Contributions to the Neyman-Scott and Mixture Problems

Joydeep Bhanja

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements
for the award of the degree of
Doctor of Philosophy
CALCUTTA
1991

Acknowledgements

This work is done under the guidance of Prof. J. K. Ghosh. Mere words can not express my sense of gratitude for the encouragement, help and guidance he has given me. In spite of his heavy burden of work, I had the liberty to occupy much of his precious time. During the final stage of the thesis I had a very difficult time — but for his help and unlimited patience this dissertation would not have materialised.

Prof. S. C. Bagchi has helped me in obtaining the counterexample given in Chapter 9 and Prof. B. V. Rao supplied me the references to the von Neumann Selection Theorem and other selection theorems. It is a pleasure to record my thanks to them.

At various stages of the preparation of the manuscript, suggestions of Dr. T. Samanta led to much improvement of the thesis. This is an opportunity to aknowledge this. Thanks are also due to Mr. S. Purkayastha for a very careful scrutiny of the final version of the manuscript.

At different stages of the computations recorded in Chapter 8, I have taken the help of Mr. S. C. Kundu, Mr. R. K. Naha, Mr. P. K. Raha and others of CSSC, Mr. S. Nandy and Mr. N. R. Pal of CSU, Dr. A. Bagchi of ECSU and Prof. S. B. Rao of Stat-Math. I would like to thank all of them.

In my extra-academic matters, I have got constant encouragement from my friend Dr. N. Sarkar and others from ERU and Prof. Bikas K. Sinha of Stat-Math. I am indebted to them beyond limits. In this respect, I would also like to mention the names of Prof. P. P. Majumder from CSU and Dr. S. K. Parui and Mr. M. K. Kundu and others from ECSU.

Finally, I would like to thank Prof. A. R. Rao and Mr. P. Bandyopadhyaya for their help in using the LaTeX, Prof. A. K. Adhikari and Prof. T. Krishnan for their help in using the Laser Printer of CSU.

Calcutta, February 1991 Joydeep Bhanja

Contents

Chapter 1
Introduction

Neyman and Scott (1948) were the first to point out that the method of maximum likelihood fails to provide efficient estimates when the number of parameters grows with the sample size n. Consider the following examples introduced by them:

Example 1.1 Let $\{ \boldsymbol{X}_i \}$ be a sequence of independent random vectors in $\mathbb{R}^p$, components X_{ij} of $\boldsymbol{X}_i$ being independent normal with mean μ_i and variance σ^2. Here σ^2 is the parameter of interest. It is easy to see that the maximum likelihood estimate for σ^2 is not even consistent. It is also known [see Lindsay (1980), Pfanzagl (1982), van der Vaart (1987)] that if $p \geq 2$, the maximum partial likelihood estimate based on $X_{ij} - \bar{X}_i$ is efficient.

Example 1.2 This is similar to Example 1.1 except that the components of $\boldsymbol{X}_i$ being independent normal with mean μ and variance σ_i^2. Here μ is the parameter of interest. It can be shown that the maximum likelihood estimate $\hat{\mu}$ is consistent and asymptotically normal provided $p \geq 3$ and $n^{-1} \sum_{i=1}^{n} \sigma_i^2$ is bounded away from zero, but it is not efficient. For $p = 1$, Bickel and Klaassen (1986), and for general p, van der Vaart (1987) show how an efficient, asymptotically normal estimate can be constructed.

Lindsay (1980) and Bickel and Klaassen (1986) provide an extremely useful general discussion of such problems. See also van der Vaart (1987, 1988).

A little reflection shows in most problems of this type the m.l.e. of the parameter of interest (θ) will be inconsistent as in Example 1.1. However, it is not easy to construct examples where this can be demonstrated mathematically. The following is a new class of such examples.

Example 1.3 Let $X_1, X_2, \ldots, X_n$ be independent random variables with $X_i \sim f(\cdot, \theta, \xi_i)$ where the function f is given by

$$f(x, \theta, \xi) := A(\theta, \xi) \exp\{\theta\psi(x) + \xi x\}$$

for any real number x and (θ, ξ) in Ω, where

$$\Omega := \{(\theta, \xi) : \int \exp\{\theta\psi(x) + \xi x\} dx < \infty\}$$

and the real valued function ψ is strictly convex or strictly concave.

Here one can show that the m.l.e. $\hat{\theta}$ is inconsistent. The details of the verification are given in Chapter 2.

The construction of efficient estimates in Examples 1.1 and 1.2 follow quite different routes. In the following pages we develop a general theory for constructing efficient estimates which is applicable to both these examples. However the efficient estimate constructed this way for Example 1.2 would differ from those in Bickel and Klaassen (1986) and van der Vaart (1987).

We first formulate a general model. For this purpose we shall use the following notations. Let Θ be an open subset of $\mathbb{R}$ with compact closure $\bar{\Theta}$, Ξ a compact metric space and $\mathcal{G}$ the set of all Borel probability measures on Ξ. The requirement of compact closure of Θ can be dropped when θ is a location parameter, as in Example 1.2, and, more generally, when there is a uniformly consistent estimate of θ. Note that $\mathcal{G}$ is weakly compact. Equip $\bar{\Theta}$ with the Euclidean metric topology and $\mathcal{G}$ with the weak topology. Let $(S, \mathcal{S})$ be an arbitrary measurable space. let $\mathbb{F}_n$ denote the empirical distribution function (e.d.f.) or the empirical probability measure based on n observations of a random variable. In particular, for n elements $\xi_1, \xi_2, \ldots, \xi_n$ from Ξ, denote $\mathbb{F}_n(\cdot, \xi_1, \xi_2, \cdots, \xi_n)$ by $\underline{G}_n$.

Model I : Let $\{ X_i \}$ be a sequence of independent random variables taking values in $(S, \mathcal{S})$ with the distributon of X_i given by $P_{\theta_o, \xi_i}, \theta_o \in \Theta, \xi_i \in \Xi$. (The probability measure $P_{\theta,\xi}$ is assumed to be well defined for $\theta_o \in \bar{\Theta}$, $\xi \in \Xi$.) The object is to estimate the so-called structural parameter θ_o.

In Model I, also called the fixed set-up model by Bickel and Klaassen (1986), invariance of the estimation problem under permutations suggests restriction to the symmetric, *i.e.*, permutation invariant, sub-σ-field of $\mathcal{S}^n$. If one restricts $\prod_1^n P_{\theta_o,\xi_i}$ to this sub-σ-field, $\prod_1^n P_{\theta_o,\xi_i}$ can be replaced by

$$\frac{1}{n!} \sum_{\text{set of all permutations } \pi \text{ of } \{1,2,\cdots,n\}} \prod_1^n P_{\theta_o,\xi_{\pi(i)}}$$

and one would expect, heuristically, on the basis of the analogy between simple random sampling with and without replacement, that Model I can be approximated by the following:

Model II : Let $\{ X_i \}$ be a sequence of i.i.d. random variables taking values in $(S, \mathcal{S})$ with common distribution P_{θ_o,G_o} where for $A \in \mathcal{S}$,

$$P_{\theta_o,G_o}(A) := \int P_{\theta_o,\xi}(A) dG_o(\xi).$$

Model II, which is often called the mixed or mixture set up, was first proposed in the present context by Kiefer and Wolfowitz (1956). As pointed out by Bickel and Klaassen (1986), an analogous idea underlies Robbins's development of empirical Bayes methods to solve compound decision problems. A mathematical justification in the latter context is provided by Hannan and Robbins (1955) and Hannan and Huang (1972).

The heuristic argument leading to Model II from Model I can be made rigorous in our problem if the error of approximation in L_1-norm of the symmetrized measures in Model I and Model II (with $G \equiv \underline{G}_n$) tends to zero. Unfortunately, it is easy to show that this is not true. However, the approximation can be verified directly for the special class of estimates which Bickel and Klaassen (1986) called "regular". The following definition gives the notion of regularity and efficiency in a suitably modified form that ensures uniformity.

Definition 1.1 (i) *Regularity :* An estimate T_n of θ_o is called

(a) *regular in Model I* if there is $\sigma_T : \Theta \times \mathcal{G} \to \mathbb{R}^+$ continuous s.t.

$$\mathcal{L}(\sqrt{n}(T_n - \theta_o)\, \sigma_T^{-1}(\theta_o, \underline{G}_n) \mid \prod_1^n P_{\theta_o,\xi_i}) \Longrightarrow \mathcal{L}(\mathcal{N}(0,1))$$

uniformly on compact subsets of $\Theta \times \Xi^\infty$

and (b) *regular in Model II* if there is $\sigma_T : \Theta \times \mathcal{G} \to I\!R^+$ continuous s.t.

$$\mathcal{L}(\sqrt{n}(T_n - \theta_o)\,\sigma_T^{-1}(\theta_o, G_o) \mid P^n_{\theta_o, G_o}) \Longrightarrow \mathcal{L}(\mathcal{N}(0,1))$$

uniformly on compact subsets of $\Theta \times \mathcal{G}$.

(ii) *Efficiency :* Among the regular estimates in a particular model any one for which the asymptotic variance is minimum is called an *efficient* estimate in the relevant model.

As pointed out by Bickel and Klaassen (1986) if T_n is regular in Model I and efficient in Model II, then it is efficient in Model I. Thus it is enough to discuss efficiency problems in Model II.

In a thesis, van der Vaart (1987, pp 103-104) has pointed out that this formulation of efficiency is not wholly satisfactory, see in this connection our Remark 5.4.

The estimates introduced by Neyman and Scott (1948) which are analoguous to Huber's M-estimates and referred to as C_1-estimates by Kumon and Amari (1984) are regular both in Model I and Model II.

These estimates are defined as a solution of

$$\sum \psi(X_i, \theta) = 0 \tag{1.1}$$

with the function ψ satisfying certain regularity conditions. More precisely, to ensure uniform asymptotic normality we strengthen the conditions given in Amari and Kumon (1988) as follows :

Definition 1.2 Any Borel measurable map ψ from $S \times \Theta$ to $I\!R$ is called a C_1*-kernel* if

(i) for each x in S, $\psi(x,\cdot)$ is continuously differentiable on Θ, with the derivative given by the function $\psi'(x,\cdot)$ and both $\psi(x,\cdot)$ and $\psi'(x,\cdot)$ have continuous extensions on $\bar{\Theta}$,

(ii) $\int \psi(\cdot,\theta)\,dP_{\theta,\xi} = 0 \;\; \forall(\theta,\xi)$,

(iii) $\sup_{(\theta,\xi)\in\bar{\Theta}\times\Xi} \int 1_{\{|\psi(\cdot,\theta)|\geq\alpha\}}\psi^2(\cdot,\theta)\,dP_{\theta,\xi} \to 0$ as $\alpha \to \infty$

and (iv) (a) $\int \psi'(\cdot,\theta)\,dP_{\theta,\xi} \neq 0 \;\; \forall(\theta,\xi)$

and (b) $\sup_{(\theta,\xi)\in\bar{\Theta}\times\Xi} \int \sup_{\theta'\in\bar{\Theta}} |\psi'(\cdot,\theta')|\,dP_{\theta,\xi} < \infty$.

Given any C_1-kernel ψ, an estimate T_n which is, uniformly on compact subsets of $\Theta \times \Xi$, both $\sqrt{n}$-consistent for θ_o and a solution of (1.1) with probability tending to one, is called a *C_1-estimate corresponding to ψ*.

A brief motivation behind the C_1-estimates and thier asymptotic properties are given in Chapter 2.

For a fixed G_o, according to semiparametric theory, there is a function $\bar{\psi}(\cdot,\cdot,G_o)$ (vide (3.20)-(3.22)) along with an estimate $T_n(G_o)$ which solves

$$\sum_{i=1}^{n} \bar{\psi}(X_i, \theta, G_o) = 0 \tag{1.2}$$

with probability tending to one and is efficient at (θ_o, G_o), provided certain regularity conditions holds. If G_o is unknown, a natural thing to do is to solve

$$\sum_{i=1}^{n} \bar{\psi}(X_i, \theta, \widehat{G}_n) = 0 \tag{1.3}$$

where $\widehat{G}_n$ is a consistent estimate of G_o.

Using a heuristic Taylor series expansion of the L.H.S. of (1.3) w.r.t. θ and G, one can show that (1.3) provides an efficient estimate if $\widehat{G}_n$ is either an $n^{\frac{1}{4}}$-consistent estimate or a consistent estimate independent of the X_i's, $\bar{\psi}$ is a "nice" function of (θ, G) and

$$\int \bar{\psi}(\cdot, \theta, G)\, dP_{\theta,G'} = 0 \ \ \forall(\theta, G, G') \tag{1.4}$$

holds, which is very similar to condition (ii) of Definition 1.2 and plays a similar role.

For the mixture models (1.4) always holds, but unfortunately, in general, it is very difficult to prove the existence of an $n^{\frac{1}{4}}$-consistent estimate of G_o.

In our original work done before the publication of Schick (1986), we were able to resolve the problem only for the examples of Chapters 6 and 7, with stronger regularity conditions than in the present version.

However, the requirement of $n^{\frac{1}{4}}$-consistency of $\widehat{G}_n$ can be dropped by the following idea of Bickel (1982) and Schick (1986), who show how, in effect, one can use an independent estimate of G_o. Thus instead of (1.3) one solves

$$\sum_{i=1}^{n_1} \bar{\psi}(X_i, \theta, G_{n1}) + \sum_{i=n_1+1}^{n} \bar{\psi}(X_i, \theta, G_{n2}) = 0 \tag{1.5}$$

where G_{n1}, G_{n2} are consistent estimates of G_o, G_{n1} is independent of $X_1, X_2, \ldots, X_{n_1}$ and G_{n2} is independent of $X_{n_1+1}, \ldots, X_n$ and $n_1 \to \infty$, $n - n_1 \to \infty$.

It is clear that such a method will also provide an efficient estimate in a general semiparametric problem if a condition like (1.4) holds. This equation can be shown to hold quite generally in models satisfying Bickel's Condition C (vide Remark 3.6) or models considered in Hasminskii and Ibragimov (1983, §3). It seems that our construction of this sort is a part of the folklore of the subject. Certainly a streamlined version of it, using one-step discretized Newton-Raphson methods, seems implicit in Bickel's construction of adaptive estimates in orthogonal cases and has appeared recently in an explicit form in Schick (1986, pp 1142, 1144) who also points out the importance of (1.4). See also van der Vaart (1987), who introduces (1.3) but abandons it in favour of the alternative one-step discretized method. We have given in Chapter 4 both our original version, in which one solves (1.5) leading to an intuitively plausible estimate but requiring stronger conditions as well as the streamlined discretized version of Schick (1986) in which estimates are less easy to interpret.

The corresponding results of Model I are summerized in Chapter 5. The main new feature is that a preliminary randomisation over indices is needed before applying the techniques of Chapter 4. The assumptions for Chapter 5 are *not* stronger than those for Chapter 4.

In Chapter 5 we also indicate briefly (vide Remark 5.6) how the results of Chapters 4 and 5 can be modified, when the dimension of X_i changes with i. Such problems were also first posed by Neyman and Scott (1948). Recent references are Lindsay (1982), Kumon and Amari (1984) and Amari and Kumon (1985), who show how one can get better lower bounds for asymptotic variance of certain proper subclasses of C_1-estimates. Our treatment is quite different in that instead of considering proper subclasses of

C_1-estimates we extend the class to a suitably modified class of regular estimates.

The main conditions that are hard to check are those imposed on $\bar{\psi}$. In the discretized version, $\bar{\psi}$ must be continuous in θ & G and in the other version, one needs also something like differentiability in θ. In our problem, as well as, other semiparametric problems, it is not clear how to check this or even whether such conditions are expected to hold in general. This point is illustrated with an example in Chapter 9. For our problem this can be checked in two special cases illustrated by Examples 1.1 and 1.2, where one either has a special factorization of the density or θ and G are orthogonal in the sense of semiparametric theory. This two cases are discussed in Chapter 6.

It turns out that the conditions on $\bar{\psi}$, can also be checked (via results on compact operators acting on a Banach space) if one has in addition independent observations with distribution G. This allow us to provide a direct application of the results in Chapter 4 to the following problem solved in a different way in Hasminskii and Ibragimov (1983, §3). Suppose one has a channel in which θ is the input, ξ is the noise and $\boldsymbol{X}$ is the observable output. In such a case $\boldsymbol{X}$ will have the distribution in Model II. But while the ξ associated with a particular $\boldsymbol{X}$ will be unknown, one can get independent observations to study directly the distribution of noise. In other words one has in addition to $\boldsymbol{X}_i$'s, independent observations $\boldsymbol{Y}_i$'s which are i.i.d. with distribution G^{q_i}. This problem is solved in Chapter 7.

The results of a simulation study of the grand mean $\bar{\bar{X}}$, the m.l.e. $\hat{\theta}$ and our estimate $T_n(\bar{\psi})$ is included in Chapter 8.

Chapter 2

More on Inconsistency and C_1-estimates

In this chapter, we shall prove the inconsistency of the m.l.e. $\hat{\theta}$ in Example 1.3, that of the Bayes estimates in Example 1.1 and give a brief motivation behind the C_1-estimates.

Assume that $P_{\theta,G}$'s (equivalently, $P_{\theta,\xi}$'s) are dominated by a σ-finite measure μ on $(S, \mathcal{S})$ with the corresponding density function denoted by f. (The formal statement is given in the next chapter.)

For the time being, let us restrict our attention to Model I or the fixed set-up.

Note that, under easy regularity condition on f, one can get, for any θ in Θ, an m.l.e. $\hat{\xi}_i(\theta)$ $(= \hat{\xi}(X_i, \theta))$ for ξ_i, which is a "smooth" solution of

$$\frac{\partial}{\partial \xi} \log\{f(X_i, \theta, \xi)\} = 0 \text{ for } i = 1, 2, \cdots, n. \tag{2.1}$$

The m.l.e. $\hat{\theta}$ will then be a solution of

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log\{f(X_i, \theta, \hat{\xi}_i(\theta))\} = 0. \tag{2.2}$$

For the special case of Example 1.3,

$$\left.\begin{array}{rcl} \dfrac{\partial}{\partial \theta} \log f(x, \theta, \xi) & = & \psi(x) + \dfrac{\partial}{\partial \theta} \log A(\theta, \xi) \\ & = & \psi(x) - E_{\theta,\xi}(\psi(X_1)) \\ & = & \psi(x) - \mu_1(\theta, \xi) \text{ (say)} \\ \text{and } \dfrac{\partial}{\partial \xi} \log f(x, \theta, \xi) & = & x + \dfrac{\partial}{\partial \xi} \log A(\theta, \xi) \\ & = & x - E_{\theta,\xi}(X_1) \\ & = & x - \mu_2(\theta, \xi) \text{ (say)} \end{array}\right\} \forall (x, \theta, \xi). \tag{2.3}$$

From (2.1)-(2.3), we get that the m.l.e. $\hat{\theta}$ of θ is a solution of

$$\sum_{i=1}^{n} \{\psi(X_i) - \mu_1(\theta, \hat{\xi}(X_i, \theta))\} = 0 \tag{2.4}$$

where for any θ in Θ and $1 \leq i \leq n$, $\hat{\xi}(X_i, \theta)$ solves

$$X_i - \mu_2(\theta, \xi) = 0.$$

Let us assume that ψ is continuous. Using the properties of exponential families one can easily get simple regularity conditions on A guaranteeing the continuity of $(x, \theta) \mapsto \mu_1(\theta, \hat{\xi}(x, \theta))$ and the tightness of $\{P_{\theta,G} : (\theta, G) \in \Theta \times \mathcal{G}\}$ so that $\hat{\theta}$ is consistent only if, for any $\epsilon > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} \mu_1(\hat{\theta}, \hat{\xi}(X_i, \hat{\theta})) - \frac{1}{n}\sum_{i=1}^{n} \mu_1(\theta, \hat{\xi}(X_i, \theta)) \overset{\prod_{i=1}^{n} P_{\theta,\xi_i}}{\longrightarrow} 0 \;\; \forall(\theta, \{\xi_i\}_{1 \leq i \leq n}). \quad (2.5)$$

From relations (2.4) and (2.5), we get that $\hat{\theta}$ is consistent only if

$$\frac{1}{n}\sum_{i=1}^{n} \psi(X_i) - \frac{1}{n}\sum_{i=1}^{n} \mu_1(\theta, \hat{\xi}(X_i, \theta)) \overset{\prod_{i=1}^{n} P_{\theta,\xi_i}}{\longrightarrow} 0 \;\; \forall(\theta, \{\xi_i\}_{1 \leq i \leq n}). \quad (2.6)$$

Assume that $E_{\theta,\xi}(\mu_1(\theta, \hat{\xi}(X_1, \theta)))$ is finite for all (θ, ξ). Then, from (2.6) with $\xi_i \equiv \xi$ and SLLN, we get

$$E_{\theta,\xi}(\psi(X_1)) = E_{\theta,\xi}(\mu_1(\theta, \hat{\xi}(X_1, \theta))) \;\; \forall(\theta, \xi).$$

Therefore, by completeness of X_1 and continuity of ψ and $\mu_1(\theta, \hat{\xi}(\cdot, \theta))$,

$$\psi(x) = \mu_1(\theta, \hat{\xi}(x, \theta)) \;\; \forall(x, \theta). \quad (2.7)$$

Again, by definition of $\hat{\xi}$,

$$x = \mu_2(\theta, \hat{\xi}(x, \theta)) \;\; \forall(x, \theta). \quad (2.8)$$

From (2.7) and (2.8) it follows that there exists (θ, ξ) such that

$$\psi(\mu_2(\theta, \xi)) = \mu_1(\theta, \xi). \quad (2.9)$$

But by definition of μ_i's, L.H.S. of (2.9) $= \psi(E_{\theta,\xi}(X_1))$ and R.H.S. of (2.9) $= E_{\theta,\xi}(\psi(X_1))$.

This contradicts the fact that ψ is srtictly concave (convex) and X_1 is non-degenarate. □

Later, in Chapter 4, we shall exihibit a consistent estimate for θ in Example 1.3 under general regularity conditions. However, we don't know whether our recipe for consistent estimates work for this example.

Let us now consider Example 1.1. We know that the ordinary m.l.e. is inconsistent. What happens if we introduce a prior $\pi(\mu)$ for the nuisance parameters such that μ_i's are i.i.d. with common density $\pi(\mu)$, integrate out μ_i's and then maximise with respect to σ? This is often done by the Baysians. We show, for reasonably "smooth" priors, that intregrating out μ_i's gives no better result as far as consistency is concerned, unless we take $\pi(\mu) \equiv$ constant, the improper noninformative prior for this problem.

The integrated likelihood is of the form $L = \prod_{i=1}^n L_i$, with

$$L_i = \frac{e^{-\frac{S_i^2}{2\sigma^2}}}{\sigma^{k-1}} \int_{-\infty}^{\infty} \frac{1}{\sigma} e^{-\frac{k}{2\sigma^2}(\bar{X}_i-\mu_i)^2} \pi(\mu_i)\, d\mu_i$$

where $\bar{X}_i = \frac{1}{k}\sum_{j=1}^k X_{ij}$ and $S_i^2 = \sum_{j=1}^k (X_{ij} - \bar{X}_i)^2$ and our estimate solves the equation

$$\begin{aligned} 0 &= \frac{d\log L}{d\sigma} \\ &= \sum_{i=1}^n \{\frac{(-S_i^2)}{2}\frac{(-2)}{\sigma^3} - \frac{(k-1)}{\sigma} \\ &\quad + \frac{\int_{-\infty}^{\infty} -\frac{k}{2}(\bar{X}_i - \mu_i)^2 \frac{(-2)}{\sigma^3} e^{-\frac{k}{2\sigma^2}(\bar{X}_i-\mu_i)^2}\pi(\mu_i)\,d\mu_i}{\int_{-\infty}^{\infty} \frac{1}{\sigma} e^{-\frac{k}{2\sigma^2}(\bar{X}_i-\mu_i)^2}\pi(\mu_i)\,d\mu_i} - \frac{1}{\sigma}\} \\ &= \sum_{i=1}^n \{\frac{S_i^2}{\sigma^3} - \frac{(k-1)}{\sigma} + \frac{\int_{-\infty}^{\infty} \frac{k}{\sigma^3}(\bar{X}_i - \mu_i)^2 e^{-\frac{k}{2\sigma^2}(\bar{X}_i-\mu_i)^2}\pi(\mu_i)\,d\mu_i}{\int_{-\infty}^{\infty} \frac{1}{\sigma} e^{-\frac{k}{2\sigma^2}(\bar{X}_i-\mu_i)^2}\pi(\mu_i)\,d\mu_i} - \frac{1}{\sigma}\}. \end{aligned} \tag{2.10}$$

But (2.10) is of the form

$$\sum_{i=1}^n \psi(X_i, \theta) = 0 \tag{2.11}$$

where $\theta = \sigma^2$, so that we have, according to Kumon and Amari (1984), a C_0-estimate.

Note the following result containing a sufficient condition for inconsistency of C_0-estimates.

Proposition 2.1 *Let $\hat{\theta}$ be a C_0-estimate. Assume that there is a point (θ, ξ) and $\alpha > 0$ such that*

(i) $E_{\theta,\xi}(\psi(X_1, \theta)) \neq 0$

and (ii) (a) for any x in S, $\psi(x, \cdot)$ is a continuous function on $[\theta - \alpha, \theta + \alpha]$

and (b) there is h in $L_1(P_{\theta,\xi})$ such that

$$|\psi(x, \theta')| \leq h(x) \ \forall (x, \theta') \in S \times [\theta - \alpha, \theta + \alpha].$$

Then there is no consistent solution to (2.11) at (θ, ξ).

Proof : By the uniform strong law of large numbers,

$$\sup_{\theta' \in [\theta - \alpha, \theta + \alpha]} \left| \frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \theta') - E_{\theta,\xi}(\psi(X_1, \theta')) \right| \to 0 \text{ a.e.}[P_{\theta,\xi}]. \qquad (2.12)$$

Also, by condition (ii) and DCT, $E_{\theta,\xi}(\psi_1(X_1, \theta))$ is continuous on $[\theta - \alpha, \theta + \alpha]$. Hence for a suitable neighbourhood N of θ contained in $[\theta - \alpha, \theta + \alpha]$, $E_{\theta,\xi}(\psi(X_1, \cdot))$ is non-zero and have the same sign. Therefore, in view of relation (2.12), with probability one, $\frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \cdot)$ is bounded away from zero on N, for all sufficiently large n. This completes the proof.

□

The following corollary gives conditions on the prior π guaranteeing inconsistency of the Bayes estimates in Example 1.1.

Corollary 2.1.1 *Consider Example 1.1 where $\theta = \sigma^2$ and $\xi = \mu$. Assume that π satisfies the property that the function $\phi_\pi(\boldsymbol{x}, \sigma) := E_{\pi,\sigma^2}((\bar{X}_1 - \mu)^2 | \boldsymbol{X}_1 = \boldsymbol{x})$ is continuous in σ and for any compact subset C of $\mathbb{R}^+$, it is dominated by a polynomial. Then conditions (i)-(ii) of the proposition holds with $\psi = \frac{S^2}{\sigma^3} - \frac{(k-1)}{\sigma} + \frac{k}{\sigma^3} E_{\pi,\sigma^2}((\bar{X}_1 - \mu)^2 | \boldsymbol{X}_1) - \frac{1}{\sigma}$ where $S^2 = \sum_{j=1}^{k} (x_j - \bar{x})^2$ and $\bar{x} = \frac{1}{k} \sum_{j=1}^{k} x_j$, unless $\pi \equiv$ constant.*

Proof : Let π and ψ be as given in the statement of the result. Then one can easily check condition (ii) of Proposition 2.1 for ψ.

We are now going to check condition (i) for it for any prior π which is not a constant. Suppose that condition (i) does not hold.

Then

$$E_{\mu',\sigma^2}\{\frac{E_{\pi,\sigma^2}((\bar{X}_1-\mu)^2|\bar{X}_1)}{\sigma^3}-\frac{1}{\sigma}\}=0 \ \forall(\mu',\sigma). \quad (2.13)$$

Fix any σ in $\mathbb{R}^+$. By condition (ii) and completeness of $\bar{X}_1$, equation (2.13) reduces to the following:

For any x in $\mathbb{R}$,

$$\frac{1}{\sigma^3}E_{\pi,\sigma^2}\{(\bar{X}_1-\mu)^2|\bar{X}_1=x\} = \frac{1}{\sigma}$$

or,

$$\frac{1}{\sigma^3}\int_{-\infty}^{\infty}(x-\mu)^2\frac{\sqrt{k}}{\sqrt{2\pi}\sigma}e^{-\frac{k}{2\sigma^2}(x-\mu)^2}\pi(\mu)\,d\mu = \frac{1}{\sigma}\int_{-\infty}^{\infty}\frac{\sqrt{k}}{\sqrt{2\pi}\sigma}e^{-\frac{k}{2\sigma^2}(x-\mu)^2}\pi(\mu)\,d\mu$$

or,

$$\frac{1}{\sigma^2}\int_{-\infty}^{\infty}(x-\mu)^2e^{-\frac{k}{2\sigma^2}(x-\mu)^2}\pi(\mu)\,d\mu = \int_{-\infty}^{\infty}e^{-\frac{k}{2\sigma^2}(x-\mu)^2}\pi(\mu)\,d\mu = a(x) \text{ (say)}. \quad (2.14)$$

We can rewrite (2.14) as

$$a''(x)=0$$

the general solution of which is of the form $a(x)=cx+d$. Since we must have $a(x)\geq 0$, $c=0$ and hence $a(x)$ is constant. It is well-known and can be proved using Laplace Transform that this implies that $\pi(\mu)\equiv$ constant. This is a contradiction to the choice of π. □

Remark 2.1 One can easily check the condition of the corollary for the normal prior of the form $\pi(\mu)=\frac{1}{\sqrt{2\pi}\eta}e^{-\frac{\mu^2}{2\eta^2}}$ for some $\eta>0$. For Cauchy prior $\pi(\mu)=\frac{1}{\pi(1+\mu^2)}$ or more generally for bounded priors satisfying the additional condition that for $\mu\in[-\delta,\delta]$, $\pi(z-\mu)\geq\frac{1}{p(z)}$ for all z where p denotes a polynomial, one can check the condition as follows :

$$\phi_\pi(x,\sigma) = \frac{\int_{-\infty}^{\infty}\frac{\sqrt{k}}{\sqrt{2\pi}\sigma}(x-\mu)^2e^{-\frac{k(x-\mu)^2}{2\sigma^2}}\pi(\mu)\,d\mu}{\int_{-\infty}^{\infty}\frac{\sqrt{k}}{\sqrt{2\pi}\sigma}e^{-\frac{k(x-\mu)^2}{2\sigma^2}}\pi(\mu)\,d\mu}$$

$$= \frac{\int_{-\infty}^{\infty}(x-\mu)^2 e^{-\frac{k(x-\mu)^2}{2\sigma^2}}\pi(\mu)d\mu}{\int_{-\infty}^{\infty} e^{-\frac{k(x-\mu)^2}{2\sigma^2}}\pi(\mu)d\mu}$$
$$= \frac{p(x,\sigma)}{q(x,\sigma)} \text{ (say)}.$$

Therefore, the continuity of ϕ_π in σ will follow from boundedness of π. Again

$$\begin{aligned}
p(x,\sigma) &= \int_{-\infty}^{\infty}(x-\mu)^2 e^{-\frac{k(x-\mu)^2}{2\sigma^2}}\pi(\mu)d\mu \\
&= \int_{-\infty}^{\infty} u^2 e^{-\frac{ku^2}{2\sigma^2}}\pi(x-u)du \text{ [Putting } u = x-\mu] \\
&\le \|\pi\|_{\sup}\int_{-\infty}^{\infty} u^2 e^{-\frac{ku^2}{2\sigma^2}}du \text{ [By boundedness of } \pi] \quad (2.15) \\
\text{and } q(x,\sigma) &= \int_{-\infty}^{\infty} e^{-\frac{k(x-\mu)^2}{2\sigma^2}}\pi(\mu)d\mu \\
&= \int_{-\infty}^{\infty} e^{-\frac{ku^2}{2\sigma^2}}\pi(x-u)du \\
&\ge \int_{-\delta}^{\delta} e^{-\frac{ku^2}{2\sigma^2}}\pi(x-u)du \ge \left(\int_{-\delta}^{\delta} e^{-\frac{ku^2}{2\sigma^2}}du\right)\frac{1}{p(x)}. \quad (2.16)
\end{aligned}$$

In view of relations (2.15) and (2.16), the conditions of the corollary holds for the given prior π. □

Remark 2.2 Intuitively the reason for the inconsistency is that we are specifying the joint distribution of n parameters, so that the magnitude of "prior information" is comparable to the information in the observations. Hence it is not possible for observations eventually to dominate the prior and direct the posterior to converge towards the true σ. This is similar to the infinite-cell multinomial where typically Bayes estimates are inconsistent (vide Freedman (1963)). See also Diaconis and Freedman (1986) for delicate examples involving unknown symmetric distribution around an unknown location parameter.

Remark 2.3 Professor James O. Berger had wondered whether consistency would be achieved if we take π to be a Cauchy prior, which has long tails. We have seen in Remark 2.1 that consistent Bayes estimation is still impossible.

Remark 2.4 Professor Peter Bickel had wanted to know if a conjugate prior for μ_i with mean λ and variance τ^2 and Empirical Bayes estimation of λ and τ^2 will lead to a consistent empirical Bayes estimate of σ^2. We note the following.

In this case, the log-likelihood function is of the form

$$-\frac{n(k-1)}{2}\log\sigma^2 - \sum_{i=1}^{n}\frac{(k-1)S_i^2}{2\sigma^2} - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\frac{1}{2(\frac{\sigma^2}{k}+\tau^2)}(\bar{X}_i-\lambda)^2.$$

Clearly, $\hat{\lambda} = \bar{\bar{X}}$.

We shall now obtain $\widehat{\sigma^2}$ and $\widehat{\tau^2}$, as follows.

First observe that for any $\sigma^2 > 0$, the m.l.e. $\widehat{\tau^2}(\sigma^2)$ can be obtained by the formula

$$\widehat{\tau^2}(\sigma^2) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(\bar{X}_i-\bar{\bar{X}})^2 - \frac{\sigma^2}{k} & \text{if } 0<\sigma^2 \le \frac{k}{n}\sum_{i=1}^{n}(\bar{X}_i-\bar{\bar{X}})^2 \\ 0 & \text{otherwise.} \end{cases}$$

One can then maximise the log-likelihood evaluated at $\lambda = \bar{\bar{X}}$ and $\tau^2 = \widehat{\tau^2}(\sigma^2)$ to get

$$\begin{aligned} \widehat{\sigma^2} &= \frac{1}{n(k-1)}\sum_{i=1}^{n} S_i^2 \\ \text{and } \widehat{\tau^2} &= \widehat{\tau^2}(\widehat{\sigma^2}), \textit{ i.e.,} \\ \widehat{\tau^2} &= \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(\bar{X}_i-\bar{\bar{X}})^2 - \frac{\widehat{\sigma^2}}{k} & \text{if } \widehat{\sigma^2} \le \frac{k}{n}\sum_{i=1}^{n}(\bar{X}_i-\bar{\bar{X}})^2 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus the empirical Bayes estimate is consistent. This may mean a similar result would be obtained if instead of i.i.d. μ_i's we take a hierarchical prior, or, a special case of that, exchangeable μ_i's. Intuitively, it seems this may be so because the amount of information in such a prior is less than the amount of information in i.i.d's.

We shall now discuss the C_1-estimates. First note that in view of relations (2.1)-(2.2), the m.l.e. is a C_0-estimate. Proposition 2.1 gives sufficient conditions for *inconsistency* of C_0-estimates but these conditions are *not* necessary. A set of sufficient conditions for consistency of C_0-estimates at some point $(\theta, \{\xi_n\}_{n\geq 1})$ are conditions (i)-(ii) and (iv) of Definition 1.2 with Θ replaced by a compact subset C of Θ containing θ.

