Sankhyā : The Indian Journal of Statistics
1992, Volume **54**, Series A, Pt. 2, pp. 135-156.

EFFICIENT ESTIMATION WITH MANY NUISANCE PARAMETERS

(Part II)

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

SUMMARY. In Part II, we shall construct an efficient estimate for the fixed set-up Neyman-Scott model where the nuisance parameters are unknown constants. This part also contains two special cases where we have orthogonality of θ and $G(\underline{G}_n)$ or partial likelihood factorisation of f. A summary of the main results appear in the Introduction to Part I.

4*. Fixed set-up

In this section, we shall state the analogues of Lemma 3.1, Theorem 3.2 and Theorem 3.3 in the fixed set-up. However, we apply a random permutation Π to the original sample $(X_1, X_2, ..., X_n)$ and base the analysis on $(X_{\Pi(1)}, X_{\Pi(2)}, ..., X_{\Pi(n)})$. Let s_n denote the group of all permutations of $\{1, 2, ..., n\}$ and P_n denote the probability distribution of Π. Later we shall make an appropriate 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 4.1. Let $(Y, \mathscr{Y})$, $(Z, \mathscr{Z})$ be measurable spaces. For any $n \geqslant 1$, Z-valued statistic V_n on $(Y, \mathscr{Y})^n$ and probability measure P_n on s_n, call the statistic sending $(y_1, y_2, ..., y_n)$ to $V_n(y_{\Pi(1)}, y_{\Pi(2)}, ..., y_{\Pi(n)})$ *the randomisation of the statistic* V_n *corresponding to* P_n and denote it by $V_n^*(P_n)$.

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

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.

* Sections 1 to 3 appeared in Part 1, February 1992 issue of Sankhyā.

Definition 4.2. Let (Y, ρ) and ϕ be as considered in Definition 2.1. For any $n \geqslant 1$, estimate V_n of $\phi(\theta_0, \underline{G}_n)$ in Model I ($\phi(\theta_0, G_0)$ 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 4.1 a *randomised estimate of* $\phi(\theta_0, \underline{G}_n)$ *in Model I* ($\phi(\theta_0, G_0)$ *in Model II*).

As a special case of the above definition, we can define the notions of randomised estimates of θ_0, $\underline{G}_n$ or $(\theta_0, \underline{G}_n)$ in Model I (θ_0, G_0 or (θ_0, G_0) in Model II) (cf Definition 2.1).

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

$$\left(\prod_{i=1}^{n} P_{\theta_0, \xi_i}\right)(\{V_n^*(P_n) \in A\}) = \int \left(\prod_{i=1}^{n} P_{\theta_0, \xi_{\pi(i)}}\right)(\{V_n \in A\})\, dP_n(\pi) \quad \ldots \quad (4.1)$$

for all A in $\mathcal{Z}$, θ_0 in $\bar{\Theta}$ and $\{\xi_i\}_{1 \leqslant i \leqslant n}$ in Ξ^n and

$$P^n_{\theta_0, G_0}(\{V_n^*(P_n) \in A\}) = P^n_{\theta_0, G_0}(\{V_n \in A\}) \quad \ldots \quad (4.2)$$

for all A in $\mathcal{Z}$, θ_0 in $\bar{\Theta}$ and G_0 in $\mathcal{G}$.

(2) In view of relations (4.1)—(4.2), there are extensions of Definitions 2.1—2.4 for randomised estimates and in view of observation (1), for any property P defined in Definitions 2.1—2.4 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 and Model II.

(3) As in observation (2), the notion of efficiency (I) ((II)) has obvious extensions for randomised estimates and one can easily prove that in the extended sense, regularity (I) implies regularity (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 section, 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_0 (vide Definition 2.2) and (b) there is a uniformly consistent (I) estimate $\hat{G}_n$ of $\underline{G}_n$ (vide Definition 2.1).

Convention 1 : For any $n \geqslant 1$, 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 (3.1).

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

where $(\hat{G}_n^O)^*$ and $(\hat{G}_n^E)^*$ are obtained from $\hat{G}_n$ using (11) of Section 2 and Definition 4.1 with $P_n = P_n^u$. $T_n^*(\psi)$ is defined in analogy with $T_n(\psi)$ by replacing (3.1) and U_n by (3.1)* and U_n^*, respectively in Definition 3.1. Clearly $T_n^*(\psi)$ equals $(T_n(\psi))^*$. Note that

(4) Theorems 3.2—3.3 and relation (4.2) together imply that Z_n^* is 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 3.1 when we have Model I instead of Model II and randomised estimates. Before stating the required lemma we need two more auxiliary results namely the folowing proposition and Lemma 4.1(t).

Proposition 4.1. *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 (II) of Section 2, of course they are not observable since ξ_i's are unknown constants). For any $\epsilon > 0$*

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

where d denotes the Prohorov metric on $\mathcal{G}$ as defined in (10) *of Section* 2.

The proof is given in Appendix C.

Corollary 4.1.1. *There is a sequence $\{\epsilon_n^o\}_{n \geqslant 1}$ decreasing to zero such that*

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

Proof. The result follows trivially from the proposition.

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

$$\alpha_n(\epsilon) := \{\{\xi_i\}_{1 \leqslant i \leqslant n} : d(\underline{G}_n^O, \underline{G}_n^E) \leqslant \epsilon\} \qquad \ldots \quad (4.3)$$

Fix any sequence $\{\epsilon_n\}_{n \geqslant 1}$ decreasing to zero. Let $\theta_0 \in \Theta$. Let $\{\xi_{ni}\}_{1 \leqslant i \leqslant n, n \geqslant 1}$ be a triangular array of elements in Ξ such that

$$\{\xi_{ni}\}_{1 \leqslant i \leqslant n} \in \alpha_n(\epsilon_n) \quad \forall n \qquad \ldots \quad (4.4)$$

Corollary 4.1.1 leads to an analysis of the following triangular array version of Model 1.

Model I(t). Let $\{X_{ni}\}_{1 \leqslant i \leqslant n,\, n \geqslant 1}$ be a triangular array of rowwise independent random variables with X_{nt} following the distribution $P_{\theta_0,\, \xi_{ni}}$, where $\theta_0 \epsilon \Theta$ and the triangular array $\{\xi_{ni}\}_{1 \leqslant i \leqslant n,\, n \geqslant 1}$ satisfies (4.4).

