}{\partial \theta^2} \right| \leq D(\theta_0)$$

for all x for which the derivative exists.

5. For all $\theta \in \textcircled{H}$, $E_\theta \frac{\partial \log f(X,\theta)}{\partial \theta} = c(\theta)$ is finite and not equal to zero.

In all the above non-regular cases we can obtain an asymptotic expansion of the likelihood ratio $\frac{dP^n_{\theta_0 + \lambda n^{-1}}}{dP^n_{\theta_0}}$ at any $\theta_0 \in \textcircled{H}$ and for all λ in an appropriate subset Λ of $\mathbb{R}$. Here P^n_θ is the n-fold product of the measure P_θ with density $f(x,\theta)$. For Case I, $\Lambda = [0,\infty)$

and for Case II, $= (-\infty, 0]$. In either of the cases, for all $\theta_o \varepsilon (\bar{H})$ and $\lambda \varepsilon \Lambda$, $P^n_{\theta_o + \lambda n^{-1}}$ is absolutely continuous with respect to $P^n_{\theta_o}$.

Expanding at θ_o by Taylor's theorem we get

$$\log \frac{dP^n_{\theta_o + \lambda n^{-1}}}{dP^n_{\theta_o}} = \lambda \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(X_i, \theta)}{\partial \theta}\Big|_{\theta_o} + \frac{\lambda^2}{2n^2} \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i, \theta)}{\partial \theta^2}\Big|_{\theta'_n(\underset{\sim}{X})}$$

on $B_{n,\lambda} = \{(X_1, \ldots, X_n) : \text{each } X_i \varepsilon (A_1(\theta_o), A_2(\theta_o)) \cap (A_1(\theta_o + \frac{\lambda}{n}), A_2(\theta_o + \frac{\lambda}{n}))\}$

$= \{\underset{\sim}{X} : \text{each } X_i \varepsilon (A_1(\theta_o + \frac{\lambda}{n}), A_2(\theta_o + \frac{\lambda}{n}))\}$

(i.e, on the set where the Taylor's expansion is possible)

where $\theta'_n(\underset{\sim}{X})$ lies between θ_o and $\theta_o + \frac{\lambda}{n}$,

and $$\frac{dP^n_{\theta_o + \lambda n^{-1}}}{dP^n_{\theta_o}} = 0 \quad \text{a.e. } P^n_{\theta_o} \text{ on } B^c_{n,\lambda}.$$

Also, $$\Delta_n(\theta_o) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(X_i, \theta)}{\partial \theta}\Big|_{\theta_o} \longrightarrow c(\theta_o) \quad \text{a.s. } P_{\theta_o}$$

by strong law of large numbers.

By assumption 4,

$$\frac{1}{n^2} \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i, \theta)}{\partial \theta^2}\Big|_{\theta'_n} \longrightarrow 0 \quad \text{a.s. } P_{\theta_o}$$

The set $B_{n,\lambda}$ can be expressed as $\{n(Z_n(X) - \theta_o) > \lambda\}$ or $\{n(Z^*_n(X) - \theta_o) < \lambda\}$ but the form of Z_n or Z^*_n depends on $A_1(\theta)$ and $A_2(\theta)$.

We shall now consider cases with different possible A_1, A_2.

Case I(a). $S(\theta)$ is an unbounded interval

i.e., $S(\theta) = [A_1(\theta), \infty)$ or $S(\theta) = (-\infty, A_2(\theta)]$

where A_1 is a monotonic nondecreasing function of θ and A_2 is a monotonic nonincreasing function of θ. For simplicity, let us first consider the simple case where $S(\theta) = [\theta, \infty)$.

$$\text{In this case, } B_{n,\lambda} = \{(X_1,\ldots,X_n): \text{each } X_i \in (\theta_0 + \tfrac{\lambda}{n}, \infty)\}$$
$$= \{ \underset{\sim}{X} : n(W_n - \theta_0) > \lambda \} , \lambda \geq 0 ,$$

where $W_n = \min(X_1, X_2, \ldots, X_n)$.

Thus the asymptotic expansion (1.1) of Chapter 1 holds. For any sequence $\{\theta_n\}$ satisfying $|n(\theta_n - \theta_0)| \leq C$ and for any $t > 0$,

$$P^n_{\theta_n}\left[n(W_n - \theta_n) > t \right] = \left[1 - \int_{\theta_n}^{\theta_n + \frac{t}{n}} f(x,\theta_n)dx \right]^n$$

and
$$n \int_{\theta_n}^{\theta_n + \frac{t}{n}} f(x,\theta_n)dx \longrightarrow f(\theta_0,\theta_0)\,t.$$

Thus assumption (A1) of Chapter 1 holds with $Z_n = W_n$ which is regular and $c(\theta) = f(\theta, \theta)$ and hence conditions of all the theorems in Chapter 1 are satisfied. For arbitrary A_1, A_2 we can define A_1^{-1}, A_2^{-1} as in case I(b) or II(b) and proceed in a similar manner.

Case I(b). $S(\theta) = [A_1(\theta), A_2(\theta)]$ with $A_1'(\theta) \geq 0$ and $A_2'(\theta) \leq 0$.

Here

$$B_{n,\lambda} = \{ \underset{\sim}{X} : A_1(\theta_0 + \tfrac{\lambda}{n}) < x_i < A_2(\theta_0 + \tfrac{\lambda}{n}) \text{ for } i = 1,2,\ldots,n \}$$
$$= \{ \underset{\sim}{X} : W_n > A_1(\theta_0 + \tfrac{\lambda}{n}), V_n < A_2(\theta_0 + \tfrac{\lambda}{n}) \} , \lambda \geq 0$$

where $W_n = \min(X_1, X_2, \ldots, X_n)$, $V_n = \max(X_1, X_2, \ldots, X_n)$.

If A_1, A_2 are strictly monotonic functions, they possess unique inverse A_1^{-1}, A_2^{-1} and $B_{n,\lambda}$ can be expressed as $\left\{ \underset{\sim}{X} : n(Z_n - \theta_o) > \lambda \right\}$ with $Z_n = \min \left\{ A_1^{-1}(W_n), A_2^{-1}(V_n) \right\}$

Here $c(\theta) = A_1'(\theta)f(A_1(\theta),\theta) - A_2'(\theta)f(A_2(\theta),\theta) > 0$.

For arbitrary A_1, A_2 we define $A_1^{-1}(w) = \sup \left\{ \theta : A_1(\theta) \leq w \right\}$

$$\text{and} \quad A_2^{-1}(v) = \sup \left\{ \theta : A_2(\theta) \geq v \right\} .$$

$$\begin{aligned} \text{Then } B'_{n,\lambda} &= \left\{ \underset{\sim}{X} : A_1(\theta_o + \tfrac{\lambda}{n}) \leq X_i \leq A_2(\theta_o + \tfrac{\lambda}{n}) \quad \text{for } i = 1,2,\ldots,n \right\} \\ &= \left\{ \underset{\sim}{X} : W_n \geq A_1(\theta_o + \tfrac{\lambda}{n}),\ V_n \leq A_2(\theta_o + \tfrac{\lambda}{n}) \right\} \\ &= \left\{ A_1^{-1}(W_n) \geq \theta_o + \tfrac{\lambda}{n},\ A_2^{-1}(V_n) \leq \theta_o + \tfrac{\lambda}{n} \right\} \\ &= \left\{ n(Z_n - \theta_o) \geq \lambda \right\} \text{ where } Z_n = \min \left\{ A_1^{-1}(W_n), A_2^{-1}(V_n) \right\} . \end{aligned}$$

Thus the asymptotic expansion (1.1) holds a.e. $P^n_{\theta_o}$.

For arbitrary A_1, A_2, $c(\theta)$ may not be nonzero for all θ . We consider only the case where $c(\theta) > 0$ for all θ . If, for example, at least for one i, $A_i'(\theta) > 0$ for all θ , this condition is satisfied.

Now for any sequence $\left\{ \theta_n \right\}$ satisfying $|n(\theta_n - \theta_o)| \leq C$ for any $C > 0$, and for any $t \geq 0$,

$$P^n_{\theta_n}\left[n(Z_n - \theta_n) \geq t \right]$$

$$= \left[1 - \int_{A_1(\theta_n)}^{A_1(\theta_n + \frac{t}{n})} f(x,\theta_n)dx - \int_{A_2(\theta_n + \frac{t}{n})}^{A_2(\theta_n)} f(x,\theta_n)dx \right]^n$$

$$\longrightarrow e^{-c(\theta_o)t} \quad \text{as} \quad n \longrightarrow \infty ,$$

because

$$\lim_{n \to \infty} n \left[\int_{A_1(\theta_n)}^{A_1(\theta_n + \frac{t}{n})} f(x,\theta_n)dx + \int_{A_2(\theta_n + \frac{t}{n})}^{A_2(\theta_n)} f(x,\theta_n)dx \right]$$

$$= t\, A_1'(\theta_0) f(A_1(\theta_0),\theta_0) - t\, A_2'(\theta_0) f(A_2(\theta_0),\theta_0)$$
$$= t\, c(\theta_0).$$

Thus assumption (A1) of Chapter 1 and the assumption of regularity of Z_n hold and hence the conclusions of all the theorems in Chapter 1 hold.

Case II(a). $S(\theta)$ is an unbounded interval, i.e., $S(\theta) = [A_1(\theta), \infty)$ or $S(\theta) = (-\infty, A_2(\theta)]$ and $A_1'(\theta) \leq 0$, $A_2'(\theta) \geq 0$ for all $\theta \in (H)$.

Proceeding as in case I(a) we can prove that condition (A2) of Chapter1is satisfied for some Z_n which is regular.

Case II(b). $S(\theta) = [A_1(\theta), A_2(\theta)]$ with $A_1'(\theta) \leq 0$ and $A_2'(\theta) \geq 0$ for all $\theta \in (H)$. Here $B_{n,\lambda} = \{\underset{\sim}{X} : W_n > A_1(\theta_0 + \frac{\lambda}{n}),\ V_n < A_2(\theta_0 + \frac{\lambda}{n})\}$, $\lambda \leq 0$, where W_n, V_n are as defined earlier and

$$c(\theta) = A_1'(\theta) f(A_1(\theta),\theta) - A_2'(\theta) f(A_2(\theta),\theta) \leq 0 \quad \text{for all } \theta \in (H) .$$

We consider only the cases where $c(\theta) < 0$ for all $\theta \in (H)$.

We define
$$A_1^{-1}(w) = \inf \{ \theta : A_1(\theta) \leq w \} ,$$
$$A_2^{-1}(v) = \inf \{ \theta : A_2(\theta) \geq v \} .$$

Then proceeding as in Case I(b) we can prove that condition A(2) is satisfied with $Z_n = \max \{ A_1^{-1}(W_n), A_2^{-1}(V_n) \}$ and this Z_n is regular.

2.3 REGRESSION TYPE MODEL

We now consider an example where the observations $X_1, X_2, \ldots, X_n$ are independent but not identically distributed. We consider the model

$$X_t = g(t)\,\theta + e_t , \qquad t = 1,2,\ldots,n$$

where e_t's are i.i.d. random variables having a common density $f(x)$

such that $f(x) > 0$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$, and $g(t)$, $t = 1,2,\ldots$ are values of a non-stochastic variable. We consider only the case where $g(t)$'s are positive. Let $K_n = \sum_{t=1}^{n} g(t)$. We make the following assumptions.

R1 $f(x)$ is continuous on $[0,\infty)$ and twice differentiable on $(0,\infty)$.

R2 (a) $\int |(\log f)'(x)| f(x)dx < \infty$

(b) $\int |(\log f)''(x)| f(x)dx < \infty$

R3 For all $\lambda \geq 0$,

$$\frac{1}{K_n^2} \sum_{t=1}^{n} g^2(t) \, \mathrm{Sup}\left\{ |(\log f)''(e_t+\alpha) - (\log f)''(e_t)| : 0 \leq \alpha \leq \lambda \max_{1\leq t\leq n} g(t) K_n^{-1} \right\}$$

converges to 0 in probability.

R4 As $n \to \infty$,

(a) $\max_{1\leq t\leq n} g(t) / \sum_{t=1}^{n} g(t) \to 0$.

and (b) $\dfrac{\Sigma g^2(t)}{K_n^2} \to 0$.

Assumption R4 is satisfied if, for example, we take $g(t) \equiv t$ or any polynomial in t. Assumption R3 is satisfied for almost all the usual cases.

We fix $\theta_0 \in Ⓗ$, the parameter space. Let P_θ^n be the joint probability distribution of $X_1,\ldots,X_n$ under θ. Expanding at θ_0 by Taylor's theorem we get for all $\lambda \geq 0$,

$$\log \frac{dP^n_{\theta_0 + \lambda K_n^{-1}}}{dP^n_{\theta_0}} = \frac{\lambda}{K_n} \sum_{t=1}^{n} (-g(t))(\log f)'(X_t - g(t)\theta_0)$$

$$+ \frac{\lambda^2}{2K_n^2} \sum_{t=1}^{n} g^2(t)(\log f)''(X_t - g(t)\theta_n'),$$

$$= \lambda \Delta_n + \varepsilon_n \text{ , say}$$

$$\text{on } B_{n,\lambda} = \left[X_t > g(t)(\theta_0 + \lambda K_n^{-1}),\ t = 1,2,\ldots,n\right]$$

$$= \left[K_n\left(\min_{1 \le t \le n} X_t / g(t) - \theta_0\right) > \lambda\right] ,$$

where θ_n' lies between θ_0 and $\theta_0 + \lambda K_n^{-1}$

and $\frac{dP^n_{\theta_0 + \lambda K_n^{-1}}}{dP^n_{\theta_0}} = 0$ a.e. $P^n_{\theta_0}$ on $B^c_{n,\lambda}$.

We shall now verify the following :

(A) $\Delta_n \xrightarrow{P^n_{\theta_0}} f(0)$

(B) $\varepsilon_n \xrightarrow{P^n_{\theta_0}} 0$

(C) $P^n_{\theta_0}(B_{n,\lambda}) \longrightarrow e^{-\lambda f(0)}$ for all $\lambda \geq 0$.

where Δ_n, ε_n and $B_{n,\lambda}$ are as above.

(A) follows from condition R2(a), the law of large numbers for weighted average (see, for example, Jamison, Orey and Pruitt (1965)), condition R4(b) and the fact that $-\int(\log f)'(x)f(x)dx = f(0)$. Condition R2(b) implies that

$$\frac{1}{K_n^2} \Sigma\, g^2(t)(\log f)''\,(X_t - g(t)\theta_0) \xrightarrow{P^n_{\theta_0}} 0 .$$

(B) now follows from condition R3.

To prove (c) we use the following result :

Lemma. Consider a double sequence of real numbers $\{a_{in}\}_n$. If (i) $\sup_{1\le i\le n} |a_{in}| \longrightarrow 0$, (ii) $\sum_{i=1}^{n} |a_{in}|$ is bounded and (iii) $\Sigma_{i=1} a_{in} \longrightarrow a$, then $\prod_{i=1}^{n} (1 - a_{in}) \longrightarrow e^{-a}$.

$$\text{Now, } P^n_{\theta_0}(B_{n,\lambda}) = \prod_{t=1}^{n} \left[1 - F\left(\frac{g(t)u}{\Sigma g(t)}\right)\right] \text{ where } F \text{ is the distribution}$$

function for f. Using continuity of f at 0 and condition R4(a) we can prove that

$$\sum_{t=1}^{n} F\left(\frac{g(t)}{\Sigma g(t)}\, u\right) - u\, f(0) \longrightarrow 0 .$$

Thus (C) is verified to be true.

Also the random variable $Z_n = \min_{1\le t\le n} X_t/g(t)$ is obviously regular since in this case the distribution of $Z_n - \theta$ does not depend on θ.

2.4 ASYMPTOTIC PROPERTIES OF MAXIMUM PROBABILITY ESTIMATORS

Weiss and Wolfowitz (1974) studied the efficiency of maximum probability estimators (m.p.e) for many non-regular cases. They also considered a general case and indeed proved that the m.p.e. is LAM under certain reasonable assumptions. In this section we shall first prove the same result for the family of non-regular cases given in the previous sections by showing that the lower bound to the local asymptotic minimax risk is attained by the m.p.e. We shall consider only 0-1 loss functions;

$$L(X) = \begin{cases} 0, \text{ if } |X| \leq r, \\ 1, \text{ otherwise}, \end{cases} \qquad (2.1)$$

where r is some positive number.

For all the nonregular cases given in the previous sections, the set on which the joint density of the observations $X_1, \ldots, X_n$ under θ is positive, can be expressed as either (a) $\{X : Z_n \geq \theta\}$ or (b) $\{X : Z_n^* \leq \theta\}$. Proceeding as in Weiss and Wolfowitz (1974) we can find statistics $\tilde{\theta}_n$ which are asymptotically " equivalent " (see discussion following (3.4) in Weiss and Wolfowitz (1974)) to the m.p.e.

For Case (a), $\tilde{\theta}_n = Z_n - r K_n^{-1}$,

and for Case (b), $\tilde{\theta}_n = Z_n^* + r K_n^{-1}$,

where K_n is the normalizing factor.

We shall consider only case (a). For case (a), $\int_0^\infty L(x-b)c(\theta_0)e^{-c(\theta_0)x}\,dx$ is minimized at $b = r$. Thus, using results of the previous sections, for $\theta_0 \in (H)$ and all $A > 0$,

$$\lim_{n \to \infty} \sup_{|\theta - \theta_0| \leq AK_n^{-1}} E_\theta L\left[n(Z_n - rK_n^{-1} - \theta)\right]$$

$$= \int_0^\infty L(x - r)c(\theta_0)e^{-c(\theta_0)x}\,dx$$

$$= \int_{2r}^\infty c(\theta_0)e^{-c(\theta_0)x}\,dx$$

and hence the estimator $\tilde{\theta}_n$ is LAM. The treatment of Case (b) is similar.

We shall now prove that the m.p.e. $\tilde{\theta}_n(r)$, if it exists, is equivalent to the estimator $\hat{\theta}_n (= Z_n - rK_n^{-1})$ suggested in Chapter 1 in the sense that their difference converges to zero in probability (see Theorem 1, below).

