The Annals of Statistics
1982, Vol 10, No. 3, 647–671

THE 1980 WALD MEMORIAL LECTURES

ON ADAPTIVE ESTIMATION

BY P. J. BICKEL[1]

University of California, Berkeley

We simplify a general heuristic necessary condition of Stein's for adaptive estimation of a Euclidean parameter in the presence of an infinite dimensional shape nuisance parameter and other Euclidean nuisance parameters. We derive sufficient conditions and apply them in the construction of adaptive estimates for the parameters of linear models and multivariate elliptic distributions. We conclude with a review of issues in adaptive estimation.

1. Introduction. In 1956, C. Stein published a paper in the Third Berkeley Symposium which deserves to be as well known as its celebrated companion piece on the inadmissibility of the normal mean. In this work Stein dealt with the problem of estimating and testing hypotheses about a Euclidean parameter θ or, more generally, a function $q(\theta)$ in the presence of an infinite dimensional "nuisance" shape parameter G. The question he asked (framed in estimation terms) was, "When can one estimate θ as well asymptotically not knowing G as knowing G?" He gave a simple necessary condition, which he checked in several important examples and, in one of these—testing that the center of symmetry has a specified value—he indicated a procedure that should work.

In recent years there has been considerable interest in an important situation where Stein's condition is satisfied, estimating the center of symmetry of an unknown symmetric distribution. Completely definitive results for this problem were obtained by Beran (1974) and Stone (1975). In this paper we return to Stein's original general formulation in the i.i.d. case. Motivated by his necessary condition for existence of adaptive estimates we obtain a simple sufficient condition for adaptation and apply it to a variety of important examples.

The paper is organized as follows. In Section 2 we define what we mean by adaptive estimation of θ; more precisely, we review some known results in the area and introduce the examples with which we will deal. In Section 3 we recall Stein's necessary condition for adaptation, and introduce a condition which we prove is sufficient. In Section 4 we check that our sufficient condition is satisfied in our examples. Section 5 contains a discussion of the connections between our work and recent research of Lindsay (1978, 1980), Hammerstrom (1978), Levitt (1974) and others, as well as a discussion of open questions. Finally, in Section 6, we gather technical parts of the proofs of our results.

2. What is adaptation? For simplicity we restrict ourselves throughout to the i.i.d. case. This is quite unnecessary for the heuristics of the paper. However, at least some of our proofs employ the assumed independence of the observations quite heavily.

Let $X_1, \cdots, X_n$ be i.i.d. k dimensional vectors with common distribution F. Let us recall the basic facts about the asymptotic theory of estimation when F ranges over a parametric model as put into their most elegant form by Le Cam.

Suppose that F is of the form F_θ where $\theta \in \Theta$, an open subset of R^p, and the F_θ have densities which we denote by $f(\cdot, \theta)$ with respect to a sigma-finite measure μ on R^k. Write

Received March 1981; revised November 1981.

[1] Research partially supported by ONR Grant No. N00014-80-C-0163 and the Adolph C. and Mary Sprague Miller Foundation for Basic Research in Science.

AMS 1970 *subject classification.* 62F20, 62G20.

Key words and phrases. Adaptation, efficient estimation, linear models, elliptic distributions.

E_θ, P_θ, $\mathscr{L}_\theta$ respectively for expectations, probabilities, and laws when θ holds. Let $\ell(x, \theta) = \log f(x, \theta)$, and define the following regularity conditions.

CONDITIONS R. *For all* $\theta \in \Theta$,

(i) $\ell(\cdot, \theta)$ *is differentiable in (the components of)* θ a.e. P_θ *and* $\dot\ell = (\partial\ell/\partial\theta_1, \cdots, \partial\ell/\partial\theta_p)$.

(ii) *The Fisher information matrix* $I(\theta)$ *exists*, $I(\theta) = E_\theta\{\dot\ell^T\dot\ell(X_1, \theta)\} < \infty$;

(iii) *Square root likelihood is differentiable in quadratic means, i.e. as* $t \to 0$,

$$E_\theta\left[\left\{\frac{f(X_1, \theta + t)}{f(X_1, \theta)}\right\}^{1/2} - 1 - \frac{t}{2}\dot\ell^T(X_1, \theta)\right]^2 = o(|t|^2),$$

and

$$P_{\theta+t}\{f(X_1, \theta) = 0\} = o(|t|^2),$$

where $|\cdot|$ *denotes the Euclidean norm (cf.* b_1 *and* b_2 *on page* 10 *of Le Cam,* 1969).

(iv) *There exist* $n^{1/2}$*-consistent estimates of* θ, *i.e.* $\{\tilde\theta_n(X_1, \cdots, X_n)\}$ *such that* $n^{1/2}(\tilde\theta_n - \theta) = O_{P_\theta}(1)$.

