Sankhyā : The Indian Journal of Statistics
1992, Volume 54, Series A, Pt. 1, pp. 1–39.

EFFICIENT ESTIMATION WITH MANY NUISANCE PARAMETERS

(Part I)

By J. BHANJA and J. K. GHOSH
Indian Statistical Institute

SUMMARY. Consider the Neyman-Scott problem where we want to estimate the common parameter θ based on the sequence $\{X_n\}_{n \geqslant 1}$, of independent random variables with X_i having the density $f(., \theta, \xi_i)$. Depending on the nature of the sequence $\{\xi_n\}_{n \geqslant 1}$, there are two set-ups, viz., the fixed set up where ξ_i's are treated as unknown constants and the mixed set up or the mixture model where ξ_i's are i.i.d with common distribution G. In this paper, we shall define a criterion for efficiency in the first model in terms of that in the second one and find an efficient estimate in the mixture model. The same problem in the other model will be discussed in part II.

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 $\boldsymbol{R}^p$, 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 $p \geqslant 2$, the maximum partial likelihood estimate based on $X_{ij} - \overline{\boldsymbol{X}}_i$ is efficient.

Example 1.2. This is similar to Example 1.1 except that the components of $\boldsymbol{X}_i$ are 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 $p \geqslant 3$ and $n^{-1} \sum_{i=1}^{n} \sigma_i^2$ is bounded away from zero, but it is not efficient. Bickel and Klaassen (1986) for $p = 1$, Bhanja and Ghosh (1987) and van der Vaart (1987) for general p show how an efficient, asymptotically normal estimate can be constructed.

Paper received. June 1988 ; revised July 1990.

AMS (1985) *subject classification*. 62G05, 62G20.

Key words and phrases. Estimation, semiparametric models, mixture models, fixed set-up models, asymptotic efficiency.

Lindsay (1980) and Bickel and Klaassen (1986) provide an extremely useful general discussion of such problems. See also van der Vaart (1987). (After this paper was submitted much of van der Vaart (1987) has appeared in print, vide van der Vaart (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 $X_i \sim f(., \theta, \xi_i)$ 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 for any (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 verification will appear elsewhere.

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 $\boldsymbol{R}$ 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 $\boldsymbol{F}_n$ denote the empirical distribution function (e.d.f.) or the empirical probability measure based on n values of a random variable, In particular, for n elements $\xi_1, \xi_2, \ldots \xi_n$ from Ξ, denote $\boldsymbol{F}_n(., \xi_1, \xi_2, \ldots, \xi_n)$ by $\underline{G}_n$.

Model I. Let $\{X_i\}$ be a sequence of independent random variables taking values in $(S, \mathcal{S})$ with the distribution of X_i given by P_{θ_0, ξ_i}, $\theta_0 \in \Theta$, $\xi_i \in \Xi$. (The probability measure $P_{\theta, \xi}$ is assumed to be well defined for $\theta \in \bar{\Theta}$, $\xi \in \Xi$). The object is to estimate the so-called structural parameter θ_0.

In Model I, called the fixed set-up model by Bickel and Klaassen, 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 $\prod_1^n P_{\theta_0,\xi_i}$ to this sub-σ-field, $\prod_1^n P_{\theta_0,\xi_i}$ can be replaced by

$\sum_{\substack{\text{set of all} \\ \text{permutations} \\ \pi \text{ of } \{1,2,\ldots,n\}}} \prod_1^n P_{\theta_0,\xi_{\pi(i)}}/n!$ and one would expect, heuristically, on the basis of the analogy between simple random sampling with and without replacement, that Model 1 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_0,G_0} where for $A \in \mathcal{S}$,

$$P_{\theta_0,G_0}(A) := \int P_{\theta_0,\xi}(A)\, d\, G_0(\xi).$$

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, 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 call "regular". The following definition gives the notion of regularity and efficiency considered by them in a suitably modified form that ensures uniformity.

Definition 1.1. (i) *Regularity* : An estimate T_n of θ_0 is called

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

$$\mathcal{L}(\sqrt{n}(T_n - \theta_0)\sigma_T^{-1}(\theta_0, \underline{G}_n) \mid \prod_1^n P_{\theta_0,\xi_i}) \Rightarrow \mathcal{L}(N(0, 1))$$

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

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

$$\mathcal{L}(\sqrt{n}(T_n - \theta_0)\sigma_T^{-1}(\theta_0, G_0) \mid P^n_{\theta_0,G_0}) \Rightarrow \mathcal{L}(N(0, 1))$$

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, 103-104) has pointed out that this formulation of efficiency in Model I is not wholly satisfactory, see in this connection our Remark 4.4.

The estimates introduced by Neyman and Scott (1948) which are analogous to Huber's M-estimates and referred to as C_1-estimates by Kumon and Amari (1984) are regular both in Model 1 and Model II.

These estimates are defined as a solution of

$$\Sigma\, \psi(X_i, \theta) = 0 \qquad \ldots \quad (1.1)$$

with the function ψ satisfying certain regularity conditions. Moro precisely, to ensure uniform asymptotic normality we strengthen the conditions given in Amari and Kumon (1985) as follows :

Definition 1.2. Any Borel-measurable map ψ from $S \times \Theta$ to $\boldsymbol{R}$ is called a C_1-*kernel*, if

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

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

(iii) $\sup\limits_{(\theta,\xi) \in \bar{\Theta} \times \Xi} \int\limits_{\{|\psi(.,\theta)| \geqslant \alpha\}} \psi^2(.,\theta)\, dP_{\theta,\xi} \to 0$ as $\alpha \to \infty$

and

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

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

Given any C_1-kernel ψ, any estimate T_n, which is, simultaneously, $\sqrt{n}$ consistent for θ_0 and a solution of (1.1) with probabiliity tending to one, uniformly on compact subsets of $\Theta \times \Xi$ is called a C_1-*estimate corresponding to* ψ.

For a fixed G_0, according to semiparametric theory, there is a function $\bar{\psi}(.\,,.\,, G_0)$ (vide (2.2)–(2.4)) along with an estimate $T_n\,(G_0)$ which solves

$$\sum_{i=1}^{n} \bar{\psi}\,(X_i, \theta, G_0) = 0 \qquad \ldots \quad (1.2)$$

with probability tending to one and is efficient at (θ_0, G_0), provided certain regularity condition hold. If G_0 is unknown, a natural thing to do is to solve

$$\sum_{i=1}^{n} \bar{\psi}(X_i, \theta, \hat{G}_n) = 0 \qquad \ldots \quad (1.3)$$

where $\hat{G}_n$ is a consistent estimate of G_0.

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 $\hat{G}_n$ is n^{δ} consistent, for a suitable $\delta > 0$ or $\hat{G}_n$ is a consistent estimate independent of the X_i's $\bar{\psi}$ is a "nice" function of (θ, G) and

$$\int \bar{\psi}(., \theta, G)\, d\, P_{\theta, G'} = 0 \;\forall\; (\theta, G, G') \qquad \ldots \quad (1.4)$$

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

For mixture models (1.4) always holds but unfortunately, in general, it is very difficult to prove the existence of an n^{δ} consistent estimate of G_0.

In our original unpublished work done before the publication of Schick (1986), we were able to resolve the problem only for the examples of Section 5 and Section 6, with stronger regularity condition than in the present version.

However, the requirement of n^{δ} consistency of $\hat{G}_n$ can be dropped using the following idea of Bickel (1982) and Schick (1986), who show how, in effect, one can use an idnependent estimate of G_0. Thus instead of (1.3) one solves

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

where G_{n1}, G_{n2} are consistent estimates, G_{n1} is independent of $X_1, \ldots, X_{n_1}$ 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. The equation can be shown to hold quite generally in models satisfying Bickel's condition C (vide Remark 2.3) 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 method, seems implicit in Bickel's construction of adaptive estimates in orthogonal cases and has appeared recently in an explicit form in Schick (1986, 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 Section 3 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).

Section 3 consists of a uniform version of Schick's results under stronger conditions. Under same conditions we get analogous results for Model I, i.e., the original Neyman-Scott formulation, in Section 4. Unlike the earlier treatments, e.g., van der Vaart (1987, 1988), we do not assume that the empirical d.f. of the nuisance parameters converges weakly. It is the relaxation of this assumption which requires that we have a uniform version of of Schick's results. As to the proof, the main new feature is that a preliminary randomization over indices is needed before applying the techniques of Section 3.

In Section 4* we also indicate briefly (vide Remark 4.6) how the results of Section 3 and 4 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).

It must be admitted that our conditions are somewhat ugly. Technically speaking, they ensure uniform continuity or uniform integrability of $f'/f^{1/2}$ or the optimal score function $\bar{\psi}$. 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. 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. These two cases are discussed in Section 5.

It turns out that the conditions of $\bar{\psi}$, 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 allows us to provide a direct application of the results in Section 3 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 X is the observable output.

* to appear in the next issue of *Sankhyā*.

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 observartions 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}. The problem of estimating θ is solved in Section 6.

A preliminary announcement of these results appeared in Bhanja and Ghosh (1987).

In subsequent communications we hope to study the problem of estimation of ξ_i's and to report some simulation studies to evaluate the behaviour of our estimate of θ. The simulation studies, made for Example 1.2 with $p \geqslant 3$, are quite promising and indicate the asymptotics works well for $n = 100$.

Let us summarise the major new contributions in this paper. We provide, a uniform version of Schick's results and then apply them to the original formulation of Neyman-Scott problem where the convergence of empirical distribution of ξ_i's is not assumed. For Example 1.2 we get new asymptotically efficient estimates for all p. Using new techniques, a new solution is proposed for an interesting problem of Hasminskii and Ibragimov.

2. Notations and preliminaries

In this section, we shall introduce some notations and give some preliminary assumptions, definitions and results most of them in a form applicable to any semiparametric family involving (θ, G). See in this connection Remark 3.5 in next section. From Section 3 onwards Θ is usually a compact subset of a Euclidean space and $\mathcal{G}$ is in the space of distribution functions equipped with the weak topology.

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 for any 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,

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

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

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

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

(4) For any function $\phi : \theta \to \boldsymbol{R}$ 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_0}, is a closed subset of the real line. Let $\boldsymbol{X} = \prod_{i=1}^{m} X_i$. Let r_0 be a positive integer. Define $\{s_i\}_{1 \leqslant i \leqslant m}$ by

$$s_i = \begin{cases} r_0 & \text{if } i = i_0 \\ 0 & \text{otherwise} \end{cases}$$

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

Let $C_{s_1, s_2, \ldots, s_m}(\boldsymbol{X})$ be the set of all continuous functions ϕ from $\boldsymbol{X}$ to $\boldsymbol{R}$ such that for any $1 \leqslant j \leqslant r_0$, $\frac{\partial^j}{\partial x_{i_0}^j} \phi$ exists on int $(\boldsymbol{X})$ with a continuous extension on $\boldsymbol{X}$.