Following Neyman and Scott (1948), Kumon and Amari (1984) has defined C_1-estimates as the subclass of C_0-estimates which are asymptotically normal on $\Theta \times \Xi$. In Chapters 4 and 5, we shall prove that under easy regularity conditions on f, the C_1-estimates are *regular* both in Model I and Model II (vide Example 4.1 and Remark 5.1).

Chapter 3

Notations and Preliminaries

In this chapter, we shall introduce some notations and give some preliminary assumptions, definitions and results in a form applicable to any semiparametric family involving (θ, G). See in this connection Remark 4.7 in next chapter.

To start with let us introduce some notations which will be used later in appropriate situations :

(1) *Let (X, d) be a metric space. For any x in X and positive number δ, we shall use the symbol $B(x, \delta)$ to denote the open ball of radius δ around the point x. In symbols,*

$$B(x, \delta) := \{y \in X : d(x, y) < \delta\} \;\; \forall x \in X \;\; \forall \delta > 0.$$

(2) *If X is a Banach space, we shall denote the unit sphere around the point zero by $S(X)$. In symbols,*

$$S(X) := \{x \in X : \|x\| = 1\}.$$

(3) *For any real-valued function ϕ on $X \times Y$, we shall denote the extended real-valued functions $\sup_{y \in Y} \phi(\cdot, y)$ and $\inf_{y \in Y} \phi(\cdot, y)$ by $\overline{\phi}(\cdot, Y)$ and $\underline{\phi}(\cdot, Y)$, respectively. In symbols,*

$$\overline{\phi}(x, Y) := \sup_{y \in Y} \phi(x, y) \text{ and } \underline{\phi}(x, Y) := \inf_{y \in Y} \phi(x, y) \text{ for all } x \in X.$$

Similar notations are used for the functions $\sup_{x \in X} \phi(x, \cdot)$, $\inf_{x \in X} \phi(x, \cdot)$ etc..

(4) *For any function $\phi : \Theta \to I\!R$ which is differentiable on Θ, we shall denote the function $\frac{\partial}{\partial \theta}\phi$ by ϕ'. In symbols,*

$$\phi'(\theta) := \frac{\partial}{\partial \theta}\phi(\theta), \text{ for all } \theta \text{ in } \Theta.$$

(5) *Let $X_1, X_2, \ldots, X_m$ be topological spaces one of which, say X_{i_o}, is a closed subset of the real line. Let $\boldsymbol{X} = \prod_{i=1}^m X_i$. Let r_o be a positive integer. Define $\{s_i\}_{1\le i\le m}$ by*

$$s_i = \begin{cases} r_o & \text{if } i = i_o \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \cdots, m$.

Let $C_{s_1,s_2,\ldots,s_m}(\boldsymbol{X})$ be the set of all continuous functions ϕ from $\boldsymbol{X}$ to $\mathbb{R}$ such that for any $1 \le j \le r_o$, $\frac{\partial^j}{\partial x_{i_o}^j}\phi$ exists on $\mathrm{int}(\boldsymbol{X})$ with a continuous extension on $\boldsymbol{X}$.

Remark 3.1 For the special case where $X_1, X_2, \ldots, X_m$ are compact, define $\|\ \|_{s_1,s_2,\ldots,s_m}$ from $C_{s_1,s_2,\ldots,s_m}(\boldsymbol{X})$ to $\mathbb{R}^+$ by

$$\|\phi\|_{s_1,s_2,\ldots,s_m} := \sum_{j=0}^{r_o} \|\frac{\partial^j}{\partial x_{i_o}^j}\phi\|_{\sup} \text{ for all } \phi \text{ in } C_{s_1,s_2,\ldots,s_m}(\boldsymbol{X}).$$

Then one can easily show that

(i) $\|\ \|_{s_1,s_2,\ldots,s_m}$ is a norm on $C_{s_1,s_2,\ldots,s_m}(\boldsymbol{X})$

and (ii) $(C_{s_1,s_2,\ldots,s_m}, \|\ \|_{s_1,s_2,\ldots,s_m})$ is a Banach space.

In practice, we take $\boldsymbol{X}$ to be $\bar{\Theta}$ or $\bar{\Theta} \times \mathcal{G}$ or $S_2 \times \bar{\Theta} \times \mathcal{G}$ or $S_2 \times S_2 \times \bar{\Theta} \times \mathcal{G}$ where S_2 is a compact metric space (vide Model III of Chapter 7), with obvious choice of i_o and $r_o = 1, 2$.

(6) *For any probability space $(\Omega, \mathcal{A}, P)$ we shall denote the space of all square integrable functions whose expectations are zero by $L_2^o(P)$. In symbols,*

$$L_2^o(P) := \{\phi \in L_2(P) : E_P(\phi) = 0\}.$$

Convention : If $P \ll Q, L_2(P) \equiv L_2(\frac{dP}{dQ}), L_2^o(P) \equiv L_2^o(\frac{dP}{dQ})$.

We shall need the following useful definition.

Definition 3.1 Let (Y, ρ) be a metric space. Let ϕ be a continuous map from $\Theta \times \mathcal{G}$ to Y. Call a Y-valued statistic T_n *a uniformly consistent estimate* *of* $\phi(\theta_o, \underline{G}_n)$ *in Model I* ($\phi(\theta_o, G_o)$ *in Model II*) if for any compact subset Θ_o of Θ, $\epsilon > 0$ and $0 < \delta < 1$, there is $N_o \geq 1$ such that for all $n \geq N_o$,

$$\sup_{(\theta_o,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{\rho(T_n, \phi(\theta_o, \underline{G}_n)) > \epsilon\}) < \delta$$
$$(\sup_{(\theta_o,G_o)\in\Theta_o\times\mathcal{G}} P^n_{\theta_o,G_o}(\{\rho(T_n, \phi(\theta_o, G_o)) > \epsilon\}) < \delta).$$

As a special case of the above definition, we can define the notions of uniformly consistent estimates of $\theta_o, \underline{G}_n$ or $(\theta_o, \underline{G}_n)$ in Model I and θ_o, G_o or (θ_o, G_o) in Model II.

Convention : Throughout the following discussion we shall abbreviate the phrase *in Model I (II)* by *(I) ((II))*.

Consider the following generalisation of the Glivenko-Cantelli Lemma.

Proposition 3.1 *Let* $X_1, X_2, \ldots, X_n, \ldots$ *be a sequence of independent random vectors in* $\mathbb{R}^p$, *with* X_i *having the distribution function* F_i, *then, for any* $\epsilon > 0$,

$$\sup_{\{F_k\}_{k=1}^n\in\mathcal{F}^n} P_{F_1,F_2,\cdots,F_n}(\{\|\mathbb{F}_n(\cdot, X_1, X_2, \cdots, X_n) - \frac{1}{n}\sum_{i=1}^{n} F_i\|_{\sup} > \epsilon\}) \to 0$$

as $n \to \infty$, *where* $\mathcal{F}$ *denotes the set of all (probability) distribution functions on* $\mathbb{R}^p$.

One can prove this by an easy modification of the argument in Loéve (1963,p. 20).

As a corollary to Proposition 3.1, we shall now prove, using Robbin's method (Robbins (1964)), the existence of a uniformly consistent estimate of $(\theta_o, \underline{G}_n)$ in Model I and (θ_o, G_o) in Model II.

For this purpose, we shall need the following assumption.

(A1) For each $n \geq 1, \boldsymbol{P}^n_F$ and $\boldsymbol{P}^n_M$ are identifiable families in $(\theta, \underline{G}_n)$ and (θ, G), respectively, where

$$\boldsymbol{P}^n_F := \{\prod_{1}^{n} P_{\theta,\xi_i} : (\theta, \{\xi_i\}_{1\leq i\leq n}) \in \bar{\Theta} \times \Xi^n\}$$
$$\text{and } \boldsymbol{P}^n_M := \{P^n_{\theta,G} : (\theta, G) \in \bar{\Theta} \times \mathcal{G}\}.$$

Let us now state the corollary.

Corollary 3.1.1 *If (i) $(S, \mathcal{S}) = (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, (ii) $P_{\theta,G}$'s are dominated by the Lebesgue measure and (iii) $(\theta, G) \mapsto F(\cdot, \theta, G)$ is continuous, where for any $(\theta, G) \in \bar{\Theta} \times \mathcal{G}$, $F(\cdot, \theta, G)$ denote the distribution function corresponding to $P_{\theta,G}$ and the topology on $\mathcal{F}$, as considered in Proposition 3.1, is generated by the sup-norm, then under assumption (A1) the following holds.*

There is a statistic $(\hat{\theta}_n, \hat{G}_n)$ which is a uniformly consistent estimate of $(\theta_o, \underline{G_n})$ in Model I and (θ_o, G_o) in Model II.

Proof : One uses an idea implicit in Robbins (1964).

Fix $n \geq 1$. Define

$$a_n(\boldsymbol{x}_{pn\times 1}, \theta, G) = \sup_{\boldsymbol{y}\in\mathbb{R}^p} |\mathbb{F}_n(\boldsymbol{y}, \boldsymbol{x}) - F(\boldsymbol{y}, \theta, G)| \text{ for } \boldsymbol{x} \in \mathbb{R}^{pn}, (\theta, G) \in \bar{\Theta} \times \mathcal{G}$$

then

$$\text{(i) } a_n : (\mathbb{R}^{pn} \times \bar{\Theta} \times \mathcal{G}, \mathcal{B}(\mathbb{R}^{pn} \times \bar{\Theta} \times \mathcal{G})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ is measureable}$$

and (ii) for each $\boldsymbol{x} \in \mathbb{R}^{pn}, a_n(\boldsymbol{x}, \cdot, \cdot) \in C(\bar{\Theta} \times \mathcal{G})$.

Therefore, the set

$$D := \{(\boldsymbol{x}, \theta, G) : a_n(\boldsymbol{x}, \theta, G) = \sup_{(\theta', G')\in\bar{\Theta}\times\mathcal{G}} a_n(\boldsymbol{x}, \theta', G')\}$$

is measurable.

So, by von-Neumann selection theorem [vide Theorem 7.2 of Parthasarathy (1972, p. 69)], there is a Borel-measurable map $(\hat{\theta}_n, \hat{G}_n)$ from $\mathbb{R}^{pn}$ to $\bar{\Theta} \times \mathcal{G}$ satisfying

$$a_n(\boldsymbol{x}, \hat{\theta}_n(\boldsymbol{x}), \hat{G}_n(\boldsymbol{x})) = \inf_{(\theta, G)\in\bar{\Theta}\times\mathcal{G}} a_n(\boldsymbol{x}, \theta, G)$$

outside a Lebesgue-null set.

Therefore, $a_n(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n, \hat{\theta}_n(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n), \hat{G}_n(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n))$

$$\leq \begin{cases} a_n((\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n), \theta_o, \underline{G_n}) & \text{in Model I} \\ a_n((\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n), \theta_o, G_o) & \text{in Model II} \end{cases} \tag{3.1}$$

outside a Lebesgue-null set.

But by Proposition 3.1

$$\left.\begin{array}{l} a_n((X_1, X_2, \cdots, X_n), \theta_o, \underline{G}_n) \overset{\prod_1^n P_{\theta_o,\xi_i}}{\longrightarrow} 0 \text{ uniformly on } \bar{\Theta} \times \Xi^n \text{ in Model I} \\ \text{and} \\ a_n((X_1, X_2, \cdots, X_n), \theta_o, G_o) \overset{P^n_{\theta_o,G_o}}{\longrightarrow} 0 \text{ uniformly on } \bar{\Theta} \times \mathcal{G} \text{ in Model II.} \end{array}\right\} \tag{3.2}$$

From (3.1),(3.2) and condition (ii), $\|F(\cdot, \hat{\theta}_n, \hat{G}_n) - F(\cdot, \theta_o, \underline{G}_n)\|_{\sup} \to 0$ uniformly on $\bar{\Theta} \times \Xi^n$ in Model I and $\|F(\cdot, \hat{\theta}_n, \hat{G}_n) - F(\cdot, \theta_o, G_o)\|_{\sup} \to 0$ uniformly on $\bar{\Theta} \times \mathcal{G}$ in Model II.

Let us now observe that assumption (A1) and condition (iii) together imply that the inverse map $F(\cdot, \theta, G) \mapsto (\theta, G)$ is well-defined and compactness of $\bar{\Theta} \times \mathcal{G}$ implies that it is continuous. The rest is easy. □

Remark 3.2 Note that the boundedness of Θ is needed only to prove the continuity of the inverse map $F(\cdot, \theta, G) \mapsto (\theta, G)$.

Remark 3.3 It is interesting to note that the null set of Corollary 3.1.1 can be dropped in the following manner. First note that the compactness of $\bar{\Theta} \times \mathcal{G}$ and continuity of $a_n(\boldsymbol{x}, \cdot, \cdot)$ for all $\boldsymbol{x}$ together imply that the $\boldsymbol{x}$-sections of D are compact. Next apply Corollary 3 of Maitra and Rao (1975) to get the required selection. See also Theorem 4.4.3 of Srivastava (1982, p. 106).

Convention : For any $k \geq 1$ such that $\Theta_k := \Theta \bigcap [k, k+1] \neq \emptyset$, we shall use the notation $(\tilde{\theta}_n(k), \tilde{G}_n(k))$ to denote the minimum distance estimates considered in Corollary 3.1.1 for the models

$$\begin{aligned} \boldsymbol{P}_F^{n,k} &:= \{\prod_1^n P_{\theta,\xi_i} : (\theta, \{\xi_i\}_{1 \leq i \leq n}) \in \bar{\Theta}_k \times \Xi^n\} \\ \text{and } \boldsymbol{P}_M^{n,k} &:= \{P^n_{\theta,G} : (\theta, G) \in \bar{\Theta}_k \times \mathcal{G}\}. \end{aligned}$$

The following result shows that we can drop the condition of compactness of $\bar{\Theta}$ at the cost of the condition of existence of a uniformly consistent estimate of θ_o.

Corollary 3.1.2 *Consider Model I and Model II, as defined in Chapter 1, with the only exception that Θ is allowed to be unbounded. Assume (A1). If* *conditions (i)-(iii) of Corollary 3.1.1 hold and there is an estimate of T_n of θ_o which is uniformly consistent in Model I (II), then there is a uniformly consistent estimate $(\hat{\theta}_n, \hat{\hat{G}}_n) := (\tilde{\theta}_n([T_n]), (\tilde{G}_n([T_n]))$ of $(\theta_o, \underline{G}_n)$ in Model I ((θ_o, G_o) in Model II).*

Proof : Let Θ_o be a given compact subset of the Θ. Let $0 < \delta < 1$ and $\epsilon > 0$ be given. We want to show that there is $N \geq 1$ such that for all $n \geq N$,

$$\begin{aligned} &\sup_{(\theta,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta,\xi_i})(\{|\hat{\theta}_n - \theta| + d(\hat{\hat{G}}_n, \underline{G}_n) > \epsilon\}) < \delta \\ &(\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta,G}(\{|\hat{\theta}_n - \theta| + d(\hat{\hat{G}}_n, G) > \epsilon\}) < \delta). \end{aligned} \tag{3.3}$$

Fix an η in the open interval $(0, 0.5)$. Using uniform consistency of T_n choose and fix $N_0 \geq 1$ such that for any $n \geq N_0$,

$$\begin{aligned} &\sup_{(\theta,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta,\xi_i})(\{|T_n - \theta| > \eta\}) < \delta/2 \\ &(\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta,G}(\{|T_n - \theta| > \eta\}) < \delta/2). \end{aligned} \tag{3.4}$$

Let us now observe that by compactness of Θ_o, there are integers k and l with $l \geq 1$, such that $\Theta_o \subseteq \bigcup_{j=1}^{l} \Theta_{k-j+1}$.

Define

$$l_o := \min\{l \geq 1 : \exists k \in \mathbb{Z} \ni \Theta_o \subseteq \bigcup_{j=1}^{l} \Theta_{k-j+1}\}. \tag{3.5}$$

Then, there is a unique integer k_o such that $\Theta_o \subseteq \bigcup_{j=1}^{l_o} \Theta_{k_o-j+1}$.

Using Corollary 3.1.1 choose and fix $N_1 \geq 1$ such that for any $n \geq N_1$,

$$\begin{aligned} &\sup_{1\leq j\leq l_o} \sup_{(\theta,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta,\xi_i})(\{|\tilde{\theta}_n(k_o - j + 1) - \theta| + \\ &\qquad d(\tilde{G}_n(k_o - j + 1), \underline{G}_n) > \epsilon\}) < \delta/8 \\ &(\sup_{1\leq j\leq l_o} \sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta,G}(\{|\tilde{\theta}_n(k_o - j + 1) - \theta| + \\ &\qquad d(\tilde{G}_n(k_o - j + 1), G) > \epsilon\}) < \delta/8). \end{aligned} \tag{3.6}$$

Let $N = N_0 \vee N_1$.

Then, for all $n \geq N$,

L.H.S. of (3.3)

$$= \sup_{(\theta,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta,\xi_i})(\{|\hat{\theta}_n - \theta| + d(\hat{\widetilde{G}}_n, \underline{G}_n) > \epsilon\})$$

$$\stackrel{(3.5)}{\leq} \sup_{(\theta,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta,\xi_i})(\{T_n \notin ([\theta]-1-\eta, [\theta]+1+\eta)\})$$

$$+ \sum_{j=0}^{3} \sup_{(\theta,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta,\xi_i})(\{|\tilde{\theta}_n([\theta]-2+j) - \theta| + d(\widetilde{G}_n([\theta]-2+j), \underline{G}_n) > \epsilon\})$$

$$\stackrel{(3.6)}{\leq} \sup_{(\theta,\{\xi_i\}_{1\leq i\leq n})\in\Theta_o\times\Xi^n} (\prod_{i=1}^{n} P_{\theta,\xi_i})(\{|T_n - \theta| > \eta\}) + 4.\delta/8$$

$$(= \sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta,G}(\{|\hat{\theta}_n - \theta| + d(\hat{\widetilde{G}}_n, G) > \epsilon\})$$

$$\stackrel{(3.5)}{\leq} \sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta,G}(\{T_n \notin ([\theta]-1-\eta, [\theta]+1+\eta)\})$$

$$+ \sum_{j=0}^{3} \sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta,G}(\{|\tilde{\theta}_n([\theta]-2+j) - \theta| + d(\widetilde{G}_n([\theta]-2+j), G) > \epsilon\})$$

$$\stackrel{(3.6)}{\leq} \sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta,G}(\{|T_n - \theta| > \eta\}) + 4.\delta/8)$$

$$\stackrel{(3.4)}{\leq} \delta/2 + \delta/2 = \delta$$

proving (3.3). □

Remark 3.4 In view of Remark 3.3, we can drop the condition (ii) from Corollaries 3.1.1 and 3.1.2.

We shall also need the following definitions.

Definition 3.2 Call an estimate T_n of θ_o a *uniformly $\sqrt{n}$-consistent estimate of θ_o in Model I (II)* if for any compact subset Θ_o of Θ the family of laws

$$\{\mathcal{L}(\sqrt{n}(T_n - \theta_o)|\prod_{i=1}^{n} P_{\theta_o,\xi_i}) : (\theta_o, \{\xi_i\}_{1\leq i\leq n}) \in \Theta_o \times \Xi^n, n \geq 1\}$$

$$(\{\mathcal{L}(\sqrt{n}(T_n - \theta_o)|P^n_{\theta_o,G_o}) : (\theta_o, G_o) \in \Theta_o \times \mathcal{G}, n \geq 1\})$$

is tight.

Definition 3.3 Let T_n be an estimate of θ_o and let Ψ be a Borel-measureable map from $S^n \times \Theta$ to $I\!R$. Consider the equation

$$\Psi((X_1, X_2, \cdots, X_n), \theta) = 0. \tag{3.7}$$

(a) Call T_n a *$\sqrt{n}$-consistent solution of (3.7) in Model I (II)* if for any $(\theta_o, \{\xi_i\}_{1 \leq i \leq n}) \in \Theta_o \times \Xi^n$ $((\theta_o, G_o) \in \Theta_o \times \mathcal{G})$ the following hold.

(i) $(\prod_{i=1}^{n} P_{\theta_o, \xi_i})(T_n \text{ solves } (3.7)) = 1 + o(1)$

$(P^n_{\theta_o, G_o}(T_n \text{ solves } (3.7)) = 1 + o(1))$

and (ii) T_n is a $\sqrt{n}$-consistent estimate of θ_o in Model I (II).

(b) Call T_n a *uniformly $\sqrt{n}$-consistent solution of (3.7) in Model I (II)* if for any compact subset Θ_o of Θ, condition (a) holds uniformly on $\Theta_o \times \Xi^n$ $(\Theta_o \times \mathcal{G})$.

Definition 3.4 Call an estimate T_n of θ_o *regular (I) ((II))*, or, more accurately, *uniformly asymptotically normal in Model I (II) with asymptotic variance* σ_T^2 [in short, *UAN (I) ((II)) with AV* σ_T^2], where σ_T is a continuous function from $\Theta \times \mathcal{G}$ to $I\!R$, if

$$\sup_{(\theta_o, \{\xi_i\}_{1 \leq i \leq n}) \in \Theta_o \times \Xi^n} \sup_{x \in I\!R} |(\prod_{i=1}^{n} P_{\theta_o, \xi_i})(\{\sqrt{n}(T_n - \theta_o) \leq x\}) - \Phi(x \sigma_T^{-1}(\theta_o, \underline{G}_n))|$$

$$(\sup_{(\theta_o, G_o) \in \Theta_o \times \mathcal{G}} \sup_{x \in I\!R} |P^n_{\theta_o, G_o}(\{\sqrt{n}(T_n - \theta_o) \leq x\}) - \Phi(x \sigma_T^{-1}(\theta_o, G_o))|)$$

$\to 0$ as $n \to \infty$, for any compact subset Θ_o of Θ.

Note that

1) As expected, for any concept defined through Definitions 3.1-3.4, the Model I-version is stronger than the Model II-version.

Let us now state and prove two auxiliary results followed by a generalised version of the Lindeberg-Feller central limit theorem where the convergence is uniform in sup-norm. We shall need the last result in the proof of our basic result Lemma 4.1.

Lemma 3.2 *Let A be a nonempty set. Consider the following two families of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

$$\begin{aligned} \mathbb{P}_\infty &:= \{P_n(\cdot,\alpha) : \alpha \in A, n \geq 1\} \\ \text{and } \mathbb{P} &:= \{P(\cdot,\alpha) : \alpha \in A\}. \end{aligned}$$

Assume that the following conditions hold.

(i) $\mathbb{P}_\infty$ is tight,

(ii) $\mathbb{P}$ is tight as well as uniformly absolutely continuous with respect to the Lebesgue measure

and (iii) for any bounded continuous function g from $\mathbb{R}$ to $\mathbb{R}$

$$\sup_{\alpha\in A} \left| \int g(\cdot)dP_n(\cdot,\alpha) - \int g(\cdot)dP(\cdot,\alpha) \right| \to 0 \text{ as } n \to \infty.$$

Then,

$$\sup_{\alpha\in A} \sup_{x\in\mathbb{R}} |F_n(x,\alpha) - F(x,\alpha)| \to 0 \text{ as } n \to \infty \tag{3.8}$$

where $F_n(\cdot,\alpha)$ and $F(\cdot,\alpha)$ denote the distribution functions corresponding to $P_n(\cdot,\alpha)$ and $P(\cdot,\alpha)$, respectively.

Proof : Let us first show that for any x in $\mathbb{R}$,

$$\sup_{\alpha\in A} |F_n(x,\alpha) - F(x,\alpha)| \to 0 \text{ as } n \to \infty. \tag{3.9}$$

Let $\epsilon > 0$ be given. Using uniform absolute continuity of $\mathbb{P}$ choose and fix $\delta > 0$ such that

$$\sup_{\alpha\in A} |F(x+\delta,\alpha) - F(x-\delta,\alpha)| < \epsilon/4 \tag{3.10}$$

Define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(y) = \begin{cases} 1 & \text{if } y \leq x-\delta \\ \frac{x+\delta-y}{2\delta} & \text{if } x-\delta \leq y \leq x+\delta \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, g is a bounded continuous function from $\mathbb{R}$ to $\mathbb{R}$. Therefore by condition (iii), there is $n_1 \geq 1$ such that for all $n \geq n_1$,

$$\sup_{\alpha\in A} \left| \int g(\cdot)dP_n(\cdot,\alpha) - \int g(\cdot)dP(\cdot,\alpha) \right| < \epsilon/4. \tag{3.11}$$

Therefore, for $n \geq n_1$,

$$
\begin{aligned}
\text{L.H.S. of (3.9)} &= \sup_{\alpha\in A} |F_n(x,\alpha) - F(x,\alpha)| \\
&= \sup_{\alpha\in A} \left| \int 1_{(-\infty,x]}(\cdot)dP_n(\cdot,\alpha) - \int 1_{(-\infty,x]}(\cdot)dP(\cdot,\alpha) \right| \\
&= \sup_{\alpha\in A} \left| \int \{1_{(-\infty,x]}(\cdot) - g(\cdot)\}dP_n(\cdot,\alpha) \right. \\
&\qquad - \int \{1_{(-\infty,x]}(\cdot) - g(\cdot)\}dP(\cdot,\alpha) \\
&\qquad \left. + \int g(\cdot)dP_n(\cdot,\alpha) - \int g(\cdot)dP(\cdot,\alpha) \right| \\
&\overset{(3.11)}{\leq} \sup_{\alpha\in A} \left[\int |1_{(-\infty,x]}(\cdot) - g(\cdot)| dP_n(\cdot,\alpha) \right] \\
&\qquad + \sup_{\alpha\in A} \left[\int |1_{(-\infty,x]}(\cdot) - g(\cdot)| dP(\cdot,\alpha) \right] + \epsilon/4 \\
&\overset{(3.10)}{\leq} \sup_{\alpha\in A} \left[\int h(\cdot,\alpha) dP_n(\cdot,\alpha) \right] + \epsilon/4 + \epsilon/4 \qquad (3.12)
\end{aligned}
$$

where

$$
h(y) = \begin{cases} 0 & \text{if } |y-x| \geq \delta \\ \frac{\delta - |y-x|}{2\delta} & \text{otherwise.} \end{cases}
$$

Clearly, h is also a bounded continuous function from $\mathbb{R}$ to $\mathbb{R}$. Hence, by condition (iii), let us choose and fix an $n_2 \geq 1$ such that for all $n \geq n_2$,

$$
\sup_{\alpha\in A} \left| \int h(\cdot)dP_n(\cdot,\alpha) - \int h(\cdot)dP(\cdot,\alpha) \right| < \epsilon/4. \qquad (3.13)
$$

Let $n_o = n_1 \vee n_2$. Then for any $n \geq n_o$,

$$
\begin{aligned}
\text{L.H.S. of (3.9)} &\overset{(3.12)}{<} \sup_{\alpha\in A} \left| \int h(\cdot)dP_n(\cdot,\alpha) - \int h(\cdot)dP(\cdot,\alpha) \right| \\
&\qquad + \sup_{\alpha\in A} \left[\int h(\cdot)dP(\cdot,\alpha) \right] + \epsilon/2 \\
&\overset{(3.10)\text{ and }(3.13)}{<} \epsilon/4 + \epsilon/4 + \epsilon/2 = \epsilon
\end{aligned}
$$

proving (3.9).

Let us now prove relation (3.8) from relation (3.9).

Using tightness of $\mathbb{P}_\infty$ and $\mathbb{P}$, choose and fix $K > 0$ such that

$$
\left.
\begin{array}{ll}
\sup_{n\geq 1} \sup_{\alpha\in A} F_n(-K,\alpha) < \epsilon/4, & \sup_{\alpha\in A} F(-K,\alpha) < \epsilon/4, \\
\inf_{n\geq 1} \inf_{\alpha\in A} \{1 - F_n(K,\alpha)\} < \epsilon/4 \text{ and } & \inf_{\alpha\in A} \{1 - F(K,\alpha)\} < \epsilon/4.
\end{array}
\right\} \qquad (3.14)
$$

Then,

$$\begin{aligned}
& \sup_{\alpha\in A}\sup_{|x|\geq K}|F_n(x,\alpha)-F(x,\alpha)| \\
\leq\ & \sup_{\alpha\in A}\sup_{x\leq -K}|F_n(x,\alpha)-F(x,\alpha)|+\sup_{\alpha\in A}\sup_{x\geq K}|F_n(x,\alpha)-F(x,\alpha)| \\
\leq\ & \max\{\sup_{\alpha\in A}\sup_{n\geq 1}F_n(-K,\alpha),\sup_{\alpha\in A}F(-K,\alpha)\} \\
& +\max[\sup_{\alpha\in A}\sup_{n\geq 1}\{1-F_n(K,\alpha)\},\sup_{\alpha\in A}\{1-F(K,\alpha)\}] \\
\overset{(3.14)}{<}\ & \epsilon/4+\epsilon/4=\epsilon/2. \qquad (3.15)
\end{aligned}$$

Using uniform absolute continuity of $I\!P$ and compactness of $[-K,K]$ choose and fix $m\geq 1$ such that

$$\sup_{\alpha\in A}\sup_{i=0,1,\cdots,2m-1}|F(x_{i+1},\alpha)-F(x_i,\alpha)|<\epsilon/4 \qquad (3.16)$$

where $x_i=-K+\frac{iK}{m}=(\frac{i-m}{m})K$ for $i=0,1,\cdots,2m$.

Using (3.9) choose and fix $N\geq 1$ such that $n\geq N$ implies

$$\sup_{\alpha\in A}\sup_{i=0,1,\cdots,2m}|F_n(x_i,\alpha)-F(x_i,\alpha)|<\epsilon/4. \qquad (3.17)$$

Then, for $n\geq N$,

$$\begin{aligned}
& \sup_{\alpha\in A}\sup_{|x|\leq K}|F_n(x,\alpha)-F(x,\alpha)| \\
\leq\ & \sup_{\alpha\in A}\sup_{i=0,1,\cdots,2m}|F_n(x_i,\alpha)-F(x_i,\alpha)| \\
& +\sup_{\alpha\in A}\sup_{i=0,1,\cdots,2m-1}|F(x_{i+1},\alpha)-F(x_i,\alpha)| \\
\overset{(3.16)\text{ and }(3.17)}{<}\ & \epsilon/4+\epsilon/4=\epsilon/2. \qquad (3.18)
\end{aligned}$$

From (3.15) and (3.18) it follows that, for any $n\geq N$,

$$\sup_{\alpha\in A}\sup_{x\in I\!R}|F_n(x,\alpha)-F(x,\alpha)|<\epsilon$$

proving (3.8). □

Lemma 3.3 *(Theorem 7 of Ibragimov and Hasminskii (1981, p. 365)). Let A, $I\!P_\infty$ and $I\!P$ be as in Lemma 3.2. Assume that the following conditions hold.*

(i) $\mathbb{P}_\infty$ *is tight*

and (ii) $\sup_{\alpha\in A} |\int e^{itx} dP_n(x,\alpha) - \int e^{itx} dP(x,\alpha)| \to 0$ *as* $n \to \infty$.

Then, for any bounded continuous function g from $\mathbb{R}$ *to* $\mathbb{R}$

$$\sup_{\alpha\in A} |\int g(x) dP_n(x,\alpha) - \int g(x) dP(x,\alpha)| \to 0 \text{ as } n \to \infty.$$

A proof of this result is given in Ibragimov and Hasminskii (1981, pp 365-366).

Proposition 3.4 *Let A be a non-empty set. For each* α *in A, let* $\{X_n(\alpha)\}_{n\geq 1}$ *be a sequence of independent random variables with mean zero and finite variance. For each* α *in A and* $n \geq 1$, *define* $S_n(\alpha)$ *and* $s_n(\alpha)$ *by*

$$S_n(\alpha) := \sum_{i=1}^{n} X_i(\alpha) \text{ and } s_n(\alpha) := \sqrt{\mathrm{Var}(\{S_n(\alpha)\})} = \sqrt{\sum_{i=1}^{n} \mathrm{Var}(\{X_i(\alpha)\})}$$

and denote the probability distribution functions induced by $X_n(\alpha)$ *and* $S_n(\alpha)/s_n(\alpha)$ *by* $G_n(\cdot,\alpha)$ *and* $F_n(\cdot,\alpha)$, *respectively. If*

(i) $\inf_{\alpha\in A} \liminf_{n\to\infty} [\int x^2 d\bar{G}_n(x,\alpha)] > 0$

and (ii) $\sup_{\alpha\in A} \limsup_{n\to\infty} [\int_{|x|\geq K} x^2 d\bar{G}_n(x,\alpha)] \to 0$ *as* $K \to \infty$

where $\bar{G}_n := \frac{1}{n}\sum_{i=1}^{n} G_i$, *for all* $n \geq 1$, *then*

$$\sup_{\alpha\in A} \sup_{x\in\mathbb{R}} |F_n(x,\alpha) - \Phi(x)| \to 0 \text{ as } n \to \infty.$$

Proof : For any α in A, let $P_n(\cdot,\alpha)$ stand for the probability measure corresponding to the distribution $F_n(\cdot,\alpha)$. Then condition (i), which is common to both Lemmas 3.2 and 3.3, follows from the definition of $F_n(\cdot,\alpha)$'s. Next, condition (ii) of Lemma 3.3 follows from conditions (i)-(ii) of the proposition and the definition of $F_n(\cdot,\alpha)$'s by an application of a uniform version of the proof of Theorem 2.7.2 of Billingsley (1979, pp 310-312). Again, $\mathbb{P}$ being a singleton containing the standard normal probability measure, condition (ii) of Lemma 3.2 holds for it. The proposition follows by an application of Lemma 3.3 followed by Lemma 3.2. □

Note that

2) Instead of assuming the obvious uniform version of the Lindeberg's condition, we are assuming a stronger but more easily verifiable pair of conditions, *viz.*, conditions (i) and (ii).

We shall need one more definition.

Definition 3.5 We shall call a function $\psi : S \times \bar{\Theta} \times \mathcal{G} \to I\!R$ a *kernel* if $\psi(\cdot, \theta, G) \in L_2^o(P_{\theta,G})$ for all (θ, G) in $\bar{\Theta} \times \mathcal{G}$, and denote the set of all kernels by K.

Convention : Given any two kernels ψ, ψ' such that

$$\psi(\cdot, \theta, G) = \psi'(\cdot, \theta, G) \text{ a.e.}[P_{\theta,G}] \ \forall(\theta, G),$$

we shall call each a ***version*** of the other one.

Consider the following assumption

(A2) There is a σ-finite measure μ on $(S, \mathcal{S})$ such that

$$P_{\theta,G} \ll \mu \ \forall(\theta, G) \in \bar{\Theta} \times \mathcal{G}.$$

Define $f : \bar{\Theta} \times \mathcal{G} \to L_1^+(\mu)$ by

$$f(\cdot, \theta, G) := \frac{dP_{\theta,G}}{d\mu} \ \forall(\theta, G) \in \bar{\Theta} \times \mathcal{G}. \tag{3.19}$$

Note that

3) In available semiparametric literatures, (A2) is always assumed. So, we shall assume it for the remaining part of the thesis, with the only exception of Chapter 7 where this condition will be dropped.

4) For the special case of the mixture models, (A2) is equivalent to

$$P_{\theta,\xi} \ll \mu \ \forall(\theta, \xi) \in \bar{\Theta} \times \Xi.$$

Convention : For $\xi \in \Xi$, we shall use the notations $f(\cdot, \theta, \delta_\xi)$ and $f(\cdot, \theta, \xi)$, interchangeably, where δ_ξ denote the point mass at $\{\xi\}$.

From the general semiparametric theory, the θ-score $s_\theta : S \times \bar{\Theta} \times \mathcal{G} \to I\!R$ should be defined by

$$s_\theta(x, \theta, G) := \frac{f'(x, \theta, G)}{f(x, \theta, G)} 1_{\{f(\cdot,\theta,G)>0\}}(x) \ \forall(x, \theta, G) \in S \times \bar{\Theta} \times \mathcal{G}. \tag{3.20}$$

Under the following assumption s_θ is well-defined and belongs to K.