Under these conditions the following theorem holds (Le Cam, 1969; Fabian and Hannan, 1980). Call θ a *regular* point if $I(\theta)$ is nonsingular and if $I(\cdot)$ is continuous at θ.

THEOREM 2.1. *Under Conditions* R *there exist estimates* $\{\hat\theta_n\}$ *such that*

(a) *For all regular,* θ, $\mathscr{L}_{\theta_n}\{n^{1/2}(\hat\theta_n - \theta_n)\} \to \mathscr{N}(0, I^{-1}(\theta))$ *whenever* $n^{1/2}|\theta_n - \theta| \le M$ *for all* n, $M < \infty$.

(b) *The estimates* $\{\hat\theta_n\}$ *are asymptotically locally sufficient in the sense of Le Cam* (1969) *and locally asymptotically minimax in the sense of Hájek* (1972) *as modified by Fabian and Hannan* (1980).

Statement (a) says that $\{\hat\theta_n\}$ are efficient in the usual sense. Hájek (1972) also establishes, for $k = 1$, that any estimates satisfying (a) also are efficient in the sense of Rao. That is, if we define $\Delta_n(\cdot)$ by

$$\hat\theta_n = \theta + n^{-1}\textstyle\sum_{i=1}^n \dot\ell(X_i, \theta)I^{-1}(\theta) + \Delta_n(\theta), \tag{2.1}$$

then

$$n^{1/2}\Delta_n(\theta) \to_{P_\theta} 0, \tag{2.2}$$

for θ_n as in the theorem. In Theorem 6.1 (Section 6.4) we extend this result to general k.

REMARK 1. The construction of $\hat\theta_n$ used by Le Cam will prove useful to us later. Let $R_n^k = \{n^{-1/2}(i_1, \cdots, i_k), i_1, \cdots, i_k$ are arbitrary integers$\}$, and let

$$\bar\theta_n = \text{the point in } R_n^k \text{ closest to } \tilde\theta_n. \tag{2.3}$$

If $\dot\ell^*(\mathrm{x}, \theta)$ has the property that

$$n^{-1/2}\textstyle\sum_{i=1}^n \{\dot\ell^*(X_i, \theta_n) - \dot\ell(X_i, \theta)\} + n^{1/2}(\theta_n - \theta)I(\theta) = o_{P_\theta}(1)$$

whenever $n^{1/2}|\theta_n - \theta| \le M$, then Theorem 4 of Le Cam (1969) shows that

$$\hat\theta_n = \bar\theta_n + n^{-1}\textstyle\sum_{j=1}^n \dot\ell^*(X_j, \bar\theta_n)I^-(\bar\theta_n) \tag{2.4}$$

is efficient in the sense of Theorem 2.1; where I^- is a generalized inverse of I. Of course, this construction is not unique and has unpleasant aspects such as the "discretization" of $\tilde\theta_n$ and its non-iterative character. However, the construction works in great generality, i.e., under the mild and natural Conditions R(i)–R(iv).

We shall actually want to take $\dot\ell^* = \dot\ell$. To do so we need an inconsequential strengthening of R(iii) which is valid in all our examples. We call UR(iii) the assumption that for all θ

$\in \Theta$, the differentiability condition of R(iii) holds uniformly in some neighbourhood of θ. We show in Theorem 6.2 (Section 6.4) that R(i), R(ii) and UR(iii) enable us to take $\dot{\ell}^* = \dot{\ell}$ in (2.4).

REMARK 2. Condition R(iv), although clearly necessary, appears hard to verify. In fact, Le Cam shows that if we assume identifiability of θ and nonsingularity of $I(\theta)$ for all $\theta \in \Theta$, R(i)-R(iii) imply R(iv). We have chosen to leave R(iv) in its present form for reasons which will be apparent later.

In a preprint which we saw after our lectures were prepared, Fabian and Hannan (1980) give a very careful treatment of estimation in locally asymptotically normal families. They present, among other results, the "right" version of Hájek's local asymptotic minimaxity, as well as a rigorous discussion of Stein's (1956) necessary conditions for adaptation. Their notion of adaptation agrees with ours (in their more general framework).

The models for which we will discuss adaptation may be described as follows: The common d.f. F of the X_i ranges over a set which can be parametrized by a Euclidean parameter θ of interest, and a shape nuisance parameter G, i.e.,

$$\mathscr{F} = \{F_{(\theta,G)} : \theta \in \Theta,\ G \in \mathscr{G}\} \tag{2.5}$$

where Θ is an open subset of R^p, $\mathscr{G}$ is a set of distributions on some space, and the map $(\theta, G) \to F_{(\theta,G)}$ is known.

