SOME ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD PROCEDURES

KASALA SUBRAMANYAM

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

CALCUTTA

June, 1979

ACKNOWLEDGEMENT

I am extremely grateful to Professor J.K. Ghosh for inspiring me to work in Asymptotic Theory and helping me along at every stage of my research. I also thank him for permitting me to include our joint work in this thesis.

I wish to thank Dr. B.K. Sinha for allowing me to include here our joint work.

I am grateful to Dr. P. Bhimasankaram, Dr. Jayantha Kumar Roy and Mr. Partha Pratim Majumder for making it possible to present some of the numerical results in the last chapter of this thesis.

Finally, a special word of thanks to Mr. Kesar Singh for helping me in the preparation of this thesis.

Kasala Subramanyam

C O N T E N T S

Page

PART II

CHAPTER 0

GENERAL INTRODUCTION AND SUMMARY

This thesis consists of two parts dealing with maximum likelihood procedures in two different frameworks. In the first part (Chapters 1,2 and 3) we consider inference about a parameter which is discrete or separated in the sense that no $f(x,\theta)$ can be obtained as a "limit" of $\{f(x,\theta_i)\}$, $\theta_i \neq \theta$. (A precise definition of what is meant by "limit" is given in Chapter 1). In the second part (Chapters 4 and 5) we consider the usual estimation problem of what may be called in constrast to Part I, a continuous parameter. We assume, we have an exponential family having a density $f(x,\theta) = c(\theta) \exp\{\sum_{j=1}^{k} \beta_j(\theta)\, p_j(x)\}$ with respect to some σ-finite measure and whose natural parameters depend on the parameter of interest θ in a smooth way; such families have been called curved exponential families by Efron (Annals of Stat., 1975). Our interest in this part is to establish the "second order efficiency" of the maximum likelihood estimate (m.l.e).

A more detailed explanation of the above problems as well as a chapterwise summary of the results obtained is given below.

In Chapter 1, we define a separated parametric space where we confine our attention to the families which are homogeneous in the sense that the sample space does not depend upon the parameter. This chapter is mainly devoted to the study of asymptotic properties of the m.l.e. in the case of one parameter separated families.

Assuming the loss function to be of the form, $W(a,\theta) = 0$ if $a = \theta$ and positive otherwise, a theorem regarding asymptotic approximation of risk of the m.l.e. $R(\hat{\theta}_n,\theta)$ is established which asserts that under suitable conditions $\lim n^{-1} \log R(\hat{\theta}_n,\theta)$ exists and equals $\log \rho(\theta)$, where $1 - \rho(\theta)$ is the divergence function introduced by Chernoff (Annals of Math. Stat., 1952). The m.l.e. turns out to be asymptotically minimax under appropriate conditions; the results to this effect are contained in two theorems. The minimaxity provides a new proof for the main result of Kraft and Puri (Sankhyā, A, 1974). In general there are maximum weighted likelihood estimates (m.w.l.e) which are asymptotically better than the m.l.e $\hat{\theta}_n$, so that $\hat{\theta}_n$ is not in general asymptotically admissible. In the case when θ is assumed to be an integer and the loss to be squared error, analogues of Bhattacharya-Barankin and Cramer-Rao lower bounds for the variance of an estimate are developed and shown to be equivalent under very mild conditions. Also for any asymptotically unbiased estimate which attains the Cramer-Rao lower bound asymptotically at some θ_o, it is shown that its risk tends to infinity at some other point θ_1. Thus there is no estimate attaining the Cramer-Rao lower bound asymptotically at all θ. This provides an answer to a question raised by Hammersley (J.R.S.S., Series B, 1950). This chapter is concluded with two examples namely the normal and the Poisson families with integral means. Most of the results and the methods of this chapter have appeared in Ghosh and Subramanyam, (Sankhyā, Series A, 1975.)

The case of two parameters - one discrete and the other continuous has been taken into consideration in Chapter 2. The main result of Chapter 1 namely getting the expression for the asymptotic risk of the m.l.e has been extended here to the more general set up. The extention is applied to an example of Cox (1962) which is about deciding between Poisson and geometric distributions.

Chapter 3 is devoted to the case of non-homogeneous families of densities. The asymptotic risk of the m.l.e is given in this case also. In the case of Binomial, $B(N,p)$, p known, it is shown that the m.l.e is neither minimax nor admissible even when one restricts oneself to the case of two point parameter space only.

We now turn to Part II. Let us consider the problem of distinguishing between (asymptotically) efficient estimates. Let T_n be an estimate of θ and suppose we can expand $E_\theta(T_n)$ formally as $\theta + \frac{b(\theta)}{n} + o(n^{-1})$. Let $T_n^* = T_n - b(T_n)/n$. Then T_n^* is (formally) unbiased up to $o(n^{-1})$. We call T_n^* as the (bias) corrected estimate of θ. Let $W(a,\theta) = \min\{(a-\theta)^2, d\}$

be the squared error loss truncated at $d > 0$. Then for any (asymptotically) efficient corrected estimate T_n^*, we have under certain conditions

$$E\{W(T_n^*,\theta)\} = \frac{1}{nI} + \frac{\psi}{n^2} + o(\frac{1}{n^2})$$

where I is the (Fisher) information contained in a single observation and n is the sample size. Among all corrected estimates, ψ is minimum for the m.l.e. This was first shown by Rao (1963) for Fisher-consistent (F-c) estimates with continuous third order derivatives. Rao confined himself to random samples from multinomial populations with proportions depending on a simple parameter θ. He also considered another measure of second order efficiency based on "the loss of information" in replacing the whole sample by an estimate and arrived at the conclusion that the m.l.e is second order efficient in view of this measure also. A description of it is given in Chapter 4.

We extend the above quoted results to curved exponential families in Chapter 4. Further, these results are extended to the multiparameter case. Our proof differs from that of Rao even when specialised to the multinomial case. The result is then applied to a bio-assay problem of Berkson. An intuitive Bayesian argument for the second order efficiency of the m.l.e is given.

Most of the results and methods of this chapter have appeared in Ghosh and Subramanyam, (Sankhyā, Series A, 1974.)

Finally we turn to the Edgeworth expansions of the distributions of locally stable estimates (the definition of local stability in section 4.2) for curved exponential families which are obtained using a recent result of Bhattacharya and Ghosh (1978) in the case when the parent distribution is dominated by a measure having an absolutely continuous component. In Chapter 5, a direct comparison of the first four cumulants of an arbitrary corrected locally stable estimate with those of the corrected m.l.e. is made. This comparison yields a key probability inequality which immediately implies the second order efficiency of the m.l.e. with respect to any bounded bowl-shaped loss function. A mathematical definition of such functions is given in section 5.1. If the above said assumption regarding the dominating measure is dropped, the formal Edgeworth expansions are no longer valid. However, it turns out that if the loss function satisfies certain additional conditions the second order efficiency of the m.l.e with respect to this loss holds. The later result includes the case of the curved multinomial families.

P A R T - I

CHAPTER 1

ESTIMATION IN SEPARATED FAMILIES - ONE PARAMETER CASE

1.1 Introduction

Suppose the model consists of two separated families of densities $f(\cdot, \theta, \eta)$, $\eta \in \Omega_\theta$, $\theta = 0,1$ and we are required to test the null hypothesis $H_o(\theta = 0)$ against the alternative $H_1(\theta = 1)$. Some such problems were first pointed out by Cox(1962). These families are assumed to be separated in the sense that no density $f(\cdot,0,\eta)$ can be obtained as a limit of a sequence $\{f(\cdot,1,\eta_i)\}$ and vice versa. Cox has not explained the sense in which this limit is to be taken but the following seems to be adequate for most purposes. Denote $M = \{f(\cdot,\theta,\eta) : \eta \in \Omega_\theta, \theta=0,1\}$. Then M can be thought of a metric space with the following metric

$$d(f^{(1)}, f^{(2)}) = \int |f^{(1)} - f^{(2)}| \, d\mu$$

where μ is a σ-finite measure with respect to which the densities are taken. If $d(f^{(n)},f) \to 0$ then we say that $f^{(n)} \to f$. Thus H_o and H_1 are separated if and only if

$$\inf_{\eta, \eta'} d(f(\cdot,0,\eta), \; f(\cdot,1,\eta')) > 0 .$$

One such example is the following given by Cox (1962). Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed random variables. Let H_o be the hypothesis that the distribution is of the Poisson form and H_1 be the hypothesis that the distribution is of the geometric form. In this problem,

$$f(x,0,\eta) = e^{-\eta}\,\eta^{x}/\,x\,! \qquad x = 0,1,2,\ \ldots .$$

and

$$f(x,1,\eta') = \eta'^{x}/\,(1+\eta')^{1+x} \qquad x = 0,1,2,\ \ldots .$$

More generally let us consider decision problems where we have a countably many separated families $f(\cdot,\theta,\eta)$, $\eta \in \Omega_\theta$, $\theta \in \bar{H}$, $\bar{H}$ being a countable set. By separated we mean that for each $\theta \in \bar{H}$,

$$\inf_{\substack{\eta \in \Omega_\theta \\ \eta' \in \Omega_{\theta'} \\ \theta \neq \theta'}} d(f(\cdot,\theta,\eta),\ f(\cdot,\theta',\eta')) > 0\ .$$

We shall consider in this chapter the case where we have only one parameter θ. The general case which can be treated in a similar way is considered in Chapter 2.

Suppose we have a countable family of separated densities $\{f(\cdot,\ \theta),\ \theta \in \bar{H}\}$ with respect to a σ-finite measure μ; i.e. for each $\theta \in \bar{H}$

$$\inf_{\substack{\theta' \neq \theta \\ \theta' \in \bar{H}}} d(f(\cdot,\theta),\ f(\cdot,\theta')) > 0\ .$$

We assume that the family of densities is homogeneous in the sense that the set $\{x : f(x,\theta) > 0\}$ is independent of θ.

Define

$$\rho(\theta',\theta) = \inf_{t\geq 0} \int \{f(x,\theta')/ f(x,\theta)\}^t f(x,\theta)\, d\mu \qquad (1.1.1)$$

and

$$\rho(\theta) = \sup_{\theta'\neq\theta} \rho(\theta',\theta).$$

Note that

$$0 \leq \rho(\theta',\theta) \leq 1 .$$

Then $1 - \rho(\theta',\theta)$ is a measure of divergence introduced by Chernoff (1952). For different types of such divergence measures and their interrelationships, one can refer Khan (1973 b).

It can be shown vide Proposition 1.2.2, that $(\overline{H})$ is separated if and only if $1 - \rho(\theta) > 0$ for all θ. Whenever we say $(\overline{H})$ is separated, we mean the family of densities $\{f(\cdot,\theta) : \theta \in (\overline{H})\}$ is separated. Note that $(\overline{H})$ is separated if and only if the family of densities $\{f(\cdot,\theta) : \theta \in (\overline{H})\}$ is a discrete metric space under the metric $d(\theta', \theta) = d(f(\cdot,\theta'), f(\cdot,\theta))$ introduced earlier. The statistical problem which we consider in this chapter is to pick the correct value of θ given a random sample $X_1,X_2,\ldots,X_n$. In what follows in this chapter, we tacitly assume this set up.

In Section 1.2, some properties of the divergence function introduced in (1.1.1) are given. Let $\hat{\theta}_n$ be the maximum likelihood estimate (m.l.e) of θ based on the sample values $x_1,x_2,\ldots,x_n$. Let $w(\theta',\theta)$ be the loss in estimating the true value θ by θ'.

We assume $w(\hat{\theta}_n,\theta) = 0$ if $\hat{\theta}_n = \theta$ and positive otherwise. Let $R(\hat{\theta}_n, \theta)$ be the risk under θ. In Section 1.3, Theorem 1.3.1 provides an estimate for the asymptotic risk of the m.l.e ; it is shown that under some conditions $\lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n,\theta)$ exists and equals $\log \rho(\theta)$. Theorem 1.3.2 shows $\hat{\theta}_n$ has an asymptotic minimax property in the sense for any other estimate T_n such that $\lim_{n\to\infty} \frac{1}{n} \log R(T_n,\theta)$ exists for all θ,

$$\sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(T_n,\theta) \geq \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n,\theta).$$

From this we derive a new proof of the main result of Kraft and Puri (1974). It is shown that there are maximum weighted likelihood estimates (m.w.l.e) which are asymptotically better than $\hat{\theta}_n$, so that $\hat{\theta}_n$ is not in general asymptotically admissible. $\hat{\theta}_n$ is said to be asymptotically admissible if there cannot exist any estimate T_n such that

$$\lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta) \leq \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta)$$

with at least one strict inequality.

In Section 1.4 following Hammersley (1950) we assume that θ takes only integral values and the loss to be squared error. We develop analogues of Cramer-Rao and Bhattacharya-Barankin bounds and show that they are equivalent under very mild conditions. We show (Theorem 1.4.1) any asymptotically unbiased estimate T_n which attains Cramer-Rao bound asymptotically at some θ is bad in the sense its risk at some other point tends to infinity.

In Section 1.5 two examples are given. In the first example we consider the normal distribution with integral mean. Better estimate than the m.l.e is considered. In the second example, we consider Poisson distribution with an unknown integral mean and show that the nearest integer to the sample mean $\bar{X}_n$ (say T_n^o) is a better estimate than m.l.e. This is a rather surprising result which may be easier to accept if one notes that T_n^o is a maximum weighted likelihood estimate (m.w.l.e) if the weights π_θ are defined as follows:

$$\pi_1 = 1$$

and $\frac{1}{n} \log \pi_{\theta+1} - \frac{1}{n} \log \pi_\theta = 1 - (\theta + \frac{1}{2}) \log(1 + \frac{1}{\theta})$, for $\theta \geq 1$.

It appears that in general if the asymptotic variance is the sole criterion, m.w.l.e's rather than m.l.e's are to be preferred. This lends support to Hammersley's (1950) belief that for discrete parameters better estimates than m.l.e's may be found.

For special case of the normal there is a slight overlap between our results in this chapter and that of Laderman (1955).

A sequential approach to these problems is considered by Khan (1973 b).

1.2 Some Properties of the Divergence Function

Let $\phi(t,\theta',\theta) = E_\theta \{f(X_1,\theta')/f(X_1,\theta)\}^t$

$\equiv \phi(t)$ supressing the dependence of ϕ on θ' and θ.

Proposition 1.2.1 (a) $\rho(\theta',\theta) = \phi(t_o)$ for some $0 < t_o < 1$.

(b) $\rho(\theta',\theta) = \rho(\theta,\theta')$.

Proof: (a) This follows from the convexity of ϕ and the fact that $\phi(t) = 1$ for $t = 0,1$.

(b)
$$\rho(\theta',\theta) = \inf_{t \geq 0} E_\theta \{f(X_1,\theta')/f(X_1,\theta)\}^t$$
$$= \inf_{0 \leq t \leq 1} E_\theta \{f(X_1,\theta')/f(X_1,\theta)\}^t$$
$$= \inf_{0 \leq t \leq 1} E_{\theta'} \{f(X_1,\theta)/f(X_1,\theta')\}^t$$
$$= \rho(\theta, \theta').$$

Proposition 1.2.2 $2\{1-\rho^2(\theta',\theta)\}^{\frac{1}{2}} \geq d(\theta',\theta) \geq \{1-\rho(\theta',\theta)\}$

Proof: Khan (1973 b, Lemma 1) has shown that

$$2\{1 - \phi^2(\tfrac{1}{2})\}^{\frac{1}{2}} \geq d(\theta',\theta).$$

For, observe that

$$d^2(\theta',\theta) = [\int |f(x,\theta') - f(x,\theta)|\, d\mu]^2$$

$$\leq \int |\sqrt{f(x,\theta')} - \sqrt{f(x,\theta)}|^2 d\mu \cdot \int |\sqrt{f(x,\theta')} + \sqrt{f(x,\theta)}|^2 d\mu$$

$$= 4(1 - \phi^2(\tfrac{1}{2}))$$

$$\therefore \; 2(1 - \phi^2(\tfrac{1}{2}))^{\frac{1}{2}} \geq d(\theta',\theta).$$

This implies $2(1 - \rho^2(\theta',\theta))^{\frac{1}{2}} \geq d(\theta',\theta)$.

For the other inequality in Proposition 1.2.2 we proceed as follows Let $y \geq 0$, $0 \leq t \leq 1$. Then

$$(1+ty) - (1+y)^t = 1 + ty - \{1 + ty(1 + \xi)^{t-1}\} \qquad (1.2.1)$$

for some $0 < \xi < 1$

$$\leq y.$$

For $0 \leq y \leq 1$, $0 \leq t \leq 1$, by the concavity of $(1-y)^t$ in y, we get

$$(1-y)^t \geq (1-y).(1)^t + y(0)^t$$

$$= 1 - y \geq 1 - y - ty.$$

So, $\quad 1 - ty - (1 - y)^t \leq y \qquad (1.2.2)$

$$\text{Let} \quad y = \begin{cases} \{f(x,\theta')/ f(x,\theta)\} - 1 & \text{if} \quad f(x,\theta') \geq f(x,\theta) \\ 1 - \{f(x,\theta')/ f(x,\theta) & \text{otherwise} \end{cases}$$

Then

$$\phi(t) \geq 1 + \int_{f(x,\theta') \geq f(x,\theta)} t\, y\, f(x,\theta)\, d\mu - \int_{f(x,\theta') < f(x,\theta)} t.y\, f(x,\theta)\, d\mu$$

$$- \int y\, f(x,\theta)\, d\mu$$

$$= 1 - d(\theta',\theta) .$$

Hence $d(\theta',\theta) \geq 1 - \rho(\theta',\theta)$, completing the proof.

We shall occasionally need the Monotone Likelihood Ratio (MLR) Assumption : $(\overline{H})$ is a set of real numbers and $f(x,\theta') / f(x,\theta)$ is an increasing function of x whenever $\theta' > \theta$

Proposition 1.2.3 : Suppose the MLR assumption holds.

Then

$$\rho(\theta_1,\theta) \geq \rho(\theta_2,\theta) \quad \text{whenever} \quad \theta < \theta_1 < \theta_2 \text{ or } \theta_2 < \theta_1 < \theta .$$

Proof : We prove astronger result, namely, that for all t such that $0 \leq t \leq 1$

$$E_\theta \{ f(X , \theta_2)/f(X ,\theta)\}^t \leq E_\theta \{f(X ,\theta_1)/f(X ,\theta)\}^t$$

$$\text{if} \quad \theta < \theta_1 < \theta_2 \quad \text{or} \quad \theta_2 < \theta_1 < \theta \qquad (1.2.3)$$

Let us consider $\theta < \theta_1 < \theta_2$. Let

$$A = \frac{E_\theta\{f(X,\theta_2)/f(X,\theta)\}^t}{E_\theta\{f(X,\theta_1)/f(X,\theta)\}^t}.$$

Then

$$A = \int \{f(x,\theta_2)/f(x,\theta_1)\}^t f_2(x)\, d\mu$$

where

$$f_2(x) = \frac{\{f(x,\theta_1)/f(x,\theta)\}^t f(x,\theta)}{\int \{f(x,\theta_1)/f(x,\theta)\}^t f(x,\theta)\, d\mu}.$$

Since $0 \leq t \leq 1$ and MLR assumption holds,

$f(x,\theta_1)/f_2(x) = K\{f(x,\theta_1)/f(x,\theta)\}^{1-t}$ is an increasing function of x. Also $\{f(x,\theta_2)/f(x,\theta_1)\}^t$ is an increasing function of x. Hence

$$A \leq \int \{f(x,\theta_2)/f(x,\theta_1)\}^t f(x,\theta_1)\, d\mu \leq 1$$

which is just (1.2.3) by definition of A. We can deal with the case $\theta_2 < \theta_1 < \theta$ in a similar way. This completes the proof.

1.3 Properties of M.L.E.

Let $x = (x_1, x_2, \ldots, x_n)$ and $\hat{\theta}_n(x)$ be an m.l.e of θ. i.e. $\hat{\theta}_n$ is any element of the set $\{\theta' : \sup_{\theta} f(x,\theta'') = f(x,\theta')\}$ if this set is non-empty and $\hat{\theta}_n$ is any element of $\textcircled{H}$ otherwise.

Let $$Z_s(\theta',\theta) = \log\{f(X_s, \theta')/f(X_s, \theta)\}$$ and

$$S(\theta', \theta) = Z_1 + Z_2 + \cdots + Z_n .$$

Let $w(\theta',\theta)$ be the loss in estimating the true θ by θ'. We assume that

$$w(\theta',\theta) = 0 \quad \text{if} \quad \theta = \theta' \quad \text{and} \quad > 0 \quad \text{if} \quad \theta' \neq \theta .$$

In Theorem 1.3.1 we find an asymptotic value of the risk $R(\hat{\theta}_n, \theta) = E_\theta\{w(\hat{\theta}_n, \theta)\}$ and use it in Theorem 1.3.2 to prove a weak minimax result. The idea behind the proof of Theorem 1.3.1 is to show that $R(\hat{\theta}_n, \theta)$ behaves asymptotically like $w(\theta_1,\theta)\,[P_\theta\{\hat{\theta}_n = \theta_1\}]$ for a suitably chosen $\theta_1 \neq \theta$ so that $\frac{1}{n}\log R(\hat{\theta}_n,\theta)$ behaves like $\frac{1}{n}\log P_\theta\{\hat{\theta}_n = \theta_1\}$.

Theorem 1.3.1 : Suppose

(i) $\sum_{\theta'\neq\theta} \rho(\theta',\theta) < \infty$ and $\sum_{\theta'\neq\theta} w(\theta',\theta)\,\rho(\theta',\theta) < \infty$.

Then

$$\lim_{n\to\infty} \frac{1}{n}\log R(\hat{\theta}_n,\theta) = \log \rho(\theta).$$

Proof : Let $B_{\theta'}$ be the set $\{S(\theta',\theta) \geq 0\}$. By Chernoff's(1952) inequality

$$P_\theta(B_{\theta'}) \leq \{\rho(\theta',\theta)\}^n.$$

Hence

$$\sum_{\theta'\neq\theta} P_\theta(B_{\theta'}) \leq \sum_{\theta'\neq\theta} \{\rho(\theta',\theta)\}^n \leq \sum_{\theta'\neq\theta} \rho(\theta',\theta) < \infty \quad \text{by (i)}$$

So by the Borel-Cantelli lemma the probability that only finitely many $B_{\theta'}$'s occur is one ; i.e., with probability one

$$\hat{\theta}_n \in \{\theta' : \sup_{\theta''} f(x, \theta'') = f(x, \theta')\}.$$

Hence

$$P_\theta\{\hat{\theta}_n = \theta'\} \leq P_\theta(B_{\theta'}) \leq \{\rho(\theta',\theta)\}^n$$

and so

$$R(\hat{\theta}_n,\theta) = \sum_{\theta'\neq\theta} w(\theta',\theta)\, P\{\hat{\theta}_n = \theta'\}$$

$$\leq \sum_{\theta'\neq\theta} w(\theta', \theta)\, \{\rho(\theta',\theta)\}^n. \qquad (1.3.1)$$

Because of (i), the supremum $\rho(\theta)$ of $\rho(\theta',\theta)$, $\theta' \neq \theta$ is attained at a finite number of points $\theta' = \theta_1, \theta_2, \ldots, \theta_k$. Since

$$\sum_{\theta'\neq\theta} w(\theta',\theta)\, \{\rho(\theta',\theta)/\rho(\theta)\}^n \leq \sum_{\theta'\neq\theta} w(\theta',\theta)\, \{\rho(\theta',\theta)/\rho(\theta)\}$$

is convergent uniformly in n, we get

$$\lim_{n\to\infty} \sum_{\theta'\neq\theta} w(\theta',\theta)\{\rho(\theta',\theta)/\rho(\theta)\}^n = \sum_{\theta'\neq\theta} w(\theta',\theta) \lim_{n\to\infty}\{\rho(\theta',\theta)/\rho(\theta)\}^n$$

$$= \sum_{\theta'=\theta_1,\theta_2,\ldots,\theta_k} w(\theta',\theta) \tag{1.3.2}$$

It follows from (1.3.1) and (1.3.2) that

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log R(\hat{\theta}_n,\theta) \leq \log \rho(\theta). \tag{1.3.3}$$

Let $\theta_1,\theta_2,\ldots,\theta_k$ be as defined earlier,

$$C = \{S(\theta',\theta) > 0 \text{ for some } \theta' = \theta_1,\theta_2,\ldots,\theta_k\},$$

$$D = \bigcap_{\theta'\neq\theta,\theta_1,\theta_2,\ldots,\theta_k} B^c_{\theta'}, \quad (B^c_{\theta'} \text{ is the complement of } B_{\theta'})$$

and

$$W = \min\{w(\theta_1,\theta), w(\theta_2,\theta),\ldots,w(\theta_k,\theta)\}.$$

Then

$$R(\hat{\theta}_n,\theta) \geq W\, P_\theta\,(C \cap D)$$

$$\geq W\{P_\theta(C) - \sum_{\theta'\neq\theta,\theta_1,\theta_2,\ldots,\theta_k} P_\theta(B_{\theta'})\}$$

$$\geq W[\{\rho(\theta) - \varepsilon\}^n - \sum_{\theta'\neq\theta,\theta_1,\ldots,\theta_k}\{\rho(\theta',\theta)\}^n] \tag{1.3.4}$$

using Chernoff's (1952) theorem for any preassigned $\varepsilon > 0$ and n sufficiently large. Since by (i) $\sum_{\theta'\neq\theta}\{\rho(\theta',\theta)/\rho(\theta)\}^n$ is uniformly convergent, we get as before

$$\lim_{n\to\infty} \sum_{\theta'\neq\theta,\theta_1,\ldots,\theta_k} \{\rho(\theta',\theta)/\rho(\theta)\}^n = \sum_{\theta'\neq\theta,\theta_1,\ldots,\theta_k} \lim_{n\to\infty} \{\rho(\theta',\theta)/\rho(\theta)\}^n$$

$$= 0 . \qquad (1.3.5)$$

By (1.3.4) and (1.3,5),

$$\underline{\lim} \frac{1}{n} \log R(\hat{\theta}_n,\theta) \geq \log \{\rho(\theta) - \varepsilon\}$$

which, taken with (1.3.3) completes the proof as ε is arbitrary.

Note that the limiting value is independent of $w(\theta',\theta)$. If the loss is 0-1 then obviously assumptions (i) and (ii) are same.

Theorem 1.3.2 : Suppose the conditions of Theorem 1.3.1 hold and T_n is an estimate such that $\lim_{n\to\infty} \frac{1}{n} \log R(T_n,\theta)$ exists for all θ. Then

$$\sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(T_n,\theta) \geq \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n,\theta).$$

Proof : We consider a fixed value of θ say θ_o and define θ_1 as in the proof of Theorem 1.3.1, i.e., $\rho(\theta_1,\theta_o) = \rho(\theta_o)$. Consider a prior π which assigns positive probability $\pi_i > 0$ to θ_i, $i = 0,1$. Then the average risk is minimised by the Bayes estimate B_n which equals θ_o if

$S(\theta_1,\theta_o) < \log \{\pi_o \; w(\theta_1,\theta_o)/\pi_1 \; w(\theta_o,\theta_1)\}$ and equals θ_1 otherwise.

Since $\frac{1}{n} \log \{\pi_o\, w(\theta_1,\theta_o)/\pi_1\, w(\theta_o,\theta_1)\}$ tends to zero it follows from Chernoff's (1952) theorem that

$$\lim_{n\to\infty} \frac{1}{n} \log R(B_n,\theta_i) = \log \rho(\theta_i,\theta_j),\ i \neq j = 0,1.$$

Now using the definition of θ_1 and Proposition 1.2.1(b), we get

$$\lim_{n\to\infty} \frac{1}{n} \log R(B_n, \pi) = \log \rho(\theta_o)$$

where $R(B_n,\pi)$ is the average risk $\pi_o\, R(B_n,\theta_o) + \pi_1\, R(B_n,\theta_1)$.

Since B_n is Bayes, we get

$$\max_{i=0,1} \lim_{n\to\infty} \frac{1}{n} \log R(T_n,\theta_i) \geq \lim_{n\to\infty} \frac{1}{n} \log R(B_n, \pi)$$

$$= \log \rho(\theta_o).$$

Hence,

$$\sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(T_n,\theta) \geq \sup_{\theta} \log \rho(\theta).$$

An appeal to Theorem 1.3.1 now completes the proof.

If $\sup_{\theta} \rho(\theta) = 1$, the result is not useful. For then any estimate with $\lim_{n\to\infty} R(T_n,\theta) \leq K\ \forall\ \theta$ has this weak minimax property even though T_n need not even be consistent.

Theorem 1.3.3 gives a more natural asymptotic minimax proper of the m.l.e $\hat{\theta}_n$ if one assumes an additional condition given in (1.3.8).

<u>Theorem 1.3.3</u> Under the additional condition

$$\sup_{\theta} \left\{ \sum_{\theta' \neq \theta} w(\theta', \theta)\, \rho(\theta', \theta) \right\} < \infty \tag{1.3.7}$$

we get

$$\lim_{n\to\infty} \frac{1}{n} \sup_{\theta} \log R(\hat{\theta}_n, \theta) = \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta) \tag{1.3.8}$$

Further if (1.3.8) holds, one has a more meaningful asymptotic minimaxity of the m.l.e $\hat{\theta}_n$ namely

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \sup_{\theta} \log R(T_n, \theta) \geq \lim_{n\to\infty} \frac{1}{n} \sup_{\theta} \log R(\hat{\theta}_n, \theta) \tag{1.3.9}$$

for any estimate T_n such that $\lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta)$ exists for all θ.

<u>Proof.</u> Note that

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \sup_{\theta} \log R(\hat{\theta}_n, \theta) \geq \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta)$$

To see the other inequality we proceed as follows.

From (1.3.1) we have

$$R(\hat{\theta}_n, \theta) \leq \sum_{\theta' \neq \theta} w(\theta', \theta) \{\rho(\theta', \theta)\}^n$$

$$\leq \{\rho(\theta)\}^{n-1} \sum_{\theta' \neq \theta} w(\theta', \theta)\, \rho(\theta', \theta).$$

$$\therefore \quad \overline{\lim} \frac{1}{n} \sup_{\theta} \log R(\hat{\theta}_n, \theta) \leq \sup_{\theta} \log \rho(\theta)$$

(in view of (1.3.6))

$$= \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta)$$

using Theorem 1.3.1. When (1.3.8) is true, we get from Theorem 1.3

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \sup_{\theta} \log R(T_n,\theta) \geq \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(T_n,\theta)$$

$$\geq \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n,\theta)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sup_{\theta} \log R(\hat{\theta}_n,\theta).$$

Corollary 1.3.4: In case $w(\theta',\theta)\,\rho(\theta',\theta)$ is only a function of $|\theta' - \theta|$ then $\sum_{\theta'\neq\theta''} w(\theta',\theta'')\,\rho(\theta',\theta'') < \infty$ for any θ'' implies the stronger condition (1.3.7); and hence one has (1.3.8).

To see this corollary let $w(\theta',\theta)\,\rho(\theta',\theta) = g(|\theta' - \theta|)$. Then, for any θ ,

$$\sum_{\theta'\neq\theta} w(\theta',\theta)\,\rho(\theta',\theta) = \sum_{\theta'\neq\theta} g(|\theta'-\theta|) = 2\sum_{i=1}^{\infty} g(i) + \sum_{i=1}^{\theta} g(i) \leq 3\sum_{i=1}^{\infty} g(i)$$

$$\leq 2 \sum_{\theta'\neq\theta''} w(\theta',\theta'')\,\rho(\theta',\theta'')$$

$$= \text{a constant independent of } \theta\,;$$

this establishes the corollary.

For example in the normal case this corollary applies.

<u>Theorem 1.3.5</u> Suppose conditions (i) and (ii) of Theorem 1.3.1 hold. If moreover (1.3.8) holds and for each n there exists T_n^o such that

$$\sup_{\theta} R(T_n^o, \theta) = \inf_{T_n} \sup_{\theta} R(T_n, \theta) \qquad (1.3.10)$$

then

$$\lim_{n\to\infty} \frac{1}{n} \sup_{\theta} \log R(T_n^o, \theta) = \sup_{\theta} \log \rho(\theta). \qquad (1.3.11)$$

<u>Proof</u>: (1.3.11) follows from (1.3.10) and (1.3.9)

If $(\bar{H})$ is finite all the conditions of the Theorem 1.3.5 hold and so we get the main result of Kraft and Puri (1974).

If we make the additional assumption in Theorem 1.3.2 that $\rho(\theta_o) = \rho(\theta_1)$ for any two points θ_o and θ_1 belonging to $(\bar{H})$, then we can make the stronger assertion that

$$\lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta_o) < \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta_o)$$

at any point θ_o implies the reverse inequality

$$\lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta_1) > \lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta_1)$$

at the corresponding point θ_1, i.e., $\hat{\theta}_n$ is asymptotically admissible. This stronger result follows from the proof of Theorem 1.3.2.

Theorem 1.3.2 shows that in some sense $\hat{\theta}_n$ is asymptotically minimax. Unfortunately, $\hat{\theta}_n$ is not in general asymptotically admissible, as the following simple example shows. In fact, we can in general do better asymptotically by using some m.w.l.e. instead of m.l.e. $\hat{\theta}_n$. Suppose $(\bar{H})$ consists of just three points 1, 2 and 3 and that

$$\rho(1) = \rho(2,1) > \rho(3,1),$$

$$\rho(2) = \rho(3,2) > \rho(1,2),$$

and

$$\rho(3) = \rho(2,3) > \rho(1,3).$$

Let

$$\pi_\theta = \begin{cases} 1 & \text{if } \theta = 1 \\ \lambda^n & \text{if } \theta = 2,3 \end{cases}$$

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

Consider an estimate T_n which maximises the weighted likelihood $\pi_0 \, f(\underset{\sim}{x}, \theta)$. Let

$$\gamma(\theta, \theta') = \inf_{t \geq 0} E_\theta[\, f(X_1, \theta')\, \pi_{\theta'}^{\frac{1}{n}} / f(X_1, \theta)\, \pi_\theta^{\frac{1}{n}}]^t$$

and

$$\gamma(\theta) = \sup_{\theta' \neq \theta} \gamma(\theta', \theta).$$

Then it can be shown as in the proof of Theorem 1.3.1 that

$$\lim_{n \to \infty} \frac{1}{n} \log R(T_n, \theta) = \log \gamma(\theta).$$

Note that

$\gamma(2,3) = \rho(2,3)$ and $\gamma(3,2) = \rho(3,2)$. Also

$\gamma(1,\theta) \geq \rho(1,\theta)$ for $\theta = 2,3$

and $\gamma(\theta,1) \leq \rho(\theta,1)$ for $\theta = 2,3$.

If we choose λ sufficiently close to one so that

(i) $\gamma(2) = \max\{\gamma(1,2), \gamma(3,2)\} = \gamma(3,2)$

(since $\rho(3,2) > \rho(1,2)$)

(ii) $\gamma(3) = \max\{\gamma(1,3), \gamma(2,3)\} = \gamma(2,3)$

(since $\rho(2,3) > \rho(1,3)$),

then $\gamma(1) = \max\{\gamma(2,1), \gamma(3,1)\} < \rho(1)$.

So $\gamma(1) < \rho(1)$ and $\gamma(\theta) = \rho(\theta)$ for $\theta = 2,3$, for a suitable choice of λ. Thus T_n is asymptotically better than $\hat{\theta}_n$.

Also note that if $\bar{H}$ is finite and loss is 0-1, m.l.e is admissible for each n. Hence if any w.m.l.e T_n is asymptotically better than the m.l.e $\hat{\theta}_n$, then there exists at least one θ such that $R_n(\hat{\theta}_n,\theta) < R_n(T_n,\theta)$ which implies for each n there exists $\theta_n \in \bar{H}$ such that

$$R(\hat{\theta}_n, \theta_n) < R(T_n,\theta_n). \tag{1.3.12}$$

So there exists θ_o and a sub-sequence $\{n_i\}$ such that

$\theta_{n_i} \rightarrow \theta_o$. By (1.3.12)

$$\lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta_o) \leq \lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta_o) \qquad (1.3.13)$$

assuming $\lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta_o)$ exists. If T_n is asymptotically better than $\hat{\theta}_n$ then the inequality in (1.3.13) must be an equality.

There is no contradiction in being admissible for all n as well as being asymptotically inadmissible ; for at least one θ the m.l.e will do better than its competitor even though the asymptotic analysis will not bring this out because for this θ,

$$\lim_{n\to\infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta) = \lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta).$$

If $\underline{\overline{H}}$ is not finite but the M.L.R assumption holds, then too usually one can construct an asymptotically better m.w.l.e.

1.4 Estimation of an Integer Valued Parameter

In this section we assume $\underline{\overline{H}}$ is the set of all integers and allow an estimate to be any real number, not necessarily an integer. The loss in estimating θ by a is $(a - \theta)^2$.

1.4.1. Some results on lower bounds of the Cramer-Rao type :

Consider k+1 values $\theta_0, \theta_1, \ldots, \theta_k$ of θ.

Let $\lambda_{\theta_0,\theta}(x_1) = f(x_1,\theta)/f(x_1,\theta_0)$ for $\theta \neq \theta_0$.