(A3) (a) For each $(x, G) \in S \times \mathcal{G}$, $f(x, \cdot, G) \in C_1(\bar{\Theta})$ and (b) $\int \frac{(f')^2(\cdot,\theta,G)}{f(\cdot,\theta,G)} d\mu(\cdot) < \infty \; \forall(\theta, G) \in \bar{\Theta} \times \mathcal{G}$.

In passing, we remark that, (A3) will be assumed to hold throughout the remaining part of the thesis.

Let us now observe that f has an obvious extension on $\bar{\Theta} \times \mathcal{M}$, where $\mathcal{M}$ denotes the set of all signed measures on Ξ. Let us denote this extension also by f.

(7) *From now on, we shall denote by Λ the extension of the likelihood ratio statistic defined by*

$$\Lambda(x, \theta, G, \theta', M) := \frac{f(x, \theta', M)}{f(x, \theta, G)} 1_{\{f(\cdot,\theta,G)>0\}}(x)$$

for all $(x, \theta, G, \theta', M) \in S \times \bar{\Theta} \times \mathcal{G} \times \bar{\Theta} \times \mathcal{M}$.

Consider $\mathcal{M}_o := \{M \in \mathcal{M} : M(\Xi) = 0\}$.

For any $(\theta, G) \in \Theta \times \mathcal{G}$, define

$$\begin{aligned} \mathcal{M}_{\theta,G} &= \{M \in \mathcal{M}_o : \Lambda(\cdot, \theta, G, \theta, M) \in L_2^o(P_{\theta,G}) \text{ and} \\ &\qquad \int 1_{\{f(\cdot,\theta,G)=0\}} f(\cdot, \theta, M) d\mu(\cdot) = 0\} \\ \text{and } N_{\theta,G} &= \{\phi \in L_2^o(P_{\theta,G}) : \exists M \in \mathcal{M}_{\theta,G} \text{ such that} \\ &\qquad \phi = \Lambda(\cdot, \theta, G, \theta, M) \text{ a.e. } [P_{\theta,G}]\}. \end{aligned} \quad (3.21)$$

The elements of the space $N_{\theta,G}$ may be thought of as the 'directional scores' with respect to small variations in G.

Remark 3.5 Under assumptions (A2)-(A3), for each $(\theta, G) \in \Theta \times \mathcal{G}$, the closed linear subspace of $L_2^o(f(\cdot, \theta, G))$ obtained by taking the closure of the linear span of $s_\theta(\cdot, \theta, G)$ and $N_{\theta,G}$ gives our tangent space $\boldsymbol{T}_{\theta,G}$ at (θ, G), which is isometric to that considered in Schick (1986) and is the same as that considered in Lindsay (1980), Bickel(1982), Bickel and Klaassen (1986) or van der Vaart (1987).

Following the above authors, let us now define an *optimal kernel* ($\bar{\psi}$) and the information (I) by

$$\left.\begin{array}{rcl}\bar{\psi}(\cdot,\theta,G) & := & \mathrm{Proj}_{N^{\perp}_{\theta,G}}\{s_{\theta}(\cdot,\theta,G)\} \\ I(\theta,G) & := & \|\bar{\psi}(\cdot,\theta,G)\|^2_{L_2(f(\cdot,\theta,G))}\end{array}\right\} \quad \forall(\theta,G). \tag{3.22}$$

We shall now establish (1.4) for the general semiparametric models as well as the mixture models under different regularity conditions. Let us now write down these coditions in the form of two assumptions.

(GA4) $\int \frac{f^2(\cdot,\theta,M)}{f(\cdot,\theta,G)} d\mu(\cdot) < \infty$ for all $(\theta,G,M) \in \bar{\Theta} \times \mathcal{G} \times \mathcal{M}_o$.

(A4) For any $\theta \in \bar{\Theta}, (G,G') \mapsto \int \bar{\psi}(\cdot,\theta,G) f(\cdot,\theta,G') d\mu(\cdot)$ is a continuous map from $\mathcal{G} \times \mathcal{G}$ to $I\!R$.

We are now in a position to state the following result.

Lemma 3.5 *Consider (a) an arbitrary semiparametric model where (A2), (A3) and (GA4) hold or (b) a mixture model where (A2)-(A4) hold. In either case,*

$$\int \bar{\psi}(\cdot,\theta,G) f(\cdot,\theta,G') d\mu(\cdot) = 0 \; \forall(\theta,G,G'). \tag{3.23}$$

Proof : Let us start with the following observation which is an obvious consequence of the fact that $\bar{\psi}$ is a kernel.

$$\int \bar{\psi}(\cdot,\theta,G) f(\cdot,\theta,G) d\mu(\cdot) = 0 \; \forall(\theta,G). \tag{3.24}$$

So, it remains to show

$$\int \bar{\psi}(\cdot,\theta,G) f(\cdot,\theta,G'-G) d\mu(\cdot) = 0 \; \forall(\theta,G,G'). \tag{3.25}$$

For the general semiparametric models use (GA4) to conclude that $A(\cdot,\theta,G,\theta,G'-G) \in N_{\theta,G} \; \forall(\theta,G,G')$. Then (3.25) follows from the fact that $\bar{\psi}(\cdot,\theta,G) \in N^{\perp}_{\theta,G} \; \forall(\theta,G)$.

For the mixture models, let us observe that, for any (θ,G,ϕ) with $(\theta,G) \in \bar{\Theta} \times \mathcal{G}$ and $\phi \in L_2^o(G)$, $A(\cdot,\theta,G,\theta,\phi dG) \in N_{\theta,G}$ proving

$$\int \bar{\psi}(\cdot,\theta,G) f(\cdot,\theta,\phi dG) d\mu(\cdot) = 0 \tag{3.26}$$

for all (θ,G,ϕ) with $(\theta,G) \in \bar{\Theta} \times \mathcal{G}$ and $\phi \in L_2^o(G)$.

Now for any σ-finite measure ν on $(\Xi, \mathcal{B}(\Xi))$, let $\boldsymbol{g}_\nu$ denote the set of all probability density functions with respect to ν those are bounded and bounded away from zero. For any $g \in \boldsymbol{g}_\nu$, let G denote the corresponding probability measure and denote the set of all such G's by $\boldsymbol{G}_\nu$, *i.e.*,

$$\boldsymbol{G}_\nu := \{G : g \in \boldsymbol{g}_\nu\}.$$

Let us now consider $(\theta, G, G') \in \bar{\Theta} \times \mathcal{G} \times \mathcal{G}$. Define $\nu = \frac{G+G'}{2}$.

Case I: $G, G' \in \boldsymbol{G}_\nu$. Let g, g' be versions of $\frac{dG}{d\nu}, \frac{dG'}{d\nu}$ which belong to $\boldsymbol{g}_\nu$. Put $\phi = \frac{g'-g}{g}$ in (3.26). By an easy algebra one can show that $f(\cdot, \theta, \phi dG) = f(\cdot, \theta, G' - G)$, so that (3.25) holds for the given point (θ, G, G').

Case II: G, G' arbitrary. Let g, g' be any two versions of $\frac{dG}{d\nu}$ and $\frac{dG'}{d\nu}$, respectively. One can get two sequences $\{g_n\}_{n \geq 1}$ and $\{g'_n\}_{n \geq 1}$ of functions in $\boldsymbol{g}_\nu$ such that $\|g_n - g\|_{L_1(\nu)} \to 0$ and $\|g'_n - g'\|_{L_1(\nu)} \to 0$. Clearly this implies that $G_n \overset{w}{\Longrightarrow} G$ and $G'_n \overset{w}{\Longrightarrow} G'$. Again by Case I, (3.25) holds for (θ, G_n, G'_n), for any $n \geq 1$. Hence by assumption (A4), (3.25) holds for (θ, G, G'). □

Remark 3.6 Note that Lemma 3.5(a) holds for general semiparametric models satisfying Bickel's Condition C, *i.e.*, semiparametric models with the space $\mathcal{G}$ of nuisance parameters G convex and the density function f affine in G, with the additional condition that $\mathcal{G}$ is compact. The corresponding result for the orthogonal case was noted by Bickel (1982), vide his remark before Conditions C and S*.

In order that (1.2) makes sense, let us make the following assumption which is a local version of (A1).

(A5) $I(\theta, G) > 0$ for all (θ, G) in $\Theta \times \mathcal{G}$.

For the next two chapters, we recall Definition 3.5 and introduce the following notations.

(8) *Let* $K^* := \{\psi \in K : P_{\theta,G}(\{|\psi(\cdot,\theta,G)| > 0\}) > 0 \ \forall(\theta,G)\}$. *We shall denote by* J *the function from* $\Theta \times \mathcal{G} \times K^*$ *to* $\mathbb{R}$ *defined by,*

$$J(\theta.G,\psi) := \frac{[\int \psi(\cdot,\theta,G) f'(\cdot,\theta,G) d\mu(\cdot)]^2}{[\int \psi^2(\cdot,\theta,G) f(\cdot,\theta,G) d\mu(\cdot)]}$$

for all $(\theta, G, \psi) \in \Theta \times \mathcal{G} \times K^*$.

(9) *Let* $K^{**} := \{\psi \in K^* : J(\theta, G, \psi) > 0 \ \forall(\theta, G)\}$. *We shall denote by* V *the function from* $\Theta \times \mathcal{G} \times K^{**}$ *to* $\mathbb{R}$ *defined by*

$$V(\theta, G, \psi) := 1/J(\theta, G, \psi)$$

for all $(\theta, G, \psi) \in \Theta \times \mathcal{G} \times K^{**}$.

Note that

5) Obviously, assumption (A5) implies $\bar{\psi} \in K^{**}$ and $J(\theta, G, \bar{\psi}) = I(\theta, G)$ $\forall(\theta, G)$.

(10) *We shall denote the Prohorov metric on* G *by* d. *In other words, the metric* d *is defined as follows :*

Let ρ *denote the metric on* Ξ. *For any* $\epsilon > 0$ *and* $A \subseteq \Xi$, *let* A^ϵ *denote the set* $\{\xi \in A : \rho(\xi, A) < \epsilon\}$. *We can now define* d *by the formula*

$$d(G_1, G_2) = \inf\{\epsilon > 0 : G_1(A) \le G_2(A^\epsilon) + \epsilon \text{ and } G_2(A) \le G_1(A^\epsilon) + \epsilon, \text{ for all } A \text{ in } \mathcal{B}(\Xi)\}$$

for all $G_1, G_2 \in \mathcal{G}$.

Later we shall need an estimate of the distribution function based, say, only on X_i's i odd, or only on X_i's i even. This is formalised below.

(11) *Let* $(A, \mathcal{A}), (B, \mathcal{B})$ *be two measurable spaces. For each* $n \ge 1$, *let* ϕ_n *be a measurable map from* $(A, \mathcal{A})^n$ *to* $(B, \mathcal{B})$. *For each* $n \ge 1$, *we shall define two more measurable maps from* $(A, \mathcal{A})^n$ *to* $(B, \mathcal{B})$ *by the relation*

$$\begin{aligned} \phi_n^O(\{a_i\}_{1\le i\le n}) &= \phi_{(n-[n/2])}(\{a_i\}_{1\le i\le n,\, i \text{ odd}}) \\ \text{and } \phi_n^E(\{a_i\}_{1\le i\le n}) &= \phi_{[n/2]}(\{a_i\}_{1\le i\le n,\, i \text{ even}}) \end{aligned}$$

for all $\{a_i\}_{1\le i\le n} \in A^n$.

Chapter 4
Mixture Models

In this chapter, we shall state and prove one auxiliary result and two main results in the mixture model. The auxiliary result will give conditions on the density function f and a kernel ψ (vide Definition 3.5) so that there exists an estimate $T_n(\psi)$ of θ_o, which is a uniformly $\sqrt{n}$-consistent solution (II) of

$$\sum_{\substack{i \text{ odd} \\ 1 \le i \le n}} \psi(X_i, \theta, \widehat{G}_n^E) + \sum_{\substack{i \text{ even} \\ 1 \le i \le n}} \psi(X_i, \theta, \widehat{G}_n^O) = 0 \tag{4.1}$$

(vide Defiinition 3.3) where $\widehat{G}_n$ is a uniformly consistent (II) estimate of G_o (vide Definition 3.1) and $\widehat{G}_n^E$ and $\widehat{G}_n^O$ are obtained from $\widehat{G}_n$ using even and odd numbered observations, respectively (formal definition is given in (11) of Chapter 3). Further conditions on ψ, guaranteeing uniform asymptotic normality (II) (vide Definition 3.4) of such estimate $T_n(\psi)$'s, are also given.

The two main results will prove the optimality of, respectively, Schick's and our estimate under the assumption that a simpler version of the conditions mentioned in the last paragraph hold for f and the optimal kernel $\bar{\psi}$ (vide relations (3.20)-(3.22)).

Before stating the auxiliary result, let us note the following assumption.

(B1) (a) There is a uniformly $\sqrt{n}$-consistent (II) estimate U_n of θ_o (vide Definition 3.2)

and (b) there is a uniformly consistent (II) estimate $\widehat{G}_n$ of G_o (vide Definition 3.1).

Let us now give a rigorous definition of our estimate $T_n(\psi)$.

Definition 4.1 For any kernel ψ, we shall define the estimate $T_n(\psi)$ as a solution of (4.1) which is nearest to U_n, if there is a solution of (4.1) lying in $(U_n - \log n/\sqrt{n}, U_n + \log n/\sqrt{n})$ and equal to U_n otherwise. This can be done in a way that ensures measurability.

Let ψ be a kernel. Fix (θ_o, G_o) in $\Theta \times \mathcal{G}$. Define a stochastic process D_n indexed by θ as follows.

$$
\begin{aligned}
D_n(\theta) \quad := \quad & \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ odd}}}^{n} \{\psi(X_i, \theta, \widehat{G}_n^E) - \psi(X_i, \theta_o, G_o) \\
& \qquad + (\theta - \theta_o) \int \psi(\cdot, \theta_o, G_o) f'(\cdot, \theta_o, G_o) d\mu(\cdot)\} \\
& + \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ even}}}^{n} \{\psi(X_i, \theta, \widehat{G}_n^O) - \psi(X_i, \theta_o, G_o) \\
& \qquad + (\theta - \theta_o) \int \psi(\cdot, \theta_o, G_o) f'(\cdot, \theta_o, G_o) d\mu(\cdot)\} \qquad (4.2)
\end{aligned}
$$

for all θ in Θ.

Consider the following conditions.

(i) $\int \{ \frac{A(\cdot, \theta_o, G_o, \theta, G_o) - 1}{(\theta - \theta_o)} - s_\theta(\cdot, \theta_o, G_o)\}^2 f(\cdot, \theta_o, G_o) d\mu(\cdot) \to 0$

as $\theta \to \theta_o$, where s_θ is the kernel defined by equation (3.20).

(ii) There is $\delta^{(1)}_{\theta_o, G_o} > 0$ such that

(a) $\int \psi^2(\cdot, \theta, G) f(\cdot, \theta_o, G_o) d\mu(\cdot) < \infty$

$$\forall (\theta, G) \in B(\theta_o, \delta^{(1)}_{\theta_o, G_o}) \times B(G_o, \delta^{(1)}_{\theta_o, G_o})$$

and (b) $\lim_{(\theta, G) \to (\theta_o, G_o)} \int \{\psi(\cdot, \theta, G) - \psi(\cdot, \theta_o, G_o)\}^2 f(\cdot, \theta_o, G_o) d\mu(\cdot) = 0.$

(iii) Assumption (B1)(b) holds with a choice of $\widehat{G}_n$ so that for any $c > 0$ and $\epsilon > 0$,

$$\sup_{\{\theta : |\theta - \theta_o| \le c/\sqrt{n}\}} P^n_{\theta_o, G_o}(\{|\sqrt{n} \int \psi(\cdot, \theta, \widehat{G}_n) f(\cdot, \theta, G_o) d\mu(\cdot)| > \epsilon\}) \to 0$$

as $n \to \infty$.

(iv) (a) There is $\delta^{(2)}_{\theta_o, G_o} > 0$ such that for all x in S and G in $B(G_o, \delta^{(2)}_{\theta_o, G_o})$,

$$\psi(x, \cdot, G) \in C(B(\theta_o, \delta^{(2)}_{\theta_o, G_o})),$$

(b) $\int \psi^2(\cdot,\theta_o,G_o)f(\cdot,\theta_o,G_o)d\mu(\cdot) < \infty$ (This condition follows from condition (ii)(a) but is given separately for ease in later references.)

and (c) $\int \psi(\cdot,\theta_o,G_o)f'(\cdot,\theta_o,G_o)d\mu(\cdot) \neq 0$.

(v) There is $\delta^{(3)}_{\theta_o,G_o} > 0$ and $A(\cdot,\theta_o,G_o) \in L_1(f(\cdot,\theta_o,G_o))$ such that

$$|\psi(\cdot,\theta',G) - \psi(\cdot,\theta,G)| \leq |\theta' - \theta| A(\cdot,\theta_o,G_o)$$

for all θ,θ' in $B(\theta_o,\delta^{(3)}_{\theta_o,G_o})$ and G in $B(\theta_o,\delta^{(3)}_{\theta_o,G_o})$.

Clearly, one can, without loss of generality, assume

$$\delta^{(1)}_{\theta_o,G_o} = \delta^{(2)}_{\theta_o,G_o} = \delta^{(3)}_{\theta_o,G_o} = \delta_o(\text{say}).$$

Let δ_o be as above. For any condition C among (i)-(v), let UC denote the condition that C, with θ_o,θ,θ' replaced by θ,θ',θ'' and G_o,G replaced by G,G'', holds uniformly with respect to θ,θ',θ'' in $B(\theta_o,\delta_o)$ and G,G' in $B(G_o,\delta_o)$.

In addition to U(i)-U(v), we shall need the following condition.

U(vi)(a) $$\sup_{(\theta,G)\in B(\theta_o,\delta_o)\times B(G_o,\delta_o)} \frac{\left[\int 1_{\{\psi^2(\cdot,\theta,G)\geq K\}}\psi^2(\cdot,\theta,G)f(\cdot,\theta,G)d\mu(\cdot)\right]}{J(\theta,G,\psi)} \to 0$$

as $K \to \infty$

and (b) $(\theta,G) \mapsto J(\theta,G,\psi)$ is continuous, where J is the function defined in (8) of Chapter 3.

Note that, becuase of compactness of $\mathcal{G}$, one can without loss of generality assume that the number δ_o considered in U(i)-U(vi) depends only on θ_o.

We can now state the auxiliary result.

Lemma 4.1 *Assume (B1). Fix (θ_o,G_o) in $\Theta \times \mathcal{G}$. Let ψ be a kernel. Let D_n be as defined in the relation (4.2). Also, whenever it makes sense, let $T_n(\psi)$ be the estimate defined in Definition 4.1. We can draw the following conclusions.*

(I) If conditions (i)-(iii) hold, then for all $c > 0$ and $\epsilon > 0$

$$\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} P^n_{\theta_o,G_o}(\{|D_n(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty.$$

(II) If conditions (i)-(iv) hold, then

(A) for any sequence $\{c_n\}$ increasing to infinity,

$$P^n_{\theta_o,G_o}(E_n) \to 1 \text{ as } n \to \infty$$

where E_n denotes the event that there is a solution of (4.1) lying inside the interval $(\theta_o - c_n/\sqrt{n}, \theta_o + c_n/\sqrt{n})$

and (B) under assumption (B1)(a), $T_n(\psi)$ is a $\sqrt{n}$-consistent solution (II) of (4.1).

(III) If conditions (i)-(v) hold, then

(A) for any $c > 0$ and $\epsilon > 0$,

$$P^n_{\theta_o,G_o}(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} |D_n(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty$$

and (B) under assumption (B1)(a)

$$\sup_{x\in\mathbb{R}} |P^n_{\theta_o,G_o}(\{\sqrt{n}(T_n(\psi) - \theta_o) \leq x\}) - \Phi(x/V(\theta_o, G_o, \psi))|$$

$\to 0$ as $n \to \infty$, where V is the function defined in (9) of Chapter 3.

(IV) For any conclusion C among (I)-(III), let UC denote the conclusion that C holds uniformly with respect to (θ_o, G_o) in compact subsets of $\Theta \times \mathcal{G}$. Then U(I), U(II) and U(III)(A) hold if the relevant conditions among U(i)-U(v) hold whereas U(III)(B) holds if conditions U(i)-U(vi) hold.

Proof : (I) For any $\theta \in \Theta$, define,

$$D_{n1}(\theta) = \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ odd}}}^{n} \{\psi(X_i, \theta, \widehat{G}^E_n) - \psi(X_i, \theta_o, G_o) + (\theta - \theta_o) \int \psi(\cdot, \theta_o, G_o) f'(\cdot, \theta_o, G_o) d\mu(\cdot)\}$$

and $D_{n2}(\theta) = D_n(\theta) - D_{n1}(\theta)$.

Fix $c > 0$ and $\epsilon > 0$.

It is enough to show that,

$$\sup_{\{\theta:|\theta-\theta_o|\le c/\sqrt{n}\}} P^n_{\theta_o,G_o}(\{|D_{n1}(\theta)| > \epsilon/2\}) \to 0$$
$$\text{and} \quad \sup_{\{\theta:|\theta-\theta_o|\le c/\sqrt{n}\}} P^n_{\theta_o,G_o}(\{|D_{n2}(\theta)| > \epsilon/2\}) \to 0. \tag{4.3}$$

We shall only show that,

$$\sup_{\{\theta:|\theta-\theta_o|\le c/\sqrt{n}\}} P^n_{\theta_o,G_o}(\{|D_{n1}(\theta)| > \epsilon/2\}) \to 0. \tag{4.4}$$

The other statement will follow by a symmetrical argument.

Now, for any sequence $\{\theta_n\}$ such that $|\theta_n - \theta_o| \le c/\sqrt{n}\ \forall n$,

$$\begin{aligned}
& \mathrm{E}(D_{n1}(\theta_n)|X_i, 1 \le i \le n, i \text{ even}) \\
= \quad & \frac{(n - [\frac{n}{2}])}{\sqrt{n}}\{\int \psi(x,\theta_n,\widehat{G}^E_n) f(x,\theta_o,G_o)d\mu(x) + 0 \\
& \qquad + (\theta_n - \theta_o)\int \psi(x,\theta_o,G_o) f'(x,\theta_o,G_o)d\mu(x)\} \\
& \qquad\qquad [\text{Since } \psi \text{ is a kernel.}] \\
\overset{(iii)}{=} \quad & -\frac{(n - [\frac{n}{2}])}{\sqrt{n}}(\theta_n - \theta_o) \\
& [\int \{\psi(x,\theta_n,\widehat{G}^E_n) - \psi(x,\theta_o,G_o)\}\frac{\{f(x,\theta_n,G_o) - f(x,\theta_o,G_o)\}}{(\theta_n - \theta_o)}d\mu(x) \\
& + \int \psi(x,\theta_o,G_o)\{\frac{f(x,\theta_n,G_o) - f(x,\theta_o,G_o)}{\theta_n - \theta_o} - f'(x,\theta_o,G_o)\}d\mu(x)] \\
& + o_{P^n_{\theta_o,G_o}}(1).
\end{aligned}$$

Therefore, by conditions (i), (ii) and assumption (B1), for any $\eta > 0$,

$$P^{[\frac{n}{2}]}_{\theta_o,G_o}(\{\sup_{\{\theta:|\theta-\theta_o|\le c/\sqrt{n}\}} |\mathrm{E}_{\theta_o,G_o}(D_{n1}(\theta)|X_i, 1 \le i \le n, i \text{ even})| > \eta\}) \to 0. \tag{4.5}$$

Let us also observe that, for any sequence $\{\theta_n\}$ such that $|\theta_n - \theta_o| \le c/\sqrt{n}\ \forall n$,

$$\begin{aligned}
& \mathrm{Var}_{\theta_o,G_o}(D_{n1}(\theta_n)|X_i, 1 \le i \le n, i \text{ even}) \\
\le \quad & \frac{(n - [\frac{n}{2}])}{n}\int \{\psi(x,\theta_n,\widehat{G}^E_n) - \psi(x,\theta_o,G_o)\}^2 f(x,\theta_o,G_o)d\mu(x).
\end{aligned}$$

Therefore, by uniform continuity of ψ and uniform consistency of $\widehat{G}_n$ (and hence $\widehat{G}_n^E$), we get for any $\eta > 0$,

$$P^{[\frac{n}{2}]}_{\theta_o,G_o}(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} \mathrm{Var}_{\theta_o,G_o}(D_{n1}(\theta)|X_i, 1 \leq i \leq n, i \text{ even}) > \eta\}) \to 0. \tag{4.6}$$

From (4.5) and (4.6), we get, for any $\eta > 0$,

$$\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} P^n_{\theta_o,G_o}(\{|D_{n1}(\theta)| > \eta\}|X_i, 1 \leq i \leq n, i \text{ even}) \overset{P^{[\frac{n}{2}]}_{\theta_o,G_o}}{\longrightarrow} 0. \tag{4.7}$$

Then, (4.4) follows by D.C.T. from (4.7) with $\eta = \epsilon/2$.

(II)(A) First observe that, because of (4.3), there is a sequence $\{c_n\}$ of nonnegative real numbers increasing to infinity such that for any $\epsilon > 0$,

$$\sup_{\{\theta:|\theta-\theta_o|\leq c_n/\sqrt{n}\}} P^n_{\theta_o,G_o}(\{|D_{n1}(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty. \tag{4.8}$$

Claim : Given any sequence $\{d_n\}$ of nonnegative real numbers such that $d_n \leq c_n \, \forall n$ and $d_n \uparrow \infty$,

$$P^n_{\theta_o,G_o}(E_n) \to 1 \text{ as } n \to \infty \tag{4.9}$$

where E_n denote the event that there is a solution of (4.1) lying inside the interval $(\theta_o - d_n/\sqrt{n}, \theta_o + d_n/\sqrt{n})$.

Then (II)(A) will follow because given any arbitrary sequence $\{d_n\}$ increasing to infinity, one can always work with the sequence $d'_n = \min\{d_n, c_n\} \, \forall n$.

Proof of the claim : Fix any sequence $\{d_n\}$ such that $d_n \leq c_n$, for all n and $d_n \uparrow \infty$.

By (4.8)

$$D_n(\theta_o \pm \frac{d_n}{\sqrt{n}}) \overset{P^n_{\theta_o,G_o}}{\longrightarrow} 0. \tag{4.10}$$

Again, by condition (iv)(b),

$$\{\mathcal{L}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(X_i,\theta_o,G_o)|P^n_{\theta_o,G_o})\}_{n\geq 1}$$

is tight and by condition (iv)(c) and choice of $\{d_n\}_{n\geq 1}$,

$$d_n\left(\int \psi(x,\theta_o,G_o)f'(x,\theta_o,G_o)d\mu(x)\right) \to \infty \text{ as } n \to \infty \tag{4.11}$$

[assuming, without loss of generality, $\int \psi(x,\theta_o,G_o)f'(x,\theta_o,G_o)d\mu(x) > 0$].

From (4.10) and (4.11),

$$\frac{1}{\sqrt{n}}\{\sum_{\substack{i \text{ odd} \\ 1\leq i\leq n}} \psi(X_i,\theta_o \pm \frac{d_n}{\sqrt{n}}, \widehat{G}_n^E) + \sum_{\substack{i \text{ even} \\ 1\leq i\leq n}} \psi(X_i,\theta_o \pm \frac{d_n}{\sqrt{n}}, \widehat{G}_n^O)\}$$

$\overset{w}{\Longrightarrow}$ Point mass at $\mp\infty$, *i.e.*, for any $K > 0$,

$$P^n_{\theta_o,G_o}(\{\frac{1}{\sqrt{n}}[\sum_{\substack{i \text{ odd} \\ 1\leq i\leq n}} \psi(X_i,\theta_o + \frac{d_n}{\sqrt{n}}, \widehat{G}_n^E) + \sum_{\substack{i \text{ even} \\ 1\leq i\leq n}} \psi(X_i,\theta_o + \frac{d_n}{\sqrt{n}}, \widehat{G}_n^O)] < -K\})$$

and

$$P^n_{\theta_o,G_o}(\{\frac{1}{\sqrt{n}}[\sum_{\substack{i \text{ odd} \\ 1\leq i\leq n}} \psi(X_i,\theta_o - \frac{d_n}{\sqrt{n}}, \widehat{G}_n^E) + \sum_{\substack{i \text{ even} \\ 1\leq i\leq n}} \psi(X_i,\theta_o - \frac{d_n}{\sqrt{n}}, \widehat{G}_n^O)] > K\})$$

$$\to 1 \text{ as } n \to \infty. \tag{4.12}$$

Fix any $K > 0$. Define, $A_{n,K}(\{d_n\})$

$$= \{\frac{1}{\sqrt{n}}[\sum_{\substack{i \text{ odd} \\ 1\leq i\leq n}} \psi(X_i,\theta_o + \frac{d_n}{\sqrt{n}}, \widehat{G}_n^E) + \sum_{\substack{i \text{ even} \\ 1\leq i\leq n}} \psi(X_i,\theta_o + \frac{d_n}{\sqrt{n}}, \widehat{G}_n^O)] < -K\}$$

$$\bigcap\{\frac{1}{\sqrt{n}}[\sum_{\substack{i \text{ odd} \\ 1\leq i\leq n}} \psi(X_i,\theta_o - \frac{d_n}{\sqrt{n}}, \widehat{G}_n^E) + \sum_{\substack{i \text{ even} \\ 1\leq i\leq n}} \psi(X_i,\theta_o - \frac{d_n}{\sqrt{n}}, \widehat{G}_n^O)] > K\},$$

then, by (4.12), $P^n_{\theta_o,G_o}(A_{n,K}(\{d_n\})) \to 1$ as $n \to \infty$ and by condition (iv)(a), on $A_{n,K}(\{d_n\})$ there is a solution of (4.1) lying inside the interval $(\theta_o - d_n/\sqrt{n}, \theta_o + d_n/\sqrt{n})$.

Since $\{d_n\}$ was arbitrary, this proves (4.9).

(II)(B) Suppose not. Then there is a sequence $\{d_n\}$ of nonnegative real numbers increasing to infinity such that $d_n \leq c_n$ for all n and

$P^n_{\theta_o,G_o}(\{\sqrt{n}|T_n - \theta_o| > d_n\}) \not\to 0$, where $\{c_n\}$ is the sequence considered in (4.8). [Note that, without loss of generality, we can assume $c_1 > 0$ and $d_1 > 0$].

Choose and fix a sequence of positive real numbers $\{\alpha_n\}$ such that $\frac{2d_n}{3d_{n-1}} < \alpha_n < \frac{3d_n}{4d_{n-1}}$, for all $n \geq 1$.

Then by $\sqrt{n}$-consistency of U_n,

$$P^n_{\theta_o,G_o}(\{\sqrt{n}|T_n - \theta_o| > d_n, \sqrt{n}|U_n - \theta_o| < \min(d_n - \alpha_n d_{n-1}, \frac{1}{2}\log n)\}) \not\to 0.$$

Consider $B_{n,K} := A_{n,K}(\{\min(2\alpha_n d_{n-1} - d_n, \frac{1}{2}\log n)\})$.

Note that in (4.9) one can easily drop the assumption of increasingness of $\{d_n\}$.

By von-Neumann selection theorem, choose a measurable function S_n which solves (4.1) on $B_{n,K}$.

Define $C_n = \{\sqrt{n}|T_n - \theta_o| > d_n, \sqrt{n}|U_n - \theta_o| < \min(d_n - \alpha_n d_{n-1}, \frac{1}{2}\log n)\}$.

Then, on $B_{n,K} \cap C_n$, S_n solves (4.1), $\sqrt{n}|S_n - U_n| < \min(\alpha_n d_{n-1}, \log n)$, whereas, $\sqrt{n}|T_n - U_n| > \alpha_n d_{n-1}$ and $P^n_{\theta_o,G_o}(B_{n,K} \cap C_n) \not\to 0$ contradicting the defintion of T_n.

(III)(A) Fix $c > 0$, $\epsilon > 0$ and $\eta > 0$. To show that there is $n_o \geq 1$ such that $n \geq n_o$ implies

$$P^n_{\theta_o,G_o}(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} |D_n(\theta)| > \epsilon\}) < \eta. \tag{4.13}$$

Fix a postive number α which divides c.

For $\theta \in [\theta_o - c/\sqrt{n}, \theta_o + c/\sqrt{n}]$, define

$$D_n^{(\alpha)}(\theta) := \frac{\sqrt{n}}{\alpha}\sum_{i=-c/\alpha}^{c/\alpha-1}[(\theta - \theta_o - \frac{i\alpha}{\sqrt{n}})D_n(\theta_o + \frac{(i+1)\alpha}{\sqrt{n}}) + \{\frac{(i+1)\alpha}{\sqrt{n}} - \theta + \theta_o\}D_n(\theta_o + \frac{i\alpha}{\sqrt{n}})] \; 1_{\{\theta_o + \frac{i\alpha}{\sqrt{n}} \leq \theta \leq \theta_o + \frac{(i+1)\alpha}{\sqrt{n}}\}}.$$

Now,

$$P^n_{\theta_o,G_o}(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} |D_n(\theta)| > \epsilon\})$$

$$
\begin{aligned}
\leq\ & P^n_{\theta_o,G_o}(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} |D_n(\theta) - D_n^{(\alpha)}(\theta)| > \epsilon/2\}) \qquad (4.14)\\
& + P^n_{\theta_o,G_o}(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} |D_n^{(\alpha)}(\theta)| > \epsilon/2\})\\
\leq\ & P^n_{\theta_o,G_o}(\{\sup_{\{(c',c'',G):|c'|\leq c,|c''|\leq c,|c''-c'|\leq\alpha,G\in B(G_o,\delta_o)\}}\\
& \qquad |D_n(\theta_o + \frac{c''}{\sqrt{n}}) - D_n(\theta_o + \frac{c'}{\sqrt{n}})| > \epsilon/2\})\\
& + P^n_{\theta_o,G_o}(\{\sup_{i\in\{0,\pm1,\pm2,\cdots,\pm c/\alpha\}} |D_n(\theta_o + \frac{i\alpha}{\sqrt{n}})| > \epsilon/2\})\\
\leq\ & \frac{2}{\epsilon}\mathrm{E}_{P^n_{\theta_o,G_o}}(\sup_{\{(c',c'',G):|c'|\leq c,|c''|\leq c,|c''-c'|\leq\alpha,G\in B(G_o,\delta_o)\}}\\
& \qquad |D_n(\theta_o + \frac{c''}{\sqrt{n}}) - D_n(\theta_o + \frac{c'}{\sqrt{n}})|)\\
& + \sum_{-c/\alpha}^{c/\alpha} P^n_{\theta_o,G_o}(\{|D_n(\theta_o + \frac{i\alpha}{\sqrt{n}})| > \epsilon/2\})\\
\leq\ & \frac{2}{\epsilon}|\mathrm{E}_{P^n_{\theta_o,G_o}}(\sup_{\{(c',c'',G):|c'|\leq c,|c''|\leq c,|c''-c'|\leq\alpha,G\in B(G_o,\delta_o)\}}\\
& \qquad \sqrt{n}|\psi(X_1,\theta_o + \frac{c''}{\sqrt{n}}, G) - \psi(X_1,\theta_o + \frac{c'}{\sqrt{n}}, G)|)\\
& + \alpha|\int \psi(\cdot,\theta_o.G_o) f'(\cdot,\theta_o,G_o)d\mu(\cdot)||\\
& + \sum_{-c/\alpha}^{c/\alpha} P^n_{\theta_o,G_o}(\{|D_n(\theta_o + \frac{i\alpha}{\sqrt{n}})| > \epsilon/2\})\\
\overset{(v)(a)}{\leq}\ & \frac{2\alpha}{\epsilon}\{\int A(\cdot,\theta_o,G_o) f(\cdot,\theta_o,G_o)d\mu(\cdot)\\
& \quad + |\int \psi(\cdot,\theta_o,G_o) f'(\cdot,\theta_o,G_o)d\mu(\cdot)|\}\\
& + \sum_{-c/\alpha}^{c/\alpha} P^n_{\theta_o,G_o}(\{|D_n(\theta_o + \frac{i\alpha}{\sqrt{n}})| > \epsilon/2\})\\
=\ & I + II \text{ (say).} \qquad (4.15)
\end{aligned}
$$

Let us choose $\alpha > 0$ such that $\alpha | c$ and $I < \eta/2$.