Remark 2.1. For the special case where $X_1, \ldots, X_m$ are compact, define $\| \;\|_{s_1, s_2, \ldots, s_m}$ from $C_{s_1, s_2, \ldots, s_m}(\boldsymbol{X})$ to $\boldsymbol{R}^+$ *by*

$$\| \phi \|_{s_1, s_2, \ldots, s_m} := \sum_{j=0}^{r_0} \left\| \frac{\partial}{\partial x_{i_0}^j} \phi \right\|_{sup} \text{ for all } \phi \text{ in } C_{s_1, s_2, \ldots, s_m}(\boldsymbol{X}).$$

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 shall take $\boldsymbol{X}$ to be $\overline{\Theta}$ or $\overline{\Theta} \times \mathcal{G}$ or $S_2 \times \overline{\Theta} \times \mathcal{G}$ or $S_2 \times S_2 \times \overline{\Theta} \times \mathcal{G}$ where S_2 is a compact metric space (vide Model III of Section 6), with obvious choice of i_0 and $r_0 = 1, 2$.

(6) For any probability space $(\Omega, \mathcal{A}, P)$ we shall denote the space of all square integrable functions whose expectations are zero by $L_2^0(P)$. In symbols,

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

Convention If $P << Q$, $L_2(P) \underset{not}{\equiv} L_2\left(\frac{dP}{dQ}\right)$ and $L_2^0(P) \underset{not}{\equiv} L_2^0\left(\frac{dP}{dQ}\right)$.

We shall need the following useful definition.

Definition 2.1. 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* $\phi(\theta_0, \underline{G}_n)$ *in Model I* ($\phi(\theta_0, G_0)$ *in Model II*) if for any compact subset Θ_0 of Θ, for any $\epsilon > 0$ and for any $0 < \delta < 1$, there is $N_0 \geqslant 1$ such that for all $n \geqslant N_0$,

$$\sup_{(\theta_0, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta_0, \xi_i} \right) (\{\rho(T_n, \phi(\theta_0, \underline{G}_n)) > \epsilon\}) < \delta$$

$$\left(\sup_{(\theta_0, G_0) \in \Theta_0 \times \mathcal{G}} P^n_{\theta_0, G_0} (\{\rho(T_n, \phi(\theta_0, G_0)) > \epsilon\}) < \delta \right).$$

As a special case of the above definition, we can define the notions of uniformly consistent estimates of θ_0, $\underline{G}_n$ or $(\theta_0, \underline{G}_n)$ in Model I and θ_0, G_0 or (θ_0, G_0) 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 2.1. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ *be a sequence of independent random vectors in* R^p, *with* $\mathbf{X}_i$ *having the distribution function* F_i, *then, for any* $\epsilon > 0$,

$$\sup_{\{F_k\}_{k=1}^n \in \mathcal{F}^n} P_{F_1, \ldots, F_n} \left(\left\{ \left\| \mathbf{F}_n(\cdot, \mathbf{X}_1, \ldots, \mathbf{X}_n) - \frac{1}{n} \sum_{i=1}^{n} F_i \right\|_{sup} > \epsilon \right\} \right) \to 0 \text{ as } n \to \infty$$

where $\mathcal{F}$ *denotes the set of all (probability) distribution functions on* $\mathbf{R}^p$.

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

As a corollary to Proposition 2.1, we shall now prove, using a method of Robbins (1964), the existence of a uniformly consistent estimate of $(\theta_0, \underline{G}_n)$ in Model I and (θ_0, G_0) in Model II.

For this purpose, we shall need the following identifiability assumption.

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

$$P^n_M := \{P^n_{\theta,G} : (\theta, G) \in \bar{\Theta} \times \mathcal{G}\}$$

$$\text{and } P^n_F := \left\{\prod_1^n P_{\theta,\xi_i} : (\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \bar{\Theta} \times \Xi^n\right\}.$$

Let us now state the corollary.

Corollary 2.1.1. *If (i) $(S, \mathcal{S}) = (\boldsymbol{R}^p, \mathcal{B}(\boldsymbol{R}^p))$ (ii) $P_{\theta,G}$'s are dominated by the Lebesgue measure and (iii) $(\theta, G) \to F(., \theta, G)$ is continuous, where for any $(\theta, G) \in \bar{\Theta} \times \mathcal{G}$, $F(., \theta, G)$ denote the distribution function corresponding to $P_{\theta,G}$ and the topology on $\mathcal{F}$, as considered in Proposition 2.1, 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 uniformly consistent estimate of $(\theta_0, \underline{G}_n)$ in Model I and (θ_0, G_0) in Model II.*

The proof is given in Appendix A.

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_0.

Corollary 2.1.2. *Consider Model I and Model II, as defined in Section 1 with the only exception that Θ is allowed to be unbounded. Assume (A1). If conditions (i)—(iii) of Corollary 2.1.1 hold and there is an estimate T_n of θ_0 which is uniformly consistent in Model I (II), then there is a uniformly consistent estimate $(\hat{\theta}_n, \hat{G}_n)$ of $(\theta_0, \underline{G}_n)$ in Model I ((θ_0, G_0) in Model II).*

The proof is given in Appendix A.

We shall also need the following definitions.

Definition 2.2. Call an estimate T_n of θ_0 a *uniformly $\sqrt{n}$-consistent estimate of θ_0 in Model* I(II) if for any compact subset Θ_0 of Θ the family of laws

$$\left\{\mathcal{L}\left(\sqrt{n}(T_n - \theta_0) \mid \prod_{i=1}^n P_{\theta_0,\xi_i}\right) : (\theta_0, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n, n \geqslant 1\right\}$$

$$(\{\mathcal{L}(\sqrt{n}(T_n - \theta_0) \mid P^n_{\theta_0,G_0}) : (\theta_0, G_0) \in \Theta_0 \times \mathcal{G}, n \geqslant 1\})$$

is tight.

Definition 2.3. Let T_n be an estimate of θ_0 and let Ψ be a Borel measurable map from $S^n \times \Theta$ to R. Consider the equation

$$\Psi((X_1, X_2, \ldots, X_n), \theta) = 0. \quad \ldots \quad (2.1)$$

(a) Call T_n *a* $\sqrt{n}$*-consistent solution of* (2.1) *in Model* I (II) if for any $(\theta_0, \{\xi_i\}_{1\leq i\leq n}) \in \Theta\times\Xi^n$ (for any (θ_0, G_0) in $\Theta\times\mathcal{G}$) the following hold.

(i) $\left(\prod_{i=1}^{n} P_{\theta_0,\xi_i}\right)(T_n \text{ solves } (2.1)) \equiv 1+o(1)$

$(P^n_{\theta_0,G_0}, (T_n \text{ solves } (2.1)) = 1+o(1))$

(ii) T_n is *a* $\sqrt{n}$-consistent estimate of θ_0 in Model I(II)

and (b) call T_n a *uniformly* $\sqrt{n}$ *consistent solution of* (2.1) *in Model* I (II) if for any compact subset Θ_0 of Θ, condition (a) holds uniformly on $\Theta_0\times\Xi^n$ $(\Theta_0\times\mathcal{G})$.

Definition 2.4. Call an estimate T_n of θ_0 *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 $\boldsymbol{R}$, if

$$\sup_{(\theta_0,\{\xi_i\}_{1\leq i\leq n})\in\Theta_0\times\Xi^n}\left|\sup_{x\in\boldsymbol{R}}\left|\left(\prod_{i=1}^{n} P_{\theta_0,\xi_i}\right)(\{\sqrt{n}(T_n-\theta_0)\leqslant x\})-\Phi(x\,\sigma_T^{-1}(\theta_0, \underline{G}_n))\right|\right.$$

$$\sup_{(\theta_0,G)\in\Theta_0\times\mathcal{G}}\ \sup_{x\in\boldsymbol{R}} |P^n_{\theta_0,G_0}(\{\sqrt{n}(T_n-\theta_0)\leqslant x\})-\Phi(x\,\sigma_T^{-1}(\theta_0,G_0))| \to 0$$

as $n\to\infty$, for any compact subset Θ_0 of Θ.

We note the following.

(I) As expected, for any concept defined through Definitions 2.1—2.4, the Model I-version is stronger than the Model II-version.

Let us now state a generalized version of the Lindeberg-Lévy central limit theorem where the convergence is uniform in sup-norm. We shall need this result in the proof of our basic result Lemma 3.1 in Appendix B.

Proposition 2.2. *Let A be a non-empty set. For each α in A, let $\{X_n(\alpha)\}_{n\geqslant 1}$ be a sequence of independent random variables with mean zero and finite variance. For each α in A, for each $n\geqslant 1$, define $S_n(\alpha)$ and $s_n(\alpha)$ by*

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

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

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

and

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

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

then $\sup_{\alpha \in A} \sup_{x \in \mathbf{R}} |F_n(x, \alpha) - \Phi(x)| \to 0$ *as* $n \to \infty$

The proof is given in Appendix A.

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

Definition 2.5. We shall call a function $\psi : S \times \overline{\Theta} \times \mathcal{G} \to \mathbf{R}$ a *kernel* if $\psi(\cdot, \theta, G) \in L_2^0(P_{\theta, G})$ for all (θ, G) in $\overline{\Theta} \times \mathcal{G}$, and denote the set of all kernels by $\mathcal{K}$.

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

$$P_{\theta, G} << \mu \;\forall\; (\theta, G) \in \overline{\Theta \times \mathcal{G}}$$

(III) In available semiparametric literatures, (A2) is always assumed. So, we shall assume it for the remaining part of this section and the next three sections. However, this condition will be dropped in Section 6.

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

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

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

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

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

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

$$s_\theta(x, \theta, G) := \frac{f'(x, \theta, G)}{f(x, \theta, G)} 1_{\{f(\cdot, \theta, G) > 0\}}(x) \qquad \ldots \quad (2.2)$$

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

Under the following assumption s_θ is well-defined and belongs to $\mathcal{K}$.

(A3) (a) For each $(x, G) \in S \times \mathcal{G}, f(x, ., G) \in C_1(\bar{\Theta})$,

(b) $\int \frac{(f')^2(., \theta, G)}{f(., \theta, G)} d\mu(.) < \infty \;\forall\; (\theta, G) \in \bar{\Theta} \times \mathcal{G}$.