Suppose $\lambda_{\theta_0,\theta}(x_1)$, $\theta = \theta_1, \theta_2, \ldots, \theta_k$ have finite variance under $\theta = \theta_0$. Then $\lambda_{\theta_0,\theta,n} = \prod_{s=1}^{n} \lambda_{\theta_0,\theta}(x_s)$, $\theta = \theta_1, \theta_2, \ldots, \theta_k$ also have finite variance under $\theta = \theta_0$. Let A_n be the correlation matrix of $\lambda_{\theta_0,\theta,n}$'s under $\theta = \theta_0$. We will assume that A_1 is non-singular. Then it can be shown that A_n is non-singular for all n. Note that

$$E_{\theta_0}(\lambda_{\theta_0,\theta,n}) = 1$$

and

$$E_{\theta_0}\{\lambda_{\theta_0,\theta,n} \cdot \lambda_{\theta_0,\theta',n}\} = \{a(\theta_0,\theta,\theta')\}^n$$

where

$$a(\theta_0,\theta,\theta') = \int \{f(x_1,\theta)/f(x_1,\theta_0)\} \cdot f(x_1,\theta')\, d\mu(x_1).$$

Clearly

$$a(\theta_0,\theta,\theta') < \{a(\theta_0,\theta,\theta) \cdot a(\theta_0,\theta',\theta')\}^{\frac{1}{2}} \text{ and } a(\theta_0,\theta,\theta) > 1.$$

Proposition 1.4.1 : Suppose $f(x,\theta)$ satisfies the M.L.R property.

If $a(\theta_0) = \inf\{a(\theta_0,\theta,\theta) : \theta \neq \theta_0, \theta = 0 \pm 1, \ldots\}$

then $a(\theta_0) = \min\{a(\theta_0,\theta_0+1,\theta_0+1), a(\theta_0,\theta_0-1,\theta_0-1)\}$.

Proof : Let $\theta > \theta_o$. Then

$$a(\theta_o,\theta,\theta) = \int \{f(x,\theta)/f(x,\theta_o)\} \cdot f(x,\theta)\, d\mu(x)$$

$$> \int \{f(x,\theta)/f(x,\theta_o)\} \cdot f(x,\theta_o+1)\, d\mu(x) \quad \text{by M.L.R property}$$

$$= \int \{f(x,\theta_o+1)/f(x,\theta_o)\} \cdot f(x,\theta)\, d\mu(x)$$

$$> \int \{f(x,\theta_o+1) \cdot f(x,\theta_o)\} \cdot f(x,\theta_o+1)\, d\mu(x)$$

$$= a(\theta_o,\theta_o+1,\ \theta_o+1).$$

Similarly if $\theta < \theta_o$, one can show that

$$a(\theta_o,\theta,\theta) > a(\theta_o,\theta_o-1,\ \theta_o-1).$$

This completes the proof.

Let T_n be any estimate with finite expectation under $\theta = \theta_o,\ \theta_1,\ldots,\theta_k$. Then

$$\mathrm{Cov}_{\theta_o}[T_n,\ \lambda_{\theta_o,\theta,n}] \doteq E_\theta(T_n) - E_{\theta_o}(T_n) \quad \text{for } \theta = \theta_1,\theta_2,\ldots,\theta_k$$

Hence considering the regression of T_n on $\lambda_{\theta_o,\theta,n}$'s we obtain

$$V_{\theta_o}(T_n) \geq \sum_{\theta=\theta_1,\theta_2,\ldots,\theta_k} \beta_{\theta,n} \frac{E_\theta(T_n) - E_{\theta_o}(T_n)}{[V_\theta(\lambda_{\theta_o,\theta,n})]^{\frac{1}{2}}} \qquad (1.4.1)$$

where $\beta_{\theta,n}$'s are given by

$$A_n(\beta_{\theta_1,n}, \beta_{\theta_2,n}, \ldots, \beta_{\theta_k,n})'$$

$$= \left(\frac{E_{\theta_1}(T_n) - E_{\theta_o}(T_n)}{V_{\theta_o}(\lambda_{\theta_o,\theta_1,n})^{1/2}}, \frac{E_{\theta_2}(T_n) - E_{\theta_o}(T_n)}{V_{\theta_o}(\lambda_{\theta_o,\theta_1,n})^{1/2}}, \ldots, \frac{E_{\theta_k}(T_n) - E_{\theta_o}(T_n)}{V_{\theta_o}(\lambda_{\theta_o,\theta_k,n})^{1/2}} \right)'$$

The r.h.s of (1.4.1) may be called a Bhattacharya-Barankin(B-B) lower bound based on $\lambda_{\theta_o,\theta,n}$'s (Bhattacharya (1946), Barankin (1949)) and if $k = 1$ we may call these Cramer-Rao (C-R) lower bounds (Cramer (1946).) Thus for any estimate T_n, the C-R lower bound is given by

$$V_{\theta_o}(T_n) \geq \frac{[E_{\theta_1}(T_n) - E_{\theta_o}(T_n)]^2}{\{a(\theta_o,\theta_1,\theta_1)\}^n - 1} \qquad (1.4.2)$$

$$= \frac{(\theta_1 - \theta_o)^2}{\{a(\theta_o,\theta_1,\theta_1\}^2 - 1} \quad \text{if} \quad E_\theta(T_n) = \theta \text{ for } \theta = \theta_o, \theta_1 .$$

Clearly the best bound of this type is obtained by maximising (1.4.2) with respect to θ_1. When the M.L.R assumption holds, the maximum occurs either at $\theta_1 = \theta_o - 1$ or $\theta_o + 1$ (vide Proposition 1.4.1), answering partly a question raised by Hammersley (1950).

Let

$$a^*(\theta_o) = \text{Min} \{ a(\theta_o,\theta,\theta) : \theta = \theta_1,\theta_2,\ldots,\theta_k \}.$$

Suppose this minimum is attained if and only if $\theta = \theta_1,\theta_2,\ldots,\theta_\lambda$. It is easy to see that A_n converges to the $k \times k$ identity matrix. If $\lim_{n\to\infty} E_\theta(T_n) = \theta$ for $\theta = \theta_o, \theta_1, \ldots, \theta_k$ then

$$\lim_{n\to\infty} \beta_{\theta,n} (a^*(\theta_o))^{\frac{n}{2}} = \begin{cases} \theta - \theta_o & \text{if } \theta = \theta_1,\theta_2,\ldots,\theta_\lambda \\ 0 & \text{if } \theta = \theta_{\lambda+1},\ldots,\theta_k . \end{cases}$$

and so

$$\frac{1}{n} \log V_{\theta_o}(T_n) \gtrsim - \log \ a^*(\theta_o).$$

Hence it follows that asymptotically B - B and C - R bounds are the same. The best asymptotically lower bound of this type may be obtained as follows.

Let $a(\theta_o)$ be as defined in Proposition 1.4.1. Note that $a(\theta_o) > 1$ since $a(\theta_o,\theta,\theta) > 1$ for all θ. Then $-\log \ a(\theta_o)$ is an asymptotic lower bound for

$$\frac{1}{n} \log V_{\theta_o}(T_n) .$$

The following result shows that if an asymptotically unbias estimate attains Cramer-Rao bound (1.4.2) asymptotically at θ then its variance tends to infinity under θ_1. Thus, there does not exist an estimate attaining Cramer-Rao lower bound asymptotically at all θ_o. This solves a problem raised by Hammersley (1950).

<u>Theorem 1.4.1</u> Suppose T_n is an estimate such that

$\lim_{n\to\infty} E_\theta(T_n) = \theta$ for $\theta = \theta_o, \theta_1$. Let $a(\theta_o, \theta_1, \theta_1) < \infty$.

If $\lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta_o) = -\log a(\theta_o, \theta_1, \theta_1)$, then

$$\lim_{n\to\infty} R(T_n, \theta_1) = \infty .$$

<u>Proof</u> : Without loss of generality, we can assume $\theta_1 > \theta_o$. We construct a test of $H_o(\theta = \theta_o)$ versus $H_1(\theta = \theta_1)$ using T_n as follows.

If $T_n - \theta_o > k$ accept H_1

$\leq k$ accept H_o

where $0 < k < \theta_1 - \theta_o$. Let α_n, β_n be the errors of first and second kind of this test. Let α'_n be the error of the first kind of the most powerful test of H_o versus H_1 which has error of the second kind equal to β_n. Note that

$$\alpha'_n \leq \alpha_n \leq E_{\theta_o} (T_n - \theta_o)^2 / k^2 . \qquad (1.4.3)$$

So by our assumption on T_n,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \alpha'_n \leq -\log a(\theta_o, \theta_1, \theta_1). \tag{1.4.4}$$

Now we will show that $\lim_{n\to\infty} \beta_n = 1$.

Suppose if possible,

$$\overline{\lim_{n\to\infty}} \beta_n < 1 .$$

Then by Stein's lemma - see Rao (1962, lemma 4.2) - we can choose a subsequence n_i such that

$$\lim_{n_i\to\infty} \frac{1}{n_i} \log \alpha'_{n_i} = -I \tag{1.4.5}$$

where

$$I = E_{\theta_1} \{ \log(f(x_1,\theta_1)/f(X_1,\theta_o))\} .$$

But $I < \log a(\theta_o,\theta_1,\theta_1)$ and so (1.4.5) contradicts (1.4.4).
Therefore

$$\underline{\lim}_{n\to\infty} \beta_n = 1$$

and hence

$$\lim_{n\to\infty} \beta_n = 1 \tag{1.4.6}$$

We now show that (1.4.6) implies $\lim_{n\to\infty} E_{\theta_1}(T_n - \theta_1)^2 = \infty$.

Let

$$\lambda_1 = E_{\theta_1}(T_n | T_n - \theta_o \leq k)$$

and

$$\lambda_2 = E_{\theta_1}(T_n | T_n - \theta_o > k)$$

Then,

$$\beta_n \lambda_1 + (1 - \beta_n) \lambda_2 = E_{\theta_1}(T_n) = \theta_1 + b(n)$$

where $b(n)$ is the bias. Note that

$$\lambda_1 \leq \theta_o + k \text{ and } b(n) \rightarrow 0 \text{. Hence}$$

$$\underline{\lim} \; (\lambda_1 - \theta_1 - b(n))^2 > 0 \text{ .} \qquad (1.4.7)$$

Now,

$$E_{\theta_1}(T_n - \theta_1)^2 \geq E_{\theta_1}(T_n - \theta_1 - b(n))^2$$

$$\geq \beta_n [\lambda_1 - \theta_1 - b(n)]^2 + (1 - \beta_n) [\lambda_2 - \theta_1 - b(n)]^2$$

by Jensen's inequality.

$$= \beta_n [\lambda_1 - \theta_1 - b(n)]^2/(1 - \beta_n)$$

which tends to infinity by (1.4.6) and (1.4.7). This completes the proof.

Both the Theorems 1.3.1 and 1.3.2 are still applicable. But the proof of Theorem 1.3.2 needs some change since the form of the Bayes estimate B_n in the present set-up will be quite different from that in Section 1.3. However, one can still show that

$$\lim_{n \to \infty} \frac{1}{n} \log R(B_n, \pi) = \log \rho(\theta_o)$$ and so the proof goes through.

We omit details. Instead we give an alternative proof to Theorem 1.3.2 using Theorem 1.4.1, in the present set-up, assuming that $\lim_{n\to\infty} E_\theta(T_n) = \theta_1$. We think that this condition can be relaxed.

<u>Proof</u> : Consider a particular value of θ, say θ_o. Let θ_1 be as defined in Theorem 1.3.1 i.e. $\rho(\theta_1,\theta_o) = \rho(\theta_o)$. Suppose

$$\lim_{n\to\infty} \frac{1}{n} \log E_{\theta_o}(T_n - \theta_o)^2 < \log \rho(\theta_1,\theta_o) = \log \rho(\theta_o). \qquad (1.4.8)$$

Then arguing as in Theorem 1.4.1 we can show

$$\lim_{n\to\infty} \frac{1}{n} \log E_{\theta_1}(T_n - \theta_1)^2 \geq \lim_{n\to\infty} \frac{1}{n} \log \beta_n$$

where β_n is the error of second kind of the most powerful test

If $S(\theta_1,\theta_o) > 0$ accept H_1 $(\theta = \theta_1)$

≤ 0 accept H_o $(\theta = \theta_o)$.

Then

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n = \log \rho(\theta_o,\theta_1) = \log \rho(\theta_1,\theta_o) = \log \rho(\theta_o) \qquad (1.4.9)$$

by Proposition 1.2.1(b).

$$\therefore \sup_{\theta} \lim_{n \to \infty} \frac{1}{n} \log E_{\theta}(T_n - \theta)^2$$

$$\geq \max \left\{ \lim_{n\to\infty} \frac{1}{n} \log E_{\theta_o}(T_n - \theta_o)^2, \right.$$

$$\left. \lim_{n\to\infty} \frac{1}{n} \log E_{\theta_1}(T_n - \theta_1)^2 \right\}$$

$$\geq \log \rho(\theta_o) \text{ using (1.4.8) and (1.4.9)}$$

$$\therefore \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log E_{\theta}(T_n - \theta)^2 \geq \sup_{\theta} \log \rho(\theta)$$

$$= \sup_{\theta} \lim_{n\to\infty} \frac{1}{n} \log E_{\theta}(\hat{\theta}_n - \theta)^2$$

by Theorem 1.3.1. This completes the proof.

1.5 Two Examples

1. Normal with integral mean :

Let $X_1, X_2, \ldots, X_n$ be independent random variables having normal distribution with known variance σ^2 and unknown mean θ, $\theta = 0, \pm 1, \pm 2, \ldots$. The m.l.e of θ is $\hat{\theta}_n$ = nearest integer to the sample mean $\bar{X}_n$. By Theorem 1.3.1,

$\lim_{n\to\infty} \frac{1}{n} \log E_{\theta}(\hat{\theta}_n - \theta)^2 = -\frac{1}{8\sigma^2}$. This was shown in a different way by Hammersley (1950). By one of the remarks following

Theorem 1.3.2, $\hat{\theta}_n$ is asymptotically admissible. Consider the class of all translation invariant estimates T_n satisfying

$$T_n(x_1 + i, \ldots, x_n + i) = T_n(x_1, \ldots, x_n) + i ,$$

for all $i = 0, \pm 1, \pm 2, \ldots$.

Let T_n^o be the best estimate in this class with respect to squared error loss. Then T_n^o is given by

$$T_n^o = (\dagger)(\bar{X}_n - i) + i \qquad \text{if} \quad i - \tfrac{1}{2} \leq \bar{X}_n < i + \tfrac{1}{2}$$

where

$$(\dagger)(x) = - \left\{ \sum_i i \; e^{-n(i+x)^2/2\sigma^2} \right\} / \left\{ \sum_i e^{-n(i+x)^2/2\sigma^2} \right\} .$$

In particular T_n^o is better than the m.l.e $\hat{\theta}_n$ so that $\hat{\theta}_n$ is neither minimax nor admissible. Khan (1973a) has shown this in a somewhat different way. Of course T_n^o is minimax and probably admissible. It can be shown that

$$\lim_{n\to\infty} \frac{1}{n} \log E_\theta (T_n^o - \theta)^2 = \lim_{n\to\infty} \frac{1}{n} \log E_\theta (\hat{\theta}_n - \theta)^2 .$$

One would not recommend T_n^o if one wants integer valued estimates. In the class of integer valued translation invariant estimates, $\hat{\theta}_n$ is the best and hence minimax as stated by Stein in the discussion following Hammersley (1950). Now it is known that $\hat{\theta}_n$ is admissible among integer valued estimates. (See

Malay Ghosh and Glen Meeden (1978)and Khan (1978)).

Theorem 1.3.1 and 1.3.2 hold also for the zero-one loss. For this loss function $\hat{\theta}_n$ is the best among all translation invariant estimates and hence minimax. Khan (1973 a) has proved $\hat{\theta}_n$ is admissible.

Also it is shown by K. Unni (1978) that minimum variance unbiased estimate of θ does not exist for any n.

2. <u>Poisson with integral mean</u>

Let $f(x,\theta) = \dfrac{e^{-\theta}\ \theta^x}{x!}$, $\theta = 1,2,\ldots$ Here

$$\rho(\theta',\theta) = \inf_{t\geq 0} \exp\{-t(\theta'-\theta) - \theta + \theta(\theta'/\theta)^t\} \qquad (1.5.1)$$

$$= \inf_{t\geq 0} \rho(\theta',\theta,t) \quad \text{(say)}.$$

This is attained for $t(\theta',\theta) = \dfrac{\log\left[\{(\theta'-\theta)/\theta\}\log(\theta'/\theta)\right]}{\log(\theta'/\theta)}$

By Proposition 1.2.3,

$$\rho(\theta) = \max\{\rho(\theta-1,\theta),\ \rho(\theta+1,\theta)\}.$$

It is shown below that

$$\rho(\theta+1,\theta) > \rho(\theta-1,\theta), \text{ so that } \rho(\theta) = \rho(\theta+1,\theta).$$

Let $0 < t < 1$. Then from (1.5.1)

$$\log \rho(\theta+1, \theta, t) = -t - \theta + \theta(1 + \frac{1}{\theta})^t$$

$$\log \rho(\theta-1, \theta, t) = +t - \theta + \theta(1 - \frac{1}{\theta})^t$$

$$\therefore \log \rho(\theta+1,\theta,t) - \log \rho(\theta-1,\theta,t) = -2t+\theta[(1+\frac{1}{\theta})^t - (1-\frac{1}{\theta})^t] \qquad (1.5.2)$$

But

$$\frac{\theta}{2t}[(1 + \frac{1}{\theta})^t - (1 - \frac{1}{\theta})^t]$$

$$= \frac{\theta}{2t}[\frac{2t}{\theta} + \frac{2t(t-1)(t-2)}{3!\,\theta^3} + \ldots] > 1 \quad \text{since} \quad 0 < t < 1.$$

Now (1.5.2) gives us

$$\rho(\theta + 1, \theta) > \rho(\theta - 1, \theta).$$

Thus

$$\rho(\theta) = \rho(\theta+1, \theta)$$

$$= \exp\{-t_o - \theta + \theta(1 + \frac{1}{\theta})^{t_o}\}$$

$$= \exp\{-t_1/\log(1+\frac{1}{\theta}) + \theta(e^{t_1} - 1)\} \qquad (1.5.$$

where $t_o = -\log\{\theta \log(1+\frac{1}{\theta})\}/\log(1+\frac{1}{\theta})$ and $t_1 = t_0 \log(1+\frac{1}{\theta})$

Theorem 1.3.1 applies. Theorem 1.3.2, though true is not useful since $\sup_\theta \rho(\theta) = 1$. However one can prove a weaker minimax property that

$$\sup_{\theta} \left\{ \lim_{n\to\infty} \frac{1}{n} \log E_\theta (T_n - \theta)^2 / |\log \rho(\theta)| \right\}$$

$$\geq \sup_{\theta} \left\{ \lim_{n\to\infty} \frac{1}{n} \log E_\theta (\hat{\theta}_n - \theta)^2 / |\log \rho(\theta)| \right\}. \tag{1.5.4}$$

It is easy to check that $t_o \to \frac{1}{2}$ as $\theta \to \infty$

$$\therefore \log \rho(\theta) = \log \rho(\theta + 1) = -t_o - \theta + \theta\left(1 + \frac{1}{\theta}\right)^{t_o}$$

$$= \frac{t_o(t_o - 1)}{2\theta} + o\left(\frac{1}{\theta}\right)$$

which tends to zero as $\theta \to \infty$. Hence $\sup_{\theta} \rho(\theta) = 1.$

Moreover $\{\log \rho(\theta)/\log \rho(\theta+1)\}$ tends to 1. Now arguing as in the proof of Theorem 1.3.2, for any estimate T_n

$$\max_{\theta'=\theta,\theta+1} \left\{ \lim_{n\to\infty} \frac{1}{n} \log E_{\theta'}(T_n - \theta')^2 \right\} \geq \log \rho(\theta)$$

$$\therefore \max_{\theta'=\theta,\theta+1} \left\{ \frac{\lim_{n\to\infty} \frac{1}{n} \log E_{\theta'}(T_n - \theta')^2}{|\log \rho(\theta')|} \right\}$$

$$\geq \min \left\{ -1, \frac{\log \rho(\theta)}{|\log \rho(\theta+1)|} \right\}$$

But the r.h.s. of above inequality tends to -1 as $\theta \to \infty$.

Hence

$$\sup_{\theta} \frac{\lim_{n\to\infty} \frac{1}{n} \log E_\theta (T_n - \theta)^2}{|\log \rho(\theta)|} \geq -1 .$$

This proves the desired result (1.5.4).

Using the technique explained in Section 1.3, one can construct a m.w.l.e T_n which is asymptotically better than $\hat{\theta}_n$ as follows.

Let

$$\pi_\theta = \begin{cases} 1 & \text{if } \theta = 1 \\ \lambda^n & \text{if } \theta > 1 \end{cases}$$

where $0 < \lambda < 1$ is to be chosen later. Consider an estimate T_n which maximises the weighted likelihood $\pi_\theta\, f(x, \theta)$.

For $\theta' \neq \theta$, let

$$\gamma(\theta', \theta) = \inf_{t \geq 0} E_\theta \left\{ \frac{f(X, \theta')\, \pi_{\theta'}^{1/n}}{f(X, \theta)\, \pi_\theta^{1/n}} \right\}^t$$

and

$$\gamma(\theta) = \sup_{\theta' \neq \theta} \gamma(\theta', \theta).$$

Then it can be shown as in the proof of Theorem 1.3.1 that

$$\lim_{n\to\infty} \frac{1}{n} \log R(T_n, \theta) = \log \gamma(\theta).$$

Since $\overline{\lim}_{\theta\to\infty} \frac{1}{\theta} \cdot \log \gamma(1,\theta) < 0$ and $\lim_{\theta\to\infty} \log \rho(\theta) = 0$, there exists θ_1 such that $\gamma(1,\theta) < \rho(\theta)$ for all $\theta \geq \theta_1$. Also since $\rho(1,\theta) < \rho(\theta + 1, \theta) = \rho(\theta)$ for all $\theta > 1$ and $\gamma(1,\theta)$ tends to $\rho(1,\theta)$ as $\lambda \to 1$ it follows that by choosing λ sufficiently close to 1, we can ensure $\gamma(1,\theta) < \rho(\theta)$ for all $2 \leq \theta \leq \theta_1$. Thus, for a proper choice of λ, $\gamma(1,\theta) < \rho(\theta)$ for all $\theta > 1$.

But $\gamma(\theta',\theta) = \rho(\theta',\theta)$ for $\theta > 1$, $\theta' > 1$. Hence it follows that $\gamma(\theta) = \rho(\theta)$ if $\theta > 1$. On the other hand by Proposition 1.2.3, $\gamma(1) = \gamma(2,1)$. Also $\gamma(2,1) < \rho(2,1)$ since $\lambda < 1$. But $\rho(2,1) = \rho(1)$. Hence $\gamma(1) < \rho(1)$. Thus we have seen that one can construct a m.w.l.e which is asymptotically better than $\hat{\theta}_n$. But it may be of interest to consider a natural competitor T_n^o = nearest integer to $\bar{X}_n$ and show that it is asymptotically better than θ_n.

Let $\theta' > \theta$. Then using Chernoff's (1952) theorem,

$$\lim_{n\to\infty} \frac{1}{n} \log P\{T_n^o \geq \theta'\} = \lim_{n\to\infty} \frac{1}{n} \log P\{\bar{X}_n \geq \theta' - \tfrac{1}{2}\}$$

$$= \log \inf_{t\geq 0} E_\theta \{e^{t(X-\theta'+\frac{1}{2})}\}$$

$$= \log \inf_{t\geq 0} \psi(t) = \log \rho^o(\theta',\theta) \text{ (say)} \tag{1.5.5}$$

where $\log(\underset{+}{})(t) = \theta(e^t-1) - t(\theta' - \frac{1}{2})$. It is easily seen that

$$\log \rho^o(\theta',\theta) = \theta \{(\theta' - \tfrac{1}{2})/\theta - 1\} - (\theta' - \tfrac{1}{2}) \log \{(\theta' - \tfrac{1}{2})/\theta\}. \quad (1.5.$$

Similarly for $\theta' < \theta$ we get

$$\lim_{n\to\infty} \frac{1}{n} \log P\{T_n^o \le \theta'\} = \theta(\theta' + \tfrac{1}{2})/\theta - 1\} - (\theta' + \tfrac{1}{2}) \log\{(\theta' + \tfrac{1}{2})/\theta\} \quad (1.5.7$$

$$= \log \rho^*(\theta',\theta) \quad \text{(say)}.$$

Further, for $\theta' > \theta$,

$$P\{T_n^o = \theta'\} \le P\{T_n^o \ge \theta'\} \le \{\rho^o(\theta',\theta)\}^n$$

and

$$P\{T_n^o = \theta'\} = P\{T_n^o \ge \theta'\} - P\{T_n^o \ge \theta'+1\}$$

$\therefore \frac{1}{n} \log P\{T_n^o = \theta'\}$ tends to $\log \rho^o(\theta',\theta)$, since $\rho^o(\theta',\theta) > \rho^o(\theta'+1,\theta)$. Also note that $\overline{\lim} \frac{1}{\theta'} \log \rho^o(\theta',\theta) < 0$. Similar statements hold for $\theta' < \theta$. Using these facts and proceeding as in the proof of Theorem 1.3.1 we get

$$\lim_{n\to\infty} \frac{1}{n} \log E_\theta(T_n-\theta)^2 = \max\{\sup_{\theta'>\theta} \log \rho^o(\theta',\theta), \sup_{\theta'<\theta} \log \rho^*(\theta',\theta)\}$$

$$= \max\{\log \rho^o(\theta+1,\theta), \log \rho^*(\theta-1,\theta)\} \quad (1.5.$$

(using Proposition 1.2.3)

From (1.5.6) and (1.5.7) we have

$$\log \rho^{o}(\theta + 1, \theta) = \frac{1}{2} - (\theta + \frac{1}{2}) \log (1 + \frac{1}{2\theta})$$

$$\log \rho^{*}(\theta - 1, \theta) = -\frac{1}{2} - (\theta - \frac{1}{2}) \log (1 - \frac{1}{2\theta})$$

Expanding $\log \rho^{o}(\theta + 1,\theta)$ and $\log \rho^{*}(\theta - 1, \theta)$ and simplifying we get

$$\log \rho^{o}(\theta+1,\theta) - \log \rho^{*}(\theta-1,\theta) = \sum_{n=1}^{\infty} (1/2\theta)^{n} [\{1/2n - 1/2n+1\}] > 0 .$$

Hence,

$$\max \{ \log \rho^{o}(\theta+1, \theta), \log \rho^{*}(\theta-1, \theta)\}$$

$$= \log \rho^{o}(\theta+1, \theta) \qquad (1.5.9)$$

Combining (1.5.8) and (1.5.9) we have

$$\lim_{n\to\infty} \frac{1}{n} \log E_{\theta}(T_n - \theta)^2 = \log \rho^{o}(\theta+1, \theta)$$

$$= \inf_{t\geq 0} (\psi)^{o}(t) \qquad \text{(using (1.5.5))}$$

where $(\psi)^{o}(t) = \theta(e^{t}-1) - t(\theta + \frac{1}{2})$.

But from (1.5.1) we get

$$\log \rho(\theta) - \log \rho^{o}(\theta+1, \theta) = \log \rho(\theta) - \inf_{t\geq 0} (\psi)^{o}(t)$$

$$\geq -t_1 \{\log(1+1/\theta)\}^{-1} + \theta(e^{t_1}-1) - (\psi)^{o}(t_1)$$

(t_1 is as defined in (1.5.3))

$$= t_1 \{(\theta + 1/2) \log(1 + 1/\theta) - 1\} / \log(1 + 1/\theta) . \qquad (1.5.10)$$

Let $f(x) = \log(1 + 1/x) - 1/(x + 1/2)$. Then

$$f'(x) = -1 / x(x+1)(2x+1)^2 < 0 \quad \forall x > 0 .$$

Also

$$\lim_{x \to \infty} f(x) = 0 .$$

So $f(x) > 0$ for all $x > 0$. Hence

$$(x + \tfrac{1}{2}) \log(1 + \tfrac{1}{x}) - 1 = (x + \tfrac{1}{2}) f(x) > 0 \ \forall x > 0 \qquad (1.5.11)$$

From (1.5.10) and (1.5.11) it follows that T_n^o is asymptotically better than $\hat{\theta}_n$ under all θ. Some explanation of this phenomenon is provided below.

Note that T_n^o maximises the weighted likelihood $\pi_\theta f(\underset{\sim}{x}, \theta)$, where π_θ's are defined recursively as follows.

$$\pi_1 = 1$$

$$\frac{1}{n} \log \pi_{\theta+1} - \frac{1}{n} \log \pi_\theta = 1 - (\theta + \tfrac{1}{2}) \log(1 + \tfrac{1}{\theta}), \ \theta \geq 1.$$

By (1.5.11) π_θ is a decreasing function of θ. So it is to be expected that under θ, T_n^o takes the value $\theta + 1$ with smaller probability than $\hat{\theta}_n$. Since the biggest contribution to the variance of T_n^o and $\hat{\theta}_n$ come from $P_\theta\{T_n^o = \theta+1\}$ and $P_\theta\{\hat{\theta}_n = \theta + 1\}$ respectively, we have here a simple explanation of the better performance of T_n^o.

CHAPTER 2

ESTIMATION IN SEPARATED FAMILIES - TWO PARAMETER CASE

2.1 Introduction

In most practical cases one has several discrete and continuous parameters. For example, if our model invloves a truncated discrete distribution with an unknown point of truncation θ_1, then θ_1 is a discrete parameter and the other parameters would be discrete or continuous, depending on the problem. The problem of binomial distribution with N = unknown number of trials and p = probability of success, has been discussed by Feldman and Fox (1968) when $N \to \infty$ in a certain way and p is known. As third example suppose we have to discriminate between normal models $N(\theta,1)$ where

$$\textcircled{H} = \{ \theta : |\theta - i| < \varepsilon \text{ for some integer } i = 0, \pm 1, \pm 2, \ldots \}$$

$$0 < \varepsilon < \frac{1}{2} .$$

Then we can think of θ_1 = nearest integer to θ as a discrete parameter and $\theta - \theta_1$ as continuous parameter. It is shown in this chapter that a result similar to Theorem 1.3.1 holds in these cases also.

To get the asymptotic variance of the m.l.e of the discrete parameter, we need a result on large deviations which is given in Proposition 2.2.1. A similar but not quite the same result can be found in Sethuraman (1964, 1970). Theorem 2.3.1 gives an expression for the asymptotic variance of the m.l.e of the discre parameter. Two examples are discussed. The first example is that considered by Cox (1962), deciding between the two distributions Poisson or geometric. Our asymptotic theory is different from that of Cox. The second example is $B(N,p)$, both N and p unknown.

2.2 A Proposition

Proposition 2.2.1 Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d random variables with common density $f(x, \eta)$ with respect to some σ-finite measure μ. Let $g(x, \eta)$ be a function defined on $R \times R$ where R is the real line $(-\infty, \infty)$. Let the following conditions hold.

(i) $g(x,\eta)$ is a continuous function of x and η.

(ii) $E_\eta \{g(X,\eta)\} < 0$ and $P_\eta\{g(X,\eta) > 0\} > 0$ for all η.

(iii) Given any bounded interval $[a,b]$, $-\infty < a < b < \infty$,

let

$$\sup_{\eta \in [a,b]} g(x,\eta) = h(x,a,b) \equiv h(x),$$

suppressing the dependence of h on a and b and

$$E_\eta \{e^{th(X)}\} < \infty \quad \text{for all} \quad t \geq 0$$

(iv) Let

$$\sup_{\eta \notin [a,b]} g(x,\eta) = \bar{h}(x,a,b) = \bar{h}(x) \quad \text{(say) and}$$

$$\varliminf_{\substack{a\to-\infty\\ b\to+\infty}} \inf_{t\geq 0} E_\eta \{e^{t\,\bar{h}(X)}\} = 0$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \log P[\sup_{\eta\varepsilon R} \sum_{s=1}^{n} g(X_s,\eta) > 0]$$

$$= \log [\sup_{\eta\varepsilon R} \inf_{0\leq t<\infty} E_\eta \{e^{tg(X,\eta)}\}].$$

We need the following Lemmas to prove the proposition.

<u>Lemma 2.2.2</u> Let the assumptions (i) and (iii) of the proposition hold. Let $-\infty < a < b < \infty$ and $0 < t_1 < \infty$. For any given $\delta > 0$, there exists a finite open cover $\{0_1, 0_2, \ldots, 0_m\}$ of $[a,b]$ such that if $\eta \varepsilon 0_j$ and $0 \leq t \leq t_1$, then

$$|E_\eta(e^{tu_j(X)}) - E_\eta(e^{tg(X,\eta)})| < \delta$$

where

$$u_j(x) = \sup_{\eta\varepsilon 0_j} g(x,\eta).$$

<u>Proof</u>. Given any $\delta > 0$, we can choose a compact set K and a finite open cover $\{0_1, 0_2, \ldots, 0_m\}$ of $[a,b]$ such that

$$|e^{tu_j(x)} - e^{tg(x,\eta)}| < \delta/2$$

for all $x \varepsilon K$, $\eta \varepsilon 0_j$, $0 \leq t \leq t_1$ and

$\int_{K^c} 2\, e^{th(x)} \; f(x,\eta)\, d\mu(x) < \delta/2$ where K^c is the the complement of K. Hence

$$E_\eta \left| e^{t\, u_j(X)} - e^{tg(X,\,\eta)} \right|$$

$$= \int_K \left| e^{tu_j(x)} - e^{tg(x,\eta)} \right| f(x,\eta)\, d\mu(x)$$

$$+ \int_{K^c} \left| e^{tu_j(x)} - e^{tg(x,\eta)} \right| f(x,\eta)\, d\mu(x)$$

$\leq \delta$ if $\eta \in O_j$ and $0 \leq t \leq t_1$.

$$\therefore \left| E_\eta(e^{t\, u_j(X)}) - E_\eta(e^{tg(X,\eta)}) \right| \leq \delta$$

if $\eta \in O_j$ and $0 \leq t \leq t_1$.

<u>Lemma 2.2.3</u> Let the assumptions (i), (ii) and (iii) of the proposition hold. Let $\phi(t,\eta) = E_\eta\{e^{tg(X,\eta)}\}$. Then there exists $0 < t_1 < \infty$ such that

$$\inf\{\phi(t,\eta) : t \geq 0\} = \inf\{\phi(t,\eta) : 0 < t < t_1\}$$

for all η.

<u>Proof</u>: By condition (iii), $\phi(t,\eta)$ is finite for all $t \geq 0$. Hence $\phi(t,\eta)$ has a well defined right derivative at $t = 0$. By (ii)

$$\phi'(0, \eta) < 0 . \tag{2.2.1}$$

Also (ii) ensures that

$$\lim_{t \to \infty} \phi(t,\eta) \geq 1 . \tag{2.2.2}$$

Note that

$$\phi(0, \eta) = 1 . \tag{2.2.3}$$

By (2.2.1) and (2.2.3) there exists $0 < t < \infty$ such that $\phi(t,\eta) < 1$. This and (2.2.2) implies $\phi(t,\eta)$ attains its minimum at a finite $t(\eta)$ (which is unique by convexity of ϕ). Clearly by (2.2.1), $t(\eta) > 0$.

We shall see below that $t(\eta)$ is a continuous function of η. Let us fix some $\eta \in (a,b)$. It follows using Dominated convergence-theorem that, under the assumptions (i), (ii) and (iii), the functions $\phi(t,\eta)$ and $d/dt\ \phi(t,\eta)$ are jointly continuous in t and η. So that, for any given compact subset S of (a,b) with η_0 as an interior point, there exists a constant $c > 0$ such that $d/dt\ \phi(t,\eta) > 0$ for $\eta \in S$ and $t > c$; hence $t(\eta) \leq c$ for all $\eta \in S$. Now, since the function $\phi(t,\eta)$ is uniformly continuous on $[0,c] \times S$, we have that $\inf_{t \in [0,c]} \phi(t,\eta)$ is continuous in $\eta \in S$. But $t(\eta)$ is a unique point that minimizes $\phi(t,\eta)$ over $t \in [0,c]$ for all $\eta \in S$; hence $t(\eta)$ is continuous at $\eta = \eta_0$. This argument can be modified easily to establish right continuity and left continuity of $t(\eta)$ at a and b respectively. Thus $t(\eta)$ is a continuous function of η. Hence the range of $t(\eta) : \eta \varepsilon [a,b]$ is bounded. The Lemma 2.2.3 follows immediately from this.

Remark 2.1.4 We need (ii) and (iii) only to conclude (2.2.1) and (2.2.2). Hence we may replace $(0, \infty)$ by another interval $[0,d]$ and in place of (ii) and (iii) assume (2.2.1) and (2.2.2) with 'd' replacing '∞' in (2.2.2).