Using (I) choose $n_o \geq 1$ such that $n \geq n_o$ implies $II < \eta/2$.

Then (4.13) follows from (4.15).

(III)(B) Easy.

(IV) An easy consequence of Proposition 3.4. □

Remark 4.1 Condition U(iii) is a uniform version of condition (2.8) of Schick (1986, p. 1144). For C_1-kernels this condition holds by definition. For the optimal kernel $\tilde{\psi}$, in view of Lemma 3.5, this condition holds even for the general semiparametric families satisfying Bickel's Condition C, provided suitable regularity conditions hold (cf Remark 3.6).

Remark 4.2 Note that for any kernel ψ,

$$\frac{\partial}{\partial\theta}[\int \psi(\cdot,\theta,G)f(\cdot,\theta,G)d\mu(\cdot)] = 0 \ \ \forall(\theta,G)$$

under suitable regularity conditions, which, in turn, implies

$$\int \psi'(\cdot,\theta,G)f(\cdot,\theta,G)d\mu(\cdot) = -\int \psi(\cdot,\theta,G)f'(\cdot,\theta,G)d\mu(\cdot) \ \ \forall(\theta,G).$$

This helped in putting the Taylor's expansion used in the proof of the above result in the usual form. This idea goes back to Bickel(1975)(vide relation (2.8) of page 429).

Let us now consider the following definition.

Definition 4.2 Any kernel ψ satisfying the conditions U(ii)-U(vi) will be called an ***estimable kernel in** Model II* (or, in short, an *EK (II)*) and any uniformly $\sqrt{n}$-consistent solution (II) of (4.1) (vide Definition 3.3) will be called a *generalised C_1-estimate in Model II corresponding to ψ* (or, in short, a *GC_1 (II)* estimate).

In view of Definition 4.2, conclusion U(III)(B) of Lemma 4.1 can be restated as

Lemma 4.1a *Assume (B1). If f satisfies U(i) and ψ is an EK (II), then $T_n(\psi)$ is a GC_1 (II) estimate (corresponding to ψ) as well as a UAN (II) estimate with AV $V(\cdot,\cdot,\psi)$.*

Example 4.1 All C_1-kernels are EK (II) and all C_1-estimates corresponding to them are GC_1 (II) estimates.

Example 4.2 It can be verified in several cases that $\psi = f'/f$ is an EK (II) and $T_n(\psi)$ is a GC_1 (II) estimate.

The following is the construction of an efficient estimate as given in Schick(1986, pp 1140-1144).

Let $l^* : S \times \Theta \times \mathcal{G} \to I\!\!R$ and $Q : \Theta \times \mathcal{G} \times \mathcal{G} \to I\!\!R$ be defined by

$$\begin{aligned} l^*(x,\theta,G) &:= \psi(x,\theta,G)/I(\theta,G) \ \ \forall (x,\theta,G) \in S \times \Theta \times \mathcal{G} \\ \text{and } Q(\theta,G,G') &:= \int l^*(\cdot,\theta,G) f(\cdot,\theta,G') d\mu(\cdot) \ \ \forall (\theta,G,G') \in \Theta \times \mathcal{G} \times \mathcal{G}. \end{aligned} \tag{4.16}$$

Consider the estimate

$$Z_n := \bar{U}_n + \frac{1}{n} \sum_{\substack{1 \le i \le n \\ i \text{ odd}}} l^*(X_i, \bar{U}_n, \hat{G}_n^E) + \frac{1}{n} \sum_{\substack{1 \le i \le n \\ i \text{ even}}} l^*(X_i, \bar{U}_n, \hat{G}_n^O) \tag{4.17}$$

where $\bar{U}_n$ is a discretized version of U_n, *i.e.*, $\bar{U}_n := \frac{(\text{nearest integer to } \sqrt{n} U_n)}{\sqrt{n}}$.

Assume that

(B2) (a) For any x in S, $f(x,\cdot,\cdot) \in C(\breve{\Theta} \times \mathcal{G})$

and (b) for any compact subset Θ_o of Θ, there is $\delta_o > 0$ such that the family of functions

$$\{\frac{(f')^2(\cdot,\theta',G)}{f(\cdot,\theta,G)} : \theta,\theta' \in \Theta_o \text{ with } |\theta - \theta'| \le \delta_o, G \in \mathcal{G}\}$$

is uniformly integrable with respect to μ.

Remark 4.3 If $(S, \mathcal{S}) = (I\!\!R^p, \mathcal{B}^p)$ and assumption (B2)(a) holds, then, in view of Corollaries 3.1.1 and 3.1.2, one can easily drop assumption (B1)(b) even if Θ is unbounded.

Remark 4.4 Let ψ be a Borel-measurable function from $I\!\!R^p$ to $I\!\!R^+$. Let $s_1, s_2, \ldots, s_k$ be k Borel-measurable functions from $I\!\!R^p$ to $I\!\!R$. Define

$$\Omega = \{\omega \in I\!\!R^k : \int \psi(x) \exp\{\sum_{j=1}^{k} s_j(x)\omega_j\} dx < \infty\}.$$

Assume that

(a) $\Omega \neq \emptyset$.

Consider the exponential family of densities defined by

$$h(x,\omega) = (d_o(\omega))^{-1}\psi(x)\exp\{\sum_{j=1}^{k} s_j(x)\omega_j\}$$

for all x in $I\!R^p$ and ω in Ω, where the function d_o is given by the formula

$$d_o(\omega) = \int \psi(x)\exp\{\sum_{j=1}^{k} s_j(x)\omega_j\}dx \quad \forall\omega.$$

Consider the family of marginal distributions of s,

$$\{Q_\omega : \omega \in \Omega\}.$$

Assume that

(b) The above family is dominated by the k-dimensional Lebesgue measure.

(c) There is a k-dimensional rectangle J contained in the support of all the Q_ω's.

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k functions in $C_{2,0}(\bar{\Theta} \times \Xi)$.

Assume that

(d) $\pi := (\pi_1, \pi_2, \cdots, \pi_k)$ is one-one and bimeasurable.

(e) Range of π is contained in the interior of Ω.

Finally, let us assume that

(f) $(S, \mathcal{S}) = (I\!R^p, \mathcal{B}^p)$ and the density f is given by the formula

$$f(x,\theta,\xi) = \frac{1}{d(\theta,\xi)}\psi(x)\exp\{\sum_{j=1}^{k} s_j(x)\pi_j(\theta,\xi)\}$$

for all x in S, θ in Θ and ξ in Ξ, where d stands for the function $d_o \circ \pi$.

Assumptions (a)-(d) and (f) are needed to prove assumption (A1) whereas assumptions (a), (e) and (f) are needed to prove assumption (B2). We shall only prove the first assertion. The proof of the other one is simple and hence omitted.

In view of assumption (d), it is enough to prove the identifiability of

$$\{\int h(\cdot,\omega)dH(\omega) : H \in \mathcal{H}\}$$

where $\mathcal{H}$ denote the set of all probability measures on Ω with compact support.

Consider any two probability measures H_1 and H_2 in $\mathcal{H}$. By assumption (b), for $i = 1, 2$

$$A_i(s) := \int \exp\{\sum_{j=1}^{k} \omega_j s_j\}(d_o(\omega))^{-1} dH_i(\omega) < \infty$$

for almost all s and hence, by assumption (c), for all s in J.

Moreover, if H_1, H_2 give rise to the same marginal of X_1 then $A_1(s) = A_2(s)$ for almost all s and hence, by continuity of A_i's, for all s in J.

Therefore

$$(d_o(\omega))^{-1} dH_1(\omega) = (d_o(\omega))^{-1} dH_2(\omega)$$

by a well known result on moment generating functions. Hence by continuity of d_o and choice of $\mathcal{H}, H_1 = H_2$. (At this stage, note that in the Lindsay(1980)'s case,to be discussed in Chapter 6(b), we don't need the identifiability of G so that one can easily replace assumptions (b) and (c) by

(b)* The family $\{h(\cdot,\omega) : \omega \in \Omega\}$ of density functions is identifiable.)

Next, observe that, assumptions (b)-(c) imply that the family $\{Q_\omega : \omega \in \Omega\}$ of probability measures is identifiable. The assertion follows by Theorem 10.0.3 of Prakasa Rao (1983, p. 440) and the definition of Q_ω's.

In order to prove the efficiency of Z_n, we need one more assumption, namely,

(B3) There is a version of the optimal kernel $\bar{\psi}$ such that

(a) For all x in S, $\bar{\psi}(x,\cdot,\cdot) \in C(\bar{\Theta} \times \mathcal{G})$

and (b) for any compact subset Θ_o of Θ the following statements hold

(i) there is $\delta_o > 0$ such that the family of functions

$$\{\bar{\psi}(\cdot,\theta',G')f(\cdot,\theta,G) : (\theta,G),(\theta',G') \in \Theta_o \times \mathcal{G} \text{ with } |\theta-\theta'| + d(G,G') \le \delta_o\}$$

is uniformly integrable with respect to μ

and (ii) $\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} [\frac{\{\int 1_{\{\bar{\psi}^2(\cdot,\theta,G)\ge K\}}\bar{\psi}^2(\cdot,\theta,G)f(\cdot,\theta,G)d\mu(\cdot)\}}{I(\theta,G)}] \to 0$

as $n \to \infty$.

Observe that

(1) Assumption (B2) is a stronger version of assumption (A3) and it implies condition U(i).

(2) l^* (and hence Z_n) is well defined only under assumption (A5). Moreover if $(\theta, G) \mapsto I(\theta, G)$ is continuous, then any condition among (ii)-(v) and U(ii)-U(vi) holds for the kernel l^*, if and only if, it holds for the kernel $\bar{\psi}$.

(3) Assumptions (B2)(a), (B3)(a) and (B3)(b)(i) imply that $(\theta, G) \mapsto I(\theta, G)$ is continuous. They also imply a local version of assumption (A4) with $\bar{\Theta}$ and $\mathcal{G}$ replaced by $B(\theta_o, \delta_o^*)$ and $B(G_o, \delta_o^*)$, respectively, where $\delta_o^* = \delta_o/2$.

(4) Assumption (B3)(b)(ii) implies assumption (A5).

The relation between assumption (B3) and the relevant conditions of the lemma will become apparent from the proof of the following result which establishes the efficiency of Z_n.

Theorem 4.2 *Assume (B1)-(B3). The estimate Z_n of θ_o, as defined through the relations (4.16)-(4.17). is UAN (II) with AV $(1/I)$ (vide Definition 3.4).*

Proof : Let us start with the following simple observation.

$$\int l^*(\cdot, \theta, G) f'(\cdot, \theta, G) d\mu(\cdot) = 1 \;\; \forall (\theta, G). \tag{4.18}$$

Next, we shall show that

$$\sup_{x \in \mathbb{R}} |P^n_{\theta_o, G_o}(\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^*(X_i, \theta_o, G_o) \leq x\}) - \Phi(x\, I^{\frac{1}{2}}(\theta_o, G_o))| \to 0 \tag{4.19}$$

as $n \to \infty$, uniformly with respect to (θ_o, G_o) in compact subsets of $\Theta \times \mathcal{G}$.

We shall proceed as follows.

First observe that, $\bar{\psi}$ being a kernel, $l^*(\cdot, \theta, G)$ has zero expectation under $P_{\theta, G}$. Fix any compact subset A of $\Theta \times \mathcal{G}$. By assumption (B3), conditions (i)-(ii) of Proposition 3.4 holds with $X_n(\alpha) = l^*(X_n, \alpha)$, for all

α in A and $n \geq 1$. Hence by Proposition 3.4, L.H.S. of (4.19) goes to zero uniformly with respect to α in A. Since A was arbitrary this proves (4.19).

In view of (4.17)-(4.19), it is enough to show that

$$P^n_{\theta_o,G_o}(\{|D_n(\bar{U}_n)| > \epsilon\}) \to 0 \text{ as } n \to \infty, \tag{4.20}$$

uniformly with respect to (θ_o, G_o) in compact subsets of $\Theta \times \mathcal{G}$.

Again, as U_n (and hence $\bar{U}_n$) is a uniformly $\sqrt{n}$-consistent (II) estimate of θ_o, it is enough to show that for any $c > 0$ and $\epsilon > 0$,

$$P^n_{\theta_o,G_o}(\{|D_n(\bar{U}_n)| > \epsilon\} \bigcap \{\sqrt{n}|\bar{U}_n - \theta_o| \leq c\}) \to 0 \text{ as } n \to \infty \tag{4.21}$$

uniformly with respect to (θ_o, G_o) in compact subsets of $\Theta \times \mathcal{G}$.

Now $\sqrt{n}|\bar{U}_n - \theta_o| \leq c$ if and only if $\sqrt{n}\theta_o - c \leq \sqrt{n}\bar{U}_n \leq \sqrt{n}\theta_o + c$ and by definition of $\bar{U}_n$, $\sqrt{n}\bar{U}_n$ is an integer. Therefore, $\bar{U}_n$ can only assume values of the form $\frac{i}{\sqrt{n}}$, where $\sqrt{n}\theta_o - c \leq i \leq \sqrt{n}\theta_o + c$ and there can be at most $[2c] + 1$ such values. (This is so because given any two real numbers $a < b$, there can be at most $[b - a] + 1$ integers in $[a, b]$.) Thus (4.21) (and hence (4.20)) holds if part U(I) of Lemma 4.1 holds with $\psi = l^*$. So, in view of observation (1), it remains to check conditions U(ii) and U(iii) with $\psi = l^*$.

In view of observations (2) and (3), assumptions (B2)(a), (B3)(a) and (B3)(b)(i) imply condition U(ii) for the kernel l^*.

In view of observation (3) and a local version of Lemma 3.5, with $\bar{\Theta} \times \mathcal{G} \times \mathcal{G}$ replaced by $B(\theta_o, \delta_o^*) \times B(G_o, \delta_o^*) \times B(G_o, \delta_o^*)$, one can easily conclude that $Q = 0$ on $B(\theta_o, \delta_o^*) \times B(G_o, \delta_o^*) \times B(G_o, \delta_o^*)$ guaranteeing condition U(iii).

□

Remark 4.5 The proof of Theorem 4.2 is similar to that of Bickel (1982) or Schick (1986) but differs in many details. In particular, we need uniformity unlike them.

For the next result, we need the following stronger version of assumption (B3).

(B3s) There is a version of the optimal kernel $\bar{\psi}$ such that

(a) for all x in S, $\bar{\psi}(x,\cdot,\cdot) \in C_{1,0}(\bar{\Theta} \times \mathcal{G})$

and (b) for any compact subset Θ_o of Θ, there is $\delta_o > 0$ such that

(i) assumption (B3)(b) holds

and (ii) $\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} [\int\{\sup_{(\theta',G')\in B((\theta,G),\delta_o)} |\bar{\psi}'(\cdot,\theta',G')|\}f(\cdot,\theta,G)d\mu(\cdot)]$

$< \infty.$

We can now state the final result of this chapter.

Theorem 4.3 *Assume (B1), (B2) and (B3s). The estimate* $T_n(\bar{\psi})$ *of* θ_o*, as defined in Definition 4.1, is UAN (II) with AV* $(1/I)$.

Proof : In view of Lemma 4.1a and observation (1), we have to check conditions U(ii)-U(vi) for the kernel $\bar{\psi}$. In the proof of the last theorem, we have checked conditions U(ii)-U(iii) for the kernel l^*. Also, observation (3) guarantees the continuity of $(\theta, G) \mapsto I(\theta, G)$. Hence, in view of observation (2), conditions U(ii)-U(iii) hold for the kernel $\bar{\psi}$ also. So, it remains to check conditions U(iv)-U(vi).

Condition U(iv)(a) follows from assumption (B3)(a), U(iv)(b) from assumption (B3)(b)(i) and U(iv)(c) from assumption (B3)(b)(ii) and the definition of $\bar{\psi}$.

Condition U(v) follows from assumptions (B3s)(a) and (B3s)(b)(ii).

Condition U(vi) is a consequence of assumption (B3s)(b)(ii) and observations (3)-(4). □

Remark 4.6 In view of observations (3)-(4), assumptions (B1), (B2) and (B3s) imply that l^* is an EK (II) and for any compact subset Θ_o of Θ and $\epsilon > 0$

$$\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} P^n_{\theta_o,G_o}(\{\sqrt{n}|Z_n - T_n(l^*)| > \epsilon\}) \to 0 \text{ as } n \to \infty.$$

Remark 4.7 As indicated in Remark 3.6, all the results stated in this chapter holds for the general semiparametric families satisfying Bickel's Condition C also.

Remark 4.8 In view of Remarks 4.3 and 4.4, for Euclidean S and exponential f, it is enough to check assumption (B1)(a), *i.e.*, the existence of a uniformly $\sqrt{n}$-consistent (II) estimate of θ_o, and assumption (B3) or (B3s), *i.e.*, the smoothness properties of the optimal kernel.

Chapter 5
Fixed Set-up

In this chapter, we shall state the analogues of Lemma 4.1, Theorem 4.2 and Theorem 4.3 in the fixed set-up. However, we apply a random permutation Π to the original sample $(X_1, X_2, \cdots, X_n)$ and base the analysis on $(X_{\Pi(1)}, X_{\Pi(2)}, \cdots, X_{\Pi(n)})$. Let s_n denote the group of all permutations of $\{1, 2, \cdots, n\}$ and P_n denote the probability distribution of Π. Later we shall make an approriate choice of P_n for the asymptotically efficient estimate so that the empirical distribution functions (or, the empirical probability measures) of $\xi_{\Pi(i)}$'s based on odd and even indices will be close to each other.

Let us start with the following definitions.

Definition 5.1 Let $(Y, \mathcal{Y})$, $(Z, \mathcal{Z})$ be two measurable spaces. For any $n \geq 1$, Z-valued statistic V_n on $(Y, \mathcal{Y})^n$ and probability measure P_n on s_n, call the statistic sending $(y_1, y_2, \cdots, y_n)$ to $V_n(y_{\Pi(1)}, y_{\Pi(2)}, \cdots, y_{\Pi(n)})$ *the randomisation of the statistic V_n corresponding to P_n* and denote it by $V_n^*(P_n)$.

In practice, we shall take $(Y, \mathcal{Y})$ to be $(S, \mathcal{S})$ or $(\Xi, \mathcal{B}(\Xi))$, Z to be $\bar{\Theta}$, $\mathcal{G}$ or $\bar{\Theta} \times \mathcal{G}$ and $\mathcal{Z}$ to be $\mathcal{B}(Z)$.

Definition 5.2 Let (Y, ρ) and ϕ be as considered in Definition 3.1. For any esimate V_n of $\phi(\theta_o, \underline{G}_n)$ in Model I ($\phi(\theta_o, G_o)$ in Model II) and probability measure P_n on s_n, we shall call the Y-valued statistic $V_n^*(P_n)$ as defined in Definition 5.1 a *randomised estimate of $\phi(\theta_o, \underline{G}_n)$ in Model I ($\phi(\theta_o, G_o)$ in Model II)*.

As a special case of the above definition, we can define the notions of randomised estimates of θ_o, $\underline{G}_n$ or $(\theta_o, \underline{G}_n)$ in Model I (θ_o, G_o or (θ_o, G_o) in Model II) (*cf.* Definition 3.1).

Note that

(1) The nonrandomised estimates are special cases of randomised estimates. Also, for any Z-valued statistic V_n on $(S, \mathcal{S})^n$ and probability measure P_n on s_n, the following hold.

$$(\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{V_n^*(P_n) \in A\}) = \int (\prod_{i=1}^{n} P_{\theta_o,\xi_{\pi(i)}})(\{V_n \in A\}) dP_n(\pi) \tag{5.1}$$

for all A in $\mathcal{Z}$, θ_o in $\bar{\Theta}$ and $\{\xi_i\}_{1\leq i\leq n}$ in Ξ^n and

$$P^n_{\theta_o,G_o}(\{V_n^*(P_n) \in A\}) = P^n_{\theta_o,G_o}(\{V_n \in A\}) \tag{5.2}$$

for all A in $\mathcal{Z}$, θ_o in $\bar{\Theta}$ and G_o in $\mathcal{G}$.

(2) In view of the relations (5.1)-(5.2), there are extensions of Definitions 3.1-3.4 for randomised estimates and in view of observation (1), for any property P defined in Definitions 3.1-3.4 and statistic V_n, P holds for V_n if and only if it holds for all possible randomisation $V_n^*(P_n)$'s of it, both in Model I as well as Model II.

(3) As in observation (2), the notion of efficiency (I) ((II)) has obvious extension for randomised estimates and one can easily prove that in the extended sense, regularity (I) implies regualrity (II). So the problem of efficiency (I) reduces to finding a *randomised* estimate which is efficient (II) and regular (I).

For the remaining part of this chapter, we shall need the following Model I-analogue of assumption (B1).

(C1) (a) There is a uniformly $\sqrt{n}$-consistent (I) estimate U_n of θ_o (vide Definition 3.2)

and (b) there is a uniformly consistent (I) estimate $\widehat{G}_n$ of $\underline{G}_n$ (vide Definition 3.1).

Convention 1 For any $n \geq 1$, let P_n^u denote the uniform distribution over s_n. From now on we shall use the shorthand notation V_n^* for $V_n^*(P_n^u)$. Let ψ be a kernel. Our goal is to solve the following randomisation of equation (4.1).

$$\frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i\,\mathrm{odd}}}^{n} \psi(X_i^*, \theta, (\widehat{G}_n^E)^*) + \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i\,\mathrm{even}}}^{n} \psi(X_i^*, \theta, (\widehat{G}_n^O)^*) = 0 \tag{4.1*}$$

where $(\hat{G}_n^E)^*$ and $(\hat{G}_n^O)^*$ are obtained from $\hat{G}_n$ using (11) of Chapter 3 and Definition 5.1 with $P_n = P_n^u$. $T_n^*(\psi)$ is defined in analogy with $T_n(\psi)$ by replacing (4.1) and U_n by (4.1)* and U_n^*, respectively. Clearly $T_n^*(\psi)$ equals $(T_n(\psi))^*$.

Note that

(4) Theorems 4.2-4.3 and relation (5.2) together imply that Z_n^* is efficient (II) under assumptions (B1)-(B3) and $T_n^*(\bar{\psi})$ is efficient (II) under assumptions (B1), (B2) and (B3s).

In view of observations (1)-(4), it remains to show that Z_n^* and $T_n^*(\bar{\psi})$ are regular (I). Naturally, we shall prove an analogue of Lemma 4.1 when we have Model I instead of Model II and randomised estimates. Before stating the required lemma we need three more auxiliary results namely the following two propositions and Lemma 5.1(t).

Proposition 5.1 *Let (Y, ρ) be a compact metric space. Let $I\!\!P$ denote the set of all Borel probability measures on Y. Let $\xi_1, \xi_2, \ldots, \xi_n$ be n independent Y-valued random variables with ξ_i following the distribution P_i. Then for all $\epsilon > 0$,*

$$\sup_{\{P_i\}_{1\le i\le n}\in I\!\!P^n} \left(\prod_{i=1}^n P_i\right)(\{d(I\!\!F_n, \bar{P}_n) > \epsilon\}) \to 0 \text{ as } n \to \infty$$

where $\bar{P}_n$ denote the measure $\frac{1}{n}\sum_{i=1}^n P_i$ and d denote the Prohorov metric on $I\!\!P$ as defined in (10) of Chapter 3.

Proof : First let us observe that for any function f in $C(Y)$ and $\epsilon > 0$,

$$\sup_{\{P_i\}_{1\le i\le n}\in I\!\!P^n} \left(\prod_{i=1}^n P_i\right)\left(\left\{\left|\int f d(I\!\!F_n - \bar{P}_n)\right| > \epsilon\right\}\right) \to 0 \text{ as } n \to \infty. \tag{5.3}$$

Next, we shall extend (5.3) to the following.
For any compact subset $\mathcal{F}$ of $C(Y)$ and $\epsilon > 0$,

$$\sup_{\{P_i\}_{1\le i\le n}\in I\!\!P^n} \left(\prod_{i=1}^n P_i\right)\left(\left\{\sup_{f\in\mathcal{F}}\left|\int f d(I\!\!F_n - \bar{P}_n)\right| > \epsilon\right\}\right) \to 0 \text{ as } n \to \infty. \tag{5.4}$$

This can be proved as follows.

Let $\mathcal{F}$ be a given compact subset of $C(Y)$ and ϵ be a given positive real number. Using compactness of $\mathcal{F}$ get hold of an $\frac{\epsilon}{4}$-net $\{f_1, f_2, \cdots, f_k\}$ of $\mathcal{F}$. Then

$$\sup_{f\in\mathcal{F}} |\int f d(I\!F_n - \bar{P}_n)| \le \epsilon/2 + \max_{1\le j\le k} |\int f_j d(I\!F_n - \bar{P}_n)|.$$

Therefore

$$\sup_{\{P_i\}_{1\le i\le n}\in\mathcal{P}^n} (\prod_{i=1}^{n} P_i)(\{\sup_{f\in\mathcal{F}} |\int f d(I\!F_n - \bar{P}_n)| > \epsilon\})$$

$$\le \sup_{\{P_i\}_{1\le i\le n}\in\mathcal{P}^n} (\prod_{i=1}^{n} P_i)(\{\max_{1\le j\le k} |\int f_j d(I\!F_n - \bar{P}_n)| > \epsilon/2\}) \to 0 \text{ as } n \to \infty,$$

by (5.3).

As $\mathcal{F}$ and ϵ were arbitrary, this proves (5.4).

Let us now consider the function $\phi : I\!R \to [0,1]$ defined by,

$$\phi(t) = \begin{cases} 1 & \text{if } t \le 0 \\ 1-t & \text{if } 0 \le t \le 1 \\ 0 & \text{if } 1 \le t. \end{cases} \tag{5.5}$$

Then ϕ is bounded and uniformly continuous as it is a continuous function with a compact support.

For any $\epsilon > 0$ and closed subset F of Y, we shall denote the function $\phi(\frac{d(\cdot,F)}{\epsilon})$ from Y to $[0,1]$ by $f_{\epsilon,F}$ and consider

$$\mathcal{F}_\epsilon := \{f_{\epsilon,F} : F \text{ a closed subset of } Y\}. \tag{5.6}$$

Let us now observe that for any x, y in Y and closed subset F of it,

$$\left.\begin{array}{rl} & d(x,z) \le d(x,y) + d(y,z) \\ \text{and} & d(y,z) \le d(y,x) + d(x,z) \end{array}\right\} \text{ for all } z \text{ in } Y.$$

Therefore taking the infemum over z in F and using the symmetry of d,

$$\begin{array}{rl} & d(x,F) \le d(x,y) + d(y,F) \\ \text{and} & d(y,F) \le d(x,y) + d(x,F). \end{array}$$

Hence $|d(x,F) - d(y,F)| \le d(x,y)$.

As x, y and F were arbitrary this proves that the family of functions

$$\{d(\cdot, F) : F \text{ a closed subset of } Y\} \tag{5.7}$$

is equicontinuous on Y.

From now on, we shall assume that ϵ is a preassigned positive number.

From (5.5)-(5.7) and boundedness and uniform continuity of ϕ, we can easily conclude that $\mathcal{F}_\epsilon$ is uniformly bounded and equicontinuous.

Therefore by Arzéla-Ascoli Theorem $\bar{\mathcal{F}}_\epsilon$ is compact.

Hence, by (5.4),

$$\sup_{\{P_i\}_{1\le i\le n}\in \mathbb{P}^n} (\prod_{i=1}^{n} P_i)(\{\sup_{f\in\mathcal{F}_\epsilon} |\int f d(\mathbb{F}_n - \bar{P}_n)| > \epsilon\}) \to 0 \text{ as } n \to \infty. \tag{5.8}$$

Let us now observe that the Prohorov metric d as defined in (10) of Chapter 3 can easily be redefined using closed sets only, *i.e.*, for any $P, Q \in \mathbb{P}$,

$$d(P,Q) = \inf\{\eta > 0 : P(F) \le Q(F^\eta) + \eta, Q(F) \le P(F^\eta) + \eta \ \ \forall F \subseteq Y, \ F \text{ closed}\}.$$

Therefore, for any P, Q in $\mathbb{P}$,

$d(P,Q) > \epsilon$

$\Longrightarrow$ there is a closed subset F of Y (possibly depending on P, Q and ϵ) such that $P(F) > Q(F^\epsilon) + \epsilon$ or $Q(F) > P(F^\epsilon) + \epsilon$

$\overset{(5.6)}{\Longrightarrow}$ there is a closed subset F of Y such that

$$\int f_{\epsilon,F}dP \overset{(5.6)}{\ge} P(F) > Q(F^\epsilon) + \epsilon \overset{(5.6)}{\ge} \int f_{\epsilon,F}dQ + \epsilon$$

$$\text{or } \int f_{\epsilon,F}dQ \overset{(5.6)}{\ge} Q(F) > P(F^\epsilon) + \epsilon \overset{(5.6)}{\ge} \int f_{\epsilon,F}dP + \epsilon$$

$\Longrightarrow$ there is a closed subset F of Y such that

$|\int f_{\epsilon,F}dP - \int f_{\epsilon,F}dQ| > \epsilon$

$\Longrightarrow$ $\sup_{f\in\mathcal{F}_\epsilon} |\int f d(P-Q)| > \epsilon,$

where $\mathcal{F}_\epsilon$ is the family of continuous functions defined in (5.6).

Therefore, for any P, Q in $\mathbb{P}$ and $\{P_i\}_{1\le i\le n}$ in $\mathbb{P}^n$,

$$(\prod_{i=1}^{n} P_i)(\{d(P,Q) > \epsilon\}) \le (\prod_{i=1}^{n} P_i)(\{\sup_{f\in\mathcal{F}_\epsilon} |\int f d(P-Q)| > \epsilon\}). \tag{5.9}$$

Taking supremum over $\{P_i\}_{1\le i\le n}$ in $\mathbb{P}^n$ we get for any P, Q in $\mathbb{P}$,

$$\begin{aligned} &\sup_{\{P_i\}_{1\le i\le n}\in\mathbb{P}^n} (\prod_{i=1}^{n} P_i)(\{d(P,Q) > \epsilon\}) \\ \le &\sup_{\{P_i\}_{1\le i\le n}\in\mathbb{P}^n} (\prod_{i=1}^{n} P)(\{\sup_{f\in\mathcal{F}_\epsilon} |\int f d(P-Q)| > \epsilon\}). \end{aligned} \tag{5.10}$$

The result follows from (5.8) and (5.10) with $P = \mathbb{F}_n$ and $Q = \bar{P}_n$. □

The following is an immediate corollary to the proposition.

Corollary 5.1.1 *Let $(Y,\rho), \mathbb{P}, (\xi_1, \xi_2, \cdots, \xi_n)$ and d be as considered in Proposition 5.1. Let P_n^u denote the uniform distribution on s_n as defined in Convention 1. Then for any $\epsilon > 0$,*

$$\sup_{\{P_i\}_{1\le i\le n}\in\mathbb{P}^n} \int (\prod_{i=1}^{n} P_{\pi(i)})(\{d(\mathbb{F}_n^O, \mathbb{F}_n^E) > \epsilon\}) dP_n^u(\pi) \to 0 \text{ as } n \to \infty.$$

Proof : In view of Proposition 5.1, it is enough to show that for any $\epsilon > 0$,

$$P_n^u(\{d((\bar{P}_n^O)^*, (\bar{P}_n^E)^*) > \epsilon\}) \to 0 \text{ as } n \to \infty. \tag{5.11}$$

Fix $f \in C(Y)$. Let us denote $\int f dP_i$ by a_i and $\frac{1}{n}\sum_{i=1}^{n} a_i$ by $\bar{a}$. Then

$$\begin{aligned} &\int \{\frac{1}{(n-[n/2])} \sum_{\substack{i=1\\ i \text{ odd}}}^{n} (\int f dP_{\pi(i)}) - \frac{1}{[n/2]} \sum_{\substack{i=1\\ i \text{ even}}}^{n} (\int f dP_{\pi(i)})\}^2 dP_n^u(\pi) \\ = &\int \{\frac{1}{(n-[n/2])} \sum_{\substack{i=1\\ i \text{ odd}}}^{n} a_{\pi(i)} - \frac{1}{[n/2]} \sum_{\substack{i=1\\ i \text{ even}}}^{n} a_{\pi(i)}\}^2 dP_n^u(\pi) \\ = &\frac{1}{4} \int \{\frac{1}{(n-[n/2])} \sum_{\substack{i=1\\ i \text{ odd}}}^{n} a_{\pi(i)} - \bar{a}\}^2 dP_n^u(\pi) \\ \le &\frac{1}{4(n-[n/2])} \{\frac{1}{n}\sum_{i=1}^{n} (a_i - \bar{a})^2\} \end{aligned} \tag{5.12}$$

since the variance under sampling without replacement is less than or equal to the variance under sampling with replacement.

By (5.12) with arbitrary f, we note that the following analogue of (5.3) holds : For any $f \in C(X)$ and $\epsilon > 0$,

$$P_n^u(\{|\int f d((\bar{P}_n^O)^* - (\bar{P}_n^E)^*)| > \epsilon\}) \to 0 \text{ as } n \to \infty. \tag{5.13}$$

(5.11) follows from (5.13) exactly as in Proposition 5.1. □

Proposition 5.2 *Let $\underline{G}_n^O$ and $\underline{G}_n^E$ be empirical distributions of ξ_i's based on odd and even numbered observations (vide (11) of Chapter 3, of course they are not observable since ξ_i's are unknown constants). For any $\epsilon > 0$,*

$$\sup_{\{\xi_i\}_{1\le i\le n}\in \Xi^n} P_n^u(\{d((\underline{G}_n^O)^*, (\underline{G}_n^E)^*) > \epsilon\}) \to 0 \text{ as } n \to \infty$$

where d denotes the Prohorov metric on $\mathcal{G}$ as defined in (10) of Chapter 3.

Proof : The given expression

$$= \sup_{\{\xi_i\}_{1\le i\le n}\in \Xi^n} \int 1_{\{d((\underline{G}_n^O)^*, (\underline{G}_n^E)^*) > \epsilon\}}(\pi) dP_n^u(\pi)$$

$$= \sup_{\{\xi_i\}_{1\le i\le n}\in \Xi^n} \int (\prod_{i=1}^n \delta_{\xi_{\pi(i)}})(\{d(\underline{G}_n^O, \underline{G}_n^E) > \epsilon\}) dP_n^u(\pi)$$

(where δ_ξ denote the degenerate distribution at $\{\xi\}$)

$$\le \sup_{\{G_i\}_{1\le i\le n}\in \mathcal{G}^n} \int (\prod_{i=1}^n G_{\pi(i)})(\{d(\underline{G}_n^O, \underline{G}_n^E) > \epsilon\}) dP_n^u(\pi)$$

$\to 0$ as $n \to \infty$ by Corollary 5.1.1 with $Y = \Xi$ and (hence) $\mathbb{P} = \mathcal{G}$. □

Corollary 5.2.1 *There is a sequence $\{\epsilon_n^\circ\}_{n\ge 1}$ decreasing to zero such that*

$$\sup_{\{\xi_i\}_{1\le i\le n}\in \Xi^n} P_n^u(\{d((\underline{G}_n^O)^*, (\underline{G}_n^E)^*) > \epsilon_n^\circ\}) \to 0 \text{ as } n \to \infty.$$

Proof : The result follows trivially from the proposition. □

In view of the corollary it is natural to consider for any $n \ge 1$ and $\epsilon > 0$,

$$\alpha_n(\epsilon) := \{\{\xi_i\}_{1\le i\le n} : d(\underline{G}_n^O, \underline{G}_n^E) \le \epsilon\} \tag{5.14}$$

Fix any sequence $\{\epsilon_n\}_{n\geq 1}$ decreasing to zero. Let $\theta_o \in \Theta$. Let $\{\xi_{ni}\}_{1\leq i\leq n, n\geq 1}$ be a triangular array of elements in Ξ such that

$$\{\xi_{ni}\}_{1\leq i\leq n} \in \alpha_n(\epsilon_n) \ \forall n. \tag{5.15}$$

Corollary 5.2.1 leads to an analysis of the following triangular array version of Model I.