For each $G \in \mathscr{G}$, define

$$\mathscr{F}_G = \{F_{(\theta,G)} : \theta \in \Theta\}. \tag{2.6}$$

The models $\mathscr{F}_G$ are parametric models. Suppose that $\mathscr{F}_G$ satisfies R(i), R(ii) and UR(iii) for each $G \in \mathscr{G}$. Define $f(\cdot, \theta, G)$, $\ell(\cdot, \theta, G)$, $I(\theta, G)$ respectively as density, log likelihood, and information in $\mathscr{F}_G$. Call (θ, G) *regular* if θ is regular in $\mathscr{F}_G$. Finally, in view of the Le Cam theorem, we can state the following definition.

DEFINITION. *A sequence of estimates $\{\hat{\theta}_n\}$ is adaptive if and only if, for every regular (θ, G),*

$$\mathscr{L}_{\theta_n}\{n^{1/2}(\hat{\theta}_n - \theta_n)\} \to \mathscr{N}(0, I^{-1}(\theta, G)) \tag{2.7}$$

whenever $n^{1/2} | \theta_n - \theta |$ stays bounded. Thus adaptive estimates, if they exist, are efficient for every $\mathscr{F}_G$ even though knowledge of the true G may not be used in the construction of the estimates.

Adaptive estimates of θ have been constructed in the first of our examples.

EXAMPLE 1. *Estimation of the center of symmetry.* Let $k = p = 1$. Take $\Theta = R$, $\mathscr{G}$ = {All distributions symmetric about 0}, $F_{(\theta,G)}(x) = G(x - \theta)$.

The problem of adaptive estimation of θ in this model began to be studied by van Eeden (1970) and Takeuchi (1971), although the corresponding testing problem was earlier considered by Stein (1956) and solved by Hájek (1962). The definitive theorem was obtained by Beran (1974) and Stone (1975).

Let

$$I(G) = \int \{g'(x)\}^2/g(x)\, dx \tag{2.8}$$

whenever g, the density of G, is absolutely continuous, and let $I(G) = \infty$ otherwise.

THEOREM 2.2. *There exist translation and scale equivariant estimates, $\{\hat{\theta}_n\}$ such that*

$$\mathscr{L}_{(0,G)}(n^{1/2}\hat{\theta}_n) \to \mathscr{N}(0, I^{-1}(G)) \tag{2.9}$$

for all $G \in \mathscr{G}$ with $I(G) < \infty$.

Hájek (1962) has shown that for this model (θ, G) is regular if $I(\theta, G) = I(G) < \infty$. The converse is also true. Thus $\{\hat{\theta}_n\}$ are adaptive according to our general definition. In fact, Stone (1975) shows that the estimates he constructs satisfy (2.9) with $I^{-1}(G) = 0$ whenever $I(G) = \infty$. □

We will construct adaptive estimates of θ in the following generalization of Example 1.

EXAMPLE 2. *Estimation of regression with symmetric errors.* We describe the model structurally in terms of a variable $X \sim F_{(\theta, G)}$. Here $k = p + 1$ and $\Theta = R^p$. Let

$$X = (C, Y) \tag{2.10}$$

where C is a p dimensional random vector and Y a scalar. Further,

$$Y = C\theta^T + \varepsilon \tag{2.11}$$

where $\varepsilon \sim G$, and ε and C are independent. We again take

$$\mathscr{G} = \{\text{All distributions } G \text{ on } R \text{ symmetric about } 0\}.$$

Finally, we suppose

$$E(C^T C) \text{ is nonsingular.} \tag{2.12}$$

This is just a stochastic version of the usual multiple regression model,

$$X_i = C_i\theta^T + \varepsilon_i, \qquad i = 1, \cdots, n,$$

where $C_1, \cdots, C_n$ are p dimensional vectors of constants such that $C^T = (C_1^T, \cdots, C_n^T)$ and $C^T C$ is nonsingular.

We deliberately do not specify that the distribution of C is known. The adaptive estimates we construct depend only on the data and work for any distribution of C satisfying (2.12). □

In many interesting situations a parameter θ for which efficient estimates exist in every model $\mathscr{F}_G$ cannot be consistently estimated in $\mathscr{F}$ because the parameter becomes unidentifiable. This is true in the next two examples. However, in both, natural functions $q(\theta)$ can be so estimated. In fact, adaptive estimation of these functions is possible. The definition of adaptive estimation of q is straightforward:

DEFINITION. *Suppose $q : \Theta \to R^d$, $d \le p$, has a total differential $\dot{q}(\theta)$, a $d \times p$ matrix. A sequence of estimates $\{\hat{q}_n\}$ of q is adaptive if and only if, for every regular (θ, G),*

$$\mathscr{L}_{\theta_n}\{n^{1/2}(q_n - q(\theta_n))\} \to \mathscr{N}(0, \dot{q}(\theta)I^{-1}(\theta, G)\dot{q}(\theta)^T) \tag{2.13}$$

whenever $n^{1/2} \mid \theta_n - \theta \mid$ stays bounded.