Proof of the Proposition 2.2.1 Fix $-\infty < a < b < \infty$ and let t_1 be as in Lemma 2.2.3. Using Lemma 2.2.2, for any given $\delta > 0$, we can choose a finite open cover $\{0_1, 0_2, \ldots, 0_m\}$ of $[a, b]$ such that if $\eta \varepsilon 0_j$ and $0 \leq t \leq t_1$

$$|E_\eta e^{t\, u_j(X)} - E_\eta (e^{t\, g(X,\eta)})| < \delta \tag{2.2.4}$$

where $u_j(x)$ is as defined in Lemma 2.2.2. Let

$$0_{m+1} = R - [a,b] = \text{the complement of } [a,b].$$

Define

$$u_{m+1}(x) = \sup_{\eta \varepsilon 0_{m+1}} g(x,\eta) = \overline{h}(x).$$

Now,

$$P\left\{\sup_{\eta\in R}\sum_{s=1}^{n} g(X_s, \eta) > 0\right\}$$

$$\leq P\left\{\max_{1\leq j\leq m+1}\sum_{s=1}^{n} u_j(X_s) > 0\right\}$$

$$\leq \sum_{j=1}^{m+1} P\left\{\sum_{s=1}^{n} u_j(X_s) > 0\right\}$$

$$\leq \sum_{j=1}^{m+1}\left[\inf_{t\geq 0} E_\eta(e^{t\,u_j(X_s)})\right]^n \quad \text{(using Chernoff's inequality)}$$

$$\leq \sum_{j=1}^{m}\left[\inf_{t\geq 0} E_\eta(e^{t\,g(X_s,\eta)}) + \delta\right]^n + \left[\inf_{t\geq 0} E_\eta(e^{t\,\bar{h}(x)})\right]^n \quad \text{using (2.2.4)}$$

The second term can be made arbitraily small because of assumption (iv). Therefore

$$P\left\{\sup_{\eta\in R}\sum_{s=1}^{n} g(X_s,\eta) > 0\right\}$$

$$\leq (m+1)\left[\sup_{\eta\in R}\inf_{t\geq 0} E_\eta(e^{t\,g(X_s,\eta)}) + \delta\right]^n$$

$$\therefore \overline{\lim_{n\to\infty}}\frac{1}{n}\log P\left\{\sup_{\eta\in R}\sum_{s=1}^{n} g(X_s,\eta) > 0\right\}$$

$$\leq \log\left[\sup_{\eta\in R}\inf_{0<t<\infty} E_\eta(e^{t\,g(X_s,\eta)}) + \delta\right] \qquad (2.2.5$$

On the other hand,

$$P\left\{\sup_{\eta\varepsilon R}\sum_{s=1}^{n} g(X_s,\eta) > 0\right\} \geq P\left\{\sum_{s=1}^{n} g(X_s,\eta) > 0\right\}$$

$$\therefore \underline{\lim}\ \frac{1}{n}\log P\left\{\sup_{\eta\varepsilon R}\sum_{s=1}^{n} g(X_s,\eta) > 0\right\}$$

$$\geq \lim_{n\to\infty}\frac{1}{n}\log P\left\{\sum_{s=1}^{n} g(X_s,\eta) > 0\right\}$$

$$= \log\left[\inf_{t\geq 0} E_\eta(e^{t\,g(X_s,\eta)})\right]$$

using Chernoff's theorem. Hence

$$\underline{\lim}\ \frac{1}{n}\log P\left\{\sup_{\eta\varepsilon R}\sum_{s=1}^{n} g(X_s,\ \eta) > 0\right\}$$

$$\geq \log\left[\sup_{\eta\varepsilon R}\inf_{t\geq 0} E_\eta(e^{tg(X_s,\eta)})\right] \tag{2.2 6}$$

$$= \log\left[\sup_{\eta\varepsilon R}\inf_{0<t<\infty} E_\eta(e^{tg(X_s,\eta)})\right] \quad \text{(using Lemma 2.2.3).}$$

As δ is arbitrary, from (2.2.5) and (2.2.6) the proof of the Proposition 2.2.1 is complete.

2.3 The Main Result

Suppose $X, X_1, X_2, \ldots, X_n$ are i.i.d random variables with common density $f(x,\theta,\eta)$ where $\eta\ \varepsilon\ R$ and $\theta = 0, \pm 1, \pm 2, \ldots,$ with respect to some σ-finite measure μ. Let θ_o and η_o

be fixed values of θ and η respectively. Also, let for $\theta \neq \theta_o$

$$\psi(x, \theta, \eta_o) = \log f(x, \theta, \eta_o) - \log f(x, \theta_o, \eta_o)$$

and

$$\rho(\theta, \theta_o, \eta_o) = \sup_{\eta \in R} \inf_{0<t<\infty} E_{\theta_o, \eta_o} \{e^{t\psi(x,\theta,\eta_o)}\}.$$

Define

$$\rho(\theta_o, \eta_o) = \sup_{\theta \neq \theta_o} \rho(\theta, \theta_o, \eta_o).$$

Suppose $w(\theta, \theta_o, \eta_o)$ is the loss in estimating the true value θ_o by θ, ...

$$w(\theta, \theta_o, \eta_o) = 0 \text{ if } \theta = \theta_o \text{ and } > 0 \text{ if } \theta \neq \theta_o.$$

Let $\hat{\theta}_n$ denote the m.l.e of the discrete parameter θ based on the sample values $(x_1, x_2, \ldots, x_n)$ and $R(\hat{\theta}_n, \theta_o, \eta_o)$ be the risk under θ_o.

Theorem 2.3.1 gives the asymptotic risk of the m.l.e $\hat{\theta}_n$.

<u>Theorem 2.3.1</u> Suppose the following conditions hold.

(i) For each $\theta \neq \theta_o$, $\psi(x, \theta, \eta_o)$ satisfies the conditions of Proposition 2.2.1 with $h(\theta, x)$ in place of $h(x)$. Also let $h^*(\theta, x) = \sup_{\eta \in R} \psi(x, \theta, \eta_o)$ and assume that $E_{\theta_o}\{e^{t\, h^*(\theta,x)}\} < \infty$ for all $t > 0$.

(ii) $\sum_{\theta\neq\theta_o} \rho^*(\theta,\theta_o) < \infty$

and

(iii) $\sum_{\theta\neq\theta_o} w(\theta,\theta_o,\eta_o)\, \rho^*(\theta,\theta_o) < \infty$

where

$$\rho^*(\theta,\theta_o) = \inf_{t\geq 0} E_{\theta_o,\eta_o}\{e^{th^*(\theta,X_1)}\}.$$

Then

$$\lim_{n\to\infty} \frac{1}{n} R(\hat{\theta}_n,\theta_o,\eta_o) = \log \rho(\theta_o, \eta_o).$$

<u>Proof</u>. Let $B_\theta = \{\theta : \sup_{\eta\in R} \sum_{s=1}^{n} (\dagger)(X_s,\theta,\eta_o) > 0\}$ so that

$$P(B_\theta) \leq P\{\sum_{s=1}^{n} h^*(\theta,X_s) > 0\} \leq \{\rho^*(\theta,\theta_o)\}^n$$

by Chernoff's (1952) inequality. This leads to

$$\sum_{\theta\neq\theta_o} P(B_\theta) \leq \sum_{\theta\neq\theta_o} \{\rho^*(\theta,\theta_o)\}^n < \infty \quad \text{(by (ii).}$$

So, by the Borel-Cantelli lemma, the probability that only finitely many B_θ's occur is one. This implies $\hat{\theta}_n$ is well defined with probability one. The condition $\sum_{\theta\neq\theta_o} \rho^*(\theta,\theta_o) < \infty$ implies that $\sum_{\theta\neq\theta_o} \rho(\theta,\theta_o,\eta_o) < \infty$. Hence the supremum $\rho(\theta_o,\eta_o)$ of $\rho(\theta,\theta_o,\eta_o)$,

$\theta \neq \theta_o$ is attained at a finite number of points, say $\theta = \theta_1, \theta_2, \ldots, \theta_k$.

Let δ be any preassigned positive number. Then by Proposition 2.2.1 , for any θ_{k+1} there exists $n_o = n_o(\theta_{k+1})$ such that if $n \geq n_o$

$$P(B_\theta) \leq [\rho(\theta_o,\eta_o) + \delta]^n, \quad \text{for } \theta \leq \theta_{k+1}.$$

we choose θ_{k+1} such that

$$\rho^*(\theta,\theta_o) < \rho(\theta_o,\eta_o) \quad \text{if } |\theta| \geq \theta_{k+1} .$$

Then

$$R(\hat{\theta}_n,\theta_o,\eta_o) = \sum_{\theta \neq \theta_o} w(\theta,\theta_o,\eta_o)\, P\{\hat{\theta}_n = \theta\}$$

$$\leq \sum_{\theta \neq \theta_o} w(\theta,\theta_o,\eta_o)\, P(B_\theta)$$

$$\leq [\rho(\theta_o,\eta_o) + \delta]^n \sum_{\theta=\theta_1,\theta_2,\ldots,\theta_k} w(\theta,\theta_o,\eta_o)$$

$$+ \sum_{\substack{\theta \neq \theta_1,\theta_2,\ldots,\theta_k \\ \text{and} \\ |\theta| < \theta_{k+1}}} [\rho(\theta,\theta_o,\eta_o)+\delta]^n\, w(\theta,\theta_o,\eta_o) +$$

$$+ \sum_{|\theta| > \theta_{k+1}} w(\theta,\theta_o,\eta_o)\, \{\rho^*(\theta,\theta_o)\}^n \qquad (2.3.$$

Also $\sum\limits_{\theta \neq \theta_0} w(\theta, \theta_0, \eta_0) \{ \rho^*(\theta, \theta_0)/\rho(\theta_0, \eta_0) \}^n < \infty$ because of assumption (iii). Thus $\sum\limits_{\theta \neq \theta_0} w(\theta, \theta_0, \eta_0)(\rho^*(\theta, \theta_0)/ \rho(\theta_0, \eta_0))^n$ is convergent uniformly in n and hence we get

$$\lim_{n \to \infty} \sum_{|\theta| > \theta_{k+1}} w(\theta, \theta_0, \eta_0) \{\rho^*(\theta, \theta_0)/\rho(\theta_0, \eta_0)\}^n = 0 \qquad (2.3.2)$$

So, from (2.3.1) and (2.3.2) it follows that

$$\overline{\lim} \frac{1}{n} \log R(\hat{\theta}_n, \theta_0, \eta_0) \leq \log [\rho(\theta_0, \eta_0) + \delta] \qquad (2.3.3)$$

To get the other inequality, let

$$C = \{\theta : \sup_{\eta \in R} \sum_{s=1}^{n} \psi(X_s, \theta, \eta_0) > 0 \text{ for some } \theta_1, \theta_2, \ldots, \theta_k\},$$

$$D = \bigcap_{\theta \neq \theta_1, \theta_2, \ldots, \theta_k} B_\theta^c \qquad (B_\theta^c = \text{the complement of } B_\theta),$$

and

$$w = \min \{w(\theta_1, \theta_0, \eta_0), w(\theta_2, \theta_0, \eta_0), \ldots, w(\theta_k, \theta_0, \eta_0)\}$$

Then it follows that

$$R(\hat{\theta}_n, \theta_0, \eta_0) \geq w\, P(C \cap D)$$

$$\geq w\, [P(C) - \sum_{\theta \neq \theta_1, \theta_2, \ldots, \theta_k} P(B_\theta)]$$

Let θ_{k+1} be as before. By Proposition 2.2.1, there exists

$n_1 = n_1(\theta_{k+1})$ such that for $n > n_1$

$$R(\hat{\theta}_n, \theta_o, \eta_o) \geq w_1\{[\rho(\theta_o, \eta_o) - \delta]^n - \sum_{\substack{\theta \neq \theta_1, \theta_2, \ldots, \theta_k \\ \text{and} \\ |\theta| < \theta_{k+1}}} [\rho(\theta, \theta_o, \eta_o) + \delta]^n$$

$$- \sum_{|\theta| > \theta_{k+1}} [\rho^*(\theta, \theta_o)]^n\}$$

which yields, as in the proof of (2.3.3),

$$\underline{\lim} \frac{1}{n} \log R(\hat{\theta}_n, \theta_o, \eta_o) \leq \log [\rho(\theta_o, \eta_o) - \delta] \qquad (2.3.4)$$

Since δ is arbitrary, the proof of the theorem is complete in view of (2.3.3) and (2.3.4).

2.4 Examples

1. Consider an example given by Cox (1962). Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables. Let H_f be the hypothesis that the distribution of X_1, is Poisson with probability function $e^{-\alpha} \alpha^x / x!$, $x = 0, 1, \ldots,$ $\alpha > 0$ and H_g be the hypothesis that the distribution of X_1 is geometric with probability function $\beta^\alpha/(1+\beta)^{1+x}$, $x = 0, 1, \ldots,$ $\beta > 0$. Now we reformulate the problem as follows. Suppose the discrete parameter space has only two points say 1 and 2.

H_f : $\theta = 1$ implies X_1 is Poisson

H_g : $\theta = 2$ implies X_1 is geometric.

Let $\hat{\theta}_n$ denote the m.l.e of the parameter θ based on the sample values $x_1, x_2, \ldots, x_n$. We will get the asymptotic risk of the m.l.e $\hat{\theta}_n$ when the true value of the parameter is $\theta = 1$ or 2. To do this, we introduce the following notation so that we can apply Theorem 2 .1. Denoting

$$f(x,1,\alpha) = \frac{e^{-\alpha}\alpha^x}{x!} \quad \text{and} \quad f(x,2,\beta) = \beta^x/(1+\beta)^{1+x}$$

define $\psi(x,1,\alpha)$ and $\psi(x,2,\beta)$ for a fixed value of α_o and β_o as follows.

$$\psi(x,1,\alpha) = \log f(x,2,\beta) - \log f(x,1,\alpha_c)$$

and

$$\psi(x,2,\beta) = \log f(x,1,\alpha) - \log f(x,2,\beta_o).$$

Then

$$\phi(t,\alpha) = E_{\beta_o}\{e^{t\psi(x,1,\alpha)}\}$$

$$= \sum_{x=0}^{\infty}\left(\frac{e^{-\alpha}\alpha^x/x!}{(\frac{\beta_o}{1+\beta_o})^x \frac{1}{1+\beta_o}}\right)^t (\frac{\beta_o}{1+\beta_o})^x \frac{1}{1+\beta_o} < \infty \;\forall\; 0 \leq t \leq 1.$$

Also

$$\phi(t,\alpha) \le \sum_{x=o}^{\infty} \left[\left(\frac{\beta_o}{1+\beta_o}\right)^x \frac{1}{1+\beta_o} \right]^{1-t} < \infty \ \forall \ 0 < t < 1 .$$

Therefore, using Dominated convergence theorem it follows that

$$\phi(t,\alpha) \to 0 \quad \text{as} \quad \alpha \to \infty \qquad (2.4.1$$

for all $0 < t < 1$. Similarly one can show that

$$\phi(t,\beta) = E_{\alpha_o} \{ e^{t\varphi(x,2,\beta)} \} < \infty \text{ for all } 0 \le t \le 1$$

and $\to 0$ as $\beta \to \infty$ for $0 < t < 1$. (2.4.

For any given bounded interval $[a,b]$, $-\infty < a < b < \infty$, let

$$h_1(x) = \sup_{\alpha \in [a,b]} \varphi(x,1,\alpha), \qquad \bar{h}_1(x) = \sup_{\alpha \notin [a,b]} \varphi(x,1,\alpha),$$

$$h_2(x) = \sup_{\beta \in [a,b]} \varphi(x,2,\beta) \qquad \bar{h}_2(x) = \sup_{\beta \notin [a,b]} \varphi(x,2,\beta),$$

$$h_1^*(x) = \sup_{\alpha \in (0,\infty)} \varphi(x,1,\alpha) \text{ and } h_2^*(x) = \sup_{\beta \in (0,\infty)} \varphi(x,2,\beta).$$

Then

$$E(e^{t\, h_1^*(X)}) = \sum_{x=o}^{\infty} \left(\frac{e^{-x}\, x^x/x!}{\left(\frac{\beta_o}{1+\beta_o}\right)^x \frac{1}{1+\beta_o}} \right)^t \left(\frac{\beta_o}{1+\beta_o}\right)^x \frac{1}{1+\beta_o} < \infty \qquad (2.4.$$

for $0 \le t < 1$

and

$$E(e^{t\, h_2^*(X)}) = \sum_{x=0}^{\infty} \left(\frac{(\frac{x}{1+x})^x \frac{1}{1+x}}{e^{-\alpha_0} \alpha_0^x / x!} \right)^t \frac{e^{-\alpha_0} \alpha_0^x}{x\,!} < \infty \qquad (2.4.4)$$

for $0 \leq t < 1$. Let $a = 0$ and $b = \alpha$, then for any $0 < t < 1$

$$E(e^{t\, \bar{h}_1(X)}) = \sum_{x=0}^{\alpha} \left(\frac{e^{-\alpha} \alpha^x / x!}{(\frac{\beta_0}{1+\beta_0})^x \frac{1}{1+\beta_0}} \right)^t (\frac{\beta_0}{1+\beta_0})^x \frac{1}{1+\beta_0}$$

$$+ \sum_{x=\alpha+1}^{\infty} \left(\frac{e^{-x} x^x / x!}{(\frac{\beta_0}{1+\beta_0})^x \frac{1}{1+\beta_0}} \right)^t (\frac{\beta_0}{1+\beta_0})^x \frac{1}{1+\beta_0} .$$

As $\alpha \to \infty$, the first term goes to zero because of (2.4.1) and the second term goes to zero because of (2.4.3).

Hence

$$\underset{\substack{a \to 0 \\ b \to \infty}}{\underline{\lim}} \; \inf_{t \geq 0} E(e^{t\, \bar{h}_1(X)}) = 0$$

Similarly one can show that this condition is true for $\bar{h}_2(x)$.

Also observe that

$$\phi'(0, \alpha) < 0 \;, \quad \phi'(0, \beta) < 0$$

and

$$\lim_{t \to 1} \phi(t, \alpha) \geq 1$$

$$\lim_{t \to 1} \phi(t, \beta) \geq 1$$

Hence conditions (2.2.1) and (2.2.3) of Lemma 2.2.3 are satisfied replacing the interval $[0, \infty)$ by another interval $[0,1)$. So by Remark 2.2.4 the conditions of Theorem 2.3.1 hold. Let

$$\rho(1,\alpha_o) = \sup_{\beta} \inf_{0<t<1} E \left\{ e^{t \psi(x,2,\beta)} \right\}$$

and

$$\rho(2,\beta_o) = \sup_{\alpha} \inf_{0<t<1} E \left\{ e^{t \psi(x,1,\alpha)} \right\}.$$

Then

$$\lim_{n \to \infty} \frac{1}{n} \log R(\hat{\theta}_n, 1, \alpha_o) = \log \rho(1, \alpha_o)$$

$$\lim_{n \to \infty} \frac{1}{n} \log R(\hat{\theta}_n, 2, \beta_o) = \log \rho(2, \beta_o)$$

where

$$R(\hat{\theta}_n, 1, \alpha_o) = P\left\{ T_n = 2 / i = 1, \alpha = \alpha_o \right\}$$

$$R(\hat{\theta}_n, 2, \beta_o) = P\left\{ T_n = 1 / i = 2, \beta = \beta_o \right\}.$$

2. Consider $B(N,p)$ when both N and p are unknown. Unfortunately $\psi(x,j,p_o)$ does not satisfy the condition (i)

of Proposition 2.2.1. So Theorem 2.3.1 does not apply. However one can check that one does not need this condition to show that

$$\limsup_{n\to\infty} \frac{1}{n} E[(T_n-i)^2/i, p_o] \geq \log \rho(i, p_o).$$

CHAPTER 3

ESTIMATION IN SEPARATED FAMILIES - NON HOMOGENEOUS CASE

3.1 Introduction

In Chapter 1 we have assumed that the family of densities $\{f(x,\theta) : \theta \in \bar{\bar{H}}\}$ is homogeneous. In this chapter we consider the case when they are not homogenous. One such example is $B(N, p)$, p known. In what follows we give a theorem similar to that of Theorem 1.3.1 in the nonhomogeneous case and apply this to $B(N, p)$. However, Theorem 1.3.2 does not seem to hold without additional conditions. We show that the m.l.e of N is inadmissible and not minimax even when the parameter space has only two points say $N = 1,2$. These results were taken from Ghosh and Subramanyam (1971).

3.2 The Main Theorem

The set up and the terminology is as given in Section 1.3 except that the family of densities $\{f(x,\theta) : \theta \in \bar{\bar{H}}\}$ is non-homogeneous. Theorem 3.2 gives an expression for the asymptotic risk of the m.l.e. $\hat{\theta}_n$ of θ. In deriving the upper bound for the asymptotic risk of the m.l.e, we need Chernoff's (1952) theorem whereas to get the lower bound we need Chernoff's theorem for an extended random variable.

Here we state a theorem due to Herman Chernoff (1952) in a slightly modified form and briefly indicate its extension to include the case of extended real random variables.

<u>Theorem (C)</u>. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables such that $E(X_1) < \infty$ and $S_n = X_1 + X_2 + \ldots + X_n$. Let 'a' be a constant such that $a > E(X_1)$. Let

$P_n = P\{\frac{S_n}{n} > a\}$ and $\phi(t) = E[e^{t(X_1 - a)}]$. Then $\phi(t)$ is defined for all t even though it is allowed to take the value infinity; $0 < \phi \leq \infty$ and $\phi(0) = 1$.

Let $\rho(a) = \inf\{\phi(t) : t \geq 0\}$ so that ρ lies between 0 and 1. Then

(i) $P_n \leq \rho^n(a)$

(ii) $\frac{1}{n} \log P_n \to \log \rho(a)$ as $n \to \infty$ if $P[X_1 > a] > 0$.

We observe that if $P_n' = P[\frac{S_n}{n} > a_n]$ and $a_n \to a$ as $n \to \infty$ then $\frac{1}{n} \log P_n' \to \log \rho(a)$, provided the hypotheses of Chernoff's theorem are valid.

If X_1 is an extended random variable such that $P(X_1 = \infty) = 0$ but $P(X_1 = -\infty) > 0$, then also Theorem (C) holds. However one then defines $\exp(-t. \infty) = 0$ for all $t \geq 0$. To see this we write

$$P(S_n \geq na) = P(S_n > na | S_n > -\infty) . P[(X_1 > -\infty)]^n$$

and apply Chernoff's theorem to the first quantity in the produc

This extension of Chernoff's theorem also appears in Bahadur (1969).

To apply this result, we need assumption (iii) given below, in the statement of Theorem 3.2 . Since the proof of this theore runs along the same lines as that of Theorem 1.3.1, we omit many details and briefly indicate the proof.

Theorem 3.2 Suppose the following conditions hold.

(i) $\sum_{\theta' \neq \theta} \rho(\theta', \theta) < \infty$

(ii) $\sum_{\theta' \neq \theta} w(\theta',\theta) \rho(\theta',\theta) < \infty$ and

(iii) $\rho(\theta',\theta) = \rho(\theta)$ implies $P_\theta\{Z(\theta',\theta) > 0\} > 0$,

where $Z(\theta',\theta) = \log f(x,\theta') - f(x,\theta)$. Then

$$\lim_{n \to \infty} \frac{1}{n} \log R(\hat{\theta}_n, \theta) = \log \rho(\theta)$$

Proof : Let $B_{\theta'} = \{S(\theta', \theta) \geq 0\}$.

Then $P_\theta\{\hat{\theta}_n = \theta'\} \leq P_\theta(B_{\theta'}) \leq \rho^n(\theta', \theta)$ and so

$$R(\hat{\theta}_n, \theta) \leq \sum_{\theta' \neq \theta} w(\theta', \theta)\, \rho^n(\theta', \theta) \qquad (3.2.1)$$

Because of assumption (i) the supremum $\rho(\theta)$ of $\rho(\theta', \theta)$, $\theta' \neq \theta$ is attained at a finite number of points $\theta' = \theta_1, \theta_2, \ldots, \theta_k$.

Since

$$\sum_{\theta' \neq \theta} w(\theta', \theta)\, \{\rho(\theta', \theta)/\rho(\theta)\}^n \leq \sum_{\theta' \neq \theta} w(\theta', \theta)\{\rho(\theta', \theta)/\rho(\theta)\}$$

is convergent uniformly in n (because of (ii)), we get

$$\lim_{n \to \infty} \sum_{\theta' \neq \theta} w(\theta', \theta)\, \{\rho(\theta', \theta)/\rho(\theta)\}^n = \sum_{\theta' = \theta_1, \theta_2, \ldots, \theta_k} w(\theta', \theta) \qquad (3.2.2)$$

It follows from (3.2.1) and (3.2.2) that

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log R(\hat{\theta}_n, \theta) \leq \log \rho(\theta) \qquad (3.2.3)$$

Let $\theta_1, \theta_2, \ldots, \theta_k$ be as above and

$$C = \{S(\theta', \theta) > 0 \quad \text{for some } \theta' = \theta_1, \theta_2, \ldots, \theta_k\}$$

$$D = \bigcap_{\theta' \neq \theta_1, \theta_2, \ldots, \theta_k} B^c_{\theta'} \qquad (B^c_{\theta'} \equiv \text{the complement of } B_{\theta'}) \qquad (\;)$$

$$w = \min \{ w(\theta_1, \theta), w(\theta_2, \theta), \ldots, w(\theta_k, \theta) \}.$$

Then

$$R(\hat{\theta}_n, \theta) \geq w P_\theta \{C \cap D\}$$

$$\geq w \{P_\theta(C) - \sum_{\theta' \neq \theta_1, \theta_2, \ldots, \theta_k} P_\theta(B_{\theta'}) \}$$

$$\geq w [(\rho(\theta) - \varepsilon)^n - \sum_{\theta' \neq \theta_1, \theta_2, \ldots, \theta_k} \rho^n(\theta', \theta)]$$

using Theorem (C) (we need assumption (iii) here) for any preassigned $\varepsilon > 0$ and sufficiently large n. As before, we get

$$\underline{\lim} \frac{1}{n} \log R(\hat{\theta}_n, \theta) \geq \log [\rho(\theta) - \varepsilon] \qquad (3.2.4)$$

Since ε is arbitrary, (3.2.3) and (3.2.4) complete the proof.

3.3 An Example

Let us consider the example $B(N, p)$ with unknown N and p is known $(0 < p < 1)$.

Let

$$f(x, N) = \binom{N}{x} p^x (1-p)^{N-x}, \qquad N = 1, 2, 3, \ldots$$
$$x = 0, 1, \ldots, N$$

If $b(N) = \frac{1}{a^n(N) - 1}$ is the Cramer-Rao lower bound based on

$\lambda_{N-1,N,n}$, then it is easy to see that $\frac{b(N)}{N} \rightarrow \frac{q}{np}$ as $N \rightarrow \infty$

where

$$\lambda_{N-1,N,n} = \frac{f(\underline{x}, N-1)}{f(\underline{x}, N)}, \quad \underline{x} = (x_1, x_2, \ldots, x_n)$$

and

$$a^n(N) = E_N(\lambda_{N-1,N,n}) = (1 + \frac{p}{nq})^n .$$

Thus for large N, the simple unbiased estimate

$T_n = \sum_{j=1}^{n} X_j/np$ very nearly attains this bound. However T_n is not integer valued. It would be interesting to study the behaviour of the estimate T_n^c = nearest integer to T_n.

For any fixed $N = N_o$, it is easy to check that

$\lim_{N \rightarrow \infty} f(x, N) = 0$ for $x = 0, 1, \ldots, N_o$. So there exists a finite N_1 depending on N_o such that $P_{N_o}\{\hat{N}_n > N_1\} = 0$ where $\hat{N}_n$ is the m.l.e of N based on the sample $x_1, x_2, \ldots, x_n$. For convenience we shall take the smallest member of the set of N's where the likelihood function attains its maximum. Now we can restrict our attention to $N \leq N_1$. Obviously $P_{N_o}\{Z(N, N_o) > 0\} > 0$ for $N < N_o$. Choose the largest $N_2(N_o < N_2 \leq N_1)$ such that $P_{N_o}\{Z(N_2, N_o) > 0\} > 0$. Then it can be shown that for all N_o such that $N_o < N \leq N_2$

$$P_{N_o}\{Z(N, N_o) > 0\} > 0 \qquad (3.3.1)$$

For, consider

$$\frac{f(x, N_2)}{f(x, N_o)} = \frac{\binom{N_2}{x} p^x (1-p)^{N_2-x}}{\binom{N_o}{x} p^x (1-p)^{N_o-x}}, \qquad x = 0,1,2,\ldots,N_o . \qquad (3.3.2)$$

We are given that there exists at least one value of x for which the ratio given in (3.3.2) is greater than 1. Note that

$$\frac{f(x, N_2)}{f(x, N_o)} = \frac{N_2!}{N_o!} \; \frac{(1-p)^{N_2-N_o}}{(N_2-x)\ldots(N_o+1-x)}$$

which is an increasing function of x and for $0 \leq x \leq N_o$ has the maximum value when $x = N_o$. Hence

$$\frac{f(N_o, N_2)}{f(N_o, N_o)} > 1 .$$

Now we show that

$$\frac{f(N_o, N_2-1)}{f(N_o, N_o)} > 1.$$

Note that

$$\frac{f(N_o, N_2)}{f(N_o, N_o)} > 1 \quad \text{implies} \quad (1-p) > \left[\frac{N_o\,!\,(N_2 - N_o)!}{N_2!} \right]^{\frac{1}{N_2 - N_o}}$$

which in turn implies

$$\frac{f(N_o,\ N_2-1)}{f(N_o,\ N_o)} > \left[\frac{N_2 - N_o}{N_2} \right] \left[\binom{N_2}{N_o} \right]^{\frac{1}{N_2-N_o}}$$

$$= \left\{ \left[\frac{N_2 - N_o}{N_2} \right]^{N_2 - N_o} \binom{N_2}{N_o} \right\}^{\frac{1}{N_2-N_o}} \geq 1$$

since $\left[\frac{N_2 - N_o}{N_2} \right]^{N_2-N_o} \geq \binom{N_2}{N_o}$ and $\frac{f(N_o, N_2-1)}{f(N_o,\ N_o)}$ is an increasing function of (1-p). We can now complete the proof of (3.3.1) by induction.

Define $\rho(N_o) = \max \{ \rho(N,\ N_o) : N \leq N_2 \}$

Now all the conditions of Theorem 3.2 are satisfied. Hence the asymptotic variance of the m.l.e $\hat{N}_n$ is $\rho^n(N_o)$.

Since

$$E_{N_o-1}\{ Z(N_o-1, N_o) \} = \sum_{x=1}^{N_o-1} \log \frac{f(x, N_o-1)}{f(x,\ N_o)} \; f(x, N_o-1)$$

$$\geq -\log \sum_{x=0}^{N_o-1} f(x, N_o) > 0$$

we have

$$\rho(N_o-1, N_o) = \inf_{0 \le t \le 1} E_{N_o} \{ e^{t\, Z(N_o-1, N_o)} \}.$$

Hence by Proposition 1.2.3

$$\max_{N<N_o} \rho(N, N_o) = \rho(N_o-1, N_o).$$

Note that in general

$$\rho(N_o+1, N_o) \neq \max_{N > N_o} \rho(N, N_o)$$

but equality holds only if

$$E_{N_o} \{ Z(N_o+1, N_o) \} > 0 .$$

Observe that the conditions of Theorem 1.3.2 do not hold. Moreover $\rho(N) \to 1$ as $N \to \infty$. For $\log \rho(N) \geq -\log a(N)$ where $a(N) = 1 + \frac{p}{Nq}$ as defined earlier so that $a(N) \to 1$ as $N \to \infty$. Moreover the conclusion of Theorem 1.3.2 does not seem to hold without additional conditions. It seems probable that a weaker minimax property of the type considered in example 2 of Section 1.5 holds. On the other hand even for a parametric space consisting of only two points $N = 1, 2$, it is not true that the corresponding m.l.e $\hat{N}_n$ (taking only two values 1 and 2) is minimax if $p \geq \frac{1}{2}$.. By taking resort to randomised estimates, it is enough to show that $\rho(2,1) < \rho(1,2)$.

To see this, note that

$$\rho(1,2) = \inf_{t \geq 0} \phi(t) \qquad (3.3.3)$$

where

$$\phi(t) = \sum_{x=0}^{1} \left(\frac{f(x,1)}{f(x,2)}\right)^{t} \cdot f(x,2) .$$

Since $\phi'(0) = q^2 \log q - 2pq \log(2q) > 0$ for $p \geq \frac{1}{2}$ (where $q = 1 - p$) the infimum in (3.3.3) is attained at $t = 0$. On the other hand

$$\rho(2,1) = \inf_{t \geq 0} \sum_{x=0}^{1} \left(\frac{f(x,2)}{f(x,1)}\right)^{t} f(x,1) \qquad (3.3.4)$$

$$= \inf_{-\infty \leq t \leq 1} \phi(t)$$

where $\phi(t)$ is as defined in (3.3.3). Since $\phi'(0) > 0$, the infimum in (3.3.4) for $\rho(2,1)$ is attained at a $t < 0$. This means $\rho(2,1) < \rho(1,2)$.

In the present context we can also show that the m.l.e $\hat{N}_n$ is not admissible. For, define an estimate $T_n(h)$ as follows.

$$T_n(h) = \begin{cases} 2 & \text{if } \frac{1}{n} \sum_{s=1}^{n} Z_{21}(x_s) > h \\ 1 & \text{if } \frac{1}{n} \sum_{s=1}^{n} Z_{21}(x_s) \leq h \end{cases}$$

where h is a positive constant and

$$Z_{21}(x_s) = \log \frac{f(x_s,2)}{f(x_s,1)} .$$

Then it is easy to see that

$$\rho_h(2,1) = \inf_{t\geq 0} e^{-th} \sum_{x=0}^{1} \left(\frac{f(x,2)}{f(x,1)}\right)^t f(x,1)$$

$$\rho_h(1,2) = \inf_{t\geq 0} e^{th} \sum_{x=0}^{1} \left(\frac{f(x,1)}{f(x,2)}\right)^t f(x,2).$$

Let us consider when $p = q = \frac{1}{2}$. Then we have

$$\rho_h(2,1) = \inf_{t\geq 0} e^{-th} [(\tfrac{1}{2})^{t+1} + \tfrac{1}{2}] = 0$$

$$\rho_h(1,2) = \inf_{t\geq 0} e^{th} (\frac{2^t}{4} + \frac{1}{2}) = 0.75$$

But the corresponding values for the m.l.e $\hat{N}_n$ are given by $\rho(2,1) = 0.5$ and $\rho(1,2) = 0.75$. So $\rho_h(2,1) < \rho(2,1)$ and $\rho_h(1,2) = \rho(1,2)$ which shows that $\hat{N}_n$ is not admissible.

CHAPTER 4

SECOND ORDER EFFICENCY OF MAXIMUM LIKELIHOOD ESTIMATES

4.1 Introduction

How can one distinguish between (asymptotically) efficient estimates ? Fisher (1925) proposed the calculation of

$$E_2' = \lim (nI - I_{T_n})$$

where I is the (Fisher) information contained in a single observation, I_{T_n} is the (Fisher) information contained in an estimate T_n and n is the sample size and the limit is to be taken in a suitable sense to be explained later. The quantity E_2' may be interpreted as the loss of information in replacing the sample by T_n. Smaller the value of E_2', better is the estimate. Fisher stated without any proof that maximum likelihood estimate minimises E_2'. (For some clarification of Fisher's calculation of E_2' for special estimates see Kendall (1946) and Nandi (1956)). Fisher's assertion was proved by Rao (1961) who restricts attention to the so called Fisher - consistent estimates. The term " second order efficiency" was first introduced by Rao (1961). However the result actually proved differs in two ways from what Fisher stated. Firstly Rao introduces a more easily computed and a more useful measure E_2 and secondly he restricts attention to Fisher consistent estimates with continuous second order derivatives. This result will be referred to as the Fisher-Rao theorem. We shall call Fisher consistent estimates with continuous second order derivatives, or rather, a slightly wider class, locally stable (II). The definitions of E_2 and local stability (II, III and IV) are given in Section 4.2.

The Fisher-Rao theorem has one unpleasant feature - its decision theoretic implications are far from clear. In fact this has been the main criticism against its use to justify the use of maximum likelihood estimates. Rao (1963) has, therefore, sought a direct comparision of the truncated mean squares. Let $w(a, \theta) = \min \{(a - \theta)^2, d\}$ be the squared error loss truncated at $d > 0$. (Actually Rao's loss is slightly different ; see the remark after Proposition 4.2.4. Suppose T_n is an asymptotically efficient estimate with

$$E_\theta \{ w(T_n, \theta) \} = \frac{1}{n I} + \frac{\psi}{n^2} + o(\frac{1}{n^2})$$

then ψ may be taken as another measure of second order efficiency. Again restricting to Fisher consistent estimates, (with third order continuous derivatives instead of second order and applying a bias correction to the estimates considered, he shows ψ is minimised by the (corrected) maximum likelihood estimate. The effect of the bias correction is to make the estimates unbiased up to terms of $o(\frac{1}{n})$; we shall call this Rao's theorem.

Both these theorems pertain to the case of independent random samples from a multinomial population with proportions depending on an unknown parameter θ. Data which appears to contradict this sort of a result in a particular bio-assay problem, has been presented by Berkson (1955). Berkson's data seems to indicate that for moderate sample size his minimum logit-chisquare estimate performs better than the maximum likelihood estimate as regards mean square error. In this connection see also Berkson and Hodges (1961). A summary of the results of Berkson is available in Ferguson(1967). Since the population which Berkson considers is not multinomial, but belongs to the Korpman-Darmois exponential family it seemed to us worth extending the results of Rao (1961, 1963) to exponential families to see what is really happening in Berkson's problem.