Model I(t) : Let $\{X_{ni}\}_{1\leq i\leq n, n\geq 1}$ be a triangular array of rowwise independent random variables with X_{ni} following the distribution $P_{\theta_o, \xi_{ni}}$, where $\theta_o \in \Theta$ and the triangular array $\{\xi_{ni}\}_{1\leq i\leq n, n\geq 1}$ satisfies (5.15).

Convention 2 Let $(Y, \mathcal{Y})$ and $(Z, \mathcal{Z})$ be as considered in Definition 5.1. Let $\{y_{ni}\}_{1\leq i\leq n, n\geq 1}$ be a triangular array of elements in Y. For any Z-valued statistic V_n on $(Y, \mathcal{Y})^n$ we shall denote the statistic $V_n(\{y_{ni}\}_{i\leq i\leq n})$ by $V_{n,n}$.

The above convention suggests obvious Model I(t)-analogue of equation (4.1) which we shall denote by $(4.1)^t$.

(5) As in observation (2), Definitions 3.1-3.4 have obvious Model I(t)-analogues, and for any property P defined in Definitions 3.1-3.4 and statistic V_n, V_n satisfies P (I) *only if* $V_{n,n}$ satisfies P (I(t)).

Let ψ be a k Fix θ_o in Θ and $\{\xi_{ni}\}_{1\leq i\leq n, n\geq 1}$ satisfying relation (5.15). The following is the Model I(t)-analogue of relation (4.2).

$$\begin{aligned}\widetilde{D}_{n,n}(\theta) \;:=\; & \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ odd}}}^{n} \{\psi(X_{ni}, \theta, \widehat{G}^E_{n,n}) - \psi(X_{ni}, \theta_o, \underline{G}_{n,n}) \\ & \qquad + (\theta - \theta_o) \int \psi(\cdot, \theta_o, \underline{G}_{n,n}) f'(\cdot, \theta_o, \underline{G}_{n,n}) d\mu(\cdot)\} \\ & + \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ even}}}^{n} \{\psi(X_{ni}, \theta, \widehat{G}^O_{n,n}) - \psi(X_{ni}, \theta_o, \underline{G}_{n,n}) \\ & \qquad + (\theta - \theta_o) \int \psi(\cdot, \theta_o, \underline{G}_{n,n}) f'(\cdot, \theta_o, \underline{G}_{n,n}) d\mu(\cdot)\}\end{aligned} \tag{5.16}$$

for all θ in Θ.

The following conditions are the Model I(t)-analogues of conditions (i)-(v) and U(i)-U(vi) of Chapter 4.

(i)^t (a) Condition (i) of Chapter 4 holds, with G_o replaced by $\underline{G}^O_{n,n}$ and $\underline{G}^E_{n,n}$, uniformly with respect to $n \geq 1$

and (b) $\int \frac{\{f'(\cdot,\theta_o,G)-f'(\cdot,\theta_o,\underline{G}_{n,n})\}^2}{f(\cdot,\theta_o,\underline{G}_{n,n})} d\mu(\cdot)|_{G=\underline{G}^O_{n,n} \text{ or } \underline{G}^E_{n,n}} \to 0$ as $n \to \infty$.

(ii)^t The following two statements hold uniformly in $n \geq 1$

(a) there is $\delta_o > 0$ such that condition (ii)(a) of Chapter 4 holds with G_o replaced by $\underline{G}^O_{n,n}$ and $\underline{G}^E_{n,n}$

and (b) i) $\int\{\psi(\cdot,\theta,G) - \psi(\cdot,\theta_o,\underline{G}^E_{n,n})\}^2 f(\cdot,\theta_o,\underline{G}^O_{n,n})d\mu(\cdot) \to 0$ as $(\theta,G) \to (\theta_o,\underline{G}^E_{n,n})$

and ii) $\int\{\psi(\cdot,\theta,G) - \psi(\cdot,\theta_o,\underline{G}^O_{n,n})\}^2 f(\cdot,\theta_o,\underline{G}^E_{n,n})d\mu(\cdot) \to 0$ as $(\theta,G) \to (\theta_o,\underline{G}^O_{n,n})$.

(iii)^t Condition (iii) of Chapter 4 holds with $(\hat{G}_n, G_o)$ replaced by $(\hat{G}^E_{n,n}, \underline{G}^O_{n,n})$, $(\underline{G}_{n,n}, \underline{G}^O_{n,n})$, $(\hat{G}^O_{n,n}, \underline{G}^E_{n,n})$ and $(\underline{G}_{n,n}, \underline{G}^E_{n,n})$.

(iv)^t (a) There is $\delta_o > 0$ such that condition (iv)(a) of Chapter 4 holds with G_o replaced by $\underline{G}_{n,n}$,

(b) condition (iv)(b) of Chapter 4 holds, with G_o replaced by $\underline{G}_{n,n}$, uniformly in $n \geq 1$

and (c) $\{\int \psi(\cdot,\theta_o,\underline{G}_{n,n}) f'(\cdot,\theta_o,\underline{G}_{n,n}) d\mu(\cdot) : n \geq 1\}$ does not contain zero as a limit point.

(v)^t There is $\delta_o > 0$ such that condition (v) of Chapter 4 holds, with G_o replaced by $\underline{G}_{n,n}$, uniformly in $n \geq 1$.

Let $\delta_o > 0$ be as considered in (ii)^t, (iv)^t and (v)^t. As before, for any condition C among (i)^t-(v)^t, UC denotes the condition that condition C holds, with θ_o, θ and θ' replaced by θ, θ' and θ'', respectively, uniformly with respect to θ, θ' and θ'' in $B(\theta_o, \delta_o)$, G in $B(\underline{G}_{n,n}, \delta_o)$ and $\{\xi_{ni}\}_{1\leq i\leq n}$ in $\alpha_n(\epsilon_n)$. Condition U(vi)^t is given below.

U(vi)^t (a) $\sup_{n\geq 1} \sup_{(\theta,\{\xi_{ni}\}_{1\leq i\leq n})\in B(\theta_o,\delta_o)\times\alpha_n(\epsilon_n)}$

$$\left[\frac{\int 1_{\{|\psi(\cdot,\theta,\underline{G}_{n,n})|\geq K\}}(x)\psi^2(x,\theta,\underline{G}_{n,n})f(x,\theta,\underline{G}_{n,n})d\mu(x)}{J(\theta,\underline{G}_{n,n},\psi)}\right] \to 0$$

as $K \to \infty$

and (b) condition U(vi)(b) of Chapter 4 holds.

Lemma 5.1(t) *Assume (C1)(b). Fix any sequence* $\{\epsilon_n\}_{n\geq 1}$ *decreasing to zero. Fix* θ_o *in* Θ *and* $\{\xi_{ni}\}_{1\leq i\leq n, n\geq 1}$ *satisfying relation (5.15). Let* ψ *be a kernel. Let* $\widetilde{D}_{n,n}$ *be as defined by relation (5.16). Also, whenever it makes sense, let* $T_{n,n}(\psi)$ *be the estimate defined through Definition 4.1 and Convention 2. We can conclude the following.*

(I) If conditions (i)^t-(iii)^t hold, then for all $c > 0$ *and* $\epsilon > 0$,

$$\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} \left(\prod_{i=1}^{n} P_{\theta_o,\xi_{ni}}\right)(\{\widetilde{D}_{n,n}(\theta) > \epsilon\}) \to 0 \text{ as } n \to \infty.$$

(II) If conditions (i)^t-(iv)^t hold, then

(A) for any sequence $\{c_n\}_{n\geq 1}$ *increasing to infinity*

$$\left(\prod_{i=1}^{n} P_{\theta_o,\xi_{ni}}\right)(E_n) \to 1 \text{ as } n \to \infty$$

where E_n *denotes the event that there is a solution of (4.1)^t lying inside the interval* $(\theta_o - c_n/\sqrt{n}, \theta_o + c_n/\sqrt{n})$

and (B) under assumption (C1)(a), $T_{n,n}(\psi)$ *is a* $\sqrt{n}$*-consistent solution (I(t)) of (4.1)^t.*

(III) If conditions (i)^t-(v)^t hold, then

(A) for any $c > 0$ *and* $\epsilon > 0$,

$$\left(\prod_{i=1}^{n} P_{\theta_o,\xi_{ni}}\right)(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} |\widetilde{D}_{n,n}(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty$$

and (B) under assumption (C1)(a),

$$\sup_{x\in \mathbb{R}} |(\prod_{i=1}^{n} P_{\theta_o,\xi_{ni}})(\{\sqrt{n}(T_{n,n}(\psi)-\theta_o)\leq x\}) - \Phi(x/V(\theta_o, \underline{G}_{n,n}, \psi))|$$

$\to 0$ *as* $n\to\infty$, *where* V *denote the function defined in (9) of Chapter 3.*

(IV) *As in Lemma 4.1(IV), for any conclusion* C *among (I)-(III), let* UC *denote the conclusion that* C *holds unformly with respect to* θ_o *in compact subsets of* Θ *and* $\{\xi_{ni}\}_{1\leq i\leq n, n\geq 1}$ *satisfying (5.15). Then* $U(I)$, $U(II)$ *and* $U(III)(A)$ *hold if the relevant conditions among* $U(i)^t$-$U(v)^t$ *hold whereas* $U(III)(B)$ *holds if conditions* $U(i)^t$-$U(vi)^t$ *hold.*

We can prove this by an easy modification of the proof of Lemma 4.1. Note that

(6) Lemma 5.1(t) is only an auxiliary result needed to prove our main result, namely Lemma 5.1(IV), the assumptions for which are only slightly stronger than those of Lemma 4.1(IV), vide observation (7) preceeding Lemma 5.1.

Let us now consider the original set-up, namely Model I with randomised estimates.

For any $n\geq 1$ and $\{\xi_i\}_{1\leq i\leq n}$ in Ξ^n, define

$$\beta_n(\{\xi_i\}_{1\leq i\leq n}) = \{\pi\in s_n : \{\xi_{\pi(i)}\}_{1\leq i\leq n}\in \alpha_n(\epsilon_n^o)\}, \tag{5.17}$$

where the sequence $\{\epsilon_n^o\}_{n\geq 1}$ is defined in Corollary 5.2.1.

Let ψ be a kernel. Fix θ_o in Θ and $\{\xi_n\}_{n\geq 1}$ in Ξ^∞. Consider the following analogue of (5.16) in the present context.

$$\widetilde{D}_n^*(\theta) \quad := \quad \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ odd}}}^{n} \{\psi(X_i^*, \theta, (\widehat{G}_n^E)^*) - \psi(X_i^*, \theta_o, \underline{G}_n)$$

$$+(\theta - \theta_o) \int \psi(\cdot, \theta_o, \underline{G}_n) f'(\cdot, \theta_o, \underline{G}_n) d\mu(\cdot)\}$$

$$+\frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ even}}}^{n} \{\psi(X_i^*, \theta, (\widehat{G}_n^O)^*) - \psi(X_i^*, \theta_o, \underline{G}_n)$$

$$+(\theta - \theta_o) \int \psi(\cdot, \theta_o, \underline{G}_n) f'(\cdot, \theta_o, \underline{G}_n) d\mu(\cdot)\} \tag{5.18}$$

for all θ in Θ.

Consider the following conditions uniformly with respect to π in $\beta_n(\{\xi_i\}_{1 \leq i \leq n})$.

(i)* (a) $$\lim_{\theta \to \theta_o} \limsup_{n \to \infty} \int [\{\frac{A(\cdot, \theta_o, G, \theta, G) - 1}{(\theta - \theta_o)} - s_\theta(\cdot, \theta_o, G)\}^2$$

$$f(\cdot, \theta_o, G)]_{G = (\underline{G}_n^O)^* \text{ or } (\underline{G}_n^E)^*} d\mu(\cdot) = 0,$$

where s_θ denotes the kernel defined by relation (3.20)

and (b) $$\int \frac{\{f'(\cdot, \theta_o, G) - f'(\cdot, \theta_o, \underline{G}_n)\}^2}{f(\cdot, \theta_o, \underline{G}_n)} d\mu(\cdot)|_{G = (\underline{G}_n^O)^* \text{ or } (\underline{G}_n^E)^*} \to 0$$

as $n \to \infty$.

(ii)* (a) There is $\delta_o > 0$ such that

$$\limsup_{n \to \infty} \sup_{(\theta, G') \in B((\theta_o, G), \delta_o)} \int \psi^2(\cdot, \theta, G')$$

$$f(\cdot, \theta_o, G) d\mu(\cdot)|_{G = (\underline{G}_n^O)^* \text{ or } (\underline{G}_n^E)^*} < \infty$$

and (b) $$\limsup_{n \to \infty} \sup_{(\theta, G') \in B((\theta_o, G), \epsilon_n^o)} \int \{\psi(\cdot, \theta, G') - \psi(\cdot, \theta_o, G)\}^2$$

$$f(\cdot, \theta_o, G) d\mu(\cdot)|_{G = (\underline{G}_n^O)^* \text{ or } (\underline{G}_n^E)^*} = 0.$$

(iii)* Assumption (C1)(b) holds with a choice of $\widehat{G}_n$ so that

$$\limsup_{n \to \infty} [\sup_{\{\theta : |\theta - \theta_o| \leq c/\sqrt{n}\}} (\prod_{i=1}^{n} P_{\theta_o, \xi_i})(\{\sqrt{n} | \int \psi(\cdot, \theta, G') f(\cdot, \theta, G) d\mu(\cdot)| > \epsilon\})]$$

$= 0$, where $(G, G') = ((\underline{G}_n^O)^*, (\widehat{G}_n^E)^*), ((\underline{G}_n^O)^*, \underline{G}_n), ((\underline{G}_n^E)^*, (\widehat{G}_n^O)^*)$ or $((\underline{G}_n^E)^*, \underline{G}_n)$.

(iv)* (a) There is $\delta_o > 0$ and $n_o \geq 1$ such that for all $n \geq n_o$, x in S and G in $B(\underline{G}_n, \delta_o)$,

$$\psi(x, \cdot, G) \in C(B(\theta_o, \delta_o)),$$

(b) $\limsup_{n\to\infty}[\int \psi^2(\cdot, \theta_o, \underline{G}_n) f(\cdot, \theta_o, \underline{G}_n) d\mu(\cdot)] < \infty$ (This condition follows from condition (ii)*(a) but is given separately for ease in later referrences.)

and (c) $\liminf_{n\to\infty} | \int \psi(\cdot, \theta_o, \underline{G}_n) f'(\cdot, \theta_o, \underline{G}_n) d\mu(\cdot)| > 0$.

(v)* There is $\delta_o > 0$ and $A(\cdot, \theta_o, \underline{G}_n) \in L_1(f(\cdot, \theta_o, \underline{G}_n))$ such that

$$|\psi(\cdot, \theta', G) - \psi(\cdot, \theta, G)| \leq |\theta' - \theta| A(\cdot, \theta_o, \underline{G}_n)$$

for all θ, θ' in $B(\theta_o, \delta_o)$ and G in $B(\underline{G}_n, \delta_o)$.

Analogous to the formulation of the conditions U(i)-U(v) on the basis of the conditions (i)-(v) in Chapter 4, we formulate the conditions U(i)*-U(v)*. An additional condition U(vi)* is given below.

U(vi)* (a) There is $\delta_o > 0$ such that

$$\limsup_{n\to\infty} \sup_{(\theta,G)\in B((\theta_o,\underline{G}_n),\delta_o)} \left[\frac{\int 1_{\{\psi^2(\cdot,\theta,G)\geq K\}}(x)\psi^2(x,\theta,G) f(x,\theta,G) d\mu(x)}{J(\theta, G, \psi)}\right]$$

$\to 0$ as $K \to \infty$

and (b) condition U(vi)(b) of Chapter 4 holds.

Note that

(6) Any condition among U(ii)*-U(vi)* is equivalent to the corresponding condition among U(ii)-U(vi) of Chapter 4 whereas condition U(i)* is a stronger version of condition U(i) of Chapter 4 with U(i)*(a) equivalent to it.

The following is the required analogue of Lemma 4.1.

Lemma 5.1 *Assume (C1)(b). Fix θ_o in Θ and $\{\xi_n\}_{n\geq 1}$ in Ξ^∞. Let ψ be a kernel. Let $\widetilde{D}_n^*$ be as defined in relation (5.18). Also, whenever it makes sense, let $T_n^*(\psi)$ be the estimate defined in Convention 1. We can draw the following conclusions.*

(I) If conditions (i)-(iii)* hold, then for all $c > 0$ and $\epsilon > 0$,*

$$\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} (\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{\widetilde{D}_n^*(\theta) > \epsilon\}) \to 0 \text{ as } n \to \infty$$

(II) If conditions (i)-(iv)* hold, then*

(A) for any sequence $\{c_n\}_{n\geq 1}$ increasing to infinity

$$(\prod_{i=1}^{n} P_{\theta_o,\xi_i})(E_n) \to 1 \text{ as } n \to \infty$$

where E_n denotes the event that there is a solution of (4.1) lying inside the interval $(\theta_o - c_n/\sqrt{n}, \theta_o + c_n/\sqrt{n})$*

and (B) under assumption (C1)(a), $T_n^*(\psi)$ is a randomised $\sqrt{n}$-consistent solution (I) of (4.1).*

(III) If conditions (i)-(v)* hold, then*

(A) for any $c > 0$ and $\epsilon > 0$,

$$(\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{\sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} |\widetilde{D}_n^*(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty$$

and (B) under assumption (C1)(a),

$$\sup_{x\in\mathbb{R}} |(\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{\sqrt{n}(T_n^*(\psi) - \theta_o) \leq x\}) - \Phi(x/V(\theta_o, \underline{G}_n, \psi))|$$

$\to 0$ as $n \to \infty$, where V denote the function defined in (9) of Chapter 3.

(IV) As in Lemma 4.1(IV), for any conclusion C among (I)-(III), let UC denote the conclusion that C holds unformly with respect to $(\theta_o, \{\xi_n\}_{n\geq 1})$ in compact subsets of $\Theta \times \Xi^\infty$. Then U(I), U(II) and U(III)(A) hold if the relevant conditions among U(i)-U(v)* hold whereas U(III)(B) holds if conditions U(i)*-U(vi)* hold.*

Proof : Observe that for all $n \geq 1$, $d((\underline{G}_n^O)^*, (\underline{G}_n^E)^*) \leq \epsilon_n^o$ if and only if $\{\xi_i^*\}_{1\leq i\leq n} \in \alpha_n(\epsilon_n^o)$ (vide relation (5.14)) so that Corollary 5.2.1 can be restated as

$$\sup_{\{\xi_i\}_{1\leq i\leq n}\in\Xi^n} [1 - P_n^u(\beta_n(\{\xi_i\}_{1\leq i\leq n}))] \to 0 \text{ as } n \to \infty \tag{5.19}$$

where β_n's are as defined in relation (5.17).

We shall now prove part (I) of the lemma and then indicate a proof of part U(I) of it. The other parts can be proved similarly.

For this purpose note that conditions (i)*-(iii)* imply that for any θ_o in Θ, $\{\xi_n\}_{n\geq 1}$ in Ξ^∞ and sequence of permutations $\{\pi_n\}_{n\geq 1}$ with π_n in $\beta_n(\{\xi_i\}_{1\leq i\leq n})$, conditions (i)^t-(iii)^t with $\epsilon_n = \epsilon_n^o$ hold at the point θ_o in Θ and triangular array $\{\xi_{\pi_n(i)}\}_{1\leq i\leq n, n\geq 1}$ which satisfies (5.15) by the choice of π_n's. Hence by part (I) of Lemma 4.1(t) for any $c > 0$ and $\epsilon > 0$,

$$\sup_{\pi\in\beta_n(\{\xi_i\}_{1\leq i\leq n})} \sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} (\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{|\widetilde{D}_n^*(\theta)| > \epsilon | \Pi = \pi\}) \tag{5.20}$$

$\to 0$ as $n \to \infty$.

Let $A_n = \Xi^n$. For any α in A_n and π in $\beta_n(\alpha)$, let

$$f_n(\pi, \alpha) := \sup_{\{\theta:|\theta-\theta_o|\leq c/\sqrt{n}\}} (\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{|\widetilde{D}_n^*(\theta)| > \epsilon | \Pi = \pi\}).$$

By (5.20)

$$\sup_{\pi\in\beta_n(\alpha)} f_n(\pi, \alpha) \to 0 \text{ as } n \to \infty$$

and f_n is $[0, 1]$-valued.

Hence, by (5.19)

$$\int_{s_n} f_n(\cdot, \alpha) dP_n^u(\cdot) = \int_{\beta_n(\alpha)} f_n(\cdot, \alpha) dP_n^u(\cdot) + \int_{\beta_n^c(\alpha)} f_n(\cdot, \alpha) dP_n^u(\cdot) \to 0$$

as $n \to \infty$, proving part (I). Part U(I) follows similarly from the uniform versions of (5.19) and (5.20) with A_n replaced by relevant compact subset of $\Theta \times \Xi^n$. □

Defintion 4.2 has an obvious extension for randomised estimates. The following is the Model I-analogue of the extension.

Definition 5.3 Any kernel ψ satisfying U(ii)*-U(vi)* will be called an *estimable kernel in Model I* (or, in short, an *EK (I)*) and any randomised $\sqrt{n}$-consistent solution (I) of (4.1), *i.e.*, any uniformly $\sqrt{n}$-consistent solution (I) of a randomisation of (4.1) namely,

$$\frac{1}{\sqrt{n}}\sum_{\substack{i=1\\ i\,\mathrm{odd}}}^{n}\psi(X_i^*(P_n),\theta,(\hat{G}_n^E)^*(P_n))+\frac{1}{\sqrt{n}}\sum_{\substack{i=1\\ i\,\mathrm{even}}}^{n}\psi(X_i^*(P_n),\theta,(\hat{G}_n^O)^*(P_n))=0$$

for some probability measure P_n of s_n will be called a *generalised* C_1-*estimate in Model I corresponding to* ψ (or, in short, a GC_1 *(I) estimate*).

There is an obvious analogue of Lemma 4.1a for Model I and randomised estimates and in view of observation (7), we can make the following remark.

Remark 5.1 For any kernel ψ, ψ is EK (I) *if and only if* it is EK (II) whereas for any randomised estimate V_n of θ_o, V_n is GC_1 (I) *only if* it is GC_1 (II). Also, one can easily verify that Examples 4.1 and 4.2, with $T_n(\psi)$ replaced by $T_n^*(\psi)$ for the latter one, remain valid for Model I.

The following is the Model I-analogue of Remark 4.3.

Remark 5.2 If $(S,\mathcal{S})=(I\!R^p,\mathcal{B}^p)$ and assumptions (A1) and (B2)(a) hold then Corollaries 3.1.1 and 3.1.2 enable us to drop assumption (C1)(b) even if Θ is unbounded.

Let us now write down the analogues of Theorems 4.2 and 4.3.

Theorem 5.2 *Assume (C1), (B2) and (B3). The (randomised) estimate* Z_n^* *of* θ_o, *as defined through relations (4.15)-(4.16), Definition 4.1 and Convention 1, is UAN (I) with AV* $(1/I)$.

Theorem 5.3 *Assume (C1), (B2) and (B3s). The (randomised) estimate* $T_n^*(\bar{\psi})$ *of* θ_o, *as defined through Definitions 4.1, 5.1 and Convention 1, is UAN (I) with AV* $(1/I)$.

Remark 5.3 Theorems 5.2 and 5.3 tell us that Z_n^* and $T_n^*(\bar{\psi})$ have the most limiting concentration around θ_o among the randomised regular (I) estimates, *i.e.*, the following holds.

For any $(\theta_o, \{\xi_n\}_{n\geq 1})$ in $\Theta \times \Xi^\infty$, randomised regular (I) estimate V_n of θ_o and convex symmetric set A in $\mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
& \lim_{n\to\infty} (\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{\sqrt{n}\, I^{\frac{1}{2}}(\theta_o, \underline{G}_n)(W_n - \theta_o) \in A\}) \\
= \;& P(\mathcal{N}(0,1) \in A) \\
\geq \;& \limsup_{n\to\infty} (\prod_{i=1}^{n} P_{\theta_o,\xi_i})(\{\sqrt{n}\, I^{\frac{1}{2}}(\theta_o, \underline{G}_n)(V_n - \theta_o) \in A\})
\end{aligned}$$

where $W_n = Z_n^*$ or $T_n^*(\bar{\psi})$.

Remark 5.4 It has been pointed out by van der Vaart (1987) as criticism of the regular estimates that given any regular estimate one can construct a non-regular asymptotically normal estimate which is better. To some extent the idea of such a construction is implicit in a grouping technique introduced in a paper by Chaterjee and Das (1983) as variance estimation. However, such better estimates due to van der Vaart is, of necessity, non-symmetric in $X_1, X_2, \ldots, X_n$. This makes one reluctant to use them. Moreover, from a technical point of view, one should compare its maximum risk, over permutations of $\xi_1, \xi_2, \ldots, \xi_n$, with the risk of a regular estimate. This is a matter that requires further examination. In this connection it would be interesting to study the efficient regular estimate in Example 1.2 with the best equivariant estimate that exists if $(\xi_1, \xi_2, \cdots, \xi_n)$ is known upto a permutation. We hope to study this in a further communication.

Remark 5.5 There can be no asymptotic improvement over efficient estimates of the kind discussed in the previous paragraph if the optimal kernel does not depend on G. Typical situations where this happens are discussed in Lindsay (1980) and Pfanzagl (1982) (see also Chapter 6(b)). In particular, this holds for the estimate in Example 1.1. We omit proof.

Remark 5.6 If the dimension q_i of $\boldsymbol{X}_i$ is not constant one can group the observations according to their dimensions, Let us now consider the special case where the distinct values of q_i, i running from 1 to n, remain fixed as n tends to infinity, in other words, there are finitely many such groups. Let us rearrange the observations to get an array of independent random variables

$$\begin{array}{cccc} Y_{11} & Y_{12} & \cdots & Y_{1n_1} \\ Y_{21} & Y_{22} & \cdots & Y_{2n_2} \\ \cdots & \cdots & \cdots & \cdots \\ Y_{r1} & Y_{r2} & \cdots & Y_{rn_r} \end{array}$$

with Y_{jl}'s following $f(\cdot, \theta_o, \xi_{jl}, k_j)$ and n_j being non-negative integers with $\sum_{j=1}^r n_j = n$. Without loss of generality, let us assume that $k_1 < k_2 < \cdots < k_r$ and $\liminf(\frac{n_j}{n}) > 0$ for all j, so that each group represents a distinct fixed set-up model by itself. Call an estimate of θ_o regular (in the new model) if it is uniformly asymptotically equivalent to a pooled mean of regular estimates (including the randomised ones) as defined through Definitions 1.1, 5.1-5.2 and observation (2), corresponding to each component fixed set-up submodel. For the j-th submodel, let $\bar{\psi}_j$ denote the optimal kernel as defined through relations (3.20)-(3.22), U_{nj} and $\widehat{G}_{nj}$ denote, respectively, the uniformly $\sqrt{n}$-consistent estimate of θ_o and uniformly consistent estimate of $\underline{G}_{nj} := I\!\!F_{n_j}(\{\xi_{jl}\}_{1 \le l \le n_j})$ (vide Definitions 3.1-3.2) as considered in assumption (C1), the superscript $*j$ stands for the operation of randomisation as defined in Definition 5.1 and the superscripts O and E stand for the operations defined in (11) of Chapter 3. Then an efficient regular estimate will be a solution of

$$\sum_{j=1}^{r} \sum_{\substack{i=1 \\ i\,\mathrm{odd}}}^{n_j} \psi(Y_{jl}^{*j}, \theta, (\widehat{G}_{nj}^{E})^{*j}) + \sum_{j=1}^{r} \sum_{\substack{i=1 \\ i\,\mathrm{even}}}^{n_j} \psi(Y_{jl}^{*j}, \theta, (\widehat{G}_{nj}^{O})^{*j}) = 0 \qquad (5.21)$$

which is nearest to $\bar{U}_{n\cdot}$ if there is a solution of (5.21) lying inside $[\bar{U}_{n\cdot} - \log n/\sqrt{n}, \bar{U}_{n\cdot} + \log n/\sqrt{n}]$ and equal to $\bar{U}_{n\cdot}$ otherwise; where $\bar{U}_{n\cdot} = \frac{1}{n}\sum_{j=1}^{r} n_j U_{nj}$.

Remark 5.7 In view of Remarks 5.2 and 4.4, for Euclidean S and exponential f, it is enough to check assumption (C1)(a), *i.e.*, the existence of a uniformly $\sqrt{n}$-consistent (I) estimate of θ_o, and assumptions (B3) and (B3s), *i.e.*, smoothness properties of the optimal kernel (*cf.* Remark 4.8).

Chapter 6
Two Special Cases

In this chapter, we shall discuss two special cases referred to in Chapter 1 where the optimal kernel $\bar{\psi}$ is "smooth". Throughout the discussion, we are assuming the validity of assumptions (A2) and (A3).

(a) *Orthogonal Case :* This is a generalised version of the symmetric location-scale problem with known functional form of the density f, as in Example 1.2. Here, for all (θ, G), $s_\theta(\cdot, \theta, G)$ belongs to the orthogonal complement of the space $N_{\theta,G}$, so that s_θ itself is a version of the optimal kernel.

Let us assume that

(D1) (a) For all x in S, $f(x, \cdot, \cdot) \in C_{2,0}(\bar{\Theta} \times \Xi)$

and (b) for any compact subset Θ_o of Θ the following statements hold

(i) there is δ_o

a) the following two families of functions

$$\{\frac{(f')^2(\cdot, \theta', G)}{f(\cdot, \theta, G)} : \theta, \theta' \in \Theta_o \text{ with } |\theta - \theta'| \leq \delta_o, G \in \mathcal{G}\}$$

and

$$\{s_\theta^2(\cdot, \theta', G') f(\cdot, \theta, G) : (\theta, G), (\theta', G') \in \Theta \times \mathcal{G} \text{ with } |\theta - \theta'| + d(G, G') \leq \delta_o\}$$

are uniformly integrable with respect to μ

and b) $\sup_{(\theta,G) \in \Theta_o \times \mathcal{G}} [\int \overline{(|s'_\theta|)}(\cdot, B((\theta, G), \delta_o)) f(\cdot, \theta, G) d\mu(\cdot)] < \infty$

and (ii) $\sup_{(\theta,G) \in \Theta_o \times \mathcal{G}} [\{\int 1_{\{s_\theta^2(\cdot,\theta,G) \geq K\}} \frac{(f')^2(\cdot, \theta, G)}{f(\cdot, \theta, G)} d\mu(\cdot)\} / I(\theta, G)]$

$\to 0$ as $K \to \infty$.

Assumption (D1) and orthogonality together imply assumptions (B2) and (B3s). Hence by the theorems proved in Chapters 4 and 5, Z_n and $T_n(\bar{\psi})$ are efficient (II) and Z_n^* and $T_n^*(\bar{\psi})$ are efficient (I) as well as efficient (II), both under assumptions (C1) and (D1).

We have verified assumptions (A1) and (D1) for Euclidean S and exponential f as considered in Remark 4.4. In particular, they hold for Example 1.2 with $p \geq 2$.

Example 1.2 with $p = 1$ does not fall in the exponential families described in Remark 4.4. However in this case one can easily verify assumption (D1). The verification of assumption (A1) is as follows : Let (θ, G), (θ', G') be such that $f(\cdot, \theta, G) = f(\cdot, \theta', G')$ a.e.$[\lambda]$. By symmetry of the normal density function we get $\theta = \theta'$. So, it remains to prove, for all θ in $\bar{\Theta}$, the identifiability of G. In this respect, let us observe that conditions (b)-(d) of Remark 4.4, have obvious modifications guaranteeing, for any θ in $\bar{\Theta}$, the identifiability of G. We have verified these conditions for Example 1.2 so that $G = G'$ and hence the validity of assumption (A1).

In view of Remark 4.7 and observation 1) of Chapter 3, it remains to check assumption (C1)(a) for Example 1.2 and in this respect the grand mean $\bar{\bar{X}} = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} X_{ij}$ is a natural choice for U_n.

In view of the last two paragraphs Theorems 4.2, 4.3, 5.2 and 5.3 hold for Example 1.2 with arbitrary p. An asymptotically efficient estimate for Example 1.2 with arbitrary p can also be obtained from the results of van der Vaart (1987, pp 89-93).

(b) *Case of Partial Likelihood Factorization :* This case in the present context was first considered by Lindsay (1980). Here the likelihood function f factorizes in the following manner.

There are Borel-measurable functions $p : S \times \bar{\Theta} \to \mathbb{R}^+$ and $q : S \times \bar{\Theta} \times \Xi \to \mathbb{R}^+$ such that

$$f(x, \theta, \xi) = p(x, \theta) q(x, \theta, \xi) \text{ for all } (x, \theta, \xi) \in S \times \bar{\Theta} \times \Xi \quad (6.1)$$

$$\text{and } \int p(\cdot, \theta') q(\cdot, \theta, \xi) d\mu(\cdot) = 1 \text{ for all } (\theta, \theta', \xi) \in \bar{\Theta} \times \bar{\Theta} \times \Xi. \quad (6.2)$$

In cases where (6.1) and (6.2) hold we call p a *partial likelihood function*.

In applications, for (6.1) and (6.2) to hold one assumes the existence of either a partially sufficient statistic t for ξ or a ξ-ancillary statistic c. In the first case q is the marginal of t and in the second p is the marginal of c. Example 1.1 falls in the first case with $t(\boldsymbol{X}_i) = \bar{X}_i$. (An example of the other kind is Example 9.4 of Lindsay (1980, pp 654-655).)

Here

$$s_\theta = \frac{p'}{p} + \frac{q'}{q}. \tag{6.3}$$

Assume that

(D2) (a) For all x in S, $p(x,\cdot) \in C_2(\bar{\Theta})$ and $q(x,\cdot,\cdot) \in C_{2,0}(\bar{\Theta} \times \mathcal{G})$ and (b) for any compact subset Θ_o of Θ the following statements hold

(i) there is $\delta_o > 0$ such that

a) the following three families of functions

$$\left\{\frac{(p')^2(\cdot,\theta')q^2(\cdot,\theta',G)}{p(\cdot,\theta)q(\cdot,\theta,G)} : \theta,\theta' \in \Theta_o \text{ with } |\theta-\theta'| \le \delta_o \text{ and } G \in \mathcal{G}\right\}$$

$$\left\{\frac{p^2(\cdot,\theta')(q')^2(\cdot,\theta',G)}{p(\cdot,\theta)q(\cdot,\theta,G)} : \theta,\theta' \in \Theta_o \text{ with } |\theta-\theta'| \le \delta_o \text{ and } G \in \mathcal{G}\right\}$$

and

$$\left\{\left(\frac{p'}{p}\right)^2(\cdot,\theta')p(\cdot,\theta)q(\cdot,\theta,G) : \theta,\theta' \in \Theta_o \text{ with } |\theta-\theta'| \le \delta_o \text{ and } G \in \mathcal{G}\right\}$$

are uniformly integrable with respect to μ

and b)

$$\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} \left[\int \overline{\left(\left|\frac{p''}{p}\right|\right)}(\cdot, B(\theta,\delta_o))p(\cdot,\theta)q(\cdot,\theta,G)d\mu(\cdot)\right] < \infty$$

and

$$\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}} \left|\int \overline{\left(\left(\frac{p'}{p}\right)^2\right)}(\cdot, B(\theta,\delta_o))p(\cdot,\theta)q(\cdot,\theta,G)d\mu(\cdot)\right| < \infty$$

and (ii) $\sup_{(\theta,G)\in\Theta_o\times\mathcal{G}}[\int 1_{\{(\frac{p'}{p})^2(\cdot,\theta)\geq K\}}\frac{(p')^2(\cdot,\theta)q(\cdot,\theta,G)}{p(\cdot,\theta)}d\mu(\cdot)]$

$$[\int \frac{(p')^2(\cdot,\theta)q(\cdot,\theta,G)}{p(\cdot,\theta)}d\mu(\cdot)]^{-1} \to 0 \text{ as } K \to \infty.$$

(D3) For any $(\theta, G) \in \bar{\Theta} \times \mathcal{G}$, there is $M_{\theta,G} \in \mathcal{M}_o$ such that

$$\frac{q'(x,\theta,G)}{q(x,\theta,G)} = \frac{q(x,\theta,M_{\theta,G})}{q(x,\theta,G)} \text{ for all } (x,\theta,G) \text{ in } S \times \bar{\Theta} \times \mathcal{G}.$$

Clearly, assumption (D2) implies assumption (B2). From assumption (D3) and relation (6.3) we have $\bar{\psi} = \frac{p'}{p}$ so that assumptions (D2) and (D3) together imply assumption (B3s). Hence we get the required efficiency of Z_n and $T_n(\bar{\psi})$ in both the set-ups under assumptions (C1)(a), (D2) and (D3). (Note that in this case $Z_n^* = Z_n$ and $T_n^*(\bar{\psi}) = T_n(\bar{\psi})$.)