In passing, we remark that, (A3) will be assumed to hold throughout the remaining portion of this paper.

Let us now observe that f has an obvious linear extension on $\bar{\Theta} \times \mathcal{M}$, where $\mathcal{M}$:= 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 defined by

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

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

Consider $\mathcal{M}_0 := \{M \in \mathcal{M} \mid M(\Xi) = 0\}$,

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

$$\mathcal{M}_{\theta,G} = \{M \in \mathcal{M}_0 : \Lambda(., \theta, G, \theta, M) \in L_2^0(P_{\theta,G}) \text{ and } \int 1_{\{f(., \theta, G) = 0\}} f(., \theta, M)\, d\mu(.) = 0\}$$

and

$$\mathcal{N}_{\theta,G} := \{\phi \in L_2^0(P_{\theta,G}) : \exists\, M \in \mathcal{M}_{\theta,G} \text{ such that } \phi = \Lambda(., \theta, G, \theta, M) \text{ a.e. } [P_{\theta,G}]\}. \qquad \ldots \quad (2.3)$$

The elements of the space $\mathcal{N}_{\theta,G}$ may be thought of as the 'directional scores' with respect to small variations in G. However by no means it is the set of all directional scores with respect to G.

Remark 2.2. Under assumptions (A2)—(A3), for each $(\theta, G) \in \Theta \times \mathcal{G}$, the closed linear subspace of $L_2^0(f(., \theta, G))$ obtained by taking the closure of the linear span of $s_\theta(., \theta, G)$ and $N_{\theta,G}$ gives our tangent space $\mathcal{T}_{\theta,G}$ 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 Klaasssen (1986) or van der Vaart. (1987)

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

$$\left.\begin{aligned} \bar{\psi}(.\,,\theta, G) &:= \text{Proj}_{N^{\perp}_{\theta,G}}\{s_\theta(.\,,\theta, G)\} \\ I(\theta, G) &:= \|\bar{\psi}(.\,,\theta, G)\|^2_{L_2(f(.\,,\theta, G))} \end{aligned}\right\} \forall\, (\theta, G) \qquad \ldots \quad (2.4)$$

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 conditions in the form of two assumptions.

(GA4) $\int \dfrac{f^2(.\,,\theta, M)}{f(.\,,\theta, G)}\, d\mu(.) < \infty$ for all $(\theta, G, M) \in \bar{\Theta} \times \mathcal{G} \times \mathcal{M}_0$

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

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

Lemma 2.1. *Consider* (*a*) *an arbitrary semiparametric model where* (A2), (A3) *and* (GA4) *hold or* (*b*) *a mixture model where* (A2)—(A4) *hold. In either case,*

$$\int \bar{\psi}(.\,,\theta, G) f(.\,,\theta, G')\, d\mu(.) = 0 \;\forall\, (\theta, G, G') \qquad \ldots \quad (2.5)$$

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

$$\int \bar{\psi}(.\,,\theta, G) f(.\,,\theta, G)\, d\mu(.) = 0 \;\forall\, (\theta, G) \qquad \ldots \quad (2.6)$$

So, it remains to show

$$\int \bar{\psi}(.\,,\theta, G) f(.\,,\theta, G'-G)\, d\mu(.) = 0 \;\forall\, (\theta, G, G') \qquad \ldots \quad (2.7)$$

For the general semiparametric models use (GA4) to conclude that $\Lambda(.\,,\theta, G, \theta, G'-G) \in N_{\theta,G} \;\forall\, (\theta, G, G')$. Then (2.7) follows from that fact that $\bar{\psi}(.\,,\theta, G) \in N^{\perp}_{\theta,G} \;\forall\, (\theta, G)$.

For the mixture models, let us observe that, for any (θ, G, ϕ) with $(\theta, G) \in \Theta \times \mathcal{G}$ and $\phi \in L_2^0(G)$, $\Lambda(.\,, \theta, G, \theta, \phi\, d\, G) \in N_{\theta, G}$ proving

$$\int \bar{\psi}(.\,, \theta, G) f(.\,, \theta, \phi\, d\, G)\, d\,\mu(.) = 0 \qquad \ldots \quad (2.8)$$

for all (θ, G, ϕ) with $(\theta, G) \in \bar{\Theta} \times \mathcal{G}$ and $\phi \in L_2^0(G)$.

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 $\mathcal{G}_\nu$, i.e.

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

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

Case I. $G, G, \in \mathcal{G}_\nu$. Let g, g' be versions of $\dfrac{dG}{d\nu}, \dfrac{dG'}{d\nu}$ which belong to $\boldsymbol{g}_\nu$. Put $\phi = \dfrac{g'-g}{g}$ in (2.8). By an easy algebra one can show that $f(.\,, \theta, \phi\, dG) = f(.\,, \theta, G'-G)$, so that (2.7) holds for the given point (θ, G, G').

Case II. G, G' arbitrary. Let g, g' be any two versions of $\dfrac{dG}{d\nu}$ and $\dfrac{dG'}{d\nu}$, respectively. One can get two sequences $\{g_n\}_{n \geqslant 1}$ and $\{g_n'\}_{n \geqslant 1}$ of functionsi n $\boldsymbol{g}_\nu$ 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 1, (2.7) holds for (θ, G_n, G_n'), for any $n \geqslant 1$. Hence by assumption (A4), (2.7) holds for (θ, G, G').

Remark 2.3. Note that Lemma 2.1 (a) holds for general semiparametric models satisfying Bickel's condition C, i.e. models with the space $\mathcal{G}$ of nuisance parameters convex and for any $x, \theta \in S \times \bar{\Theta}, f(x, \theta, .)$ an affine function, 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 sections, we recall Definition 2.5 and introduce the following notations.

(8) Let $\mathcal{K}^* := \{\psi \in \mathcal{K} : P_{\theta, G}(\{|\psi(.\,, \theta, G)| > 0\}) > 0\ \forall (\theta, G)\}$.

We shall denote by J the function from $\Theta \times G \times \mathcal{K}^*$ to $\boldsymbol{R}$ defined by,

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

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

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

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

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

Note that

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

(10) We shall denote the Prohorov metric on $\mathcal{G}$ by d. In other words, the metric d is defined as follows :

Let ρ denote the metric on Ξ. For any $\epsilon > 0$, for any $A \subsetneq \Xi$, let A_ϵ denote the set $\{\xi \in A : \rho(\xi, A) < \epsilon\}$. We can now define d by the formula $d(G_1, G_2) = \inf \{\epsilon > \theta : G_1(A) \leqslant G_2(A^\epsilon) + \epsilon$ and $G_2(A) \leqslant G_1(A^\epsilon) + \epsilon$, for all

A in $\mathcal{B}(\Xi)\}$ for all $G_1, G_2 \in \mathcal{G}$.

Later we shall need an estimate of 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 $n \geqslant 1$, let ϕ_n be a measurable map from $(A, \mathcal{A})^n$ to $(B, \mathcal{B})$. For each $n \geqslant 1$, we shall define two more measurable maps from $(A, \mathcal{A})^n$ to $(B, \mathcal{B})$ by the relation

$$\phi_n^O(\{a_i\}_{1 \leqslant i \leqslant n}) = \phi_{(n-[n/2])}(\{a_i\}_{1 \leqslant i \leqslant n,}\ i \text{ odd})$$

and

$$\phi_n^E(\{a_i\}_{1 \leqslant i \leqslant n}) = \phi_{[n/2]}(\{a_i\}_{1 \leqslant i \leqslant n,}\ i \text{ even}),$$

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

3. Mixture models

In this section, we shall state one auxiliary result and prove two main results in the mixture model. The auxiliary result will give conditions on the density function f and a kernel ψ (vide Definition 2.5) so that there exists an estimate $T_n(\psi)$ of θ_0, which is a uniformly $\sqrt{n}$-consistent solution (II) of

$$\sum_{\substack{i \text{ odd} \\ 1 \leqslant i \leqslant n}} \psi(X_i, \theta, \hat{G}_n^E) + \sum_{\substack{i \text{ even} \\ 1 \leqslant i \leqslant n}} \psi(X_i, \theta, \hat{G}_n^O) = 0 \qquad \ldots \quad (3.1)$$

(vide Definition 2.3), where $\hat{G}_n$ is a uniformly consistent (II) estimate of G_0 (vide Definition 2.1) and $\hat{G}_n^E$ and $\hat{G}_n^O$ are obtained from $\hat{G}_n$ using even and odd numbered observations, respectively (formal definition is given in (11) of Section 2). Further conditions on ψ, guaranteeing uniform asymptotic normality (II) (vide Definition 2.4) of such estimate $T_n(\psi)$'s are also given.

The two main results will prove optimality of, respectively, Schick's and our estimate under the assumption that a simpler version of the coditions mentioned in the last paragraph hold for f and the optimal kernel $\bar{\psi}$ (vide relations (2.2)—(2.4)).

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_0 (vide Definition 2.2) and (b) there is a uniformly consistent (II) estimate $\hat{G}_n$ of G_0 (vide Definition 2.1).

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

Definition 3.1. For any kernel ψ, we shall define the estimate $T_n(\psi)$ as a solution of (3.1) which is nearest to U_n, if there is a solution of (3.1) 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_0, G_0) in $\Theta \times \mathcal{G}$. Define a stochastic process D_n indexed by θ as follows.

$$D_n(\theta) = \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i\ odd}}^{n} \{\psi(X_i, \theta, \hat{G}_n^E) - \psi(X_i, \theta_0, G_0) + (\theta-\theta_0)\int \psi(.,\theta_0, G_0) f'(.,\theta_0, G_0) d\mu(.)\}$$

$$+ \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i\ even}}^{n} \{\psi(X_i, \theta, \hat{G}_n^O) - \psi(X_i, \theta_0, G_0) + (\theta-\theta_0) \int \psi(.,\theta_0, G_0) f'(., \theta_0, G_0) d\mu(.)\} \quad \ldots \quad (3.2)$$

for all θ in Θ.

Consider the following conditions :

(i) $$\int \left\{ \frac{\Lambda(.\,, \theta_0, G_0, \theta, G_0) - 1}{(\theta-\theta_0)} - s_\theta(., \theta_0, G_0) \right\}^2 f(.\,, \theta_0, G_0) d\mu(.) \to 0 \text{ as } \theta \to \theta_0$$

where s_θ is the kernel defined by equation (2.2).