We consider the set up of Chapter 1. Let $f_n(\underset{\sim}{x},\theta)$ be the density of P_θ^n with respect to some dominating σ-finite measure on $\underline{A}^n$. We consider only Case (a) and assume that the following condition holds a.s. $P_{\theta_0}^n$

(A1)* For any $\lambda \in \mathbb{R}$,

$$\Lambda_n(\lambda) = \begin{cases} \exp \lambda \Delta_n(\theta_0) + \varepsilon_n(\lambda,\theta_0), & \text{if } K_n(Z_n - \theta_0) > \lambda , \\ 0 , & \text{if } K_n(Z_n - \theta_0) < \lambda , \end{cases}$$

where $\Lambda_n(\lambda) = \dfrac{dP^n_{\theta_0 + \lambda K_n^{-1}}}{dP^n_{\theta_0}}$ for $\lambda \in \mathbb{R}$, Z_n is a random variable satisfying

$$Z_n \geq \theta_0 \quad \text{a.s.} \quad P^n_{\theta_0}$$

and $$\lim_{n \to \infty} P^n_{\theta_0}\left[K_n(Z_n - \theta_0) > t\right] = e^{-tc(\theta_0)} \quad \text{for all} \quad t \geq 0 ,$$

$$\Delta_n(\theta_0) \xrightarrow{P^n_{\theta_0}} c(\theta_0) \quad \text{for some} \quad c(\theta_0) > 0$$

and $$\varepsilon_n \xrightarrow{P^n_{\theta_0}} 0 .$$

Thus under assumption (A1)*, for all $\lambda < 0$,

$$\Lambda_n(\lambda) = \exp \left\{ \lambda \Delta_n(\theta_0) + \varepsilon_n(\lambda,\theta_0) \right\} \quad \text{a.s.} \quad P^n_{\theta_0}$$

and hence for all $\lambda < 0$, $\Lambda_n(\lambda) \xrightarrow{P^n_{\theta_0}} e^{\lambda c(\theta_0)}$.

We here assume that

$$\text{(B1)} \quad E_{\theta_o}(\Lambda_n(\lambda)) \longrightarrow e^{\lambda c(\theta_o)} \quad \text{for all } \lambda < 0 .$$

It is to be noted that the above conditions hold for all the non-regular cases considered in Sections 2.2 and 2.3.

Now, the maximum probability estimator $\bar{\theta}_n(r)$ with respect to the loss function (2.1) is that value of d for which the integral

$$\int f_n(\underset{\sim}{X}_n, \theta)\, d\theta$$

over the set $[d - rK_n^{-1}, d + rK_n^{-1}]$ is a maximum.
Here $\underset{\sim}{X}_n$ denotes the observation at the n-th stage. We assume that $f_n(\underset{\sim}{x}, \theta)$ is jointly measurable in $(\underset{\sim}{x}, \theta)$ and a measurable m.p.e. $\bar{\theta}_n(r)$ exists.

<u>Theorem 1</u>. Suppose that the sequence $\{K_n(\bar{\theta}_n - \theta_o)\}$ is relatively compact under $\{P^n_{\theta_o}\}$. Then under assumptions (A1)* and (B1),

$$K_n(\bar{\theta}_n - \theta_o) - K_n(Z_n - rK_n^{-1} - \theta_o) \xrightarrow{P^n_{\theta_o}} 0 \quad \text{as} \quad n \longrightarrow \infty .$$

To prove this theorem we need the following lemma.

<u>Lemma</u>. Set for $\lambda \in \mathbb{R}$,

$$\Lambda_n^*(\lambda) = \begin{cases} \exp(c(\theta_o)\lambda), & \text{if } K_n(Z_n - \theta_o) > \lambda , \\ 0, & \text{if } K_n(Z_n - \theta_o) < \lambda . \end{cases}$$

Then for any $\lambda \in \mathbb{R}$,

$$E_{\theta_o} |\Lambda_n(\lambda) - \Lambda_n^*(\lambda)| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty .$$

Proof of the lemma. The result follows from the fact that

$$|\Lambda_n(\lambda) - \Lambda_n^*(\lambda)| \xrightarrow{P^n_{\theta_0}} 0$$

and $\Lambda_n(\lambda)$ and $\Lambda_n^*(\lambda)$ are uniformly integrable.
This is proved in Ibragimov and Hasminskii (1981) for all $\lambda \geq 0$. Using condition (B1) it can be proved for all $\lambda < 0$ in a similar manner. ///

Proof of Theorem 1. We use the idea of the proof of Theorem 4 in Jaganathan (1982). We shall prove that for all $\delta > 0$,

$$\lim_{n \to \infty} P^n_{\theta_0} \left[|K_n(\bar{\theta}_n - \theta_0) - K_n(Z_n - \bar{r}K_n^{-1} - \theta_0)| > \delta \right] = 0 .$$

Given $\varepsilon > 0$, we choose $K > 0$ sufficiently large such that for all n,

$$P^n_{\theta_0} \left[|K_n(\bar{\theta}_n - \theta_0)| > K - r \right] < \frac{\varepsilon}{4}$$

and
$$P^n_{\theta_0} \left[|K_n(Z_n - rK_n^{-1} - \theta_0)| > K - r \right] < \frac{\varepsilon}{4} .$$

Thus it is enough to prove that for all sufficiently large n,

$$P^n_{\theta_0} (A_n) < \frac{\varepsilon}{2} \tag{2.2}$$

where
$$A_n = \{ |K_n(\bar{\theta}_n - \theta_0) - K_n(Z_n - rK_n^{-1} - \theta_0)| > \delta ,$$
$$|K_n(\bar{\theta}_n - \theta_0)| \leq K-r , \quad |K_n(Z_n - rK_n^{-1} - \theta_0)| \leq K-r \}$$

Since $E_{\theta_0} |\Lambda_n(\lambda) - \Lambda_n^*(\lambda)| \leq 1 + \exp(c(\theta_0)\lambda)$

the above lemma implies that $\int_{-K}^{K} E_{\theta_0} |\Lambda_n(\lambda) - \Lambda_n^*(\lambda)| \, d\lambda \to 0$

and therefore
$$E_{\theta_0} \left[\int_{-K}^{K} |\Lambda_n(\lambda) - \Lambda_n^*(\lambda)| \, d\lambda \right] \to 0 \tag{2.3}$$

Now, if we set

$$B_1 = \left[K_n(\bar{\theta}_n - \theta_0) - r, \; K_n(\bar{\theta}_n - \theta_0) + r \right]$$

and $\quad B_2 = \left[K_n(Z_n - rK_n^{-1} - \theta_o) - r,\ K_n(Z_n - rK_n^{-1} - \theta_o) + r\right]$,

we have $B_1 \subset [-K,K]$, $B_2 \subset [-K,K]$ whenever A_n occurs and hence (2.3) implies that

$$\int_{A_n} \int_{B_i} |\Lambda_n(\lambda) - \Lambda_n^*(\lambda)|\, d\lambda\, dP_{\theta_o}^n \longrightarrow 0 \quad \text{for } i = 1,2. \tag{2.4}$$

Now suppose that (2.2) is not true.

Then $P_{\theta_o}^n(A_n) \geq \frac{\varepsilon}{2}$ $\qquad$ (2.5)

for infinitely many values of n.

From the definition of $\Lambda_n^*(\lambda)$ it can be shown that when the event A_n occurs we have

$$a_o + \int_{B_1} \Lambda_n^*(\lambda)\, d\lambda < \int_{B_2} \Lambda_n^*(\lambda)\, d\lambda$$

where a_- is a positive real number not depending on n.

Then (2.5) implies that for some $a_o > 0$,

$$a_o + \int_{A_n} \int_{B_1} \Lambda_n^*(\lambda)\, d\lambda < \int_{A_n} \int_{B_2} \Lambda_n^*(\lambda)\, d\lambda$$

for infinitely many values of n.

This together with (2.4) implies that

$$\int_{A_n} \int_{B_1} \Lambda_n(\lambda)\, d\lambda < \int_{A_n} \int_{B_2} \Lambda_n(\lambda)\, d\lambda \tag{2.6}$$

for infinitely many values of n.

On the other hand, from the definition of m.p.e.

$$\int_{B_1} \frac{f_n(\underset{\sim}{X}_n, \theta_o + \lambda K_n^{-1})}{f_n(X_n, \theta_o)}\, d\lambda \geq \int_{B_2} \frac{f_n(\underset{\sim}{X}_n, \theta_o + \lambda K_n^{-1})}{f_n(X_n, \theta_o)}\, d\lambda$$

i.e., $$\int_{A_n}\int_{B_1} \Lambda_n(\lambda)\, d\lambda \geq \int_{A_n}\int_{B_2} \Lambda_n(\lambda)\, d\lambda$$

for all n, contradicting (2.6).

Thus (2.2) is true and hence the theorem is proved. ///

Remark. The result (Theorem 1) can also be proved for any loss function of the form

$$L(X) = L(|X|) = M, \quad \text{if} \quad |X| > r ,$$
$$\leq M, \quad \text{if} \quad |X| \leq r ,$$

for any $M, r > 0$. The maximum probability estimate for such a loss function is defined to be that value of d for which the integral

$$\int [M - L(K_n(d - \theta))]\ f_n(\underset{\sim}{X}_n, \theta)\ d\theta$$

is a maximum. The proof follows the same lines as the proof of Theorem 1.

2.5 ASYMPTOTIC PROPERTIES OF BAYES ESTIMATORS FOR REGRESSION TYPE MODEL

The asymptotic properties of Bayes estimators were studied in Ibragimov and Hasminskii (1981) for a large family of non-regular cases when the observations are independent and identically distributed. In this section we consider the regression model of Section 2.3 and using a general result on the asymptotic behaviour of the Bayes estimators (Theorem I.10.2 in Ibragimov and Hasminskii (1981, Ch. I)) we prove the efficiency of the Bayes estimators. For this we make the following assumptions in addition to the assumptions R1-R4 made in Section 2.3:

R5 There exist positive constants a, M_1, M_2 such that for all $x \geq 0$,

$$f(x) \leq M_1 + M_2\, x^a .$$

R6 There exists a constant $C^* > 0$ such that for all $n \geq 1$,

$$\left(\prod_{t=1}^{n} g(t)\right)^{1/n} \Big/ \max_{1 \leq t \leq n} g(t) \geq C^* .$$

(Condition R6 is satisfied, for example, when $g(t)$ is some polynomial in t).

When the parameter set Θ is unbounded we make the following assumption :

R7 $\int f^{1/2}(x - h) f^{1/2}(x)\,dx \leq \left[C_1 |h|\right]^{-\alpha}$

for all h and for some $\alpha > 0$, $C_1 > 0$.

We now consider the family $\{\tilde{\theta}_n\}$ of Bayes estimators with respect to the loss function $L(K_n^{-1}(\theta - a))$ and some prior density q. We assume that L is a subconvex loss function possessing a polynomial majorant and satisfying the following condition :

(C) $\varphi(b) = \int_0^\infty L(x-b) f(0) e^{-f(0)x} dx$ is finite for some b and attains its minimum at the unique point b_0.

Let Q be the set of continuous positive functions on $\mathbb{R}$ possessing a polynomial majorant.

<u>Theorem 2</u>. Let $\tilde{\theta}_n$ be a Bayes estimator with respect to a prior density $q \in Q$ and a loss function $L(K_n(\theta - a))$, where L is a continuous subconvex function possessing a polynomial majorant and satisfying condition (C). Then under conditions R1-R7, the Bayes estimator $\tilde{\theta}_n$ is asymptotically efficient for estimating θ in the sense that uniformly in θ belonging to any compact subset of Θ,

$$\lim_{n \to \infty} E_\theta L\left[K_n(\tilde{\theta}_n - \theta)\right] = \int_0^\infty L(x - b_0) f(0) e^{-f(0)x} dx$$

where the right hand side is the lower bound to the asymptotic risk of an estimator obtained in Theorem 2 of Chapter 1.

To prove this theorem we shall need a general result on asymptotic behaviour of the Bayes estimators due to Ibragimov and Hasminskii (1981, Ch. I). For easy reference we state below the set up and the result of Ibragimov and Hasminskii.

Suppose we have a sequence of experiments $E^n = \{(\underline{X}^n, \underline{A}^n), P_\theta^n, \theta \in Ⓗ\}$ where Ⓗ is an open subset of $\mathbb{R}^k$, $k \geq 1$.
We set

$$\Lambda_{n,\theta}(u) = \frac{dP^n_{\theta + K_n^{-1} u}}{dP^n_\theta},$$

where $K_n(\uparrow \infty)$ is some normalizing factor. The random function $\Lambda_{n,\theta}(u)$ is defined on the set $U_n = K_n(Ⓗ - \theta)$. Below we shall denote by G the set of families $\{g_n(y)\}$ of functions g_n with the following properties :

(1) For each fixed $n \geq 1$, $g_n(y) \uparrow \infty$ is a positive function on $[0,\infty)$.

(2) For any $N > 0$, $\lim\limits_{\substack{y \to \infty \\ n \to \infty}} y^N e^{-g_n(y)} = 0$.

The following theorem is due to Ibragimov and Hasminskii (1981, Ch. I).

Theorem (Ibragimov and Hasminskii). Let $\tilde{\theta}_n$ be a Bayes estimator with respect to a prior density $q \in Q$ and the loss function $L(K_n(\theta - a))$, where L is a continuous subconvex loss function possessing a polynomial majorant. Assume that the random functions $\Lambda_{n,\theta}(u)$ possess the following properties :

(1) For any compact set $K' \subset \textcircled{H}$ there correspond nonnegative numbers $a(K) = a$ and $B(K) = B$ and functions $g_n^K(y) = g_n(y)$, $\{g_n\} \in G$ such that

(1.1) For some $\alpha > 0$ and for all $\theta \in K$,

$$\sup_{\substack{|u_1|\le R,\, |u_2|\le R \\ u_1, u_2 \in U_{n,\theta}}} |u_2 - u_1|^{-\alpha} E_\theta^{(n)} \left| \Lambda_{n,\theta}^{1/2}(u_2) - \Lambda_{n,\theta}^{1/2}(u_1) \right|^2 \le B(1+R^a).$$

(1.2) For all $\theta \in K$ and $u \in U_{n,\theta}$

$$E_\theta^{(n)} \Lambda_{n,\theta}^{1/2}(u) \le e^{-g_n(|u|)} .$$

(2) The finite dimensional distributions of the random functions $\Lambda_{n,\theta}(u)$ uniformly in $\theta \in K$ converge to the finite dimensional distributions of the random functions $\Lambda_\theta(u) = \Lambda(u)$.

(3) The random function

$$\psi(s) = \int_{\mathbb{R}^k} L(s-u) \frac{\Lambda(u)}{\int_{\mathbb{R}^k} \Lambda(v)dv} du$$

attains its (absolute) minimum at the unique point $\tau(\theta)$.

Then the distribution of $K_n(\tilde{\theta}_n - \theta)$ converges uniformly in $\theta \in K$ to the distribution of $\tau(\theta)$ and we have uniformly in $\theta \in K$,

$$\lim_{n \to \infty} E_\theta^{(n)} L\left[K_n(\tilde{\theta}_n - \theta) \right] = EL(\tau(\theta)).$$

<u>Proof of Theorem 2</u>. We verify all the conditions of the Theorem (Ibragimov and Hasminskii). We fix some $\theta \in \textcircled{H}$. For $u \in \mathbb{R}$

$$\Lambda_{n,\theta}(u) = \frac{\prod_{t=1}^{n} f(X_t - g(t)\,\theta - g(t)\, K_n^{-1}\, u)}{\prod_{t=1}^{n} f(X_t - g(t)\,\theta)} .$$

First of all we note that the marginal distributions of the process $\Lambda_{n,\theta}(u)$ do not depend on θ. Then for any $u_1 < u_2$,

$$E_\theta|\Lambda^{1/2}_{n,\theta}(u_2) - \Lambda^{1/2}_{n,\theta}(u_1)|^2$$

$$\leq 2\left[1 - \prod_{t=1}^{n}\int f^{1/2}(x_t - g(t)K_n^{-1}u_2)f^{1/2}(x_t - g(t)K_n^{-1}u_1)dx_t\right]$$

$$\leq 2\sum_{t=1}^{n}\left[1 - \int f^{1/2}(x - g(t)K_n^{-1}u_2)f^{1/2}(x - g(t)K_n^{-1}u_1)dx\right]$$

$$\left[\text{since for } 0 \leq \rho_1,\rho_2,\ldots,\rho_n \leq 1,\ 1 - \rho_1\rho_2\ldots\rho_n \leq \sum_{i=1}^{n}(1-\rho_i)\right]$$

$$= \sum_{t=1}^{n}\int |f^{1/2}(x - g(t)K_n^{-1}u_2) - f^{1/2}(x - g(t)K_n^{-1}u_1)|^2\,dx \qquad (2.7)$$

Now

$$|f^{1/2}(x - g(t)K_n^{-1}u_2) - f^{1/2}(x - g(t)K_n^{-1}u_1)|^2\,dx.$$

$$\leq \int |f(x - g(t)K_n^{-1}u_2) - f^{1/2}(x - g(t)K_n^{-1}u_1)|\,dx$$

$$\left[\text{since for any } \alpha,\beta > 0\ (\sqrt{\alpha} - \sqrt{\beta})^2 \leq |\alpha - \beta|\right]$$

$$= \int_{g(t)K_n^{-1}u_1}^{g(t)K_n^{-1}u_2} f(x - g(t)K_n^{-1}u_1)dx + \int_{g(t)K_n^{-1}u_2}^{\infty}\left|\int_{g(t)K_n^{-1}u_1}^{g(t)K_n^{-1}u_2} f'(x-s)\,ds\right|dx$$

$= I_1 + I_2$, say.

$$I_2 = \int_{g(t)K_n^{-1}u_1}^{g(t)K_n^{-1}u_2}\left\{\int_{g(t)K_n^{-1}u_2 - s}^{\infty}|f'(x)|dx\right\}ds$$

$$\leq g(t)K_n^{-1}(u_2 - u_1)\int_0^{\infty}|f'(x)|dx.$$

$$= g(t)K_n^{-1}(u_2 - u_1)\,M, \text{ say}$$

where $M = \int_0^{\infty}|f'(x)|dx < \infty$ by assumption R2(a).

By assumption R5, for all $u_1 < u_2$ such that $|u_1| \leq R$, $|u_2| \leq R$ we have

$$I_1 \leq g(t)K_n^{-1}(u_2 - u_1)\left[M_1 + M_2(2R)^a\right].$$

Therefore, from (2.7)

$$E_\theta|\Lambda_{n,\theta}^{1/2}(u_2) - \Lambda_{n,\theta}^{1/2}(u_1)|^2 \leq (u_2 - u_1)\left[M + M_1 + M_2 2^a R^a\right]$$

i.e.,
$$\sup_{|u_1|\leq R,\, |u_2|\leq R} |u_2 - u_1|^{-1} E_\theta|\Lambda_{n,\theta}^{1/2}(u_2) - \Lambda_{n,\theta}^{1/2}(u_1)|^2 \leq B(1 + R^a)$$

for some $B > 0$ and for all $\theta \in \textcircled{H}$.