Convention 2 : Let $(Y, \mathscr{Y})$ and $(Z, \mathscr{Z})$ be as considered in Definition 4.1. Let $\{y_{ni}\}_{1 \leqslant i \leqslant n,\, n \geqslant 1}$ be a triangular array of elements in Y. For any $n \geqslant 1$ and Z-valued statistic V_n on $(Y, \mathscr{Y})^n$, we shall denote $V_n(\{y_{ni}\}_{1 \leqslant i \leqslant n})$ by $V_{n,n}$.

The above convention suggests obvious Model I(t)-analogue of equation (3.1) which we shall denote by (3.1)(t).

(5) As in observation (2), Definitions 2.1—2.4 have obvious Model I(t)- t)- analogues, and for any property P defined in Definitions 2.1—2.4 and statistic V_n, V_n satisfies $P(I)$ *only if* $V_{n,n}$ satisfies P(I(t)).

Let ψ be a kernel. Fix θ_0 in Θ and $\{\xi_{ni}\}_{1 \leqslant i \leqslant n,\, n \geqslant 1}$ satisfying relation (4.4). The following is the Model I(t)-analogue of relation (3.2)

$$\begin{aligned}\widetilde{D}_{n,n}(\theta) := \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ odd}}}^{n} \{&\psi(X_{nt}, \theta, \hat{G}^{E}_{n,n}) - \psi(X_{nt}, \theta_0, \underline{G}_{n,n}) \\ &+ (\theta-\theta_0) \int \psi(., \theta_0, \underline{G}_{n,n}) f'(., \theta_0, \underline{G}_{n,n}) d\mu(.)\} \\ + \frac{1}{\sqrt{n}} \sum_{\substack{i=1 \\ i \text{ even}}}^{n} \{&\psi(X_{nt}, \theta, \hat{G}^{O}_{n,n}) - \psi(X_{nt}, \theta_0, \underline{G}_{n,n}) \\ &+ (\theta-\theta_0) \int \psi(., (\theta_0, \underline{G}_{n,n}) f'(., \theta_0, \underline{G}_{n,n}) d\mu(.)\} \qquad \ldots \quad (4.5)\end{aligned}$$

for all θ in Θ.

In order to state Lemma 4.1(t) we need new conditions in which G_0 has to be replaced by $\underline{G}^{E}_{n,n}$, $\underline{G}^{O}_{n,n}$ and then $\underline{G}_{n,n}$ in the conditions (i)—(v) and U(i)—U(vi) of Section 3. The exact conditions to be referred to as $(\text{i})^t$—$(\text{v})^t$ and U(i)^t—U(vi)^t which are somewhat artificial, are given in Appendix C. However Lemma 4.1(t) is only an auxiliary result needed to prove our main result, namely Lemma 4.1(IV), the assumptions for which are only slightly stronger than those of Lemma 3.1(IV), vide observation (6) preceeding Lemma 4.1.

Lemma 4.1(t). *Assume* (C1)(*b*). *Fix any sequence* $\{\epsilon_n\}_{n \geqslant 1}$ *decreasing to zero. Fix* θ_0 *in* Θ *and* $\{\xi_{nt}\}_{1 \leqslant i \leqslant n,\, n \geqslant 1}$ *satisfying* (4.4). *Let* ψ *be kernel. Let*

$\widetilde{D}_{n,n}$ be as defined by relation (4.5). Also, whenever it makes sense, let $T_{n,n}(\psi)$ be the estimate defined through Definition 3.1 and Convention 2. We can conclude the following.

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

$$\sup_{\{\theta : |\theta-\theta_0| \leqslant c/\sqrt{n}\}} \left(\prod_{i=1}^{n} P_{\theta_0, \xi_{ni}}\right) (\{|\widetilde{D}_{n,n}(\theta)| > \epsilon\}) \to 0 \text{ as } n \to \infty.$$

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

A) for any sequence $\{c_n\}_{n \geqslant 1}$ increasing to infinity $\left(\prod_{i=1}^{n} P_{\theta_0, \xi_{ni}}\right)$(\{There is a solution of (3.1)(t) lying in $(\theta_0 - c_n/\sqrt{n},\ \theta_0 + c_n/\sqrt{n})\}) \to 1$ as $n \to \infty$ and

(B) under assumption (C1) (a), $T_{n,n}(\psi)$ is a $\sqrt{n}$-consistent solution (I(t)) of (3.1)(t).

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

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

$$\left(\prod_{i=1}^{n} P_{\theta_0, \xi_{ni}}\right) \left(\left\{\sup_{\{\theta : |\theta-\theta_0| \leq c/\sqrt{n}\}} |\widetilde{D}_{n,n}(\theta)| > \epsilon\right\}\right) \to 0 \text{ as } n \to \infty$$

and

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

$$\sup_{x \in R} \left| \left(\prod_{i=1}^{n} P_{\theta_0, \xi_{ni}}\right) (\{\sqrt{n}(T_{n,n}(\psi) - \theta_0) \leqslant x\}) - \phi(x/V(\theta_0, \underline{G}_{n,n}, \psi))\right| \to 0 \quad \text{as } n \to \infty$$

where V is the positive real-valued function defined in (9) of Section 2.

(IV) As in Lemma 3.1 (IV), for any conclusion C among (I)—(III), let UC denote the conclusion that C holds uniformly with respect to θ_0 in compact subsets of Θ and $\{\xi_{ni}\}_{1 \leqslant i \leqslant n, n \geqslant 1}$ satisfying (4.4). 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 $U(i)^t - U(vi)^t$ hold.

We can prove this by an easy modification of the proof of Lemma 3.1.

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

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

$$\beta_n(\{\xi_i\}_{1 \leqslant i \leqslant n}) = \{\pi \in s_n : \{\xi_{\pi(i)}\}_{1 \leqslant i \leqslant n} \in \alpha_n(\epsilon_n^0)\}, \qquad \ldots (4.6)$$

where the sequence $\{\epsilon_n^0\}_{n \geqslant 1}$ is defined in Corollary 4.1.1.