(ii) There is $\delta^{(1)}_{\theta_0, G_0} > 0$ such that

(a) $\int \psi^2(., \theta, G) f(., \theta_0, G_0)\, d\mu(.) < \infty \ \forall (\theta, G) \in B(\theta_0, \delta^{(1)}_{\theta_0, G_0}) \times B(G_0, \delta^{(1)}_{\theta_0, G_0})$

and (b) $\lim_{(\theta, G) \to (\theta_0, G_0)} \int \{\psi(.\,, \theta, G) - \psi(.\,, \theta_0, G_0)\}^2 f(., \theta_0, G_0)\, d\mu(.) = 0.$

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

$$\sup_{\{\theta : |\theta-\theta_0| < c/\sqrt{n}\}} P^n_{\theta_0, G_0}(\{|\sqrt{n}\int \psi(., \theta, \hat{G}_n) f(., \theta, G_0)\, d\mu(.)| > \varepsilon\}) \to 0 \text{ as } n \to \infty.$$

(iv) (a) There is $\delta^{(2)}_{\theta_0,G_0} > 0$, such that for all x in S and G in $B(G_0, \delta^{(2)}_{\theta_0,G_0})$,

$\psi(x, . , G) \epsilon C\,(B(\theta_0, \delta^{(2)}_{\theta_0,G_0}))$,

(b) $\int \psi^2 (. , \theta_0, G_0) f(. ,\theta_0, G_0)\, d\mu(.) < \infty$.

(This condition follows from (ii) (a) but is given separately for ease in later references.)

(c) $\int \psi (. , \theta_0, G_0) f'(. , \theta_0, G_0)\, d\mu(.) \neq 0$.

(v) There is $\delta^{(3)}_{\theta_0,G_0} > 0$ and $A(. , \theta_0, G_0) \epsilon L_1(f(. , \theta_0, G_0))$ such that

$$|\psi (. , \theta', G)-\psi(. , \theta, G)| \leqslant |\theta'-\theta|\, A(. , \theta_0, G_0)$$

for all θ, θ' in $B\,(\theta_0, \delta^{(3)}_{\theta_0,G_0})$ and G in $B(G_0, \delta^{(3)}_{\theta_0,G_0})$.

Clearly, one can, without loss of generality, assume

$$\delta^{(1)}_{\theta_0,G_0} = \delta^{(2)}_{\theta_0,G_0} = \delta^{(3)}_{\theta_0,G_0} = \delta_0 \text{ (say)}.$$

Let δ_0 be as above. For any condition C among (i)—(v), let UC denote the condition that C, with $\theta_0, \theta, \theta'$ replaced by $\theta, \theta', \theta''$ and G_0, G replaced by G, G', holds uniformly with respect to $\theta, \theta', \theta''$ in $B(\theta_0, \delta_0)$ and G, G' in $B(G_0, \delta_0)$.

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

U(vi)(a) $$\sup_{(\theta,G)\epsilon B(\theta_0,\delta_0)\times B(G_0,\delta_0)} [\{\int 1_{\{\psi^2(.,\theta,G) \geqslant K\}} \psi^2(., \theta, G) f(.,\theta, G) d\mu(.)\}/J(\theta, G, \psi)] \to 0 \text{ as } K \to \infty$$

and (b) $(\theta, G) \to J(\theta, G, \psi)$ is continuous, where J is the function defined in (8) of Section 2.

Note that, because of compactness of $\mathcal{G}$, one can without loss of generality assume that the number δ_0 considered in U(i)—(vi) depends only on θ_0.

We can now state the auxiliary result.

Lemma 3.1. *Assume* (B1). *Fix* (θ_0, G_0) *in* $\Theta \times \mathcal{G}$. *Let* ψ *be a kernel. Let* D_n *be as defined in the relation* (3.2). *Also, whenever it makes sense, let* $T_n(\psi)$ *be the estimate defined in Definition* 3.1. *We can draw the following conclusions.*

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

$$\sup_{\{\theta : |\theta_0| \leqslant c/\sqrt{n}\}} P^n_{\theta_0,G_0}(\{|D_n(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty.$$

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

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

$P^n_{\theta_0,G_0}$ (*{There is a solution of* (3.1) *lying inside the interval* $(\theta_0 - c_n/\sqrt{n}$, $\theta_0 + c_n/\sqrt{n})\}) \to 1$ *as* $n \to \infty$

and (B) *under assumption* (B1) (*a*), $T_n(\psi)$ *is a* $\sqrt{n}$*-consistent solution* (*II*) of (3.1)

(III) *If conditions (i)—(v) hold then*

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

$$P^n_{\theta_0,G_0},(\{\sup_{\{\theta : |\theta - \theta_0| \leqslant c/\sqrt{n}\}} |D_n(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty$$

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

$$\sup_{x \in \boldsymbol{R}} |P^n_{\theta_0,G_0}(\{\sqrt{n}(T_n(\psi) - \theta_0) \leqslant x\}) - \Phi(x/V(\theta_0, G_0, \psi))| \to 0 \text{ as } n \to \infty,$$

where V is the function defined in (9) *of Section* 2.

(IV) *For any conclusion C among* (I)—(III) *let UC denote the conclusion that C holds uniformly with respect to* (Θ_0, G_0) *in compact subsets of* $\Theta \times \mathcal{G}$. *Then* U(1), U(II) *and* U(III) (A) *hold if the relevant conditions among* U(i)—U(v) *hold whereas* U(III) (B) *holds if* U(i)—U(vi) *hold.*

The proof is given in Appendix B.

Remark 3.1. Condition U(iii) is a uniform version of condition (2.8) Schick (1986, p. 1144). For C_1-kernels this condition holds by definition. For the optimal kernel $\bar{\psi}$, in view of Lemma 2.1, this condition holds even for the general semiparametric families satisfying Bickel's Condition C, provided suitable regularity conditions hold (cf. Remark 2.3).

Remark 3.2. Note that for any kernel ψ,

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

under suitable regularity conditions, which, in turn, implies

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

This will help 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 3.2. 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 (3.1) (vide Definition 2.3) will be called a *generalised C_1-estimate in Model II corresponding to* ψ (or in short a *GC(II) estimate*).

In view of Definition 3.2, conclusion U(III) (B) of Lemma 3.1 can be restated as

Lemma 3.1a. *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 $V(.,.,\psi)$.*

Example 3.1. All C_1-kernels are EK(II) and all C_1-estimates corresponding to it are GC_1(II) estimates.

Example 3.2. 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, 1140–1144).

Let $l^* : S \times \theta \times \mathcal{G} \rightarrow \boldsymbol{R}$ and $Q : \theta \times \mathcal{G} \times \mathcal{G} \rightarrow \boldsymbol{R}$ be defined by

$$l^*(x, \theta, G) := \bar{\psi}(x, \theta, G)/I(\theta, G) \text{ for } (x, \theta, G) \epsilon\, S \times \Theta \times \mathcal{G}$$

and $$Q(\theta, G, G') := \int l^*(., \theta, G)f(., \theta, G')d\,\mu(.) \text{ for } (\theta, G, G')\epsilon\, \theta \times \Theta \times \mathcal{G} \quad \ldots (3.3)$$

Consider the estimate

$$Z_n := \bar{U}_n + \frac{1}{n} \sum_{\substack{1 \leq i \leq n \\ i \text{ odd}}} l^*(X_i, \bar{U}_n, \hat{G}_n^E) + \frac{1}{n} \sum_{\substack{1 \leq i \leq n \\ i \text{ even}}} l^*(X_i, \bar{U}_n, \hat{G}_n^O) \ldots \quad (3.4)$$

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

Assume that

(B2) (a) For any x in S, $f(x, \ldots,) \in C_{1,0}(\overline{\Theta} \times \mathcal{G})$

and (b) for any compact subset Θ_0 of Θ, there is $\delta_0 > 0$ such that the family of functions

$$\left\{ \frac{(f')^2(., \theta', G)}{f(. , \theta, G)} : \theta, \theta' \in \Theta_0 \text{ with } |\theta - \theta'| \leqslant \delta_0, G \in \mathcal{G} \right\}$$

is uniformly integrable with respect to μ.

Remark 3.3. If $(S, \mathcal{S}) = (\boldsymbol{R}^p, \mathcal{B}^p)$ and assumption (B2)(a) holds, then, in view of Corollaries 2.1.1. and 2.1.2, one can easily drop assumption (B1)(b) even if Θ is unbounded.

Remark 3.4. Let ψ be a Borel-measurable function from $\boldsymbol{R}^p$ to $\boldsymbol{R}^+$. Let $s_1, s_2, \ldots, s_k$ be k Borel-measurable functions from $\boldsymbol{R}^p$ to $\boldsymbol{R}$. Define

$$\boldsymbol{\Omega} = \left\{ \boldsymbol{\omega} \in \boldsymbol{R}^k : \int \psi(\boldsymbol{x}) \exp \left\{ \sum_{j=1}^{k} s_j(\boldsymbol{x})\omega_j \right\} d\boldsymbol{x} < \infty \right\}.$$

Assume that

(a) $\boldsymbol{\Omega} \neq \phi$.

Consider the exponential family of densities defined by

$$h(\boldsymbol{x}, \boldsymbol{\omega}) = (d_0(\boldsymbol{\omega}))^{-1} \psi(\boldsymbol{x}) \exp \left\{ \sum_{j=1}^{k} s_j(\boldsymbol{x})\, \omega_j \right\}. \quad \forall\, \boldsymbol{\omega}.$$

for all $\boldsymbol{x}$ in $\boldsymbol{R}^p$ and $\boldsymbol{\omega}$ in $\boldsymbol{\Omega}$, where the function d_0 is given by the formula

$$d_0(\boldsymbol{\omega}) = \int \psi(\boldsymbol{x}) \exp \left\{ \sum_{j=1}^{k} s_j(\boldsymbol{x})\omega_j \right\} d\boldsymbol{x} \quad \forall\, \boldsymbol{\omega}.$$

Consider the family of marginal distributions of $\mathbf{s}$

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

Assume that

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

(c) There is 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 $C_{2,0}(\overline{\Theta} \times \Xi)$.

Assume that

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

(e) Range of $\boldsymbol{\pi}$ is contained in the interior of $\boldsymbol{\Omega}$.

Finally, let us assume that

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

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

for all $\boldsymbol{x}$ in S, θ in Θ and ξ in Ξ, where d stands for the function $d_0 \circ \pi$.