The extension to exponential families is carried out in Section 4.2. The main idea is simple. It is shown that all locally stable efficient estimates T_n which are unbiased up to $0(1/n)$ have same covariance up to $o(1/n)$ with Z_n, Z_n^2 and Z_nW_n (which are defined in Section 4.2 on pp.85,88). Moreover $\hat{\theta}_n$ even after bias correction is easily shown to be a linear function of Z_n, Z_n^2 and Z_nW_n up to $o(1/n)$. So up to $o(1/n)$ we can write T_n as a sum of two orthogonal components the first of which is the bias corrected maximum likelihood estimate. Rao's theorem is an immediate consequence. The Fisher-Rao theorem follows similarly. The expansions given in Theorem 4.2.6 try to make clear the relation between the two types of results from the present point of view. Moreover it is shown that if T_n is an efficient l.s. (III) estimate then one can find another estimate $h(\hat{\theta}_n) = \hat{\theta}_n + g(\hat{\theta}_n)/n$ such that

$$E_\theta \{W(T_n,\theta)\} \geq E_\theta \{W(h(\hat{\theta}_n), \theta)\} + o(1/n^2) \ \forall\ \theta .$$

It is pointed out in Theorem 4.2.6 that the maximum likelihood estimate is not unique in enjoying this asymptotic optimum property. Any estimate which differs from the maximum likelihood estimate up to $o(1/n)$ has the same property. Theorem 4.2.6 is extended to the multiparameter case in Theorem 4.2.11. An asymptotic Bhattacharya bound is developed and necessary and sufficient conditions are given for the maximum likelihood estimate to attain it. The calculations in this section, though similar to Rao's are, we believe, somewhat simpler and more illuminating even when specialized to the multinomial case. Some remarks are given in Section 4.3.

In Section 4.4 these results are applied to Berkson's problem. It is shown that if a correction is made to the maximum likelihood estimate so that its bias is same as that of Berkson's minimum logit-chi-square estimate up to terms $O(1/n)$, then the maximum likelihood estimate has a smaller variance up to the terms of $O(1/n^2)$.

In the next section we approach the problem from a Bayesian point of view. It is pointed out that Lindley's comments in the discussion following Rao (1962) are not justified. Using the results of Lindley (1961) a heuristic argument is given to show that these theorems hold quite generally and not merely in the restricted set-up considered in Section 4.2. Roughly speaking it turns out that the second order optimum properties of the maximum likelihood estimate are due to its being Bayes up to $o(1/n)$; this is a surprising fact since the expansion for the posterior upto $o(1/n)$ is not a function of the maximum likelihood estimate only. A rigorous proof along these lines is available in Ghosh, Sinha and Wieand (1980). (Similar results have been obtained by Pfanzagl (1975) and Pfanzagl and Wefelmeyer (1978)).

In the next chapter we shall extend our Theorem 4.2.6 to cover more general loss functions.

Our results and methods in this chapter are taken from Ghosh and Subramanyam (1974). They overlap substantially with those of Efron (1975), Takeuchi and Akahira (1978) and R. Ponnapalli (1976).

The readers should be warned that what we, following Rao, have called " second order efficiency" is termed "third order efficiency" by Pfanzagl, Akahira and Takeuchi. What they call "second order efficiency" is a different property. The reason for this different nomenclature will be explained in the next chapter.

4.2 Second Order Efficiency for Curved Exponential Families

Suppose $\{X_i\}$ is a sequence of i.i.d. r.v's taking values in some measurable space (S, A). Let R^k be the k-dimensional Euclidean space. For each $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ lying in some fixed open set $V_1 \subseteq R^k$, let x_i's have the probability density $f^*(x, \beta)$ with respect to some non-degenerate σ-finite measure μ on (S, A). We assume this is a curved exponential family, i.e.,

$$f^*(x, \beta) = c^*(\beta) \exp \left\{ \sum_{j=1}^{k} \beta_j p_j(x) \right\}$$

where p_j's are real valued measurable functions. Let

$$p_o(x) = 1 \;\forall\; x .$$

We assume that $p_o, p_1, p_2, \ldots, p_k$ are linearly independent in the sense

$$\sum_{j=0}^{k} c_j p_j = 0 \Rightarrow c_o = c_1 = \ldots = c_k = 0 . \quad (4.2.1)$$

Let

$$h_i(\beta) = E_\beta(p_i) = \int p_i(x) \, f^*(x, \beta) \, d\mu(x). \quad (4.2.2)$$

Then

$$\frac{\partial h_i}{\partial \beta_j} = - \frac{\partial^2 \log c^*(\beta)}{\partial \beta_i \, \partial \beta_j} = \text{cov}_\beta (p_i, p_j). \quad (4.2.3)$$

Thus

$$\left[\frac{\partial h_i}{\partial \beta_j}\right] \quad i,j = 1,2,\ldots,k$$

is the $k \times k$ dispersion matrix of $p_1, p_2, \ldots, p_k$ and it is positive definite since $p_0, p_1, \ldots, p_k$ are linearly independent. So for each $\beta^o \in V_1$ $\exists$ an open neighbourhood V of β^o such that it is contained in V_1 and restricted to V the map

$$\beta \xrightarrow{h} h(\beta)$$

is one-to-one and onto an open set W in R^k. We fix such a ' V ' and such a W and denote the inverse of this map as

$$\pi \xrightarrow{\beta} \beta(\pi)$$

from W onto V. We now introduce an alternative parameterization $\{f(x,\pi) ; \pi \in W\}$ for the family $\{f^*(x,\beta), \beta \in V\}$ where for $\pi \in W$, $f(x,\pi) = f^*(x, \beta(\pi))$. Writing $c(\pi) = c^*(\beta(\pi))$, we get

$$f(x,\pi) = c(\pi) \exp\left\{\sum_{j=1}^{k} \beta_j(\pi) p_j(x)\right\}.$$

The statistical problem that we consider is one where $\pi_1, \pi_2, \ldots, \pi_k$ are known functions $\pi_1(\theta), \pi_2(\theta), \ldots, \pi_k(\theta)$ of a single unknown real parameter θ lying in the parametric space $\bar{H}$ and θ has to be estimated on the basis of observations $x_1, x_2, \ldots, x_n$.

Assumption (I) : $(\bar{H})$ is an open set. For each $\theta \in (\bar{H})$, $\pi(\theta) \in W$ and $\pi_i(\theta)$ is thrice continuously differentiable on $(\bar{H})$, $i = 1,2,\ldots,k$. The rank of $(\pi_1'(\theta), \pi_2'(\theta) \ldots \pi_k'(\theta)) = 1$ $\forall\ \theta \in (\bar{H})$.

Let L denote the joint probability density for n independent observations $x_1, x_2, \ldots, x_n$. So, we have here

$$L = \{c(\pi(\theta))\}^n \cdot \exp \sum_{j=1}^{k} \beta_j(\pi(\theta))\, \bar{p}_{jn}\}$$

where

$$\bar{p}_{jn} = \sum_{i=1}^{n} p_j(x_i)/n, \quad \text{and} \quad \underline{\pi}(\theta) = (\pi_1(\theta), \pi_2(\theta), \ldots, \pi_k(\theta))$$

We shall henceforth write $c(\theta)$ for $c(\pi(\theta))$ and $\beta(\theta)$ for $\beta(\pi(\theta))$, so that L becomes

$$\{c(\theta)\}^n \cdot \exp\{n \sum_{j=1}^{k} \beta_j(\theta)\, \bar{p}_{jn}\} \cdot$$

Since $\bar{p}_n = (\bar{p}_{1n}, \bar{p}_{2n}, \ldots, \bar{p}_{kn})$ is jointly sufficient for θ, we consider only estimates of the form $T_n = T_n(\bar{p}_n)$ which depend on $x_1, x_2, \ldots, x_n$ only through $\bar{p}_n$. By an estimate T_n we shall actually mean a sequence of estimates $\{T_n\}$.

Note that the partial derivatives of all orders of $\beta_1, \beta_2, \ldots, \beta_k$ with respect to $\pi_1, \pi_2, \ldots, \pi_k$ exist at all points of W. Hence, by Assumption (I), $\beta_1, \beta_2, \ldots, \beta_k$ are thrice continuously differentiable functions of θ.

Since (i) rank of $(\pi'(\theta), \pi_2'(\theta), \ldots, \pi_k'(\theta))$ is one and (ii) $\beta(\pi(\theta))$ is an interior point of V, it is easy to see that the Fisher Information

$$n\, I(\theta) = E_\theta \left\{ \frac{d \log L}{d\theta} \right\}^2 = E_\theta \left\{ \sum_{i=1}^{k} \frac{\partial \log L}{\partial \pi_i} \cdot \pi_i'(\theta) \right\}^2$$

is finite and positive, where $\pi_i'(\theta) = \dfrac{d\pi_i(\theta)}{d\theta}$.

Also $I'(\theta) = \dfrac{dI(\theta)}{d\theta}$ exists.

Let us consider the following conditions:

(i) For each $\theta \in (\bar{H})$ there is an open neighbourhood V_θ of $\pi(\theta)$ with compact closure $\bar{V}_\theta$ such that $\bar{V}_\theta \subset W$ and the domain of definition of T_n includes $\bar{V}_\theta$. Moreover, there exists a function T (depending on V_θ) such that

$$T_n(X_1, X_2, \ldots, X_n) = T(\bar{p}_{1n}, \bar{p}_{2n}, \ldots, \bar{p}_{kn}) = T(p) \;\forall\; n \quad \text{if} \quad p \in V_\theta,$$

and $\dfrac{\partial T}{\partial p_i}$, $\dfrac{\partial^2 T}{\partial p_i \partial p_j}$ $i, j = 1, 2, \ldots, k$ exist and are continuous on $\bar{V}_\theta$.

(ii) $T_n(\pi(\theta)) = \theta \;\; \forall\; \theta \in (\bar{H})$.

<u>Definition 4.2.1</u>. If T_n satisfies conditions (i) and (ii), we shall say T_n is locally stable of order two which will

be abbreviated as l.s.(II). If T_n satisfies conditions (i) and (ii) and has third order continuous derivatives, we shall say T_n is locally stable of order three, which will be abbreviated as l.s. (III). If T_n satisfies conditions (i) and (ii) and has third and fourth order continuous derivatives, we shall say T_n is l.s.(IV). (We need this in the next chapter.) If T_n has only continuous first order derivatives but otherwise satisfies conditions (i) and (ii), we shall say T_n is l.s.(I) i.e. locally stable of order one.

Hereafter we will denote by $T^j(\theta)$, $T^{ij}(\theta)$, $T^{ij\ell}(\theta)$ etc. the appropriate partial derivatives of T of order one, two, three etc. respectively evaluated at $\bar{p}_n = \underline{\pi}(\theta)$.

Condition (i) is a stationarity and smoothness requirement which is likely to stabilize the large sample properties of T_n. For example convergence to an asymptotic distribution may be expected to be, in general, more rapid with condition (i) than without it. If T_n satisfies (i) then T_n is consistent iff it satisfies (ii). Note that the neighbourhood V_θ may be different for different estimates.

Consider the likelihood equation

$$F(\theta, \bar{p}_{1n}, \ldots, \bar{p}_{kn}) = \frac{d \log L}{d\theta} = 0$$

i.e.,

$$\sum_{i=1}^{k} \beta_i'(\theta)\ (\bar{p}_{in} - \pi_i(\theta)) = 0. \qquad (4.2.4)$$

If $\bar{p}_{in} = \pi_i(\theta_o)$ for all i then $F(\theta_o, \pi_1(\theta_o),\ldots,\pi_k(\theta_o)) = 0$.

This implies $\theta = \theta_o$ is a solution. Since, moreover

$$\frac{1}{n}\frac{d^2\log L}{d\theta^2}\Bigg|_{\bar{p}_{in}=\pi_i(\theta_o)} = \frac{d^2\log c(\theta)}{d\theta^2}\Bigg|_{\theta_o} + \sum_{i=1}^{k} \beta_i''(\theta)\cdot\pi_i(\theta)\Bigg|_{\theta_o} = -I(\theta_o)<0$$

it follows by the implicit function theorem that there exists a suitable neighbourhood V_{θ_o} of $\pi(\theta_o)$ and another suitable neighbourhood V'_{θ_o} of θ_o and a unique function $\hat{\theta}_n(\bar{p}_{1n},\ldots,\bar{p}_{kn})$ (depending on θ_o) defined on $\bar{V}_{\theta_o} \to V'_{\theta_o}$ (see the foot note below) such that

$$F(\hat{\theta}_n(\bar{p}_n),\ \bar{p}_{1n},\ldots,\bar{p}_{kn}) = 0 \quad \text{for all } \bar{p}_n \text{ in } \bar{V}_{\theta_o}$$

and

$$\hat{\theta}_n\ (\pi(\theta_o)) = \theta_o\ .$$

Also $\hat{\theta}_n(\bar{p}_n)$ is thrice continuously differentiable under Assumption (I) i.e., $\hat{\theta}_n$ is l.s.(III) under Assumption (I).

Usually one uses V_{θ_o} and V'_{θ_o} in place of their closures; we need this slight modification for a global construction.

We have defined $\hat{\theta}_n$ only locally but if $(\bar{H})$ has a compact closure, $\underline{\pi}(\theta)$ on $(\bar{H})$ has a continuous extension on the closure, and $\{\underline{\pi}(\theta)\ ;\ \theta \,\varepsilon\, (\bar{\bar{H}})\} \subset W$, then it is not hard to combine the local definitions to get a global definition on a suitable neighbourhood of the curve $\{\pi(\theta)\ ;\ \theta \,\varepsilon\, (\bar{\bar{H}})\}$. To see this define V_{θ_o} and V'_{θ_o} as above for all $\theta_o \,\varepsilon\, (\bar{\bar{H}})$ and note that a finite union of V's say $V_1, V_2, \ldots, V_k$, covers $\{\pi(\theta);\ \theta \,\varepsilon\, (\bar{\bar{H}})\}$. Let the solution in $\bar{V}_i$ be denoted by $\hat{\theta}_{ni}$. Consider a $\underline{\pi}(\theta)$ belonging to $\bar{V}_i$ and $\bar{V}_j$; then by the continuity of $\hat{\theta}_{ni}$ and $\hat{\theta}_{nj}$ at $\underline{\pi}(\theta)$ one can choose a neighbourhood $V''_\theta \subset V_\theta$ of $\underline{\pi}(\theta)$ such that for $\bar{p}_n \,\varepsilon\, V''_\theta$, $\hat{\theta}_{ni}$ and $\hat{\theta}_{nj}$ lie in V'_{θ_o}. Hence by the uniqueness part of the implicit function theorem, $\hat{\theta}_{ni} = \hat{\theta}_{nj}$ if $\bar{p}_n \,\varepsilon\, V''_\theta$. V''_θ can be defined in a similar fashion for $\underline{\pi}(\theta)$'s belonging to intersections of more than two V_i's. Suppose we choose V''_θ such that in addition to the previous conditions it also satisfies

(i) $V''_\theta \subset V_i$ if $\pi(\theta) \,\varepsilon\, V_i$ and

(ii) $V''_\theta \subset \bar{V}_i^c$ if $\pi(\theta) \not\varepsilon\, \bar{V}_i$

for all $\theta \,\varepsilon\, (\bar{H})$. Let $V_\pi = \bigcup V''_\theta$. Let $\bar{p}_n \,\varepsilon\, V_\pi$. Hence $\bar{p}_n \,\varepsilon\, V''_\theta$ for some θ. Suppose $\bar{p}_n \,\varepsilon\, \bar{V}_i$ and $\bar{V}_j$; this implies $\pi(\theta) \,\varepsilon\, \bar{V}_i$ as well as $\bar{V}_j$ which in turn implies that $\hat{\theta}_{ni} = \hat{\theta}_{nj}$ on V''_{θ}. Thus $\hat{\theta}_{ni}(\bar{p}_n)$ is the same for all $\bar{V}_i$ containing $\bar{p}_n$; let this common value be denoted by $f(\bar{p}_n)$. If we set

$$\hat{\theta}_n = f(\bar{p}_n) \qquad \text{if } \bar{p}_n \in V_\pi$$
$$= \text{arbitrary} \quad \text{if } \bar{p}_n \notin V_\pi ,$$

our task is accomplished.

Following Rao (1961) T_n is said to be efficient up to first order or asymptotically efficient or simply efficient if for some α and $\beta > 0$, which may depend on θ,

$$|n^{1/2} Z_n - \alpha - \beta\, n^{1/2}(T_n - \theta)| \longrightarrow 0 \tag{4.2.5}$$

in probability under θ, where $Z_n = \frac{1}{n} \frac{d \log L}{d\theta}$. Hajek (1972) has proved under quite general conditions that T_n has a certain locally asymptotically minimax property iff T_n is efficient up to first order and $\beta = I$ where I is Fisher information.

Suppose T_n is l.s.(I). Then

$$\sqrt{n}\,(T_n - \theta) = \sqrt{n}\,(T_n(\bar{p}_n) - T_n(\pi(\theta)))$$
$$= \sqrt{n} \sum_{j=1}^{k} (\bar{p}_{jn} - \pi_j(\theta))\, T^j + o_p(1) \tag{4.2.6}$$

where $o_p(1)$ is a term which tends to zero in probability.

Substituting the value of $n^{1/2}(T_n - \theta)$ in (4.2.5) we get

$$\left| n^{1/2} Z_n - \alpha - \beta \sqrt{n} \sum_{j=1}^{k} (\bar{p}_{jn} - \pi_j(\theta))\, T^j \right| \xrightarrow{P} 0$$

i.e.,

$$\left| n^{1/2} \sum_{j=1}^{k} \beta_j' \, (\bar{p}_{jn} - \pi_j(\theta)) - \alpha - \beta \sqrt{n} \sum_{j=1}^{k} (\bar{p}_{jn} - \pi_j(\theta))\, T^j \right| \xrightarrow{P}$$

i.e.,

$$\left| n^{1/2} \sum_{j=1}^{k} (\bar{p}_{jn} - \pi_j(\theta)) \, (\beta_j' - \beta \, T^j) - \alpha \right| \xrightarrow{P} 0 .$$

So it follows that for a l.s.(I) estimate T_n to be efficient up to first order it is necessary and sufficient that

$$\alpha = 0 \quad \text{and} \quad \beta_j' = \beta \, T^j \qquad (4.2.7)$$

To evaluate β we proceed as in Rao (1961). Since

$$T_n(\pi(\theta)) = T(\pi(\theta)) = \theta,$$

we get on differentiating with respect to θ,

$$\sum_{j=1}^{k} T^j \, \pi_j' = 1 \qquad (4.2.8)$$

where

$$\pi_j'(\theta) = \frac{d\pi_j(\theta)}{d\theta} .$$

From (4.2.7) and (4.2.8),

$$\beta = \sum_{j=1}^{k} \beta \, T^j \, \pi_j' = \sum_{j=1}^{k} \beta_j' \, \pi_j' . \qquad (4.2.9)$$

But

$$0 = \frac{d}{d\theta} E_\theta \left(\frac{d \log f(x,\theta)}{d\theta}\right)$$

$$= \frac{d}{d\theta}\left[\frac{d \log c(\theta)}{d\theta} + \sum_{j=1} \beta_j' \pi_j\right]$$

$$\frac{d^2 \log c(\theta)}{d\theta^2} + \sum_j \beta_j'' \pi_j + \sum_j \beta_j' \pi_j' .$$

So,

$$\sum_j \beta_j' \pi_j' = - \frac{d^2 \log c(\theta)}{d\theta^2} - \sum_j \beta_j'' \pi_j$$

$$= -E_\theta\left(\frac{d^2 \log f(k,\theta)}{d\theta^2}\right) = I(\theta). \qquad (4.2.10)$$

From (4.2.9) and (4.2.10) we get $\beta = I$. Hence it follows from (4.2.10) that a necessary and sufficient condition for a l.s.(I) estimate to be efficient up to first order is

$$T^j = \beta_j' \cdot \frac{1}{I} \quad \forall\, \theta \in (\bar{H}). \qquad (4.2.11)$$

Before defining second order efficiency let us state a simple lemma.

Let us fix $\theta \in (\bar{H})$. Let $U = \{\bar{p}_n : |\bar{p}_{i1} - \pi_i(\theta_0)| < \delta,\ i=1,2\ldots k\}$ where $\delta > 0$ is chosen so that $U \subset W$. Let I_U and I_{U^c} denote the indicator functions of U and its complement U^c. We shall also use I_U and I_{U^c} to denote $I_U(\bar{p}_n)$ and $I_{U^c}(\bar{p}_n)$ respectively.

Lemma 4.2.2.

$$P_\theta\{\bar{p}_n \varepsilon U^c\} < A\rho^n \tag{4.2.12}$$

$$E_\theta[\{|\bar{p}_{in}-\pi_i(\theta)|^r\}\{|\bar{p}_{jn}-\pi_j(\theta)|^s\} I_{U^c}] < B\rho^{n/2} \tag{4.2.13}$$

for some $0 < \rho < 1$, $A > 0$, $B > 0$, provided $r, s \geq 0$. Also ρ depends on θ and B depends on i, j, r and s in addition to θ.

Proof: Let

$$\rho_{i1} = \inf_{t \geq 0} E_\theta[\exp\{t(\bar{p}_{i1}-\pi_i(\theta)-\delta)\}]$$

$$\rho_{i2} = \inf_{t \leq 0} E_\theta[\exp\{t(\bar{p}_{i1}-\pi_i(\theta)+\delta)\}]$$

$$\rho_i = \max(\rho_{i1}, \rho_{i2})$$

$$\rho = \max_{1 \leq i \leq k} \rho_i .$$

Note that $0 < \rho < 1$. Clearly

$$P_\theta\{\bar{p}_n \varepsilon U^c\} \leq \sum_1^k P_\theta\{|\bar{p}_{in}-\pi_i(\theta)| \geq \delta\}$$

$$\leq 2\sum_1^k \rho_i^n(\theta) \quad \text{(by Chernoff's (1952, p.495) inequality)}$$

$$\leq 2k\rho^n .$$

Also,

$$E_\theta \{|\bar{p}_{in} - \pi_i(\theta)|^r \, |\bar{p}_{jn} - \pi_j(\theta)|^s \, I_{U^c}\}$$

$$\leq [E_\theta\{|\bar{p}_{in}-\pi_i(\theta)|^{4r}\}]^{1/4}[E_\theta\{|\bar{p}_{jn}-\pi_j(\theta)|^{4s}\}]^{1/4}[E_\theta\{I_{U^c}\}]^{1/2}$$

by two applications of Cauchy-Schwarz inequality. The first two terms on the right hand side are bounded in n (in fact, go to zero). So (4.2.13) now follows from (4.2.12). This completes the proof.

We shall now describe Rao's first measure of second order efficiency for a l.s.(II) estimate T_n, which is efficient up to first order. Fix $\theta_o \in \textcircled{H}$. We shall think of θ_o as the true value of the parameter. Let V_{θ_o} be the open neighbourhood of $\pi(\theta_o)$ which we may associate with T_n in accordance with the definition of local stability of order two and choose $\delta > 0$ such that

$$U = \{ \bar{p}_n : |\bar{p}_{in} - \pi_i(\theta_o)| < \delta\} \subset V_{\theta_o} \qquad i = 1,2,\dots,k .$$

For any random variable Z, let $E^U(Z)$ denote $E_{\theta_o}(Z \, I_U)$ where $I_U = I_U(\bar{p}_n)$. Let

$$Z_n = \frac{1}{n} \left. \frac{d \log L}{d\theta} \right|_{\theta = \theta_o} . \qquad (4.2.14)$$

Recall that $E_{\theta_o}(Z_n) = 0$ and $n\, E_{\theta_o}(Z_n^2) = I(\theta_o) = I.$

Here, we state two auxiliary propositions the proofs of which are given in the Appendix. These propositions are needed to define E_2 and $\bigoplus$. We use $\{T_n\}$ to indicate the sequence of estimates T_n.

Proposition 4.2.3 : Let T_n be l.s.(II) and first order efficient. Then

(i) $a_\lambda(\theta_o) = \lim_{n\to\infty} E^U n\{Z_n-(T_n-\theta_o)I - \lambda(T_n-\theta_o)^2\}$ exists.

(ii) $E_2(\{T_n\}, \theta_o, \lambda, U) = \lim_{n\to\infty} [n^2 E^U\{Z_n - (T_n-\theta_o)I$

$- \lambda(T_n-\theta_o)^2 - a_\lambda(\theta_o)/n\}^2]$ exists.

(iii) $E_2(\{T_n\}, \theta_o, \lambda, U_1) = E_2(\{T_n\}, \theta_o, \lambda, U_2)$

where U_1 and U_2 are any two neighbourhoods of $\underline{\pi}(\theta_o)$ contained in V_{θ_o}.

In view of (iii) we shall write

$$E_2(\{T_n\}, \theta_o, \lambda) \text{ for } E_2(\{T_n\}, \theta_o, \lambda, U).$$

Let

$$E_2(\{T_n\}, \theta_o) = \inf_\lambda E_2(\{T_n\}, \theta_o, \lambda).$$

This E_2 is Rao's first measure of second order efficiency for a l.s.(II) estimate T_n, which is efficient up to first order.

We can think of E_2 as a measure of how well a quadratic in T_n approximates Z_n. If (for each fixed θ_o) Z_n were a function of T_n, T_n would be a sufficient statistic for θ_o. So, E_2 measures, in a sense, how "nearly" sufficient T_n is. The reason for taking a quadratic in T_n is mainly one of expediency. In Rao (1961) E_{θ_o} is used instead of E^U but the calculations can be justified only with E^U. See in this connection Rao (1963) where essentially the present approach is followed. The intuitive justification for using E^U is that we do not wish our measure to be unduly affected by the tail of the distribution of the estimate. If we are comparing two l.s. (II) efficient estimates $T_n^{(1)}$ and $T_n^{(2)}$, we may take $U \subset V^1_{\theta_o} \cap V^2_{\theta_o}$ for the calculation of E_2 for both the estimates, to remove the apparent arbitrariness of U and hence of the method of comparing $T_n^{(1)}$ and $T_n^{(2)}$.

Proposition 4.2.4 : Let T_n be l.s.(II) and first order efficient. Then

(i) $b(\theta_o) = \lim_{n\to\infty} n\{E^U(T_n) - \theta_o\}$ exists.

If moreover T_n is l.s.(III) then the following results hold.

(ii) $b(\theta)$ is a continuously differentiable function on $\bar{H}$.

(iii) If $T_n^* = T_n - b(T_n)/n$ then

$$E^U(T_n^*) = \theta_o + o(1/n).$$

(iv) If $T_n' = T_n + m(T_n)/n$ where m is a continuously differentiable function in a neighbourhood of θ_o, then

$$E^U(T_n' - \theta_o)^2 = \frac{1}{nI} + \frac{\Psi(\{T_n'\}, \theta_o)}{n^2} + o(\frac{1}{n^2})$$

where $\Psi(\{T_n'\}, \theta_o)$ does not depend on U.

(v) Let $W(a,\theta) = \min\{(a - \theta)^2, d\}$ be the squared error loss truncated at $d > 0$ and let T_n' be as defined in (iv). Then

$$E_{\theta_o}\{W(T_n', \theta_o)\} - E^U(T_n' - \theta_o)^2 = o(\frac{1}{n^2}).$$

Rao (1963) takes $E^U(T_n' - \theta_o)^2$ as the risk function of T_n' but Proposition 4.2.4 shows that it does not matter up to $o(\frac{1}{n^2})$ whether we take $E^U(T_n' - \theta_o)^2$ or $E_{\theta_o}\{W(T_n', \theta_o)\}$ as our risk function. Following Rao, we take $\Psi(\{T_n'\}, \theta_o)$ as our second measure of second order efficiency of $\{T_n'\}$.

Before stating our main result, we shall now introduce a few more notations. Let

$$W_n = \frac{1}{n} \left.\frac{d^2 \log L}{d\theta^2}\right|_{\theta_o} + I(\theta_o) \tag{4.2.15}$$

$$= \frac{1}{n} \left.\frac{d^2 \log c(\theta)}{d\theta^2}\right|_{\theta_o} + \Sigma\, \beta_i''(\theta_o)\, \bar{p}_{in} + I(\theta_o). \tag{4.2.16}$$

Clearly $E_{\theta_o}(W_n) = 0$. Let

$$\mu_{rs} = E_{\theta_o}(Z_1^r W_1^s) \tag{4.2.17}$$

where Z_n is defined earlier by (4.2.14). Note that $I(\theta_o) = \mu_{20}$. As stated before we shall often write I for $I(\theta_o)$. Let

$$J = E_{\theta_o}\left\{ \frac{d^3 \log f(X,\theta)}{d\theta^3} \Big|_{\theta_o} \right\}$$

$$= \frac{d^3 \log c(\theta)}{d\theta^3} \Big|_{\theta_o} + \Sigma \beta_i'''(\theta_o)\, \pi_i(\theta_o) \tag{4.2.18}$$

where $\beta_i'''(\theta_o) = \frac{d^3 \beta_i(\theta)}{d\theta^3} \Big|_{\theta_o}$. Let S_n denote the random variable

$$(Z_n W_n - \mu_{11}/n) \cdot \frac{1}{I^2} + (Z_n^2 - I/n) \cdot J/2I^3 \tag{4.2.19}$$

which will play an important role in what follows.

<u>Definition 4.2.5</u> : If Y_n is a sequence of random variables such that $E^U\{Y_n^2\} = o(a_n^2)$ or $O(a_n^2)$, we shall write Y_n is $o_E(a_n)$ or $O_E(a_n)$ accordingly.

A random variable X will be called E^U- orthogonal to another random variable Y if the covariance under θ_o, of

$X\, I_U$ and Y is zero. i.e., $\mathrm{Cov}_{\theta_o}\{X\, I_U, Y\} = 0$; X and Y are said to be E^U-orthogonal up to $o(1/n^2)$ if the covariance of $X\, I_U$ and Y is $o(1/n^2)$ i.e.,

$$\mathrm{cov}_{\theta_o}\{X\, I_U, Y\} = o(1/n^2).$$

Also note that if T_n is efficient and l.s.(II), then $E_{\theta_o}(T_n) = \theta_o + b(\theta_o)/n + o(1/n)$ using Proposition 4.2.4.

Under Assumption (I), $\hat{\theta}_n$ is efficient and l.s.(III) and hence we may write

$$E_{\theta_o}(\hat{\theta}_n) = \theta_o + b_o(\theta_o)/n + o(1/n) \tag{4.2.20}$$

where $b_o(\theta_o) = \lim\ n\{E^U(\hat{\theta}_n) - \theta_o\}$.

We can now state our main result.

<u>Theorem 4.2.6</u> : Under Assumption (I) we have the following :

$$\text{(i)}\quad \hat{\theta}_n - \theta_o - Z_n/I = b_o(\theta_o)/n + S_n + \hat{R}_n \tag{4.2.21}$$

where $\hat{R}_n$ is $o_E(1/n)$ and E^U-orthogonal to Z_n^2 and $Z_n W_n$ up to $o(n^{-2})$ and S_n, $b_o(\theta_o)$ are as defined above.

If $\hat{\theta}^*_n = \hat{\theta}_n - b(\hat{\theta}_n)/n$ then

$$\hat{\theta}^*_n - \theta_o - Z_n/I = \frac{-Z_n}{2nI^4}\{2\mu_{21} I + J\mu_{30}\} + S_n + \hat{R}^*_n \tag{4.2.22}$$

where $\hat{R}_n^*$ is $o_E(n^{-1})$ and E^U-orthogonal to Z_n, Z_n^2 and Z_nW_n up to $o(n^{-2})$.

(ii) Let T_n be efficient and l.s.(II). Then

$$T_n - \theta_o - Z_n/I = b(\theta_o)/n + S_n + R_n \tag{4.2.23}$$

where R_n is $O_E(n^{-1})$ and E^U-orthogonal to Z_n^2 and Z_nW_n up to $o(n^{-2})$ and $E^U(R_n) = o(n^{-1})$. Also

$$E_2(\{T_n\}, \theta_o) \geq E_2(\{\hat{\theta}_n\}, \theta_o) \;\forall\; \theta_o \in \bar{H}. \tag{4.2.24}$$

Let T_n be efficient and l.s.(III). Then

$$T_n^* - \theta_o = (\hat{\theta}_n^* - \theta_o) + R_n^* \tag{4.2.25}$$

where R_n^* is $O_E(n^{-1})$, E^U-orthogonal to Z_n, Z_n^2 and Z_nW_n up to $o(n^{-2})$ and hence to $(\hat{\theta}_n^* - \theta_o)$ up to $o(n^{-2})$; T_n^* and $\hat{\theta}_n^*$ are as defined in Proposition 4.2.4 and (4.2.22) respectively. Also

$$\psi(\{T_n^*\}, \theta_o) \geq \psi(\{\hat{\theta}_n^*\}, \theta_o) \;\forall\; \theta_o \in \bar{H}. \tag{4.2.26}$$

Moreover

$$E_2(\{T_n\}, \theta_o) = I^2\psi(\{T_n^*\}, \theta_o) - \frac{2}{I^2}\{J/2 + \mu_{11}\}^2. \tag{4.2.27}$$

(iii) Let T_n be efficient and l.s.(III) and $m(\theta)$ be a continuously differentiable function on $(\bar{H})$. Let $T'_n = T_n + m(T_n)/n$. Then there exists a continuously differentiable function g on $(\bar{H})$ such that if

$$\hat{\theta}'_n = \hat{\theta}_n + g(\hat{\theta}_n)/n$$

then $\hat{\theta}_n$ is better than T'_n up to $o(n^{-2})$ in the sense that

$$\lim_{n\to\infty} n^2[E_{\theta_o}\{W(T'_n, \theta_o)\} - E_{\theta_o}\{W(\hat{\theta}'_n, \theta_o)\}] \geq 0$$

where $W(a,\theta) = \min\{(a-\theta)^2, d\}$ is the squared error loss truncated at $d > 0$.

Here o_E, O_E and E^U-orthogonal are as given in Definition 4.2.5.

We shall need a few lemmas to prove this theorem.

Lemma 4.2.7. If Assumption (I) holds and T_n is efficient and l.s.(II) then

$$T_n(\bar{p}_n) - \theta_o = \frac{Z_n}{I} + \frac{1}{2}\sum_i\sum_j (\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o))T^{ij} + o_E(\tfrac{1}{n}). \tag{4.2.28}$$

If, moreover, T_n is l.s.(III) then

$$T_n(\bar{p}_n) - \theta_o = \frac{Z_n}{I} + \frac{1}{2}\sum_i\sum_j (\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o))\, T^{ij}$$

$$+ \frac{1}{6}\sum_i\sum_j\sum_\ell (\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o))(\bar{p}_{\ell n} - \pi_\ell(\theta_o)]\,T^{ij\ell} + o_E(n^{-\frac{3}{2}})$$

$$= T_{n1} + T_{n2} + T_{n3} + o_E(n^{-3/2}) \text{ (say)} \tag{4.2.29}$$

<u>Proof</u> : For $\bar{p}_n \in U$, consider the Taylor expansion

$$T_n(\bar{p}_n) - \theta_o = T_n(\bar{p}_n) - T(\underline{\pi}(\theta_o))$$

$$= \sum_j (\bar{p}_{jn} - \pi_j(\theta_o))T^j + \sum_i \sum_j (\bar{p}_{in} - \pi_i(\theta_o)(\bar{p}_{jn} - \pi_j(\theta_o))\frac{T^{ij}}{2} + R(\bar{p}_n) \tag{4.2.30}$$

where T^j's and T^{ij}'s are the first and second order derivatives of T evaluated at $\pi(\theta_o)$ and R is the remainder term.

Note that from (4.2.11) $T^j = \beta_j^! \cdot \frac{1}{I}$ $\forall\, \theta \in (\underline{H})$. Hence (4.2.30) becomes

$$T_n(\bar{p}_n) - \theta_o = \frac{1}{I} \sum_j (\bar{p}_{jn} - \pi_j(\theta_o))\beta_j^! + \sum_i \sum_j (\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o))T^{ij}/2$$
$$+ R(\bar{p}_n).$$

Also,

$$Z_n = \frac{1}{n} \left.\frac{d \log L}{d\theta}\right|_{\theta_o} = \sum_j (\bar{p}_{jn} - \pi_j(\theta_o))\beta_j^!, \text{(from (4.2.14))}.$$

To prove (4.2.28) it remains to be shown that $R(\bar{p}_n)$ is $o_E(1/n)$ i.e., $E^U\{R^2(\bar{p}_n)\} = o(1/n^2)$.