Let ψ be a kernel. Fix θ_0 in Θ and $\{\xi_n\}_{n\geqslant 1}$ in Ξ^∞. Consider the following analogue of (4.5) in the present context.

$$\begin{aligned}\widetilde{D}_n^*(\theta) := \frac{1}{\sqrt{n}} \sum_{\substack{i=1\\ i\ odd}}^{n} \{\psi(X_i^*, \theta, (\hat{G}_n^E)^*) - \psi(X_i^*, \theta_0, \underline{G}_n)\\ + (\theta-\theta_0) \int \psi(., \theta_0, \underline{G}_n) f'(., \theta_0, \underline{G}_n)\, d\mu(.)\}\\ + \frac{1}{\sqrt{n}} \sum_{\substack{i=1\\ i\ even}}^{n} \{\psi(X_i^*, \theta, (\hat{G}_n^O)^*) - \psi(X_i^*, \theta_0, \underline{G}_n)\\ + (\theta-\theta_0) \int \psi(., \theta_0, \underline{G}_n) f'(., \theta_0, \underline{G}_n) d\mu(.)\}\end{aligned} \tag{4.7}$$

for all θ in Θ.

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

(i)* (a) $\lim\limits_{\theta\to\theta_0} \limsup\limits_{n\to\infty}$

$$\int \left[\left\{\frac{\Lambda(.,\theta_0,G,\theta,G)-1}{(\theta-\theta_0)} - s_\theta(., \theta_0, G)\right\}^2 f(.,\theta_0,G)\right]_{G=(\underline{G}_n^O)^* \text{ or } (\underline{G}_n^E)^*} d\mu(.) = 0$$

where s_θ denote the kernel defined by relation (2.2), and

(b) $$\int \frac{\{f'(., \theta_0, G) - f'(., \theta_0, \underline{G}_n)\}^2}{f(., \theta_0, \underline{G}_n)}\Bigg|_{G=(\underline{G}_n^O)^* \text{ or } (\underline{G}_n^E)^*} d\mu(.) \to 0 \text{ as } n\to\infty.$$

(ii)* (a) There is $\delta_0 > 0$ such that

$$\limsup_{n\to\infty} \sup_{(\theta,G')\in B((\theta_0,G),\, \delta_0)} \int \psi^2(., \theta, G') f(., \theta_0, G) d\mu(.)\Big|_{G=(\underline{G}_n^O)^* \text{ or } (\underline{G}_n^E)^*} < \infty$$

and

(b) $$\limsup_{n\to\infty} \sup_{(\theta,G')\in B((\theta_0,G),\epsilon_n^0)} \int \{\psi(., \theta, G') - \psi(., \theta_0, G)\}^2 d\mu(.)\Big|_{G=(\underline{G}_n^O) \text{ or } (\underline{G}_n^E)^*} = 0$$

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

$$\limsup_{n\to\infty} \left[\sup_{\{\theta\,:\,|\theta-\theta_0| < c/\sqrt{n}\}} \left(\prod_{i=1}^{n} P_{\theta_0,\xi_i}\right)(\{\sqrt{n}\,|\int\psi(., \theta, G') f(., \theta, G) d\mu(.)| > \epsilon\})\right] = 0$$

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

(iv)* (a) There $\delta_0 > 0$ and $n_0 \geqslant 1$ such that for all $n \geqslant n_0$, x in S and G in $B(\underline{G}_n, \delta_0)$,

$$\psi(x, ., G) \in C(B(\theta_0, \delta_0))$$

(b) $\limsup\limits_{n\to\infty} \left[\int \psi^2(., \theta_0, \underline{G}_n) f(., \theta_0, \underline{G}_n)\, d\mu(.)\right] < \infty$

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

and (c) $\liminf |\int \psi(.,\theta_0,\underline{G}_n)f'(.,\theta_0,\underline{G}_n)d\mu(.)| > 0$.

(v)* There is $\delta_0 > 0$ and $A(.,\theta_0,\underline{G}_n) \in L_1(f(.,\theta_0,\underline{G}_n))$ such that

$$|\psi(.,\theta',G)-\psi(.,\theta,G)| \leqslant |\theta'-\theta| A(.,\theta_0,\underline{G}_n)$$

for all θ, θ' in $B(\theta_0,\delta_0)$ and G in $B(\underline{G}_n,\delta_0)$

Analoguous to the formulation of the conditions U(i)—U(v) on the basis of the conditions (i) to (v) in Section 3, we formulate the condiions U(i)* — U(v)*. An additional condition U(vi)* is given below.

U(vi)* (a) There is $\delta_0 > 0$ such that

$$\limsup_{n\to\infty} \sup_{(\theta,G)\in B((\theta_0,\underline{G}_n),\delta_0)} \left[\left\{\int_{\{\psi^2(.,\theta,G)\geqslant K\}} \psi^2(.,\theta,G)f(.,\theta,G)d\mu(.)\right\}/J(\theta,G,\psi)\right] \to 0$$
$$\text{as } K\to\infty$$

and (b) $(\theta, G) \to J(\theta, G, \psi)$ is continuous, where J denote the non-negative real-valued function defined in (8) of Section 2 which is positive by part (a) of this condition.

Note that

(6) Any condition among U(ii)*—U(vi)* is equivalent to the corresponding condition among U(ii)—U(vi) of Section 3 whereas U(i)* is a stronger version of condition U(i) of Section 3 with U(i)* (a) equivalent to it.

The following is the required analogue of Lemma 3.1.

Lemma 4.1. *Assume (C1)(b). Fix θ_0 in Θ and $\{\xi_n\}_{\geqslant 1}$ in Ξ^∞. Let ψ be a kernel. Let $\widetilde{D}_n^*$ be as defined in relation (4.7). 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 $\epsilon > 0$*

$$\sup_{\{\theta:\,|\theta-\theta_0|\leqslant c/\sqrt{n}\}} \left(\prod_{i=1}^n P_{\theta_0,\,\xi_i}\right)(\{|\widetilde{D}_n^*(\theta)| > \epsilon\}) \to 0 \text{ as } n\to\infty.$$

(*II*) *If conditins (i)*—(iv)* hold then*

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

$$\left(\prod_{i=1}^n P_{\theta_0,\,\xi_i}\right)(\{\text{There is a solution of (3.1)* lying in } (\theta_0-c_n/\sqrt{n},\ \theta_0+c_n/\sqrt{n})\})\to 1$$
$$\text{as } n\to\infty$$

and (B) under assumption (C1)(a), $T_n^*(\psi)$ is a randomised $\sqrt{n}$-consistent solution (I) of (3.1)*.