EXAMPLE 3. *Regression with a constant and arbitrary errors.* In Example 2, let $C = (C^\circ, 1)$, C° a $p - 1$ dimensional vector. Define X, Y, ε as before and suppose ε and C are independent. However, let $\mathscr{G} = \{\text{all distributions on } R\}$, and replace (2.12) by

$$E(C^\circ - EC^\circ)^T(C^\circ - EC^\circ) \quad \text{nonsingular.} \tag{2.14}$$

Evidently θ is not identifiable in $\mathscr{F}$ since a change in the constant θ_p could equally well be a change in G. However, $q(\theta) = (\theta_1, \cdots, \theta_{p-1})$ can be adaptively estimated, as we shall see.

A special case of this model, where $p = 2$ and

$$C^\circ = \begin{cases} 1 & \text{with probability } \lambda \\ 0 & \text{with probability } 1 - \lambda, \end{cases}$$

can be thought of as a two-sample model with random sample sizes, i.e., we observe N observations with distribution $G(x - \theta_1 - \theta_2)$ and $n - N$ observations with distribution $G(x - \theta_2)$, where N has a binomial (n, λ) distribution.

Adaptation in the two-sample model with fixed sample sizes (and unknown scale) was studied by Stein (1956), Weiss and Wolfowitz (1970), and Wolfowitz (1974). A definitive result was obtained by Beran (1974). Weiss and Wolfowitz (1971) considered the fixed sample size multiple regression model and obtained partial results. □

EXAMPLE 4. *Parameters of elliptic distributions.* The following multivariate generalization of the symmetric one-sample location and scale model has been considered by Huber (1977) and others. Let

$$X = \mu + \varepsilon V^{-1/2}$$

where μ is an unknown $1 \times k$ vector, V is a positive definite $k \times k$ symmetric matrix, and $V^{-1/2}$ is the unique positive definite symmetric square root of V^{-1}. We suppose $\varepsilon \sim G$, where

$$\mathscr{G} = \{G : G \text{ absolutely continuous, spherically symmetric on } R^k\}.$$

Take $\theta = (\mu, [V])$ where for any symmetric $k \times k$ matrix $M = \| m_{ij} \|$, we define $[M]$ to be the lexicographically written row vector of the lower $k(k+1)/2$ entries of M. Thus, $p = k(k+3)/2$ and

$$\Theta = \{(\mu, [V]) : V \text{ symmetric positive definite}\}$$

is an open subset of R^p.

Here θ is efficiently estimable at regular points of $\mathscr{F}_G$ but is not identifiable in $\mathscr{F}$. A common scale change in all coordinates is ascribable to either V or G, yet $(\mu, V/\text{tr } V)$ can be estimated consistently, in fact, adaptively, as we shall see.

3. Stein's considerations and a sufficient condition for adaptation. We begin by recalling Stein's necessary condition for adaptation. Define a parametric subfamily of $\mathscr{G}$ as a set $\{\mathscr{G}_\eta\}$, $\eta \in T$, where T is an open set in R^t and the map $\eta \to G_\eta$ is smooth. The parametric submodel of $\mathscr{F}$ corresponding to the parametric subfamily $\{G_\eta\}$ is naturally defined by $\{F_{(\theta, G_\eta)} : \theta \in \Theta, \eta \in T\}$. Here is Stein's necessary condition.

CONDITION S. *For every parametric submodel obeying* R(i)–R(iv) *with* $G_{\eta_0} = G_0$

$$\int \left\{ \frac{\partial}{\partial \theta_i} \ell(x, \theta, G_\eta) \frac{\partial}{\partial \eta_j} \ell(x, \theta, G_\eta) \right\}_{\theta=\theta_0, \eta=\eta_0} f(x, \theta_0, G_0)\mu(dx) = 0 \tag{3.1}$$

$$i = 1, \cdots, p, \quad j = 1, \cdots, t.$$

Stein (1956) shows that if an adaptive estimate of θ exists and (θ_0, G_0) is regular, then Condition S must hold. The argument is simple. Let

$$I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix},$$

where I_{11} is $p \times p$ and I_{22} is $t \times t$, be the $(p+t) \times (p+t)$-dimensional Fisher information matrix of the parametric submodel $F_{(\theta, G_\eta)}$ evaluated at (θ_0, η_0), and write

$$I^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix}.$$

Now, by definition, if $\{\hat\theta_n\}$ is adaptive, then $I_{11}^{-1} = I^{-1}(\theta_0, G_0)$ is the asymptotic variance covariance matrix of $n^{1/2}(\hat\theta_n - \theta_n)$ whenever $n^{1/2} | \theta_n - \theta |$ stays bounded. But, by Hájek's (1972) theorem, I^{11} is the smallest variance covariance matrix achievable in this way. Thus $I_{11}^{-1} = I^{11}$ which is equivalent to $I_{12} = 0$, which is Condition S.