Thus condition (1.1) of the Theorem (Ibragimov and Hasminskii) is satisfied.

Now,

$$E_\theta\Lambda_{n,\theta}^{1/2}(u) = \prod_{t=1}^{n}\left\{1 - \frac{1}{2}\int|f^{1/2}(x - g(t)K_n^{-1}u) - f^{1/2}(x)|^2\,dx\right\}$$

$$\leq \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}\int|f^{1/2}(x - g(t)K_n^{-1}u) - f^{1/2}(x)|^2\,dx\right\}$$

$$(\text{since } 1 - \rho \leq e^{-\rho}).$$

We choose $A > 0$ sufficiently small such that whenever $0 \leq x \leq A$ we have $f(x) \geq \frac{1}{2}f(0)$.

For $u \geq 0$, $\int|f^{1/2}(x - g(t)K_n^{-1}u) - f^{1/2}(x)|^2\,dx$

$$\geq \int_0^{g(t)K_n^{-1}u} f(x)\,dx$$

and for $u \leq 0$, $\int|f^{1/2}(x - g(t)K_n^{-1}(u) - f^{1/2}(x)|^2\,dx$

$$\geq \int_{-g(t)K_n^{-1}|u|}^{0} |f^{1/2}(x + g(t)K_n^{-1}|u|) - f^{1/2}(x)|^2\,dx.$$

$$= \int_{-g(t)K_n^{-1}|u|}^{0} f(x + g(t)K_n^{-1}|u|)dx.$$

Thus for $\max_{1\le t\le n} g(t)K_n^{-1}|u| \le A$ we have

$$E_\theta \Lambda_{n,\theta}^{1/2}(u) \le \exp\left\{-\tfrac{1}{4} f(0)|u|\right\}. \tag{2.8}$$

Also by assumption R7, for all $u \in \mathbb{R}$,

$$E_\theta \Lambda_{n,\theta}^{1/2}(u) \le [c_1|u|]^{-n\alpha} \left[\left(\prod_{t=1}^{n} g(t)\right)^{1/n} K_n^{-1}\right]^{-n\alpha} \tag{2.9}$$

Fix any $r > 0$. We want to prove that

$$\lim_{\substack{|u|\to\infty \\ n\to\infty}} |u|^r E_\theta \Lambda_{n,\theta}^{1/2}(u) = 0. \tag{2.10}$$

From (2.9), for $\max_{1\le t\le n} g(t)K_n^{-1}|u| > A$,

$$|u|^r E_\theta \Lambda_{n,\theta}^{1/2}(u)$$

$$\le |u|^r \left[c_1|u|K_n^{-1} \max_{1\le t\le n} g(t)\right]^{-n\alpha} \left[\left(\prod g(t)\right)^{1/n} / \max_{1\le t\le n} g(t)\right]^{-n\alpha}$$

$$= c_1^{-r}\left[c_1|u|K_n^{-1} \max_{1\le t\le n} g(t)\right]^{-n\alpha+r} \left[K_n / \max_{1\le t\le n} g(t)\right]^{r} \left[\left(\prod g(t)\right)^{1/n} / \max_{1\le t\le n} g(t)\right]^{-n\alpha}$$

$$< c_1^{-r}(c_1A)^{-n\alpha+r} \left[K_n / \max_{1\le t\le n} g(t)\right]^{r} \left[\left(\prod_1^n g(t)\right)^{1/n} / \max_{1\le t\le n} g(t)\right]^{-n\alpha}$$

(we choose n so large that $-n\alpha+r < 0$)

$$= A^r \frac{n^r}{\left[c_1 A \dfrac{\left(\prod g(t)\right)^{1/n}}{\max_{1\le t\le n} g(t)}\right]^{n\alpha}} \qquad \left[\text{since } K_n \le n \max_{1\le t\le n} g(t)\right].$$

and this converges to 0 (as $n \to \infty$) by assumption R6. This result and (2.8) give (2.10).

Now proceeding as in Section 2.3 we can express $\Lambda_{n,\theta}(u)$ for all $u \in \mathbb{R}$ as

$$\Lambda_{n,\theta}(u) = \begin{cases} \exp\left\{f(0)u + \varepsilon_n\right\}, & \text{if } \tau_n > u, \\ 0, & \text{if } \tau_n \leq u, \end{cases}$$

where $\varepsilon_n \xrightarrow{P_\theta^n} 0$ and τ_n is a random variable converging in distribution to a random variable τ with density $f(0)e^{-f(0)x}$ on $(0, \infty)$. This is proved for all $u \geq 0$ in Section 2.3. The proof for $u < 0$ is similar to that for the case $u \geq 0$. Then it can be easily shown that the marginal distributions of the process $\Lambda_{n,\theta}(u)$, $u \in \mathbb{R}$ converge to the marginal distributions of the process

$$\Lambda(u) = \begin{cases} e^{f(0)u}, & \text{if } \tau > u, \\ 0, & \text{if } \tau \leq u. \end{cases}$$

Also the random function

$$\psi(s) = \int L(s-u)\Lambda(u)\,du$$

attains its minimum value at the unique point $s = \tau - b_0$. Thus all the conditions of the Theorem (Ibragimov and Hasminskii) are satisfied and Theorem 2 is proved. ///

CHAPTER 3

THE ASYMPTOTIC BEHAVIOUR OF POSTERIOR DISTRIBUTIONS AND BAYES ESTIMATORS IN NON-REGULAR CASES

3.1 INTRODUCTION

Let $\{x_1, x_2, \dots\}$ be independent observations with a common distribution having a density which depends on a real or k-dimensional parameter θ. Suppose a prior distribution of θ is given. Then under suitable conditions, the posterior distribution of θ given the observations $x_1, x_2, \dots, x_n$ is very close to normal distribution if n is large. This was first observed by Laplace in 1774 and more recently, by Bernstein (1917) and also by von Mises (1931) and the result is referred to as Bernstein - von Mises theorem. For independent and identically distributed observations a rigorous proof was given by LeCam (1953, 1958) and his result is an improvement over earlier results. Various modifications and extensions of this result have been made by several authors including Bickel and Yahav (1969), Chao (1970), Borwanker, Kallianpur and Prakasa Rao (1971). However, none of these authors considered the non-regular cases. In this chapter we study the limiting behaviour of posterior distribution and Bayes estimators for a class of non-regular cases for which the support of the density depends on the parameter θ. The limiting posterior distribution is, however, not normal. In Section 3.2 we prove the convergence of the posterior distribution to some exponential distribution. The asymptotic behaviour of Bayes estimators is studied in Section 3.3. In Section 3.4 Bernstein - von Mises theorem in the regular case is reexamined. It is shown that the results of Bickel and Yahav (1969) and Chao (1970) can be improved upon by relaxing an assumption and this increases the scope of applicability of their results to include the various standard examples.

3.2 LIMIT OF POSTERIOR DISTRIBUTIONS IN NONREGULAR CASES

We consider the set up of Section 2.2 (Chapter 2) and in Theorem 1 of this section, find the limit of posterior distributions under two additional assumptions. Let $x_1, x_2, \ldots, x_n$ be a random sample from a distribution P_θ with density function $f(x, \theta)$ on $\mathbb{R}$ with respect to Lebesgue measure where $\theta \in \Theta$, an open interval of $\mathbb{R}$. We fix $\theta_0 \in \Theta$, which may be regarded as the true parameter point. We assume that $f(x, \theta)$ is strictly positive for all x in a closed interval (bounded or unbounded) $S(\theta)$ depending on θ and is zero outside $S(\theta)$. Let $A_1(\theta)$, $A_2(\theta)(A_1 < A_2)$ be the boundaries of $S(\theta)$. As in Section 2.2, we consider the following cases :

Case I : The support $S(\theta)$ is decreasing in θ, i.e.,

$$S(\theta_2) \subset S(\theta_1) \quad \text{whenever} \quad \theta_2 > \theta_1$$

Case II: The support $S(\theta)$ is increasing in θ, i.e.,

$$S(\theta_2) \supset S(\theta_1) \quad \text{whenever} \quad \theta_2 > \theta_1 .$$

We now make the following assumptions on the density $f(x, \theta)$:

(A1) $A_1(\theta)$ and $A_2(\theta)$ are continuously differentiable functions of θ (if not infinity).

(A2) On the set $\{(x, \theta) : x \in S(\theta)\}$, $f(x, \theta)$ is jointly continuous in (x, θ).

(A3) The derivatives $\dfrac{\partial \log f(x, \theta)}{\partial \theta}$, $\dfrac{\partial^2 \log f(x, \theta)}{\partial \theta^2}$ exist for all (x, θ) in $\{(x, \theta) : A_1(\theta) < x < A_2(\theta)\}$.

(A4) For all $\theta_0 \in \textcircled{H}$, there exists a neighbourhood $N(\theta_0)$ of θ_0 and a constant $D(\theta_0) > 0$ such that for all $\theta \in N(\theta_0)$,

$$\left| \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right| \leq D(\theta_0)$$

for all x for which the derivative exists.

(A5) For all $\theta \in \textcircled{H}$, $E_\theta \left[\frac{\partial \log f(X, \theta)}{\partial \theta} \right] = c(\theta)$ is finite and not equal to zero.

(A6) For Case I:

$$E_{\theta_0} \sup \left\{ \log f(X,\theta) - \log f(X,\theta_0) : \theta < \theta_0 - \delta, \theta \in \textcircled{H} \right\} < 0$$

for sufficiently large $\delta > 0$.

For Case II :

$$E_{\theta_0} \sup \left\{ \log f(X,\theta) - \log f(X,\theta_0) : \theta > \theta_0 + \delta , \theta \in \textcircled{H} \right\} < 0$$

for sufficiently large $\delta > 0$.

(A7) $\lim_{\rho \to 0} E_{\theta_0} \log f(X, \theta, \rho) = E_{\theta_0} \log f(X, \theta)$

where $f(x, \theta, \rho)$ is the supremum of $f(x, \theta')$ with respect to $\theta' \in \textcircled{H}$ when $|\theta - \theta'| \leq \rho$.

Below we shall prove results for Case I only. The treatment of Case II is similar.

It has been shown in Section 2.2 that for Case I, the set $\left\{ (x_1,\ldots,x_n) : x_i \in S(\theta) \text{ for } i = 1,2,\ldots,n \right\}$ can be expressed as $\left\{ Z_n(\underset{\sim}{x}) \geq \theta \right\}$ for some statistic Z_n and a slight modification of Z_n is locally asymptotically minimax estimator of θ . Also for this case, $c(\theta_0) \geq 0$ and hence by (A5) $c(\theta_0) > 0$.

We now consider a prior probability distribution with density $\lambda(.)$ with respect to Lebesgue measure. The posterior density of θ given the observations $x_1, x_2, \ldots, x_n$ is

$$g_n(\theta | x_1, \ldots, x_n) = \frac{\prod_{i=1}^{n} f(x_i, \theta)\, \lambda(\theta)}{\int \prod_{i=1}^{n} f(x_i, \eta) \lambda(\eta) d\eta} .$$

We will consider the posterior density of $t = n(\theta - \hat{\theta}_n)$, where $\hat{\theta}_n = Z_n - \frac{b}{n}$ for some $b > 0$. The posterior density of $t = n(\theta - \hat{\theta}_n)$ is given by

$$g_n^*(t | x_1, x_2, \ldots, x_n) = c_n^{-1} \gamma_n(t) \lambda(\hat{\theta}_n + tn^{-1}),$$

where $\gamma_n(t) = \dfrac{\prod f(x_i, \hat{\theta}_n + tn^{-1})}{\prod f(x_i, \hat{\theta}_n)}$, $c_n = \int \gamma_n(t) \lambda(\hat{\theta}_n + tn^{-1}) dt.$

Note that $\prod f(x_i, \theta)$ is positive if each $x_i \in S(\theta)$, i.e., if $\theta \leq Z_n(\underset{\sim}{x})$. Thus, $\gamma_n(t) = 0$ for $t > b$.

We now consider a weight function $H(t) = \hat{H}(|t|)$ satisfying the following conditions:

(A8) (a) $H(t)$ is nonnegative and there exists $\varepsilon > 0$ such that for all $b > 0$,

$$\int_{-\infty}^{0} H(t) \exp\left\{t(c(\theta_0) - \varepsilon)\right\} dt + \int_{0}^{b} H(t) \exp\left\{t(c(\theta_0) + \varepsilon)\right\} dt < \infty$$

(b) For every $u > 0$ and $\delta > 0$,

$$e^{-nu} \int_{|t| > \delta n} H(t) \lambda(\hat{\theta}_n + tn^{-1}) dt \longrightarrow 0 \text{ a.s. as } n \longrightarrow \infty .$$

Unless otherwise specified, all probability statements are with respect to P_{θ_0} - measure. The phrase a.s. P_{θ_0} will be omitted if it is clear from the context.

<u>Theorem 1.</u> Consider Case I. Under assumptions (A1) - (A7), for any weight function H satisfying (A8) and any prior probability density $\lambda(.)$ over $(\bar{H})$ which is continuous and positive in an open neighbourhood of θ_o,

$$\lim_{n \to \infty} \int_{\mathbb{R}} H(t)\,|g_n^*(t|x_1,x_2,\ldots,x_n) - g_{\theta_o}(t)|\,dt = 0 \quad \text{a.s.} \quad P_{\theta_o}$$

where $g_n^*(t|x_1,\ldots,x_n)$ is the posterior density of $t = n(\theta - \hat{\theta}_n)$ given the observations $x_1,x_2,\ldots,x_n$

$$\text{and} \quad g_{\theta_o}(t) = \begin{cases} c(\theta_o)\exp\left\{ c(\theta_o)(t-b)\right\} & \text{for } t < b, \\ 0, & \text{otherwise.} \end{cases}$$

<u>Proof.</u> <u>Step 1.</u> It can be easily shown that $Z_n \to \theta_o$ a.s. as $n \to \infty$. For $t > b$, $\gamma_n(t) = 0$.

For $t < b$, $\log \gamma_n(t) = \frac{t}{n} \sum_{i=1}^{n} \frac{\partial \log f(x_i,\theta)}{\partial \theta}\Big|_{\theta_n'}$

for some θ_n' lying between $\hat{\theta}_n$ and $\hat{\theta}_n + \frac{t}{n}$.

Using assumption (A5) and strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(x_i,\theta)}{\partial \theta}\Big|_{\theta_o} \longrightarrow c(\theta_o) \quad \text{a.s.}$$

Also

$$\left|\frac{1}{n}\Sigma \frac{\partial \log f(x_i,\theta)}{\partial \theta}\Big|_{\theta_n'} - \frac{1}{n}\Sigma \frac{\partial \log f(x_i,\theta)}{\partial \theta}\Big|_{\theta_o}\right|$$

$$\leq \frac{1}{n}\Sigma \left|\frac{\partial^2 \log f(x_i,\theta)}{\partial \theta^2}\Big|_{\theta_n''}\right| \cdot |\theta_n' - \theta_o| \qquad (3.1)$$

for some θ_n'' lying between θ_n' and θ_o.

By assumption (A4), there exists $\delta > 0$ such that whenever $|t| \leq \delta n$ the right hand side of (3.1) is (for all sufficiently large n depending on the sample point) less than or equal to

$$\begin{aligned}
& D(\theta_o)\ |\theta_n' - \theta_o| \\
& \leq D(\theta_o)(|\theta_n' - \hat{\theta}_n| + |\hat{\theta}_n - \theta_o|) \\
& \leq D(\theta_o)(|t|n^{-1} + |\hat{\theta}_n - \theta_o|) \\
& \leq D(\theta_o)(\delta + |\hat{\theta}_n - \theta_o|) \ .
\end{aligned}$$

Since $\hat{\theta}_n \longrightarrow \theta_o$, given any $\varepsilon > 0$, we can choose $\delta > 0$ sufficiently small so that for all sufficiently large n,

$$\left.\begin{array}{lll}
\log \gamma_n(t) < t(c(\theta_o) - \varepsilon) & \text{whenever} & -\delta n \leq t \leq 0 \\
\text{and } \log \gamma_n(t) < t(c(\theta_o) + \varepsilon) & \text{whenever} & 0 \leq t < b
\end{array}\right\} \quad (3.2)$$

Also for each t, $\gamma_n(t) \longrightarrow \gamma(t)$ as $n \longrightarrow \infty$ where $\gamma(t)$ is given by

$$\gamma(t) = \begin{cases} \exp\,(tc(\theta_o)) & \text{for } t < b\ , \\ 0, & \text{otherwise}\ . \end{cases}$$

<u>Step 2</u>. We shall prove that for sufficiently small $\delta > 0$,

$$\lim_{n \to \infty} \int_{|t| \leq \delta n} H(t) \left| \gamma_n(t) \lambda(\hat{\theta}_n + tn^{-1}) - \lambda(\theta_o)\gamma(t) \right| dt = 0 \quad \text{a.s.}$$

We have

$$\begin{aligned}
& \int_{|t| \leq \delta n} H(t) \left| \gamma_n(t) \lambda(\hat{\theta}_n + tn^{-1}) - \lambda(\theta_o)\gamma(t) \right| dt \\
& \leq \int_{-\delta n}^{b} H(t)\, \gamma_n(t) \left| \lambda(\hat{\theta}_n + tn^{-1}) - \lambda(\theta_o) \right| dt \\
& + \int_{-\delta n}^{b} H(t)\ \lambda(\theta_o) \left| \gamma_n(t) - \gamma(t) \right| dt.
\end{aligned}$$

For the first integral the integrand is dominated by some integrable function for sufficiently small $\delta > 0$ and the integrand converges to zero for each t as $n \to \infty$. This follows from (3.2), assumption (A8) and continuity of λ at θ_0. Hence by dominated convergence theorem the first integral converges to zero. The second integral also converges to zero by similar argument.

Step 3. We shall prove that for all $\delta > 0$,

$$\lim_{n \to \infty} \int_{|t|>\delta n} H(t) \left| \mathcal{V}_n(t)\lambda(\hat{\theta}_n + tn^{-1}) - \lambda(\theta_0)\mathcal{V}(t) \right| dt = 0 \quad \text{a.s.}$$

We have

$$\int_{|t|>\delta n} H(t) \left| \mathcal{V}_n(t)\lambda(\hat{\theta}_n + tn^{-1}) - \lambda(\theta_0)\mathcal{V}(t) \right| dt$$

$$\leq \int_{t<-\delta n} \mathcal{V}_n(t)H(t)\lambda(\hat{\theta}_n + tn^{-1})dt + \int_{t<-\delta n} \lambda(\theta_0)H(t)\mathcal{V}(t)dt$$

$$= I_{1n} + I_{2n} \text{ (say).}$$

By integrability of $H(t)\mathcal{V}(t)$, the second part I_{2n} converges to zero a.s. We shall now prove that there exists $\varepsilon^*(\delta) > 0$ such that for all sufficiently large n,

$$\sup_{t<-\delta n} \mathcal{V}_n(t) < \exp\left\{ -n\varepsilon^*(\delta) \right\} \quad \text{a.s.} \tag{3.3}$$

If (3.3) is true, $I_{1n} \to 0$ a.s. by assumption (A8).