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

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

$$\left(\prod_{i=1}^{n} P_{\theta_0, \xi_i}\right)\left(\left\{\sup_{\{\theta:\, |\theta-\theta_0| \leq c/\sqrt{n}\}} |\widetilde{D}_n^*(\theta)| > \epsilon\right\}\right) \to 0 \text{ as } n \to \infty$$

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

$$\sup_{x \in \mathbf{R}} \left|\left(\prod_{i=1}^{n} P_{\theta_0, \xi_i}\right)(\{\sqrt{n}(T_n^*(\psi)-\theta_0) \leqslant x\}) - \phi(x/V(\theta_0, \underline{G}_n, \psi))\right| \to 0$$

as $n \to \infty$ where V denote the positive real-valued function defined in (9) *of Section* 2.

(IV) As in Lemma 3.1 *(IV), for any conclusion C among (I)$-$(III), let UC denote the conclusion that C holds uniformly with respect to $(\theta_0, \{\xi_n\}_{n\geqslant 1})$ in compact subsets of $\Theta \times \Xi^\infty$. Then U(I), U(II) and U(III) (A) holds 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 $n \geqslant 1$, $d((\underline{G}_n^O)^*, (\underline{G}_n^E)^*) \leqslant \varepsilon_n^o$ if and only if $\{\xi_i^*\}_{1\leqslant i\leqslant n} \in \alpha_n(\epsilon_n^o)$ (vide relation (4.3)) so that Corollary 4.1.1 can be restated as

$$\sup_{\{\xi_i\}_{1\leqslant i\leqslant n} \in \Xi^n} [1 - P_n^u(\beta_n(\{\xi_i\}_{1\leqslant i\leqslant n}))] \to 0 \text{ as } n \to \infty \quad \ldots \quad (4.8)$$

where β_n's are as defined in relation (4.6).

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 similary.

For this purpose note that conditions (i)*$-$(iii)* imply that for any θ_0 in Θ, $\{\xi_n\}_{n\geqslant 1}$ in Ξ^∞ and sequence of permutations $\{\pi_n\}_{n\geqslant 1}$ with π_n in $\beta_n(\{\xi_i\}_{1\leqslant i\leqslant n})$, conditions (i)^t$-$(iii)^t with $\varepsilon_n = \varepsilon_n^o$ hold at the point θ_0 in Θ and triangular array $\{\xi_{\pi_n(i)}\}_{1\leqslant i\leqslant n,\, n\geqslant 1}$ which statisfies (4.4) by the choice of π_n's. Hence by part (I) of Lemma 4.1(t) for any $c > 0$ and $\epsilon > 0$

$$\sup_{\pi \in \beta_n(\{\xi_i\}_{1\leqslant i\leqslant n})} \sup_{\{\theta:\, |\theta-\theta_0| < c/\sqrt{n}\}}$$

$$\left(\prod_{i=1}^{n} P_{\theta_0, \xi_i}\right)(\{|\widetilde{D}_n^*(\theta)| > \epsilon \mid \Pi = \pi\}) \to 0 \text{ as } n \to \infty. \quad \ldots \quad (4.9)$$

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

$$f_n(\pi, \alpha) := \sup_{\{\theta:\, |\theta-\theta_0| \leqslant c/\sqrt{n}\}} \left(\prod_{i=1}^{n} P_{\theta_0, \xi_i}\right)(\{|\widetilde{D}_n^*(\theta)| > \epsilon \mid \Pi = \pi\}).$$

By (4.9)

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

and f_n is [0, 1]-valued.

Hence, by (4.8)

$$\int_{s_n} f_n(., \alpha)\, d P_n^u = \int_{\beta_n(\alpha)} f_n(., \alpha)\, d P_n^u + \int_{\beta_n^c(\alpha)} f_n(., \alpha)\, d P_n^u \to 0 \text{ as } n \to \infty$$

proving part (I). Part U(I) follows similarly from the uniform versions of (4.8) and (4.9) provided A_n is replaced by relevant compact subset of $\Theta \times \Xi^n$.

Definition 3.2 has an obvious extension for randomised estimates. The following is the Model I-analogue of the extension.

Definition 4.3. 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 uniformly $\sqrt{n}$-consistent solution (I) of (3.1), i.e. any uniformly $\sqrt{n}$-consistent solution (I) of a randomisaton of (3.1) namely,

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

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

There is an obvious analogue of Lemma 3.1a for Model I and randomised estimates and in view of observation (6), we can make the following remark.

Remark 4.1. 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_0, V_n is GC_1(I) *only if* it is GC_1(II). Also, one can easily verify that Example 3.1 and 3.2, with $T_n(\psi)$ replaced by $T_u^*(\psi)$ for the latter one, remain valid for Model I.

The following is the Model I-analogue of Remark 3.3.

Remark 4.2. If $(S, \mathcal{S}) = (\mathbf{R}^p, \mathcal{B}^p)$ and assumptions (A1) and (B2) (a) hold then Corollaries 2.1.1 and 2.1.2 enable us to drop assumption (C1) (b) even if Θ is unbounded.

Let us now write down the analogues of Theorems 3.2 and 3.3.

Theorem 4.2. *Assume* (*C*1), (*B*2) *and* (*B*3). *The* (*randomised*) *estimate* Z_n^* *of* θ_0, *as defined through relations* (3.3)—(3.4), *Definition* 4.1 *and Convention* 1 *is UAN* (*I*) *with AV* (1/I).

Theorem 4.3. *Assume* (*C*1), (*B*2) *and* (*B*3*s*). *The* (*randomised*) *estimate* $T_n^*(\bar{\psi})$ *of* θ_0, *as defined through Definitions* 3.1, 4.1 *and Convention* 1 *is UAN* (*I*) *with AV* (1/I).

Remark 4.3. Theorems 4.2 and 4.3 tell us that Z_n^* and $T_n^*(\bar{\psi})$ have the most limiting concentration around θ_0 among the randomised regular (I) estimates, i.e. the following holds.