Condition S suffers from two defects: (i) it can be awkward to verify, (ii) it is unclear how to proceed from it to the construction of adaptive procedures. We now proceed to derive a simpler condition which is at least heuristically necessary and which in turn leads to a verifiable sufficient condition.

All the examples we have studied exhibit the following simple convexity structure:

CONDITION C. *$\mathscr{G}$ is convex and $G_0, G_1 \in \mathscr{G}$ implies that for $0 \le \alpha \le 1$*

$$F_{(\theta,\alpha G_0+(1-\alpha)G_1)} = \alpha F_{(\theta,G_0)} + (1-\alpha)F_{(\theta,G_1)}.$$

This structure suggests that we examine Condition S for the following $\{G_\eta\}$. Fix G_0 and G_1, take $T = (0, 1)$, and let

$$G_\eta = \eta G_0 + (1-\eta)G_1.$$

Then Condition S becomes for $\eta > 0$, $i = 1, \cdots, p$,

$$\int \frac{\partial}{\partial\theta_i} \ell(x, \theta, G_\eta)\{f(x, \theta, G_1) - f(x, \theta, G_0)\}\mu(dx) = 0.$$

Letting $\eta \to 0$ formally we get for "all" $G_0, G_1 \in \mathscr{G}$ that the following holds.

CONDITION S*.

$$\int \dot{\ell}(x, \theta, G_0)f(x, \theta, G_1)\mu(dx) = 0.$$

It may be shown formally that if Condition S* holds, so does Condition S (Bickel, 1979). Condition S* has a simple heuristic interpretation. If G_0 is a fixed shape in $\mathscr{G}$ let θ_n^* be the M-estimate corresponding to G_0, i.e., solving

$$\textstyle\sum_{i=1}^n \dot{\ell}(x_i, \theta_n^*, G_0) = 0.$$

We know that, under regularity conditions (Huber, 1967), if Condition S* holds, then $n^{1/2}(\theta_n^* - \theta)$ is asymptotically normal under $F_{(\theta,G)}$ with mean 0 and variance covariance matrix $A^{-1}B(A^T)^{-1}$, where

$$A = \left\| -\int \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell(x, \theta, G_0)f(x, \theta, G)\mu(dx)\right\|,$$

(3.2)

$$B = \int \dot{\ell}^T(x, \theta, G_0)\dot{\ell}(x, \theta, G_0)f(x, \theta, G)\mu(dx).$$

A heuristic summary of this is as follows. Firstly, M-estimates corresponding to a fixed shape G_0 should be $n^{-1/2}$ consistent for θ under every shape G_1. Secondly, suppose we can estimate the true G by data-dependent $\{G_n\}$ so that the score functions $\ell(\cdot, \cdot, G_n)$ converge to $\ell(\cdot,\cdot, G)$ and so that the matrices A_n, B_n obtained by replacing G_0 by G_n in (3.2) converge to $I(\theta, G)$. It then seems plausible that the sequence of M-estimates corresponding to G_n is adaptive.

Motivated by these considerations we now formulate two conditions, GR(iv) and H.

CONDITION GR(iv). *There exist estimates $\{\tilde{\theta}_n\}$ such that $n^{1/2}(\tilde{\theta}_n - \theta) = O_{P_{(\theta,G)}}(1)$ at all regular points (θ, G).*

Let

(3.3)

$$\mathscr{H} = \{h: h \text{ maps } R^k \times \Theta \text{ to } R^k \text{ and } \int h(x, \theta)F_{(\theta,G)}(dx) = 0 \text{ for all } \theta \in \Theta, G \in \mathscr{G}\}.$$

In view of Condition S*, $\mathscr{H}$ includes the space of possible score functions. For convenience we introduce

$$\tilde{\ell}(x, \theta, G) = \dot{\ell}(x, \theta, G)I^{-}(\theta, G), \tag{3.4}$$

where I^{-} is any generalized inverse. (In fact we only need $\tilde{\ell}$ for θ such that $I(\theta, G)$ is nonsingular.) Note that $\tilde{\ell}$ can be substituted for $\dot{\ell}$ in Condition S*. Here is our main condition:

CONDITION H. *Appropriate consistent estimation of score functions is possible. That is, there exists a sequence of maps* $\hat{\tilde{\ell}}_m : (R^k)^m \to \mathscr{H}$, $m = 1, 2, \cdots$, *taking* $(x_1, \cdots, x_m)$ *into* $\hat{\tilde{\ell}}(\cdot, \cdot; x_1, \cdots, x_m)$ *such that for all regular* (θ, G) *and any* $|\theta_m - \theta| = O(m^{-1/2})$,

$$\int |\hat{\tilde{\ell}}_m(x, \theta_m; X_1, \cdots, X_m) - \tilde{\ell}(x, \theta_m, G)|^2 F_{(\theta_m, G)}(dx) \to 0 \tag{3.5}$$

in $P_{(\theta, G)}$ *probability.*

Note that GR(iv) is evidently a necessary condition for adaptive estimation and is the natural generalization of R(iv). Under Condition S*, M-estimates corresponding to a fixed shape are natural candidates for $\bar{\theta}_n$. In view of Stein's necessary Condition S*, we conjecture that Condition H is necessary for adaptation. W. R. van Zwet pointed out a suggestive inequality bolstering this conjecture (Klaassen, 1980, Theorem 3.2.1). In any case these conditions are sufficient.