Let us note the following

Remark 6.1 If assumption (D2) holds and equation (4.1), with $\psi = \frac{p'}{p}$, has a unique solution $\hat{\theta}_n$ (say, the latter holds for Examples 9.2-9.5 of Lindsay (1980) which includes Example 1.1) , then part U(III)(B) of Lemma 4.1 (equivalently, that of Lemma 5.1) holds with $T_n(\psi)$ replaced by $\hat{\theta}_n$, in other words, $\hat{\theta}_n$ is UAN(I) with AV $V(\cdot,\cdot,\psi)$, guaranteeing assumption (C1)(a) with $U_n = \hat{\theta}_n$ (which, in turn, implies $T_n(\psi) = \hat{\theta}_n$).

We are now going to check assumptions (C1)(a), (D2) and (D3) for Example 1.1. In view of Remark 6.1, it is enough to check assumptions (D2) and (D3). We have verified assumption (D2) for the more general case of Euclidean S and exponential p, q provided assumptions (a), (b)* and (d)-(f) of Remark 4.4, with $\bar{\Theta} \times \Xi$ and $C_{2,0}(\bar{\Theta} \times \Xi)$ replaced by $\bar{\Theta}$ and $C_2(\bar{\Theta})$, respectively, hold for p and assumptions (a), (e) and (f) of this remark hold for q. A proof of assumption (D3) is given in Lindsay (1980, §8.1-§8.2).

Example 1.1 can also be handled in a slightly different way, vide Pfanzagl (1982). Pfanzagl assumes the existence of a partially sufficient statistic $t(x)$ of ξ. Instead of assumption (D3) he assumes the completeness of t with respect to the family $\{P_{\theta,\xi} : \xi \in \Xi\}$ for all θ.

Note that in this case s_θ is given by (6.3) and the functions of $N_{\theta,G}$ depends on x only through t. One can use the latter fact and partial sufficiency of t to conclude that

$$\frac{p'(\cdot,\theta)}{p(\cdot,\theta)} \in N_{\theta,G}^{\perp} \;\; \forall(\theta,G).$$

Therefore,

$$\bar{\psi} = \frac{p'}{p} + \bar{\psi}_t \tag{6.4}$$

where $\bar{\psi}_t$ denote the optimal kernel in the mixture model induced by the marginals $P_{\theta,G}^t$ of t.

Therefore, using Lemma 3.5 for the marginal model, one observes that, under assumptions (A2)-(A4),

$$E_{P_{\theta,G}^t}\{\bar{\psi}_t(\cdot,\theta,G)\} = 0 \text{ a.e.}[P_{\theta,G}^t] \;\; \forall(\theta,G,G').$$

Hence, by completeness of t,

$$\begin{aligned} \bar{\psi}_t(\cdot,\theta,G) &= 0 \text{ a.e.}[P_{\theta,G}^t] \;\; \forall(\theta,G) \\ i.e., \; \bar{\psi}_t(t(\cdot),\theta,G) &= 0 \text{ a.e.}[P_{\theta,G}] \;\; \forall(\theta,G) \end{aligned}$$

proving, in view of relation (6.4), that $\bar{\psi} = \frac{p'}{p}$.

Note that for any θ in Θ one can easily weaken the condition of completeness of $\{P_{\theta,\xi} : \xi \in \Xi\}$ by L_2-completeness of it, in other words, it is enough to assume that for any θ in Θ and function ϕ of t,

$$\phi \in L_2^{\circ}(P_{\theta,\xi}) \;\; \forall\xi \;\; \textit{only if} \;\; \phi = 0 \text{ a.e.}[P_{\theta,\xi}] \;\; \forall\xi$$

(see also Definition 5.12 of van der Vaart (1987, p. 107)).

If, in the above, one allows t to be a l-dimensional real-vector depending on θ, $i.e.$, $t = t(\cdot,\theta)$, essentially the same calculation imply that the optimal kernel is

$$\bar{\psi} = \frac{p'}{p} + \sum_{j=1}^{l}(\frac{\partial}{\partial u_j}q)\{(\frac{\partial}{\partial\theta}t) - E(\frac{\partial}{\partial\theta}t|t)\}$$

— a result due to van der Vaart (1987).

Our calculations are somewhat different from the above authors (*i.e.*, Pfanzagl and van der Vaart). Assumptions needed for applying Theorems 4.2, 4.3, 5.2 and 5.3 for Pfanzagl's case are (C1)(a) and (D2) whereas those for van der Vaart's case are (C1) and an obvious generalisation of (D2). In this connection, it may be pointed out that van der Vaart's method, based on a generalisation of Pfanzagl's model, is a powerful one yielding a solution for Examples 1.1 and 1.2 as well as Example 9.6 of Lindsay (1980, pp 656-657) and the symmetric location-scale model of Bickel and Klaassen (1986). However, his L_2-completeness condition does not apply to Example 9.4 of Lindsay (1980) mentioned earlier in this chapter. His estimate is different from ours and requires fewer regularity conditions.

Chapter 7
Mixture Models with Observations on G

So far, we had only a single set of observations, which is used for estimating both θ and G. The question now arises what happens if another (independent) set of observations giving information only on G is also available? Will the problem become simpler? Hasminskii and Ibragimov (1983, §3) has provided a positive answer to this question. In the following discussion we shall derive their results using the methods of Chapter 4 instead of the original method due to Hasminskii and Ibragimov (1983). The assumptions of direct observations on G allow a verification of the smoothness conditions of the optimal kernel. Less important, but still useful, is the fact that we also have a uniformly $\sqrt{n}$-consistent estimate of G so that the splitting technique can be avoided and the identifiability assumption becomes much weaker.

This problem is taken up mainly as a technically interesting case where the required smoothness of the optimal kernel can be demonstrated. However, it may also have some practical applications as indicated in the following scenario.

Example 7.1 Suppose there is a source sending a signal θ over time. The signal as it comes out of the source at time t is distorted to

$$Z_t = \theta + \epsilon_t$$

where ϵ_t is the noise associated with the source, ϵ_t's are, say, i.i.d. $\mathcal{N}(0, \sigma^2)$, σ^2 known. As Z_t passes through a channel there is a further distortion ξ_t leading to an observation of

$$W_t = \theta + \epsilon_t + \xi_t$$

ξ_t's are assumed to be i.i.d. with common distribution G. Clearly ξ_t's are not observable at a time a signal is being sent, since they are confounded with the signal. But the distribution G can be estimated by observing $W_t \equiv \xi_t$ when a signal of magnitude zero is being sent from another controlled source, *i.e.*, when $Z_t \equiv 0$. Suppose that two independent sets of observations are recorded, the first one $(X_1, X_2, \cdots, X_n, \cdots)$ being recorded at time instances when a signal is being sent from the original source and the other one $(Y_1, Y_2, \cdots, Y_n, \cdots)$ when sent from the alternative.

More generally, one assumes given a signal θ and an "uncontrollable" noise ξ, the channel produces response X according to the density function $f(\cdot, \theta, \xi)$. Along with observations on X, one has observations on the distribution of ξ. Let us now write down the model explicitly.

Model III : Let $(S_1, \mathcal{S}_1)$ be an arbitrary measurable space and S_2 be a compact metric space with $\mathcal{S}_2 = \mathcal{B}(S_2)$. In other words, $(S_1, \mathcal{S}_1)$ and S_2 play the roles of $(S, \mathcal{S})$ and Ξ, respectively, of Model II. Let Θ be as in Model II and $\mathcal{G}$ stand for the set of all probability measures in $(S_2, \mathcal{S}_2)$. In this model, observations $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ are i.i.d. random vectors in $(S_1, \mathcal{S}_1)^p \times (S_2, \mathcal{S}_2)^q (=(\boldsymbol{S}, \boldsymbol{\mathcal{S}}))$ with common distribution

$$P^{III}_{\theta_o, G_o} = \prod_{j=1}^{p} f(x_j, \theta_o, G_o) d\mu(x_j) \prod_{k=1}^{q} dG_o(y_k) \text{ for all } (\boldsymbol{x}, \boldsymbol{y}) \in S_1^p \times S_2^q$$

for some θ_o in Θ and G_o in $\mathcal{G}$. [As in Chapter 1, let us also assume that the probability measures are well-defined on $\bar{\Theta} \times \mathcal{G}$.]

Note that

(1) Definitions 3.1-3.5 have obvious Model III-analogues which is obtained by replacing X_i's by $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$'s and P_{θ_o, G_o} by $P^{III}_{\theta_o, G_o}$ in the relevant definition for Model II.

As in Chapter 3, we shall abbreviate the phrase 'in Model III' by '(III)'.

Notation : We shall denote the set of all Model III-kernels by $\boldsymbol{K}$.

We shall need the following definition.

Definition 7.1 A function $\psi : S_1 \times \bar{\Theta} \times \mathcal{G} \to \mathbb{R}$ $(S_2 \times \bar{\Theta} \times \mathcal{G} \to \mathbb{R})$ is called an S_1 (S_2) - *kernel* if $\psi(\cdot, \theta, G) \in L_2^o(f(\cdot, \theta, G))$ $(L_2^o(G))$ for all (θ, G) in $\bar{\Theta} \times \mathcal{G}$ and the set of all S_1 (S_2) - kernels is denoted by K_1 (K_2).

Given any S_1-kernel ψ_1 and S_2-kernel ψ_2, let us define a function Q_{ψ_1,ψ_2} from $S \times \bar{\Theta} \times \mathcal{G}$ to $I\!R$ by

$$Q_{\psi_1,\psi_2}((\boldsymbol{x},\boldsymbol{y}),\theta,G) := \sum_{j=1}^{p} \psi_1(x_j,\theta,G) + \sum_{k=1}^{q} \psi_2(y_k,\theta,G) \quad \forall((\boldsymbol{x},\boldsymbol{y}),\theta,G) \in S \times \bar{\Theta} \times \mathcal{G}. \tag{7.1}$$

Note that

(2) Relation (7.1) defines a bounded linear map from $K_1 \times K_2$ to $\boldsymbol{K}$.

As in relation (3.20), let us define the θ-score S_θ for Model III by

$$S_\theta((\boldsymbol{x},\boldsymbol{y}),\theta,G) := \sum_{j=1}^{p} \frac{f'(x_j,\theta,G)}{f(x_j,\theta,G)} = \sum_{j=1}^{p} s_\theta(x_j,\theta,G) \quad \forall((\boldsymbol{x},\boldsymbol{y}),\theta,G) \in S \times \bar{\Theta} \times \mathcal{G}. \tag{7.2}$$

(3) $S_\theta \in \boldsymbol{K}$ if and only if $s_\theta \in K_1$ so that S_θ is well-defined under assumption (A3). Also, for $p = 1$ and $q = 0$, $S_\theta \equiv s_\theta$.

Before proceeding further let us make the following convention and definition.

Convention : For any G in $\mathcal{G}$ and ϕ in $L_2^\circ(G)$, we shall denote the function $\int f(\cdot,\cdot,\xi)\phi(\xi)dG(\xi)$ by $f(\cdot,\cdot,\phi dG)$.

Definition 7.2 For any (θ,G) in $\bar{\Theta} \times \mathcal{G}$ and ϕ in $L_2^\circ(G)$, define $A_{\theta,G}(\phi)$ from S_1 to $I\!R$ by

$$A_{\theta,G}(\phi)(x) := \frac{f(x,\theta,\phi dG)}{f(x,\theta,G)} \quad \forall x \in S_1$$

and for any ϕ in K_2, define $A(\phi)$ from $S_1 \times \bar{\Theta} \times \mathcal{G}$ to $I\!R$ by

$$A(\phi)(x,\theta,G) := A_{\theta,G}(\phi(\cdot,\theta,G))(x) \quad \forall (x,\theta,G) \in S_1 \times \bar{\Theta} \times \mathcal{G}.$$

Then

(4) For any (θ,G) in $\bar{\Theta} \times \mathcal{G}$, $A_{\theta,G}$ is a bounded linear map from $L_2^\circ(G)$ to $L_2^\circ(f(\cdot,\theta,G))$ and A is a bounded linear map from K_2 to K_1.

For any (θ, G) in $\bar{\Theta} \times \mathcal{G}$, define

$$N_{\theta,G} := \{\Psi \in L_2^\circ(P_{\theta,G}^{III}) : \exists \phi \in L_2^\circ(G) \text{ with } \Psi(x,y) := \sum_{j=1}^{p} A_{\theta,G}(\phi)(x_j) + \sum_{k=1}^{q} \phi(y_k) \ \forall (x,y)\}. \quad (7.3)$$

(5) The elements of $N_{\theta,G}$ can be thought of as 'directional scores' with respect to small variations in G. However, for the special case where $p = 1$ and $q = 0$, $N_{\theta,G}$ is a proper subset of $N_{\theta,G}$ unless G has a finite support so that relation (7.3) falls short of an analogue of relation (3.21).

As in Chapter 3, one can define an *optimal kernel* $\bar{\Psi}$ and the information I^{III} by the following analogue of relation (3.22).

$$\left.\begin{array}{rcl} \bar{\Psi}(\cdot,\theta,G) & := & \mathrm{Proj}_{N_{\theta,G}^{\perp}}\{S_\theta(\cdot,\theta,G)\} \\ I^{III}(\theta,G) & := & \|\bar{\Psi}(\cdot,\theta,G)\|^2_{L_2(P_{\theta,G}^{III})} \end{array}\right\} \ \forall(\theta,G). \quad (7.4)$$

In order to get a simpler formula for evaluating $\bar{\Psi}$, we shall need the following definitions and lemma.

Definition 7.3 For any (θ, G) in $\bar{\Theta} \times \mathcal{G}$, the closed linear space in $L_2^\circ(P_{\theta,G}^{III})$ spanned by $S_\theta(\cdot,\theta,G)$ and $N_{\theta,G}$ will be called *the tangent space at* (θ, G) and will henceforth be denoted by $T_{\theta,G}^{III}$.

Remark 7.1 Our tangent space coincides with that considered in Hasminskii and Ibragimov (1983, §3).

Definition 7.4 Call a kernel Ψ a *tangent-space-kernel* if $\Psi(\cdot,\theta,G) \in T_{\theta,G}^{III}$ for all $(\theta,G) \in \bar{\Theta} \times \mathcal{G}$.

Observe that

(6) $T_{\theta,G}^{III}$ consists of functions of the form $\sum_{j=1}^{p} \phi_1(x_j) + \sum_{k=1}^{q} \phi_2(y_k)$ for some $\phi_1 \in L_2^\circ(f(\cdot,\theta,G))$ and $\phi_2 \in L_2^\circ(G)$ so that any tangent-space-kernel Ψ must be of the form Q_{ψ_1,ψ_2} for some $\psi_1 \in K_1$, $\psi_2 \in K_2$.

(7) A tangent-space-kernel Q_{ψ_1,ψ_2} is an optimal kernel if and only if

$$p\left(\int \psi_1(\cdot,\theta,G) A_{\theta,G}(\phi)(\cdot) f(\cdot,\theta,G) d\mu(\cdot)\right) + q\left(\int \psi_2(\cdot,\theta,G)\phi(\cdot) dG(\cdot)\right) = 0 \quad (7.5)$$

for all (θ, G, ϕ) with $\phi \in L_2^o(G)$.

In Lemma 7.1(a) (vide relation (7.6)), we give a simpler sufficient condition for a tangent-space-kernel to be optimal. Lemma 7.1(b) gives a sort of converse which is a Model III-analogue of Lemma 3.5. Then in Lemmas 7.2-7.4 we show a smooth solution of (7.6) exists.

Lemma 7.1 *(a) If for some tangent-space-kernel* Q_{ψ_1,ψ_2},

$$p\int \psi_1(\cdot,\theta,G)f(\cdot,\theta,G')d\mu(\cdot) + q\int \psi_2(\cdot,\theta,G)dG'(\cdot) = 0 \quad \forall(\theta,G,G') \in \bar{\Theta}\times\mathcal{G}\times\mathcal{G} \tag{7.6}$$

then Q_{ψ_1,ψ_2} *is a version of the optimal kernel.*

(b) Conversely, if Q_{ψ_1,ψ_2} *is a version of the optimal kernel defined through the relations (7.2)-(7.4) and for all* θ *the L.H.S. of (7.6) is continuous in* (G, G')*, then (7.6) holds for it.*

Proof : (a) In view of observations (4), (7) and the fact that the set of all bounded functions ϕ lying in $L_2^o(G)$ is dense in it, it is enough to show (7.5) for bounded ϕ's only.

Now, let ϕ be a bounded function in $L_2^o(G)$, then there is $\epsilon > 0$ such that $1+\eta\phi(y) \geq 0$ for all y in S_2 whenever $|\eta| < \epsilon$. Fix one such ϵ. Define the curve $G_{.} : (-\epsilon,\epsilon) \to \mathcal{G}$ by

$$dG_\eta(\cdot) = \{1+\eta\phi(\cdot)\}dG(\cdot) \;\; \forall\eta \in (-\epsilon,\epsilon).$$

The relation (7.6) with $G' = G_\eta$ implies relation (7.5) for ϕ. Since ϕ was arbitrary, this proves (a).

(b) An easy modification of the proof of Lemma 3.5(b). □

Let ψ_i be an S_i kernel, $i = 1,2$. Consider the following Model III-analogue of equation (4.1).

$$\sum_{\substack{i=1\\ i \text{ odd}}}^{n} Q_{\psi_1,\psi_2}((X_i,Y_i),\theta,\hat{\hat{G}}_n^E) + \sum_{\substack{i=1\\ i \text{ even}}}^{n} Q_{\psi_1,\psi_2}((X_i,Y_i),\theta,\hat{\hat{G}}_n^O) = 0 \tag{7.7}$$

where $\hat{\hat{G}}_n := I\!F_{nq}(Y_{11},Y_{12},\cdots,Y_{1q},Y_{21},Y_{22},\cdots,Y_{2q},\cdots,Y_{n1},Y_{n2},\cdots,Y_{nq})$ and suffixes E and O stand for the operations based on even and odd indices, respectively, as defined in (11) of Chapter 3.

Note that

(8) By Proposition 5.1, $\hat{\hat{G}}_n$ is a uniformly consistent (III) estimate of G_o (vide Definition 3.1 and observation (1)).

Assume that

(E1) There is a uniformly $\sqrt{n}$-consistent (III) estimate U_n of θ_o (vide Definition 3.2 and observation (1)).

Let $T_n(Q_{\psi_1,\psi_2})$ be the estimate defined through Definition 4.1 and observation (1). We are now going to give regularity conditions on f, ψ_1 and ψ_2 guaranteeing uniform asymptotic normality (III) (vide Definition 3.4 and observation (1)) of $T_n(Q_{\psi_1,\psi_2})$.

Fix (θ_o, G_o) in $\Theta \times \mathcal{G}$. Let δ_o denote a positive real-number which will be chosen later. The following are the required regularity conditions.

U[i] Condition U(i) of Chapter 4 holds.

For any condition U(C) among U(ii), U(iv)-U(vi) of Chapter 4, U[C] denotes the condition that U(C) holds with ψ replaced by ψ_1 or the relevant parts of it hold with (ψ, f, μ) replaced by (ψ_2, u, G) where u denotes the function identically equal to one. The condition U[iii] is given below.

U[iii] For any $c > 0$ and $\epsilon > 0$, the supremum, over $\theta \in B(\theta_o, \delta_o)$, $G \in B(G_o, \delta_o)$ and $|\theta' - \theta| \leq c/\sqrt{n}$, of

$$(P_{\theta,G}^{III})^n(\{|\sqrt{n}p \int \psi_1(\cdot, \theta, \hat{\hat{G}}_n) f(\cdot, \theta, G) d\mu(\cdot) + \sqrt{n}q \int \psi_2(\cdot, \theta, \hat{\hat{G}}_n) dG(\cdot)| > \epsilon\})$$

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

The following is the Model III-analogue of Lemma 4.1a.

Lemma 7.2 *Assume (E1). If f satisfies conditon U[i] and (ψ_1, ψ_2) satisfies conditions U[ii]-U[vi], then $T_n(Q_{\psi_1,\psi_2})$ is a uniformly $\sqrt{n}$-consistent solution (III) (vide Definition 3.3 and observation (1)) of (7.7) as well as UAN (III) with AV $V(\cdot, \cdot, \psi_1, \psi_2)$ where*

$$V(\theta, G, \psi_1, \psi_2) := \frac{p\|\psi_1(\cdot, \theta, G)\|^2_{L_2(P_{\theta,G})} + q\|\psi_2(\cdot, \theta, G)\|^2_{L_2(G)}}{p\langle \psi_1(\cdot, \theta, G), s_\theta(\cdot, \theta, G)\rangle^2_{L_2(P_{\theta,G})}} \quad \forall(\theta, G).$$

An easy modification of the proof of the parts U(II)(B) and U(III)(B) of Lemma 4.1 proves the result.

From now on, we shall assume that q is positive.

Note that

(9) If there is ϕ in K_2 such that $Q_{s_\theta + A(\phi),\phi}$ satisfies relation (7.6), then

$$I^{III}(\theta, G) = p\|s_\theta(\cdot,\theta,G) + A(\phi)(\cdot,\theta,G)\|^2_{L_2(f(\cdot,\theta,G))} + q\|\phi(\cdot,\theta,G)\|^2_{L_2(G)} \quad \forall(\theta, G).$$

Hence $I^{III}(\theta, G) > 0$ if and only if

(E0) $0 < \|s_\theta(\cdot,\theta,G)\|^2_{L_2(f(\cdot,\theta,G))} < \infty \ \forall(\theta, G)$.

(10) Assumption (D1) implies assumption (E0) and condition U[i].

Therefore, in view of observation (9), conditions U[ii]-U[vi] hold for $\psi_1 = s_\theta + A(\phi)$ and $\psi_2 = \phi$ provided there is ϕ belonging to $C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G})$ such that Q_{ψ_1,ψ_2} satisfies relation (7.6).

Remark 7.2 In view of observations (9)-(10), it remains to prove the existence of ϕ lying in $C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G})$ such that $Q_{s_\theta + A(\phi),\phi}$ satisfies relation (7.6).

Let us now observe that, for any ψ_1 in K_1 and ψ_2 in K_2, relation (7.6) is equivalent to

$$p\int \psi_1(\cdot,\theta,G) f(\cdot,\theta,y)d\mu(\cdot) + q\psi_2(y,\theta,G) = 0 \ \ \forall(\theta,G,y) \in \bar{\Theta} \times \mathcal{G} \times S_2.$$

Therefore, $Q_{s_\theta + A(\phi),\phi}$ satisfies relation (7.6) if and only if

$$\begin{aligned} p\int s_\theta(\cdot,\theta,G) f(\cdot,\theta,y)d\mu(\cdot) &= -p\int A(\phi)(\cdot,\theta,G) f(\cdot,\theta,y)d\mu(\cdot) \\ &\quad -q\phi(y,\theta,G) \qquad \forall(\theta,G,y). \end{aligned} \tag{7.8}$$

But, for any $(\theta, G, y) \in \bar{\Theta} \times \mathcal{G} \times S_2$,

$$\begin{aligned} &\int A(\phi)(\cdot,\theta,G) f(\cdot,\theta,y)d\mu(\cdot) \\ =\ &\int \frac{\int \phi(y',\theta,G) f(x,\theta,y')dG(y')}{f(x,\theta,G)} f(x,\theta,y)d\mu(x) \\ =\ &\int K(y,y',\theta,G)\phi(y',\theta,G)dG(y') \end{aligned} \tag{7.9}$$

where the function $K : S_2 \times S_2 \times \bar{\Theta} \times \mathcal{G} \to I\!R$ is defined by

$$K(y,y',\theta,G) := \int \frac{f(\cdot,\theta,y)f(\cdot,\theta,y')}{f(\cdot,\theta,G)}\, d\mu(\cdot)$$
$$\forall (y,y',\theta,G) \in S_2 \times S_2 \times \bar{\Theta} \times \mathcal{G}.$$

Assume that

(E2) (a) For any x in S_1, $f(x,\cdot,\cdot) \in C_{1,0}(\bar{\Theta} \times S_2)$
and (b) the following three families of functions

$$\{\frac{f(\cdot,\theta,y)f(\cdot,\theta,y')}{f(\cdot,\theta,G)} : (y.y',\theta,G) \in S_2 \times S_2 \times \bar{\Theta} \times \mathcal{G}\}$$
$$\{\frac{f(\cdot,\theta,y)f'(\cdot,\theta,y')}{f(\cdot,\theta,G)} : (y,y',\theta,G) \in S_2 \times S_2 \times \bar{\Theta} \times \mathcal{G}\}$$
and $$\{\frac{f(\cdot,\theta,y)f(\cdot,\theta,y')f'(\cdot,\theta,G)}{f(\cdot,\theta,G)} :$$
$$(y,y',\theta,G) \in S_2 \times S_2 \times \bar{\Theta} \times \mathcal{G}\}$$

are uniformly integrable with respect to μ.

(11) Under assumption (E2), $K \in C_{0,0,1,0}(S_2 \times S_2 \times \bar{\Theta} \times \mathcal{G})$ and therefore the R.H.S. of relation (7.9) defines a bounded operator B from CK_2 to CK_2 where CK_2 denotes the subspace $C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G}) \cap K_2$, of $C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G})$.

Define $\eta : S_2 \times \bar{\Theta} \times \mathcal{G} \to I\!R$ by

$$\eta(y,\theta,G) := \int s_\theta(x,\theta,G)f(x,\theta,y)d\mu(x)\ \ \forall (y,\theta,G). \tag{7.10}$$

(12) Under assumption (D1), η belongs to the set CK_2.
By (7.8)-(7.10) we see that for all (y,θ,G)

$$\begin{aligned}\eta(y,\theta,G) &= -(B + \frac{q}{p}I)(\phi)(y,\theta,G)\\ &= -C(\phi)(y,\theta,G)\end{aligned} \tag{7.11}$$

where C denotes the (bounded) operator $(B + \frac{q}{p}I)$ from CK_2 to itself.

In view of observations (9)-(12), Remark 7.2 and relations (7.9)-(7.11), it remains to prove that C is invertible. As a first step towards this goal let us show that

Lemma 7.3 *Under assumption (E2), (a) C is 1-1 and (b) B is a compact operator.*

Proof : (a) Let $\phi \in CK_2$ be such that $C\phi = 0$.

Then,

$$\iint \phi(y,\theta,G)K(y,y',\theta,G)\phi(y',\theta,G)dG(y)dG(y') + \frac{q}{p}\|\phi(\cdot,\theta,G)\|^2_{L_2(G)} = 0 \quad \forall(\theta,G)$$

which, in turn, implies

$$\int \frac{\{\int \phi(y,\theta,G)f(x,\theta,y)dG(y)\}^2}{f(x,\theta,G)} d\mu(x) + \frac{q}{p}\|\phi(\cdot,\theta,G)\|^2_{L_2(G)} = 0 \ \ \forall(\theta,G)$$

equivalently,

$$\phi(\cdot,\theta,G) = 0 \text{ a.e.}[G] \ \ \forall(\theta,G)$$

and hence

$$\phi \equiv 0 \ [\text{Since } \phi \in CK_2].$$

(b) Want to show that $\overline{AS}$ is compact in $\| \ \|_{0,1,0}$-norm where

$$\begin{aligned} AS &:= \{A(\phi) : \phi \in S(C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G}))\} \\ \text{and } S(C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G})) &:= \{\phi \in C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G}) : \|\phi\|_{0,1,0} = 1\}. \end{aligned}$$

Note that, $\|\phi\|_{0,1,0} = \|\phi\|_{\sup} + \|\frac{\partial}{\partial\theta}\phi\|_{\sup}$ for $\phi \in C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G})$.

Also, note that $\overline{AS}$ is compact if and only if AS and $(AS)'$ are uniformly bounded and equicontinuous where

$$(AS)' := \{\frac{\partial}{\partial\theta}(A(\phi)) : \phi \in S(C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G}))\}.$$

We shall only show that AS is uniformly bounded and equicontinuous. One can prove this fact for $(AS)'$ in a similar way.

Now, let us observe that

$$\|A\phi\|_{\sup} \leq \|K\|_{\sup}\|\phi\|_{\sup} \leq \|K\|_{\sup}\|\phi\|_{0,1,0} \ \ \forall\phi \in C_{0,1,0}(S_2 \times \bar{\Theta} \times \mathcal{G}).$$

Therefore, AS is uniformly bounded. (7.12)

Let us now fix $(y_o, \theta_o, G_o) \in S_2 \times \bar{\Theta} \times \mathcal{G}$, then

$$\begin{aligned}
& |A(\phi)(y,\theta,G) - A(\phi)(y_o,\theta_o,G_o)| \\
= & \ |\int K(y,y',\theta,G)\phi(y',\theta,G)dG(y') \\
& - \int K(y_o,y',\theta_o,G_o)\phi(y',\theta_o,G_o)dG_o(y')| \\
\leq & \ |\int \{K(y,y',\theta,G) - K(y_o,y',\theta_o,G_o)\}\phi(y',\theta,G)dG(y')| \\
& + |\int K(y_o,y',\theta_o,G_o)\{\phi(y',\theta,G) - \phi(y',\theta_o,G_o)\}dG(y')| \\
& + |\int K(y_o,y',\theta_o,G_o)\phi(y',\theta_o,G_o)d(G-G_o)(y')| \\
\leq & \ \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \text{ if } |\theta - \theta_o|, \ d(G,G_o) \text{ and } \rho(y,y_o) < \delta \qquad (7.13)
\end{aligned}$$

where ρ is some metric inducing topology on S_2 and $0 < \delta < \frac{\epsilon}{3\|K\|_{sup}}$ is chosen in such a way that for any pair (y,y',θ,G), $(\bar{y},\bar{y}',\bar{\theta},\bar{G})$; $|\theta - \bar{\theta}|$, $d(G,\bar{G})$, $\rho(y,\bar{y})$ and $\rho(y',\bar{y}') < \delta$ imply

$$|\phi(y,\theta,G) - \phi(\bar{y},\bar{\theta},\bar{G})| < \frac{\epsilon}{3\|K\|_{\sup}} \text{ and } |K(y,y',\theta,G) - K(\bar{y},\bar{y}',\bar{\theta},\bar{G})| < \frac{\epsilon}{3}.$$

From (7.12) and (7.13) it follows that AS is uniformly bounded and equicontinuous. □

As a corollary to Lemma 7.3, one can show that

Lemma 7.4 *Under assumptions (D1) and (E2) there is a function $\tilde{\phi} \in CK_2$ such that $\psi_1 = s_\theta + A(\tilde{\phi})$ and $\psi_2 = \tilde{\phi}$ together satisfy relation (7.6) and conditions U[ii -U[vi.*

Proof : In view of calculations done earlier it is enough to show that C is invertible.

Suppose not. Then using the fact that $q \neq 0$ and compactness of B conclude from Theorem 4.25(b) of Rudin (1974) that C is not 1-1, which contradicts Lemma 7.3(a). □

In view of Lemmas 7.1-7.4 one can show

Theorem 7.5 *Assume (E1), (D1) and (E2). Let $\tilde{\phi}$ be as consi ed in Lemma 7.4. Then $T_n(Q_{s_\theta + A(\tilde{\phi}),\tilde{\phi}})$ is UAN (III) with AV $(1/I^{III})$.*

Remark 7.3 It is worth pointing out that we do need the compactness of the operator B since it acts on a Banach space rather than a Hilbert space. Note also that we make use of 'non-negative'-ness of B (vide Lemma 7.3) but the associated inner product will not give the norm of the Banach space.

Remark 7.4 We have verified assumptions (D1) and (E2) for the following two cases

Case I (a) S is compact,

(b) $f \in C_{0,2,0}(S \times \bar{\Theta} \times \mathcal{G})$

(c) $f(x,\theta,G) > 0 \ \forall(x,\theta,G)$

and (d) $P_{\theta,G}(\{|f'(\cdot,\theta,G)| > 0\}) > 0 \ \forall(\theta,G)$.

Case II We have Euclidean S and exponential f following assumptions (a) and (f) of Remark 4.4 and additional assumptions

$$(d)^* \ \sum_{j=1}^{k}(\pi_j')^2(\theta,\xi) > 0 \ \forall(\theta,\xi)$$

and (e)* for any θ in Θ, both $2\underline{\pi_j}(\{\theta\}\times\Xi) - \overline{\pi_j}(\{\theta\}\times\Xi)$ and $2\overline{\pi_j}(\{\theta\}\times\Xi) - \underline{\pi_j}(\{\theta\}\times\Xi)$ belong to the interior of Ω.

In particular, Example 7.1 falls in Case II of the above remark.

Remark 7.5 To solve (7.7), we have to determine for various values of θ, $\tilde{\phi}(Y,\theta,\hat{\hat{G}}_n^O)$ and $\tilde{\phi}(Y,\theta,\hat{\hat{G}}_n^E)$ evaluated at all the observations Y_{ij}'s on Y.

Now, one can easily observe that for G's with finite support and y's restricted to the support of G, equation (7.11) can be rewritten in the from $Ax = y$ where the matrix A is positive-definite. Thus we have a unique solution for $\tilde{\phi}(Y_{ij},\theta,\hat{\hat{G}}_n^O)$ for odd i's and for $\tilde{\phi}(Y_{ij},\theta,\hat{\hat{G}}_n^E)$ for even i's.

To evaluate $\{\tilde{\phi}(Y_{ij},\theta,\hat{\hat{G}}_n^O), 1 \le i \le q, i \text{ even}\}$ and $\{\tilde{\phi}(Y_{ij},\theta,\hat{\hat{G}}_n^E), 1 \le i \le q, i \text{ odd}\}$, define $\hat{G}_{\epsilon 1} := (1-\epsilon)\hat{\hat{G}}_n^O + \epsilon\hat{\hat{G}}_n^E$ and $\hat{G}_{\epsilon 2} := \epsilon\hat{\hat{G}}_n^O + (1-\epsilon)\hat{\hat{G}}_n^E$ and solve the linear version of (7.11) with $G = \hat{G}_{\epsilon 1}$ and $\hat{G}_{\epsilon 2}$ for

$\{\tilde{\phi}(Y_{ij},\theta,\widehat{G}_{\epsilon 1}), 1 \le i \le q, i \text{ even }\}$ and $\{\tilde{\phi}(Y_{ij},\theta,\widehat{G}_{\epsilon 2}), 1 \le i \le q, i \text{ odd }\}$. Again the solutions are unique. Now let ϵ tend to zero.