For any $(\theta_0, \{\xi_n\}_{n\geq 1})$ in $\Theta\times\Xi^\infty$, randomised regular (I) estimate V_n of θ_0 and convex symmetric set A in $\mathcal{B}(\boldsymbol{R})$,

$$\lim_{n\to\infty}\left(\prod_{j=1}^{n} P_{\theta_0,\xi_i}\right)(\{\sqrt{n}\, I^{1/2}(\theta_0, \underline{G}_n)(W_n-\theta_0)\in A\}) = P(\mathcal{N}(0,1)\in A)$$

$$\geqslant \limsup_{n\to\infty}\left(\prod_{i=1}^{n} P_{\theta_0,\xi_i}\right)(\{\sqrt{n}\, I^{1/2}(\theta_0, \underline{G}_n)(V_n-\theta_0)\in A\})$$

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

Remark 4.4. It has been pointed out by van der Vaart (1987) as criticism of 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 of Chatterjee and Das (1983) as variance estimation. However, such better estimates due to van der Vaart are, of necessity, non-symmetric in $X_1, \ldots, X_n$. This makes one reluctant to use them. Moreover, from a technical point of view, one should compare its maximum risk, over permutations of $\xi_1, \ldots, \xi_n$, 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 $(\xi_1, \ldots, \xi_n)$ is known up to a permutation. We hope to study this in a further communication.

Remark 4.5. There can be no asymptotic improvement over efficient regular 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 Section 5(b)). In particular, this holds for the estimate in Example 1.1. We omit proof.

Remark 4.6. 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

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

with Y_{jl}'s following $f(\,.,\,\theta_0,\,\xi_{jl},\,k_j)$ and n_j's being non-negative integers with $\sum_{j=1}^{r} n_j = n$. Without loss of generality let us assume that $k_1 < k_2 < \ldots < k_r$ and liminf $(n_j/n) > 0$ for all j, so that each group represents a distinct fixed set-up model by itself. Call an estimate of θ_0 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, 4.1—4.2 and observation (2), corresponding to each component fixed set up submodel. For the j-th submodels let $\bar{\psi}_j$ denote optimal kernel as defined through (2.2)—(2.4) U_{nj} and $\hat{G}_{n,j}$ denote, respectively, the uniformly $\sqrt{n}$ consistent estimate of θ_0 and uniformly consistent estimate of $\underline{G}_{nj} := F_{n_j}(\{\xi_{jl}\})_{1 \leq l \leq n_j})$ (vide Definitions 2.1—2.2) as considered in assumption (C1) the superscript $^{*}j$ stands for the operation of randomisation as defined in Definition 4.1 and the superscripts O and E stand for the operations defined in (11) of Section 2. Then an efficient regular estimate will be a solution of

$$\sum_{j=1}^{r} \Big[\sum_{\substack{l=1 \\ l \text{ odd}}}^{n_j} \bar{\psi}_j(Y_{jl}^{*j},\, \theta,\, (\hat{G}_{nj}^{*j})^E) + \sum_{\substack{l=1 \\ l \text{ even}}}^{n_j} \bar{\psi}_j\,(Y_{jl}^{*j},\, \theta,\, (\hat{G}_{nj}^{*j})^O \Big] = 0 \qquad \ldots \quad (4.10)$$

which is nearest to $\bar{U}_{n.}$ if there is a solution of (4.10) lying in $[\bar{U}_{n.} - \log n/\sqrt{n},$ $\bar{U}_{n.} + \log n/\sqrt{n}]$ and equal to $\bar{U}_{n.}$ otherwise ; where $\bar{U}_{n.} := \frac{1}{n} \sum_{j=1}^{r} n\, U_{nj}$.

Remark 4.7. In view of remarks 4.2 and 3.4, for Euclidian 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_0, and (B3) or (B3s), i.e. smoothness properties of the optimal kernel (cf Remark 3.8).

5. Two special cases

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

(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), $s_\theta(., \theta, G)$ belongs to the orghogonal 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, $f(x, ., .) \epsilon\, C_{2,0}\, (\bar{\Theta} \times \Xi)$ and (b) for any compact subset Θ_0 of Θ the following statements hold

(i) there is $\delta_0 > 0$ such that

(a) the following two families of functions

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

and $\{s_\theta^2\,(. , \theta', G') f(. , \theta, G) : (\theta, G), (\theta', G') \epsilon \Theta_0 \times \mathcal{G} \text{ with } |\theta - \theta'| + d(G, G') \leqslant \delta_0\}$ are uniformly integrable with respect to μ and

$$\text{(b)} \sup_{(\theta, G) \epsilon \Theta_0 \times \mathcal{G}} [\textstyle\int (\overline{|s_\theta'|})\, (., B(\,(\theta, G), \delta_0)\,)\, f(. , \theta, G) d\mu\,(\,.\,)] < \infty$$

and

$$\text{(ii)} \sup_{(\theta, G) \epsilon \Theta_0 \times \mathcal{G}} \left[\left\{\int 1_{\{s_\theta^2(., \theta, G) \geqslant K\}} \frac{(f')^2(. , \theta, G)}{f(. , \theta, G)} d\mu(\,.\,)\right\} / I(\theta, G)\right] \to 0 \text{ as } K \to \infty$$

Assumption (D1) and orthogonality together imply assumptions (B2) and (B3s). Hence by the theorems proved in Sections 3 and 4, 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 3.4. In particular, they hold for Example 1.2 with $p \geqslant 2$.

Example 1.2 with $p = 1$ does not fall in the exponential families described in Remark 3.4. 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 $f(. , \theta, G) = f(. , \theta', G')$ a.e. $[\lambda]$. By symmetry of the normal density function we get that $\theta = \theta'$. 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 3.4, 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 Section 2, it remains to check assumption (C1) (a) for Example 1.2 and in this respect the grand mean

$$\bar{\bar{X}} = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{n} X_{ij}$$

is a natural choice for U_n.