Note that

$$R(\bar{p}_n) = \varepsilon(\bar{p}_n) \sum \sum (\bar{p}_{in} - \pi_i(\theta_o))\ (\bar{p}_{jn} - \pi_j(\theta_o))T^{ij} \tag{4.2.31}$$

where $\varepsilon(p) \to 0$ as $p \to \pi(\theta_o)$ and $|\varepsilon(p)| < M$ on U for some suitable M. Fix $\eta > 0$ and choose $\delta_1 > 0$ such that

$$U_1 = \{ \bar{p}_n : |\bar{p}_{in} - \pi_i(\theta_o)| < \delta_1, \ i = 1,2,\ldots,k\} \subset U$$

and $|\varepsilon(\bar{p}_n)| < \eta$ if $p_n \in U_1$. Therefore

$$E^U(R^2(p_n)) = E^U\{R^2 I_{U_1}\} + E^U\{R^2 I_{U_1^c}\}$$

$$= E_{\theta_o}\{R^2 I_{U_1}\} + E_{\theta_o}\{R^2 I_{U_1^c}\} + o(n^{-2})$$

$$\leq \eta^2 \sum\sum E_{\theta_o}\{|\bar{p}_{in} - \pi_i(\theta_o)|^4\}^{1/2} E_{\theta_o}\{|\bar{p}_{jn}-\pi_j(\theta_o)|^4\}^{1/2} |T^{ij}|^2$$

$$+ M^2 \sum\sum E_{\theta_o}\{|\bar{p}_{in} - \pi_i(\theta_o)|^2 |\bar{p}_{jn}-\pi_j(\theta_o)|^2 I_{U^c}\} |T^{ij}|^2 + o(n^{-2}) \qquad (4.2.32)$$

by (4.2.31). The second term in (4.2.32) is $M^2.0(\rho^{n/2})$ by Lemma 4.2.2 which is obviously valid with U_1 in place of U. The first term in (4.2.32) is $\eta\, 0(1/n^2)$. Since η is arbitrary, it follows that $E^U(R^2) = o(1/n^2)$.

Similarly, one can prove (4.2.29).

This completes the proof of Lemma 4.2.7.

In following pages when we assert that a random variable is $o_E(n^{-1})$, $0_E(n^{-1})$ etc., we shall not usually give a proof. But in each case justification is easy and involves an application of Lemma 4.2. 2 aided perhaps by the Cauchy-Schwarz inequality.

Suppose T_n is efficient and l.s.(II). Then

$$b(\theta_o) = \frac{1}{2} \Sigma \Sigma T^{ij} E_{\theta_o}(p_{i1} - \pi_i(\theta_o))(p_{j1} - \pi_j(\theta_o)) + o(n^{-1}), \tag{4.2.33}$$

follows immediately from Lemma 4.2.7. If T_n is also l.s.(III) then by Proposition 4.2.4, $b(\theta)$ is continuously differentiable. By considering the Taylor expansion around θ_o it follows that

$$\frac{b(T_n)}{n} = \frac{b(\theta_o)}{n} + (T_n - \theta_o)\frac{b'(\theta_o)}{n} + o_E(n^{-3/2})$$

$$= \frac{b(\theta_o)}{n} + \frac{Z_n}{I} \cdot \frac{b'(\theta_o)}{n} + o_E(n^{-3/2}) \tag{4.2.34}$$

applying (4.2.28) to $T_n - \theta_o$. When T_n is efficient and l.s.(III) we define

$$T_n^* = T_n - b(T_n)/n . \tag{4.2.35}$$

Then by (iii) of Proposition 4.2.4,

$$E^U(T_n^*) = \theta_o + o(1/n). \tag{4.2.36}$$

We shall now calculate the covariance of $(T_n^* - \theta_o) I_U$ with Z_n, $Z_n W_n$ and Z^2 show that these covariances are the same up to $o(1/n^2)$ in the sense that they don't depend on T_n up to $o(1/n^2)$.

<u>Lemma 4.2.8</u> : If Assumption (I) holds and T_n is efficient and l.s.(III) then

$$E^U \{(T_n^* (\bar{p}_n) - \theta_o) Z_n\} = \frac{1}{n} + o(1/n^2). \tag{4.2.37}$$

<u>Proof</u> : Since T_n is l.s.(III) we have from (4.2.34),

$$\frac{b(T_n)}{n} = \frac{b(\theta_o)}{n} + (T_n - \theta_o) \frac{b'(\theta_o)}{n} + o_E(n^{-3/2})$$

where

$$\frac{b(\theta_o)}{n} = \frac{1}{2} \sum_i \sum_j T^{ij} E_{\theta_o}\{(\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o))\}$$

$$= \frac{1}{2} \Sigma \Sigma T^{ij} . \frac{a_{ij}(\theta_o)}{n} \quad \text{(say)}$$

where

$$\frac{a_{ij}(\theta_o)}{n} = \Sigma \Sigma T^{ij} E_{\theta_o}\{(\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o))\} . \tag{4.2.38}$$

Thus

$$b'(\theta_o) = \frac{1}{2n} \sum_i \sum_j T^{ij} a'_{ij}(\theta_o) + \frac{1}{2n} \sum_i \sum_j a_{ij}(\theta_o) \sum_{\ell} T^{ij\ell} \pi'_{\ell}(\theta_o). \tag{4.2.39}$$

Also,

$$T_n^* - \theta_o = (T_n - \theta_o)\left[1 - \frac{b'(\theta_o)}{n}\right] - \frac{b(\theta_o)}{n} + o_E(n^{-3/2}),$$

so that

$$E^U_{\theta_o}\{(T_n^* - \theta_o) Z_n\} = \left[1 - \frac{b'(\theta_o)}{n}\right] E_{\theta_o}\{(T_n - \theta_o) Z_n\} + o(n^{-2})$$

(using Lemma 4.2.2)

$$= [1-\frac{b'(\theta_o)}{n}][E_{\theta_o}(Z_n^2/I) + \frac{1}{2}\sum_i\sum_j E_{\theta_o}\{(\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn}-\pi_j(\theta_o))Z_n\}$$

$$+ \frac{1}{6}\sum_i\sum_j\sum_\lambda E_{\theta_o}\{(\bar{p}_{in}-\pi_i(\theta_o))(\bar{p}_{jn}-\pi_j(\theta_o))(\bar{p}_{\lambda n}-\pi_\lambda(\theta_o))Z_n\} + o(n^{-2}). \tag{4.2.40}$$

Differentiating (4.2.38) with respect to θ_o we have

$$\frac{a'_{ij}(\theta_o)}{n} = n\, E_{\theta_o}\{(\bar{p}_{in}-\pi_i(\theta_o))\ (\bar{p}_{jn}-\pi_j(\theta_o))\ Z_n\}\ . \tag{4.2.41}$$

Denoting $E_{\theta_o}\{(\bar{p}_{in}-\pi_i(\theta_o))\ (\bar{p}_{jn}-\pi_j(\theta_o))\ (\bar{p}_{\lambda n}-\pi_\lambda(\theta_o))\}$ as $\frac{a_{ij\lambda}(\theta_o)}{n}$ and differentiating it with respect to θ_o and rearranging the terms gives us

$$E_{\theta_o}\{(\bar{p}_{in} - \pi_i(\theta_o))\ (\bar{p}_{jn} - \pi_j(\theta_o))(\bar{p}_{\lambda n} - \pi_\lambda(\theta_o))\ Z_n\}$$

$$= \frac{a'_{ij\lambda}(\theta_o)}{n^3} + \frac{1}{n^2}[\pi'_i(\theta_o)\ a_{j\lambda}(\theta_o) + \pi'_j(\theta_o)\ a_{i\lambda}(\theta_o) + \pi'_\lambda(\theta_o)a_{ij}(\theta_o)]. \tag{4.2.42}$$

Now Lemma 4.2.8 follows from (4.2.40), (4.2.41) and (4.2.42).

Note that we can get (4.2.37) formally by differentiating (4.2.36) with respect to θ. The trouble in justifying this is that one has to show that on differentiating the $o(1/n)$ term in (4.2.36) one would get a term of order $o(1/n)$. The calculations needed for this are no less cumbrous than the direct proof of Lemma 4.2.8 given above.

To calculate the covariance of $(T_n^* - \theta_o) I_U$ with Z_n^2 and $Z_n W_n$, we shall need the following results which are well known and easy to derive. The same formulas were given in Rao (1961) but we shall use them in a different way. Let $(Y_{1i}, Y_{2i}, Y_{3i}, Y_{4i})$ be i.i.d random vectors with zero expectations. Let $\bar{Y}_j = \frac{1}{n} \sum_{i=1}^{n} Y_{ji}$. Then up to $o(n^{-2})$

$$E(\bar{Y}_1^4) = 3[\mathrm{var}(Y_{11})]^2/n^2, \tag{4.2.43}$$

$$\mathrm{cov}\,(\bar{Y}_1^2, \bar{Y}_2 \bar{Y}_3) = 2\,\mathrm{cov}(Y_{11}, Y_{21}) \cdot \mathrm{cov}(Y_{11}, Y_{31}) / n^2, \tag{4.2.44}$$

$$\mathrm{cov}\,(\bar{Y}_1 \bar{Y}_2, \bar{Y}_3 \bar{Y}_4) = \{\mathrm{cov}(Y_{11}, Y_{31}) \cdot \mathrm{cov}(Y_{21}, Y_{41}) + \mathrm{cov}(Y_{11}, Y_{41}) \cdot \mathrm{cov}(Y_{21}, Y_{31})\}/n^2, \tag{4.2.45}$$

$$E(\bar{Y}_1^3\, \bar{Y}_2) = 3\,\mathrm{var}(Y_{11}) \cdot \mathrm{cov}(Y_{11}, Y_{21})/n^2, \tag{4.2.46}$$

$$E(\bar{Y}_1^2\, \bar{Y}_2^2) = \{\mathrm{var}(Y_{11}) \cdot \mathrm{var}(Y_{21}) + 2(\mathrm{cov}(Y_{11}, Y_{21}))^2\} / n^2. \tag{4.2.47}$$

<u>Lemma 4.2.9</u>: Suppose Assumption (I) holds and T_n is efficient and l.s.(II). Let

$$T_n^{**} = T_n(\bar{p}_n) - \theta_o - Z_n/I . \qquad (4.2.48)$$

Then

$$E^U \{ T_n^{**}(Z_n^2 - I/n) \} = \frac{2\mu_{II} + J}{n^2 I} + o\,(1/n^2) \qquad (4.2.49)$$

and

$$E^U \{ T_n^{**}(Z_n W_n - \mu_{11}/n) \} = \frac{\mu_{02}}{n^2 I} + \frac{\mu_{11}(J + \mu_{11})}{n^2 I^2} + o\,(1/n^2). \qquad (4.2.50)$$

Proof : $E^U \{ T_n^{**}\, (Z_n^2 - I/n) \}$

$= E^U \{ (T_n(\bar{p}_n) - \theta_o - Z_n/I)\ (Z_n^2 - I/n) \}$

$= E^U \{ \frac{1}{2} \Sigma\, \Sigma\, T^{ij}\ (\bar{p}_{in} - \pi_i(\theta_o))\ (\bar{p}_{jn} - \pi_j(\theta_o))(Z_n^2 - I/n) \} + o\,(n^{-2})$

(by (4.2.28))

$= E_{\theta_o} \{ \frac{1}{2} \Sigma\, \Sigma\, T^{ij} (\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o)) . (Z_n^2 - I/n) \} + o\,(n^{-2})$

(by Lemma 4.2.2)

$= \frac{1}{2} \Sigma\, \Sigma\, T^{ij}\ \mathrm{cov}_{\theta_o} \{ Z_n^2,\ (\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o)) \} + o\,(n^{-2})$

$= \Sigma\, \Sigma\, T^{ij}\ \mathrm{cov}_{\theta_o} \{ Z_n, \bar{p}_{in} \} ;\ \mathrm{cov}_{\theta_o} \{ Z_n, \bar{p}_{jn} \} + o\,(n^{-2})$ by (4.2.42)

$= \frac{1}{n} \Sigma\, \Sigma\, T^{ij}\ \mathrm{cov}_{\theta_o} \{ Z_n, \bar{p}_{in} \} . \pi_j'(\theta_o) + o\,(n^{-2})$

(since $\mathrm{cov}_{\theta_o}(Z_n, \bar{p}_{jn}) = E_{\theta_o}(\bar{p}_{jn} Z_n) = \pi_j'(\theta_o)/n$)

$= \frac{1}{n} \sum_i \{ \mathrm{cov}_{\theta_o}(Z_n, \bar{p}_{in}) . \sum_j T^{ij}\ \pi_j'(\theta_o) \} + o\,(n^{-2})$

$$= \frac{1}{n} \sum_i \mathrm{cov}_{\theta_o}(Z_n, \bar{p}_{in})\ d/d\theta\ (\beta_i'/I) \Big|_{\theta_o} + o\,(n^{-2}) \qquad (4.2.51)$$

Differentiating both sides of (4.2.11) with respect to θ and putting $\theta = \theta_o$ we get

$$\sum_j T^{ij} \pi_j'(\theta_o) = \frac{d}{d\theta}(\beta_i'/I)\Big|_{\theta_o} = \frac{\beta_i''(\theta_o)}{I} - \frac{\beta_i'(\theta_o)\, I'(\theta_o)}{I^2} . \qquad (4.2.52)$$

As noted earlier $I'(\theta_o)$ and $\beta_i'(\theta_o)$ exists by Assumption (I). Hence the right hand side of (4.2.51) can be written as

$$= \frac{1}{n}\Big(\sum_i \text{cov}_{\theta_o}(Z_n, \bar{p}_{in}) \cdot \Big\{\frac{\beta_i''(\theta_o)}{I} - \frac{\beta_i'(\theta_o)\, I'(\theta_o)}{I^2}\Big\} + o\,(n^{-2})$$

$$= \frac{1}{n}\Big\{\text{cov}_{\theta_o}\Big(Z_n, \sum \bar{p}_{in}\frac{\beta_i''(\theta_o)}{I}\Big) - \text{cov}_{\theta_o}\Big(Z_n, \frac{\sum \bar{p}_{in}\ \beta_i'(\theta_o) I'(\theta_o)}{I^2}\Big)\Big\} + o\,(n^{-2})$$

$$= \frac{1}{nI}\,\text{cov}(Z_n, W_n) - \frac{I'}{n\,I^2}\,\text{var}(Z_n) + o\,(n^{-2}) \quad \text{(using (4.2.14) and (4.2.15))}$$

$$= \frac{\mu_{11}}{nI} - \frac{I'}{n^2 I} + o\,(n^{-2}), \quad \text{since } \text{cov}(Z_n, W_n) = \mu_{11} \text{ and } V(Z_n) = I/n ,$$

$$= \frac{2\mu_{11} + J}{n^2 I} + o\,(n^{-2}), \qquad (4.2.53)$$

substituting $-I' = J + \mu_{11}$. This is so because

$$-I = \frac{d^2 \log c(\theta)}{d\theta^2} + \sum_{i=1}^{k} \beta_i''(\theta)\ \pi_i(\theta)$$

so that

$$-I'(\theta_o) = \frac{d^3 \log c(\theta_o)}{d\theta_o^3} + \sum \beta_i'''(\theta_o)\ \pi_i(\theta_o) + \sum \beta_i''(\theta_o)\ \pi_i'(\theta_o)$$

$$= J + \sum \beta_i''(\theta_o)\ \pi_i'(\theta_o) \quad \text{by (4.2.18)}$$

$$= J + \mu_{11}. \qquad (4.2.54)$$

This completes the proof of (4.2.49).

To prove (4.2.50), proceeding the same way as before but using (4.2.45) in place of (4.2.44), we get

$$E^U\{T_n^{**}(Z_nW_n - \mu_{11}/n)\}$$

$$= \frac{1}{2}\Sigma\,\Sigma\, T^{ij}\mathrm{cov}_{\theta_o}\{Z_nW_n, (\bar{p}_{in}-\pi_i(\theta))(\bar{p}_{jn}-\pi_j(\theta_o))\} + o\,(n^{-2})$$

$$= \frac{1}{2}\Sigma\,\Sigma\, T^{ij}\{\mathrm{cov}_{\theta}\,(Z_n,\bar{p}_{in})\cdot\ \mathrm{cov}_{\theta_o}(W_n,\bar{p}_{jn}) +$$

$$\mathrm{cov}_{\theta_o}(Z_n,\bar{p}_{jn})\cdot\ \mathrm{cov}_{\theta_o}(W_n,\bar{p}_{in})\} + o\,(n^{-2}) \text{ using } (4.2.45)$$

$$= \Sigma\,\Sigma\, T^{ij}\{\mathrm{cov}_{\theta_o}(Z_n,\bar{p}_{jn})\cdot\ \mathrm{cov}_{\theta_o}(W_n,\bar{p}_{in})\} + o\,(n^{-2})$$

$$= \frac{1}{n}\sum_i\sum_j T^{ij}\ \pi_j'(\theta_o)\cdot\mathrm{cov}_{\theta}(W_n,\bar{p}_{in})+o\,(n^{-2})$$

$$= \frac{1}{n}\sum_i \mathrm{cov}_{\theta_o}(W_n,\bar{p}_{in})\cdot\left\{\frac{\beta_i''(\theta_o)}{I} - \frac{\beta_i'(\theta_o)I'(\theta_o)}{I^2}\right\} + o_E(n^{-2}) \text{ using } (4.2.52)$$

$$= \frac{\mathrm{var}(W_n)}{nI} - \frac{I'\ \mathrm{cov}_{\theta_o}(W_n,Z_n)}{nI^2} + o\,(n^{-2})$$

$$= \frac{\mu_{02}}{nI} + \frac{\mu_{11}(J+\mu_{11})}{n^2I^2} + o\,(n^{-2})$$

(since $\mathrm{var}(W_n) = \frac{\mu_{02}}{n}$, $\mathrm{cov}_{\theta_o}(W_n,Z_n) = \frac{\mu_{11}}{n}$ and $-I' = J + \mu_{11}$).

This proves (4.2.50) and completes the proof of the lemma.

<u>Lemma 4.2.10</u> : Suppose Assumption (I) holds and T_n is efficient and l.s.(III). Let

$$T_n^{***} = T_n^* - \theta_o - Z_n/I . \qquad (4.2.55)$$

Then $E^U \{ T_n^{***}(Z_n^2 - I/n)$ and $E^U \{ T_n^{***}(Z_n W_n - \mu_{11}/n)\}$ are given by the right hand sides of (4.2.49) and (4.2.50) respectively.

<u>Proof</u> : Lemma 4.2.10 follows from Lemma 4.2.9 if we note that

(i) $T_n^{**} - T_n^{***} = \dfrac{b(\theta_o)}{n} + o_E(n^{-1})$,

(ii) $Z_n^2 - I/n = O_E(n^{-1})$,

(iii) $E^U_{\theta_o}(Z_n^2 - I/n) = o\,(n^{-2})$,

(iv) $Z_n W_n - \mu_{11}/n = O_E(n^{-1})$

and

(v) $E^U_{\theta_o}(Z_n W_n - \mu_{11}/n) = o\,(n^{-2})$.

Using the definition of T_n^{**}, T_n^{***}, (4.2.34) and (4.2.35) we get

$$T_n^{**} - T_n^{***} = \frac{b(\theta_o)}{n} + \frac{Z_n}{I}\,\frac{b'(\theta_o)}{n} + o_E(n^{-3/2})$$

$$= \frac{b(\theta_o)}{n} + o_E(n^{-1})$$

since

$$E^U \left\{ \frac{Z_n}{I}\,\frac{b'(\theta)}{n} \right\}^2 = o(n^{-2}).$$

For,

$$E^U\left\{\left(\frac{Z_n}{I}\right)\frac{b'(\theta_o)}{n}\right\}^2 = \frac{b'(\theta_o)^2}{n^2\ I^2} E_{\theta_o}(Z_n^2) + o\ (n^{-2})\ \text{(by Lemma 4.2.2)}$$

$$= \frac{b'(\theta_o)^2}{n^3 I} + o\ (n^{-2}) = o\ (n^{-2}).$$

This proves (i). To prove (ii), by using Lemma 4.2.2 we conclude

$$E^U(Z_n^2 - I/n)^2 = E_{\theta_o}\ (Z_n^2 - I/n)^2 + o\ (n^{-2})$$

$$= E_{\theta_o}(Z_n^4) - 2\ I/n\ .\ I/n + I^2/n^2 + o\ (n^{-2})$$

$$= \frac{3I}{n^2} - \frac{2I^2}{n^2} + \frac{I^2}{n^2} + o_E(n^{-2})\quad \text{using (4.2.43)}$$

$$= \frac{2I^2}{n^2} + o\ (n^{-2}) = 0\ (n^{-2})$$

so,

$$Z_n^2 - I/n\quad \text{is}\quad 0_E(n^{-1}).$$

Thus using Lemma 4.2.2 it follows that

$$E^U\ (Z_n^2 - I/n) = E_{\theta_o}(Z_n^2 - I/n) + o\ (n^{-2})$$

$$= o\ (n^{-2})$$

This proves (iii). To prove (iv), let us note that

$$E^U(Z_nW_n - \mu_{11}/n)^2 = E_{\theta_o}(Z_nW_n - \mu_{11}/n)^2 + o\ (n^{-2}) \quad \text{(by Lemma 4.2.2)}$$

$$= E_{\theta_o}(Z_n^2W_n^2) - \frac{\mu_{11}^2}{n^2} + o\ (n^{-2})$$

$$= \frac{I.\mu_{02}}{n^2} + 2\frac{\mu_{11}^2}{n^2} - \frac{\mu_{11}^2}{n^2} + o\ (n^{-2}) \text{ using (4.2.47)}$$

$$= 0\ (n^{-2}).$$

Now Lemma 4.2.2 implies that

$$E^U(Z_nW_n - \mu_{11}/n) = E_{\theta_o}(Z_nW_n - \mu_{11}/n) + o\ (n^{-2})$$

$$= o\ (n^{-2}).$$

This proves (v).

Lemmas 4.2.9 and 4.2.10 are special cases of a more general result which expresses the covariance of $T_n^{**} I_U$ and $T_n^{***} I_U$ with $Z_n \{\Sigma\, \alpha_i\ (\bar{p}_{in} - \pi_i(\theta_o))\}$ as the covariance of $\Sigma\, \alpha_i(\bar{p}_{in} - \pi_i(\theta_o))$ with $\frac{1}{nI^2}\{W_nI - Z_nI'\}$, where α_i's are constants. That is

$$\text{cov}_{\theta_o}\{T_n^{**} I_U,\ Z_n \cdot \Sigma\, \alpha_i\ (\bar{p}_{in} - \pi_i(\theta_o))\}$$

$$= \text{cov}_{\theta_o}\{T_n^{***} I_U,\ Z_n \cdot \Sigma\, \alpha_i\ (\bar{p}_{in} - \pi_i{'}\theta_o))\}$$

$$= \text{cov}_{\theta_o}\{\Sigma\, \alpha_i(\bar{p}_{in} - \pi_i(\theta_o)),\ \frac{1}{nI^2}\ (W_nI - Z_nI')\} + o\ (n^{-2}).$$

The proof of this more general result is similar to the proofs of Lemma 4.2.9 and 4.2.10. For completeness, we shall outline the proof below. Another "formal" proof is given in the third remark of Section 4.3. Using Lemma 4.2.2, we can write

$$\mathrm{cov}_{\theta_o}\{ T_n^{**} I_U, Z_n \Sigma \alpha_i(\bar{p}_{in} - \pi_i(\theta_o))\}$$

$$= \mathrm{cov}_{\theta_o} \{ T_n^{**}, Z_n \Sigma \alpha_i (\bar{p}_{in} - \pi_i(\theta_o))\} + o(n^{-2})$$

$$= \mathrm{cov}_{\theta_o}\{ \tfrac{1}{2} \Sigma \Sigma T^{ij}(\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o)),$$

$$Z_n \Sigma \alpha_i (\bar{p}_{in} - \pi_i(\theta_o))\} + o(n^{-2}) \quad \text{(using (4.2.28))}$$

$$= \tfrac{1}{2} \Sigma \Sigma T^{ij} \mathrm{cov}_{\theta_o}\{(\bar{p}_{in} - \pi_i(\theta_o))(\bar{p}_{jn} - \pi_j(\theta_o)), Z_n Q_n\}.$$

$$(\text{where } Q_n = \Sigma \alpha_i (\bar{p}_{in} - \pi_i(\theta_o)))$$

$$= \Sigma \Sigma T^{ij} [\mathrm{cov}_{\theta_o}(Q_n, \bar{p}_{in}) \cdot \mathrm{cov}_{\theta_o}(Z_n, \bar{p}_{jn})]$$

$$= \frac{1}{n} \Sigma \Sigma T^{ij} \mathrm{cov}_{\theta_o}(Q_n, \bar{p}_{in}) \pi_j'(\theta_o)$$

$$= \frac{1}{n} \sum_i \mathrm{cov}_{\theta_o}(Q_n, \bar{p}_{in}) \left\{ \frac{\beta_i''(\theta_o)}{I} - \frac{\beta_i'(\theta_o) I'(\theta_o)}{I^2} \right\} \quad \text{(using (4.2.52))}$$

$$= \mathrm{cov}_{\theta} \left\{ Q_n, \frac{1}{n} \sum_i \left(\frac{\bar{p}_{in} \beta_i''(\theta_o)}{I} - \frac{\bar{p}_{in} \beta_i'(\theta_o) I'(\theta_o)}{I^2} \right) \right\}$$

$$= \operatorname{cov}_{\theta_o}\{Q_n, \tfrac{1}{n}(W_n I - Z_n I'(\theta_o))/I^2\} \quad \text{using (4.2.14) and (4.2.15)}$$

$$= \operatorname{cov}_{\theta_o}\{\Sigma \alpha_i(\bar{p}_{in} - \pi_i(\theta_o)), \tfrac{1}{n}(W_n I - Z_n I'(\theta_o)|I^2\} .$$

This completes the proof.

We are now in a position to prove the theorem.

Proof of Theorem 4.2.6 : We first prove (4.2.21).

$$0 = n^{-1} \frac{d \log L}{d\theta}\Big|_{\hat{\theta}_n} = Z_n + (\hat{\theta}_n - \theta_o)\, n^{-1} \frac{d^2 \log L}{d\theta^2}\Big|_{\theta_o}$$

$$+ \frac{1}{2n}(\hat{\theta}_n - \theta_o)^2 \frac{d^3 \log L}{d\theta^3}\Big|_{\theta_o} + o_E(n^{-1})$$

$$= Z_n + (\hat{\theta}_n - \theta_o)(W_n - I) + \frac{1}{2}(\hat{\theta}_n - \theta_o)^2 J + o_E(n^{-1}).$$

Hence

$$(\hat{\theta}_n - \theta_o) = \frac{Z_n}{I} + \frac{(\hat{\theta}_n - \theta_o)\, W_n}{I} + \frac{1}{2I}(\hat{\theta}_n - \theta_o)^2 J + o_E(n^{-1})$$

$$= \frac{Z_n}{I} + \frac{Z_n W_n}{I^2} + \frac{J}{2I^3} Z_n^2 + o_E(n^{-1}) \qquad (4.2.56)$$

since $\theta_n - \theta_o = \frac{Z_n}{I} + O_E(n^{-1})$ by Lemma 4.2.7. It follows now from (4.2.56) that

$$b_o(\theta_o) = \frac{\mu_{11}}{I^2} + \frac{J}{2I^2} \qquad (4.2.57)$$

Therefore

$$(\hat{\theta}_n - \theta_o) - \frac{b_o(\theta_o)}{n} = \frac{Z_n}{I} + \frac{(Z_nW_n - \mu_{11}/n)}{I^2} + \frac{J(Z_n^2 - I/n)}{2I^3} + o_E(n^{-1})$$

$$= \frac{Z_n}{I} + S_n + o_E(n^{-1})$$

$$= \frac{Z_n}{I} + S_n + \hat{R}_n \qquad (4.2.59)$$

which is (4.2.21) where S_n is defined in (4.2.19). From (4.2.59) it is clear that $\hat{R}_n$ is $o_E\ (n^{-1})$. Also using Lemma 4.2.10 and (4.2.43) to (4.2.47) it can be seen that $\hat{R}_n$ is E^U orthogonal to Z_n^2 and Z_nW_n up to $o(n^{-2})$. This proves (4.2.21).

We next prove (4.2.23). This is the crucial step. Suppose T_n is efficient and l.s.(II). Then by Lemma 4.2.9, the covariance of $T_n^{**}\ I_U$ and hence of $\left\{ T_n^{**} - \frac{b(\theta_o)}{n} \right\} I_U$ with Z_n^2 and Z_nW_n does not depend on T_n. (In fact to prove this one needs only (4.2.51) and (4.2.54) rather than the more explicit formulas (4.2.49) and (4.2.50) given in Lemma 4.2.9). Hence we may write the regression equation

$$\left\{ T_n^{**} - \frac{b(\theta_o)}{n} \right\} I_U = E_{\theta_o}\left[\left\{ T_n^{**} - \frac{b(\theta_o)}{n} \right\} I_U\right] + \alpha_n\ (Z_n^2 - I/n)$$

$$+ \beta_n(Z_nW_n - \mu_{11}/n) + \eta_n \qquad (4.2.59)$$

where α_n and β_n do not depend on T_n, η_n has zero covariance with Z_n^2 and Z_nW_n up to $o(n^{-2})$.

By Lemma 4.2.2 and Proposition 4.2.3 (i), $E_{\theta_o}[\{T_n^{**} - \frac{b(\theta_o)}{n}\} I_U]$ is $o(n^{-1})$. We may therefore lump this term and η_n and rewrite (4.2.59) as

$$\{T_n^{**} - \frac{b(\theta_o)}{n}\} I_U = \alpha_n(Z_n^2 - I/n) + \beta_n(W_n Z_n - \mu_{11}/n) + \eta_n^* \qquad (4.2.60)$$

where η_n^* is orthogonal to Z_n^2 and $Z_n W_n$ up to $o(n^{-2})$.

We may write (4.2.60) as

$$T_n - \theta_o - \frac{Z_n}{I} - \frac{b(\theta_o)}{n} = \alpha_n(Z_n^2 - I/n) + \beta_n(Z_n W_n - \frac{\mu_{11}}{n}) + \eta_n^{**} \qquad (4.2.61)$$

where $\eta_n^* = \eta_n^{**} I_U$ and so η_n^{**} is E^U-orthogonal to $(Z_n^2 - I/n)$ and $(Z_n W_n - \frac{\mu_{11}}{n})$ up to $o(n^{-2})$. We can calculate α_n and β_n directly but it may be illuminating to get it in an indirect but some what easier method. Since $\hat{\theta}_n$ is efficient and l.s. (II) (in fact l.s.(III)), we get on comparing (4.2.58) and (4.2.61) that

$$\alpha_n = J/2I^3 \quad \text{and} \quad \beta_n = \frac{1}{I^2} . \qquad (4.2.62)$$

From (4.2.61) and (4.2.62) we get

$$T_n - \theta_o - \frac{Z_n}{I} - \frac{b(\theta_o)}{n} = \frac{J(Z_n^2 - I/n)}{2I^3} + \frac{(Z_n W_n - \mu_{11}/n)}{I^2} + \eta_n^{**}$$

$$= S_n + R_n \qquad (4.2.63)$$

where R_n is E^U-orthogonal to Z_n^2 and $Z_n W_n$ up to $o(n^{-2})$.

Since

$$R_n = T_n - \theta_o - \frac{b(\theta_o)}{n} - \frac{Z_n}{I} - S_n \tag{4.2.64}$$

it follows from Lemma 4.2.7, applied to T_n, that R_n is $O_E(n^{-1})$. That $E^U(R_n)$ is of order $o(n^{-1})$ follows from (4.2.64) and the definition of $b(\theta_o)$ and S_n. Thus R_n satisfies all the conditions stated in Theorem 4.2.6, completing the proof of (4.2.23).

To prove (4.2.24), let us recall that

$$E_2(\{T_n\}, \theta_o, \lambda) = \lim_{n\to\infty} n^2 E^U\{Z_n - (T_n - \theta_o) I - \lambda (T_n - \theta_o)^2 - a_\lambda(\theta_o)/n\}^2$$

where $a_\lambda(\theta_o)$ is defined in Proposition 4.2.3 as

$$a_\lambda(\theta_o) = \lim_{n\to\infty} E^U n \{Z_n - (T_n - \theta_o) I - \lambda(T_n - \theta_o)^2\}$$

$$= - b(\theta_o). I - \lambda/I, \quad \text{since } T_n - \theta_o = Z_n/I + o_E(n^{-1}).$$

So

$$E_2(\{T_n\}, \theta_o, \lambda) = \lim n^2 E^U \{Z_n - (T_n - \theta_o - \frac{b(\theta_o)}{n}) I - \frac{\lambda}{I^2}(Z_n^2 - I/n)\}^2$$

$$= \lim n^2 E^U \{I S_n + I R_n + \frac{\lambda}{I^2} (Z_n^2 - I/n)\}^2 \text{ using (4.2.63))}$$

$$= \lim n^2 E^U \{I S_n + \frac{\lambda}{I^2} (Z_n^2 - I/n)\}^2 + \lim n^2 E^U\{I R_n\}^2$$

(since R_n is E^U-orthogonal to S_n and Z_n^2 up to $o(n^{-2})$).

$$= E_2(\{\hat{\theta}_n\}, \theta_o, \lambda) + \lim n^2 E^U(I R_n)^2$$

Now since

$$E_2(\{\hat{\theta}_n\},\theta_o,\lambda) = \lim n^2 E^U \{Z_n - (\hat{\theta}_n - \theta_o) - \frac{b_o(\theta_o)}{n}) I - \frac{\lambda}{I^2}(Z_n^2 - I/n)\}^2$$

$$= \lim n^2 E^U \{I\ S_n + \frac{\lambda}{I^2}(Z_n^2 - I/n)\}^2 \text{(using(4.2.59))}$$

it follows that

$$\inf_\lambda E_2(\{T_n\},\theta_o,\lambda) = \inf_\lambda E_2(\{\hat{\theta}_n\},\theta_o,\lambda) + \lim n^2 E^U (I\ R_n)^2$$

and

$$E_2(\{T_n\},\theta_o) \geq E_2(\{\hat{\theta}_n\},\theta_o)$$

with equality iff

$$E^U(I\ R_n)^2 = o\ (n^{-2}).$$

This completes the proof of (4.2.24).

We now derive (4.2.22).

$$\hat{\theta}_n^* - \theta_o - Z_n/I = \hat{\theta}_n - \frac{b(\hat{\theta}_n)}{n} - \theta_o - Z_n/I \quad \text{(from (4.2.35))}$$

$$= \hat{\theta}_n - \{\frac{b_o(\theta_o)}{n} + \frac{Z_n}{I} \cdot \frac{b_o'(\theta_o)}{n}\} - \theta_o - \frac{Z_n}{I} + o_E(n^{-3/2})$$

(using (4.2.34))

$$= (\hat{\theta}_n - \theta_o - \frac{b_o(\theta_o)}{n} - \frac{Z_n}{I}) - \frac{b_o'(\theta_o)}{n} Z_n/I + o_E(n^{-3/2})$$

$$= S_n + \hat{R}_n - \frac{b_o'(\theta_o)}{n}\frac{Z_n}{I} + o_E(n^{-3/2}) \quad \text{(using (4.2.21))}$$

$$= \frac{(Z_n W_n - \mu_{11}/n)}{I^2} + \frac{J(Z_n^2 - I/n)}{2I^3} - \frac{b_o'(\theta_o)}{n}\frac{Z_n}{I} + \hat{R}_n + o_E(n^{-3/2})$$

$$= \frac{J}{2I^3}\{Z_n^2 - \frac{I}{n} - \frac{\mu_{30} Z_n}{n\mu_{20}}\} + \frac{1}{I^2}\{Z_n W_n - \frac{\mu_{11}}{n} - \frac{\mu_{21} Z_n}{n\mu_{20}}\} + \hat{R}_n^* \tag{4.2.65}$$

where

$$\hat{R}_n^* = \hat{R}_n + \frac{Z_n}{n}\left\{\frac{-b_o'(\theta_o)}{I} + \frac{J\,\mu_{30}}{2I^4} + \frac{\mu_{21}}{I^3}\right\} + o_E(n^{-3/2})$$

$$= \hat{R}_n + o_E(n^{-1})$$

and so $\hat{R}_n^*$ is E^U-orthogonal to Z_n^2 and $Z_n W_n$ up to $o(n^{-2})$. By Lemma 4.2.8 the left hand side of (4.2.65) is E^U-orthogonal to Z_n up to $o(n^{-2})$. By easy direct computation the same result is true of the first two terms on the right hand side of (4.2.65). Hence $\hat{R}_n^*$ is also E^U-orthogonal to Z_n up to $o(n^{-2})$. So, $\hat{R}_n^*$ has the properties asserted in Theorem 4.2.6. Also observe that (4.2.65) can be rewritten as

$$\hat{\theta}_n^* - \theta_o - \frac{Z_n}{I} = -\frac{Z_n}{2nI^4}\left\{2\,\mu_{21} I + J.\mu_{30}\right\} + S_n + \hat{R}_n^*$$

which is nothing but (4.2.22). This completes the proof of (4.2.22).

To prove (4.2.25) we proceed as follows.

By Lemmas 4.2.8 and 4.2.10, $(T_n^* - \theta_o)I_U$ has the same covariance up to $o(n^{-2})$ with Z_n, $Z_n W_n$ and Z_n^2 as $(\hat{\theta}_n^* - \theta_o)I_U$. So (4.2.25) can be deduced from (4.2.22) in the same way as (4.2.23) was deduced from (4.2.21). It is easy to check that $R_n - R_n^* = o_E(n^{-1})$. But from (4.2.25) we have

$$E(T_n^* - \theta_o)^2 = E(\hat{\theta}_n^* - \theta_o)^2 + E(R_n^*)^2 + o\ (n^{-2})$$

Hence

$$n^2[E(T_n^* - \theta_o)^2 - E(\hat{\theta}_n^* - \theta_o)^2] = n^2\ E(R_n^{*2}) + o\ (1) > 0,$$

since

$$n^2\ E_{\theta_o}(R_n^{*2}) = 0\ (1),$$

i.e.,

$$\psi(\{T_n^*\}, \theta_o) = \psi(\{\hat{\theta}_n^*\}, \theta_o) + \lim n^2\ E^U(R_n^{*2}) \qquad (4.2.66)$$

which implies

$$\psi(\{T_n^*\}, \theta_o) > \psi(\{\hat{\theta}_n^*\}, \theta_o)$$ proving (4.2.26).