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 now 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

$$\{\int h(., \boldsymbol{\omega})\, d\, H(\boldsymbol{\omega}) : H \in \mathcal{H}\}$$

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

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

$$A_i(\boldsymbol{s}) := \int \exp\left\{\sum_{j=1}^{k} \omega_j s_j\right\} d_0\,(\boldsymbol{\omega}))^{-1}\, dH_i(\boldsymbol{\omega}) < \infty$$

for almost all $\boldsymbol{s}$ and hence, by assumption (c), for all $\boldsymbol{s}$ in J.

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

Therefore

$$(d_0(\boldsymbol{\omega}))^{-1}\, dH_1\,(\boldsymbol{\omega}) = (d_0(\boldsymbol{\omega}))^{-1}\, dH_2(\boldsymbol{\omega})$$

by a well known result on moment generating functions. Hence by continuity of d_0 and choice of $\mathcal{H}$, $H_1 = H_2$. (At this stage, note that in the case of Lindsay (1980), to be discussed in Section 5(b), we don't need the identifiability of G so that one can easily replace assumptions (b) and (c) by

(b) The family $\{h(., \boldsymbol{\omega}) : \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$ is identifiable.)

Next, observe that, assumptions (b)—(c) imply that the family $\{Q_{\boldsymbol{\omega}} : \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$ of probability measures is identifiable. The assertion follows by Theorem 10.0.3 of Prakasa Rao (1983, p. 440) and the definition of $Q_{\boldsymbol{\omega}}$'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, $\bar{\psi}(x, ., .) \in \bar{C}(\Theta \times \mathcal{G})$

and (b) for any compact subset Θ_0 of Θ the following statements hold

(i) there is $\delta_0 > 0$ such that the family of functions

$\{\bar{\psi}^2(., \theta', G')f(., \theta, G) : (\theta', G') \in \Theta_0 \times \mathcal{G}$ with $|\theta - \theta'| + d(G, G') \leqslant \delta_0\}$

is uniformly integrable with respect to μ and

(ii) $$\sup_{(\theta,G)\in\Theta_0\times\mathcal{G}} \left[\left\{\int_{\{\bar{\psi}^2(., \theta, G) > K\}} \bar{\psi}^2(., \theta, G) f(., \theta, G) d\mu(.)\right\} / I(\theta, G)\right]$$

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

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 $(\theta, G) \to I(\theta, G)$ 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 $(\theta, G) \to I(\theta, G)$ is continuous. They also imply a local version of assumption (A4) with $\bar{\Theta}$ and G replaced by B (θ_0, δ_0^*) and $B(G_0, \delta_0^*)$, respectively, where $\delta_0^* = \delta_0/2$.

(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 3.2. *Assume* (B1)—(B3). *The estimate* Z_n *of* θ_0, *as defined through* (3.3)—(3.4), *is UAN (II) with A V* $(1/I)$ (*vide Definition* 2.4).

Proof. Let us start with the following simple observation.

$$\int l^*(., \theta, G)\, f'(., \theta, G)\, d\mu(.) = 1 \;\; \forall (\theta, G). \qquad \ldots \quad (3.5)$$

Next, we shall show that

$$\sup_{x\in\boldsymbol{R}} \left| P^n_{\theta_0, G_0}\left(\left\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^*(X_i, \theta_0, G_0) \leqslant x\right\}\right) - \Phi(x\, I^{1/2}(\theta_0, G_0))\right| \to 0 \qquad \ldots \quad (3.6)$$

as $n \to \infty$, uniformly with respect to (θ_0, G_0) in compact subsets of $\Theta \times \mathcal{G}$.

We shall proceed as follows.

First observe that, $\bar{\psi}$ being a kernel, $l^*(., \theta, G)$ 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 2.2 holds with $X_n(\alpha) = l^*(X_n, \alpha)$, for all α in A and for all $n \geqslant 1$. Hence by Proposition 2.2, L. H. S. of (3.6) goes to zero uniformly with respect to α in A. Since A is arbitrary this proves (3.6).

In view of (3.4)-(3.6), it is enough to show that

$$P^n_{\theta_0, G_0}(\{|D_n(\bar{U}_n)| > \epsilon\}) \to 0 \text{ as } n \to \infty, \qquad \dots \quad (3.7)$$

uniformly with respect to (θ_0, G_0) 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_0, it is enough to show that for any $c > 0$ and $\epsilon > 0$,

$$P^n_{\theta_0, G_0}(\{|D_n(\bar{U}_n)| > \epsilon\} \cap \{\sqrt{n}\,|\bar{U}_n - \theta_0| \leqslant c\}) \to 0 \text{ as } n \to \infty \qquad \dots \quad (3.8)$$

uniformly with respect to (θ_0, G_0) in compact subsets of $\Theta \times \mathcal{G}$.

Now $\sqrt{n}\,|\bar{U}_n - \theta_0| \leqslant c$ if and only if $\sqrt{n}\,\theta_0 - c \leqslant \sqrt{n}\,\bar{U}_n \leqslant \sqrt{n}\,\theta_0 + c$ 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 $\sqrt{n}\,\theta_0 - c \leqslant i \leqslant \sqrt{n}\,\theta_0 + c$ and there can be at most $[2c]+1$ such values. (This is so because given any two real numbers $a < b$, there can at most $[b-a]+1$ integers in $[a, b]$.) Thus (3.8) (and hence (3.7)) holds if U(I) of Lemma 3.1 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 2.1, with $\bar{\Theta} \times \mathcal{G} \times \mathcal{G}$ replaced by $B(\theta_0, \delta_0^*) \times B(G_0, \delta_0^*) \times B(G_0, \delta_0^*)$, one can easily conclude that $Q = 0$ on $B(\theta_0, \delta_0^*) \times B(G_0, \delta_0^*) \times B(G_0, \delta_0^*)$ guaranteeing U(iii).

Remark 3.5. The proof of Theorem 3.2 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, $\bar{\psi}(x, ., .) \in C_{1,0}(\bar{\Theta} \times \mathcal{G})$

(b) for any compact subset Θ_0 of Θ, there is $\delta_0 > 0$ such that

(i) (B3)(b) holds and

(ii) $$\sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} \left[\int \left\{ \sup_{(\theta', G') \in B((\theta, G), \delta_0)} |\bar{\psi}'(., \theta', G'| \, f(., \theta, G) d\mu(.) \right] < \infty.$$

We can now state the final result of this section.

Theorem 3.3. *Assume* (B1), (B2) *and* (B3). *The estimate* $T_n(\bar{\psi})$ *of* θ_0, *as defined in Definition* 3.1, *is UAN(II) with AV*(1/*I*).

Proof. In view of Lemma 3.1a 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 $(\theta, G) \rightarrow I(\theta, G)$. Hence, in view of observation (2), the conditions U(ii)—U(iii) hold for the kernel $\bar{\psi}$ also. So, it remains to check conditions U(iv)—U(vi).

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 $\bar{\psi}$.

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

U(vi) is a consequence of assumption (B3)(b)(ii) and observations (3)—(4).

Remark 3.6. In view of observations (3)—(4), assumptions (B1), (B2) and (B3s) imply l^* is an *EK* (*II*) and for any compact subset Θ_0 of Θ and $\epsilon > 0$

$$\sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} P^n_{\theta_0, G_0} (\{\sqrt{n} \, | Z_n - T_n(l^*) | > \epsilon\}) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Remark 3.7. As indicated in Remark 2.3, all the results stated in this section hold for the general semiparametric families satisfying Bickel's Condition *C* also.

Remark 3.8. In view of Remarks 3.3 and 3.4, for Euclidian 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_0, and assumption (B3) or (B3s), i.e. smoothness properties of the optimal kernel.

Appendix A

Proof of Corollary 2.1.1. One uses an idea implicit in Robbins (1964).

Fix $n \geqslant 1$. Define

$$a_n(\boldsymbol{x}_{pn\times 1}, \theta, G) = \sup_{\boldsymbol{y}\in \boldsymbol{R}^p} |F_n(\boldsymbol{y}, \boldsymbol{x}) - F(\boldsymbol{y}, \theta, G)|$$

$$\text{for } \boldsymbol{x} \in \boldsymbol{R}^{pn}, (\theta, G) \in \bar{\Theta} \times \mathcal{G}.$$

then (i) $a_n : (\boldsymbol{R}^{pn} \times \bar{\Theta} \times \mathcal{G}, \mathcal{B}(\boldsymbol{R}^{pn} \times \bar{\Theta} \times \mathcal{G})) \to (\boldsymbol{R}, \mathcal{B}(\boldsymbol{R}))$ is measurable and (ii) for each $\boldsymbol{x} \in \boldsymbol{R}^{pn}$, $a_n(\boldsymbol{x}, ., .) \in C(\bar{\Theta} \times \mathcal{G})$.

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

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

outside a Lebesgue null set.

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

$$\begin{cases} a_n((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n), \theta_0, \underline{G_n}) \text{ in Model I} \\ a_n((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n), \theta_0, G_0) \text{ in Model II} \end{cases} \quad \ldots \text{(A.1)}$$

outside a Lebesgue-null set.

But, by Proposition 2.1

$$\left.\begin{array}{l} a_n((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n), \theta_0, \underline{G_n}) \xrightarrow{\prod_1^n P_{\theta_0, \xi_i}} 0 \text{ uniformly on } \bar{\Theta} \times \Xi^n \text{ in Model I} \\ \text{and} \\ a_n((\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n), \theta_0, G_0) \xrightarrow{P^n_{\theta_0, G_0}} 0 \text{ uniformly on } \bar{\Theta} \times \mathcal{G} \text{ in Model II} \end{array}\right\} \ldots \text{(A.2)}$$

From (A.1), (A.2) and condition (iii),

$$\|F(\cdot, \hat{\boldsymbol{\theta}}_n, \hat{G}_n) - F(\cdot, \theta_0, \underline{G_n})\|_{sup} \to 0 \text{ uniformly on } \bar{\Theta} \times \Xi^n \text{ in Model I.}$$

and

$$\|F(\cdot, \hat{\boldsymbol{\theta}}_n, \hat{G}_n) - F(\cdot, \theta_0, G_0)\|_{sup} \to 0 \text{ uniformly on } \bar{\Theta} \times \mathcal{G} \text{ in Model II.}$$

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

Remark A.1. Note that the boundedness of Θ is needed only to ensure the continuity of the inverse map $F(., \theta, G) \to (\theta, G)$.