In view of the last paragraphs, Theorems 3.2, 3.3, 4.2 and 4.3 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, 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 $p: S\times\bar{\Theta}\to \boldsymbol{R}^+$ and $q: S\times\bar{\Theta}\times\Xi\to\boldsymbol{R}^+$ such that

$$f(x, \theta, \xi) = p(x, \theta)\, q(x, \theta, \xi) \text{ for all } (x, \theta, \xi) \in S\times\bar{\Theta}\times\Xi \quad \ldots (5.1)$$

and

$$\int p(\cdot\,, \theta')\, q(\cdot\,, \theta, \xi)\, d\mu(\cdot) = 1 \text{ for all } (\theta, \theta', \xi) \in \bar{\Theta}\times\bar{\Theta}\times\Xi \quad \ldots (5.2)$$

In cases where (5.1) and (5.2) hold we call p a *partial-likelihood function.*

In applications, for (5.1) and (5.2) 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 $t(\boldsymbol{X}_i) = \bar{\boldsymbol{X}}_i$. (An example of the other kind is Example 9.4 of Lindsay, 1980, 654–655).

Here

$$s_\theta = \frac{p'}{p} + \frac{q'}{q}. \quad \ldots (5.3)$$

Assume that

(D2) (a) For all x in S, $p(x, .) \in C_2(\bar{\Theta})$ and $q(x, ., .) \in C_{2,0}(\bar{\Theta}\times\mathcal{G})$

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

(i) there is $\delta_0 > 0$ such that

(a) the following three families of functions

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

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

and

$$\{(p'/p)^2 (., \theta')\, p(., \theta)\ \ q(., \theta, G) : \theta, \theta' \in \Theta_0 \text{ with } |\theta-\theta'| \leqslant \delta_0 \text{ and } G \in \mathcal{G}\}$$

are uniformly integrable with respect to μ, and

(b) $$\sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} |\int(\overline{|p''/p|})(., B(\theta, \delta_0))p(., \theta)q(., \theta, G)d\,\mu(.) < \infty$$

and $$\sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} [\int(\overline{(p'/p)^2})(., B(\theta, \delta_0))\, p(., \theta)q(., \theta, G)d\,\mu(.)] < \infty$$

and (ii) $$\sup_{(\theta, G) \in \Theta_0 \times \mathcal{G}} [\int 1_{\{(p'/p)^2(., \theta) \geqslant K\}} \frac{(p')^2(., \theta)q(., \theta, G)}{p(., \theta)}\, d\,\mu(.)]$$

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

(D3) For any (θ, G) in $\bar{\theta} \times \mathcal{G}$, there is $M_{\theta, G} \in \mathcal{M}_0$ 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 (5.3) we have $\bar{\psi} = p'/p$ 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 of 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 5.1. If assumption (D2) holds and equation (3.1) with $\psi = p'/p$, has a unique solution (the latter holds for Examples 9.2–9.5 of Lindsay (1980) which includes Example 1.1) $\hat{\theta}_n$ (say), then part U(III)(B) of Lemma 3.1 (equivalently, that of Lemma 4.1) holds with $T_n(\psi)$ replaced by $\hat{\theta}_n$, in other words, $\hat{\theta}_n$ is UAN(I) with AV $V(., ., \psi)$, 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 5.1, it is enough to check assumptions (D2) and (D3). We have verified assumption (D2) for more general case of Euclidean S and exponential p, q provided assumptions (a), (b)* and (d)—(f) of Remark 3.4, 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 statistics $t(x)$ 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 (5.3) and the functions of $N_{\theta,G}$ depend on x only through t. One can use the latter fact and sufficiency of t to conclude that

$$p'(. , \theta)/p(. , \theta) \in N^{\perp}_{\theta,G} \quad \forall (\theta, G).$$

Therefore,

$$\bar{\psi} = (p'/p) + \bar{\psi}_t \qquad \ldots \quad (5.4)$$

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

Therefore, using Lemma 2.1 for the marginal model, one observes that, under assumptions (A2)–(A4)

$$E_{P^t_{\theta,G'}}\{\bar{\psi}_t(., \theta, G)\} = 0 \quad \text{a.e.} [P^t_{\theta,G}] \ \forall (\theta, G, G')$$

Hence, by completeness of t,

$$\bar{\psi}_t(. , \theta, G) = 0 \text{ a.e. } [P^t_{\theta,G}] \ \forall (\theta, G)$$

$$\text{i.e. } \bar{\psi}_t(t(.), \theta, G) = 0 \text{ a.e. } [P_{\theta,G}] \ \forall (\theta, G)$$

proving, in view of (5.4), that $\bar{\psi} = p'/p$.

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,

$$\phi \in L^0_2(P_{\theta,\xi}) \ \forall \xi$$

only if

$$\phi = 0 \quad \text{a.e} [P_{\theta,\xi}] \ \forall \xi$$

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

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

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

—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 3.2, 3.3, 4.2 and 4.3 for Pfanzagl's case are (Cl)(a) and (D2) whereas those for van der Vaart's case are (Cl) 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, 1.2 as well as Example 9.6 of Lindsay (1980, 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 section. His estimate is different from ours and requires fewer regularity conditions.

Appendix C*

We shall need the following auxiliary result.

Lemma C.1. *Let (Y, ρ) be a compact metric space. Let $\mathcal{P}$ 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 $\epsilon > 0$,*

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

where $\bar{P}_n$ denote the measure $\frac{1}{n}\sum_{i=1}^{n} P_i$ on $\mathcal{P}$ and d denote the Prohorov metric on $\mathcal{P}$ as defined in (10) *of Section* 2.

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

$$\sup_{\{P_i\}_{1\leq i\leq n}\in\mathcal{P}^n} \left(\prod_{i=1}^{n} P_i\right)\left(\left\{\left|\int f\, d(F_n-\bar{P}_n)\right| > \epsilon\right\}\right) \to 0 \text{ as } n\to\infty \quad \ldots \quad (C.1)$$

Next, we shall extend (C.1) to the following

For any compact subset $\mathcal{F}$ of $C(Y)$ and $\epsilon > 0$,

$$\sup_{\{P_i\}_{1\leq i\leq n}\in\mathcal{P}^n} \left(\prod_{i=1}^{n} P_i\right)\left(\left\{\sup_{f\in\mathcal{F}}\left|\int f\, d(F_n-\bar{P}_n)\right| > \epsilon\right\}\right) \to 0 \text{ as } n\to\infty \quad \ldots \quad (C.2)$$

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 $(\epsilon/4)$-net $\{f_1, f_2, \ldots, f_k\}$ of $\mathcal{F}$. Then

$$\sup_{f\in\mathcal{F}} \left|\int f\, d(F_n-\bar{P}_n)\right| < \epsilon/2 + \max_{1\leq j\leq k}\left|\int f_j\, d(F_n-\bar{P}_n)\right|$$

*Appendices A and B appeared in Part I, February 1992 issue of Sankhyā.