We next prove (4.2.27). From (4.2.66) we have

$$\psi(\{T_n^*\}, \theta_o) = \psi(\{\hat{\theta}_n^*\}, \theta_o) + \lim n^2\ E^U(R_n^2) \qquad (4.2.67)$$

since

$$R_n - R_n^* = o_E(n^{-1}) \quad \text{and} \quad R_n = 0_E(n^{-1}).$$

Clearly,

$$\psi(\{\hat{\theta}_n^*\}, \theta_o) = \lim n^2\ E^U(S_n^2) = \frac{1}{I^4}\{I\ \mu_{02} - \mu_{11}^2\} + \frac{2}{I^4}\{\frac{J}{2} + \mu_{11}\}^2 \qquad (4.2.68)$$

Also from the proof of (4.2.24) one has

$$E_2(\{T_n\}, \theta_o) = \lim n^2 I^2 E^U(R_n^2) + \lim n^2 E^U\{Z_n W_n - \mu_{11}/n)\frac{1}{I^2} - \gamma\ (Z_n^2 - I/n)\}^2$$

where γ is the limiting regression coefficient of $I^{1/2}(Z_n W_n - \mu_{11}/n)$ on $(Z_n^2 - I/n)$ and is found to be μ_{11}/I^3 applying (4.2.43) and (4.2.47). So

$$E_2(\{T_n\}, \theta_o) = \lim n^2 I^2 E^U(R_n^2) + \frac{1}{I^2}\{I \cdot \mu_{02} - \mu_{11}^2\} \qquad (4.2.69)$$

using (4.2.47). So by (4.2.67),(4.2.68) and (4.2.69) it follows that

$$E_2(\{T_n\}, \theta_o) = I^2 \cdot \varphi(\{T_n\}, \theta_o) - \frac{2}{I^2}\{J/2 + \mu_{11}\}^2$$

proving (4.2.27). This completes the proofs of (i) and (ii).

Proof of (iii) : Let T_n be efficient and l.s.(III). Let m and T_n' be as in the statement of Theorem 4.2.6(iii). Let

$$\hat{\theta}_n' = \hat{\theta}_n - \{b_o(\hat{\theta}_n) - b(\hat{\theta}_n) - m(\hat{\theta}_n)\}/n$$

where $b(\theta_o)$ is defined in Proposition 4.2.4.

Then as in the proof of Lemma 4.2.8 it can be shown that $\hat{\theta}_n'$ and T_n' have the same covariance with Z_n up to $o(n^{-2})$. Since

$$[\hat{\theta}_n - \{b_o(\theta) - b(\theta) - m(\theta)\}/n - \hat{\theta}_n'] \text{ and } (T_n + m(\theta)/n - T_n']$$

are $o_E(n^{-1})$, we can apply Lemma 4.2.9 to conclude that $\hat{\theta}_n'$ and T_n' have the same covariance with $Z_n W_n$ and Z_n^2 up to $o(n^{-2})$.

It follows as in the proof of (4.2.26) that

$$\varphi(\{T_n'\}, \theta_o) \geq \varphi(\{\hat{\theta}_n'\}, \theta_o)$$

which leads to the desired conclusion by Proposition 4.2.4.

This completes the proof of Theorem 4.2.6.

Note that the main difference between (4.2.21) and (4.2.23) is that $\hat{R}_n$ is $o_E^U(n^{-1})$ where as R_n is only $O_E(n^{-1})$. This is at the root of a result like (4.2.24). The main difference between $\hat{R}_n$ and $\hat{R}^*_n$ is that $\hat{R}^*$ is E^U-orthogonal to Z_n up to $o(n^{-2})$ but $\hat{R}_n$ is not. A similar remark applies to R_n and R^*_n. The significance of (4.2.25) for proving (4.2.26) should be self evident.

We conclude this section by considering the case when we have two or more parameters to be estimated. In what follows we indicate briefly that Theorem 4.2.6 can be generalised to include the case of two or more parameters. For simplicity, we state the result for two parameters.

Suppose $\theta \in (\bar{H})$ is a vector with two real coordinates $\theta = (\theta^{(1)}, \theta^{(2)})$ and $T_n = (T_n^{(1)}, T_n^{(2)})$ is an efficient estimate. We can define the local stability of orders II and III and efficiency of an estimate T_n exactly in the same way as we have done in (4.2.5). We make the following assumption.

Assumption (A). Assume that $k \geq 2$, $\pi_1(\theta^{(1)},\theta^{(2)})$, $\pi_2(\theta^{(1)},\theta^{(2)})$, ..., $\pi_k(\theta^{(1)}, \theta^{(2)})$ are thrice continuously differentiable and that the rank of

$$\begin{pmatrix} \dfrac{\partial \pi_1(\theta^{(1)}, \theta^{(2)})}{\partial \theta^{(1)}}, \ldots, & \dfrac{\partial \pi_k(\theta^{(1)},\theta^{(2)})}{\partial \theta^{(1)}} \\[2ex] \dfrac{\partial \pi_1(\theta^{(1)}, \theta^{(2)})}{\partial \theta^{(2)}}, \ldots, & \dfrac{\partial \pi_k(\theta^{(1)}, \theta^{(2)})}{\partial \theta^{(2)}} \end{pmatrix}$$

is 2 for all $\theta \varepsilon (\bar{H})$. Suppose $\theta_o = (\theta_o^{(1)}, \theta_o^{(2)})$ is the true value of the parameter. For some $\lambda = (\lambda_{ij})_{2\times 2}$ and $\gamma = (\gamma_{ij})_{2\times 2}$, let

$$\phi_1(\lambda \{T_n\}, \theta_o) = n\{\frac{1}{n} \frac{\partial \log L}{\partial \theta^{(1)}}\Big|_{\theta_o} - (T_n^{(1)} - \theta_o^{(1)}) I_{11}$$

$$- (T_n^{(2)} - \theta_o^{(2)}) I_{12} - \Sigma \lambda_{ij} (T_n^{(i)} - \theta_o^{(i)}) (T_n^{(j)} - \theta_o^{(j)})\}$$

$$\phi_2(\gamma, \{T_n\}, \theta_o) = n \{\frac{1}{n} \frac{\partial \log L}{\partial \theta^{(2)}}\Big|_{\theta_o} - (T_n^{(1)} - \theta_o^{(1)}) I_{11}$$

$$- (T_n^{(2)} - \theta_o^{(2)}) I_{22} - \Sigma \gamma_{ij} (T_n^{(i)} - \theta_o^{(i)}) (T_n^{(j)} - \theta_o^{(j)})\}$$

where $[I_{ij}]$ is the 2×2 information matrix defined by

$$I_{ij}(\theta_o) = \frac{1}{n} E_{\theta_o}\{\frac{\partial \log L}{\partial \theta_o^{(i)}} \frac{\partial \log L}{\partial \theta_o^{(j)}}\}, \quad i,j = 1,2.$$

Let

$$a_\lambda(\theta_o) = \lim E^U \phi_1(\lambda, \{T_n\}, \theta_o)$$

and

$$a_\gamma(\theta_o) = \lim E^U \phi_2(\gamma, \{T_n\}, \theta_o).$$

Define

$$E_2^{(1)}(\{T_n\}, \theta_o, \lambda, U) = \lim[n^2 E^U\{\phi_1(\lambda, \{T_n\}, \theta_o) - a_\lambda(\theta_o)/n\}^2]$$

and

$$E_2^{(2)}(\{T_n\}, \theta_o, \gamma, U) = \lim[n^2 E^U\{\phi_2, (\gamma, \{T_n\}, \theta_o) - a_\gamma(\theta_o)/n\}^2].$$

Then it can be shown that (as in Section 4.2) $E_2^{(1)}$ and $E_2^{(2)}$ are independent of U, where U is a neighbourhood of $\pi(\theta_o)$. Now consider the limiting dispersion matrix of $\phi_1(\lambda, \{T_n\}, \theta_o)$ and $\phi_2(\gamma, \{T_n\}, \theta_o)$ where the values of λ_{ij} and γ_{ij} are chosen such that $E_2^{(1)}$ and $E_2^{(2)}$ are minimum. This limiting ~~dispersion~~ matrix can be defined as second order efficiency of T_n (Rao's first measure).

To define Rao's second measure we proceed as follows. Define $T_n^* = T_n - \frac{b(T_n)}{n}$. Now consider the expansion for the dispersion matrix of T_n^* and look at the coefficient matrix of $1/n^2$. Suppose we denote this coefficient matrix by $\psi(\{T_n^*\}, \theta_o)$. Suppose $\hat{\theta}_n = (\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})$ is the m.l.e of θ_o. Then we have the following theorem which can be proved exactly on the same lines as that of the one parameter case.

Theorem 4.2.11 (i) If Assumption (A) holds and T_n is l.s.(II) and first order efficient then the difference between the limiting dispersion matrix of $\phi_1(\lambda, \{T_n\}, \theta_o)$, $\phi_2(\gamma, \{T_n\}, \theta_o)$ and that of $\phi_1(\lambda, \{\hat{\theta}_n\}, \theta_o)$, $\phi_2(\gamma, \{\hat{\theta}_n\}, \theta_o)$ is positive semidefinite.

(ii) If Assumption (A) holds and T_n is l.s (III) and first order efficient then

$$\psi(\{T_n^*\}, \theta_o) - \psi(\{\hat{\theta}_n^*\}, \theta_o)$$

is a positive semidefinite matrix.

4.3 Some Remarks

Remark 4.3.1. The expansions obtained above are not affected by the singularity of the dispersion matrix of Z_n, $Z_n W_n$, Z_n^2 up to $o(n^{-2})$ but it is worth pointing out that singularity up to $o(n^{-2})$ obtains if and only if there is a linear relation between $Z_n W_n$ and Z_n^2 up to $o(n^{-1})$, which can be true if and only if there is a linear relation between Z_1 and W_1. If such a relation holds for all θ in some open set then $f(x,\theta)$ is essentially a one dimensional exponential family. For, the fact that Z_1 and W_1 are linearly related taken with the linear independence of $\bar{p}_{11},\dots,\bar{p}_{k1}$ gives us

$$\beta_i''(\theta)/\beta_i'(\theta) = g(\theta)$$

the solution of which can be written in the form $\beta_i(\theta) = a_i\beta_1(\theta)+b_i$ where a_i and b_i are constants. Hence $\log f(x,\theta) = c(\theta) + \beta_1(\theta) \sum_{i=1}^{k} a_i \bar{p}_{i1} + \sum_{i=1}^{k} b_i$. Of course if $f(x,\theta)$ is one dimensional exponential density then there is a linear relation between Z_1 and W_1.

Remark 4.3.2. $E_2(\{\hat{\theta}_n\}, \theta_o) = \frac{1}{I^2}(I\mu_{02} - \mu_{11}^2) = 0$ if and only if there is a linear relation between Z_1 and W_1. We have seen in Remark 4.2.1 that if this result holds for all θ_o in some open set, then $f(x,\theta)$ is essentially a one dimensional exponential density. Moreover if there is a linear relationship between Z_1 and W_1 and $J + 2\mu_{11} = 0$, then $\varphi(\{\hat{\theta}_n^*\}, \theta_o) = 0$.

Remark 4.3.3. It may be illuminating to give a "formal" proof of the following result

$$I. E^U(T_n^* Z_n M_n) = \frac{1}{n} E^U(H_n M_n)$$

where $M_n = \Sigma \alpha_i(\bar{p}_{in} - \pi_i(\theta_o))$ and $H_n = (W_n - Z_n I')/I$.

Note that

$$\frac{1}{n} E^U \{(I. T_n^*) M_n\} = o(n^{-2}).$$

Hence differentiating this w.r.t. θ we get formally

$$E^U\{(I T_n^*) Z_n M_n\} = -\frac{1}{n} E^U \{\frac{d}{d\theta}(I T_n^*) M_n\}$$

$$+ \frac{1}{n} E^U(T_n^*) \Sigma \alpha_i \pi_i' + o(n^{-2})$$

$$= -\frac{1}{n} E^U \{\frac{d}{d\theta}(I T_n^*) M_n\} + o(n^{-2})$$

$$= \frac{1}{n} E^U(H_n M_n)$$

since $H_n + I'T_n^* = o_E(n^{-1/2})$.

Remark 4.3.4. From the view point of second order efficiency, we have seen that smaller the measures E_2 and Ψ the better is the estimate. In fact they are really measures of deficiency as defined in a more general context by Hodges and Lehmann (1970). Deficiency of $\{T_n\}$ relative to $\{\hat{\theta}_n\}$ is

$$[\Psi(\{T^*\}, \theta_o - (\{\hat{\theta}_n^*\}, \theta_o)]/I.$$

Remark 4.3.5. To facilitate comparison with Rao's (1961, 1963) results for the multinomial distribution, let us denote by m_{ij} what Rao (1963) denotes as μ_{ij}. We shall show how to express m_{ij} in terms of μ_{ij} and vice versa. Let the multinomial population consists of k+1 classes with probabilities $\pi_1, \pi_2, \ldots, \pi_{k+1}(\theta)$. Let

$$U_i = \begin{cases} 1 & \text{if first observation falls in the i-th class} \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$Z_1 = \sum_{i=1}^{k+1} \left(\frac{\pi_i'}{\pi_i}\right) U_i \quad \text{and} \quad W_1 = \sum_{i=1}^{k+1} \frac{d^2 \log \pi_i}{d\theta^2} U_i + I.$$

Define $\mu_{ij} = E_{\theta_0}(Z_1^i W_1^j)$. Also denote $Y = \Sigma\left(\frac{\pi_i''}{\pi_i}\right)U_i$. Then, following Rao (1963), $m_{ij} = E_{\theta_0}(Z_1^i Y^j)$. Clearly $W_1 = Y-(Z_1^2)+ I$. Using this we can express m_{ij} in terms of μ_{ij} and vice versa. For example, $\mu_{11} = m_{11} - m_{30}$.

Remark 4.3.6. Let us try to understand the calculations of second order efficiency given by Fisher (1925). As Rao (1961, equation 5.14) has pointed out that Fisher's measure $E_2' = \lim (nI-I_{T_n})$ can be shown to be equal to the limit of the expectation of the conditional variance of nZ_n given T_n. Consider the expansion

$$Y_n = Z_n + (T_n - \theta_o) \frac{1}{n} \left. \frac{d \log L}{d\theta} \right|_{\theta_o}$$

where

$$Y_n = \frac{1}{n} \left. \frac{d \log L}{d\theta} \right|_{T_n} .$$

Suppose T_n is efficient and l.s.(II). Note that this expansion is not correct up to $0_E(n^{-1})$ but the missing term of $0_E(n^{-1})$ is a function of T_n and bynot including that term does not cause any error in the calculation of conditional variance. Thus, a correct expansion is

$$Y_n = Z_n + (T_n - \theta_o) \frac{1}{n} \left. \frac{d^2 \log L}{d\theta^2} \right|_{\theta_o} + \frac{(T_n - \theta_o)^2}{2} \frac{1}{n} \left. \frac{d^3 \log L}{d\theta^2} \right|_{\theta_o} + o_E(n^{-1})$$

which can be rewritten as

$$Z_n = Y_n - \frac{Z_n W_n}{I} + (T_n - \theta_o) I + (T_n - \theta_o)^2 J/2 + o_E(n^{-1}). \qquad (4.3.1)$$

Fisher now takes the conditional expectation using the joint asymptotic normal distribution of the $\bar{p}_{in}$'s and replaces the condition " $T_n - \theta_o$ = constant " by " Z_n = constant ". Let us note that these calculations lead to Rao's measure E_2. For, the " conditional expectation " of $Y_n - \frac{Z_n W_n}{I}$ is easily seen to be of the form $\lambda_o Z_n^2 + c/n$ since by (4.3.1) and Lemma 4.2.7, $Y_n - Z_n$ can be written in the form $\sum_i \sum_j \alpha_{ij} (\bar{p}_{in} - \pi_i)(\bar{p}_{jn} - \pi_j)$.

Also by (4.3.1)

$$T_n - \theta_o = \frac{Z_n}{I} - \frac{Y_n}{I} + \frac{Z_n W_n}{I^2} + \frac{Z_n^2 J}{2I^3} + o_E(n^{-1})$$

so that by Theorem 4.2.6,

$$\frac{Y_n}{I} = - R_n + o_E(n^{-1})$$

where R_n is defined in (4.2.23).

Evaluating λ_o and c one finds, the " conditional variance " of nZ_n given $T_n^{\cdot}$ is the " expectation " of

$$\{ Y_n - \frac{Z_n W_n}{I} - \lambda_o Z_n^2 - c/n \}^2$$

which equals

$$E^U [I^2 \{- R_n - \frac{Z_n W_n}{I^2} - \lambda_o Z_n^2 - c/n\}^2]$$

$$= E_2 (\{ T_n \}, \theta_o).$$

Thus the measure that Fisher calculates is exactly the measure E_2 of Rao.

Fisher seems to believe, wrongly as it turns out, that Y_n is independent of $(T_n - \theta_o)$ or Z_n up to second order terms. But it is true that Y_n is E^U - orthogonal to Z_n^2 and $Z_n W_n$ up to $o(n^{-2})$. Thus

$$\text{Fisher's measure} = E_2 = \lim E^U(Y_n^2) + E_2(\{\hat{\theta}_n\}, \theta_o).$$

Fisher gets wrong result for the minimum chi-square method because he substitutes the variance of

$$Y_n' = \frac{1}{2} \Sigma \frac{(\bar{p}_{in} - \pi_i(\theta_o))^2}{\pi_i^2(\theta_o)} \pi_i'(\theta_o)$$

for the varinace of

$$Y_n = \frac{1}{2} \Sigma \frac{(\bar{p}_{in} - \pi_i(T_n))^2}{\pi_i^2(T_n)} \pi_i'(T_n).$$

Remark 4.3.7. Suppose that instead of the criterion E_2 one considers

$$E_2'' = \inf_{\lambda_1, \lambda_2} \lim n^2 E^U \{Z_n - (T_n - \theta_o)\lambda_1 - (T_n - \theta_o)^2 \lambda_2 - b_{\lambda_1, \lambda_2}(\theta_o)/n\}^2$$

where

$$b_{\lambda_1,\lambda_2}(\theta_o) = \lim n E^U \{ Z_n - (T_n - \theta_o) \lambda_1 - (T_n - \theta_o)^2 \lambda_2\}.$$

Then for E_2'' we have the following result.

Suppose Assumption (I) holds and T_n is efficient and l.s.(II). Let $\hat{\theta}_n'$ be defined as in Theorem 4.2.6(iii) with $m \equiv 0$. Then

$$E_2''(\{T_n\}, \theta_o) \geq E_2''(\{\hat{\theta}_n'\}, \theta_o).$$

The proof is similar to that of Theorem 4.2.6(iii).

<u>Remark 4.3.8</u>. Suppose the likelihood equation is not linear in θ. Then to get the m.l.e one applies the method of scoring proposed by Fisher (See Rao (1973), p.366). The main criticism against the m.l.e is that it is difficult to compute. We show now that two iterations in the method of scoring is good enough up to $o(n^{-1})$. Let $\hat{\theta}(0) = \hat{\theta}_n(0)$ denote the initial value such that $\hat{\theta}(0) - \theta_o = O_E(n^{-1/2})$. Let $\hat{\theta}(1)$ and $\hat{\theta}(2)$ be the values at the first and second iterations. Denote

$$L_{(i)} = \frac{1}{n} \left. \frac{d^i \log L}{d\theta^i} \right|_{\hat{\theta}(0)} .$$

Then we define $\hat{\theta}(1)$ by

$$0 = L_{(1)} + (\hat{\theta}(1) - \hat{\theta}(0))\, L_{(2)}$$

such that

$$\hat{\theta}(1) - \hat{\theta}(0) = - L_{(1)}/L_{(2)} + o_E(n^{-1/2}). \tag{4.3.2}$$

Similarly,

$$0 = \frac{1}{n} \left. \frac{d \log L}{d\theta} \right|_{\hat{\theta}(1)} + (\hat{\theta}(2) - \hat{\theta}(1)) \frac{1}{n} \left. \frac{d^2 \log L}{d\theta^2} \right|_{\hat{\theta}(1)}$$

$$= L_{(1)} + (\hat{\theta}(1) - \hat{\theta}(0))\, L_{(2)} + \frac{(\hat{\theta}(1) - \hat{\theta}(0))^2}{2} L_{(3)}$$

$$+ (\hat{\theta}(2) - \hat{\theta}(1)) \{L_{(2)} + (\hat{\theta}(1) - \hat{\theta}(0))\, L_{(3)}\} + o_E(n^{-1})$$

$$= \frac{1}{2} \frac{L_{(1)}^2}{L_{(2)}^2} \cdot L_{(3)} + (\hat{\theta}(2) - \hat{\theta}(1)) \{L_{(2)} - \frac{L_{(1)}}{L_{(2)}} L_{(3)}\} + o_E(n^{-1})$$

using (4.3.2).

Thus we arrive at

$$\hat{\theta}(2) - \hat{\theta}(1) = -\frac{1}{2}\frac{L_{(1)}^2}{L_{(2)}^2} \cdot L_{(3)} \cdot \left[L_{(2)} - \frac{L_{(1)}}{L_{(2)}}\right]^{-1} + o_E(n^{-1})$$

$$= -\frac{1}{2}\frac{L_{(1)}^2}{L_{(2)}^3} L_{(3)} + o_E(n^{-1}).$$

Hence

$$\hat{\theta}(2) - \hat{\theta}(0) = -\frac{L_{(1)}}{L_{(2)}} - \frac{L_{(1)}^2 L_{(3)}}{2L_{(2)}^3} + o_E(n^{-1}). \qquad (4.3.3)$$

Further note that

$$0 = \frac{1}{n}\frac{d \log L}{d\theta}\bigg|_{\hat{\theta}_n} = L_{(1)} + (\hat{\theta}_n - \hat{\theta}(0)) L_{(2)} + \frac{(\hat{\theta}_n - \hat{\theta}(0))^2}{2} L_{(3)} + o_E(n^{-1})$$

which gives,

$$\hat{\theta}_n - \hat{\theta}(0) = \frac{-L_{(2)} + \{L_{(2)}^2 - 2L_{(1)}L_{(3)}\}^{\frac{1}{2}}}{L_{(3)}} + o_E(n^{-1})$$

$$= \frac{-L_{(2)} + L_{(2)}\left\{1 - \frac{2L_{(1)}L_{(3)}}{L_{(2)}^2}\right\}^{\frac{1}{2}}}{L_{(3)}} + o_E(n^{-1})$$

$$= -\frac{L_{(1)}}{L_{(2)}} - \frac{1}{2}\frac{L_{(1)}^2 L_{(3)}}{L_{(2)}^3} + o_E(n^{-1}).$$

From (4.3.3) and (4.3.4) we have

$$\hat{\theta}_n - \hat{\theta}(2) = o_E(n^{-1}).$$

The other root of the quadratic in (4.3.4) is not consistent with the assumption $(\hat{\theta}_n - \hat{\theta}(0) = 0_E(n^{-1/2})$.

Remark 4.3.9. We conclude this section by developing an asymptotic Bhattacharya bound for efficient estimates.

Suppose T_n is an estimate such that

$$E^U(T_n) = \theta + o\ (n^{-1}) \tag{4.3.5}$$

and

$$E^U(T_n \cdot \frac{1}{n^2} \frac{1}{L} \frac{d^2L}{d\theta^2} = o\ (n^{-2}) \tag{4.3.6}$$

((4.3.6) follows from (4.3.5) differentiating it formally).

If T_n is efficient we may expect

$$E^U \{(T_n - \theta_o - Z_n/I) \cdot W_n/n\} = o(n^{-2}). \tag{4.3.7}$$

An estimate T_n is regular if T_n is efficient and (4.3.5), (4.3.6) and (4.3.7) hold.

If T_n is efficient and l.s.(III), then it can be shown easily that T_n is regular, since

$$\frac{1}{n^2} \frac{1}{L} \frac{d^2L}{d\theta^2} = W_n/n + (Z_n^2 - I/n). \tag{4.3.8}$$

If T_n is regular then by (4.3.6), (4.3.7) and (4.3.8), we have

$$E^U \{(T_n - \theta_o - Z_n/I)\ (Z_n^2 - I/n)\} = \frac{2\mu_{11} + J}{n^2 I} + o\ (n^{-2}).$$

Using the regression of $T_n I_U$ on Z_n and $\frac{1}{n^2}\frac{d^2L}{d\theta^2}$ we get

$$E^U_{\theta_o}(T_n - \theta_o)^2 \geq \frac{I}{nI} + \frac{1}{n^2 I^4}\left\{\frac{J}{2} + \mu_{11}\right\} + o\,(n^{-2}).$$

It is easy to show that $\hat{\theta}^*_n$ attains this bound if and only if there is a linear relation between Z_1 and W_1. The implication of this last relation has been discussed in Remark 4.3.1.

4.4 A Problem of Berkson

Suppose a dose d_i of some drug is given to the j-th animal, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,n$. Let π_i denote the probability of death at dose d_i and $p_i(x_j) = 1$ if the j-th animal receiving dose d_i dies, $= 0$ otherwise. Suppose that an experiment with doses $d_1,d_2,\ldots,d_k$ is repeated n times with a total of $n.k$ animals. Then the likelihood function is given by

$$L = \prod_{i=1}^{k} \pi_i^{n\,\bar{p}_{in}} (1-\pi_i)^{n(1-\bar{p}_{in})}$$

where

$$\bar{p}_{in} = \frac{1}{n}\sum_{j=1}^{n} p_i(x_j).$$

Then

$$L = c(\pi)\exp\left\{n\sum_{i=1}^{k}\left\{\log\left(\frac{\pi_i}{1-\pi_i}\right)\right\}\bar{p}_{in}\right.$$

Berkson now assumes a logistic model for the π_i's i.e., Probability of death = $\pi_i(\alpha, \beta) = 1/e^{-(\alpha+\beta d_i)}$, $i = 1,2,\ldots,k$

where α and β are usually unknown. Let us assume β to be known and $\alpha = \theta$ to be unknown. Then

$$\frac{\pi_i(\theta)}{1 - \pi_i(\theta)} = e^{\theta + \beta d_i}.$$

This implies

$$L = c(\theta) \exp\left\{n \sum_{i=1}^{n} (\theta + \beta d_i) \bar{p}_{in}\right\}$$

which is the likelihood function for one parameter exponential family. Hence this is a special case of the problem considered in Section 4.2.

The likelihood equation is

$$0 = \frac{d \log L}{d\theta} = n \sum_{i=1}^{k} (\bar{p}_{in} - \pi_i(\theta)) . \qquad (4.4.1)$$

Let $\hat{\theta}_n$ be the maximum likelihood estimate of θ,

$$L_i = \log\{\pi_i/1 - \pi_i\} = \theta + \beta d_i ,$$

and

$$\Lambda_i^n = \log\{\bar{p}_{in}/1 - \bar{p}_{in}\} .$$

Minimising $\sum_1^k n\, \bar{p}_{in}(1 - \bar{p}_{in}) (\Lambda_i^n - L_i)^2$ with respect to θ one gets the minimum logit chi-square estimate T_n. T_n is to be found from

$$n \sum_{1}^{k} \bar{p}_{in} (1 - \bar{p}_{in}) (\lambda_i^n - L_i^*) = 0 \qquad (4.4.2)$$

where $L_i^* = T_n + \beta d_i$.

Note that here Assumption (I) holds and both T_n and $\hat{\theta}_n$ are efficient and l.s. (III). The expansions for these estimates become, after some simplications

$$\hat{\theta}_n - \theta_o - \frac{Z_n}{I} = \frac{\sum_{i=1}^{k} \pi_i(1-\pi_i)(2\pi_i-1)}{2I^3} Z_n^2 + o_E(1/n) \qquad (4.4.3)$$

and

$$T_n - \theta_o - \frac{Z_n}{I} = \sum_{i=1}^{k} (2\pi_i - 1) (\bar{p}_{in}-\pi_i) Z_n/I^2$$

$$- \frac{1}{2I} \sum_{i=1}^{k} \frac{(2\pi_i-1)}{\pi_i(1 - \pi_i)} (\bar{p}_{in} - \pi_i)^2 + o_E(1/n) \qquad (4.4.4)$$

where

$$I = \sum_{i=1}^{k} \pi_i(1 - \pi_i) \quad \text{and} \quad Z_n = \frac{1}{n} \frac{d \log L}{d\theta} .$$

Also note that

$$E_{\theta_o}^U (T_n) = \theta_o + \frac{b(\theta_o)}{n} + o\,(n^{-1})$$

and

$$E_{\theta_o}^U (\theta_o) = \theta_o + \frac{b_o(\theta_o)}{n} + o\,(n^{-1})$$

where

$$b(\theta_o) = \Sigma\, \pi_i(1-\pi_i)(2\pi_i-1)/I^2 - \Sigma(2\pi_i-1)/2I$$

and

$$b_o(\theta_o) = \Sigma \pi_i(1 - \pi_i)(2\pi_i - 1)/2I^2 . \qquad (4.4.6)$$

If we consider the corresponding estimates $\hat{\theta}_n^*$ and T_n^* which are corrected for bias up to $o(n^{-1})$, we get

$$E^U_{\theta_o}(\hat{\theta}^*_n - \theta_o)^2 = \frac{1}{nI} + \frac{[\Sigma \pi_i(1-\pi_i)(2\pi_i-1)]^2}{2n^2 I^4} + o(n^{-2}) \qquad (4.4.7)$$

$$E^U_{\theta_o}(T^*_n - \theta_o)^2 = \frac{1}{nI} + \frac{[\Sigma \pi_i(1-\pi_i)(2\pi_i-1)]^2}{n^2 I^4}$$
$$+ \frac{\Sigma(2\pi_i-1)^2}{2n^2I^2} - \frac{\Sigma \pi_i(1-\pi_i)(2\pi_i-1)^2}{n^2I^3} + o(n^{-2}). \qquad (4.4.8)$$

It follows from the theorem of Section (4.2) or can be checked directly using (4.4.5) and (4.4.6) that

$$E^U_{\theta_o}(T^*_n - \theta_o)^2 - E^U_{\theta_o}(\hat{\theta}^*_n - \theta_o)^2 = E^U_{\theta_o}(T^*_n - \hat{\theta}^*_n)^2 + o(n^{-2}).$$

Define

$$\hat{\hat{\theta}}_n = \hat{\theta}_n + \frac{\{b(\hat{\theta}_n) - b_o(\hat{\theta}_n)\}}{n}$$

where $b(\theta)$ and $b_o(\theta)$ are as defined in (4.4.5) and (4.4.6). Then the corrected m.l.e may be taken to be the truncated version of $\hat{\hat{\theta}}_n$. To truncate $\hat{\hat{\theta}}_n$ choose some $d > 0$ such that the true θ may be assumed to lie in $(-d, d)$ and then replace $\hat{\hat{\theta}}_n$ by d or $-d$ according as it exceeds d or falls below d. (The asymptotic theory is insensitive to the choice of d). Let the estimate T_n be truncated in a similar way. Then the mean square error of the truncated $\hat{\theta}_n$ is strictly smaller than that of the truncated T_n if terms of order $o(n^{-2})$ are neglected. This result remains true for quite general loss functions vide Ghosh, Sinha and Wieand (1980).

Here one has a complete sufficient statistic, namely $\Sigma \bar{p}_{in}$, but T_n^* is not a function of it. If one considers the so-called Rao-Blackwellized $T_n' = E(T_n^* | \Sigma \bar{p}_{in})$ then it is indistinguishable from $\hat{\theta}_n$ up to $0_E(1/n)$.

Our second order expansions seemed to agree quite well with the Monte-Carlo values in a few examples of Berkson that we studied. In the examples T_n had lower bias as well as lower variance than $\hat{\theta}_n$ but $b'(\theta_o)$ for T_n was also smaller than the corresponding quantity $b'(\theta_o)$ for $\hat{\theta}_n$. This last fact explains why $\hat{\theta}_n^*$ performs better than T_n^*, since

$$E^U(T_n^*-\theta_o)^2 = E^U(T_n-\theta_o-b(\theta_o)/n)^2 - \frac{2b'(\theta_o)}{n^2 I} + o\,(n^{-2}).$$

Silverstone (1957) and Rao (1960) have defended the use of the maximum likelihood estimates from certain other points of view.

4.5 Bayesian Approach to Second Order Efficiency

In an important pioneering paper on Bayesian analysis, Lindley (1961) has considered an expansion for the aposteriori risk and obtained from it an expansion for a Bayes estimate in powers of $1/n$. In the discussion following Rao (1962) he seeks a Bayesian justification of Rao's (1962) results. Lindley considers a loss function, depending on observations, which is proportional to

$$\lambda(d,\theta) = \left(Z_n - \frac{1}{n}\frac{d \log L}{d\theta}\Big|_d\right)^2 \qquad (4.5.1)$$

and a uniform prior measure. Actually his terminology is slightly different. He considers the product of prior and loss function and calls it a weight function. Lindley shows that $\hat{\theta}_n$ is Bayes up to $o(n^{-1})$ for this prior and loss function. He also claims that the loss function given in (4.5.1) is equivalent to the measure E_2 of Rao and that the Bayes property of $\hat{\theta}_n$ explains its second order efficiency. It seems to us that both these claims are unjustified.

For example, consider, the special case of i.i.d. $N(\theta,1)$ random variables and note that here (4.5.1) reduces to $(d - \theta)^2$. Then, presumably, one would evaluate an efficient estimate T_n by calculating $E_{\theta_o}(T_n - \theta_o)^2$ if the loss function (4.5.1) were used. This seems to have no relation with

$$E_{\theta_o}\{(\bar{X}_n-\theta_o) - (T_n-\theta_o) - \lambda(T_n-\theta_o)^2 - a_\lambda(\theta_o)/n\}^2$$

(where $a_\lambda(\theta_o)$ is defined as in Proposition 4.2.3 and $\bar{X}_n$ is the sample mean) which one to consider for Rao's measure. Moreover for the loss function(4.5.1)the Bayes property of $\hat{\theta}_n$ does not imply that for every efficient estimate T_n,

$$E_{\theta_o}(T_n - \theta_o)^2 \geq E_{\theta_o}(\theta_o - \theta_o)^2 + o_E(n^{-2}) \text{ for all } \theta_o. \tag{4.5.2}$$

In fact it is easy to see that (4.5.2) is false. Note that $\hat{\theta}_n = \bar{x}_n$, the sample mean and so if we take $T_n = \bar{x}_n + b(\bar{x}_n)/n$ where $b(\theta_o)$ and $b'(\theta_o) < 0$, then (4.5.2) is violated. If the prior is the Lebesgue measure, then the approximate Bayes property for $\hat{\theta}_n$ becomes an exact one in the sense $\int [(d-\theta)^2 f(x_1,\theta)\ldots f(x_n,\theta)]\, d\theta$ is minimised at $d = \hat{\theta}_n$. This result is known to be at the root of minimaxity and admissibility of $\hat{\theta}_n$ with respect to the loss (4.5.1) but it cannot imply any uniformly best property like (4.5.2).

The remarks regarding Lindley's loss for the special case considered above are true for the general problem with slight modification but we shall not pursue this matter further. Let us now proceed to show that a Bayesian proof of results on second order efficiency is indeed possible though not on the lines of

Lindley outlined above. Our arguments will be heuristic but for a rigorous treatment, refer to Ghosh, Sinha and Wieand (1980).