THEOREM 3.1. *If Conditions* GR(iv) *and* H *hold, then adaptive estimates exist.*

NOTE. The construction is closely related to that given for adaptive rank tests in the linear model by Hájek (1962). A related construction for Example 1 has been given by Bretagnolle (private communication). See also Hasminskii and Ibragimov (1978).

PROOF. Define $\bar{\theta}_n$ as in (2.3). Let $\{m(n)\}$ be a sequence of subsample sizes with $m(n) = o(n)$. Write m for $m(n)$ and let $\bar{n} = n - m$.
Define

$$\hat{\theta}_n = \bar{\theta}_n + \bar{n}^{-1} \textstyle\sum_{i=m+1}^{n} \hat{\tilde{\ell}}(X_i, \bar{\theta}_n; X_1, \cdots, X_m). \tag{3.6}$$

We claim $\{\hat{\theta}_n\}$ is adaptive. By Theorem 6.2,

$$\bar{\theta}_n + \bar{n}^{-1} \textstyle\sum_{i=m+1}^{n} \tilde{\ell}(X_i, \bar{\theta}_n, G)$$

is efficient for every regular (θ, G). Write P_θ for $P_{(\theta, G)}$. Then to prove the theorem it is enough to show

$$\bar{n}^{-1/2} \textstyle\sum_{i=m+1}^{n} \{\hat{\tilde{\ell}}_m(X_i, \bar{\theta}_n; X_1, \cdots, X_m) - \tilde{\ell}(X_i, \bar{\theta}_n, G)\} = o_{P_\theta}(1). \tag{3.7}$$

Now we use a trick of Le Cam's and note that we need only establish (3.7) with $\bar{\theta}_n$ replaced by $\theta_n = \theta + t_n n^{-1/2}$, where t_n is an arbitrary convergent deterministic sequence. This follows since $\bar{\theta}_n$ is $\sqrt{n}$ – consistent and the intersection of its range with any sphere of radius $Mn^{-1/2}$ about θ is finite with cardinality bounded independent of n. Having made the replacement, we prove (3.7). Note that R(i) – R(iii) imply that the $\bar{n}$ dimensional product measures of $X_{m+1}, \cdots, X_n$ under P_θ and under P_{θ_n} are contiguous. Therefore, it suffices to prove (3.7) in P_{θ_n} probability. Condition on $X_1, \cdots, X_m$ for this probability. Since $\hat{\tilde{\ell}}(\cdot, \cdot; X_1, \cdots, X_m) \in \mathscr{H}$,

$$\int \hat{\tilde{\ell}}_m(x, \theta_n; X_1, \cdots, X_m) f(x, \theta_n, G)\mu(dx) = 0 \tag{3.8}$$

and by R(i) − R(iii),

$$\int \tilde{\ell}(x, \theta_n, G) f(x, \theta_n, G)\mu(dx) = 0. \tag{3.9}$$

Therefore

$$\begin{aligned} &E_{\theta_n}[\,|\bar{n}^{-1/2} \textstyle\sum_{i=m+1}^{n} \{\hat{\tilde{\ell}}_m(X_i, \theta_n; X_1, \cdots, X_m) - \tilde{\ell}(X_i, \theta_n, G)\}|^2 \,|\, X_1, \cdots, X_m] \\ &\quad = \int |\hat{\tilde{\ell}}_m(x, \theta_n; X_1, \cdots, X_m) - \tilde{\ell}(x, \theta_n, G)|^2 f(x, \theta_n, G)\mu(dx) \to 0 \end{aligned} \tag{3.10}$$

in P_θ probability by Condition H and hence, by contiguity again, in P_{θ_n} probability. Claim (3.7) is proved, and the theorem follows. □

NOTES. It is possible to replace Condition H by the following condition H′ which permits separate estimation of $\dot{\ell}$ and I^{-1}.

CONDITION *H′*. (a) *There exist maps* $\hat{\dot{\ell}}_m (R^k)^m \to \mathscr{H}$ *such that for all regular* (θ, G), $|\theta_m - \theta| = O(m^{-1/2})$

$$\int |\hat{\dot{\ell}}_m(x, \theta_m; X_1, \cdots, X_m) - \dot{\ell}(x, \theta_m, G)|^2 f(x, \theta_m, G)\mu(dx) = o_{P_\theta}(1). \tag{3.11}$$

(b) *There exist estimates* $\hat{I}_m(X_1, \cdots, X_m)$ *of* $I(\theta, G)$ *consistent for all regular* (θ, G).