Therefore

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

$$\leqslant \sup_{\{P_i\}_{1\leqslant i\leqslant n}\in\mathcal{P}^n}\left(\prod_{i=1}^{n}P_i\right)\left(\left\{\max_{1<j<k}\left|\int f_j\,d(F_n-\bar{P}_n)\right|>\epsilon/2\right\}\right)\to 0$$

as $n\to\infty$, by (C.1).

As $\mathcal{F}$ and ϵ were arbitrary, this proves (C.2).

Let us now consider the function $\phi:\mathbf{R}\to[0,1]$ defined by,

$$\phi(t)=\begin{cases}1 & \text{if } t\leqslant 0\\ 1-t & \text{if } 0\leqslant t\leqslant 1\\ 0 & \text{if } 1\leqslant t\end{cases} \quad \dots \quad (C.3)$$

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 $\phi\left(\frac{d(.,F)}{\epsilon}\right)$ from Y to $[0, 1]$ by $f_{\epsilon,F}$ and consider

$$\mathcal{F}:=\{f_{\epsilon,F}: F \text{ a closed subset of } Y\} \quad \dots \quad (C.4)$$

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

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

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

$$d(x,F)\leqslant d(x,y)+d(y,F)$$

and

$$d(y,F)\leqslant d(x,y)+d(x,F).$$

Hence

$$|d(x\;F)-d(y,F)|\leqslant d(x,y).$$

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

$$\{d(.\,,F): F \text{ a closed subset of } Y\} \quad \dots \quad (C.5)$$

is equicontinuous on Y.

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

From (C.3)—(C.5) 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 $\bar{\mathscr{F}}_\varepsilon$ is compact. Hence, by (C.2),

$$\sup_{\{P_i\}_{1\leqslant i\leqslant n}\in\mathscr{P}^n}\left(\prod_{i=1}^n P_i\right)\left(\left\{\sup_{f\in\mathscr{F}_\varepsilon}\left|\int d(F_n-\bar{P}_n)\right|>\epsilon\right\}\right)\to 0 \text{ as } n\to\infty \quad \ldots \quad (C.6)$$

Let us now observe that the Prohorov metric d as defined in (10) of Section 2 can easily be redefined using closed sets only, i.e., for any $P, Q\in\mathscr{P}$,

$$d(P, Q) = \inf\{\eta>0 : P(F)\leqslant Q(F^\eta)+\eta,\ Q(F)\leqslant p(F^\eta)+\eta\ \forall\ F,\ F \text{ closed}\},$$

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

$$d(P, Q)>\epsilon$$

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

$$P(F)>Q(F^\varepsilon)+\epsilon \text{ or } Q(F)>P(F^\varepsilon)+\epsilon \quad \ldots \quad (C.7)$$

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

$$\int f_{\varepsilon,F}dP \overset{(C.4)}{\geqslant} P(F) \overset{(C.7)}{>} Q(F^\varepsilon)+\epsilon \overset{(C.4)}{\geqslant} \int f_{\varepsilon,F}dQ+\epsilon$$

$$\text{or}\quad \int f_{\varepsilon,F}dQ \overset{(C.4)}{\geqslant} Q(F) \overset{(C.7)}{>} P(F^\varepsilon)+\epsilon \overset{(C.4)}{\geqslant} \int f_{\varepsilon,F}dP+\epsilon$$

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

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

$\Longrightarrow \sup_{f\in\mathscr{F}_\varepsilon}\left|\int fd(P-Q)\right|>\epsilon$, where $\mathscr{F}_\varepsilon$ is the family of continuous functions defined by (C.4).

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

$$\left(\prod_{i=1}^n P_i\right)(\{d(P, Q)>\epsilon\})\leqslant\left(\prod_{i=1}^n P_i\right)\left(\left\{\sup_{f\in\mathscr{F}_\varepsilon}\left|\int f\,d(P-Q)\right|>\epsilon\right\}\right) \quad \ldots \quad (C.8)$$

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

$$\sup_{\{P_i\}_{1\leqslant i\leqslant n}\in\mathscr{P}^n}\left(\prod_{i=1}^n P_i\right)(\{d(P, Q)>\epsilon\})$$

$$\leqslant \sup_{\{P_i\}_{1\leqslant i\leqslant n}\in\mathscr{P}^n}\left(\prod_{i=1}^n P_i\right)\left(\left\{\sup_{f\in\mathscr{F}_\varepsilon}\left|\int f\,d(P-Q)\right|>\epsilon\right\}\right) \quad \ldots \quad (C.9)$$

The result follows from (C.6) and (C.9) with $P=F_n$ and $Q=\bar{P}_n$.

The following is an immediate corollary to the lemma.

Corollary C.1.1. *Let (Y, ρ), $\mathcal{P}$, $(\xi_1, \xi_2, \ldots, \xi_n)$ and d be as considered in Lemma C.1. Let P_n^u denote the uniform distribution on s_n as defined in Convention 1 of Section 4. Then for any $\epsilon > 0$,*

$$\sup_{\{P_i\}_{1 \leqslant i \leqslant n} \in \mathcal{P}^n} \int \left(\prod_{i=1}^{n} P_{\pi(i)} \right) (\{d(F_n^O, F_n^E) > \epsilon\}) \, dP_n^u(\pi) \to 0$$

as $n \to \infty$.