Remark A.2. It is interesting to note that the null set of Corollary 2.1.1 can be dropped in the following manner. First note that the compacntness of $\bar{\Theta} \times \mathcal{G}$ and the continuity of $a_n(\boldsymbol{x}, ., .)$ 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 $k \geqslant 1$ such that $\Theta_k := \Theta \cap [k, k+1] \neq \phi$, we shall use the notation $(\tilde{\theta}_n(k), \tilde{G}_n(k))$ to denote the minimum distance estimates considered in Corollary 2.1.1 for the models

$$P_F^{n,k} := \left\{ \prod_{i=1}^{n} P_{\theta_0, \xi_i} : (\theta_0, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_k \times \Xi^n \right\}$$

and

$$P_M^{n,k} := \left\{ P^n_{\theta_0, G_0} : (\theta_0, G_0) \in \Theta_k \times \mathcal{G} \right\}.$$

Proof of *Corollary* 2.1.2. Let Θ_0 be a given compact subset of Θ. Let $0 < \delta < 1$ and $\epsilon > 0$ be given. We want to show that there is $N \geqslant 1$ such that for all $n \geqslant N$,

$$\sup_{(\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta, \xi_i} \right) (\{ |\hat{\theta}_n - \theta| + d(\hat{G}_n, \underline{G_n}) > \epsilon \}) < \delta \quad \ldots \quad \text{(A.3)}$$

$$\left(\sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} P^n_{\theta, G} (\{ |\hat{\theta}_n - \theta| + d(\hat{G}_n, G) > \epsilon \}) < \delta \right)$$

where

$$(\hat{\theta}_n, \hat{G}_n) := (\tilde{\theta}_n([T_n]), \tilde{G}_n([T_n])).$$

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

$$\sup_{(\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta, \xi_i} \right) (\{ |T_n - \theta| > \eta \}) < \delta/2 \quad \ldots \quad \text{(A.4)}$$

$$\left(\sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}_n} P^n_{\theta, G} (\{ |T_n - \theta| > \eta \}) < \delta/2 \right)$$

Let us observe that by compactness of Θ_0, there are integers k and l with $l \geqslant 1$, such that $\Theta_0 \subseteq \bigcup_{j=1}^{l} \Theta_{k-j+1}$

Define

$$l_0 = \min \left\{ l \geqslant 1 : \exists\, k \in \mathbf{Z} \ni \Theta_0 \subseteq \bigcup_{j=1}^{l} \Theta_{k+j-1} \right\} \qquad \ldots \quad \text{(A.5)}$$

Then, there is a unique integer k_0 such that $\Theta_0 \subseteq \bigcup_{j=1}^{l_0} \Theta_{k_0+j-1}$.

Using Orollary 2.1.1 choose and fix $N_1 \geqslant 1$ such that for any $n \geqslant N_1$,

$$\sup_{1 < j < l_0} \sup_{(\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta, \xi_i} \right) \left(\{ |\tilde{\theta}_n(k_0-j+1) - \theta| + d(\tilde{G}_n(k_0-j+1), \underline{G_n}) > \epsilon \} \right) < \delta/8 \ \ldots \quad \text{(A.6)}$$

$$\Bigg(\sup_{1 < j < l_0} \sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} P^n_{\theta, G} | (\{ \tilde{\theta}_n(k_0-j+1) - \theta | + d(\tilde{G}_n(k_0-j+1), G) > \epsilon \}) < \delta/8 \Bigg)$$

Let $N = N_0 \vee N_1$. Then, for all $n \geqslant N$,
L.H.S. of (A.3)

$$= \sup_{(\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta, \xi_i} \right) (\{ |\hat{\theta}_n - \theta| + d(\hat{G}_n, G) > \epsilon \})$$

$$\leqslant \sup_{(\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta, \xi_i} \right) (\{ T_n \notin ([\theta]-1-\eta, [\theta]+1+\eta) \})$$

$$+ \sum_{j=0}^{3} \sup_{(\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta, \xi_i} \right)$$

$$\times (\{ |\tilde{\theta}_n(\theta]-2+j) - \theta| + d(\tilde{G}_n([\theta]-2+j), G) > \epsilon \})$$

$$\leqslant \sup_{(\theta, \{\xi_i\}_{1 \leqslant i \leqslant n}) \in \Theta_0 \times \Xi^n} \left(\prod_{i=1}^{n} P_{\theta, \xi_i} \right) (\{ |T_n - \theta| > \eta \}) + 4.\delta/8$$

$$\Bigg(= \sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} P^n_{\theta, G} (\{ |\hat{\theta}_n - \theta| + d(\hat{G}_n, G) > \epsilon \})$$

$$\leqslant \sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} P^n_{\theta, G} (\{ T_n \notin ([\theta]-1-\eta, [\theta]+1+\eta) \})$$

$$+ \sum_{j=0}^{3} \sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} P^n_{\theta, G} (\{ |\tilde{\theta}_n([\theta]-2+j) - \theta| + d(\tilde{G}_n([\theta]-2+j), G) > \epsilon \})$$

$$\leqslant \sup_{(\theta, G)\epsilon\Theta_0 \times \mathcal{G}} P^n_{\theta, G}(\{|T_n-\theta| > \eta\})+4.\delta/8) \text{ by (A.5) and (A.6))}$$

$$\leqslant \delta/2+\delta/2 = \delta \text{ by (A.6) proving (A.3).}$$

Remark A.3. In view of Remark A.2, we can drop condition (ii) from Corollaries 2.1.1 and 2.1.2.

For the proof of Proposition 2.2, we shall need the following two auxiliary results

Lemma A.1. *Let A be a nonempty set. Consider the following two families of probability measures on* $(\boldsymbol{R}, \mathcal{B}(\boldsymbol{R}))$.

$$\mathcal{P}_\infty : = \{P_n(., \alpha) : \alpha \epsilon A, n \geqslant 1\}$$

and

$$\mathcal{P} : = \{P(., \alpha) : \alpha A\}.$$

Assume that the following conditons hold.

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

(ii) $\mathcal{P}$ *is tight as well as unifomly absolutely continuous with respect to the Lebesgue measure and*

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

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

Then,

$$\sup_{\alpha \epsilon A} \sup_{x \epsilon \boldsymbol{R}} |F_n(x, \alpha)-F(x, \alpha)| \to 0 \text{ as } n\to\infty. \quad \ldots \quad (A.7)$$

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

Proof. Let us first show that for any x in $\boldsymbol{R}$.

$$\sup_{\alpha \epsilon A} |F_n(x, \alpha)-F(x, \alpha)| \to 0 \text{ as } n\to\infty \quad \ldots \quad (A.8)$$

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

$$\sup_{\alpha \epsilon A} |F(x+\delta, \alpha)-F(x-\delta, \alpha)| < \epsilon/4 \quad \ldots \quad (A.9)$$

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

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

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

$$\sup_{\alpha \in A} \mid \int g(.)dP_n(., \alpha) - \int g(.)dP(., \alpha) \mid < \epsilon/4 \quad \ldots \quad (A.10)$$

Therefore, for $n \geqslant n_1$,

$$\text{LHS of (A.8)} = \sup_{\alpha \in A} \mid F_n(x, \alpha) - F(x, \alpha) \mid$$

$$= \sup_{\alpha \in A} \mid \int 1_{(-\infty, x]} dP_n(., \alpha) - \int 1_{(-\alpha, x]} dP(., \alpha) \mid$$

$$= \sup_{\alpha \in A} \mid \int \{1_{(-\infty, x]}(\cdot) - g(\cdot)\} dP(., \alpha)$$

$$- \int \{1_{(-\infty, x]}(\cdot) - g(\cdot)\} dP_n(., \alpha)$$

$$+ \int g(\cdot) dP_n(., \alpha) - \int g(\cdot) dP(. \alpha) \mid$$

(A. 10)

$$\leqslant \sup_{\alpha \in A} [\int \mid 1_{(-\infty, x]}(\cdot) - g(\cdot) \mid dP_n(., \alpha)]$$

$$+ \sup_{\alpha \in A} [\int \mid 1_{(-\infty, x]}(\cdot) g(\cdot) \mid dP(., \alpha)] + \frac{\epsilon}{4}$$

(A.9)

$$\leqslant \sup_{\alpha \in A} [\int h(., \alpha) dP_n(., \alpha)] + \frac{\epsilon}{4} + \frac{\epsilon}{4} \quad \ldots \quad (A.11)$$

here

$$h(y) = \begin{cases} 0 & \text{if } \mid y - x \mid \geqslant \delta \\ \dfrac{\delta - \mid y - x \mid}{2\delta} & \text{otherwise} \end{cases}$$

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

$$\sup_{\alpha \in A} \mid \int h(.)dP_n(., \alpha) - \int h(.)dP(., \alpha) \mid < \epsilon/4 \quad \ldots \quad (A.12)$$

Let $n_0 = n_1 \vee n_2$. Then for any $n \geqslant n_0$

$$\text{LHS of (A.8) and (A.11)} < \sup_{\alpha \in A} \mid \int h(\cdot) dP_n(\cdot, \alpha) - \int h(\cdot) dP(.), (\alpha) \mid$$

$$+ \sup_{\alpha \in A} [\int h(\cdot) dP(., \alpha)] + \epsilon/2$$

(A.9) and (A.12) $< \epsilon/4 + \epsilon/4 + \epsilon/2 = \epsilon$ proving (A.8).

Let us now prove relation (A.7) from (A.8).

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

$$\sup_{n \geqslant 1} \sup_{\alpha \in A} F_n(-K, \alpha) < \epsilon/4 \sup_{\alpha \in A} F(-K, \alpha) < \epsilon/4, \quad \ldots \quad (A.13)$$

$$\left(\inf_{n \geqslant 1} \inf_{\alpha \in A} F_n(K, \alpha) > 1-\epsilon/4 \text{ and } \inf_{\alpha \in A} F(K, \alpha) > 1-\epsilon/4\right)$$

Then,

$$\sup_{\alpha \in A} \sup_{|x| \geqslant K} |F_n(x \ \alpha)-F(x \ \alpha)|$$

$$\leqslant \sup_{\alpha \in A} \sup_{x \leqslant -K} |F(x, \alpha)-F_n(x, \alpha)| + \sup_{\alpha \in A} \sup_{x \geqslant K} |F_n(x, \alpha)-F(x, \alpha)|$$

$$\leqslant \max \left\{\sup_{\alpha \in A} \sup_{n \geqslant 1} F_n(-K, \alpha), \sup_{\alpha \in A}, F(-K, \alpha)\right\}$$

$$+\max \left[\sup_{\alpha \in A} \sup_{n \geqslant 1} \{1-F_n(K, \alpha)\}, \sup_{\alpha \in A} \{1-F(K, \alpha)\}\right]$$

$$< \epsilon/4 + \epsilon/4 = \epsilon/2 \text{ by (A.13)} \quad \ldots \quad (A.14)$$

Using uniform absolute continuity of $\mathscr{P}$ and compactness of $[-K, K]$ choose and fix $m \geqslant 1$ such that

$$\sup_{\alpha \in A} \sup_{i=0, 1, \ldots, 2m-1} |F(x_{i+1}, \alpha)-F(x_i, \alpha)| < \epsilon/4 \quad \ldots \quad (A.15)$$

where $x_i = -K + \dfrac{iK}{m} = \left(\dfrac{i-m}{m}\right) K$ for $i = 0, 1, \ldots, 2m$.