To prove (2.3) we write

$$\frac{1}{n} \log \mathcal{V}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \log f(x_i, \hat{\theta}_n + \tfrac{t}{n}) - \log f(x_i, \theta_0) \right\}$$

$$- \frac{1}{n} \sum_{i=1}^{n} \left\{ \log f(x_i, \hat{\theta}_n) - \log f(x_i, \theta_0) \right\}$$

$$= A_n + B_n \text{ (say).}$$

It is easy to prove that $B_n \longrightarrow 0$ a.s.

For $t < -\delta n$, $\hat{\theta}_n + tn^{-1} - \theta_o < -\frac{\delta}{2}$ for all sufficiently large n and therefore $A_n \leq \underset{\theta - \theta_o < -\frac{\delta}{2}}{\text{Sup}} \frac{1}{n} \sum_{i=1}^{n} \left\{ \log f(x_i, \theta) - \log f(x_i, \theta_o) \right\}$.

By assumption (A6) we can get $\delta_o > \delta$ such that

$$E_{\theta_o} \text{ Sup} \left\{ \log f(X, \theta) - \log f(X, \theta_o) : \theta < \theta_o - \delta_o, \theta \in \textcircled{H} \right\} < 0.$$

Set $\textcircled{H}_o = \left\{ \theta \in \textcircled{H} : \theta < \theta_o - \delta_o \right\}$,

$\textcircled{H}_1 = \left\{ \theta \in \textcircled{H} : \theta_o - \delta_o \leq \theta \leq \theta_o - \frac{\delta}{2} \right\}$.

For each $\theta \in \textcircled{H}_1$, we can get $\rho_\theta > 0$ such that

$$E_{\theta_o} \log f(X, \theta, \rho_\theta) < E_{\theta_o} \log f(X, \theta_o)$$

where $f(x, \theta, \rho_\theta)$ is as defined earlier (see (A7)). This is possible by assumption (A7) and the fact that

$$E_{\theta_o} \log f(X, \theta) < E_{\theta_o} \log f(X, \theta_o) \quad \text{for } \theta \neq \theta_o .$$

Since the set $\textcircled{H}_1$ is compact there exists a finite number of points $\theta_1, \ldots, \theta_k \in \textcircled{H}$ such that $\bigcup_{j=1}^{k} (\theta_j - \rho_{\theta_j}, \theta_j + \rho_{\theta_j})$ covers $\textcircled{H}_1$ and

$$E_{\theta_o} \log f(X, \theta_j, \rho_{\theta_j}) < E_{\theta_o} \log f(X, \theta_o) \text{ for } j = 1, 2, \ldots, k \qquad (3.4)$$

Now for all $t < -\delta n$ and for all sufficiently large n,

$$A_n \leq \text{Sup} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\log f(x_i, \theta) - \log f(x_i, \theta_o) \right] \right.$$

$$\left. \theta \in \textcircled{H}_o \cup \left\{ \bigcup_{j=1}^{k} (\theta_j - \rho_{\theta_j}, \theta_j + \rho_{\theta_j}) \right\} \right\}$$

$$\leq \operatorname{Max} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\sup_{\theta \in (H)_0} \log f(x_i, \theta) - \log f(x_i, \theta_0) \right], \right.$$

$$\left. \frac{1}{n} \sum_{i=1}^{n} \left[\log f(x_i, \theta_j, \rho_{\theta_j}) - \log f(x_i, \theta_0) \right], \ j = 1,2,\ldots,k \right\}.$$

By assumption (A6) and (3.4), using strong law of large number we can get $\varepsilon(\delta) > 0$ such that the right hand side is less than $-\varepsilon(\delta)$ a.s. for all sufficiently large n. Since $\beta_n \longrightarrow 0$ a.s. this proves (3.3).

Step 4. We shall now prove the theorem using the results proved in step 2 and step 3. The results proved in step 2 and step 3 imply that

$$\int H(t) \left| \gamma_n(t)\lambda(\hat{\theta}_n + \tfrac{t}{n}) - \lambda(\theta_0)\gamma(t) \right| dt = 0 \quad \text{a.s.} \tag{3.5}$$

Putting $H(t) \equiv 1$ which satisfies (A8) trivially, we get

$$c_n = \int \gamma_n(t)\lambda(\hat{\theta}_n + \tfrac{t}{n})dt \longrightarrow \lambda(\theta_0)\int \gamma(t)dt$$
$$= \frac{\lambda(\theta_0)}{c(\theta_0)} \exp\left\{ bc(\theta_0) \right\}. \tag{3.6}$$

Hence

$$\int_{\mathbb{R}} H(t) \left| g_n^*(t|x_1,x_2,\ldots,x_n) - g_{\theta_0}(t) \right| dt$$
$$\leq \int H(t)c_n^{-1} \left| \gamma_n(t)\lambda(\hat{\theta}_n + tn^{-1}) - \lambda(\theta_0)\gamma(t) \right| dt$$
$$+ \int H(t) \left| c_n^{-1}\lambda(\theta_0) - c(\theta_0)\exp\left\{ -bc(\theta_0) \right\} \right| \gamma(t)dt$$

and these two terms converge to zero a.s. by (3.5) and (3.6). ///

Remark 1. If condition (A6) is not satisfied, we can still get a weaker version of the theorem. In such cases it can be proved that under assumptions (A1) – (A5), there exists $\delta > 0$ such that for any prior probability density $\lambda(.)$ over $(\theta_0 - \delta, \theta_0 + \delta)$ which is continuous and positive in a neighbourhood of θ_0, and for any weight function H satisfying (A8) (a), we have

$$\lim_{n \to \infty} \int_{\mathbb{R}} H(t) \left| g_n^*(t|x_1,\ldots,x_n) - g_{\theta_0}(t) \right| dt = 0 \quad \text{a.s.}$$

where $g_n^*(t|x_1,\ldots,x_n)$ and $g_{\theta_0}(t)$ are as defined earlier. See Theorem 1(a) of Section 3.3 in this context.

Remark 2. Assumption (A7) is given in Wald (1949) as a lemma and is proved therein under mild conditions. The idea of the proof of (3.3) in step 3 is essentially due to Wald (1949).

Remark 3. Assumption (A8) (a) is satisfied for the functions $H(t) \equiv |t|^m$, $m \geq 0$. If $\int |\theta|^m \lambda(\theta) d\theta < \infty$ for some integer $m \geq 0$, then (A8) (b) is satisfied for the function $H(t) \equiv |t|^m$ (see, for example, Borwanker et al. (1971) or Basawa and Prakasa Rao (1980)).

3.3 ASYMPTOTIC BEHAVIOUR OF BAYES ESTIMATORS

In this section, we shall give applications of Theorem 1 to the asymptotic theory of Bayes estimation. We consider the set up of Section 3.2. Let $L(.)$ be a loss function. Since n is the normalizing factor in this case, it is reasonable to specify the loss by $L(n(T_n - \theta))$ when T_n is the estimator. Now a Bayes estimator $\tilde{\theta}_n$ is an estimator which minimizes

$$\int L(n(a - \theta)) g_n(\theta|x_1,\ldots,x_n) d\theta \qquad (3.7)$$

with respect to $a \in (\overline{H})$ for all sequences $(x_1,x_2,\ldots)$. Here $g_n(\theta|x_1,\ldots,x_n)$ denotes the posterior density of θ given $x_1,\ldots,x_n$. We shall here assume that such a measurable Bayes estimator $\tilde{\theta}_n$ of θ exists.

Theorem 2. Consider Case I and suppose that assumptions (A1) - (A7) are satisfied. Let the prior probability density be positive and continuous in an open neighbourhood of θ_0 . Let $\tilde{\theta}_n$ be a Bayes estimator of θ with respect to a loss function $L(.)$ satisfying the following conditions :

(i) $L(t) = L(-t) \geq 0$ for all t

and $L(t)$ is a nondecreasing function of $|t|$

(ii) L is lower semi-continuous

i.e., $\{t : L(t) \leq c\}$ is closed for all $c \geq 0$

(iii) $\int_0^\infty L(t-b)c(\theta) \exp\left\{-c(\theta)t\right\} dt$ has a strict minimum at $b = b(\theta) > 0$ and $b(\theta)$ is a continuous function of θ

(iv) Condition (A8) is satisfied with $H(t)$ replaced by $L(t)$.

Then as $n \to \infty$

(a) $n(\tilde{\theta}_n - \hat{\theta}_n) \to 0$ a.s.

where $\hat{\theta}_n = Z_n - \frac{b(Z_n)}{n}$ and Z_n is as defined in Section 3.2.

(b) $\tilde{\theta}_n \to \theta_0$ a.s.

and $\mathcal{L}\left\{n(\tilde{\theta}_n - \theta_0) | P^n_{\theta_0}\right\} \Rightarrow \mathcal{L}\left\{X - b(\theta_0)\right\}$

where X is a random variable with a distribution having density

$$f(x) = \begin{cases} c(\theta_0)\exp\left\{-c(\theta_0)x\right\}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(c) $\int L\left[n(\tilde{\theta}_n - \theta)\right] g_n(\theta | x_1, x_2, \ldots, x_n) d\theta$
$\to \int_0^\infty L(t - b(\theta_0)) c(\theta_0) \exp\left\{-c(\theta_0)t\right\} dt$

where $g_n(\cdot | x_1, \ldots, x_n)$ is the posterior density as defined in Section 3.2.

Theorem 2. Consider Case I and suppose that assumptions (A1) - (A7) are satisfied. Let the prior probability density be positive and continuous in an open neighbourhood of θ_0. Let $\widetilde{\theta}_n$ be a Bayes estimator of θ with respect to a loss function $L(.)$ satisfying the following conditions :

(i) $L(t) = L(-t) \geq 0$ for all t
and $L(t)$ is a nondecreasing function of $|t|$

(ii) L is lower semi-continuous
i.e., $\{t : L(t) \leq c\}$ is closed for all $c \geq 0$

(iii) $\int_0^\infty L(t-b)c(\theta) \exp\{-c(\theta)t\}\, dt$ has a strict minimum at $b = b(\theta) > 0$ and $b(\theta)$ is a continuous function of θ

(iv) Condition (A8) is satisfied with $H(t)$ replaced by $L(t)$.

Then as $n \to \infty$

(a) $n(\widetilde{\theta}_n - \hat{\theta}_n) \longrightarrow 0$ a.s.

where $\hat{\theta}_n = Z_n - \frac{b(Z_n)}{n}$ and Z_n is as defined in Section 3.2.

(b) $\widetilde{\theta}_n \longrightarrow \theta_0$ a.s.
and $\mathcal{L}\{n(\widetilde{\theta}_n - \theta_0) | P^n_{\theta_0}\} \Rightarrow \mathcal{L}\{X - b(\theta_0)\}$

where X is a random variable with a distribution having density

$$f(x) = \begin{cases} c(\theta_0)\exp\{-c(\theta_0)x\}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(c) $\int L\left[n(\widetilde{\theta}_n - \theta)\right] g_n(\theta | x_1, x_2, \ldots, x_n) d\theta$
$\longrightarrow \int_0^\infty L(t - b(\theta_0)) c(\theta_0) \exp\{-c(\theta_0)t\}\, dt$

where $g_n(\cdot | x_1, \ldots, x_n)$ is the posterior density as defined in Section 3.2.

Proof. The proof is similar to the proof given in Borwanker, Kallianpur and Prakasa Rao (1971).

By Theorem 1, we have

$$\lim_{n \to \infty} \int L(t) \left| g_n^*(t|x_1, x_2, \ldots, x_n) - g_{\theta_0}(t) \right| dt = 0 \quad \text{a.s.}$$

where $g_n^*(t|x_1,\ldots,x_n)$ is the posterior density of $t = n(\theta - \theta_n^*)$ given $x_1,\ldots,x_n$, $\theta_n^* = Z_n - \frac{b(\theta_0)}{n}$ and $g_{\theta_0}(t) = \begin{cases} c(\theta_0)\exp\left\{ c(\theta_0)(t-b(\theta_0)) \right\} & \text{for } t < b(\theta_0), \\ 0, & \text{otherwise.} \end{cases}$

Since $\widetilde{\theta}_n$ minimizes (3.7) with respect to a, using the above fact we get

$$\overline{\lim_{n \to \infty}} \int L\left[n(\widetilde{\theta}_n - \theta) \right] g_n(\theta|x_1,\ldots,x_n) d\theta$$

$$\leq \overline{\lim_{n \to \infty}} \int L\left[n(\theta_n^* - \theta) \right] g_n(\theta|x_1,\ldots,x_n) d\theta$$

$$= \overline{\lim_{n \to \infty}} \int L(t) g_n^*(t|x_1,\ldots,x_n) dt$$

$$= \int_{-\infty}^{b(\theta_0)} L(t) c(\theta_0) \exp\left\{ c(\theta_0)(t - b(\theta_0)) \right\} dt$$

$$= \int_0^{\infty} L(t - b(\theta_0)) c(\theta_0) \exp\left\{ -c(\theta_0) t \right\} dt. \qquad (3.8)$$

To prove (c) it is now enough to prove

$$\underline{\lim_{n \to \infty}} \int L\left[n(\widetilde{\theta}_n - \theta) \right] g_n(\theta|x_1,\ldots,x_n) d\theta$$

$$\geq \int_0^{\infty} L(t - b(\theta_0)) c(\theta_0) \exp\left\{ -c(\theta_0) t \right\} dt \qquad (3.9)$$

We write $V_n = n(\widetilde{\theta}_n - \theta_n^*)$.

We first prove the theorem assuming that

$$\overline{\lim}\ |V_n| < \infty \quad \text{a.s.} \tag{3.10}$$

$$\text{i.e., } P_{\theta_0}\left[\ \overline{\lim}\ |V_n| < \infty\right] = 1 .$$

We take any sequence $(x_1, x_2, \dots)$ from this set of P_{θ_0} - probability one.

Let v be a point in the limit set of $V_n(\underset{\sim}{x})$. Suppose $v \neq 0$. Let $\{n_i\}$ be a subsequence such that $\lim V_{n_i}(\underset{\sim}{x}) = v$. Then

$$\underline{\lim} \int L\left[n_i(\theta_{n_i} - \theta)\right] g_{n_i}(\theta | x_1, \dots, x_{n_i}) d\theta$$

$$= \underline{\lim} \int L(t + V_{n_i})\, g^*_{n_i}(t | x_1, \dots, x_{n_i}) dt$$

$$\geq \int \underline{\lim}\ L(t + V_{n_i})\, g^*_{n_i}(t | x_1, \dots, x_{n_i}) dt$$

$$\geq \int_{-\infty}^{b(\theta_0)} L(t + v) c(\theta_0) \exp\left\{c(\theta_0)(t - b(\theta_0))\right\} dt$$

$$= \int_0^{\infty} L\left[t - (b(\theta_0) + v)\right]\ c(\theta_0) \exp\left\{-c(\theta_0)t\right\} dt$$

$$> \int_0^{\infty} L(t - b(\theta_0)) c(\theta_0)) \exp\left\{-c(\theta_0)t\right\} dt. \text{ by assumption (iii).}$$

Thus, because of (3.8) v must be equal to zero.

Therefore we have

$$V_n \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty \quad \text{a.s.} \tag{3.11}$$

and hence (3.9) is proved (Proceeding as above, replacing $\{n_i\}$ by $\{n\}$).

Since $b(.)$ is continuous $n(\hat{\theta}_n - \theta^*_n) \longrightarrow 0$ a.s.

This together with (3.11) proves (a).

Since $Z_n \to \theta_0$ a.s. and $n(Z_n - \theta_0)$ converges in distribution to the random variable X described in Theorem 2 (see Chapter 2), (b) is an easy consequence of (a).

We shall now prove (3.10).

Suppose (3.10) is not true. i.e., $P_{\theta_0}\left[\overline{\lim}\ |V_n| = \infty\right] > 0.$

Take any sequence $(x_1, x_2, \ldots)$ from the above set of positive probability. Then given any $M > 0$, there exists a subsequence V_{n_i} such that $|V_{n_i}| > M$ for all $i \geq 1$.

Then

$$\int L\left[n_i(\tilde{\theta}_{n_i} - \theta)\right]\ g_{n_i}(\theta \mid x_1, \ldots, x_{n_i})d\theta$$

$$\geq \int_{|t| \leq \frac{M}{4}} L(t + V_{n_i})g^*_{n_i}(t \mid x_1, \ldots, x_{n_i})dt$$

$$\geq \int_{|t| \leq \frac{M}{4}} L(|t| + \tfrac{M}{4})g^*_{n_i}(t \mid x_1, \ldots, x_{n_i})dt$$

$$\left[\text{Since } |t| \leq \tfrac{M}{4},\ |V_{n_i}| > M, \text{ we have } |t + V_{n_i}| \geq |t| + \tfrac{M}{4}\right]$$

$$\geq \int_{|t| \leq \frac{M}{4}} L(t+1)g^*_{n_i}(t \mid x_1, \ldots, x_{n_i})dt \quad \text{(Choosing } M > 1\text{)}$$

Therefore,

$$\overline{\lim_{n \to \infty}} \int L\left[n(\tilde{\theta}_n - \theta)\right]\ g_n(\theta \mid x_1, \ldots, x_n)d\theta.$$

$$\geq \underline{\lim} \int L\left[n_i(\tilde{\theta}_{n_i} - \theta)\right]\ g_{n_i}(\theta \mid x_1, \ldots, x_{n_i})d\theta$$

$$\geq \int_{|t| \leq \frac{M}{4}} L(t+1)g_{\theta_0}(t)dt\ .$$

$$\lim_{M \to \infty} \int_{|t| \leq \frac{M}{4}} L(t+1) g_{\theta_0}(t)\,dt$$

$$= \int_{-\infty}^{b(\theta_0)} L(t+1) c(\theta_0) \exp\left\{ c(\theta_0)(t - b(\theta_0)) \right\} dt$$

$$> \int_{-\infty}^{b(\theta_0)} L(t) c(\theta_0) \exp\left\{ c(\theta_0)(t - b(\theta_0)) \right\} dt \quad \text{by assumption (iii).}$$

Thus, for all sequences $(x_1, x_2, \ldots)$ in a set of positive P_{θ_0} - probability,

$$\overline{\lim_{n \to \infty}} \int L\left[n(\tilde{\theta}_n - \theta) \right] g_n(\theta \mid x_1, \ldots, x_n)\,d\theta$$

$$> \int_{-\infty}^{b(\theta_0)} L(t) c(\theta_0) \exp\left\{ c(\theta_0)(t - b(\theta_0)) \right\} dt$$

$$= \int_0^{\infty} L(t - b(\theta_0)) c(\theta_0) \exp\left\{ -t\, c(\theta_0) \right\} dt.$$

But this is impossible by relation (3.8).