It is easy to show that if GR(iv) and H′ both hold, and if we define

$$\theta_n^* = \bar{\theta}_n + \bar{n}^{-1} \textstyle\sum_{i=m+1}^{n} \hat{\dot{\ell}}(X_i, \bar{\theta}_n; X_1, \cdots, X_m)\hat{I}_n^{-} \tag{3.12}$$

then

$$\theta_n^* = \bar{\theta}_n + \bar{n}^{-1} \textstyle\sum_{i=m+1}^{n} \hat{\dot{\ell}}(X_i, \bar{\theta}_n; X_1, \cdots, X_m) I^{-1}(\theta, G) + o_{P_\theta}(n^{-1/2}) \tag{3.13}$$

and θ_n^* is adaptive.

A natural choice of $\hat{I}_n$ is provided by

$$\hat{I}_n = \bar{n}^{-1} \textstyle\sum_{i=m+1}^{n} \hat{\dot{\ell}}^T \hat{\dot{\ell}}(X_i, \theta_n; X_1, \cdots, X_m) \tag{3.14}$$

We show in Section 6.2 that this choice of $\hat{I}_n$ is consistent for regular (θ, G) provided that GR(iv) and (3.11) hold, and if

$$m^{-1} \textstyle\sum_{i=1}^{m} \dot{\ell}^T \dot{\ell}(X_i, \theta_m, G) \to I(\theta, G) \tag{3.15}$$

in P_θ probability for all regular (θ, G).

These are the results we will apply to Example 2 and which are applicable to other situations where all of θ is estimable. To deal with Examples 3 and 4 we need an extension of our theory. First we study the analogue of Condition S* when we only ask that $q(\theta)$, rather than all of θ, be estimated adaptively. Stein considers this question in a slightly different formulation. He writes $\theta = (q, t)$ with $q = q(\theta)$ and t, the rest of θ, is a nuisance parameter, and he introduces the model $\{F_{(\theta, G_\eta)}\}$. He notes that adaptive estimation of q is possible only if the upper left-hand corner of the inverse of the information matrix for (q, t) with $\eta = \eta_0$ fixed is the same as the upper left-hand corner of the inverse of the information matrix for (q, t, η) evaluated at η_0. We do not pursue further his matrix formulation of this condition, but only note that in the presence of convexity Condition C, Stein's condition is heuristically equivalent to the d equations

CONDITION S* (*generalized*).

$$\int \dot{\ell}(x, \theta, G_0) I^{-1}(\theta, G_0) \dot{q}^T(\theta) f(x, \theta, G_1)\mu(dx) = 0$$

for every shape $G_0, G_1 \in \mathscr{G}$. For $q(\theta) = \theta$, $\dot q$ is the identity and our more general formulation of S* agrees with our old one.

New difficulties are introduced by the possible lack of identifiability of θ. Of course we need to have q identifiable. That is, if

$$F_{(\theta_0, G_0)} = F_{(\theta_1, G_1)} = F \tag{3.16}$$

then

$$q(\theta_0) = q(\theta_1).$$

But adaptation requires more. If F can be embedded in both $\mathscr{F}_{G_0}$ and $\mathscr{F}_{G_1}$ as in (3.16), then the information bound for estimation of q must be the same in both parametric families. That is, (3.16) implies

$$\dot q(\theta_0) I^-(\theta_0, G_0)\dot q^T(\theta_0) = \dot q(\theta_1) I^-(\theta_1, G_1)\dot q^T(\theta_1). \tag{3.17}$$

This condition is satisfied in all our examples because if $\mathscr{F}_{G_0}$ and $\mathscr{F}_{G_1}$ have a member in common then they are the same, or, rather, one is a smooth relabelling of the other. For instance, in Example 3, (3.16) holds if and only if G_1 is obtained from G_0 by a translation. We shall use this structural feature in a stronger way to reduce $\mathscr{G}$ and make θ identifiable. Here is a formal statement of our structural assumptions. They are obviously satisfied in Examples 3 and 4.

ASSUMPTION A1. *Either* $\mathscr{F}_{G_0} = \mathscr{F}_{G_1}$ *or* $\mathscr{F}_{G_0} \cap \mathscr{F}_{G_1} = \varnothing$, *for all* $G_0, G_1 \in \mathscr{G}$.

ASSUMPTION A2. *There exists* $T \subset R^{p-d}$ *and a smoothly invertible map from* Θ *to* $Q \times T$ *where* $Q = q(\Theta)$ *which carries* θ *into* $(q(\theta), t(\theta))$. *That is, we can identify* q *with a piece of* θ.