We first approach Rao's result. Let $\{X_i\}$ be a sequence of i.i.d. random variables with density $f(x, \theta)$ and the loss function be $(d-\theta)^2$. Let the prior have a density $q(\theta)$ with respect to the Lebesgue measure and suppose $q(\theta)$ is twice continuously differentiable and positive everywhere. Then the Bayes solution is, using Lindley (1961),

$$B_n = \hat{\theta}_n + \frac{1}{2}\frac{L_3}{L_2^2} - \frac{1}{L_2}\frac{q'(\hat{\theta}_n')}{q(\theta_n)} + o_E(n^{-3/2}) \qquad (4.5.3)$$

where

$$L_i = \left.\frac{d^i \log L}{d\theta^i}\right|_{\hat{\theta}_n}.$$

Let

$$B_n' = \hat{\theta}_n + \frac{1}{2}\frac{L_3}{L_2^2} - \frac{1}{L_2}\frac{q'(\hat{\theta}_n)}{q(\hat{\theta}_n)}. \qquad (4.5.4)$$

We can consider B_n' as an approximate Bayes estimate. Then we note that

$$E_\theta(B_n' - \theta)^2 = E_\theta(B_n - \theta)^2 + o\ (n^{-2}).$$

Also from (4.5.4) we have

$$E_\theta(B_n') = \theta + \frac{b_o(\theta)}{n} + \frac{c(\theta)}{n} + o\ (n^{-1})$$

where

$$E_\theta\ (\hat{\theta}_n) = \theta + \frac{b_o(\theta)}{n} + o\ (n^{-1})$$

and

$$E_\theta(B_n' - \hat{\theta}_n) = \frac{c(\theta)}{n} + o(n^{-1}) .$$

Let us assume that $c(\theta)$ is continuously differentiable and consider another estimate

$$B_n'' = \hat{\theta}_n + \frac{c(\hat{\theta}_n)}{n} .$$

Then

$$E_\theta(B_n'' - \hat{\theta}_n) = \frac{c(\theta)}{n} + o(n^{-1}) .$$

Clearly $B_n'' = B_n + o_E(n^{-1})$. Now we will show that this property implies that B_n'' is Bayes up to $o(n^{-2})$ in the sense of (4.5.8) given below. Note that

$$E_\theta(B_n'-\theta)^2 = E_\theta(\hat{\theta}_n-\theta)^2 + \frac{\{c(\theta)\}^2}{n^2} + \frac{2c(\theta)}{n} E_\theta(\hat{\theta}_n - \theta)$$

$$+ 2E_\theta\{(\hat{\theta}_n-\theta)(B_n' - \hat{\theta}_n - c(\theta)/n)\} + o(n^{-2}). \quad (4.5.5)$$

Similarly

$$E_\theta(B_n'' -\theta)^2 = E_\theta(\hat{\theta}_n-\theta)^2 + \frac{\{c(\theta)\}^2}{n^2} + \frac{2c(\theta)}{n} E_\theta(\hat{\theta}_n - \theta)^2$$

$$+ 2E_\theta\{(\hat{\theta}_n - \theta) (B_n'' - \hat{\theta}_n - c(\theta)/n)\} + o(n^{-2}). \quad (4.5.6)$$

But

$$E_\theta(\hat{\theta}_n-\theta)\{B_n'-\hat{\theta}_n-c(\theta)/n\} - E_\theta(\hat{\theta}_n-\theta)\{B_n''-\hat{\theta}_n-c(\theta)/n\}$$

$$= E_\theta\{(\hat{\theta}_n-\theta)(B_n'-B_n'')\} + o(n^{-2})$$

$$= E_\theta\{(Z_n/I)(B_n'-B_n'')\} + o(n^{-2})$$

$$(\text{since } \hat{\theta}_n-\theta-Z_n/I = O_E(n^{-1}))$$

$$= o(n^{-2}). \qquad (4.5.7)$$

The last step above is obtained by differentiating the relation

$$E_\theta(B_n' - B_n'') = o(n^{-1}).$$

It follows now from (4.5.5), (4.5.6) and (4.5.7) that

$$E_\theta(B_n''-\theta)^2 = E_\theta(B_n'-\theta)^2 + o(n^{-2}) = E_\theta(B_n-\theta)^2 + o(n^{-2}). \qquad (4.5.8)$$

We have now arrived at a remarkable fact. From (2.18) of Lindley (1961) we notice that up to $O(n^{-1})$ the posterior depends on $\hat{\theta}_n$, L_2, L_3 and L_4; in a sense, therefore, they are sufficient to $O(n^{-1})$. Nevertheless (4.5.8) show that for the loss function $(d-\theta)^2$ and all smooth priors B_n'' is a Bayes solution to the degree of accuracy specified in (4.5.8). Thus $\hat{\theta}_n$ alone is not sufficient to $O(n^{-1})$ but this Bayes solution B_n'' is a function of $\hat{\theta}_n$ alone. Incidentally, the Bayes property (4.5.8) would hold for any $T_n = B_n'' + o_E(1/n)$.

Consider now any efficient estimate T_n, such that

$$E_\theta(T_n) = \theta + \frac{b(\theta)}{n} .$$

Let

$$T_n' = T_n - \{ b(T_n) - b_o(T_n) - c(T_n) \} /n .$$

Then some easy calculations lead to

$$E_\theta(T_n'-\theta)^2 - E_\theta(B_n''-\theta)^2 = E_\theta(T_n^*-\theta)^2 - E_\theta(\hat{\theta}_n^*-\theta)^2 + o\,(n^{-2}) \tag{4.5.9}$$

where $T_n^* = T_n - \frac{b(T_n)}{n}$ and $\theta_n^* = \hat{\theta}_n - \frac{b_o(\hat{\theta}_n)}{n}$. Thus using (4.5.8), (4.5.9) and the definition of B_n'', we get

$$\int E_\theta(T_n^*-\theta)^2\, q(\theta)d\theta \geq \int E_\theta(\hat{\theta}_n^*-\theta)^2\, q(\theta)d\theta + o\,(n^{-2}). \tag{4.5.10}$$

Since (4.5.10) is true for all q, Rao's result follows.

We now turn to Rao-Fisher result. Consider a fixed λ and a fixed efficient estimate T_n. Let $a_\lambda(\theta)$ be such that

$$E_\theta \{ Z_n - (T_n-\theta)I - \lambda(T_n-\theta)^2 \} = a_\lambda(\theta)/n + o\,(n^{-1}). \tag{4.5.11}$$

Let $a_{o\lambda}(\theta)$ be defined similarly for $\hat{\theta}_n$.

Consider the loss function

$$\ell(d,\theta) = \{ Z_n - (d-\theta)I - \lambda Z_n^2 /I^2 - a_\lambda(\theta)/n \}^2 \tag{4.5.12}$$

and a prior $q(\theta)$ satisfying the same restrictions as above. One can show as before that an estimate of the form $B_n'' = \hat{\theta}_n + c(\hat{\theta}_n)/n$ satisfies

$$E_\theta\{ \lambda (B_n'', \theta)\} = E_\theta\{ \lambda (B_n, \theta)\} + o(n^{-2}) \qquad (4.5.13)$$

where B_n is the Bayes solution for the loss function given in (4.5.12). Now it is easy to show that

$$E_\theta \{ Z_n - (\hat{\theta}_n - \theta)I - \lambda(\hat{\theta}_n - \theta)^2 - a_{o\lambda}(\theta)/n\}^2 = E_\theta\{ \lambda (B_n'', \theta)\} + o(n^{-2}) \qquad (4.5.14)$$

Also

$$E_\theta \{ Z_n - (T_n - \theta)I - \lambda(T_n - \theta)^2 - a_\lambda(\theta)/n\}^2 = E_\theta\{ \lambda (T_n, \theta)\} + o(n^{-2}). \qquad (4.5.15)$$

Since q is arbitrary, we get from (4.5.13), (4.5.14) and (4.5.15) that

$$E_2(\{ T_n\}, \theta, \lambda) \geq E_2(\{ \hat{\theta}_n\}, \theta, \lambda) \quad \text{for all } \theta$$

which gives the Rao-Fisher result.

To justify these heuristic arguments one would of course need various restrictions on $f(x,\theta)$ and T_n but one would expect that the restrictions would be nuch milder than those considered in Section 4.2.

APPENDIX

Proof of Proposition 4.2.3 :

(i) By Lemma 4.2.7,

$$a_\lambda(\theta_o) = \lim_{n\to\infty} E^U n\{ Z_n-(T_n-\theta_o)I - \lambda(T_n-\theta_o)^2\}$$

$$= \lim_{n\to\infty} n E^U\{-\tfrac{1}{2}\Sigma\Sigma T^{ij}(\bar{p}_{in}-\pi_i(\theta_o))(\bar{p}_{jn}-\pi_j(\theta_o))$$

$$- \lambda Z_n^2/I^2\}$$

$$= \lim_{n\to\infty} n E_{\theta_o}\{-\tfrac{1}{2}\Sigma\Sigma T^{ij}(\bar{p}_{in}-\pi_i(\theta_o))(\bar{p}_{jn}-\pi_j(\theta_o))$$

$$- \lambda Z_n^2/I^2\} \quad \text{(by Lemma 4.2.2)}$$

$$= -\tfrac{1}{2}\Sigma\Sigma T^{ij} E_{\theta_o}(p_{i1}-\pi_i(\theta_o))(p_{j1}-\pi_j(\theta_o)-\lambda/I$$

which is seen to exist.

(ii) By Lemma 4.2.7,

$$E_2(\{T_n\}, \theta_o, \lambda, U\} = \lim_{n\to\infty} n^2 E^U \{Z_n - (T_n - \theta_o)I -$$

$$\lambda(T_n-\theta_o)^2 - a_\lambda(\theta_o)/n\}^2$$

$$= \lim_{n\to\infty} n^2 E^U \{-\tfrac{1}{2}\Sigma\Sigma T^{ij}(\bar{p}_{in}-\pi_i(\theta_o))(\bar{p}_{jn}-\pi_j(\theta_o)) -$$

$$\lambda Z_n^2/I^2 - a_\lambda(\theta_o)/n\}^2$$

$$= \lim_{n\to\infty} n^2 E_{\theta_o}\{-\tfrac{1}{2}\Sigma\Sigma T^{ij}(\bar{p}_{in}-\pi_i(\theta_o))(\bar{p}_{jn}-\pi_j(\theta_o)) -$$

$$\lambda Z_n^2/I^2 - a_\lambda(\theta_o)/n\}^2 \quad \text{(by Lemma 4.2.2)}$$

which is easily seen to exist.

(iii) The required result follows since the limit obtained in the proof of (ii) does not depend on U.

Proof of Proposition 4.2.4 :

(i) By Lemma 4.2.7 and Lemma 4.2.2,

$$b(\theta_o) = \frac{1}{2} \Sigma \Sigma T^{ij} E_{\theta_o}(p_{i1}-\pi_i(\theta_o))(p_{j1}-\pi_j(\theta_o)).$$

(ii) $b(\theta)$ is continuously differentiable if T^{ij}'s are, since

$E_\theta(p_{i1}-\pi_i(\theta_o))(p_{j1}-\pi_j(\theta_o))$ is differentiable.

(iii) Since $b(\theta)$ is continuously differentiable, we have

$$b(T_n) = b(\theta_o) + (T_n-\theta_o)\frac{b'(\theta_o)}{n} + o_E(n^{-1/2}).$$

Now (iii) follows immediately from Lemma 4.2.7.

(iv) Proceeding on the same lines as in the proof of Proposition 4.2.3(ii) we can show that

$$\psi(\{T_n'\},\theta_o) = \lim_{n\to\infty} n^2\{E^U(T_n' - \theta_o)^2 - 1/nI\}$$

exists. However one needs an analogue of Lemma 4.2.8 in addition to Lemmas 4.2.2 and 4.2.6. Details are omitted.

(v) We can choose U_1 so small that

$$(T_n' - \theta_o)^2 < d \qquad \text{if } \bar{p}_n \in U_1 .$$

Then

$$|E_{\theta_o}\{W(T_n',\theta_o)\} - E^{U_1}(T_n'-\theta_o)^2| \leq d.\ P_{\theta_o}\{\bar{p}_n \in U_1^c\} = o\ (n^{-2})$$

by Lemma 4.2.2. But

$$|E^U(T_n' - \theta_o)^2 - E^{U_1}(T_n' - \theta_o)^2| = o\ (n^{-2}) \text{ by (iv).}$$

This completes the proof.

CHAPTER 5

EDGEWORTH EXPANSIONS FOR LOCALLY STABLE ESTIMATES AND SECOND ORDER EFFICIENCY

5.1 Introduction

As in the previous chapter, suppose we have a curved exponential family and assume it is dominated by the Lebesgue measure (or slightly more generally, by a measure with an absolutely continuous component). Consider estimates T_n which are l.s.(IV) and efficient, described in Section 4.2 . Then using a result of Bhattacharya and Ghosh (1978) one can obtain under suitable regularity conditions an Edgeworth expansion for the normalised variables $\sqrt{n}\,(T_n - \theta)\sqrt{I(\theta)}$ which is valid up to $o(n^{-1})$, uniformly on compact θ - sets. Here $I(\theta)$ stands for Fisher Information in a single observation. It was shown in Bhattacharya and Ghosh (1978), (Remark 1.4) that the moments of $\sqrt{n}(T_n-\theta)\sqrt{I(\theta)}$ obtained by the so called delta method agree with the moments of the Edgeworth expansion up to $o(n^{-1})$. We shall refer to these two results as Proposition 5.1 and assume below sufficient conditions for it to hold with $s = 4$.

Let T_n be l.s.(IV) and efficient, $\theta_o \in$ the parameter space $\overline{H}$ and $(\overline{H})_o$ a compact neighbourhood of θ_o. Then for $\theta \in (\overline{H})_o$, the Edgeworth expansion up to $o(n^{-1})$ for $\sqrt{n}\,(T_n - \theta)\sqrt{I(\theta)}$ may be expressed in the following form :

$$P_\theta\{\sqrt{n}\,(T_n-\theta)\sqrt{I(\theta)} \le x\} = \Phi(x) + \frac{\phi_{1,x}(\theta)}{\sqrt{n}} + \frac{\phi_{2,x}(\theta)}{n} + o(n^{-1}) \qquad (5.1.1)$$

where

$$\Phi(x) = \int_{-\infty}^{x} e^{-t^2/2}\, dt/\sqrt{2\pi}\ , \quad \phi(x) = \Phi'(x),$$

$$\phi_{1,x}(\theta) = \int_{-\infty}^{x} \{K_{11}(\theta)\, H_1(z) + \frac{K_{21}(\theta)}{2} H_2(z) + \frac{K_{31}(\theta)}{6} H_3(z)\}\, \phi(z)dz\ , \qquad (5.1.2)$$

$$\phi_{2,x}(\theta) = \int_{-\infty}^{x} [K_{12}(\theta)\, H_1(z) + \frac{K_{22}(\theta)}{2} H_2(z) + \frac{K_{32}(\theta)}{6} H_3(z) +$$

$$+ \frac{K_{41}(\theta)}{24} H_4(z) + \frac{1}{2}\{K_{11}^2(\theta)\, H_2(z) + \frac{K_{21}^2(\theta)}{4} H_4(z) +$$

$$+ \frac{K_{31}^2(\theta)}{36} H_6(z) + K_{11}(\theta)\, K_{21}(\theta)\, H_3(z) +$$

$$+ \frac{1}{3}K_{11}(\theta)K_{31}(\theta)H_4(z) + \frac{1}{6}K_{21}(\theta)K_{31}(\theta)H_5(z)\}]\ \phi(z)dz,$$

$$H_p(x)\,\phi(x) = (-\frac{d}{dx})^p\,\phi(x) \qquad (5.1.3)$$

and the terms $o(n^{-1})$ in (5.1.1) is uniform in $\theta \varepsilon \textcircled{H}_o$ and x, the K_{ij}'s are as defined in Remark 5.1.1 below.

<u>Remark 5.1.1</u>. The various $K_{ij}(\theta)$'s appearing in the Edgeworth expansion given above can be obtained by the delta method as follows. The estimate T_n is expanded in a Taylor's series as given in Lemma 4.2.7, yielding an expansion of $\sqrt{n}\,(T_n-\theta)\sqrt{I(\theta)}$; this expansion is then raised to a positive integral power say

' r ', expectations are taken term by term and terms of order $o(n^{-1})$ are neglected. By this we get the rth(formal)moments $\mu_r(\theta, T_n) = E_\theta\{\sqrt{n}(T_n - \theta)\sqrt{I(\theta)}\}^r$ of $\sqrt{n}(T_n - \theta)\sqrt{I(\theta)}$ up to $o(n^{-1})$. (It is easy to see that $\mu_r(\theta, T_n) = E^U\{\sqrt{n}(T_n - \theta)\sqrt{I(\theta)}\}^r + o(n^{-1})$; vide Remark 5.1.4). Finally cumulants $K_r(\theta, T_n)$ are calculated from these (formal) moments by standard formulae (Kendall (1952), p. 63) neglecting the terms of order $o(n^{-1})$. For example,

$$K_2(\theta, T_n) = \mu_2(\theta, T_n)$$

$$K_3(\theta, T_n) = \mu_3(\theta, T_n)$$

$$K_4(\theta, T_n) = \mu_4(\theta, T_n) - 3\mu_2^2(\theta, T_n) \quad (\text{up to } o(n^{-1})).$$

Let $K_{ri}(\theta)$, $i = 0,1,2$ denote the coefficient of $n^{-i/2}$ in $K_r(\theta, T_n)$. Then the following relation is valid.

$$\begin{aligned} K_r(\theta, T_n) &= \frac{K_{r1}(\theta)}{\sqrt{n}} + \frac{K_{r2}(\theta)}{n}, && r = 1,3 \\ &= 1 + \frac{K_{r1}(\theta)}{\sqrt{n}} + \frac{K_{r2}(\theta)}{n}, && r = 2 \\ &= \frac{K_{r2}(\theta)}{n}, && r = 4 \\ &= 0, && r \geq 5. \end{aligned} \qquad (5.1.4)$$

Suppose $b(\theta)$ is twice continuously differentiable in $(\bar{H})_o$. Define $T_n^* = T_n - \frac{b(T_n)}{n}$. Then, as proved in Section 5.2 (vide Property 5.3.5), $\sqrt{n}\,(T_n^* - \theta)\,\sqrt{I(\theta)}$ has also an Edgeworth expansion up to $o(n^{-1})$ which is similar to (5.1.1) except that $K_{11}(\theta)$ and $K_{22}(\theta)$ in (5.1.2) and (5.1.3) are to be replaced by

$$\text{(a)}\quad K_{11}^*(\theta) = K_{11}(\theta) - b(\theta)\sqrt{I(\theta)}$$

and

$$\text{(b)}\quad K_{22}^*(\theta) = K_{22}(\theta) - 2b'(\theta), \tag{5.1.5}$$

respectively. Moreover, one can use the delta method to calculate the formal r-th moments

$$\mu_2^*(\theta, T_n^*) = E_\theta(\sqrt{n}\,(T_n^* - \theta)\sqrt{I(\theta)})^r = \mu_r^*(\theta) \quad \text{(say)}$$

and hence the formal cumulants $K_r(\theta, T_n^*) = K_r^*(\theta)$ (say) of $\sqrt{n}\,(T_n^* - \theta)\,\sqrt{I(\theta)}$, and K_{rj}^*'s from a relation analogous to (5.1.4).

Suppose $\hat{\theta}_n$ is the m.l.e of θ. Then $\hat{\theta}_n$ is l.s(IV) (vide Section 4.2). Hence Edgeworth expansion and the remark 5.1.1 apply if we specialize T_n to $\hat{\theta}_n$. In this case b will be denoted by b_o and we will write $\hat{\theta}_n^* = \hat{\theta}_n - \frac{b_o(\hat{\theta}_n)}{n}$. Also the (formal) moments (cumulants) of $\sqrt{n}\,(\hat{\theta}_n - \theta)\,\sqrt{I(\theta)}$ will be denoted by $\bar{\mu}_r(\theta)$ $(\bar{K}_r(\theta))$, those of $\sqrt{n}\,(\hat{\theta}_n^* - \theta)\,\sqrt{I(\theta)}$ by

$\bar{\mu}_r^*(\theta)(\bar{K}_r^*(\theta))$ and the quantities analogous to K_{rj}'s (K_{rj}^*'s) by $\bar{K}_{rj}$'s ($\bar{K}_{rj}^*$'s). From (5.1.5) (a) it follows that one can always choose b and b_o as

$$b(\theta) = K_{11}(\theta) / \sqrt{I(\theta)} \quad \text{and} \quad b_o(\theta) = \bar{K}_{11}(\theta)/\sqrt{I(\theta)}$$

which will imply

$$K_{11}^*(\theta) = \bar{K}_{11}^*(\theta)=0 \ \forall\ \theta \in (\bar{H})_o . \tag{5.1.6}$$

The corresponding T_n^* and $\hat{\theta}_n^*$ will thus be "unbiased" up to $o(n^{-1})$. In what follows we assume that b and b_o are chosen so that (5.1.6) is satisfied. This definition is identical to that given in Chapter 4, vide Remark 5.1.4. Our main result can then be stated as follows.

<u>Theorem 5.1.2</u>: Let T_n be l.s (IV) and efficient, $\hat{\theta}_n$ be the m.l.e and b and b_o be chosen so that (5.1.6) holds. Then

$$K_1^*(\theta) = \bar{K}_1^*(\theta) = 0 \tag{5.1.7}$$

$$K_2^*(\theta) > \bar{K}_2^*(\theta) \tag{5.1.8}$$

$$K_3^*(\theta) = \bar{K}_3^*(\theta) \tag{5.1.9}$$

$$K_4^*(\theta) = \bar{K}_4^*(\theta) \tag{5.1.10}$$

for all $\theta \in (\bar{H})_o$.

In this connection it is worth remarking that (5.1.7) follows immediately from the definition of b and b_o, and (5.1.8) is essentially a reformulation of the main result (Theorem 4.2.6 (iii) of Chapter 4). This will be clear from the proof. Hence the really new facts are (5.1.9) and (5.1.10). From Theorem 5.1.2 we derive the following

Corollary 5.1.3 :

$$P_\theta \{- x_1 \leq \sqrt{n}\, (\hat{\theta}_n^* - \theta)\, \sqrt{I(\theta)} \leq x_2\}$$

$$\geq P_\theta \{- x_1 \leq \sqrt{n}\, (T_n^* - \theta)\, \sqrt{I(\theta)} \leq x_2\} + o(n^{-1})$$

for all $\theta \in (\bar{H})_o$ and all $x_1, x_2 \geq 0$ (at least one of x_1, x_2 being positive); the term $o(n^{-1})$ is of smaller order than n^{-1} uniformly in x_1, x_2.

This immediately implies the second order efficiency of the m.l.e with respect to any bounded loss function $L_n(a, \theta) = h(\sqrt{n}\,(a - \theta))$ which is bowlshaped (i.e., whose minimum value is zero at $a - \theta = 0$ and which increases as $|a - \theta|$ increases) i.e., the following inequality holds :

$$E_\theta \{ L_n(T_n^*, \theta)\} \geq E_\theta \{ L_n(\hat{\theta}_n^*, \theta)\} + o(n^{-1}), \qquad (5.1.11)$$

for all $\theta \in (\bar{H})_o$.

Remark 5.1.4. Note that

$$E^U(\sqrt{n}\,(T_n-\theta_o)\sqrt{I(\theta_o)})^r = \mu_r(\theta_o, T_n) + o(n^{-1}).$$

For, by (4.2.29),

$$E^U(\sqrt{n}(T_n-\theta_o)\sqrt{I(\theta_o)})^r = E^U\{\sqrt{n}(T_{n1}+T_{n2}+T_{n3})\sqrt{I(\theta_o)}\}^r + o(n^{-1})$$

$$= E_{\theta_o}\{\sqrt{n}(T_{n1}+T_{n2}+T_{n3})\sqrt{I(\theta_o)}\}^r + o(n^{-1})$$

(by Lemma 4.2.1)

$$= \mu_r(\theta_o, T_n) + o(n^{-1})$$

by definition of μ_r. Hence in particular the $b(\theta_o)$ and $b_o(\theta_o)$ of Theorem 5.1.1 satisfy

$$b(\theta_o) = K_{11}(\theta_o)/\sqrt{I(\theta_o)}$$

and

$$b_o(\theta_o) = \bar{K}_{11}(\theta_o)/\sqrt{I(\theta_o)}$$

i.e., $b(\theta_o)$ and $b_o(\theta_o)$ would be alternatively defined as in Chapter 4, namely

$$b(\theta_o) = \lim n\,E^U(T_n-\theta_o) \text{ and } b_o(\theta_o) = \lim n\,E^U(\hat{\theta}_n-\theta_o).$$

An indirect proof of Corollary (5.1.3) (but not of the theorem) or rather a different version of it appears in Ghosh, Sinha and Wieand (1980) (vide Remarks 3.1 and 3.2). Similar results are available in Pfanzagl (1975), Pfanzagl and Wefelmeyer (1978) and Takeuchi and Akahira (1978). (Our results appear in Ghosh, Sinha and Subramanyam (1979); they were obtained independently by the first two authors in the first half of 1977 and by the third author a little later).

So far we have the set up of a curved exponential family and the assumption that the dominating measure has an absolutely continuous component. If in the above set up we drop this assumption, the formal Edgeworth expansions are no longer valid. This does not affect the theorem but does affect Corollary 5.1.3 and hence the inequality (5.1.11) is no longer true. However it turns out (vide Corollary 5.4.2) that if the loss function satisfies certain additional conditions (vide (5.4.2) and (5.4.5)) the inequality (5.1.11) remains valid even when the dominating measure does not have an absolutely continuous component. This modification takes care of the curved multinomial for which Corollary 5.1.3 fails.

The discerning author may have guessed by now why second order efficiency has also been called third order efficiency. The Edgeworth expansions of $\sqrt{n}\,(T_n^* - \theta_o)$ is a series in powers of $n^{-1/2}$ whereas $E^U\, n(T_n^* - \theta_o)^2$ is a series in power of n^{-1}. To get the latter up to $o(n^{-1})$ from the Edgeworth expansions one needs three terms. Thus if one thinks of (5.1.1) it is natural to follow Takeuchi and Pfanzagl and call it the third order efficiency of the m.l.e. On the other hand when one's primary interest is in $E^U n(T_n^* - \theta_o)^2 \geq E^U n(\hat{\theta}_n^* - \theta_o)^2 + o(n^{-1})$ it is equally natural to refer to it as second order efficiency.

Note that Corollary 5.1.3 evidently remains true if we replace $o(n^{-1})$ by $o(n^{-1/2})$. Results of this type have been called second order efficiency by Akahira and Takeuchi(1976a,1976b) and Pfanzagl(1973). (To use Rao's scale for efficiency, one should call it the efficiency of order 3/2). In the present context of l.s.(IV) efficient estimates Corollary 5.1.3 with $o(n^{-1/2})$ in place of $o(n^{-1})$ turns out be be uninteresting, as pointed out by Ghosh and Subramanyam (1974). For, the proof of Theorem 5.1.1 shows that the inequality in Corollary 5.1.3 is an equality up to $o(n^{-1/2})$ and hence up to $o(n^{-1/2})$ it offers no discrimination among l.s.(IV) efficient estimates.

In Section 5.2, we adopt the notations, terminology and assumptions of Chapter 4. In Chapter 4 it suffices to assume that T is thrice continuously differentiable. We need one more derivative in order to get Proposition 5.2.1. In this section, the proof of the theorem and Corollary 5.1.3 are also given. The statement and the proof of Corollary 5.3.4 apear in Section 5.3. We consider a numerical example in the final section.

5.2 Proof of the theorem

Our set up in this section is the same as in Chapter 4. Our assumptions are the following :

The dominating measure μ has an absolutely continuous component. (5.2.1)

With $p_o(x) = 1$, $p_o(x), p_1(x), \dots, p_k(x)$ are linearly independent a.e(μ). (5.2.2)

Assumption (I) of Section 4.2. (5.2.3)

T_n is l.s(IV) (See definition 4.2.1) (5.2.4)

The following proposition is an immediate consequence of Theorem 2 of Bhattacharya and Ghosh (1978).

Proposition 5.2.1. Suppose conditions (5.2.1) to (5.2.4) hold. Then (5.1.1) holds.

Remark 5.2.2 : The following result extends Remark 5.1.4 :

$$E^U_{\theta_o}\{(T_n-\theta_o)^r Z_n^s W_n^t\} n^{(r+s+t)/2}$$

$$= E_{\theta_o}\{(T_{n1}+T_{n2}+T_{n3})^r Z_n^s W_n^t\} n^{(r+s+t)/2} + o(n^{-1}).$$

This follows immediately from Lemma 4.2.1 and Lemma 4.2.7 noting that

$$(T_n-\theta_o)^r - (T_{n1} + T_{n2} + T_{n3})^r = o_E(n^{-3r/2})$$

Z_n and W_n are each $o_E(n^{-1/2})$.

As in Lemma 4.2.7 the quantities $\hat{\theta}_{n_i}$'s, T^*_{ni}'s and $\hat{\theta}^*_{ni}$'s relating to the expansion of $\hat{\theta}_n - \theta_o$, $T^*_n - \theta_o$ and $\hat{\theta}^*_n - \theta_o$ respectively (similar to $T_n-\theta_o = T_{n1}+T_{n2}+T_{n3}+o_E(n^{-3/2})$) can be defined analogously.

We state below a set of properties whose applications provide a direct proof of the theorem. Proofs are given whenever necessary.

Let

$$P_1(\bar{p}_{1n},\ldots,\bar{p}_{kn}) \text{ and } P_2(\bar{p}_{1n},\ldots,\bar{p}_{kn})$$

be two homogeneous polynomials in $\bar{p}_{1n},\ldots,\bar{p}_{kn}$ of degree r and s respectively.

<u>Property 5.2.3.</u> The bivariate moments μ_{k_1,k_2} of $(\sqrt{n})^r P_1$ and $(\sqrt{n})^s P_2$ have expansions in powers of n^{-1}.

<u>Property 5.2.4.</u> The leading term of μ_{k_1,k_2} can be obtained by assuming $\bar{p}_{1n},\ldots,\bar{p}_{kn}$ to have a multinormal distribution with means $\pi_1,\ldots,\pi_k$ and dispersion matrix Σ/n where Σ is as defined in (4.2.3).

<u>Property 5.2.5.</u> $Z_n = \sum_{i=1}^{k} \beta_i'(\theta)(\bar{p}_{in} - \pi_i(\theta))$ and

$$Q_n = \sum_i \sum_j (T^{ij} - \hat{\theta}^{ij})(\bar{p}_{in} - \pi_i(\theta))(\bar{p}_{jn} - \pi_j(\theta))$$

are independent when $\bar{p}_{1n},\ldots,\bar{p}_{kn}$ have a multinormal distribution.

<u>Proof</u>. By a standard theorem [Searle (1971), page 59], it is enough to show

$$(\Delta^{ij}) \Sigma \underline{\beta}' = \underline{0} \tag{5.2.5}$$

where $\Delta^{ij} = T^{ij} - \hat{\theta}_n^{ij}$ and $\underline{\beta}' = (\beta_1'(\theta),\ldots,\beta_k'(\theta))'$.

From (4.2.11), upon differentiating both sides with respect to θ and noting that $\Delta^{ij} = \Delta^{ji}$ we get

$$\sum_{i=1}^{k} \Delta^{ij} \pi_i'(\theta) = 0 \quad \text{for all} \quad i = 1,2,\ldots,k \,. \tag{5.2.6}$$

On the other hand, direct calculations show

$$\pi_i'(\theta) = \sum_{j=1}^{k} \frac{\partial \pi_i(\theta)}{\partial \beta_j} \beta_j'(\theta)$$

and

$$\frac{\partial \pi_i(\theta)}{\partial \beta_j} = \text{cov}[p_i(x), p_j(x)], \quad i,j = 1,2,\ldots,k,$$

giving $(\pi_1'(\theta),\ldots, \pi_k'(\theta))' = \Sigma \underline{\beta}'$ which together with (5.2.6) proves (5.2.5).

Property 5.2.6. Any odd order multivariate central moment of $\bar{p}_{1n},\ldots,\bar{p}_{kn}$ is zero whenever $\bar{p}_{1n},\ldots,\bar{p}_{kn}$ have a multinormal distribution.

Property 5.2.7. Let $T_n^* = T_n - b(T_n)/n$ where T_n is l.s.(IV) and efficient and $b(\cdot)$ is a twice differentiable function. For $\theta \in (\bar{H})_0$, $P_\theta\{\sqrt{n}(T_n^*-\theta)\sqrt{I(\theta)} \leq x\}$ is given by (5.1.1) with $\phi_{1,x}(\theta)$, $\phi_{2,x}(\theta)$ as given in (5.1.2) and (5.1.3) except that $K_{11}(\theta)$ and $K_{22}(\theta)$ are to be replaced by $K_{11}^*(\theta) = K_{11}(\theta) - b(\theta)\sqrt{I(\theta)}$ and $K_{22}^*(\theta) = K_{22}(\theta) - 2b'(\theta)$ respectively.

Proof. The proof appears in Ghosh, Sinha and Wieand (1980). However, for the sake of completeness, we give it here.

$$P_\theta\{\sqrt{n}\,(T_n^* - \theta)\,\sqrt{I(\theta)} \leq x\}$$

$$= P_\theta\{\sqrt{n}\,(T_n-\theta-\frac{b(\theta)}{n} - \frac{(T_n-\theta)b'(\theta)}{n} - \frac{(T_n-\theta)^2}{2!}\frac{b''(\bar{\theta})}{n})\sqrt{I(\theta)} \leq x\}$$

where $\bar{\theta} \in (T_n,\theta)$

$$= P_\theta\{\sqrt{n}\,(T_n-\theta)(1-\frac{b'(\theta)}{n} - \frac{(T_n-\theta)b''(\bar{\theta})}{2n})\sqrt{I(\theta)} \leq x + \frac{b(\theta)\sqrt{I(\theta)}}{\sqrt{n}}\}$$

$$= P_\theta\{\sqrt{n}\,(T_n-\theta)\sqrt{I(\theta)} \leq (x + \frac{b(\theta)\sqrt{I(\theta)}}{\sqrt{n}})(1-\frac{b'(\theta)}{n} - \frac{(T_n-\theta)b''(\bar{\theta})}{2n})^{-1}\}$$

$$= P_\theta \{\sqrt{n}\,(T_n-\theta)\sqrt{I(\theta)} \le (x + \frac{b(\theta)\sqrt{I(\theta)}}{\sqrt{n}})(1 - \frac{b'(\theta)}{n} - \frac{(T_n-\theta)b''(\bar\theta)}{2n})^{-1},$$

$$|T_n - \theta| \le \frac{c \log n}{\sqrt{n}}\} + o(n^{-1})$$

$$= P_\theta \{\sqrt{n}\,(T_n-\theta)\sqrt{I(\theta)} \le x + \frac{xb'(\theta)}{n} +$$

$$\frac{b(\theta)\sqrt{I(\theta)}}{\sqrt{n}} + o(n^{-1})\} + o(n^{-1})$$

$$= P_\theta \{\sqrt{n}\,(T_n-\theta)\sqrt{I(\theta)} \le x + \frac{b(\theta)\sqrt{I(\theta)}}{\sqrt{n}} + \frac{x\,b'(\theta)}{n}\} + o(n^{-1})$$

$$= \Phi(x + \frac{b(\theta)\sqrt{I(\theta)}}{\sqrt{n}} + \frac{x\,b'(\theta)}{n}) + \frac{\phi_{1,x+b(\theta)\sqrt{I(\theta)}/\sqrt{n}+x\,b'(\theta)/n}(\theta)}{\sqrt{n}} +$$

$$+ \frac{\phi_{2,x+b(\theta)\sqrt{I(\theta)}/\sqrt{n}+xb'(\theta)/n}(\theta)}{n} + o(n^{-1}).$$

After straightforward simplication this can be written as

$$= \Phi(x) + \frac{1}{\sqrt{n}} \int_{-\infty}^{x} \{(K_{11}(\theta)-b(\theta)\sqrt{I(\theta)})\,H_1(z) + \frac{K_{21}(\theta)}{2} H_2(z)$$

$$+ \frac{K_{31}(\theta)}{6} H_3(z)\}\,\phi(z)dz + \frac{1}{n} \int_{-\infty}^{x} [K_{12}(\theta)\,H_1(z) +$$

$$+ \frac{K_{22}(\theta)-2b'(\theta)}{2} H_2(z) + \frac{K_{32}(\theta)}{6} H_3(z) + \frac{K_{41}(\theta)}{24} H_4(z)$$

$$+ \frac{1}{2} \{(K_{11}(\theta)-b(\theta)\sqrt{I(\theta)})^2 H_2(z) + \frac{K_{21}^2(\theta)}{4} H_4(z) +$$

$$+ \frac{K_{31}^2(\theta)}{36} H_6(z) + (K_{11}(\theta) - b(\theta)\sqrt{I(\theta)})\, K_{21}(\theta)\, H_3(z) +$$

$$+ \frac{1}{3}(K_{11}(\theta) - b(\theta)\sqrt{I(\theta)})\, K_{31}(\theta)\, H_4(z) + \frac{1}{6} K_{21}(\theta)\, K_{31}(\theta)\, H_5(z)\}]$$

$$\times\ \phi(z)\, dz + o(n^{-1}),$$

which proves the property.

<u>Proof of the theorem</u> : We now calculate K_{ij}'s and the analogous quantities appearing in the Theorem 5.1.2, using Remarks 5.1.1 and 5.2.2 and Properties 5.2.3 to 5.2.7.

Let $\theta_o \in (\bar{H})_o$ be arbitrarily chosen and fixed. Using (4.2.29) and (4.2.25), it is easy to show that

$$\bar{K}_{12}(\theta_o) = K_{12}(\theta_o) = 0 \tag{5.2.7}$$

which along with (5.1.6) establishes (5.1.7) for $\theta = \theta_o$.

Again, from (4.2.26), it follows immediately that

$$\mu_2^*(\theta_o) = \bar{\mu}_2^*(\theta_o) + E^U_{\theta_o}(n\, R_n^{*2})\, I(\theta_o) \tag{5.2.8}$$

which is (5.1.8) for $\theta = \theta_o$ because $K_2^* = \mu_2^*$ and $\bar{K}_2^* = \bar{\mu}_2^*$.