Thus (3.10) is proved and this completes the proof. ///

Remark. It is already mentioned in Remark 1 that even if assumption (A6) is not satisfied, we can get a weaker version of Theorem 1. Using this result and proceeding as in the proof of Theorem 2 we can prove that if $\tilde{\theta}_n$ is a Bayes estimator with respect to any prior over a small neighbourhood of the true parameter point θ_0 (as described in Remark 1), then it is asymptotically equivalent to $\hat{\theta}_n$ (as defined in Theorem 2) in the sense that

$$n(\tilde{\theta}_n - \hat{\theta}_n) \longrightarrow 0 \quad \text{as} \quad n \to \infty \quad \text{a.s.}$$

We shall now give another application of the result on the limiting behaviour of the posterior distribution and the posterior risk. A lower bound to the local asymptotic minimax risk was obtained in Chapter 1. We can obtain the same lower bound using the following theorem (Theorem 2(a)) on the limit of posterior risk.

We assume that assumptions (A1) - (A5) hold. We consider a loss function $L(.)$ satisfying conditions (i) - (iii) of Theorem 2. We also assume

(A8)(a'). There exists $\varepsilon > 0$ such that for all $b > 0$ and all θ in a neighbourhood of θ_0,

$$\int_{-\infty}^{0} L(t)\exp\left\{t(c(\theta)-\varepsilon)\right\} dt + \int_{0}^{b} L(t)\exp\left\{t(c(\theta)+\varepsilon)\right\} dt < \infty .$$

Then we have the following result.

<u>Theorem 2(a)</u>. There exists $\alpha_0 > 0$ such that if $\tilde{\theta}_n$ is a Bayes estimator with respect to any prior probability density over $(\theta_0-\alpha, \theta_0+\alpha)$ which is positive, bounded and continuous on $(\theta_0-\alpha, \theta_0+\alpha)$, where α is any number in $(0, \alpha_0)$, then for all $\theta \in (\theta_0-\alpha, \theta_0+\alpha)$,

$$n(\tilde{\theta}_n - \hat{\theta}_n) \longrightarrow 0 \quad \text{a.s. } P_\theta$$

and $$\int L\left[n(\tilde{\theta}_n-\beta)\right] g_n(\beta \mid x_1,\ldots,x_n)\,d\beta \longrightarrow \int_0^\infty L(t-b(\theta))c(\theta)\exp\left\{-c(\theta)t\right\} dt \quad \text{a.s. } P_\theta$$

where $\hat{\theta}_n = Z_n - \frac{b(Z_n)}{n}$ and $g_n(\beta \mid x_1,\ldots,x_n)$ is the posterior density as defined earlier.

The proof of this result is exactly similar to the proof of Theorem 2 except that in stead of using Theorem 1, we here use the following result.

Theorem 1(a). Under the above assumptions, there exists $\alpha_0 > 0$ such that for all $\alpha \in (0, \alpha_0)$ and for any prior probability density λ over $(\theta_0 - \alpha, \theta_0 + \alpha)$ which is positive, bounded and continuous on $(\theta_0 - \alpha, \theta_0 + \alpha)$, we have for all $\theta \in (\theta_0 - \alpha, \theta_0 + \alpha)$,

$$\lim_{n \to \infty} \int_{\mathbb{R}} L(t) \left| g_n^*(t \mid x_1, \ldots, x_n) - g_\theta(t) \right| dt = 0 \quad \text{a.s. } P_\theta$$

where $g_n^*(t \mid x_1, \ldots, x_n)$ and $g_\theta(t)$ are as defined earlier.

Proof. The proof is similar to the proof of Theorem 1 and we use the same notations.

For $t > b$, $\gamma_n(t) = 0$

For $t < b$, $\log \gamma_n(t) = t \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta} \Big|_{\theta_n'}$

where θ_n' lies between $\hat{\theta}_n$ and $\hat{\theta}_n + t n^{-1}$.

By strong law of large numbers,

$$\frac{1}{n} \Sigma \frac{\partial \log f(x_i, \theta)}{\partial \theta} \longrightarrow c(\theta) \quad \text{a.s. } P_\theta .$$

$$\left| \frac{1}{n} \Sigma \frac{\partial \log f(x_i, \theta)}{\partial \theta} \Big|_{\theta_n'} - \frac{1}{n} \Sigma \frac{\partial \log f(x_i, \theta)}{\partial \theta} \right|$$

$$\leq \frac{1}{n} \Sigma \left| \frac{\partial^2 \log f(x_i, \theta)}{\partial \theta^2} \Big|_{\theta_n''} \right| \cdot |\theta_n' - \theta|$$

where θ_n'' lies between θ_n' and θ.

Now, $|\theta_n'' - \theta_0| \leq |t| n^{-1} + |\hat{\theta}_n - \theta| + |\theta - \theta_0|$,

$$|\theta_n' - \theta| \leq |t| n^{-1} + |\hat{\theta}_n - \theta| .$$

Let $\varepsilon > 0$ be as in condition (A8)(a'). Since $\hat{\theta}_n \to \theta$ a.s. P_θ, we can choose $\delta > 0$ such that for all $\theta \in (\theta_0 - \delta, \theta_0 + \delta)$ and for all sufficiently large n (depending on the sample point),

$$\left.\begin{array}{lll} \log \mathcal{V}_n(t) < t(c(\theta) - \varepsilon) & \text{whenever} & -\delta n \leq t \leq 0 \\ \log \mathcal{V}_n(t) < t(c(\theta) + \varepsilon) & \text{whenever} & 0 \leq t < b \end{array}\right\} \text{ a.s. } P_\theta .$$

Also for each t, $\mathcal{V}_n(t) \to \mathcal{V}(t)$ a.s. P_θ

where $\mathcal{V}(t)$ is given by

$$\mathcal{V}(t) = \begin{cases} \exp(tc(\theta)) & \text{for } t < b , \\ 0 , & \text{otherwise.} \end{cases}$$

Then as in step 2 of the proof of Theorem 1,

$$\lim_{n \to \infty} \int_{|t| \leq \delta n} L(t) \left| \mathcal{V}_n(t) \lambda(\hat{\theta}_n + t n^{-1}) - \lambda(\theta) \mathcal{V}(t) \right| dt = 0 \quad \text{a.s. } P_\theta .$$

for all $\theta \in (\theta_0 - \delta, \theta_0 + \delta)$.

If we take $\alpha < \delta/2$, then for any prior λ over $(\theta_0 - \alpha, \theta_0 + \alpha)$, the set on which $\mathcal{V}_n(t)$ is defined will be a subset of $\left\{ |t| \leq \delta n \right\}$ for all sufficiently large n a.s. P_θ, where θ is any number in $(\theta_0 - \alpha, \theta_0 + \alpha)$ and therefore we have

$$\int L(t) \left| \mathcal{V}_n(t) \lambda(\hat{\theta}_n + t n^{-1}) - \lambda(\theta) \mathcal{V}(t) \right| dt = 0 \quad \text{a.s. } P_\theta .$$

The rest of the proof is exactly the same as step 4 of the proof of Theorem 1. ///

We shall now use Theorem 2(a) to find a lower bound to the local asymptotic minimax risk. Let α_0 be as in Theorem 2(a). For any $\alpha \in (0, \alpha_0)$,

$$\underset{n \to \infty}{\underline{\lim}} \inf_{T_n} \sup_{|\theta - \theta_0| \leq \alpha} E_\theta L \left\{ n(T_n - \theta) \right\}$$

$$\geq \underset{n \to \infty}{\underline{\lim}} \int_{\theta_0 - \alpha}^{\theta_0 + \alpha} E_\theta L \left\{ n(\tilde{\theta}_n - \theta) \right\} \lambda(\theta) \, d\theta$$

where the infimum in the left hand side is over all estimators T_n of θ, $\lambda(\theta)$ is any prior density over $(\theta_0 - \alpha, \theta_0 + \alpha)$ as stated in Theorem 2(a) and $\tilde{\theta}_n$ is the corresponding Bayes estimator.

Let $g_n(\beta | x_1, \ldots, x_n)$ be the posterior density and $m(\underset{\sim}{x})$ be the marginal density of $\underset{\sim}{x} = (x_1, \ldots, x_n)$. We write

$$\xi_n(\underset{\sim}{x}) = \int L \left\{ n(\tilde{\theta}_n - \beta) \right\} g_n(\beta | x_1, \ldots, x_n) \, d\beta .$$

Then we have

$$\underset{n \to \infty}{\underline{\lim}} \inf_{T_n} \sup_{|\theta - \theta_0| \leq \alpha} E_\theta L \left\{ n(T_n - \theta) \right\}$$

$$\geq \underset{n \to \infty}{\underline{\lim}} \int \xi_n(x) \, dm(\underset{\sim}{x})$$

$$= \underset{n \to \infty}{\underline{\lim}} \int_{\theta_0 - \alpha}^{\theta_0 + \alpha} E_\theta \left\{ \xi_n(\underset{\sim}{x}) \right\} \lambda(\theta) \, d\theta$$

$$\geq \int_{\theta_0 - \alpha}^{\theta_0 + \alpha} E_\theta \left\{ \underset{n \to \infty}{\underline{\lim}} \xi_n(\underset{\sim}{x}) \right\} \lambda(\theta) \, d\theta \quad \left[\text{By Fatou's lemma} \right]$$

$$= \int_{\theta_0 - \alpha}^{\theta_0 + \alpha} A(\theta) \lambda(\theta) \, d\theta \quad \left[\text{By Theorem 2(a)} \right]$$

where $A(\theta) = \int_0^\infty L(t - b(\theta)) \, c(\theta) \exp \left\{ - c(\theta) t \right\} dt.$

Under mild condition, $\lim_{\alpha \to 0} \int_{\theta_0 - \alpha}^{\theta_0 + \alpha} A(\theta)\lambda(\theta)d\theta = A(\theta_0)$

(we here choose appropriate λ).

Thus we have the following lower bound to the local asymptotic minimax risk :

$$\lim_{\alpha \to 0} \varliminf_{n \to \infty} \inf_{T_n} \sup_{|\theta - \theta_0| \le \alpha} E_\theta L \left\{ n(T_n - \theta) \right\}$$

$$\ge \int_0^\infty L(t - b(\theta_0))\, c(\theta_0) \exp\left\{ - c(\theta_0)t \right\} dt.$$

3.4 REGULAR CASE

We now reexamine the Bernstein-von Mises theorem in regular cases. Consider the set up of Bickel and Yahav (1969) or Chao (1970). We have a random sample $X_1, X_2, \ldots, X_n$ from a distribution P_θ having a density $f(x, \theta)$ depending on a real parameter θ, where $\theta \in \Theta$, an open interval of $\mathbb{R}$. Let θ_0 be the true parameter point. We make the following assumptions.

(B1) We are given a Bayes prior measure Λ on Θ and Λ has a density λ with respect to the Lebesgue measure which is continuous and positive in an open neighbourhood of θ_0.

(B2) $\dfrac{\partial \log f(x, \theta)}{\partial \theta}$ and $\dfrac{\partial^2 \log f(x, \theta)}{\partial \theta^2}$ exist and are continuous in θ for almost all x.

(B3) $E_{\theta_0} \sup \left\{ \left| \dfrac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right| : |\theta - \theta_0| < \varepsilon, \theta \in \Theta \right\} < \infty$

for some $\varepsilon > 0$.

(B4) $\hat{\theta}_n \longrightarrow \theta_o$ a.s. P_{θ_o} where $\hat{\theta}_n$ is a maximum likelihood estimator.

(B5) $I(\theta_o) = -E_{\theta_o}\left(\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}\Big|_{\theta_o}\right)$ is a finite positive number.

(B6) $\lim_{\rho \to 0} E_{\theta_o} \log f(x, \theta, \rho) = E_{\theta_o} \log f(X, \theta)$

where $f(x, \theta, \rho)$ is as defined in assumption (A6).

(B7) $E_{\theta_o} \operatorname{Sup}\left\{\log f(X, \theta) - \log f(X, \theta_o) : |\theta - \theta_o| > \delta, \theta \in \Theta\right\} < 0$

for sufficiently large $\delta > 0$.

We consider a nonnegative weight function $H(t) = \hat{H}(|t|)$ satisfying the following conditions.

(B8)(a) There exists $\varepsilon > 0$ such that

$$\int H(t) \exp\left\{-\frac{t^2}{2}(I(\theta_o) - \varepsilon)\right\} dt < \infty .$$

(b) For all $u > 0$ and $\delta > 0$,

$$e^{nu} \int_{|t| > \delta\sqrt{n}} H(t) \lambda(\hat{\theta}_n + \frac{t}{\sqrt{n}}) dt \longrightarrow 0 \quad \text{a.s.} \quad \text{as } n \to \infty .$$

The following theorem is an improvement over the results of Bickel and Yahav (1969) and Chao (1970) (see discussions following the proof).

Theorem 3. Under assumptions (B1) - (B7), for any weight function H satisfying (B8),

$$\lim_{n \to \infty} \int_{\mathbb{R}} H(t) |g_n^*(t|x_1,\ldots,x_n) - \varphi(I^{-1}(\theta_o), t)| dt = 0 \quad \text{a.s. } P_{\theta_o}$$

where $g_n^*(t|x_1,\ldots,x_n)$ is the posterior density of $t = \sqrt{n}(\theta - \hat{\theta}_n)$ given observations $x_1,\ldots,x_n$ and $\varphi(v, t)$ is the density of the normal distribution with mean zero and variance v.

Proof. The posterior distribution can be written as

$$g_n^*(t|x_1,\dots,x_n) = C_n^{-1}\, \nu_n(t)\lambda(\hat{\theta}_n + tn^{-1/2})$$

where
$$\nu_n(t) = \frac{\prod_{i=1}^{n} f(x_i, \hat{\theta}_n + tn^{-1/2})}{\prod_{i=1}^{n} f(x_i, \hat{\theta}_n)}, \quad C_n = \int \nu_n(t)\lambda(\hat{\theta}_n + tn^{-1/2})\,dt.$$

We shall only prove that for all $\delta > 0$, there exists $\varepsilon(\delta) > 0$ such that whenever $|t| > \delta\sqrt{n}$, for all sufficiently large n,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\log f(x_i, \hat{\theta}_n + tn^{-1/2}) - \log f(x_i, \theta_0)\right\} < -\varepsilon(\delta) \quad \text{a.s.} \qquad (3.12)$$

The remaining part of the proof is similar to that given in Bickel and Yahav (1969) or Borwanker, Kallianpur and Prakasa Rao (1971). To prove (3.12) we use the same argument as is used in the proof for non-regular case (see step 3 in the proof of Theorem 1). We first note that for $|t| > \delta\sqrt{n}$, $|\hat{\theta}_n + \frac{t}{\sqrt{n}} - \theta_0| > \frac{\delta}{2}$ for all sufficiently large n and therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\log f(x_i, \hat{\theta}_n + \frac{t}{\sqrt{n}}) - \log f(x_i, \theta_0)\right\}$$

$$\leq \sup_{|\theta - \theta_0| > \frac{\delta}{2}} \left\{\frac{1}{n}\Sigma \log f(x_i, \theta) - \log f(x_i, \theta_0)\right\}.$$

We then get $\delta_0 > \delta$ such that

$$E_{\theta_0} \sup\left\{\log f(X,\theta) - \log f(X,\theta_0) : |\theta - \theta_0| > \delta_0, \theta \in (H)\right\} < 0$$

Setting $(H)_0 = \left\{\theta \in (H) : |\theta - \theta_0| > \delta_0\right\}$

and $(H)_1 = \left\{\theta \in (H) : \frac{\delta}{2} \leq |\theta - \theta_0| \leq \delta_0\right\}$

and proceeding as in step 3 in the proof of Theorem 1, we can get a finite number of points $\theta_1, \theta_2, \ldots, \theta_k \in \textcircled{H}_1$ and open neighbourhoods $(\theta_j - \rho_{\theta_j}, \theta_j + \rho_{\theta_j})$, $j = 1,2,\ldots,k$ forming a cover of $\textcircled{H}_1$ such that

$$E_{\theta_0} \log f(X, \theta_j, \rho_{\theta_j}) < E_{\theta_0} \log f(X, \theta_0) \text{ for } j = 1,2,\ldots,k. \qquad (3.13)$$

Then for all t in $\{ |t| > \delta\sqrt{n} \}$ and all sufficiently large n,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\log f(x_i, \hat{\theta}_n + tn^{-1/2}) - \log f(x_i, \theta_0)\right\}$$

$$\leq \operatorname{Max}\left\{ \frac{1}{n}\sum_{i=1}^{n}\left\{\sup_{\theta \in \textcircled{H}_0} \log f(x_i, \theta) - \log f(x_i, \theta_0)\right\},\right.$$

$$\left. \frac{1}{n}\sum_{i=1}^{n}\left\{\log f(x_i, \theta_j, \rho_{\theta_j}) - \log f(x_i, \theta_0)\right\},\ j = 1,2,\ldots,k \right\}$$

From this,(3.12) follows by assumption (B7), relation (3.13) and strong law of large number. ///

The proofs of Bernstein-von Mises theorem given in Bickel and Yahav (1969) and Chao (1970) are based on the assumption

$$E_{\theta_0} \sup\left\{\log f(X,\theta) - \log f(X,\theta_0) : |\theta - \theta_0| > \delta,\ \theta \in \textcircled{H}\right\} < 0$$
$$\text{for all } \delta > 0. \qquad (3.14)$$

Borwanker, Kallianpur and Prakasa Rao (1971) proved their results under a Markov-process analogue of the above assumption. Here the above assumption is replaced by the weaker assumption (B7) and a reasonable assumption (B6) (which is proved in Wald (1949) under mild conditions). The assumption (3.14) is not satisfied for the usual regular cases. Consider, for example, the normal distribution with mean θ and variance 1, which is a standard example of regular case.

$$f(x,\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\theta)^2\right\}, \quad -\infty < x < \infty, \; -\infty < \theta < \infty .$$

Let us take $\theta_0 = 0$. Then

$$\text{Sup}\left\{\log f(X,\theta) - \log f(X,\theta_0) : \theta < \theta_0 - \delta\right\}$$
$$= \begin{cases} \frac{1}{2} X^2, & \text{if } X < -\delta, \\ -X\delta - \frac{1}{2}\delta^2, & \text{if } X > -\delta . \end{cases}$$

This implies that

$$E_{\theta_0} \text{Sup}\left\{\log f(X,\theta) - \log f(X,\theta_0) : \theta < \theta_0 - \delta\right\} > 0$$

and hence $E_{\theta_0} \text{Sup}\left\{\log f(X,\theta) - \log f(X,\theta_0) : |\theta - \theta_0| > \delta\right\} > 0$

for all sufficiently small $\delta > 0$.