Using (A.8) choose and fix $N \geqslant 1$ such that $n \geqslant N$ implies

$$\sup_{\alpha \in A} \sup_{i=0, 1, 2, \ldots, 2m} |F_n(x_i \ \alpha)-F(x_i \ \alpha)| < \epsilon/4 \quad \ldots \quad (A.16)$$

Then, for $n \geqslant N$

$$\sup_{\alpha \in A} \sup_{|x| \leqslant K} |F_n(x, \alpha)-F(x, \alpha)|$$

$$< \sup_{\alpha \in A} \sup_{i=0, 1, \ldots, 2m} |F_n(x_i, \alpha)-F(x_i, \alpha)|$$

$$+ \sup_{\alpha \in A} \sup_{i=0, 1, \ldots, 2m-1} |F(x_{i+1} \ \alpha)-F(x_i \ \alpha)|$$

$$< \epsilon/4 + \epsilon/4 = \epsilon/2 \text{ by (A.15) and (A.16).} \quad \ldots \quad (A.17)$$

From (A.14) and (A.17) it follows that, for any $n \geqslant N$,

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

proving (A. 7).

Lemma A.2 (*Theorem* 7 *of Ibragimov and Hasminskii*, **1981, 365**). *Let* A, $\mathscr{P}_\infty$ *and* $\mathscr{P}$ *be as in Lemma* A.1. *Assume that the following conditions hold.*

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

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

Thus, for any bounded continuous function g from $\boldsymbol{R}$ to $\boldsymbol{R}$

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

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

Proof of Proposition 2.2. For any α in A, let $P_n(., \alpha)$ stand for the probability measure corresponding to the distribution function $F_n(., \alpha)$. Then condition (i), which is common to both Lemmas A.1 and A.2, follows from the definition of $F_n(., \alpha)$'s. Next, condition (ii) of Lemma A.2 follows from conditions (i)—(ii) of the proposition and the definition of $F_n(., \alpha)$'s by an application of a uniform version of the proof of Theorem 2.7.2 of Billingsley (1979, 310-312). Again, $\mathscr{P}$ being a singleton containing the standard normal probability measure, condition (iii) of Lemma A.1 holds for it. The proposition follows by an application of Lemma A.2 followed by Lemma A.1.

Appendix B

Proof of Lemma 3.1. (*I*) For $\theta \epsilon \Theta$, define,

$$D_{n1}(\theta) = \frac{1}{\sqrt{n}} \sum_{i \epsilon \left\{1,\, 3,\, \ldots,\, 2\left[\frac{n-1}{2}\right]+1\right\}} \{\psi(X_i, \theta, \hat{G}_n^E) - \psi(X_i, \theta_0, G_0)$$

$$+ (\theta - \theta_0) \int \psi(., \theta_0, G_0) f'(., \theta_0, G_0) d\mu(.)\}$$

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

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

It is enough to show that,

$$\sup_{\{\theta: |\theta-\theta_0| \leqslant c/\sqrt{n}\}} P^n_{\theta_0, G_0}(\{|D_{n1}(\theta)| > \epsilon/2\}) \to 0 \quad \dots \quad \text{(B.1)}$$

and

$$\sup_{\{\theta: |\theta-\theta_0| \leqslant c/\sqrt{n}\}} P^n_{\theta_0, G_0}(\{|D_{n2}(\theta)| > \epsilon/2\}) \to 0.$$

We shall only show that

$$\sup_{\{\theta: |\theta-\theta_0| \leqslant c/\sqrt{n}\}} P^n_{\theta_0, G_0}(\{|D_{n1}(\theta)| > \epsilon/2\}) \to 0. \quad \dots \quad \text{(B.2)}$$

The other statement will follow by a symmetrical argument. Now, for any sequence $\{\theta_n\}$ s.t. $|\theta_n - \theta_0| \leqslant c/\sqrt{n} \;\forall\; n$,

$$E\left(D_{n1}(\theta_n) \,|\, X_2, \dots, X_{2\left[\frac{n}{2}\right]}\right)$$

$$= \frac{\left(n - \left[\frac{n}{2}\right]\right)}{\sqrt{n}} \{\int \psi(x, \theta_n, \hat{G}^E_n) f(x, \theta_0, G_0) d\mu(x) + 0$$

$$+ (\theta_n - \theta_0) \int \psi(x, \theta_0, G_0) f'(x, \theta_0, G_0) d\mu(x)\} \text{ [since } \psi \text{ is a kernel]}$$

$$= -\frac{\left(n - \left[\frac{n}{2}\right]\right)}{\sqrt{n}} (\theta_n - \theta_0)$$

$$\left[\int \{\psi(x, \theta_n, \hat{G}^E_n) - \psi(x, \theta_0, G_0)\} \frac{\{f(x, \theta_n, G_0) - f(x, \theta_0, G_0)\}}{(\theta_n - \theta_0)} d\mu(x)\right.$$

$$\left. + \int \psi(x, \theta_0, G_0) \left\{\frac{f(x, \theta_n, G_0) - f(x, \theta_0, G_0)}{\theta_n - \theta_0} - f'(x, \theta_0, G_0)\right\} d\mu(x)\right]$$

$$+ O_{P^n_{\theta_0, G_0}}(1) \text{ by (iii).}$$

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

$$P^{[n/2]}_{\theta_0, G_0}\left(\left\{\sup_{\{\theta\,:\,|\theta-\theta_0| \leqslant c/\sqrt{n}\}} |E_{\theta_0, G_0}(D_{n1}(\theta) \,|\, X_2\, X_4, \dots, X_{2[n/2]})| > \eta\right\}\right) \to 0. \quad \dots \quad \text{(B.3)}$$

Let us also observe that, for any sequence $\{\theta_n\}$ s.t. $|\theta_n - \theta_0| \leqslant c/\sqrt{n} \;\forall\; n$,

$$\text{Var}_{\theta_0, G_0}\left(D_{n1}(\theta_n) \,|\, X_2, X_4, \dots, X_{2\left[\frac{n}{2}\right]}\right)$$

$$\leqslant \frac{\left(n - \left[\frac{n}{2}\right]\right)}{n} \int \{\psi(x, \theta_n, \hat{G}^E_n) - \psi(x, \theta_0, G_0)\}^2 f(x, \theta_0, G_0) d\mu(x).$$

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

$$P^{[n/2]}_{\theta_0, G_0}\left(\left\{\sup_{\{\theta : |\theta-\theta_0| \leqslant c/\sqrt{n}\}} \mathrm{Var}_{\theta_0, G_0}\left(D_{n1}(\theta) \mid X_2, X_4, \ldots, X_{2\left[\frac{n}{2}\right]}\right) > \eta\right\}\right) \to 0. \qquad \ldots \text{(B.4)}$$

From (B.3) and (B.4), we get, for any $\eta > 0$

$$\sup_{\{\theta : |\theta-\theta_0| \leqslant c/\sqrt{n}\}} P^n_{\theta_0, G_0}(\{|D_{n1}(\theta)| > \eta\} \mid X_2, X_4, \ldots, X_{2[n/2]}) \xrightarrow{P^{[n/2]}_{\theta_0, G_0}} 0. \qquad \ldots \text{(B.5)}$$

Then (B.2) follows by D.C.T. from (B.5) with $\eta = \epsilon/2$.

(II) (A) First observe that, because of (B.1), there is a sequence $\{c_n\}$ of nonnegative real numbers increasing to infinity such that for any $\epsilon > 0$

$$\sup_{\{\theta : |\theta-\theta_0| \leqslant c_n/\sqrt{n}\}} P^n_{\theta_0, G_0}(\{|D_n(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty. \qquad \ldots \text{(B.6)}$$

Claim : Given any sequence $\{d_n\}$ of non-negative real numbers s.t. $d_n \leqslant c_n \;\forall\; n$ and $d_n \uparrow \infty$,

$$P^n_{\theta_0, G_0}\left(\left\{\text{There is a solution of (3.1) lying inside } \left(\theta_0 - \frac{d_n}{\sqrt{n}}, \theta_0 + \frac{d_n}{\sqrt{n}}\right)\right\}\right) \to 1 \text{ as } n \to \infty. \qquad \ldots \text{(B.7)}$$

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

Proof of the Claim Fix any sequence $\{d_n\}$ s.t. $d_n \leqslant c_n$, for all n and $d_n \uparrow \infty$. By (B.6)

$$D_n\left(\theta_0 \pm \frac{d_n}{\sqrt{n}}\right) \xrightarrow{P^n_{\theta_0, G_0}} 0. \qquad \ldots \text{(B.8)}$$

Again, by condition (iv)(b),

$$\left\{\mathscr{L}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i, \theta_0, G_0) \,\middle|\, P^n_{\theta_0, G_0}\right)\right\}_{n \geqslant 1} \text{ is tight}$$

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

$$d_n\left(\int \psi(x, \theta_0, G_0) f'(x, \theta_0, G_0) d\mu(x)\right) \to \infty \text{ as } n \to \infty \qquad \ldots \text{(B.9)}$$

[assuming, without loss of generality,

$$\int \psi(x, \theta_0, G_0) f'(x, \theta_0, G_0) d\mu(x) > 0].$$

From (B.8) and (B.9),

$$\frac{1}{\sqrt{n}} \sum_{\substack{i\ odd\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0 \pm \frac{d_n}{\sqrt{n}}, \hat{G}_n^E\right) + \sum_{\substack{i\ even\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0 + \frac{d_n}{\sqrt{n}}, \hat{G}_n^O\right)$$

$\overset{w}{\Rightarrow}$ Point mass at $\mp\infty$,

$$\text{i.e., } P^n_{\theta_0,G_0}\left(\left\{\frac{1}{\sqrt{n}}\left[\sum_{\substack{i\ odd\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0+\frac{d_n}{\sqrt{n}}, \hat{G}_n^E\right) + \sum_{\substack{i\ even\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0+\frac{d_n}{\sqrt{n}}, \hat{G}_n^O\right)\right] < -K\right\}\right) \to 1 \quad \ldots \quad (B.10)$$

and

$$P^n_{\theta_0,G_0}\left(\left\{\frac{1}{\sqrt{n}}\left[\sum_{\substack{i\ odd\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0-\frac{d_n}{\sqrt{n}}, \hat{G}_n^E\right) + \sum_{\substack{i\ even\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0-\frac{d_n}{\sqrt{n}}, \hat{G}_n^O\right)\right] > K\right\}\right) \to 1,$$

for any $K > 0$.