In actual practice one would solve

$$\sum_{\substack{i=1\\ i \text{ odd}}}^{n} [\sum_{j=1}^{p}\{s_\theta(X_{ij},\theta,\hat{\widehat{G}}_n^E) + A(\tilde{\phi})(X_{ij},\theta,\hat{\widehat{G}}_n^E)\} + \sum_{k=1}^{q}\tilde{\phi}(Y_{ik},\theta,\widehat{G}_{\epsilon 2})]$$

$$+ \sum_{\substack{i=1\\ i \text{ even}}}^{n} [\sum_{j=1}^{p}\{s_\theta(X_{ij},\theta,\hat{\widehat{G}}_n^O) + A(\tilde{\phi})(X_{ij},\theta,\hat{\widehat{G}}_n^O)\} + \sum_{k=1}^{q}\tilde{\phi}(Y_{ik},\theta,\widehat{G}_{\epsilon 1})] = 0$$

for various values of ϵ and stop when two consecutive values differ insignificantly.

Chapter 8
The Results of A Simulation Study

In this chapter, we shall compare first theoretically and then numerically both the small and large sample behaviour of our estimate with two more estimates — the grand mean $\bar{\bar{X}}$ and the (fixed set-up) m.l.e. in Example 1.2 or the symmetric location-scale problem. The first estimate or the grand mean is the most obvious estimate here and the second one is suggested by Neyman and Scott (1948).

From relations (2.1)-(2.2), one can easily conclude that the second estimate is a solution of

$$\sum_{i=1}^{n} \frac{(\bar{X}_i - \mu)}{S_i^2 + k(\bar{X}_i - \mu)^2} = 0$$

where $\bar{X}_i = \frac{1}{k}\sum_{j=1}^{k} X_{ij}$ and $S_i^2 = \sum_{j=1}^{k} X_{ij}^2 - k\bar{X}_i^2$.

Convention : Throughout this chapter, we shall use τ_i to denote the quantity $(\sigma_i^2)^{-1}$, H to denote the common distribution of τ_i's and refer to the second estimate as the 'm.l.e.'. We shall also use the term 'asymptotic variance' of an estimate T_n to mean the variance of the asymptotic distribution of $\sqrt{n}(T_n - \mu)$ (*cf.* Definitions 1.1 and 3.4).

Note that

(1) The m.l.e. is asymptotically normal *only if* $k \geq 3$ and in that case it is a C_1-estimate (as $\mathcal{G}$ is compact).

The first theoretical result shows that, as expected, for H close to the degenerate distribution the grand mean behaves like the efficient estimate, whereas, as Neyman and Scott (1948) predicted, for large k, the m.l.e. behaves similarly.

Proposition 8.1 *(a) For any $k \geq 1$, the ratio of the asymptotic variances of $\bar{\bar{X}}$ and $T_n(\bar{\psi})$ tends to one as $E(\tau)E(\frac{1}{\tau})$ approaches one,* i.e., *τ approaches the degenerate random variable.*

(b) For any choice of H, the ratio of the asymptotic varances of the m.l.e. and $T_n(\bar{\psi})$ approaches one as k becomes large.

Proof : Let us observe that,

$$
\begin{aligned}
(\text{AV } (T_n(\bar{\psi}))^{-1} &= E[S(\boldsymbol{X}_1,\mu)\{E(\tau|\boldsymbol{X}_1)\}^2] \\
&\quad \text{where } S(\boldsymbol{X}_1,\mu) := \sum_{j=1}^{k}(X_{ij}-\mu)^2 \\
&\leq E[S(\boldsymbol{X}_1,\mu)E(\tau^2|\boldsymbol{X}_1)] = E[S(\boldsymbol{X}_1,\mu)\tau^2] \\
&= E[E\{S(\boldsymbol{X}_1,\mu)|\tau\}\tau^2] = E[\frac{k}{\tau}\cdot\tau^2] = kE(\tau). \quad (8.1)\\
(\text{AV m.l.e.})^{-1} &= (k-2)E(\tau). \quad (8.2)\\
(\text{AV } \bar{\bar{X}})^{-1} &= \frac{k}{E(\frac{1}{\tau})}. \quad (8.3)
\end{aligned}
$$

From (8.2) and (8.3),

$$\frac{\text{AV } \bar{\bar{X}}}{\text{AV } T_n(\bar{\psi})} \leq \frac{kE(\tau)}{\frac{k}{E(\frac{1}{\tau})}} = kE(\tau)\frac{E(\frac{1}{\tau})}{k} = E(\tau)E(\frac{1}{\tau}) \quad (8.4)$$

and from (8.1) and (8.2),

$$\frac{\text{AV m.l.e.}}{\text{AV } T_n(\bar{\psi})} \leq \frac{kE(\tau)}{(k-2)E(\tau)} = \frac{k}{k-2}. \quad (8.5)$$

Again,

$$\frac{\text{AV } \bar{\bar{X}}}{\text{AV } T_n(\bar{\psi})} \geq \frac{\text{AV } \bar{\bar{X}}}{\text{AV m.l.e.}} = \frac{(k-2)E(\tau)}{\frac{k}{E(\frac{1}{\tau})}} = \frac{(k-2)}{k}E(\tau)E(\frac{1}{\tau}) \quad (8.6)$$

$$\text{and } \frac{\text{AV m.l.e.}}{\text{AV } T_n(\bar{\psi})} \geq \frac{\text{AV m.l.e.}}{\text{AV } \bar{\bar{X}}} = \frac{k}{(k-2)}\frac{1}{E(\tau)E(\frac{1}{\tau})}. \quad (8.7)$$

From (8.4) and (8.6),

$$\{\frac{(k-2)}{k}E(\tau)E(\frac{1}{\tau})\} \vee 1 \leq \frac{\text{AV } \bar{\bar{X}}}{\text{AV } T_n(\bar{\psi})} \leq E(\tau)E(\frac{1}{\tau}) \quad (8.8)$$

and from (8.5) and (8.7),

$$\left\{\frac{k}{(k-2)}\frac{1}{E(\tau)E(\frac{1}{\tau})}\right\}\bigvee 1 \le \frac{\text{AV m.l.e.}}{\text{AV } T_n(\bar{\psi})} \le \frac{k}{k-2}. \tag{8.9}$$

From (8.8), $\frac{\text{AV }\bar{\bar{X}}}{\text{AV } T_n(\bar{\psi})} \to 1$ as $E(\tau)E(\frac{1}{\tau}) \to 1$ proving part (a) of the proposition and from (8.9), $\frac{\text{AV m.l.e.}}{\text{AV } T_n(\bar{\psi})} \to 1$ as $k \to \infty$ proving part (b) of it.

□

The following is the second and the remaining theoretical result.

Proposition 8.2 *Fix any $k \ge 3$. Let τ_i follow the gamma distribution $\Gamma(\alpha, \lambda)$ with $\alpha > 1$. Then as α approaches one, the ratio of the asymptotic variances of $\bar{\bar{X}}$ and $T_n(\bar{\psi})$ tends to infinity whereas that of the m.l.e. and $T_n(\bar{\psi})$ tends to $\frac{k^2+2k}{k^2+2k-8}$ which is a number strictly greater than 1.*

However, for large values of α, the grand mean $\bar{\bar{X}}$ behaves like an efficient estimate whereas the asymptotic variance of the m.l.e. is approximately $\frac{k}{k-2}$ times that of our estimate.

Proof : By an easy algebra, one can show that

$$(\text{AV }\bar{\bar{X}})^{-1} = \frac{k}{E(\frac{1}{\tau})} = \frac{k}{\frac{\lambda}{(\alpha-1)}} = \frac{k(\alpha-1)}{\lambda}.$$

Again,

$$\begin{aligned}(\text{AV m.l.e.})^{-1} &= (k-2)E(\tau) = (k-2)\frac{\alpha}{\lambda}\\ \text{and } (\text{AV } T_n(\bar{\psi}))^{-1} &= \frac{k(\alpha+\frac{k}{2})}{(\alpha+\frac{k}{2}+1)}\cdot\frac{\alpha}{\lambda}.\end{aligned}$$

Therefore,

$$\frac{\text{AV }\bar{\bar{X}}}{\text{AV } T_n(\bar{\psi})} = \frac{k(\alpha+\frac{k}{2})}{(\alpha+\frac{k}{2}+1)}\cdot\frac{\alpha}{\lambda}\cdot\frac{\lambda}{k(\alpha-1)} = \frac{(\alpha+\frac{k}{2})\alpha}{(\alpha+\frac{k}{2}+1)(\alpha-1)} \tag{8.10}$$

$$\text{and } \frac{\text{AV m.l.e.}}{\text{AV } T_n(\bar{\psi})} = \frac{k(\alpha+\frac{k}{2})}{(\alpha+\frac{k}{2}+1)}\cdot\frac{\alpha}{\lambda}\cdot\frac{1}{(k-2)}\cdot\frac{\lambda}{\alpha} = \frac{k(\alpha+\frac{k}{2})}{(\alpha+\frac{k}{2}+1)(k-2)}. \tag{8.11}$$

The result follows from relations (8.10) and (8.11). □

Before proceeding further, we shall obtain a more accurate relation between the asymptotic variances of the three estimates. First observe that

$$\{E(\tau^2|\boldsymbol{X}) - E^2(\tau|\boldsymbol{X})\}\{\sum_{j=1}^{k}(X_j-\mu)^2\} \le E[\{\tau - E(\tau)\}^2|\boldsymbol{X}]\{\sum_{j=1}^{k}(X_j-\mu)^2\}$$

by the theory of least squares.

Therefore, by taking the expectation over the distribution of $\boldsymbol{X}$,

$$\begin{aligned}
& E[\{E(\tau^2|\boldsymbol{X}) - E^2(\tau|\boldsymbol{X})\}\{\sum_{j=1}^{k}(X_j-\mu)^2\}] \\
\le\ & E[E\{(\tau - E(\tau))^2|\boldsymbol{X}\}\{\sum_{j=1}^{k}(X_j-\mu)^2\}] \\
=\ & E[\{\tau - E(\tau)\}^2\{\sum_{j=1}^{k}(X_j-\mu)^2\}] \\
=\ & E[\{\tau - E(\tau)\}^2\frac{k}{\tau}] \\
=\ & kE[\tau^2\frac{1}{\tau} - 2\tau E(\tau)\frac{1}{\tau} + \{E(\tau)\}^2\frac{1}{\tau}] \\
=\ & k[E(\tau) - 2E(\tau) + \{E(\tau)\}^2\{E(\frac{1}{\tau})\}] \\
=\ & kE(\tau)\{E(\tau)E(\frac{1}{\tau}) - 1\}.
\end{aligned}$$

Therefore

$$\begin{aligned}
(\text{AV } T_n(\bar{\psi}))^{-1} &= E[\{E(\tau|\boldsymbol{X})\}^2\{\sum_{j=1}^{k}(X_j-\mu)^2\}] \\
&\ge E[E(\tau^2|\boldsymbol{X})\{\sum_{j=1}^{k}(X_j-\mu)^2\}] - kE(\tau)\{E(\tau)E(\frac{1}{\tau}) - 1\} \\
&= E[\tau^2\{\sum_{j=1}^{k}(X_j-\mu)^2\}] - kE(\tau)\{E(\tau)E(\frac{1}{\tau}) - 1\} \\
&= E(\tau^2\frac{k}{\tau}) - kE(\tau)\{E(\tau)E(\frac{1}{\tau}) - 1\} \\
&= kE(\tau)\{\quad - E(\tau)E(\frac{1}{\tau})\} \\
&= \frac{k}{(k-2)}\{2 - E(\tau)E(\frac{1}{\tau})\}(\text{AV m.l.e.})^{-1}
\end{aligned}$$

and hence

$$(\text{AV } T_n(\bar{\psi})) \leq [\frac{k}{(k-2)}\{2 - E(\tau)E(\frac{1}{\tau})\}]^{-1}(\text{AV m.l.e.})$$

provided $E(\tau)E(\frac{1}{\tau}) < 2$.

Also, in view of relation (8.9),

$$\text{AV } T_n(\bar{\psi}) \geq \frac{(k-2)}{k}(\text{AV m.l.e.}).$$

Therefore,

$$\frac{(k-2)}{k}(\text{AV m.l.e.}) \leq \text{AV } T_n(\bar{\psi}) \leq \frac{(k-2)}{k}\frac{1}{\{2 - E(\tau)E(\frac{1}{\tau})\}}(\text{AV m.l.e.}). \tag{8.12}$$

Suppose that $k = 3$, the smallest k for which the m.l.e. is UAN. Suppose that τ_i's follow the uniform distribution over $(\epsilon, \frac{1}{\epsilon})$ for $\epsilon > 0$ (*i.e.*, we have a sort of noninformative prior on σ_i^2's). A difficulty arises due to the fact that the integrals involved are mathematically intractible. So we resorted to numerical techniques for $\epsilon = 0.03125(0.03125)0.96875$. The results of this effort is given in Table 1.

For small values of ϵ, the asymptotic variance of the grand mean $\bar{\bar{X}}$ is twice the asymptotic variance of the efficient estimate $T_n(\bar{\psi})$ whereas for ϵ's close to 1 they are nearly the same (*cf.* Proposition 8.1) and the asymptotic variance of the m.l.e. is roughly three times the mixture lower-bound. We have also evaluated the quantities given in relation (8.12) which become close for those values of ϵ which are close to 1.

To study the small sample behaviour of these estimates we have also made a series of simulations. For each of these simulation 1000 samples each of size $n = 100$ are generated from the population considered in Example 1.2 for different choices of H and common value of $k = 3$ and $\mu = 1.0$. The only difficult part was the estimation of H. For this purpose, we have used the nonparametric m.l.e. based on the sample $\{S_i^2\}_{1 \leq i \leq n} := \sum_{j=1}^{k}(X_{ij} - \bar{X}_i)^2$, coming from a mixture of exponential distributions and used the algorithm given in Jewell (1982, p. 483). Roughly what this algorithm does is the following.

Table 1

ϵ	AV $T_n(\bar{\psi})$	AV m.l.e.	$\frac{\text{AV m.l.e.}}{3}$	$\frac{\text{AV m.l.e.}}{[3\{2-E(\tau)E(\frac{1}{\tau})\}]}$	AV $\bar{\bar{X}}$
0.03125	0.0238516	0.0624390	0.0208130	—	0.0722734
0.06250	0.0471831	0.1245136	0.0415045	—	0.1159776
0.09375	0.0698811	0.1858664	0.0619555	—	0.1492571
0.12500	0.0920136	0.2461538	0.0820513	—	0.1760374
0.15625	0.1135486	0.3050524	0.1016841	2.0018846	0.1982033
0.18750	0.1343806	0.3622641	0.1207547	0.5918380	0.2168714
0.21875	0.1543583	0.4175209	0.1391736	0.4250715	0.2327801
0.25000	0.1733220	0.4705882	0.1568627	0.3657613	0.2464523
0.28125	0.1911400	0.5212669	0.1737556	0.3383300	0.2582759
0.31250	0.2076875	0.5693950	0.1897983	0.3243945	0.2685485
0.34375	0.2228800	0.6148471	0.2049490	0.3172670	0.2775045
0.37500	0.2366652	0.6575342	0.2191781	0.3139313	0.2853321
0.40625	0.2490469	0.6974015	0.2324672	0.3128288	0.2921850
0.43750	0.2600419	0.7344262	0.2448087	0.3130759	0.2981900
0.46875	0.2697123	0.7686149	0.2562050	0.3141373	0.3034536
0.50000	0.2781355	0.8000000	0.2666667	0.3156735	0.3080654
0.53125	0.2854070	0.8286366	0.2762122	0.3174615	0.3121018
0.56250	0.2916332	0.8545994	0.2848665	0.3193524	0.3156283
0.59375	0.2969261	0.8779783	0.2926594	0.3212450	0.3187014
0.62500	0.3013433	0.8988764	0.2996255	0.3230715	0.3213700
0.65625	0.3130538	0.9174061	0.3058020	0.3247866	0.3236769
0.68750	0.3159907	0.9336870	0.3112290	0.3263611	0.3256595
0.71875	0.3183959	0.9478428	0.3159476	0.3277776	0.3273507
0.75000	0.3203480	0.9599999	0.3200000	0.3290271	0.3287795
0.78125	0.3257487	0.9702851	0.3234284	0.3301068	0.3299717
0.81250	0.3269285	0.9788234	0.3262745	0.3310180	0.3309500

Suppose that, we have a set of i.i.d. random variables $Y_1, Y_2, \ldots, Y_n$ with the common distribution of the form $\int_0^\infty e^{-t\tau} dH(\tau)$ for some distribution H on $\mathbb{R}^+$. It is easy to prove that the m.l.e. has a finite support containing at most n points lying in the interval $[\frac{1}{Y_{(n)}}, \frac{1}{Y_{(1)}}]$. Jewell (1982) used an EM-type algorithm to find for $r = 1, 2, 3, \cdots$ the m.l.e. $\widehat{H}_r$ when H is restricted to the set of all r-point distributions on $\mathbb{R}^+$ as long as the log-likelihood increases. For each $r = 2, 3, \cdots$ he used the uniform distribution over r equi-spaced points starting with $\frac{1}{Y_{(n)}}$ and ending with $\frac{1}{Y_{(1)}}$.

In our case, we found that with this starting distribution, the algorithm runs too slowly. Instead, we have obtained a starting point from $\widehat{H}_{r-1}$ by introducing a new point in the middle of two successive points which fall most widely apart from each other.

The complete computer program written in FORTRAN for VAX 8650 which we have used to get our results is given in Appendix A.

We shall now discuss, one by one, the results of these simulations. For each of these simulations, a comparison of the given estimates based on the simulated random sample is made in terms of the (suitably normalised) bias (*i.e.*, the average distance from the true value of μ), the standard error (s.e.) and the asymptotic variance in Tables 2(a)-(e), respectively; also note that due to the limitation of the CPU time we have obtained 1000 samples in two groups of equal size.

Simulation 1 Here we took H to be a discrete distribution taking values 1.0 and 15.0 with probabilities 0.2 and 0.8, respectively. This distribution is one of the mixing distributions considered by Jewell (1982). Note that the s.e. of the m.l.e. for two groups differ widely. We attribute this fact to the failure of the linearisation technique.

Simulation 2 Here H is a discrete distribution taking values 0.1 and 1.0 with equal probabilities 0.5.

Simulation 3 Here H is a discrete distribution taking values 0.2 and 1.0 with probabilities 0.4 and 0.6, respectively. We choose this distribution as a special case where $\bar{\bar{X}}$ is better than the m.l.e..

Table 2(a)

Group	Estimate	$\bar{\bar{X}}$	m.l.e.	$T_n(\bar{\psi})$
I	$\sqrt{n}$ bias	−0.0021574	0.0536739	−0.0008184
	s.e.	0.0911756	1.0408700	0.0536020
	AV	0.0844444	0.0819672	0.0296475
II	$\sqrt{n}$ bias	−0.0092238	0.1244006	−0.0016165
	s.e.	0.0716781	4.5568588	0.0394353
	AV	0.0844444	0.0819672	0.0296475
Combined	$\sqrt{n}$ bias	−0.0056906	0.0890372	−0.0012174
	s.e.	0.0814270	2.7988769	0.0465187
	AV	0.0844444	0.0819672	0.0296475

Table 2(b)

Group	Estimate	$\bar{\bar{X}}$	m.l.e.	$T_n(\bar{\psi})$
I	$\sqrt{n}$ bias	0.0147551	0.0912993	−0.0441951
	s.e.	1.8469815	2.2123895	1.9045249
	AV	1.8333333	1.8181818	0.7604553
II	$\sqrt{n}$ bias	−0.0381543	−0.0997603	−0.0085974
	s.e.	1.5362952	2.2512605	1.2087399
	AV	1.8333333	1.8181818	0.7604553
Combined	$\sqrt{n}$ bias	0.0116996	−0.0042305	−0.0263962
	s.e.	1.6916454	2.2319164	1.5566356
	AV	1.8333333	1.8181818	0.7604553

Table 2(c)

Group	Estimate	$\bar{\bar{X}}$	m.l.e.	$T_n(\bar{\psi})$
I	$\sqrt{n}$ bias	0.0053186	0.0270631	0.0127041
	s.e.	1.0257751	1.4094767	1.3838135
	AV	0.8666667	1.4705882	0.6027694
II	$\sqrt{n}$ bias	−0.0217991	−0.0709684	−0.0201541
	s.e.	0.8449745	1.7684397	0.8942075
	AV	0.8666667	1.4705882	0.6027694
Combined	$\sqrt{n}$ bias	−0.0082402	−0.0219526	−0.0037250
	s.e.	0.9353766	1.5889823	1.1390132
	AV	0.8666667	1.4705882	0.6027694

Simulation 4 Here H is a discrete distribution taking the values 0.05 and 1.0 with equal probabilities 0.5. Contrast this case with Simulation 3.

Simulation 5 Finally, we take H to be a gamma distribution with $\alpha = 2.0$ and $\lambda = 1.0$.

Table 2(d)

Group	Estimate	$\bar{\bar{X}}$	m.l.e.	$T_n(\bar{\psi})$
I	$\sqrt{n}$ bias	0.0281004	0.0184313	−0.0740670
	s.e.	3.4746987	3.9005452	1.6857447
	AV	3.5000000	1.9047619	0.7436790
II	$\sqrt{n}$ bias	−0.0612840	−0.1136078	−0.0126448
	s.e.	2.927079	3.4211260	1.6804266
	AV	3.5000000	1.9047619	0.7436790
Combined	$\sqrt{n}$ bias	0.0165918	−0.0475882	−0.0433559
	s.e.	3.2009089	3.6608793	1.6830951
	AV	3.5000000	1.9047619	0.7436790

Table 2(e)

Group	Estimate	$\bar{\bar{X}}$	m.l.e.	$T_n(\bar{\psi})$
I	$\sqrt{n}$ bias	0.0223542	−0.0203280	−0.0046954
	s.e.	0.3629048	0.5808004	0.4678451
	AV	0.3333333	0.5000000	0.2142857
II	$\sqrt{n}$ bias	0.0122539	0.0548677	−0.0061972
	s.e	0.3510266	0.9207742	0.3168615
	AV	0.3333333	0.5000000	0.2142857
Combined	$\sqrt{n}$ bias	0.0173040	0.0172698	−0.0007509
	s.e.	0.3569660	0.7507874	0.3923536
	AV	0.3333333	0.5000000	0.2142857

Chapter 9
The Counterexample

In this chapter, we shall state the counterexample mentioned in Chapter 1 suggesting that it may not be easy to get simple regularity conditions on f gauranteeing the continuity of $\bar{\psi}$.

First, observe that condition U(i) of Chapter 4 is equivalent to the following :

(F1) For any (θ, G) in $\Theta \times \mathcal{G}$, the likelihood ratio $\Lambda(\cdot, \theta, G, \theta', G)$ is quadratic-mean-differentiable with respect to θ' at the point θ. In other words, there is $s_\theta(\cdot, \theta, G)$ in $L_2^\circ(P_{\theta,G})$ such that

$$\|\{\Lambda(\cdot, \theta, G, \theta + h, G) - 1\} - hs_\theta(\cdot, \theta, G)\|_{L_2(P_{\theta,G})} = o(|h|).$$

For any (θ, G) in $\Theta \times \mathcal{G}$, let $\bar{N}_{\theta,G}$ denote the set of all functions ψ in $L_2(P_{\theta,G})$ such that there is a map ν from $[-1, 1]$ to $\mathcal{G}$ with $\nu(0) = G$ such that

$$\|\{\Lambda(\cdot, \theta, G, \theta + h, \nu(k)) - 1\} - hs_\theta(\cdot, \theta, G) - k\psi\|_{L_2(P_{\theta,G})} = o(\sqrt{h^2 + k^2}).$$

Then $\bar{N}_{\theta,G}$ is dense in $N_{\theta,G}$ so that one can define $\bar{\psi}$ using $\bar{N}_{\theta,G}$ instead of $N_{\theta,G}$.

Note that for any (θ, G), $L_2^\circ(P_{\theta,G})$ is isometric to $L_2(\mu) \cap \{\phi : \langle \phi, f^{\frac{1}{2}}(\cdot, \theta, G)\rangle_{L_2(\mu)} = 0\}$ and $\bar{N}_{\theta,G}$ is isometric to $\bar{N}^1_{\theta,G} := \{\frac{1}{2}\psi f^{\frac{1}{2}}(\cdot, \theta, G) : \psi \in \bar{N}_{\theta,G}\}$.

Define the functions ρ_θ and ρ_θ^* from $\Theta \times \mathcal{G}$ to $L_2^\circ(\mu)$ by

$$\left.\begin{array}{rcl} \rho_\theta(\cdot, \theta, G) & = & \frac{1}{2}s_\theta(\cdot, \theta, G) f^{\frac{1}{2}}(\cdot, \theta, G) \\ \text{and } \rho_\theta^*(\cdot, \theta, G) & = & \text{Proj}_{(\bar{N}^1_{\theta,G})^\perp}\, \rho_\theta(\cdot, \theta, G) \end{array}\right\} \forall(\theta, G).$$

Then

$$\bar{\psi} = \frac{2\rho_\theta^*}{f^{\frac{1}{2}}}. \tag{9.1}$$

The following is a weaker version of assumption (B3) as required by Schick (1986).

(F2) For any (θ, G) in $\Theta \times \mathcal{G}$

$$\lim_{(\theta',G')\to(\theta,G)} \|\bar{\psi}(\cdot,\theta',G') - \bar{\psi}(\cdot,\theta,G)\|_{L_2^s(P_{\theta,G})} = 0.$$

In view of relation (9.1), one can easily prove that

Proposition 9.1 *If assumptions (A2), (F1) and (F2) hold, Then ρ_θ^* is a continuous map from $\Theta \times \mathcal{G}$ to $L_2(\mu)$.*

Usual choices of $\mathcal{G}$ and regularity conditions on f guarantee assumptions (A2) and (F1). However, in what follows, we shall give an example suggesting that the continuity conditions insured by assumptions (A2) and (F1) may not suffice for the continuity of ρ_θ^*. Specifically, we show that, if there is a function taking values in a Hilbert space $\mathcal{H}$ and we consider its projection on linear spaces which are generated by continuous functions the projection of the continuous function will not in general be continuous.

The reason why a counterexample is non-trivial is that the (closed) linear spaces on which the projection is made are isomorphic to a fixed set L. If L were finite-dimensional, projections would have been continuous.

Before stating the example let us make a brief mathematical formulation of the problem.

Let X be a compact topological space. Let $\mathcal{H}$ be an infinte-dimensional Hilbert space. Let $f : X \to \mathcal{H} \oplus \mathbb{C}$ and $g : X \times \mathcal{H} \to \mathcal{H} \oplus \mathbb{C}$ be two continuous maps with g satisfying the additional property that for any x in X, the map $x \mapsto g(x, \cdot)$ is one-one.

For any x in X, let $\mathcal{M}(x) := \overline{\{g(x,h) : h \in \mathcal{H}\}}$.

The (topological) spaces X and $\mathcal{H} \oplus \mathbb{C}$ are like our $\Theta \times \mathcal{G}$ and $L_2(\mu)$, respectively, the function f is like our ρ_θ and the (linear) spaces $\mathcal{M}(x)$ and $\mathcal{H}$ are like our $(\dot{N}_{\theta,G}^1)^\perp$ and L, respectively.

The question is whether $x \mapsto \mathrm{Proj}_{\mathcal{M}(x)} f(x)$ is continuous or not.

Here is an example providing negative answer to the above question.

Example 9.1 Let $\mathcal{H}$ be an infinite-dimensional Hilbert space. Let $T : \mathcal{H} \to \mathcal{H}$ be a compact operator satisfying the additional property that

(a) T is one-one (equivalently, 0 is not an eigenvalue of T).

Claim : $\exists h_n \in \mathcal{H}$ such that $\|h_n\| = 1$ but $\|Th_n\| \leq \frac{1}{n^2}$.

First observe that, because of (a), $T\mathcal{H}$ is an infinite-dimensional topological vector space. Clearly, it is enough to show that $\nexists\lambda > 0$ such that

$$\|x\| \leq \lambda \|Tx\| \quad \forall x \in \mathcal{H}.$$

If there were such a λ, define, once again by (a), $S : T\mathcal{H} \to \mathcal{H}$ by

$$S(Tx) = x \quad \forall x.$$

Then

$$\|Sy\| = \|x\| \leq \lambda \|y\|$$

putting $y = Tx$ so that S extends to $\mathcal{H}_1 := \overline{T\mathcal{H}}$ — an infinite-dimensional Hilbert space and

$$T(Sy) = y \text{ on } \mathcal{H}_1. \tag{9.2}$$

Let B_1 denote the unit-ball on $\mathcal{H}_1$. Then, there is $r > 0$ such that

$$SB_1 \subseteq B_2^r$$

— a ball of radius r in $\mathcal{H}$.

But $\overline{T(B_2^r)}$ is compact.

Therefore, by relation (9.2), $\overline{T(SB_1)}$ is a compact set containing B_1.

This is a contradiction to the fact that $\mathcal{H}_1$ is infinite-dimensional.

Hence the claim follows.

For any integer $n \geq 1$, using Hahn-Banach Theorem get hold of a function $\lambda_n : \mathcal{H} \to \mathbb{C}$ such that $\lambda_n(h_n^\circ) = 1$ and $\|\lambda_n\|_{\text{op}} = 1$ and define a map $T_n : \mathcal{H} \to \mathcal{H} \oplus \mathbb{C}$ by $T_n := T(h) \oplus \frac{\lambda_n(h)}{h}$ for all h, Define $T_\circ$ by $T_\circ(h) = T(h) \oplus 0$ for all h. Let $X = \{1, \frac{1}{2}, \cdots, \frac{1}{n}, \cdots\} \cup \{0\}$.

Define $f : X \to \mathcal{H} \oplus \mathbb{C}$ by $f(x) := 0 \oplus 1$ for all x. (*i.e.*, f takes a constant value on $\mathcal{H} \oplus \mathbb{C}$.)

Define $g : X \times \mathcal{H} \to \mathcal{H} \oplus \mathbb{C}$ by

$$g(x,h) := \sum_{n=1}^{\infty} 1_{\{x:x=\frac{1}{n}\}} T_n(h) + 1_{\{x:x=0\}} T_o(h)$$

for all (x,h) in $X \times \mathcal{H}$.

One can easily check that f and g are continuous and for all x, $g(x,\cdot)$ is a one-one map from $\mathcal{H}$ to $\mathcal{H} \oplus \mathbb{C}$.

Let, for any x in X,

$$\mathcal{N}(x) = \overline{\{g(x,h) : h \in \mathcal{H}\}}.$$

Claim : $x \mapsto \mathrm{Proj}_{\mathcal{N}(x)} f(x)$ is *not* continuous.

Proof of the claim :

Clearly,

$$\begin{aligned}
\mathrm{Proj}_{\mathcal{N}(0)} f(0) &= \mathrm{Proj}_{\overline{\{T_o(h): h\in\mathcal{H}\}}}(0 \oplus 1) \\
&= \mathrm{Proj}_{\{T(h)\oplus 0: h\in\mathcal{H}\}}(0 \oplus 1) \\
&= 0 \oplus 0 = 0 \text{ (of } \mathcal{H} \bigoplus \mathbb{C}). \qquad (9.3)
\end{aligned}$$

Now,

$$\begin{aligned}
\mathrm{Proj}_{\mathcal{N}(\frac{1}{n})} f(\frac{1}{n}) &= \mathrm{Proj}_{\overline{\{T_n(h): h\in\mathcal{H}\}}}(0 \oplus 1) \\
&= y_n \oplus \mu_n \qquad \text{(say)}. \qquad (9.4)
\end{aligned}$$

Note that

$$\|y_n\|^2 + \|\mu_n\|^2 \le \|f(\frac{1}{n})\|^2 = 1. \qquad (9.5)$$

Again,

$$(0 \oplus 1 - y_n \oplus \mu_n) \perp T(h) \oplus \frac{\lambda_n(h)}{h} \text{ for all } h.$$

In particular,

$$\langle y_n \oplus (\mu_n - 1), T_n(h_n) \oplus \frac{1}{n}\rangle_{\mathcal{H} \bigoplus \mathbb{C}} = 0$$

equivalently,

$$\langle y_n, T_n(h_n)\rangle_{\mathcal{H}} + \frac{(\mu_n - 1)}{n} = 0.$$

Therefore

$$|\frac{\mu_n - 1}{n}| = |\langle y_n, T_n(h_n)\rangle_{\mathcal{H}}| \leq \frac{1}{n^2}$$

by relation (9.5) and choice of h_n.

Therefore

$$|\mu_n - 1| \leq \frac{1}{n} \to 0 \text{ as } n \to \infty$$

implying

$$\mu_n \to 1 \text{ as } n \to \infty. \tag{9.6}$$

Claim follows from relations (9.3), (9.4) and (9.6). □

We shall now give a brief motivation behind the choice of g and $\mathcal{H}$ for the special case of semiparametric models satisfying Bickel's Condition C, *i.e.*, semiparametric models where $\mathcal{G}$ is convex and f is affine in G. This family includes the so-called mixture models.

Assume that

(F3) $\mathcal{G}$ is a compact subset of a topological vector space $\mathcal{L}$.

Remark 9.1 For the special case of mixture models a natural choice of $\mathcal{L}$ is the set of all signed measures $\mathcal{M}$ on Ξ.

The affinity of f enables us to extend the definition of f and A to the spaces $S \times \Theta \times \mathcal{L}$ and $S \times \Theta \times \mathcal{G} \times \Theta \times \mathcal{L}$, respectively.

In view of assumption (F3), for any (θ, G), $\bar{N}_{\theta,G}$ can be replaced by

$$\{A(\cdot, \theta, G, \theta, L) : L \in \mathcal{L}\}$$

and $\bar{N}^1_{\theta,G}$ can be replaced by

$$\{g(\theta, G, L) : L \in \mathcal{L}\}$$

where $g(\theta, G, L) := \frac{f(\cdot,\theta,L)}{\frac{1}{2}f^{\frac{1}{2}}(\cdot,\theta,G)}\ \forall(\theta, G, L)$.