As another example consider

$$f(x,\theta) = \theta e^{-\theta x}, \quad 0 < x < \infty \; ; \; \theta > 0 .$$

Take $\theta_0 = 1$. Then

$$\text{Sup}\left\{\log f(X,\theta) - \log f(X,\theta_0) : \theta < \theta_0 - \delta\right\}$$
$$= \begin{cases} -\log X - 1 + X, & \text{if } X > \frac{1}{1-\delta}, \\ \log(1-\delta) + \delta X, & \text{if } X < \frac{1}{1-\delta} \end{cases} \quad (\text{for } \delta < 1).$$

It is now easy to show that

$$E_{\theta_0} \text{Sup}\left\{\log f(X,\theta) - \log f(X,\theta_0) : \theta < \theta_0 - \delta\right\} > 0$$

for all sufficiently small $\delta > 0$.

Similarly for many other regular cases, the left hand side of (3.14) can be shown to be greater than zero for sufficiently small $\delta > 0$. However, condition (B7) is satisfied in most of the usual situations.

CHAPTER 4

ESTIMATION IN MULTIPARAMETER CASE

4.1 INTRODUCTION

In the previous chapters, we considered the case where there is only one unknown real parameter θ with respect to which the problem is non-regular. In this chapter we consider the case in which there is an additional unknown parameter, say, φ. A typical example is the case of i.i.d. observations from a distribution with density

$$f(x,\theta,\varphi) = g(x-\theta,\varphi)$$

where $g(x,\varphi)$ is, for every φ, a density on $[0,\infty)$(we assume $g(0,\varphi) > 0$) This type of problems were studied by Smith (1985), Cheng and Iles (1987) and others but these authors were concerned mainly with the problem of obtaining the asymptotic distribution of the maximum likelihood estimators or its alternatives. We here study the problem of efficient estimation from the Hajek-Le Cam-Millar point of view as outlined in Chapter 1 (and also in the introductory chapter), that is, we obtain a lower bound to the (local) asymptotic risk and suggest an estimator which attains this lower bound. It is assumed that the usual regularity conditions are satisfied with respect to the additional parameter φ. For simplicity, we consider only the case in which φ is a real parameter. An important result in this situation is that the problem of estimation of θ and φ, when considered together, are asymptotically independent and the limiting experiment is a product of a regular one and a non-regular one. In Section 4.2 we obtain a limiting experiment which is the product of the Gaussian shift experiment and the limiting experiment obtained in Chapter 1. Using this we also obtain a lower bound to the asymptotic risk. In Section 4.3 we consider the example of independent and identically distributed observations and suggest an efficient estimator.

4.2 LOWER BOUND FOR ASYMPTOTIC RISK UNDER AN ASYMPTOTIC EXPANSION OF LIKELIHOOD RATIO

Let $\left\{ (\underline{X}^n, \underline{A}^n), P^n_{\theta,\varphi}, \theta \in Ⓗ, \varphi \in \Phi \right\}$, $n \geq 1$, be a sequence of statistical experiments where Ⓗ and Φ are open subsets of $\mathbb{R}$. We fix $\theta_0 \in$ Ⓗ and $\varphi_0 \in \Phi$.

We set

$$\Lambda_{n,\theta_0,\varphi_0}(u,v) = \Lambda_n(u,v) = \frac{dP^n_{\theta_0 + un^{-1},\, \varphi_0 + vn^{-1/2}}}{dP^n_{\theta_0,\, \varphi_0}}, \quad u \geq 0, \; v \in \mathbb{R},$$

and make the following assumption :

(A) For any $u \geq 0$ and $v \in \mathbb{R}$, we have a.s. $P^n_{\theta_0,\varphi_0}$

$$\Lambda_n(u,v) = \begin{cases} \exp\left\{ u\, c(\theta_0,\varphi_0) + v\Delta_n(\theta_0,\varphi_0) - \frac{1}{2} v^2 I_{\theta_0}(\varphi_0) + \varepsilon_n \right\}, & \text{if } \tau_n > u, \\ 0 & \text{if } \tau_n < u, \end{cases}$$

where $c(\theta_0,\varphi_0) > 0$, $0 < I_{\theta_0}(\varphi_0) < \infty$ are constants, ε_n, Δ_n and τ_n are random variables such that

$$\varepsilon_n \xrightarrow{P^n_{\theta_0,\varphi_0}} 0$$

$$(\Delta_n, \tau_n) \xrightarrow{\mathcal{L}} (\Delta, \tau)$$

where $\Delta \sim N(0, I_{\theta_0}(\varphi_0))$, τ has a distribution with density $c(\theta_0,\varphi_0) \exp\left\{ -c(\theta_0,\varphi_0)x \right\}$ on $(0,\infty)$ and Δ and τ are independent.

We set $Q^n_{u,v} = P^n_{\theta_o + un^{-1},\ \varphi_o + vn^{-1/2}}$.

From now onwards, we shall write just c, Δ_n and I in place of $c(\theta_o, \varphi_o)$, $\Delta_n(\theta_o, \varphi_o)$ and $I_{\theta_o}(\varphi_o)$ respectively. Unless otherwise stated, all probability statements are with respect to $P^n_{\theta_o, \varphi_o}$.

Lemma 1. Under assumption (A), for any $u \geq 0$ and $v \in \mathbb{R}$, $Q^n_{u,v}$ is contiguous to $Q^n_{o,o}$.

Proof. For any $u \geq 0$ and $v \in \mathbb{R}$, by assumption (A) we have

$$\Lambda_n(u,v) \xrightarrow{\mathcal{L}} \exp(u\,c + v\,\Delta - \tfrac{1}{2}v^2 I)\ 1_{(\tau > u)}$$

and

$$\begin{aligned} &E\left[\exp(u\,c + v\,\Delta - \tfrac{1}{2}v^2 I)\ 1_{(\tau > u)}\right] \\ &= e^{uc}\ E\left[\exp(v\,\Delta - \tfrac{1}{2}v^2 I)\right]\ e^{-uc} \\ &= 1 . \end{aligned}$$

Hence by a result on contiguity (referred to as Le Cam's 1st lemma in Hajek and Sidak (1967)) the result follows. ///

Remark. It is more natural to replace the assumption of asymptotic independence of Δ_n and τ_n by the assumption of contiguity of $Q^n_{u,v}$ to $Q^n_{o,o}$ because it is this result of contiguity which is used to obtain a limiting experiment (see Theorem 1 below). Indeed, we have the following result :

Lemma 2. Suppose that for any $u \geq 0$ and $v \in \mathbb{R}$, we have a.s. $P^n_{\theta_o,\varphi_o}$

$$\Lambda_n = \begin{cases} \exp\left\{u\,c + v\Delta_n - \tfrac{1}{2}v^2 I + \varepsilon_n\right\}, & \text{if } \tau_n > u \\ 0, & \text{if } \tau_n < u, \end{cases}$$

where $c > 0$, $0 < I < \infty$ are constants and ε_n, Δ_n and τ_n are random variables such that ε_n converges in $P^n_{\theta_o,\varphi_o}$ - probability to zero and the distribution of (Δ_n, τ_n) converges weakly to some bivariate distribution. Then the following two statements are equivalent.

(i) For all $u \geq 0$, $v \in \mathbb{R}$,

$$Q^n_{u,v} \text{ is contiguous to } Q^n_{o,o} .$$

(ii) $(\Delta_n, \tau_n) \xrightarrow{\mathcal{L}} (\Delta, \tau)$

where Δ and τ are independent random variables as described in assumption (A).

<u>Proof</u>. That (ii) implies (i) is proved above (Lemma 1). Suppose now (i) holds. Putting $v = 0$ and using a result on contiguity (a converse of Le Cam's 1st lemma) one can easily show that

$$\tau_n \xrightarrow{\mathcal{L}} \tau .$$

Suppose $(\Delta_n, \tau_n) \xrightarrow{\mathcal{L}} (\Delta^*, \tau)$.

We shall prove that $\Delta^* \sim N(0,I)$ and is independent of τ. By statement (i) and a result on contiguity (used above), for all $u \geq 0$ and $v \in \mathbb{R}$,

$$E\left[\exp(u\,c + v\,\Delta^* - \tfrac{1}{2}v^2 I)\, 1_{(\tau > u)}\right] = 1$$

$$\text{i.e., } E\left[e^{v\,\Delta^*}\, 1_{(\tau > u)}\right] = \exp(-u\,c + \tfrac{1}{2}v^2 I)$$

$$\text{i.e., } E\left[e^{v\,\Delta^*}\, 1_{(\tau > u)}\right] = E\left[\exp(\tfrac{1}{2}v^2 I)\, 1_{(\tau > u)}\right] .$$

This implies

$$E(e^{v\,\Delta^*} \mid \tau) = \exp(\tfrac{1}{2}v^2 I) \quad \text{for all } u \geq 0,\ v \in \mathbb{R}.$$

Thus, Δ^* and τ are independent and $\Delta^* \sim N(0,I)$. ///

Let us now denote by $Q_{u,v}(u \geq 0,\ v \in \mathbb{R})$ a probability on $\mathbb{R}^2$ with density

$$q_{u,v}(x,y) = \begin{cases} c \exp\left\{-c(x-u)\right\} (2\pi)^{-1/2}\, I^{1/2} \exp\left\{-\frac{1}{2}(y-v)^2\right\}, & \text{if } x > u, \\ 0, & \text{otherwise.} \end{cases}$$

We define experiments

$$E^n = \left\{ Q^n_{u,v} : u \geq 0,\ v \in \mathbb{R} \right\},\ n \geq 1,$$
$$E = \left\{ Q_{u,v} : u \geq 0,\ v \in \mathbb{R} \right\}.$$

Then we have the following result :

Theorem 1. Under assumption (A), the sequence of experiments E^n converges to E.

Proof. To prove this we use Millar's proposition stated in Section 1.2. Hence $Q^n_{u,v}$ is contiguous to $Q^n_{o,o}$ and $Q_{u,v}$ is absolutely continuous with respect to $Q_{o,o}$ for all $u \geq 0$ and $v \in \mathbb{R}$. By assumption (A) for any $(u,v) \in \mathbb{R}^+ \times \mathbb{R}$,

$$\mathcal{L}\left\{ \frac{dQ^n_{u,v}}{dQ^n_{o,o}} \mid Q^n_{o,o} \right\} \Rightarrow \mathcal{L}\left\{ \exp(u\, c + v\Delta - \tfrac{1}{2} v^2 I)\, 1_{(\tau > u)} \right\}.$$

Since the distribution of $\frac{dQ_{u,v}}{dQ_{o,o}}$ under $Q_{o,o}$ is same as that of $\exp(u\, c + v\,\Delta - \frac{1}{2} v^2 I)\, 1_{(\tau > u)}$ we have

$$\mathcal{L}\left\{ \frac{dQ^n_{u,v}}{dQ^n_{o,o}} \mid Q^n_{o,o} \right\} \Rightarrow \mathcal{L}\left\{ \frac{dQ_{u,v}}{dQ_{o,o}} \mid Q_{o,o} \right\}.$$

Similarly it is easy to verify that for $(u_1,v_1),\ldots,(u_k,v_k) \in \mathbb{R}^+ \times \mathbb{R}$,

$$\mathcal{L}\left\{\left(\frac{dQ^n_{u_1,v_1}}{dQ^n_{o,o}}, \ldots, \frac{dQ^n_{u_k,v_k}}{dQ^n_{o,o}}\right) \Big| \, Q^n_{o,o}\right\} \Rightarrow \mathcal{L}\left\{\left(\frac{dQ_{u_1,v_1}}{dQ_{o,o}}, \ldots, \frac{dQ_{u_k,v_k}}{dQ_{o,o}}\right) \Big| \, Q_{o,o}\right\}$$

The result now follows from Millar's proposition. ///

We shall now obtain a lower bound to the local asymptotic minimax risk using Theorem 1 and the Hajek-Le Cam asymptotic minimax theorem (stated in Section 1.3). We consider a subconvex loss function which is defined as follows :

Definition. A loss function $L(\underset{\sim}{t}, \underset{\sim}{a}) = L(\underset{\sim}{t} - \underset{\sim}{a})$, $\underset{\sim}{t}, \underset{\sim}{a} \in \mathbb{R}^2$, is said to be subconvex if L satisfies the following conditions :

(i) $L(\underset{\sim}{x}) \geq 0$ for all $\underset{\sim}{x} \in \mathbb{R}^2$

(ii) $L(x_1, x_2) = L(|x_1|, |x_2|)$ for all $\underset{\sim}{x} = (x_1, x_2) \in \mathbb{R}^2$

(iii) $\{\underset{\sim}{x} : L(\underset{\sim}{x}) \leq c\}$ is closed and convex for all $c \geq 0$.

We have the following lemma :

Lemma 3. Under assumption (A), for any subconvex loss function L,

$$\underset{\substack{A_1 \to \infty \\ A_2 \to \infty}}{\underline{\lim}} \; \underset{n \to \infty}{\underline{\lim}} \; \inf_{(\hat{\theta}_n, \hat{\varphi}_n)} \; \sup_{\substack{\theta_o \leq \theta \leq \theta_o + A_1 n^{-1} \\ |\varphi - \varphi_o| \leq A_2 n^{-1/2}}} E_\theta L\left(n(\hat{\theta}_n - \theta), \sqrt{n}(\hat{\varphi}_n - \varphi)\right)$$

$$\geq \inf_{\delta} \; \sup_{\underset{\sim}{u} \in \mathbb{R}^+ \times \mathbb{R}} R(\delta, \underset{\sim}{u}) \qquad (4.1)$$

where the infimum in the left hand side is over all estimators $(\hat{\theta}_n, \hat{\varphi}_n)$ of (θ, φ), the infimum in the right hand side is over all randomized (Markov kernel) procedures for the experiment E and $R(\delta, \lambda)$ is the risk of the procedure δ at $\underset{\sim}{u} = (u_1, u_2)$ with loss function L.

We omit the proof since it is similar to the proof of Lemma 1 of Section 1.3.

We shall now compute the minimax risk given in the right hand side of (4.1). We use the same technique as was used in Section 1.3.

We assume that

C(i) $E\,L(X - b, Y)$ exists and is finite for some b, where X has a distribution with density $c\,e^{-cx}$ on $(0,\infty)$ and $Y \sim N(0, I^{-1})$ (c, I are as in assumption (A)).

Also there exists $b_0 = b_0(\theta_0, \varphi_0)$ such that

$$E\,L(X - b_0, Y) = \inf_b E\,L(X - b, Y) = R_0 \text{ , say .}$$

C(ii) For every $\varepsilon > 0$, there exists $N > 0$ such that for all $b_1, b_2 \in \mathbb{R}$,

$$\int_{-N}^{N} \int_{0}^{N} L(x - b_1, y - b_2)\, dF_X(x)\, dF_Y(y) \geq R_0 - \varepsilon .$$

Remark 1. Conditions C(i) and C(ii) hold if, for example, L is bounded. For an unbounded subconvex loss function, the conditions are satisfied if we assume that L is continuous and nondecreasing in $|x_1|$ and $|x_2|$.

Remark 2. We note that

$$\inf_{b_1, b_2} E\,L(X - b_1, Y - b_2) = \inf_b E\,L(X - b, Y) .$$

This can be proved using Anderson's lemma (see, for example, Ibragimov and Hasminskii, p. 157).

Lemma 4. For any subconvex loss function satisfying conditions C(i) and C(ii), we have

$$\inf_{\delta} \sup_{\underset{\sim}{u} \in R^+ \times \mathbb{R}} R(\delta, \underset{\sim}{u}) = E\,L(X - b_0, Y)$$

where the minimax risk in the left hand side is as described in Lemma 3.

Proof. As in the proof of Lemma 2 of Section 1.3 we shall exhibit a sequenee τ_M of prior distributions on $\mathbb{R}^+ \times \mathbb{R}$ and show that

$$\lim_{M \to \infty} \inf_{\delta} r(\delta, \tau_M) \geq R_0$$

where the infimum in the left hand side is over all non-randomized decision rule $\delta(X,Y) = (T(X,Y), V(X,Y))$ and $r(\delta,\tau_M)$ is the Bayes risk of δ with respect to the prior τ_M. This will prove the result (as in the proof of Lemma 2 of Section 1.3).

We choose τ_M as the uniform distribution over the set $(0,M)\times(-M,M)$. Let ε be any positive number and N be such that

$$\int_{-N}^{N}\int_{0}^{N} L(x - b_1, y - b_2)\, dF_X(x)\, dF_Y(y) \geq R_0 - \varepsilon$$

for all $b_1, b_2 \in \mathbb{R}$.