Proof. In view of Lemma C.1, it is enough to show that for any $\epsilon > 0$,

$$P_n^u(\{d((\bar{P}_n^O)^*, (\bar{P}_n^E)^*) > \epsilon\}) \to 0 \text{ as } n \to \infty. \quad \ldots \text{ (C.10)}$$

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

$$\int \left\{ \frac{1}{(n-[n/2])} \sum_{\substack{i=1 \\ i \text{ odd}}}^{n} (\int f dP_{\pi(i)}) - \frac{1}{[n/2]} \sum_{\substack{i=1 \\ i \text{ even}}}^{n} (\int f dP_{\pi(i)}) \right\}^2 dP_n^u(\pi)$$

$$= \int \left\{ \frac{1}{(n-[n/2])} \sum_{i \text{ odd}} a_{\pi(i)} - \frac{1}{[n/2]} \sum_{i \text{ odd}} a_{\pi(i)} \right\}^2 dP_n^u(\pi)$$

$$\leqslant 4 \int \left\{ \frac{1}{(n-[n/2])} \sum_{i \text{ odd}} a_{\pi(i)} - \bar{a} \right\}^2 dP_n^u(\pi)$$

$$\leqslant \frac{4}{(n-[n/2])} \left\{ \frac{1}{n} \sum_{i=1}^{n} (a_i - \bar{a})^2 \right\} \quad \ldots \text{ (C.11)}$$

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

By (C.11) with arbitrary f, we note that the following analogue of (C.1) holds : namely, for any $f \in C(X)$ and $\epsilon > 0$

$$P_n^u(\{ | \int f \, d((\bar{P}_n^O)^* - (\bar{P}_n^E)^*) | > \epsilon\}) \to 0 \text{ as } n \to \infty. \quad \ldots \text{ (C.12)}$$

(C.10) follows from (C.12) exactly as in Lemma C.I.

Proof of Proposition 4.1. The given expression

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

(where * denote the operation of randomisation as defined in Definition 4.1)

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

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

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

$$\leqslant \sup_{\{G_i\}_{1 \leqslant i \leqslant n} \in G^n} \int \left(\prod_{i=1}^{n} G_{n(i)} \right) (\{d(\underline{G}_n^0, \underline{G}_n^E) > \epsilon\}) \, dP_n^u(\pi) \longrightarrow 0$$

as $n \to \infty$ by Corollary C.1.1 with $Y = \Xi$ and (hence) $P = G$.

Let ψ be a kernel. Fix θ_0 in Θ and triangular array $\{\xi_{ni}\}_{1 \leqslant i \leqslant n, n \geqslant 1}$ of elements in Ξ satisfying relation (4.4). The following are the conditions $(i)^t$–$(v)^t$ and $U(i)^t - U(vi)^t$, referred to in the discussion preceeding Lemma 4.1(t).

(i)^t (a) Condition (i) of Section 3 holds, with G_0 replaced by $\underline{G}_{n,n}^O$ or $\underline{G}_{n,n}^E$, uniformly with respect to $n \geqslant 1$ and

(b) $$\int \frac{\{f'(., \theta_0, G) - f'(., \theta_0, \underline{G}_{n,n})\}^2}{f(. \theta_0, \underline{G}_{n,n})} \, d\mu(\,.\,) \Bigg|_{G = \underline{G}_{n,n}^O \text{ or } \underline{G}_{n,n}^E} \to 0 \text{ as } n \to \infty.$$

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

(a) there is $\delta_0 > 0$ such that condition (ii) (a) of Section 3 holds with G_0 replaced by $\underline{G}_{n,n}^O$ or $\underline{G}_{n,n}^E$ and

(b) (i) $\int \{\psi(.,\theta,G) - \psi(., \theta_0, \underline{G}_{n,n}^E)\}^2 f(.,\theta_0,\underline{G}_{n,n}^O) \, d\mu(\,.\,) \longrightarrow 0$ as $(\theta, G) \longrightarrow (\theta_0, \underline{G}_{n,n}^E)$ and

(ii) $\int \{\psi(., \theta, G) - \psi(., \theta_0, \underline{G}_{n,n}^O)\}^2 f(., \theta_0, \underline{G}_{n,n}^E) \, d\mu(\,.\,) \longrightarrow 0$ as $(\theta, G) \longrightarrow (\theta_0, \underline{G}_{n,n}^O)$.

(iii)^t Condition (iii) of Section 3 holds with $(\hat{G}_n, G_0)$ replaced by $(\hat{G}_{n,n}^E, \underline{G}_{n,n}^O)$ or $(\underline{G}_{n,n}, \underline{G}_{n,n}^O)$ or $(\hat{G}_{n,n}^O, \underline{G}_{n,n}^E)$ or $(\underline{G}_{n,n}, \underline{G}_{n,n}^E)$.

(iv)^t (a) There is $\delta_0 > 0$ such that condition (iv) (a) of Section 3 holds with G_0 replaced by $\underline{G}_{n,n}$,

(b) condition (iv) (b) of Section 3 holds, with G_0 replaced by $\underline{G}_{n,n}$, by G„jn, uniformly in $n \geqslant 1$ and

(c) $\{\int \psi(., \Theta_0, \underline{G}_{n,n}) f'(., \Theta_0, \underline{G}_{n,n}) \, d\mu(.) : n \geqslant 1\}$ does not contain zero as a limit point.

(v)^t there is $\delta_0 > 0$ such that condition (v) of Section 3 holds, with G_0 replaced by $\underline{G}_{n,n}$, uniformly in $n \geqslant 1$.

Let $\delta_0 > 0$ 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_0, θ and θ' replaced by θ, θ' and θ'', respectively, uniformly with respect to θ, θ' and θ'' in $B(\theta_0, \delta_0)$, G in $B(\underline{G}_{n,n}, \delta_0)$ and $\{\xi_{ni}\}_{1 \leqslant i \leqslant n}$ in $\alpha_n(\epsilon_n)$. The Condition U(vi)^t is given below.

U(vi)^t (a) $\sup\limits_{n \geqslant 1} \sup\limits_{(\theta, \{\xi_{ni}\}_{1 \leqslant i \leqslant n}) \epsilon B(\theta_o, \delta_o) \times \alpha_n(e_n)}$

$$[\int 1_{\{|\psi(\cdot, \theta, \underline{G}_{n,n})| \geqslant K\}} (x) \psi^2 (x, \theta, \underline{G}_{n,n}) f(x, \theta, \underline{G}_{n,n}) \, d\mu(\pi) \; / J(\theta, \underline{G}_{n,n}, \psi)] \to 0$$

as $K \to \infty$ and (b) U(vi) (b) holds.

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 : Asymptotic 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, E. (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. (1987). 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.