Appendix A
Computer Programs

This chapter contains the computer program mentioned in Chapter 8. All the pseudorandom number generation algorithms are taken from Fishman (1978).

```
      PROGRAM SIMULATION1
      INTEGER*2 ONE,TWO,THREE,FOUR,FIVE,SIX,SEVEN,EIGHT,TEN,SIXTY
      REAL*8 ZERO,ONE_HALF
      REAL*8 EPS
      INTEGER RE,RO
      REAL*8 PSI2A,PSI2B,PHI
      REAL*8 BIAS1,BIAS2,BIAS3,C,D1,DN,DSQRTN,PHIE,PHIO,PSI2EMAX,
     $PSI2EMIN,PSI2OMAX,PSI2OMIN,Q,QO,S,SS,SUM1,SUM2,SUM3,SUM4,SUM5,T,
     ¢TAU,THETA,TN1,TN2,TN2O,TN3,U,V1,V2,V3,VAR1,VAR2,VAR3
      REAL*8 LAMDAE(5000),LAMDAO(5000),PE(5000),PO(5000),PSI1(10000),
     $PSI2(10000),PSI2E(5000),PSI2O(5000),X(10000,3)
      REAL*8 FUNC,MIXPOS,NORMAL,RANDOM
      COMMON NERR /BLK1/ PSI2A,PSI2B,PHI
      PARAMETER (ONE=1,TWO=2,THREE=3,FOUR=4,FIVE=5,SIX=6,SEVEN=7)
      PARAMETER (EIGHT=8,TEN=10,SIXTY=60)
      PARAMETER (ZERO=0.0D0,ONE_HALF=1.5D0)
      PARAMETER (EPS='0000000000003D00'X)
      PARAMETER (MAXITER=1000)
      DATA NBAD1_1,NBAD1_2,NBAD2_1,NBAD2_2 /4*0/
      DATA BIAS1,BIAS2,BIAS3,VAR1,VAR2,VAR3 /6*0.0D0/
      DATA LAMDAE,LAMDAO,PE,PO,PSI1,PSI2,PSI2E,PSI2O,X /80000*0.0D0/
      FUNC(S,INIT)=NORMAL(INIT)/DSQRT(S)
      WRITE (*,*) 'GIVE THE VALUE OF THE PARAMETER OF INTEREST.'
1     READ (*,*,ERR=14) THETA
      WRITE (*,*) 'GIVE YOUR CHOICE OF THE MIXING DISTRIBUTION FROM THE'
     $,' SET {1,2,3}.'
2     READ (*,*,ERR=15) NDIST
      IF ((NDIST.NE.ONE).AND.(NDIST.NE.TWO).AND.(NDIST.NE.THREE))
     $GO TO 15
```

```
      WRITE (*,*) 'GIVE THE SAMPLE SIZE.'
3     READ (*,*,ERR=16) N
      IF (N.LE.0) THEN
      GO TO 16
      ELSE IF (N.GT.10000) THEN
      WRITE (*,*) 'SAMPLE SIZE IS TRUNCATED TO 10000.'
      N=10000
      END IF
      WRITE (*,*) 'GIVE THE NUMBER OF SAMPLES TO BE GENERATED.'
4     READ (*,*,ERR=17) NITER
      NITER=MAX0(MIN0(NITER,MAXITER),ONE)
      OPEN (1,FILE='OUT3')
      WRITE (1,'(1H1,40X,I4,A,I6,A)') NITER,
     $' SAMPLES TO BE GENERATED EACH OF SIZE',N,'.'
      WRITE (*,*) 'GIVE AN INTEGER BETWEEN 1 AND 2147483646.'
5     READ (*,*,ERR=18) INIT
      IF ((INIT.LE.0).OR.(INIT.GT.2147483646)) GO TO 18
      WRITE (1,'(1H0,46X,A,I10,A)') 'THE SEED PRIOR TO SAMPLING IS ',
     $INIT,'.'
      WRITE (1,'(///11X,A,12X,A,13X,A,18X,A,16X,A)') 'SAMPLE NUMBER',
     $'SEED AT END','GRAND MEAN','MLE','ONE STEP MLE'
      LINE=EIGHT
      C=ONE/ONE_HALF
      NE=N/TWO
      NO=N-NE
      DN=DBLE(N)
      DSQRTN=DSQRT(DN)
      DO 13 ITER=ONE,NITER
      WRITE (*,*) 'ITER=',ITER
      SUM3=ZERO
      DO 7 I=ONE,N
      SUM1=ZERO
      SUM2=ZERO
      TAU=MIXPOS(NDIST,INIT)
      DO 6 J=ONE,THREE
      X(I,J)=FUNC(TAU,INIT)+THETA
      S=X(I,J)
      SUM1=SUM1+S
6     SUM2=SUM2+S*S
      S=SUM1/THREE
```

```
      T=(SUM2-S*SUM1)/TWO
      PSI1(I)=S
      PSI2(I)=T
      IF (MOD(I,TWO).EQ.ONE) THEN
      PSI2O((I+ONE)/TWO)=T
      IF (I.EQ.ONE) THEN
      PSI2OMIN=T
      PSI2OMAX=T
      ELSE
      PSI2OMIN=DMIN1(PSI2OMIN,T)
      PSI2OMAX=DMAX1(PSI2OMAX,T)
      END IF
      ELSE
      PSI2E(I/TWO)=T
      IF (I.EQ.TWO) THEN
      PSI2EMIN=T
      PSI2EMAX=T
      ELSE
      PSI2EMIN=DMIN1(PSI2EMIN,T)
      PSI2EMAX=DMAX1(PSI2EMAX,T)
      END IF
      END IF
7     SUM3=SUM3+S
      TN1=SUM3/N
      V1=TN1-THETA
      BIAS1=BIAS1+V1
      VAR1=VAR1+V1*V1
      WRITE (*,*) 'THE GRAND MEAN IS ',TN1
      SUM1=ZERO
      DO 8 I=ONE,N
      S=PSI1(I)-TN1
      T=PSI2(I)*C
8     SUM1=SUM1+S/(S*S+T)
      Q=DABS(SUM1)
      M2=0
      TN2=TN1
9     TN2O=TN2
      QO=Q
      SUM1=ZERO
      SUM2=ZERO
```

```
      SUM3=ZERO
      DO 10 I=ONE,N
      S=PSI1(I)-TN2
      SS=S*S
      T=PSI2(I)*C
      U=ONE/(SS+T)
      SUM1=SUM1+S*U
      SUM2=SUM2+U
10    SUM3=SUM3+T*U*U
      TN2=TN2-SUM1/(SUM2-TWO*SUM3)
      Q=DABS(SUM1)
      D1=DABS(TN2-TN20)
      M2=M2+ONE
      IF (MOD(M2,1000).EQ.0) WRITE (*,*) 'M2=',M2,' TN2=',TN2,' Q=',Q
      IF ((D1.GE.EPS).AND.(D1.LT.1.0D0).AND.(Q.LT.Q0).AND.(M2.LT.1))
     $ GO TO 9
      WRITE (*,*) 'M2=',M2,' TN2=',TN2,' Q=',Q
      IF ((Q0.LE.Q).OR.(D1.GE.1.0D0)) TN2=TN20
      V2=TN2-THETA
      IF (DABS(V1).LT.DABS(V2)) THEN
      IF (V1*V2.GT.ZERO) THEN
      NBAD1_1=NBAD1_1+ONE
      ELSE
      NBAD1_2=NBAD1_2+ONE
      END IF
      END IF
      BIAS2=BIAS2+V2
      VAR2=VAR2+V2*V2
      WRITE (*,*) 'THE MLE IS ',TN2
      PSI2A=ONE/PSI20MAX
      PSI2B=ONE/PSI20MIN
      CALL MAXLHD(N0,PSI20,R0,LAMDA0,P0)
      PHI0=PHI
      PSI2A=ONE/PSI2EMAX
      PSI2B=ONE/PSI2EMIN
      CALL MAXLHD(NE,PSI2E,RE,LAMDAE,PE)
      PHIE=PHI
      SUM1=ZERO
      SUM2=ZERO
      DO 12 I=ONE,N
```

```
      S=PSI1(I)-TN1
      SUM3=ZERO
      SUM4=ZERO
      SUM5=ZERO
      IF (MOD(I,TWO).EQ.ONE) THEN
      DO K=ONE,RE
      T=LAMDAE(K)
      U=DSQRT(T)*DEXP(-ONE_HALF*T*S*S)*PE(K)
      SUM3=SUM3+U
      U=T*U
      SUM4=SUM4+U
      U=T*U
      SUM5=SUM5+U
      END DO
      ELSE
      DO K=ONE,RO
      T=LAMDAO(K)
      U=DSQRT(T)*DEXP(-ONE_HALF*T*S*S)*PO(K)
      SUM3=SUM3+U
      U=T*U
      SUM4=SUM4+U
      U=T*U
      SUM5=SUM5+U
      END DO
      END IF
      IF (SUM3.NE.ZERO) THEN
      T=SUM4/SUM3
      U=SUM5/SUM3
      ELSE
      T=ZERO
      U=ZERO
      END IF
      SUM1=SUM1+S*T
12    SUM2=SUM2+(3.ODO*S*S*(U-T*T)-T)
      TN3=TN1-SUM1/SUM2
      IF (DABS(TN3-TN1).GT.1.ODO) TN3=TN1
      V3=TN3-THETA
      IF (DABS(V1).LT.DABS(V3)) THEN
      IF (V1*V3.GT.ZERO) THEN
      NBAD2_1=NBAD2_1+ONE
```

```
      ELSE
      NBAD2_2=NBAD2_2+ONE
      END IF
      END IF
      BIAS3=BIAS3+V3
      VAR3=VAR3+V3*V3
      WRITE (*,*) 'THE ONE STEP ESTIMATE IS ',TN3
      LL=LINE+ONE
      IF (LL.GT.SIXTY) THEN
      WRITE (1,'(1H1,10X,A,12X,A,13X,A,18X,A,16X,A)') 'SAMPLE NUMBER',
     $'SEED AT END','GRAND MEAN','MLE','ONE STEP MLE'
      LL=THREE
      END IF
      WRITE (1,'(1H0,13X,I4,17X,I10,7X,3G24.16))') ITER,INIT,TN1,TN2,TN3
      LINE=MOD(LL,SIXTY)+ONE
      DO 13 I=ONE,N
      IF (I.LE.RE) THEN
      LAMDAE(I)=ZERO
      PE(I)=ZERO
      END IF
      IF (I.LE.RO) THEN
      LAMDAO(I)=ZERO
      PO(I)=ZERO
      END IF
      DO 13 J=ONE,THREE
13    X(I,J)=ZERO
      WRITE (*,*) 'NBAD1_1=',NBAD1_1,' NBAD1_2=',NBAD1_2,' NBAD2_1=',
     $NBAD2_1,' NBAD2_2=',NBAD2_2
      LINE=MOD(LL,SIXTY)+ONE
      LL=LINE+THREE
      IF (LL.GT.SIXTY) THEN
      WRITE (1,'(1H1)')
      LL=FOUR
      END IF
      WRITE (1,'(1H0,9X,A,I5,A,I5,2A/1H0,3X,A,I5,2A,I5,A)')
     $'FOR THE MLE THE NUMBER OF BAD SAMPLES OF TYPE ONE IS',NBAD1_1,
     @' AND THAT OF TYPE TWO IS',NBAD1_2,'.FOR THE ONE STEP MLE THE ',
     #'NUMBER OF',' BAD SAMPLES OF TYPE ONE IS',NBAD2_1,' AND THAT OF',
     $' TYPE TWO IS',NBAD2_2,'.'
      LINE=MOD(LL,SIXTY)+ONE
```

```
      S=DN/(NITER-ONE)
      T=BIAS1/NITER
      VAR1=(VAR1-BIAS1*T)*S
      BIAS1=T*DSQRTN
      T=BIAS2/NITER
      VAR2=(VAR2-BIAS2*T)*S
      BIAS2=T*DSQRTN
      T=BIAS3/NITER
      VAR3=(VAR3-BIAS3*T)*S
      BIAS3=T*DSQRTN
      WRITE (*,*) 'BIAS1=',BIAS1,' BIAS2=',BIAS2,' BIAS3=',BIAS3
      LL=LINE+THREE
      IF (LL.GT.SIXTY) THEN
      WRITE (1,'(1H1)')
      LL=FOUR
      END IF
      WRITE (1,'(1HO,9X,2(A,G23.15),A/1HO,4X,G23.15,A)')
     $'THE ESTIMATED BIAS OF THE GRAND MEAN IS',BIAS1,',THE MLE IS',
     @BIAS2,' AND THE ONE STEP MLE IS',BIAS3,'.'
      LINE=MOD(LL,SIXTY)+ONE
      WRITE (*,*) 'VAR1=',VAR1,' VAR2=',VAR2,' VAR3=',VAR3
      LL=LINE+THREE
      IF (LL.GT.SIXTY) THEN
      WRITE (1,'(1H1)')
      LL=FOUR
      END IF
      WRITE (1,'(1HO,9X,2(A,G23.15),A/1HO,4X,A,G23.15,A)')
     $'THE ESTIMATED ASYMPTOTIC VARIANCE OF THE GRAND MEAN IS',VAR1,
     @',THE MLE IS',VAR2,' AND THE ','ONE STEP MLE IS',VAR3,'.'
      STOP
14    WRITE (*,*) 'ERROR:THETA MUST BE A REAL NUMBER.'
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 1
      GO TO 19
15    WRITE (*,*) 'ERROR:CHOICE OF THE MIXING DISTRIBUTION LIMITED TO',
     $' THE SET {1,2,3}.'
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 2
      GO TO 19
16    WRITE (*,*) 'ERROR:SAMPLE SIZE MUST BE A POSITIVE INTEGER.'
```

```
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 3
      GO TO 19
17    WRITE (*,*) 'ERROR:THE NUMBER OF ITERATIONS MUST BE A POSITIVE',
     $' INTEGER.'
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 4
      GO TO 19
18    WRITE (*,*) 'ERROR:THE SEED MUST LIE BETWEEN 1 AND 2147483646.'
      IF (NERR.LE.TEN) GO TO 5
19    STOP
     $'THE PROGRAM IS TERMINATING DUE TO MORE THAN TEN I/O ERRORS.'
      END
      REAL*8 FUNCTION RANDOM(INIT)
      INTEGER*2 ZERO,ONE,TWO,THREE,FOUR,FIVE,SIX,SEVEN,EIGHT
      INTEGER CARRY,PRDT
      DIMENSION M(4),I(4),IM(8)
      REAL*8 C
      PARAMETER (ZERO=0,ONE=1,TWO=2,THREE=3,FOUR=4,FIVE=5,SIX=6,SEVEN=7)
      PARAMETER (EIGHT=8)
      PARAMETER (MAXPRM=2147483647)
      PARAMETER (C=2.147483648D+9)
      DATA M /126,218,172,23/
*     M IS A BINARY REPRESENTATION OF THE NUMBER 397204094 WRITTEN BYTEWISE
      IF (INIT.LE.ZERO) THEN
      WRITE (*,*) 'ERROR:THE SEED IS NEGATIVE.'
      STOP
      END IF
      I(FOUR)=INIT/16777216
      J=INIT-I(FOUR)*16777216
      I(THREE)=J/65536
      J=J-I(THREE)*65536
      I(TWO)=J/256
      I(ONE)=J-I(TWO)*256
      CARRY=ZERO
      DO 2 J=TWO,EIGHT
      LMIN=MAX0(ONE,J-FOUR)
      LMAX=MIN0(FOUR,J-ONE)
      PRDT=CARRY
      DO 1 L=LMIN,LMAX
```

```
1     PRDT=PRDT+I(L)*M(J-L)
      CARRY=PRDT/256
2     IM(J-ONE)=PRDT-CARRY*256
      IM(EIGHT)=CARRY
      K=33554432*IM(EIGHT)+131072*IM(SEVEN)+512*IM(SIX)+2*IM(FIVE)+
     $IM(FOUR)/128
      INIT=MOD(IM(FOUR),128)*16777216+IM(THREE)*65536+IM(TWO)*256+
     $IM(ONE)
      MK=MAXPRM-K
      IF (INIT.LT.MK) THEN
      INIT=INIT+K
      ELSE
      INIT=INIT-MK
      END IF
      RANDOM=INIT/C
      RETURN
      END

      REAL*8 FUNCTION NORMAL(INIT)
      LOGICAL*2 IND1,IND2,IND3
      REAL*8 A,B,RANDOM,U1,U2,V,W1,W2
      REAL*8 ZERO,ONE,TWO
      DATA ZERO,ONE,TWO /0.0D0,1.0D0,2.0D0/
      DATA IND1 /.FALSE./
      IF (IND1) THEN
      NORMAL=W2
      ELSE
1     U1=RANDOM(INIT)
      U1=TWO*U1
      IND2=(U1.GT.ONE)
      IF (IND2) U1=U1-ONE
      U2=RANDOM(INIT)
      U2=TWO*U2
      IND3=(U2.GT.ONE)
      IF (IND3) U2=U2-ONE
      B=U1*U1+U2*U2
      IF (B.GT.ONE) GO TO 1
      V=-DLOG(RANDOM(INIT))
      A=DSQRT(TWO*V/B)
      W1=U1*A
      W2=U2*A
```

```
IF (IND2) W1=-W1
IF (IND3) W2=-W2
NORMAL=W1
END IF
IND1=.NOT.IND1
RETURN
END

REAL*8 FUNCTION U(A,B,INIT)
REAL*8 A,B,C
REAL*8 RANDOM
IF (A.GT.B) THEN
WRITE (*,*) 'ERROR:UPPER LIMIT < LOWER LIMIT.'
C=A
A=B
B=C
END IF
U=A
IF (A.NE.B) U=U+(B-A)*RANDOM(INIT)
RETURN
END

REAL*8 FUNCTION D(K,X,P,INIT)
INTEGER*2 ONE
LOGICAL IND
REAL*8 SUM,T
REAL*8 P(K),X(K)
REAL*8 RANDOM
PARAMETER (ONE=1)
DATA IND /.TRUE./
IF (IND) THEN
DO 1 I=ONE,K-ONE
DO 1 J=I+ONE,K
IF (X(I).GT.X(J)) THEN
T=X(I)
X(I)=X(J)
X(J)=T
T=P(I)
P(I)=P(J)
P(J)=T
END IF
```

```
1     CONTINUE
      IND=.FALSE.
      END IF
      SUM=0.0D0
      T=RANDOM(INIT)
      DO 2 I=ONE,K
      SUM=SUM+P(I)
      IF (SUM.GE.T) GO TO 3
2     CONTINUE
3     D=X(I)
      RETURN
      END

      REAL*8 FUNCTION GAMMA(ALPHA,LAMDA,INIT)
      REAL*8 ONE
      REAL*8 ALPHA,LAMDA
      REAL*8 A,B,C,U1,U2,V
      REAL*8 RANDOM
      LOGICAL*1 IND /.TRUE./
      PARAMETER (ONE=1.0D0)
      IF ((ALPHA.LT.ONE).OR.(LAMDA.LE.0.0D0)) GO TO 2
      IF (ALPHA.EQ.ONE) THEN
      GAMMA=-DLOG(RANDOM(INIT))
      ELSE
      IF (IND) THEN
      A=DSQRT(2.0D0*ALPHA-ONE)
      B=ALPHA-DLOG(4.0D0)
      C=ALPHA+A
      A=ONE/A
      IND=.FALSE.
      END IF
1     U1=RANDOM(INIT)
      U2=RANDOM(INIT)
      V=A*DLOG(U1/(ONE-U1))
      GAMMA=ALPHA*DEXP(V)
      IF ((B+C-GAMMA).LT.DLOG(U1*U1*U2)) GO TO 1
      END IF
      IF (LAMDA.NE.ONE) GAMMA=GAMMA/LAMDA
      RETURN
2     WRITE (*,*) 'ERROR:ALPHA MUST BE AT LEAST ONE AND LAMDA MUST BE',
     $' POSITIVE.'
```

```
      STOP
      END
      REAL*8 FUNCTION MIXPOS(NDIST,INIT)
      INTEGER*2 ONE,TWO,THREE,TEN
      REAL*8 ZERO
      CHARACTER CHR1*80,CHR2*30
      LOGICAL IND*1 /.TRUE./
      REAL*8 A,ALPHA,B,C,LAMDA,T
      REAL*8 P(50),X(50)
      REAL*8 D,GAMMA,RANDOM,U
      COMMON NERR
      PARAMETER (ONE=1,TWO=2,THREE=3,TEN=10)
      PARAMETER (ZERO=0.0D0)
      PARAMETER (CHR1=' THE SUPPORT OF THE MIXING DISTRIBUTION CAN NOT
     $CONTAIN ZERO:TYPE FRESH INPUT.')
      PARAMETER (CHR2='INPUT-OUTPUT ERROR:TRY AGAIN.')
      DATA P,X /100*0.0D0/
      IF (IND) THEN
      OPEN (2,FILE='IN2')
      IF (NDIST.EQ.ONE) THEN
      WRITE (*,*) 'GIVE THE LOWER LIMIT (A) AND THE UPPER LIMIT (B).'
1     READ (*,*,ERR=2) A,B
      IF (A.GT.B) THEN
      WRITE (*,*) 'ERROR:UPPER LIMIT < LOWER LIMIT.'
      C=A
      A=B
      B=A
      END IF
      IF (B.LT.ZERO) THEN
      WRITE (*,*) 'ERROR:SIGNS OF BOTH A AND B ARE WRONG.'
      C=A
      A=-B
      B=-C
      ELSE IF (A.LE.ZERO) THEN
      WRITE (*,*) CHR1
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 1
      GO TO 10
      END IF
      WRITE (2,'(1H1///14X,2A,G23.16,A,G23.16,A)') 'THE MIXING',
```

```
     $' DISTRIBUTION IS THE UNIFORM DISTRIBUTION OVER(',A,',',B,').'
      CLOSE (2)
      IND=.FALSE.
      GO TO 9
2     WRITE (*,*) CHR2
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 1
      GO TO 10
      ELSE IF (NDIST.EQ.TWO) THEN
      WRITE (*,*) 'GIVE VALUES OF K,X AND P.'
3     READ (*,*,ERR=6) K,(X(I),P(I),I=ONE,K)
      DO 4 I=ONE,K
      IF (X(I).EQ.ZERO) THEN
      WRITE (*,*) CHR1
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 3
      GO TO 10
      ELSE IF (X(I).LT.ZERO) THEN
      WRITE (*,*) 'NEGATIVE SAMPLE POINT:DROPPING THE NEGATIVE SIGN.'
      X(I)=-X(I)
      END IF
      IF ((P(I).LE.ZERO).OR.(P(I).GT.1.ODO).OR.
     $((P(I).EQ.1.ODO).AND.(K.GT.ONE))) THEN
      WRITE (*,'(A,I4,A)') ' THE CHOICE OF THE PROBABILITY OF THE ',I,
     $'-TH SAMPLE POINT IS WRONG:GIVE THE CORRECT VALUE.'
      NERR=NERR+ONE
      IF (NERR.GT.TEN) GO TO 10
      READ (*,*) P(I)
      END IF
4     CONTINUE
      DO 5 I=ONE,K-ONE
      DO 5 J=I+ONE,K
      IF (X(I).GT.X(J)) THEN
      T=X(I)
      X(I)=X(J)
      X(J)=T
      T=P(I)
      P(I)=P(J)
      P(J)=T
      END IF
```

```
5     CONTINUE
      WRITE (2,'(1H1///38X,2A/(7X,4(A,I2,A,G23.16)))') 'THE MIXING',
     $' DISTRIBUTION IS A DISCRETE DISTRIBUTION WITH',
     @(' X(',I,')=',X(I),' P(',I,')=',P(I),I=ONE,K)
      CLOSE (2)
      IND=.FALSE.
      GO TO 9
6     WRITE (*,*) CHR2
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 3
      GO TO 10
      ELSE IF (NDIST.EQ.THREE) THEN
      WRITE (*,*) 'GIVE THE VALUES OF ALPHA AND LAMDA.'
7     READ (*,*,ERR=8) ALPHA,LAMDA
      IF ((ALPHA.LT.ONE).OR.(LAMDA.LE.ZERO)) GO TO 8
      WRITE (2,'(1H1///8X,2A,G23.16,A,G23.16,A)') 'THE MIXING',
     $' DISTRIBUTION IS A GAMMA DISTRIBUTION WITH ALPHA=',ALPHA,
     @' AND LAMDA=',LAMDA,'.'
      IND=.FALSE.
      GO TO 9
8     WRITE (*,*) CHR2
      NERR=NERR+ONE
      IF (NERR.LE.TEN) GO TO 7
      GO TO 10
      ELSE
      WRITE (*,*) 'CHOICE OF THE MIXING DISTRIBUTION IS WRONG.'
      STOP
      END IF
      END IF
9     IF (NDIST.EQ.ONE) THEN
      MIXPOS=U(A,B,INIT)
      ELSE IF (NDIST.EQ.TWO) THEN
      MIXPOS=D(K,X,P,INIT)
      ELSE
      MIXPOS=GAMMA(ALPHA,LAMDA,INIT)
      END IF
      RETURN
10    STOP
     $'THE PROGRAM IS TERMINATING DUE TO MORE THAN TEN I/O ERRORS.'
      END
```

```
      REAL*8 FUNCTION DSUM(A,N)
      REAL*8 A(N)
      DSUM=0.0D0
      DO 1 I=1,N
1     DSUM=DSUM+A(I)
      RETURN
      END

      REAL*8 FUNCTION DPRD(A,N)
      REAL*8 A(N)
      DPRD=1.0D0
      DO 1 I=1,N
1     DPRD=DPRD*A(I)
      RETURN
      END

      REAL*8 FUNCTION DINPRD(A,B,N)
      REAL*8 A(N),B(N)
      DINPRD=0.0D0
      DO 1 I=1,N
1     DINPRD=DINPRD+A(I)*B(I)
      RETURN
      END

      REAL*8 FUNCTION DSQNRM(A,N)
      REAL*8 A(N)
      DSQNRM=DINPRD(A,A,N)
      RETURN
      END

      REAL*8 FUNCTION DSUMLG(A,N)
      REAL*8 A(N)
      DSUMLG=0.0D0
      DO 1 I=1,N
1     DSUMLG=DSUMLG+DLOG(A(I))
      RETURN
      END

      REAL*8 FUNCTION DDSGNDTS(X,Y)
      REAL*8 ZERO,HALF,ONE,TWO
      REAL*8 X,Y,Z,W,WHALF
```

```
PARAMETER (ZERO=0.0D0,HALF=0.5D0,ONE=1.0D0,TWO=2.0D0)
IF ((X.EQ.ZERO).AND.(Y.EQ.ZERO)) THEN
DDSGNDTS=ZERO
RETURN
END IF
Z=DMAX1(DABS(X),DABS(Y))
W=ONE
IF (Z.LT.HALF) THEN
WHALF=HALF
DO WHILE (Z.LT.WHALF)
W=WHALF
WHALF=W*HALF
END DO
ELSE IF (Z.GE.ONE) THEN
DO WHILE (Z.GE.W)
W=W*TWO
END DO
END IF
DDSGNDTS=DABS(X-Y)/W
RETURN
END

SUBROUTINE MAXLHD(N,X,R,LAMDA,P)
INTEGER*2 ONE,TWO,THREE,FOUR
REAL*8 ZERO,HALF
INTEGER R,R1
REAL*8 E,EPS1,EPS2,EPS3,G,PHIOLD,PHITMP,RE,S,T,U,Y,Z
REAL*8 XA,XB,PHINEW
REAL*8 LAMDA(N),P(N),X(N)
REAL*8 F(10000),V(10000),W(10000)
REAL*8 DDSGNDTS,DSUM
COMMON /BLK1/ XA,XB,PHINEW
PARAMETER (ONE=1,TWO=2,THREE=3,FOUR=4)
PARAMETER (ZERO=0.0D0,HALF=0.5D0)
PARAMETER (MMAX=1000000)
PARAMETER (EPS1='0000000000003980'X,EPS2='0000000000003680'X)
PARAMETER (EPS3='0000000000003300'X)
DATA F,V,W /30000*0.0D0/
S=N/DSUM(X,N)
LAMDA(ONE)=S
P(ONE)=ONE
```

```
      PHINEW=N*(DLOG(S)-ONE)
      OPEN (2,STATUS='SCRATCH')
      DO 7 R=TWO,N
      PHIOLD=PHINEW
      R1=R-ONE
      IF (R.GT.TWO) REWIND(2)
      WRITE (2,*) (LAMDA(J),P(J),J=ONE,R1)
      IF (R.EQ.TWO) THEN
      T=XB-S
      S=S-XA
      IF (S.LT.T) THEN
      INDEX=ONE
      U=LAMDA(ONE)
      LAMDA(ONE)=U-HALF*S
      P(ONE)=HALF
      LAMDA(TWO)=U
      P(TWO)=HALF
      ELSE
      INDEX=TWO
      P(ONE)=HALF
      LAMDA(TWO)=LAMDA(ONE)+HALF*T
      P(TWO)=HALF
      END IF
      ELSE
      S=LAMDA(TWO)-LAMDA(ONE)
      INDEX=TWO
      DO 1 J=THREE,R1
      T=LAMDA(J)-LAMDA(J-ONE)
      IF (T.GT.S) THEN
      S=T
      INDEX=J
      END IF
1     CONTINUE
      DO 2 J=R,ONE,-ONE
      IF (J.GT.INDEX) THEN
      J1=J-ONE
      LAMDA(J)=LAMDA(J1)
      P(J)=HALF*P(J1)
      ELSE IF (J.EQ.INDEX) THEN
      LAMDA(INDEX)=LAMDA(INDEX)-HALF*S
```

```
      P(INDEX)=HALF
      ELSE
      P(J)=HALF*P(J)
      END IF
2     CONTINUE
      END IF
      M=O
3     E=ZERO
      RE=ZERO
      IF (M.EQ.O) PHINEW=ZERO
      DO 5 I=ONE,N
      G=ZERO
      S=X(I)
      DO 4 K=ONE,R
      T=LAMDA(K)
      F(K)=T*DEXP(-S*T)
      T=F(K)*P(K)
4     G=G+T
      IF (M.EQ.O) PHINEW=PHINEW+DLOG(G)
      DO 5 J=ONE,R
      T=F(J)/G
      V(J)=V(J)+T
      T=S*T
5     W(J)=W(J)+T
      PHITMP=PHINEW
      PHINEW=ZERO
      DO I=ONE,N
      S=X(I)
      G=ZERO
      DO 6 J=ONE,R
      IF (I.EQ.ONE) THEN
      T=P(J)*V(J)/N
      U=V(J)/W(J)
      Y=P(J)
      Z=LAMDA(J)
      E=DMAX1(E,DABS(Y-T),DABS(Z-U))
      RE=DMAX1(RE,DDSGNDTS(Y,T),DDSGNDTS(Z,U))
      V(J)=ZERO
      W(J)=ZERO
      P(J)=T
```

```
      LAMDA(J)=U
      ELSE
      T=P(J)
      U=LAMDA(J)
      END IF
6     G=G+T*U*DEXP(-S*U)
      PHINEW=PHINEW+DLOG(G)
      END DO
      M=M+ONE
      IF (M.LE.5000) THEN
      IF ((E.GE.EPS1).AND.(RE.GE.EPS2)) GO TO 3
      ELSE IF (M.LE.MMAX) THEN
      IF ((E.GE.EPS1).AND.(RE.GE.EPS2).AND.
     $(DABS(PHINEW-PHITMP).GE.EPS3)) GO TO 3
      END IF
      IF (((PHINEW-PHIOLD).LT.EPS1).OR.(P(INDEX).LT.EPS1)) GO TO 8
7     CONTINUE
8     IF ((PHINEW.LE.PHIOLD).OR.(P(INDEX).LT.EPS1)) THEN
      LAMDA(R)=ZERO
      P(R)=ZERO
      R=R1
      REWIND (2)
      READ (2,*) (LAMDA(J),P(J),J=ONE,R)
      END IF
      CLOSE(2)
      JSHIFT=0
      DO 9 J=ONE,R
      JJ=J+JSHIFT
      IF (JJ.LE.R) THEN
      IF (P(JJ).EQ.ZERO) THEN
      DO WHILE ((P(JJ).EQ.ZERO).AND.(JJ.LE.R))
      JJ=JJ+ONE
      END DO
      JSHIFT=JJ-J
      IF (JJ.LE.R) THEN
      LAMDA(J)=LAMDA(JJ)
      P(J)=P(JJ)
      ELSE
      LAMDA(J)=ZERO
      P(J)=ZERO
```

```
      END IF
      ELSE IF (J.NE.JJ) THEN
      LAMDA(J)=LAMDA(JJ)
      P(J)=P(JJ)
      END IF
      IF (JJ.EQ.R) GO TO 9
      JJ=JJ+ONE
      IF (LAMDA(J).EQ.LAMDA(JJ)) THEN
      S=P(J)
      DO WHILE ((LAMDA(J).EQ.LAMDA(JJ)).AND.(JJ.LE.R))
      S=S+P(JJ)
      JJ=JJ+ONE
      END DO
      JSHIFT=JJ-J-ONE
      P(J)=S
      END IF
      ELSE
      LAMDA(J)=ZERO
      P(J)=ZERO
      END IF
9     CONTINUE
      R=R-JSHIFT
      RETURN
      END
```

References

AMARI, S. and KUMON M. (1988) : Estimation in the presence of infinitely many nuisance parameters — Geometry of estimating functions. *Ann. Statist.*, **16**, 1044-1068.

BEGUN, J. M., HALL, W. J., HUANG, W. M. and WELLNER, J. A. (1983) : Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.*, **11**, 432-452.

BICKEL, P. J. (1975) : One-step Huber estimates in the linear model. *Jour. Amer. Statist. Assoc.*, **70**, 428-434.

BICKEL, P. J. (1982) : On adaptive estimation. *Ann. Statist.*, **10**, 647-671.

BICKEL, P. J. and KLAASSEN, C. A. J. (1986) : Empirical Bayes estimation in functional and structural models, and uniformly adaptive estimation of location. *Adv. Appl. Math.*, **7**, 55-69.

BILLINGSLEY, P. (1979) : *Probability and Measure. First Edition.* John Wiley & Sons, New York.

CHATTERJEE, S. K. and DAS K. (1983) : Estimation of variance component for one-way classification with heteroscedastic error. *Calcutta Statist. Assoc. Bull.*, **32**, 57-78.

DIACONIS, P. and FREEDMAN, D. (1986) : On the consistency of Bayes estimates. *Ann. Statist.*, **14**, 1-26.

FISHMAN, G. S. (1978) : *Principles of Discrete Event Simulation.* Academic Press, New York.

FREEDMAN, D. (1963) : On the asymptotic behavior of Bayes estimate in the discrete case I. *Ann. Math. Statist.*, **34**, 1386-1403.

HANNAN, J. and HUANG, J. S. (1972) : A stability of symmetrization of product measures with few distinct factors. *Ann. Math. Statist.*, **43**, 308-319.

HANNAN, J. F. and ROBBINS, H. (1955) : Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.*, **26**, 37-51.

HASMINSKII, R. Z. and IBRAGIMOV, I. A. (1983) : On asymptotic efficiency in the presence of an infinitedimensional nuisance parameter. *Probability Theory and Mathematical Statistics (Fourth USSR-Japan symposium Proceedings, 1982. Edited by itô and Prohorov) : Lecture Notes in Mathematics*, **1021**, 195-229, Springer Verlag, New York.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981) : *Statistical Estimation. Asymptotic Theory.* Springer Verlag, New York.

JEWELL, N. P. (1982) : Mixtures of exponential distributions. *Ann. Statist.*, **10**, 479-484.

KIEFER, J. and WOLFOWITZ, J. (1956) : Consistency of the maximum likelihood estimators in the presence of infinitely many nuisance parameters. *Ann. Math. Statist.*, **27**, 887-906.

KUMON, M. and AMARI, S. (1984) : Estimation of a structural parameter in the presence of a large number of nuisance parameters. *Biometrica*, **71**, 445-459.

LEVIT B. YA. (1978): Infinite-dimensional informational lower bounds. *Theor. Probab. Appl.*, **20**, 388-394.

LINDSAY, B. G. (1980) : Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators. *Phil. Trans. Roy. Soc. London Ser. A*, **296**, 639-662.

LINDSAY, B. G. (1982) : Conditional score functions : some optimality results. *Biometrica*, **69**, 503-512.

LOÉVE, M. (1963) : *Probability Theory, 3rd Edition.* D. van-Nostard, Princeton.

MAITRA, A. and RAO, B. V. (1975) : Selection Theorems and the reduction principle. *Trans. Amer. Math. Soc.*, **20**, 57-66.

NEYMANN, J. and SCOTT, E. (1948) : Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1-32.

PARTHASARATHY, T. (1972) : *Selection Theorems and Their Applications. Lecture Notes in Mathematics*, **263**, Springer Verlag, New York.

PFANZAGL, J. (1982) (with W. Wefelmeyer) : *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statistics*, **13**, Springer Verlag, New York.

PRAKASA RAO, B. L. S. (1983) : *Nonparametric Functional Estimation.* Academic Press, New York.

ROBBINS, H. (1964) : The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.*, **35**, 1-20.

RUDIN, W. (1974) : *Functional Analysis.* Tata-Mc-Graw-Hill.

SCHICK A. (1986) : On asymptotically efficient estimation in semiparametric models. *Ann. Statist.*, **14**, 1139-1151.

SRIVASTAVA S. M. (1982) : *Lecture Notes in Descriptive Set Theory.* Indian Statistical Institute.

VAN DER VAART, A. W. (1987) : Statistical estimation in large parameter spaces. Ph.D. Thesis, Univ. Leiden.

VAN DER VAART, A. W. (1988) : Estimating a real parameter in a class of semiparametric models. *Ann. Statist.*, **16**, 1450-1474.