For any non-randomized decision rule $(T(X,Y),V(X,Y))$ and any $M > N$,

$r(\tau_M,(T,V))$

$$= \frac{1}{2M^2}\int_{-M}^{M}\int_{0}^{M}\int L(T(x,y)-u,\ V(x,y)-v)\ dF_{X,Y}(x - u,\ y - v)$$

$$= \frac{1}{2M^2}\int_{-M}^{M}\int_{0}^{M}\int L\left[T(x+u,y+v) - u,\ V(x+u,y+v) - v\right] dF_{X,Y}(x,y)\ du\ dv$$

$$= \frac{1}{2M^2}\int\int_{-M+y}^{M+y}\int_{x}^{M+x} L\left[T(z_1,z_2) - z_1+x,\ V(z_1,z_2) - z_2+y\right] dz_1\ dz_2\ dF_{X,Y}(x,y)$$

(interchanging the integrals and putting $x+u = z_1$, $y+v = z_2$)

$$= \frac{1}{2M^2}\iint \left[\int_{z_2-M}^{z_2+M}\int_{z_1-M}^{z_1} L\left[x+T(z_1,z_2)-z_1,\ y+V(z_1,z_2)-z_2\right] dF_X(x)\ dF_Y(y)\right] dz_1\ dz_2$$

$$\geq \frac{1}{2M^2}\int_{-(M-N)}^{M-N}\int_N^M\left[\int_{z_2-M}^{z_2+M}\int_{z_1-M}^{z_1} L\left[x+T(z_1,z_2)-z_1,\ y+V(z_1,z_2)-z_2\right] dF_X(x)\ dF_Y(y)\right] dz_1\ dz_2$$

$$\geq \frac{1}{2M^2}\int_{-(M-N)}^{M-N}\int_N^M\left[\int_{-N}^{N}\int_0^{N} L\left[x+T(z_1,z_2)-z_1,\ y+V(z_1,z_2)-z_2\right] dF_X(x)\ dF_Y(y)\right] dz_1\ dz_2$$

$$\geq \frac{(M-N)^2}{M^2}(R_0-\varepsilon)\ .$$

Since $\varepsilon > 0$ arbitrary, this proves the result. ///

Now, from Lemma 3 and Lemma 4 we get the following result :

Theorem 2. Under assumption (A), for any subconvex loss function L satisfying C(i) and C(ii),

$$\lim_{\substack{A_1\to\infty \\ A_2\to\infty}}\ \lim_{n\to\infty}\ \inf_{(\hat\theta_n,\hat\varphi_n)}\ \sup_{\substack{\theta_0\leq\theta\leq\theta_0+A_1n^{-1} \\ |\varphi-\varphi_0|\leq A_2n^{-1/2}}} E_{\theta,\varphi}L\left(n(\hat\theta_n-\theta),\ \sqrt{n}\,(\hat\varphi_n-\varphi)\right)$$

$$\geq E\ L(X-b_0,\ Y)$$

where X, Y, b_0 are as given above.

4.3 EFFICIENT ESTIMATION IN I.I.D. CASE

In this section we consider a specific family of non-regular cases and apply the results of the previous section to solve the problem of efficient estimation in these cases. Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations, each X_i having distribution $P_{\theta,\varphi}$, $\theta \in (\Theta)$, $\varphi \in \bar\Phi$, with density $f(x,\theta,\varphi)$ on $\mathbb{R}$ with respect to Lebesgue measure, where

$$f(x,\theta,\varphi) > 0 \quad \text{for} \quad x \geq \theta$$
$$= 0 \quad \text{for} \quad x < \theta$$

and $\textcircled{H}$ and Φ are open subsets of $\mathbb{R}$.

We make the following assumptions on the density $f(x,\theta,\varphi)$:

(B1) $f(x,\theta,\varphi)$ is jointly continuous in (x,θ,φ) on the set $\{(x,\theta,\varphi) : x \geq \theta\}$.

(B2) All partial derivatives upto the second order of $f(x,\theta,\varphi)$ with respect to θ and φ and the third derivatives

$$\frac{\partial^3 \log f(x,\theta,\varphi)}{\partial\theta\,\partial\varphi^2} \quad \text{and} \quad \frac{\partial^3 \log f(x,\theta,\varphi)}{\partial\varphi^3}$$

exist for all $x > \theta$.

(B3) For all $(\theta,\varphi) \in (\bar{\textcircled{H}}) \times \Phi$,

(a) $E_{\theta,\varphi} \dfrac{\partial \log f(X_1,\theta,\varphi)}{\partial\varphi} = 0$,

and (b) $0 < E_{\theta,\varphi} \left[\dfrac{\partial \log f(X_1,\theta,\varphi)}{\partial\varphi} \right]^2 = - E_{\theta,\varphi} \left[\dfrac{\partial^2 \log f(X_1,\theta,\varphi)}{\partial\varphi^2} \right] < \infty$.

(B4) For all $(\theta,\varphi) \in (\bar{\textcircled{H}}) \times \Phi$,

$$\int_{\theta}^{\theta+h} \frac{\partial f(x,\theta,\varphi)}{\partial\theta}\, dx = o(h^{1/2}), \quad h > 0$$

(B5) For any $(\theta_0,\varphi_0) \in (\bar{\textcircled{H}}) \times \Phi$, there exists a neighbourhood $N(\theta_0,\varphi_0)$ of (θ_0,φ_0) and P_{θ_0,φ_0} - integrable functions $H_i(x)$, $i=1,\ldots,4$, such that for all $(\theta,\varphi) \in N(\theta_0,\varphi_0)$ and all $x > \theta$,

(a) $\left| \dfrac{\partial^2 \log f(x,\theta,\varphi)}{\partial\theta^2} \right| \leq H_1(x)$

(b) $\left| \dfrac{\partial^2 \log f(x,\theta,\varphi)}{\partial\theta\,\partial\varphi} \right| \leq H_2(x)$

(c) $\left| \dfrac{\partial^3 \log f(x,\theta,\varphi)}{\partial\theta\, \partial\varphi^2} \right| \le H_3(x)$

(d) $\left| \dfrac{\partial^3 \log f(x,\theta,\varphi)}{\partial\varphi^3} \right| \le H_4(x)$

Let $P^n_{\theta,\varphi}$ denote the n fold product of the probability measure $P_{\theta,\varphi}$. We fix $(\theta_0,\varphi_0) \in \textcircled{H} \times \Phi$. Unless otherwise specified all probability statements are with respect to P_{θ_0,φ_0}. We shall show that under assumptions (B1) - (B4), the asymptotic expansion as given in (A) of Section 4.2 holds.

Expanding at (θ_0,φ_0) by Taylor's theorem, we get for all $u \ge 0$, $v \in \mathbb{R}$,
$\log \Lambda_n(u,v)$

$$= \sum_{i=1}^{n} \log f(X_i,\theta_0 + un^{-1}, \theta_0 + vn^{-1/2}) - \sum_{i=1}^{n} \log f(X_i,\theta_0,\varphi_0)$$

$$= \frac{u}{n}\sum \frac{\partial \log f(X_i,\theta,\varphi)}{\partial\theta}\Big|_{(\theta_0,\varphi_0)} + \frac{v}{\sqrt{n}}\sum \frac{\partial \log f(x,\theta,\varphi)}{\partial\varphi}\Big|_{(\theta_0,\varphi_0)} + \frac{v^2}{2n}\sum \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial\varphi^2}\Big|_{(\theta_n,\varphi_n)}$$

$$+ \left\{ \frac{u^2}{2n^2}\sum \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial\theta^2}\Big|_{(\theta_n,\varphi_n)} + \frac{uv}{n\sqrt{n}}\sum \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial\theta\,\partial\varphi}\Big|_{(\theta_n,\varphi_n)} \right\}$$

on the set $B_{n,u} = \left[X_i > \theta_0 + un^{-1} \text{ for } i = 1,2,\ldots,n \right]$
$= \left[n(Z_n - \theta_0) > u \right]$

where (θ_n,φ_n) lies in the interior of the line segment joining (θ_0,φ_0) and $(\theta_0 + un^{-1}, \varphi_0 + vn^{-1/2})$, $Z_n = \min(X_1,X_2,\ldots,X_n)$.
Also, $\Lambda_n(u,v) = 0$ on the set $\left[n(Z_n - \theta_0) < u \right]$

It is now easy to show that

$$E_{\theta_0,\varphi_0} \frac{\partial \log f(X_1,\theta,\varphi)}{\partial \theta}\Big|_{(\theta_0,\varphi_0)} = f(\theta_0,\theta_0,\varphi_0) > 0$$

and therefore

$$\frac{1}{n}\Sigma \frac{\partial \log f(X_i,\theta,\varphi)}{\partial \theta}\Big|_{(\theta_0,\varphi_0)} \xrightarrow{P^n_{\theta_0,\varphi_0}} f(\theta_0,\theta_0,\varphi_0) .$$

By assumption (B3) we have

$$\mathcal{L}\left\{ \frac{1}{\sqrt{n}}\Sigma \frac{\partial \log f(X_i,\theta,\varphi)}{\partial \varphi}\Big|_{(\theta_0,\varphi_0)} \Big| P^n_{(\theta_0,\varphi_0)}\right\} \Rightarrow N(0,I_{\theta_0}(\varphi_0))$$

and $$\frac{1}{n}\Sigma \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial \varphi^2}\Big|_{(\theta_0,\varphi_0)} \xrightarrow{P^n_{\theta_0,\varphi_0}} - I_{\theta_0}(\varphi_0)$$

where $I_{\theta}(\varphi) = E_{\theta,\varphi}\left[\frac{\partial \log f(X_1,\theta,\varphi)}{\partial \varphi}\right]^2$.

By assumption (B5) and the law of large numbers

$$\left\{ \frac{1}{n}\Sigma \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial \varphi^2}\Big|_{(\theta_n,\varphi_n)} - \frac{1}{n}\Sigma \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial \varphi^2}\Big|_{(\theta_0,\varphi_0)}\right\} 1(n(Z_n-\theta_0)>u) \xrightarrow{P^n_{\theta_0,\varphi_0}} 0 ,$$

$$\left\{ \frac{1}{n^2}\Sigma \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial \theta^2}\Big|_{(\theta_n,\varphi_n)}\right\} 1(n(Z_n-\theta_0) > u) \xrightarrow{P^n_{\theta_0,\varphi_0}} 0$$

and $$\left\{ \frac{1}{n\sqrt{n}}\Sigma \frac{\partial^2 \log f(X_i,\theta,\varphi)}{\partial \theta\, \partial \varphi}\Big|_{(\theta_n,\varphi_n)}\right\} 1(n(Z_n-\theta_0) > u) \xrightarrow{P^n_{\theta_0,\varphi_0}} 0 .$$

Now to verify that the asymptotic expansion (A) of Section 4.2 holds with

$$\Delta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left. \frac{\partial \log f(X_i,\theta,\varphi)}{\partial \varphi} \right|_{(\theta_o,\varphi_o)} \quad \text{and} \quad \tau_n = n(Z_n - \dot{\theta}_o) \; ,$$

it remains to show that

$$\text{(i)} \quad \tau_n \xrightarrow{\mathcal{L}} \tau$$

where τ is a random variable with density $f(\dot{\theta}_o,\theta_o,\varphi_o)\exp\left\{-f(\dot{\theta}_o,\theta_o,\varphi_o)x\right\}$ on $(0,\infty)$

and (ii) Δ_n and τ_n are asymptotically independent.

The proof of the convergence of $n(W_n - \theta_o)$ of Case I(a) of Section 2.2 allows us to conclude (i).

To prove (ii), we write

$$H(x) = \left. \frac{\partial \log f(x,\dot{\theta},\varphi)}{\partial \varphi} \right|_{(\dot{\theta}_o,\varphi_o)}$$

and prove for all $a \geq 0$, $b \in \mathbb{R}$,

$$\lim_{n \to \infty} P^n_{\dot{\theta}_o,\varphi_o}\left[\frac{1}{\sqrt{n}} \Sigma H(X_i) \leq b \;\middle|\; n(Z_n - \dot{\theta}_o) \geq a \right]$$

$$= \lim_{n \to \infty} P^n_{\dot{\theta}_o,\varphi_o}\left[\frac{1}{\sqrt{n}} \Sigma H(X_i) \leq b \right] \tag{4.2}$$

in three steps:

Let $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ be independent random variables each having distribution same as the conditional distribution of $H(X_1)$ given $X_1 \geq \theta_o + \frac{a}{n}$. Then the left hand side of (4.2) is equal to

$$\lim_{n \to \infty} P\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{ni} \leq b \right] .$$

Step 1. Let $E(Y_{ni}) = \mu_n$, $Var(Y_{ni}) = \sigma_n^2$, $i = 1,2,\dots,n$.

We shall prove that

$$\sqrt{n}\,\mu_n \longrightarrow 0 \quad \text{and} \quad \sigma_n^2 \longrightarrow I_{\theta_0}(\varphi_0) .$$

$$\sqrt{n}\,\mu_n = \sqrt{n}\,E\left[H(X_1) \mid X_1 \geq \theta_0 + an^{-1}\right]$$

$$= \frac{\sqrt{n}\int_{\theta_0 + an^{-1}}^{\infty} H(x)\, f(x, \theta_0, \varphi_0)\, dx}{P_{\theta_0,\varphi_0}\left(X_1 \geq \theta_0 + an^{-1}\right)}$$

$$= - \frac{\sqrt{n}\int_{\theta_0}^{\theta_0 + an^{-1}} H(x)\, f(x, \theta_0, \varphi_0)\, dx}{P_{\theta_0,\varphi_0}(X_1 \geq \theta_0 + an^{-1})} \qquad \left(\text{since } EH(X_1) = 0\right)$$

$$= \frac{-\sqrt{n}\int_{\theta_0}^{\theta_0 + an^{-1}} \left.\frac{\partial f(x,\theta,\varphi)}{\partial \theta}\right|_{(\theta_0,\varphi_0)} dx}{P_{\theta_0,\varphi_0}(X_1 \geq \theta_0 + an^{-1})}$$

$$\longrightarrow 0 \qquad \text{by assumption (B4).}$$

Also, $\sigma_n^2 = E(H^2(X_1) \mid X_1 \geq \theta_0 + an^{-1}) - \mu_n^2$.

$$\longrightarrow E\, H^2(X_1) = I_{\theta_0}(\varphi_0) .$$

Step 2. We shall prove that

$$\mathcal{L}\left\{\sum_{i=1}^{n}(Y_{ni} - \mu_n)/\sqrt{n}\,\sigma_n\right\} \Rightarrow N(0,1) .$$

Let $V_{ni} = Y_{ni} - \mu_n$, $i = 1,2,\ldots,n$.

Then $V_{n1}, V_{n2}, \ldots, V_{nn}$ are i.i.d. with $E(V_{ni}) = 0$, $E(V_{ni}^2) = \sigma_n^2$.

Also, for any $\varepsilon > 0$,

$$\frac{1}{\sigma_n^2} E\left[V_{n1}^2 \, 1_{(|V_{n1}| \geq \varepsilon \sigma_n \sqrt{n})}\right]$$

$$= \frac{E\left[(H(X_1) - \mu_n)^2 \, 1_{(|H(X_1) - \mu_n| \geq \varepsilon \sigma_n \sqrt{n} \, , \, X_1 \geq \theta_0 + a n^{-1})}\right]}{\sigma_n^2 \, P_{\theta_0,\varphi_0}\left[X_1 \geq \theta_0 + a n^{-1}\right]}$$

$\longrightarrow 0$ by dominated convergence theorem.

Hence by the central limit theorem

$$\mathcal{L}\left\{\frac{\Sigma V_{ni}}{\sqrt{n}\,\sigma_n}\right\} \Rightarrow N(0,1) .$$

<u>Step 3</u>. From the results proved in step 1 and step 2, we have

$$\mathcal{L}\left\{\frac{1}{\sqrt{n}} \Sigma Y_{ni}\right\} \Rightarrow N(0, I_{\theta_0}(\varphi_0)),$$

Since the asymptotic distribution of $\frac{1}{\sqrt{n}} \Sigma H(X_i)$ is also $N(0, I_{\theta_0}(\varphi))$ this proves (4.2).

Theorem 2 now gives a lower bound to the local asymptotic minimax risk. Our next problem is to find efficient estimators of θ and φ . A natural estimator of θ is the sample minimum Z_n . As an estimator of φ , we suggest any value $\hat{\varphi}_n$ which maximizes

$$\hat{L}(\varphi) = \prod_{i=1}^{n} f(X_i, Z_n, \varphi)$$

with respect to $\varphi \in \Phi$. Our estimator then satisfies the equation

$$\frac{\partial \log \hat{L}(\varphi)}{\partial \varphi} = 0 \, . \tag{4.3}$$

We now have the following theorem :

Theorem 3. Let $X_1, \ldots, X_n$ be i.i.d. observations from a distribution $P_{\theta,\varphi}$ with density $f(x,\theta,\varphi)$ satisfying assumptions (B1) – (B5). We also assume that for any $(\theta_0,\varphi_0) \in \textcircled{H} \times \Phi$, there exists a neighbourhood $N(\theta_0,\varphi_0)$ of (θ_0,φ_0) and a Lebesgue-integrable function $H(x)$ such that for all $(\theta,\varphi) \in N(\theta_0,\varphi_0)$ and all $x > \theta$,

$$|f(x,\theta,\varphi)| \leq H(x) \, .$$

We consider a loss function $L(.)$ for which $b_0 = b_0(\theta,\varphi)$ (as defined in condition C(i) of Section 4.2) is continuous in (θ,φ) and set

$$\hat{\theta}_n = Z_n - \frac{b(Z_n, \hat{\varphi}_n)}{n} \, .$$

Then for all (θ_0,φ_0), with P_{θ_0,φ_0} - probability 1, the equation (4.3) admits a solution $\hat{\varphi}_n$ such that

(i) $\hat{\varphi}_n \longrightarrow \varphi_0$ a.s. P_{θ_0,φ_0}

(ii) $\sqrt{n}\,(\hat{\varphi}_n - \varphi_0) - I^{-1}_{\theta_0}(\varphi_0)\,\Delta_n \xrightarrow{P^n_{\theta_0,\varphi_0}} 0$

(where $I_{\theta_0}(\varphi_0)$ and $\Delta_n = \Delta_n(\theta_0,\varphi_0)$ are as defined above)

and hence $\mathcal{L}\left\{\sqrt{n}\,(\hat{\varphi}_n - \varphi_0) \mid P^n_{\theta_0,\varphi_0}\right\} \Rightarrow N(0, I^{-1}_{\theta_0}(\varphi_0))$

(iii) for any sequence $\{\theta_n, \varphi_n\}$ satisfying $\theta_0 \leq \theta_n \leq \theta_0 + A_1 n^{-1}$ and $|\varphi_n - \varphi_0| \leq A_2 n^{-1/2}$ for any $A_1, A_2 > 0$,

$$\mathcal{L}\left\{\left(n(\hat{\theta}_n - \theta_n),\ \sqrt{n}\,(\hat{\varphi}_n - \varphi_n)\right) \mid P^n_{\theta_n,\varphi_n}\right\} \Rightarrow \mathcal{L}\left\{(X - b_o(\theta_o,\varphi_o)\ ,\ Y)\right\}$$

(where X and Y are random variables as described in Theorem 2) and hence the estimator $(\hat{\theta}_n, \hat{\varphi}_n)$ is efficient for any bounded subconvex loss function L in the sense that

$$\lim_{n\to\infty} \sup_{\substack{\theta_o \le \theta \le \theta_o + A_1 n^{-1} \\ |\varphi - \varphi_o| \le A_2 n^{-1/2}}} E_{\theta,\varphi} L\left(n(\hat{\theta}_n - \theta),\ \sqrt{n}\,(\hat{\varphi}_n - \varphi)\right)$$

$$= E\,L\left(X - b_o(\theta_o, \varphi_o)\ ,\ Y\right) .$$

Remark. The existence of the integrable function $H(x)$ is assumed to ensure that the map $(\theta,\varphi) \longrightarrow F_{\theta,\varphi}$ is continuous with respect to the Kolmogorov Smirnov distance so that there exists a strongly consistent sequence of estimators of (θ,φ) (see, Ghosh (1983)). Here $F_{\theta,\varphi}$ denotes the distribution function of $P_{\theta,\varphi}$.