To prove (5.1.9) for $\theta = \theta_o$, note from (4.2.26) that

$$\mu_3^*(\theta_o) = \bar{\mu}_3^*(\theta_o) + 3\{I(\theta_o)\}^{3/2}\, E_{\theta_o}[\{\sqrt{n}(\hat{\theta}^*_{n1}+\hat{\theta}^*_{n2}+\hat{\theta}^*_{n3})\}^2(\sqrt{n}\, R_n^*)]$$

$$+ 3E_{\theta_o}[\{\sqrt{n}\,(\hat{\theta}^*_{n1}+\hat{\theta}^*_{n2}+\hat{\theta}^*_{n3})\}\,(\sqrt{n}\, R_n^*)^2]\,\{I(\theta_o)\}^{3/2}. \tag{5.2.9}$$

Now

$$E_{\theta_o}[\{\sqrt{n}(\hat{\theta}^*_{n1}+\hat{\theta}^*_{n2}+\hat{\theta}^*_{n3})\}^2 (\sqrt{n}\ R^*_n)]$$

$$= E_{\theta_o}[\{\sqrt{n}(\hat{\theta}^*_{n1}+\hat{\theta}^*_{n2}+\hat{\theta}^*_{n3})\}^2 (\sqrt{n}\ .(T^*_n-\hat{\theta}^*_n)\}] \quad \text{using (4.2.25)}$$

$$= E_{\theta_o}[\{(\frac{\sqrt{n}\ Z_n}{I(\theta_o)})^2 + 2\frac{\sqrt{n}\ Z_n}{I(\theta_o)} \cdot o_E(n^{-1/2})\} \cdot$$

$$\{\frac{\sqrt{n}\ Z_n}{I(\theta_o)} \cdot (\frac{b_o'(\theta_o)-b'(\theta_o)}{n}) + \frac{b_o(\theta_o)-b(\theta_o)}{n} +$$

$$+\frac{1}{2}\sqrt{n}\ Q_n + \frac{1}{6}\sqrt{n}\ (T_{n3} - \hat{\theta}_{n3}) + o\ (n^{-1})\}] \quad \text{by (4.2.29)}$$

$$= \frac{1}{2} E_{\theta_o} \{(\frac{\sqrt{n}\ Z_n}{I(\theta_o)})^2 \sqrt{n}\ Q_n\} + o_E(n^{-1}) \quad \text{by Property 5.2.6 .}$$

$$= \frac{n^{3/2}}{2I^2(\theta_o)} E_{\theta_o}(Z_n^2\ Q_n) + o_E(n^{-1})$$

$= o\ (n^{-1})$ since $E_{\theta_o}(Z_n^2\ R^*_n) = o_E(n^{-2})$ by (4.2.25), which implies $E_{\theta_o}(Z_n^2\ Q_n) = o_E(n^{-2})$, yielding $E_{\theta_o}(Z_n^2\ Q_n) = 0\ (n^{-3})$ by Property 5.2.3.

Again,

$$E_{\theta_o}[\sqrt{n}\ (\hat{\theta}^*_{n1} + \hat{\theta}^*_{n2} + \hat{\theta}^*_{n3})\ (\sqrt{n}\ R^*_n)^2]$$

$$= E_{\theta_o}[\sqrt{n}\ (\hat{\theta}^*_{n1}+\hat{\theta}^*_{n2}+\hat{\theta}^*_{n3})\ \{\sqrt{n}(T^*_n-\hat{\theta}^*_n)\}^2] + o(n^{-1})$$

$$= E_{\theta_o}\left[\left(\frac{\sqrt{n}\, Z_n}{I(\theta_o)}\right)\left\{\frac{\sqrt{n}\, Z_n}{I(\theta_o)}\left(\frac{b_o'(\theta_o)-b'(\theta_o)}{n}\right) + \frac{(b_o(\theta_o)-b(\theta_o))}{\sqrt{n}}\right.\right.$$

$$\left.\left. + \frac{1}{2}\sqrt{n}\quad Q_n + \frac{\sqrt{n}}{6}(T_{n3} - \hat{\theta}_{n3})\right\}^2\right] + o\ (n^{-1})$$

$$= E_{\theta_o}\left[\left(\frac{\sqrt{n}\, Z_n}{I(\theta_o)}\right)\left(\tfrac{1}{4}n\ \ Q_n^2\right)\right] + o\ (n^{-1})$$

$= o\ (n^{-1})$ by Properties 5.2.4 and 5.2.5, establishing (5.1.9) for $\theta = \theta_o$ because $\mu_3^* = K_3^*$ and $\bar{\mu}_3^* = \bar{K}_3^*$.

To prove (5.1.10) for $\theta = \theta_o$, note again from (4.2.25) that

$$\mu_4^*(\theta_o) = \bar{\mu}_4^*(\theta_o) + 4I^2(\theta_o)\ E_{\theta_o}\left[\left\{\sqrt{n}(\hat{\theta}_{n1}^*+\hat{\theta}_{n2}^*+\hat{\theta}_{n3}^*)\right\}^3(\sqrt{n}\ R_n^*)\right]$$

$$+ 6I^2(\theta_o)\ E_{\theta_o}\left[\left\{\sqrt{n}\ (\hat{\theta}_{n1}+\hat{\theta}_{n2}^*+\hat{\theta}_{n3}^*)\right\}^2(\sqrt{n}\ R_n^*)^2\right] + o\ (n^{-1}). \qquad (5.2.10)$$

Now, by (4.2.25) and Property 5.2.3,

$$\int R_n^*\ Z_n^2(\theta)\ L(\underline{x},\ \theta)\ d\underline{x} = \frac{c(\theta)}{n^{5/2}} + o\ (n^{-5/2}) \qquad (5.2.11)$$

and

$$\int R_n^*\ Z_n^2(\theta)\ W_n(\theta)\ L(\underline{x},\ \theta)\ d\underline{x} = \frac{d(\theta)}{n^{5/2}} + o\ (n^{-5/2}) \qquad (5.2.12)$$

where $L(\underline{x},\ \theta)$ is the joint density function of $\underline{x} = (x_1, x_2, \ldots, x_n)$ and

$$W_n(\theta) = \frac{1}{n}\ \frac{d^2\log L}{d\theta_o^2} + I(\theta_o).$$

Differentiating both sides of (5.2.11) with respect to θ, we get

$$2 \int R_n^* Z_n(\theta) W_n(\theta) L(\underline{x}, \theta) d\underline{x}$$

$$+ \int R_n^* Z_n^2(\theta) n Z_n L(\underline{x},\theta) d\underline{x} = o_{-}(n^{-5/2})$$

implying

$$\int R_n^* Z_n^3(\theta) L(\underline{x}, \theta) = o (n^{-7/2}), \text{ using (5.2.12).} \qquad (5.2.13)$$

[As in the proof of Lemma 4.2.8, (5.2.13) can be justified by direct calculations.] (5.2.13) gives

$$E_{\theta_o}[\{\sqrt{n}(\hat{\theta}^*_{n1}+\hat{\theta}^*_{n2}+\hat{\theta}^*_{n3})\}^3 \cdot (\sqrt{n} R_n^*)] = o_{-}(n^{-1}).$$

On the other hand, by using Properties 5.2.4 and 5.2.5,

$$E_{\theta_o}[\{\sqrt{n} (\hat{\theta}^*_{n1}+\hat{\theta}^*_{n2}+\hat{\theta}^*_{n3})\}^2 (\sqrt{n} R_n^*)^2]$$

$$= E_{\theta_o}[\{(\frac{\sqrt{n} Z_n}{I(\theta_o)})^2 + \frac{\sqrt{n} Z_n}{I(\theta_o)} \cdot O_E(n^{-1/2})\} (\sqrt{n} R_n^*)^2] \qquad (5.2.14)$$

$$= E_{\theta_o}[\{\frac{\sqrt{n} Z_n}{I(\theta_o)}\}^2 \cdot \{\frac{1}{4} n Q_n^2 - \frac{(b_o(\theta_o)-b(\theta_o))^2}{n}\}] + o (n^{-1})$$

$$= E_{\theta_o}\{\frac{\sqrt{n} Z_n}{I(\theta_o)}\}^2 \cdot E_{\theta_o}\{\frac{n}{4} Q_n - \frac{(b_o(\theta_o)-b(\theta_o))^2}{n}\}] + o_{-}(n^{-1})$$

$$= \frac{1}{I(\theta_o)} \cdot \{1+ \frac{\bar{K}^*_{21}(\theta_o)}{\sqrt{n}} + \frac{\bar{K}^*_{22}(\theta_o)}{n}\} E_{\theta_o}\{(\sqrt{n} R_n^*)^2\} + o (n^{-1})$$

(5.2.15)

where in the above expression the term $\left\{\frac{\bar{K}^*_{21}(\theta_o)}{\sqrt{n}} + \frac{\bar{K}^*_{22}(\theta_o)}{n}\right\} \times$

$\times E_{\theta_o}\{\sqrt{n}\, R^*_n\}^2$ is actually $o(n^{-1})$ but we have introduced it there to bring the expression in the following form :

$$\frac{1}{I(\theta_o)} \{\bar{K}^*_2(\theta_o)\} \left\{\frac{\mu^*_2(\theta_o) - \bar{\mu}^*_2(\theta_o)}{I(\theta_o)}\right\} + o\,(n^{-1}) . \qquad (5.2.16)$$

To get (5.2.15) from (5.2.16) we have used (5.2.8) and (5.1.4).

Combining (5.2.10), (5.2.11) and (5.2.16) we get

$$\mu^*_4(\theta_o) = \bar{\mu}^*_4(\theta_o) + 6\,\bar{\mu}^*_2(\theta_o)\,\{\mu^*_2(\theta_o) - \bar{\mu}^*_2(\theta_o)\} + o(n^{-1})$$

(since $\bar{K}^*_2(\theta_o) = \bar{\mu}^*_2(\theta_o)$).

Hence

$$\mu^*_4(\theta_o) - 3\,\mu^{*2}_2(\theta_o) = \bar{\mu}^*_4(\theta_o) - 3\mu^{*2}_2(\theta_o) + 6\bar{\mu}^*(\theta_o)\mu^*_2(\theta_o) - 6\bar{\mu}^{*2}_2(\theta_o) + o(n^{-1})$$

$$K^*_4(\theta_o) = \bar{K}^*_4(\theta_o) - 3\{\mu^*_2(\theta_o) - \bar{\mu}^*_2(\theta_o)\}^2 + o(n^{-1})$$

$$= \bar{K}^*_4(\theta_o)$$

(since $\{\mu^*_2(\theta_o) - \bar{\mu}^*_2(\theta_o)\}^2 = I^2(\theta_o)[E^U(\sqrt{n}\, R^*_n)^2]^2 + o(n^{-1}) = o(n^{-1})$).

This completes the proof of the theorem.

<u>Proof of Corollary 5.1.2.</u> Choose $\theta_o \varepsilon (\bar{\underline{H}})_o$ arbitrarily and fix it. Using (4.2.46) we get on direct computation

$$K_{21}(\theta_o) = \bar{K}_{21}(\theta_o) \text{ and } K_{32}(\theta_o) = \bar{K}_{32}(\theta_o) = 0 . \qquad (5.2.17)$$

In view of (5.2.7), (5.2.17) and the theorem, it follows from (5.1.1) and Property 5.3.7 that in the Edgeworth expansion (under θ_o) for $\sqrt{n}(T_n^* - \theta_o)\sqrt{I(\theta_o)}$ and $\sqrt{n}(\hat{\theta}_n^* - \theta_o)\sqrt{I(\theta_o)}$ up to $o(n^{-1})$ the only coefficients which differ are $K_{22}^*(\theta_o)$ and $\bar{K}_{22}^*(\theta_o)$ and $K_{22}^*(\theta_o) > \bar{K}_{22}^*(\theta_o)$ by (5.1.8). This immediately implies the corollary.

<u>Remark 5.2.1.</u> The corollary remains valid if one replaces T_n^* by T_n and $\hat{\theta}_n^*$ by $\hat{\theta}_n'$ where $\hat{\theta}_n' = \hat{\theta}_n - c(\hat{\theta}_n)/n$ with $c(\theta) = \{\bar{K}_{11}(\theta) - K_{11}(\theta)\} / I(\theta)$ so that $\hat{\theta}_n'$ and T_n have the same "bias" up to $o(n^{-1})$.

<u>Remark 5.2.2.</u> The probability inequality connecting T_n and $\hat{\theta}_n'$ has also been recently proved by Ghosh, Sinha and Wieand (1980) for an arbitrary one-parameter family of distributions by a different approach. As mentioned there, the technique of proof depends on comparing the performance of a natural test based on an efficient estimate for a certain hypothesis testing problem with that of the Bayes test under a suitably chosen prior and

noting that a test based on the m.l.e is Bayes up to a certain order.

5.3 Non-absolutely continuous case

We assume the same set up as in Section 5.2 except that (5.2.1) is now dropped. As mentioned in the introduction, the formal Edgeworth expansion (5.1.1) is then no longer valid. However, we shall show that if the loss function satisfies certain additional conditions, then the inequality (5.1.11) is still true.

Fix $\theta_o \in \textcircled{H}$. We shall regard θ_o as the true value of the parameter. Consider the curved exponential density

$$f(x,\pi(\theta_o)) = c(\pi(\theta_o)) \exp \left\{ \sum_{j=1}^{k} \beta_j(\pi(\theta_o))\, p_j(x) \right\}. \qquad (5.3.1)$$

Denote $\underline{Y}^* = (p_1(X) - \pi_1(\theta_o), \ldots, p_k(X) - \pi_k(\theta_o))$ and assume without loss of generality that $\mathrm{cov}_{\theta_o}(\underline{Y}^*) = I$, the $k \times k$ identity matrix. We will denote by $\| \underline{Y} \|^{s}$ the random variable $\sum_{i=1}^{k} Y_i^{s}$. In view of (5.3.1) it follows that

$$E_{\theta_o} \| \underline{Y} \|^{s} < \infty \quad \text{for all } s \geq 3. \qquad (5.3.2)$$

Given $X_1, X_2, \ldots, X_n$ i.i.d random variables according to (5.3.1), define

$$y_i^n = y_i = \sqrt{n}\,(\bar{p}_{in} - \pi_i(\theta_o)), \quad i = 1,2,\ldots, k$$

and

$$\underline{y} = (y_1, y_2, \ldots, y_k)$$

where $\bar{p}_{in} = \frac{1}{n} \sum_{j=1}^{n} p_i(x_j)$.

Let P_n stand for the distribution of $\underline{Y}^*$ and ψ_{s_o-2} for the formal multivariate Edgeworth expansion of length (s_o-2) for $\underline{Y}^*$ where $s_o \geq 3$ is an integer (vide Götze and Hipp (1978), for details). For any function $f : R^k \to R$, define for $r \geq 0$,

$$M_r(f) = \sup \{(1 + \| \underline{y} \|^r)^{-1} \, |f(\underline{y})| : \underline{y} \in R^k\}.$$

For any positive integer m, let $C^m(R^k)$ denote the set of all functions on R^k with continuous derivatives of order m. Finally, for k-dimensional non-negative integral vector $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, let

$$D^{\alpha} f = \frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_k}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_k^{\alpha_k}} f(\underline{x})$$

and $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_k$.

Then the following propositions hold.

<u>Proposition 5.3.1.</u> Let $f : R^k \to R$. Assume the following :

(i) $M_s(f) < \infty$ for some $s \geq 3$.

(ii) $f \in C^{s_o-2}(R^k)$ with s_o = integral part of s.

(iii) $M_p(D^\alpha f) < \infty$ for $|\alpha| = s_o-2$ and for some positive integer p.

Then

$$\int f\, d(P_n - \Psi_{s_o-2}) = o(n^{-(s_o-2)/2}).$$

Proposition 5.3.2 For all $s \geq 3$,

$$\int \|\underline{x}\|^s I_{\{\|x\|>[(s-2)\log n]^{1/2}\}}(\underline{x})\, P_n(d\underline{x}) = o(n^{-(s-2)/2}).$$

Remark 5.3.3. Propositions 5.3.1 and 5.3.2 follow from Theorems 3.6 and 3.2.1 and Remarks 3.9 and 3.10 of Götze and Hipp (1978) coupled with (5.3.2).

For any function $f(\underline{y})$, let $E\, f(\underline{Y})$ and $E'_{s_o-2}\, f(\underline{Y})$ denote the expectations of $f(\underline{Y})$ under P_n and Ψ_{s_o-2} respectively.

Let T_n be efficient and l.s.(IV) and T_n^* be as defined in Section 5.2. Assume that the loss function $L(T_n^*, \theta_o) = h(\sqrt{n}\,(T_n^* - \theta_o))$ satisfies the following condition.

For some $r \geq 2$

$$h(Z) = |Z|^r \quad \text{if} \quad \|\underline{y}\| \leq c\sqrt{\log n}$$

$$= K_n \,, \quad \text{otherwise} \qquad (5.3.3)$$

where K_n and c are constants to be suitably chosen. Note from (4.2.28) and the definition of T_n^* that

$$T_n^* - \theta_o = \bar{f}_n(\underline{y}) + 0(\sqrt{\frac{\log n}{n}})^4 \quad \text{if } \|\underline{y}\| \leq c\sqrt{\log n} \qquad (5.3.4)$$

where

$$\bar{f}_n(\underline{y}) = \sum_1^k a_i \frac{y_i}{\sqrt{n}} + \frac{1}{n} \sum_i \sum_j a_{ij}\, y_i\, y_j$$

$$+ \frac{1}{n^{3/2}} \sum_i \sum_j \sum_\ell a_{ij\ell}\, y_i\, y_j\, y_\ell \qquad (5.3.5)$$

where $y_i = (\bar{p}_{in} - \pi_i(\theta_o))\sqrt{n}$ and a_i's etc., are constants (their explicit expressions are not needed for our purpose) and the term $0(\sqrt{\frac{\log n}{n}})^4$ in (5.3.3) is uniformly so over $\|\underline{y}\| \leq c\sqrt{\log n}$.

Define

$$f_n(\underline{y}) = |\sqrt{n}\, \bar{f}_n(\underline{y})|^r : R^k \longrightarrow R$$

and assume that f_n satisfies

(i) $M_4(f_n) \leq A < \infty$

(ii) $f_n \in C^2(R^k)$

(iii) for $|\alpha| = 2$, $\{D^\alpha f_n, n \geq 1\}$ is equicontinuous on compact sets of $\underline{y}$

(iv) for $|\alpha| = 2$ and some positive integer p, $M_p(D^\alpha f_n) \leq B < \infty$.

(5.3.6)

We will prove

Corollary 5.3.4 : Under (5.3.3) and (5.3.6),

$$E\{L(T_n^*, \theta_o)\} \geq E\{L(\hat{\theta}_n^*, \theta_o)\} + o(n^{-1}).$$

Before proving the corollary, let us establish a few auxiliary results. Given L satisfying (5.3.3), we define L' as

$$L'(\underline{y}, \theta_o) = f_n(\underline{y}) \quad \text{if} \quad \|\underline{y}\| \leq c\sqrt{\log n}$$

$$= K_n, \quad \text{otherwise} \tag{5.3.7}$$

and claim that

$$E\{L(\underline{y}, \theta_o) - L'(\underline{y}, \theta_o)\} = o(n^{-1}). \tag{5.3.8}$$

To prove (5.3.8), note that

$$E\{L(\underline{y},\theta_o)-L'(\underline{y},\theta_o)\} = E[\{\left|\sqrt{n}\,\bar{f}_n(\underline{y})+ O\left(\frac{(\sqrt{\log n})^4}{n^{3/2}}\right)\right|^r - |\sqrt{n}\,\bar{f}_n(\underline{y})|^r\} \cdot I_{\|\underline{y}\|\leq c\sqrt{\log n}}]. \tag{5.3.9}$$

Denote

$$X_n = \sqrt{n}\,\bar{f}_n(\underline{y}) = \Sigma a_i y_i + \frac{1}{\sqrt{n}} \Sigma\Sigma a_{ij}\, y_i\, y_j + \frac{1}{n} \sum_i \sum_j \sum_\lambda a_{ij\lambda}\, y_i\, y_j\, y_\lambda \quad \text{(using 5.3.5)}$$

(5.3.10)

and observe that

$$\Sigma\, a_i y_i = 0_p(1), \quad \frac{1}{\sqrt{n}} \Sigma \Sigma\, a_{ij}\, y_i\, y_j = 0_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{and}$$

$$\frac{1}{n} \Sigma \Sigma \Sigma\, a_{ij\ell}\, y_i y_j y_\ell = 0_p(1/n). \tag{5.3.11}$$

Hence, (5.3.9) can be written as

$$E\left[\left\{\left|X_n + 0\left(\frac{(\sqrt{\log n})^4}{n^{3/2}}\right)\right|^r - |X_n|^r\right\} \cdot I_{\|\underline{y}\| \le c\sqrt{\log n}}\right]$$

$$= E\left[\left\{\left|X_n + 0\left(\frac{(\sqrt{\log n})^4}{n^{3/2}}\right)\right|^r - |X_n|^r\right\} \cdot I_{\{\|\underline{y}\| \le c\sqrt{\log n}\, \cap\, |X_n| \ge n^{-3/4}\}}\right.$$

$$\left. + E\left[\left\{\left|X_n + 0\left(\frac{(\sqrt{\log n})^4}{n^{3/2}}\right)\right|^r - |X_n|^r\right\} \cdot I_{\{\|\underline{y}\| \le c\sqrt{\log n}\, \cap\, |X_n| < n^{-3/4}\}}\right]. \tag{5.3.12}$$

Now

$$\text{the first term} \le C \cdot E\left\{|X_n|^{r-1}\right\} \cdot 0\left(\frac{(\sqrt{\log n})^4}{n^{3/2}}\right) \text{ for some } C$$

$$= o(n^{-1}), \text{ by (5.3.10), (5.3.11) and (5.3.2).}$$

and

$$\text{the second term} \le n^{-\frac{3r}{4}} + \left\{n^{-3/4} + 0\left(\frac{(\sqrt{\log n})^4}{n^{3/2}}\right)\right\}^r$$

$$= o(n^{-1}), \quad \text{since } r \ge 2.$$

This establishes (5.3.8).

We next define L'' as

$$L''(\underset{\sim}{y}, \theta_o) = |X_n|^r \text{ for all } \underset{\sim}{y} \tag{5.3.13}$$

and claim that

$$E\{L'(\underset{\sim}{y}, \theta_o) - L''(\underset{\sim}{y}, \theta_o)\} = o(n^{-1}). \tag{5.3.14}$$

To prove (5.3.13), note that

$$E\{L'(\underset{\sim}{y},\theta_o)-L''(\underset{\sim}{y},\theta_o)\} = E[K_n I_{\|\underset{\sim}{y}\|> c\sqrt{\log n}} - |X_n|^r I_{\|\underset{\sim}{y}\| > c\sqrt{\log n}}].$$

Using a result of Bhattacharya and Ranga Rao [(1976), p.179, Corollary 17.3], which is essentially a weaker version of Proposition 5.3.2, it follows that

$$E[I_{\|\underset{\sim}{y}\| > c\sqrt{\log n}}] = o(n^{-c^2/2}). \tag{5.3.15}$$

Also, from (5.3.10) and (5.3.11), it follows upon applying a suitable moment inequality and using Proposition 5.3.2 that

$$E[|X_n|^r I_{\|\underset{\sim}{y}\| > c\sqrt{\log n}}] = o(n^{-1}) \tag{5.3.16}$$

for an appropriately chosen c. Combining (5.3.15) and (5.3.16) and choosing c and K_n suitably, (5.3.14) follows.

In view of (5.3.8) and (5.3.14), it follows that

$$E\{L(\underline{y},\theta_o)\} = E\{L''(\underline{y},\theta_o)\} + o(n^{-1}). \qquad (5.3.17)$$

To compute $E\{L''(\underline{y},\theta_o)\}$, we identify $|X_n|^r$ as the function $f_n(\underline{y})$ and use the easily verifiable fact that Proposition 5.3.1 (with $s = 4$) is also valid if one replaces f by f_n satisfying (5.3.6).

This leads to

$$E\{L''(\underline{y},\theta_o)\} = E_2^! \{L''(\underline{y},\theta_o)\} + o(n^{-1}) \qquad (5.3.18)$$

where $E_2^!$, as mentioned before, refers to the expectation taken under Ψ_{s_o-2} (here $s_o = 4$).

We now apply Lemma 2.1 and Remark 1.4 (p.438) of Bhattacharya and Ghosh (1978) to conclude that

$$E_2^! \{L''(\underline{y},\theta_o)\} = \int_{-\infty}^{\infty} |u|^r \, dF_n(u) + o(n^{-1}) \qquad (5.3.19)$$

where F_n is the (formal) Edgeworth expansion (of length 2) for $\sqrt{n}\,(T_n^* - \theta_o)$ (or, what is the same, that of its Taylor's expansion X_n. To justify (5.3.19), it is enough to note that (i) Lemma 2.1 of Bhattacharya and Ghosh (1978) does not assume Cramer's condition ; (ii) the proof of their Theorem 2 given on pages 445-446 that the F_n of Lemma 2.1 coincides with the formal Edgeworth expansion does not require Cramer's condition and (iii) the second identity in their (2.10) is true if we replace $h^{j}_{s-1}(z)$ by any function $L(h_{s-1}(z))$ provided L has at most a polynomial growth as $u \to \pm\infty$.

Having established (5.3.19), we are now in a position to provide a proof of the corollary.

Proof of Corollary 5.3.4. By (5.3.17), (5.3.18) and (5.3.19),

$$E\{L(T_n^*, \theta_o)\} = \int_{-\infty}^{\infty} |u|^r \, dF_n(u) + o(n^{-1})$$

and exactly analogously

$$E\{L(\hat{\theta}_n^*, \theta_o)\} = \int_{-\infty}^{\infty} |u|^r \, dF_n^*(u) + o(n^{-1})$$

where $F_n^*(u)$ stands for the (formal) Edgeworth expansion (of length 2) for $\sqrt{n}\,(\hat{\theta}_n^* - \theta_o)$. The (formal) Edgeworth expansions $F_n(u)$ and $F_n^*(u)$ are readily obtained from the right hand side of (5.1.1) by replacing the K_{ij}'s by the corresponding cumulants of $\sqrt{n}\,(T_n^* - \theta_o)$ and $\sqrt{n}\,(\hat{\theta}_n^* - \theta_o)$ respectively. Upon evaluating the integrals involved above and applying the theorem, the corollary follows.

5.4 An example of Berkson - revisited

The following example was studied to get an idea about the numerical accuracy of the approximations used in Chapters 4 and 5.

Following Berkson (1955) consider 3 doses d_1, d_2 and d_3 equally spaced, unit distance apart, say -1, 0, and 1 respectively; 10 animals are exposed to each dose and the true probabilities of an animal responding to doses d_1, d_2 and d_3 are taken to be $\pi_1 = 0.3$, $\pi_2 = 0.5$ and $\pi_3 = 0.7$, respectively which in turn determines the values of the parameters as $\theta = 0$ and $\beta = 0.8473$ (using the notation of Section 4.4). When we expose 10 animals for each dose, there are 11 possibilities of animal response at each dose and hence with 3 doses we get $11^3 = 1331$ possible samples. We took only 1329 of these samples because at the two remaining cases namely when all the animals died or survived i.e., $p_i = 0$ or 1 the m.l.e. of θ is not finite. In all the other cases, the m.l.e. is the unique solution of the likelihood equation. (In the language of Chapter 4, this amounts to taking a suitable $U = (p_1, p_2, p_3)$ which excludes $(0,0,0)$ and $(1,1,1)$). We computed the m.l.e. using the method of scoring (i) with two iterations $(\hat{\theta}_n(1))$ and three iterations $(\hat{\theta}_n(2))$. The minimum logit chisquare was computed for each sample. Whenever $p_i = 0$ or 1, the minimum logit chisquare estimate is not well defined. As suggested by Berkson we used the

2n rule to compute the estimate in these cases. As given in (4.4.9) we computed the bias adjusted m.l.e.'s namely $\hat{\theta}_n(2)$ and $\hat{\theta}_n(3)$. (It turns out that in this example these two agree in most cases up to three decimal places).

The mean square error and the fourth moment of $\hat{\theta}_n(2)$ are 0.13428 and 0.05691 respectively and the corresponding values for the minimum logit chisquare are 0.13644 and 0.0589 respectively. (The mean square error of the minimum logit chisquare estimate agrees quite well with that given in Berkson (1955).) This shows that the m.l.e. has a slight edge over the minimum logit chisquare estimate w.r.t. the squared error loss and its square.

According to formula (4.4.9) the mean square error of the bias adjusted m.l.e. is given by (up to $o(n^{-2})$)

$$\frac{1}{nI} - \frac{2}{n^2 I} + \frac{3}{n^2 I^4}[\Sigma\, \pi_i(1-\pi_i)(2\pi_i-1)]^2 + \frac{2}{n^2 I^3}\, \Sigma\, \pi_i^2(1-\pi_i)^2$$

$$- \frac{1}{n^2 I^3}\, \Sigma\, \pi_i(1-\pi_i)(2\pi_i-1)^2 - \frac{3}{2n^2 I^3}[\Sigma\, \pi_i(1-\pi_i)(2\pi_i-1)][\Sigma(2\pi_i-1)]$$

$$+ \frac{1}{4n^2 I^4}[\Sigma\, \pi_i(1-\pi_i)(2\pi_i-1) - I\, \Sigma(2\pi_i-1)]^2$$

$$= 0.12719$$

and that for the minimum logit chisquare estimate using (4.4.4) is given by (up to $o(n^{-2})$)

$$\frac{1}{nI} + \frac{2}{n^2I^4} \Sigma (2\pi_i - 1)\, \pi_i^2 (1-\pi_i)^2 + \frac{1}{n^2I^4} [\Sigma (2\pi_i-1)\, \pi_i(1-\pi_i)]^2$$

$$+ \frac{1}{n^2I^4} \sum_{i\neq i}\sum \pi_i\pi_j(1-\pi_i)(1-\pi_j)(2\pi_i-1)^2 + \left(\frac{3}{2n^2I^3} - \frac{2}{n^2I^4}\right) \Sigma (2\pi_i-1)^2$$

$$+ \left(\frac{1}{4n^2I^2} - \frac{1}{n^2I^4}\right) [\Sigma(2\pi_i-1)]^2 - \frac{5}{n^2I^3} \Sigma(2\pi_i-1)^2\, \pi_i(1-\pi_i)$$

$$- \frac{1}{n^2I^3} \sum_{i\neq j}\sum (2\pi_i-1)(2\pi_j-1)\, \pi_j(1-\pi_j)$$

$$= .13154$$

So the approximations are in error by 0.00709 for the m.l.e. and .00490 for the minimum logit chisquare estimate.

REFERENCES

Akahira, M., and Takeuchi, K. (1976a). On the second-order asymptotic efficiencies of estimators. Proceedings of the Third Japan-USSR Symposium on Probability Theory (G.Maruyama and J.V. Prokhorov, Eds.), 604-638. Lecture Notes in Mathematics 550, Springer-Verlag, Berlin.

Akahira, M., and Takeuchi, K. (1976b). On the second-order asymptotic efficiency of estimators in multiparameter cases. Rep. Univ. Electro-Comm. 26, 261-269.

Bahadur, R.R. (1969). Some limit theorems in statistics. Regional Conference series in Applied Mathematics, Published by SIAM, Philadelphia, PA.

Barankin, E.W. (1949). Locally best unbiased estimates, Ann. Math. Statist. 20, 477-501.

Berkson, J. (1955). Maximum likelihood and minimum χ^2 estimates of the logistic function. J. Amer. Statist. Assoc., 50, 130-126.

Berkson, J., and Hodges, J.L. Jr. (1961). A minimum estimator for the logistic function. Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, 4, 77-86, Univ. of California Press.

Bhattacharya, A. (1946). On some analogues of the amount of information and their uses in statistical estimation, Sankhyā, 8, 1-14, 201-218, 315-328.

Bhattacharya, R.N., and Ghosh, J.K.(1978). On the validity of the formal Edgeworth expansion. Ann. Statist., 6, 434-451.

Bhattacharya, R.N., and Rao, R. Ranga (1976). Normal Approximation and Asymptotic Expansions. Wiley, New York.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist., 23, 493-507.

Cox, D.R. (1962). Further results on tests of separate families of hypothesis. J.R. Statist. Soc., Series B, 24, 406-424.

Cramer, H. (1946). Mathematical Methods of Statistics, Princeton University Press, Princeton.

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). Ann. Statist. 3, 1189-1242.

Feldman, D. and Fox, M. (1968). Estimation of the parameter n in the binominal distribution. J. Amer. Statist. Assoc., 63, 150-158.

Ferguson, T. S. (1967). Mathematical Statistics A Decision Theoretical Approach. Academic Press, New York and London.

Fisher, R.A. (1925). Theory of statistical estimation. Proceedings Cambridge Phil. Soc., 22, 700-725.

Gotze, F., and Hipp, C. (1978). Asymptotic expansions in the central limit theorem under moment conditions. Z. Wahrscheinlichkeitstheorie Verw. Gebiete 42, 67-87.

Ghosh, J.K., Sinha, B.K., and Subramanyam, K. (1979). Edgeworth expansions for Fisher-consistent estimators and second order efficiency. Indian Statistical Institute, Tech. Report No. Math-Stat. 7/79.

Ghosh, J.K., Sinha, B.K., and Wieand, H.S. (1980). Second order efficiency of the mle with respect to any bounded bowl-shaped loss function. To appear in Ann. Statist.

Ghosh, J.K., and Subramanyam, K. (1971). Estimation of discrete parameters in large samples. Indian Statistical Institute, Tech. Report No. Math-Stat. 7/71.

Ghosh, J.K., and Subramanyam, K. (1974). Second order efficiency of maximum likelihood estimators. Sankhyā, Series A, 36, 325-358.

Ghosh, J.K. and Subramanyam, K. (1975). Inference about separated families in large samples. Sankhyā, Series A, 37, 502-513.

Ghosh, M. and Meeden, G. (1978). Admissibility of the mle of the normal integer mean. Sankhyā, Series B, 40, 1-10.

Hajek, J. (1972). Local asymptotic minimax and admissibility in estimation. Proceedings of the Sixth Berkeley Symposium, 1, 175-194. Univ. of California Press.

Hammersley, J.M. (1950). On estimating restricted parameters. J.R. Statist., Soc., Series B, 12, 192-240.

Hodges, J.L., Jr., and Lehmann, E.L. (1970). Deficiency. Ann. Math. Statist., 41, 783-801.

Kendall, M.G. (1946). The Advanced Theory of Statistics. First edition, 2, Griffin and Co., London.

Kendall, M.G. (1952). The Advanced Theory of Statistics. (Volume I). Charles Griffin and Company Ltd., London.

Khan, R.A. (1973a). On some properties of Hammersley's estimator of an integer mean. Ann. Statist., 1, 756-762.

Khan, R.A. (1973b). On sequential distinguishability. Ann. Statist., 1, 838-850.

Khan, R. A.(1978). A note on the admissibility of Hammersley's estimator of an integer mean. Canad. J. Statist., 6, 113-117.

Kraft, O., and Puri, M.L. (1974). The asymptotic behaviour of minimax risk for multiple decision problems. Sankhyā, Series A, 36, 1-12.

Laderman, L. (1955). On the asymptotic behaviour of decision procedures, Ann. Math. Statist., 26, 551-575.

Lindley, D.V. (1961). The use of prior probability distribution in statistical inference and decision. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics, 1, 453-468. Univ. of California Press.

Nandi, H.K. (1956). Variance of a linear function of constrained variables. Cal. Stat. Assoc. Bulletin, 7, 41-43.

Pfanzagl, J. (1973). Asymptotic expansions related to minimum contrast estimators. Ann. Statist., 1, 993-1026.

Pfanzagl, J. (1975). On asymptotically complete classes. In Statistical Inference and Related Topics. Proceedings of the Summer Research Institute on Statistical Inference for Stochastic Processes, Vol.2 (M.L.Puri, Ed.), 1-43, Academic Press, New York.

Pfanzagl, J. and Wefelmeyer, W. (1978). A third-order optimum property of the maximum likelihood estimator. J. Multivariate Analysis, 8, 1-29.

Ponnapalli, R. (1976). Deficiencies of minimum discrepancy estimators. Canad. J. Statist., 4, 33-50.

Rao, C.R. (1960). Apparent anomalies and irregularities of the method of maximum likelihood. Proceedings of the 32nd Session of the Int. Stat. Inst., Tokyo, Reprinted with discussion in Sankhyā. 24.73-102.

Rao, C.R. (1961). Asymptotic efficiency and limiting information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1, 531-546. Univ. of California Press.

Rao, C.R. (1962). Efficient estimates and optimum inference procedures in large samples. J. Roy. Statist., Soc., Ser. B, 24, 46-72.

Rao, C.R. (1963). Criteria of estimation in large samples. Sankhyā, Ser. A, 25, 189-206. Also included in Contributions to Statistics presented to Professor P.C. Mahalanobis on the occasion of his 70-th birthday.

Rao, C.R. (1973). Linear Statistical Inference and its Applications. 2nd edition. Wiley, New York.

Searle, S.R. (1971). Linear Models. John Wiley and Sons, Inc., New York.

Sethuraman, J. (1964). On the probability of large deviations of families of sample means. Ann. Math. Statist., 35, 1304-1316.

Sethuraman, J. (1970). Correction to " On the probability of large deviations of families of sample means". Ann. Math. Statist., 41, 1376-1380.

Silverstone, H. (1957). Estimating the logistic curve. J. Amer. Statist. Assoc., 52, 567-577.

Takeuchi, K. and Akahira, M. (1978). Third order asymptotic efficiency of maximum likelihood estimator for multiparameter exponential case. Reprinted from Rep. Univ. Electro-Comm., 28-2, (Sci. and Tech. Sect.), 271-293.

Unni, K. (1978). The theory of estimation in algebraic and analytic exponential families with applications to variance component models. Ph.D. Thesis, Indian Statistical Institute.