Define, $A_{n,K}(\{d_n\})$

$$= \left\{\frac{1}{\sqrt{n}}\left[\sum_{\substack{i\ odd\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0+\frac{d_n}{\sqrt{n}}, \hat{G}_n^E\right) + \sum_{\substack{i\ even\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0+\frac{d_n}{\sqrt{n}}, \hat{G}_n^O\right)\right] < -K\right\}$$

$$\cap \left\{\frac{1}{\sqrt{n}}\left[\sum_{\substack{i\ odd\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0-\frac{d_n}{\sqrt{n}}, \hat{G}_n^E\right) + \sum_{\substack{i\ even\\ 1\leqslant i\leqslant n}} \psi\left(X_i, \theta_0-\frac{d_n}{\sqrt{n}}, \hat{G}_n^O\right)\right] > K\right\},$$

then by (B.10), $P^n_{\theta_0,G_0}$, $(A_{n,K}(\{d_n\})) \to 1$ as $n \to \infty$ and by condition (iv) (a) on $A_{n,K}(\{d_n\}\}$ there is a solution of (3.1) laying inside $\left(\theta_0 - \frac{d_n}{\sqrt{n}}, \theta_0 + \frac{d_n}{\sqrt{n}}\right)$.

Since $\{d_n\}$ was arbitrary, this proves (B.7).

(II) (B) Suppose not. Then there is a sequence $\{d_n\}$ of nonnegative real numbers increasing to infinity s.t. $d_n \leqslant c_n$ for all n and $P^n_{\theta_0,G_0}(\{\sqrt{n}\ |T_n - \theta_0| > d_n\}) \not\to 0$, where $\{c_n\}$ is the sequence considered in (B. 6). [Note that, without loss of generality, we can assume $c_1 > 0$, $d_1 > 0$].

Choose and fix a sequence of positive real numbers

$$\{\alpha_n\} \text{ s.t. } \frac{2d_n}{3d_{n-1}} < \alpha_n < \frac{3d_n}{4d_{n-1}}, \text{ for all } n \geqslant 1.$$

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

$$P^n_{\theta_0,G_0}(\{\sqrt{n}\,|T_n-\theta_0| > d_n, \sqrt{n}\,|U_n-\theta_0|$$

$$< \min(d_n-\alpha_n d_{n-1}, \tfrac{1}{2}\log n)\}) \not\to 0.$$

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

Note that in (B.7) one can easily drop the assumption of increasingness of $\{d_n\}$

By von-Neumann selection theorem choose a measurable function S_n which solves (3.1) on $B_{n,K}$.

Define $C_n = \Big\{\sqrt{n}\,|T_n-\theta_0| > d_n,$

$$\sqrt{n}\,|U_n-\theta_0| < \min\left(d_n-\alpha_n\, d_{n-1}, \frac{1}{2}\log n\right)\Big\}.$$

Then on $B_{n,K}\cap C_n$, S_n solves (3.1), $\sqrt{n}\,|S_n-U_n| < \min(d_n-\alpha_n d_{n-1}, \log n)$ whereas $\sqrt{n}\,|T_n-U_n| > \alpha_n d_{n-1}$ and $P^n_{\theta_0,G_0}(B_{n,K}\cap C_n) \not\to 0$ contradicting the definition of T_n.

(III) (A) Fix $c > 0$, $\epsilon > 0$ and $\eta > 0$. To show that there is $n_0 \geqslant 1$ s.t. $n \geqslant n_0$ implies

$$P^n_{\theta_0,G_0}\left(\left\{\sup_{\{\theta\,:\,|\theta-\theta_0| < c/\sqrt{n}\}} |D_n(\theta)| > \epsilon\right\}\right) < \eta \qquad \ldots \text{(B.11)}$$

Fix a positive number α which divides c. For $\theta \in \left[\theta_0-\frac{c}{\sqrt{n}}, \theta_0+\frac{c}{\sqrt{n}}\right]$, define

$$D_n^{(\alpha)}(\theta) := \frac{\sqrt{n}}{\alpha}\sum_{i=-c/\alpha}^{c/\alpha-1}\Bigg[\left(\theta-\theta_0-\frac{i\alpha}{\sqrt{n}}\right) D_n\left(\theta_0+\frac{(i+1)\alpha}{\sqrt{n}}\right)$$

$$+\left\{\frac{(i+1)\alpha}{\sqrt{n}}-\theta+\theta_0\right\} D_n\left(\theta_0+\frac{i\alpha}{\sqrt{n}}\right)\Bigg]$$

$$1_{\left\{\theta_0+\frac{i\alpha}{\sqrt{n}} \leqslant \theta \leqslant \theta_0+\frac{(i+1)\alpha}{\sqrt{n}}\right\}}$$

Now,

$$P^n_{\theta_0, G_0}\left(\left\{\sup_{\{\theta : |\theta-\theta_0| \leqslant c/\sqrt{n}\}} |D_n(\theta)| > \epsilon\right\}\right)$$

$$\leqslant P^n_{\theta_0, G_0}\left(\left\{\sup_{\{\theta : |\theta-\theta_0| < c/\sqrt{n}\}} |D_n(\theta)-D_n^{(\alpha)}(\theta)| > \epsilon/2\right\}\right)$$

$$+P^n_{\theta_0, G_0}\left(\left\{\sup_{\{\theta : |\theta-\theta_0| \leqslant c/\sqrt{n}\}} |D_n^{(\alpha)}(\theta)| > \epsilon/2\right\}\right)$$

$$\leqslant P^n_{\theta_0, G_0}\left(\left\{\sup_{\{(c', c'', G): |c'|, |c''| \leqslant c, |c''-c'| \leqslant \alpha, G \in B(G_0, \delta_0)\}} \left|D_n\left(\theta_0+\frac{c''}{\sqrt{n}}\right)-D_n\left(\theta_0+\frac{c'}{\sqrt{n}}\right)\right| > \epsilon/2\right\}\right)$$

$$+P^n_{\theta_0, G_0}\left(\left\{\sup_{i \in \{0, \pm 1, \pm 2, \ldots, \pm c/\alpha\}} \left|D_n\left(\theta_0+\frac{i\alpha}{\sqrt{n}}\right)\right| > \epsilon/2\right\}\right)$$

$$\leqslant \frac{2}{\epsilon} E_{P^n_{\theta_0, G_0}}\left(\sup_{\{(c', c'', G): |c'|, |c''| \leqslant |c''-c'| \leqslant \alpha, G \in B(G_0, \delta_0)\}} \left|D_n\left(\theta_0+\frac{c''}{\sqrt{n}}\right)-D_n\left(\theta_0+\frac{c'}{\sqrt{n}}\right)\right|\right)+$$

$$\sum_{i=-c/\alpha}^{c/\alpha} P^n_{\theta_0, G_0}\left(\left\{\left|D_n\left(\theta_0+\frac{i\alpha}{\sqrt{n}}\right)\right| > \epsilon/2\right\}\right)$$

$$\leqslant \frac{2}{\epsilon}\left[E_{P_{\theta_0, G_0}}\left(\sup_{\{(c', c'', G): |c'|, |c''| \leqslant |c''-c'| < \alpha, G \in B(G_0, \delta_0)\}} \sqrt{n}\left|\psi\left(X_1, \theta_0+\frac{c''}{\sqrt{n}}, G\right)-\psi\left(X_1, \theta_0+\frac{c'}{\sqrt{n}}, G\right)\right|\right)\right.$$

$$\left.+\alpha\left|\int \psi(\cdot, \theta_0, G_0) f'(\cdot, \theta_0, G_0)\, d\mu(\cdot)\right|\right]$$

$$+\sum_{i=-c/\alpha}^{c/\alpha} P^n_{\theta_0, G_0}\left(\left\{\left|D_n\left(\theta_0+\frac{i\alpha}{\sqrt{n}}\right)\right| > \epsilon/2\right\}\right)$$

$$< \frac{2\alpha}{\epsilon}\left\{\int A(\cdot, \theta_0, G_0) f(\cdot, \theta_0, G_0) d\mu(\cdot)+\int \psi(\cdot, \theta_0, G_0) f'(\cdot, \theta_0, G_0) d\mu(\cdot)\right\}$$

$$+\sum_{i=-c/\alpha}^{c/\alpha} P^n_{\theta_0, G_0}\left(\left\{\left|D_n\left(\theta_0+\frac{i\alpha}{\sqrt{n}}\right)\right| > \epsilon/2\right\}\right) = I+II \text{ by (v) (a)}$$

Let us choose $\alpha > 0$ s.t. $\alpha | c$ and $I < \eta/2$. ... (B. 12)

Using (I) choose $n_0 \geqslant 1$ s.t. $n \geqslant n_0$ implies $\text{II} < \eta/2$. ... (B. 13)

Then (B. 11) follows from (B. 12) and (B. 13).

(III) (B) Easy.

(IV) An easy consequence of Proposition 2.2.

References

AMARI, S. and KUMON, M. (1985). Optimal estimation in the presence of infinitely many nuisance parameters—geometry of estimating functions. *Tech. Report* no. METR 85-2, Univ. Tokyo, Japan.

BHANJA, J. and GHOSH, J. K. (1987). Inference about a parameter in the presence of a large number of nuisance parameters. *Bull. Inst. Math. Stat.* **16**, 73.

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.

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.

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 infinite dimensional nuisance parameter. Probability Theory and Mathematical Statistics (Fourth USSR-Japan Symposium Proceedings, 1982. Edited by Itô and Prohorov) : *Lecture Notes in Math.*, 1021, 195-229, Springer-Verlag, New York.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981). *Statistical Estimation : Estimation Theory*, Springer Verlag, New York.

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. *Biometrika*, **71**, 445-459.

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.

——— (1982). Conditional score functions : some optimality results. *Biometrika*, **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, B. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, **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.

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 Edition.

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. (1986). Statistical estimation in large parameter spaces. *Thesis*, University of Leiden.

——— (1988). Estimating a real parameter in a class of semiparametric models. *Ann. Statist.*, **16**, 1450-1474.

Theoretical Statistics and Mathematics Division
Indian Statistical Institute
203 B. T. Road
Calcutta 700 035
India.