Proof of Theorem 3. The proofs of (i) and (ii) are similar to the usual proof of consistency and asymptotic normality of the maximum likelihood estimators (one can see, for example, Serfling (1980)). However, the usual proofs have a small gap which can be removed by an argument given in Ghosh (1983).

We fix (θ_o,φ_o) which is regarded as the true parameter point. By assumptions (B2) and (B5), we have for all φ in a neighbourhood $S(\varphi_o)$ of φ_o ,

$$\frac{1}{n}\ \frac{\partial \log \hat{L}(\varphi)}{\partial \varphi} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(X_i,Z_n,\varphi)}{\partial \varphi}\Bigg|_{\varphi_o} + \frac{1}{n}(\varphi - \varphi_o)\sum_{i=1}^{n} \frac{\partial^2 \log f(X_i,Z_n,\varphi)}{\partial \varphi^2}\Bigg|_{\varphi_o}$$

$$+ \frac{1}{2}(\varphi - \varphi_o)^2\ \xi\ \frac{1}{n}\sum_{i=1}^{n} H_4(X_i)$$

where ξ is a random variable such that $|\xi| \leq 1$.

We write

$$\frac{1}{n} \frac{\partial \log \hat{L}(\varphi)}{\partial \varphi} = A_n + (\varphi - \varphi_o) B_n + \frac{1}{2} (\varphi - \varphi_o)^2 \xi C_n + \varepsilon_n(\varphi)$$

where

$$A_n = \frac{1}{n} \Sigma \left. \frac{\partial \log f(X_i, \theta_o, \varphi)}{\partial \varphi} \right|_{\varphi_o} \longrightarrow 0 \quad \text{a.s.},$$

$$B_n = \frac{1}{n} \Sigma \left. \frac{\partial^2 \log f(X_i, \theta_o, \varphi)}{\partial \varphi^2} \right|_{\varphi_o} \longrightarrow - I_{\theta_o}(\varphi_o) \quad \text{a.s.},$$

$$C_n = \frac{1}{n} \Sigma H_4(X_i) \longrightarrow E\, H_4(X_1) \quad \text{a.s.},$$

$$\varepsilon_n(\varphi) = \frac{1}{n} \Sigma \left. \frac{\partial \log f(X_i, Z_n, \varphi)}{\partial \varphi} \right|_{\varphi_o} - \frac{1}{n} \Sigma \left. \frac{\partial \log f(X_i, \theta_o, \varphi)}{\partial \varphi} \right|_{\varphi_o}$$

$$+ \frac{1}{n} (\varphi - \varphi_o) \Sigma \left. \frac{\partial^2 \log f(X_i, Z_n, \varphi)}{\partial \varphi^2} \right|_{\varphi_o} - \frac{1}{n} (\varphi - \varphi_o) \Sigma \left. \frac{\partial^2 \log f(X_i, \theta, \varphi)}{\partial \varphi^2} \right|_{\varphi_o}.$$

Since $Z_n \longrightarrow \theta_o$ a.s., for any fixed φ in $S(\varphi_o)$, $\varepsilon_n \longrightarrow 0$ a.s. by assumption (B5).

Let $\varepsilon > 0$ be given such that $\varepsilon < I_{\theta_o}(\varphi_o) / E\, H(X_1)$ and $\varphi_1 = \varphi_o - \varepsilon$, $\varphi_2 = \varphi_o + \varepsilon$ lie in $S(\varphi_o)$.

Then

$$\left| \frac{1}{n} \left. \frac{\partial \log \hat{L}(\varphi)}{\partial \varphi} \right|_{\varphi_1} - I_{\theta_o}(\varphi_o)\, \varepsilon \right| \leq |A_n| + \varepsilon\, |B_n + I_{\theta_o}(\varphi_o)| + \frac{1}{2} \varepsilon^2 C_n + |\varepsilon_n|$$

and

$$\left| \frac{1}{n} \left. \frac{\partial \log \hat{L}(\varphi)}{\partial \varphi} \right|_{\varphi_2} + I_{\theta_o}(\varphi_o)\, \varepsilon \right| \leq |A_n| + \varepsilon\, |B_n + I_{\theta_o}(\varphi_o)| + \frac{1}{2} \varepsilon^2 C_n + |\varepsilon_n|.$$

Now, on a set of probability one, for all sufficiently large n,

$$|A_n| + \varepsilon \, |B_n + I_{\theta_0}(\varphi_0)| + \frac{1}{2}\varepsilon^2 C_n + |\varepsilon_n| < \frac{3}{4} I_{\theta_0}(\varphi_0)\, \varepsilon$$

and therefore,

$$\frac{1}{n} \left.\frac{\partial \log \hat{L}(\varphi)}{\partial \varphi}\right|_{\varphi_0 - \varepsilon} > 0 \quad \text{and} \quad \frac{1}{n} \left.\frac{\partial \log \hat{L}(\varphi)}{\partial \varphi}\right|_{\varphi_0 + \varepsilon} < 0 .$$

Since $\dfrac{\partial \log \hat{L}(\varphi)}{\partial \varphi}$ is continuous in φ, the interval $(\varphi_0 - \varepsilon, \varphi_0 + \varepsilon)$ contains a solution of the equation (4.3).

Following Serfling (1980) we shall now choose a particular solution $\hat{\varphi}_{n,\varepsilon}$ of the equation (4.3), lying in $[\varphi_0 - \varepsilon, \varphi_0 + \varepsilon]$ and construct a sequence of estimators $\hat{\varphi}_n$ using these $\{\hat{\varphi}_{n,\varepsilon}, \varepsilon > 0\}$. We remove the gap in the proof given in Serfling (as pointed out in Ghosh (1983)) by using a strongly consistent sequence of estimators φ_n^*. Under our assumptions, such a sequence φ_n^* may be obtained by using Le Cam's construction (Ghosh (1983)).

We define

$$\hat{\varphi}_{n,\varepsilon} = \inf \left\{ \varphi : \varphi_n^* - 2\varepsilon \le \varphi \le \varphi_n^* + 2\varepsilon , \; \frac{\partial \log \hat{L}(\varphi)}{\partial \varphi} = 0 \right\} .$$

We note that on a set of probability one, for all sufficiently large n,

$$\left[\varphi_n^* - 2\varepsilon, \varphi_n^* + 2\varepsilon\right] \supset \left[\varphi_0 - \varepsilon, \varphi_0 + \varepsilon\right]$$

and hence the set

$$\left\{ \varphi : \varphi_n^* - 2\varepsilon \le \varphi \le \varphi_n^* + 2\varepsilon , \; \frac{\partial \log \hat{L}(\varphi)}{\partial \varphi} = 0 \right\}$$

is nonempty. We can now define a sequence $\hat{\varphi}_n$ as in Serfling (1980) such that $\hat{\varphi}_n \longrightarrow \varphi_0$ a.s. P_{θ_0,φ_0}. This proves (i).

To prove (ii) we write

$$0 = \frac{1}{n} \frac{\partial \log \hat{L}(\varphi)}{\partial \varphi}\Big|_{\hat{\varphi}_n} = A_n + B_n(\hat{\varphi}_n - \varphi_o) + \frac{1}{2} \xi \, C_n(\hat{\varphi}_n - \varphi_o)^2 + \varepsilon_n(\hat{\varphi}_n) \quad (4.4)$$

Since

$$\sqrt{n}\,(Z_n - \theta_o) \xrightarrow{P^n_{\theta_o,\varphi_o}} 0 ,$$

by assumption (A5) we have

$$\sqrt{n}\,\varepsilon_n(\hat{\varphi}_n) \xrightarrow{P^n_{\theta_o,\varphi_o}} 0 .$$

Also, since

$$\hat{\varphi}_n \longrightarrow \varphi_o \text{ a.s.}, \quad B_n \longrightarrow -I_{\theta_o}(\varphi_o) \text{ a.s.}, \quad C_n \longrightarrow E\, H_4(X_1) \text{ a.s.},$$

we have from (4.4)

$$\sqrt{n}\,(\hat{\varphi}_n - \varphi_o) - I^{-1}_{\theta_o}(\varphi_o)\,\Delta_n \xrightarrow{P^n_{\theta_o,\varphi_o}} 0 .$$

This proves (ii) .

We shall now prove (iii). We first note that for any $A_1, A_2 > 0$,

$$\lim_{n \to \infty} \sup_{\substack{\theta_o \le \theta \le \theta_o + A_1 n^{-1} \\ |\varphi - \varphi_o| \le A_2 n^{-1/2}}} E_{\theta,\varphi} L(n(\hat{\theta}_n - \theta), \sqrt{n}\,(\hat{\varphi}_n - \varphi))$$

$$= \lim_{n \to \infty} E_{\theta_n,\varphi_n} L(n(\hat{\theta}_n - \theta_n), \sqrt{n}\,(\hat{\varphi}_n - \varphi_n)) \quad (4.5)$$

for some sequence (θ_n, φ_n) satisfying

$$\theta_o \le \theta_n \le \theta_o + A_1 n^{-1} \quad \text{and} \quad |\varphi_n - \varphi_o| \le A_2 n^{-1/2} .$$

We write

$$\theta_n = \theta_o + u_n n^{-1} , \quad \varphi_n = \varphi_o + v_n n^{-1/2}$$

where $\quad 0 \le u_n \le A_1 , \quad |v_n| \le A_2$.

We can now easily verify

(1) $\mathcal{L}\left\{\Lambda_n(u_n,v_n) \mid P^n_{\theta_o,\varphi_o}\right\}$, $n \geq 1$ is relatively compact

and

(2) for any subsequence $\{m\} \subset \{n\}$ for which $\mathcal{L}\left\{\Lambda_m(u_m,v_m) \mid P^n_{\theta_o,\varphi_o}\right\}$ converges to some distribution F, we have $\int x \, dF(x) = 1$.

To do this, we use the Taylor's series expansion of $\log \Lambda_n(u_n,v_n)$ and proceed exactly as is done in the verification of the asymptotic expansion (A).

Hence $P^n_{\theta_n,\varphi_n}$ is contiguous to $P^n_{\theta_o,\varphi_o}$.

From (ii) we have

$$\mathcal{L}\left\{\left(n(\hat{\theta}_n-\theta_o) , \sqrt{n}\,(\hat{\varphi}_n-\varphi_o)\right) \mid P^n_{\theta_o,\varphi_o}\right\} \Rightarrow \mathcal{L}\left\{(X - b_o,Y)\right\} .$$

We can find the limit of the joint distribution of $n(\hat{\theta}_n-\theta_n)$, $\sqrt{n}\,(\hat{\varphi}_n-\varphi_n)$ and $\Lambda_n(u_n,v_n)$ under $P^n_{\theta_o,\varphi_o}$. We assume without loss of generality that $\{u_n\}$ and $\{v_n\}$ are convergent sequences and use the fact that $\sqrt{n}\,(\hat{\varphi}_n - \varphi_o)$ is asymptotically equivalent to $I^{-1}_{\theta_o}(\varphi_o)\,\Delta_n$. Since $P^n_{\theta_n,\varphi_n}$ is contiguous to $P^n_{\theta_o,\varphi_o}$, by a well known result on contiguity (Roussas (1972, p.33, Theorem 7.1)) we have

$$\mathcal{L}\left\{\left(n(\hat{\theta}_n - \theta_n), \sqrt{n}\,(\hat{\varphi}_n - \varphi_n)\right) \mid P^n_{\theta_n,\varphi_n}\right\} \Rightarrow \mathcal{L}\left\{(X - b_o, Y)\right\}$$

(The theorem in Roussas (1972) states the result for the case when two sequences of probabilities $\{P_n\}$ and $\{P'_n\}$ are contiguous to each other. It is easy to see that his proof works also for the case when only $\{P'_n\}$ is contiguous to $\{P_n\}$.)

Since L is bounded and subconvex, the result now follows from (4.5). ///

REFERENCES

Basawa, I.V. and Prakasa Rao, B.L.S. (1980). Statistical Inference for Stochastic Processes. Academic Press, New York.

Bernstein, S. (1937). Theory of Probability (Russian).

Bickel, P.J. and Yahav, J. (1969). Some contributions to the asymptotic theory of Bayes solutions. Z. Wahrsch. verw. Gebiete, 11, 257-275.

Blackwell, D. (1953). Equivalent comparisons of experiments. Ann. Math. Statist., 24, 265-272.

Borwanker, J.D., Kallianpur, G. and Prakasa Rao, B.L.S. (1971). The Bernstein-von Mises theorem for Markov Processes. Ann. Math. Statist., 42, 1241-1253.

Chao, M.T. (1970). The asymptotic behaviour of Bayes estimators. Ann. Math. Statist., 41, 601-609.

Cheng, R.C.H. and Iles, T.C. (1987). Corrected maximum likelihood in non-regular problems. J. Roy. Statist. Soc. Ser. B, 49, 95-101.

Dugue, D. (1936a). Sur le maximum de precision des estimations gaussiennes a la limite. Comptes Rendus, Paris, Vol. 202, 193-196.

Dugue, D. (1936b). Sur le maximum de precision des lois limites d'estimations. Comptes Rendus, Paris, Vol, 202, 452-454.

Dugue, D. (1937). Application des proprietes de la limite au sens du calcul des probabilites a l'etude de diverses questions d'estimation. J. Ecole Polytechnique, Vol. 3, 305-374.

Dunford, N. and Schwartz, J.T. (1958). Linear Operators, Part I : General Theory. Interscience, N.Y.

Fabian, V. and Hannan, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. Z. Wahrsch. verw. Gebiete, 59, 459-478.

Ferguson, T.S. (1967). Mathematical Statistics : A Decision Theoretic Approach. Academic Press, New York.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. Phil. Trans. Roy. Soc. Ser. A, Vol. 222, 309-368.

Fisher, R.A. (1925). Theory of statistical estimation. Proc. Cambridge Phil. Soc., Vol. 22, 700-725.

Gauss, C.F. (1809). Theoria motus corporum coelestium in sectionibus conicis solem ambientium. Hamburg.

Ghosh, J.K. (1983). Review of " Approximation Theorems of Mathematical Statistics " , by R.J. Serfling. J. Amer. Statist. Assoc., 78, 731-732.

Ghosh, J.K. (1985). Efficiency of estimates - Part I. Sankhyā , Ser. A, 47, 310-325.

Hajek, J. (1970). A characterization of limiting distributions of regular estimates. Z. Wahrsch. verw. Gebiete, 14, 323-330.

Hajek, J. (1972). Local asymptotic minimax and admissibility in estimation. Proc. 6th Berkeley Symp. Vol. 1, 175-194 (L.M. Le Cam, J. Neyman and E.J. Scott, eds.) University of California Press, Berkeley and Los Angeles.

Hajek, J. and Sidak, Z. (1967). Theory of Rank Tests. Academic Press, New York.

Ibragimov, I.A. and Hasminskii, R.Z. (1981). Statistical Estimation : Asymptotic Theory. Springer, New York.

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property. Ann. Inst. Statist. Math., 22, 1-13.

Jamison, B., Orey, S. and Pruitt, W. (1965). Convergence of weighted averages of independent random variables. Z. Wahrsch. verw. Gebiete, 4, 40-44.

Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log likelihood ratios is mixed normal. Sankhyā, Ser. A, 44, 173-212.

Laplace, P.S. (1774). Memoire sur la probabilite des causes par les evenements. Memoires de mathematique et de physique presentes a l'academie royale des sciences, par divers savants, and lus dans ses assembles 6 (Reprinted in Laplace's Oeuvres completes 8 27-65, English translation by S. Stigler, Technical Report 164, Dept. Statist., Univ. Chicago (September, 1984)).

Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. Univ. Calif. Publ. Statist., 1, 277-330.

Le Cam, L. (1955). An extension of Wald's theory of statistical decision functions. Ann. Math. Statist., 26, 69-81.

Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypotheses. Proc. 3rd Berkeley Symp., Vol. 1, 129-156.

Le Cam, L. (1958). Les proprietes asymptotiques des solutions de Bayes. Publ. Inst. Statist. Univ. Paris, 7, 17-35.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. Univ. Calif. Publ. Statist., 3, 37-98.

Le Cam, L. (1964). Sufficiency and approximate sufficiency. Ann. Math. Statist., 35, 1419-1455.

Le Cam, L. (1972). Limits of experiments. Proc. 6th Berkeley Symp. Vol. 1, 245-261 (L.M. Le Cam, J. Neyman and E.L. Scott, eds.) University of California Press, Berkeley and Los Angeles.

Millar, P.W. (1983). The minimax principle in asymptotic statistical theory. Ecole d'Ete de Probabilites de Saint-Flour XI (P.L. Hennequin, ed.), Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Neyman, J. (1949). Contribution to the theory of the χ^2- test. Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, 239-273.

Roussas, G.G. (1972). Contiguity of Probability Measures : Applications in Statistics, Cambridge University Press, Cambridge.

Samanta, T. (1986a). Local asymptotic minimax estimation in non-regular cases. Technical Report No. 19/86, Stat-Math. Div., ISI, Calcutta (To appear in Sankhyā).

Samanta, T. (1986b). The asymptotic behaviour of posterior distribution and Bayes estimators in non-reguler cases. Technical Report No. 25/86, Stat-Math. Div., ISI, Calcutta (Submitted for publication).

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

Smith, R.L. (1985). Maximum likelihood estimation in a class of non-regular cases. Biometrika, 72, 67-90.

von Mises, R. (1931). Wahrscheinlichkeitsrechnung. Springer, Berlin.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist., 20, 595-601.

Wald, A. and Wolfowitz, J. (1951). Two methods of randomization in statistics and the theory of games. Ann. Math., 53, 581-586.

Weiss, L. and Wolfowitz, J. (1974). Maximum probability estimators and related topics. Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Wilks, S.S. (1938). Shortest average confidence intervals from large samples, Ann. Math. Statist., 9, 166-175.

Wolfowitz, J. (1965). Asymptotic efficiency of the maximum likelihood estimator. Theor. Probability Appl., 10, 247-260.