ASSUMPTION A3. *If we replace* θ *by* (q, t) *and* $\mathscr{F}_{G_0} = \mathscr{F}_{G_1}$ *there exists a unique smoothly invertible mapping* $\tau(q, \cdot)$ *of* T *into itself defined by* $F_{(q,t,G_0)} = F_{(q,\tau,G_1)}$.

Assumption A1 implies that there exists an "identifying subset" $\mathscr{G}_0 \subset \mathscr{G}$ such that (i) $\mathscr{F} = \{F_{(\theta,G)}: G \in \mathscr{G}_0, \theta \in \Theta\}$, and (ii) θ is identifiable when G is restricted to $\mathscr{G}_0$ provided that it is identifiable in each $\mathscr{F}_G$. We can select $\mathscr{G}_0$ as a set of representatives of the equivalence classes generated by the relation $G_1 \equiv G_2 \Leftrightarrow \mathscr{F}_{G_1} = \mathscr{F}_{G_2}$. For instance, in Example 3 we can take $\mathscr{G}_0 = \{G: \mu(G) = 0\}$ where μ is a location parameter. As we noted, Assumptions A2 and A3 imply that if (a) $\mathscr{F}_G = \mathscr{F}_{G_0}$, $G_0 \in \mathscr{G}_0$, and (b) $F_{(\theta,G)} = F_{(\theta_0,G_0)}$, then $q(\theta) = q(\theta_0)$ and (3.16) holds. That is, it does not matter in which parametric model $\mathscr{F}_G$ we embed a distribution F. The value of q and the ease with which q can be estimated remain the same. Since we can talk about estimation of θ for $(\theta, G) \in \Theta \times \mathscr{G}_0$ it is natural to propose the following extensions of the conditions for $\sqrt{n}$-consistency and appropriate consistent estimation of score functions.

GENERALIZED CONDITION GR(iv). *There exists* $\mathscr{G}_0$ *satisfying* (i) *and* (ii) *above and estimates* $\{\tilde\theta_n\}$ *such that*

$$n^{1/2}(\tilde\theta_n - \theta) = O_{P_{(\theta,G)}}(1)$$

for all (θ, G), $G \in \mathscr{G}_0$.

We now redefine $\tilde\ell$, $\mathscr{H}$ *for given* q. Our definitions agree with the old ones when q is the identity. Let

$$\mathscr{H} = \left\{ h: h \text{ maps } R^k \times \Theta \text{ into } R^d \text{ so that} \right.$$
$$\left. \int h(x, \theta) f(x, \theta, G)\mu(dx) = 0 \quad \text{for all} \quad (\theta, G) \right\}.$$

$$\tilde\ell(x, \theta, G) = \ell(x, \theta, G) I^-(\theta, G)\dot q^T(\theta). \tag{3.18}$$

Condition H is now generalized as was condition GR(iv), merely by substituting $\mathscr{G}_0$ for $\mathscr{G}$. The easy extension of Theorem 3.1 is as follows.

THEOREM 3.2. *If Assumptions* A1–A3 *and the generalized conditions* GR(iv) *and* H *hold, then adaptive estimates* $\{\hat{q}_n\}$ *of* $q(\theta)$ *exist.*

The proof is the same as for Theorem 3.1 when we propose as estimate

$$\hat{q}_n = q(\bar{\theta}_n) + \bar{n}^{-1} \sum_{i=m+1}^{n} \hat{\dot{\ell}}_m(X_i, \bar{\theta}_n; X_1, \cdots, X_m). \tag{3.19}$$

4. Adaptation in Examples 1–4. For the examples we leave verification of the trivial structural Assumptions A1 through A3 to the reader. In each example we shall proceed through the following steps:

Step A. Formally verify Stein's orthogonality Condition S* and in the process construct what we can think of as the "space of possible score functions" $\mathscr{H}$ or a suitable subset $\mathscr{H}_0$.

Step B. Find a suitable identifying subset $\mathscr{B}_0$ and construct $\sqrt{n}$-consistent estimates $\{\bar{\theta}_n\}$ so as to satisfy GR(iv).

Step C. Construct score function estimates $\hat{\dot{\ell}}$ satisfying (3.5) and taking values in $\mathscr{H}_0$ i.e. satisfy Condition H for the appropriate consistent estimation of score functions, or satisfy its modification H′ providing for separate estimation of $\dot{\ell}$ and I.

Since Example 1 is a special case of Example 2 and has already been dealt with satisfactorily, we begin with Example 2. For convenience from now on we write P for P_θ.

EXAMPLE 2. *Step A.* If the distribution of C has density r with respect to some ν, and if G has density g, then $X = (C, Y)$ has density (with respect to the product measure)

$$f(c, y, \theta, G) = r(c)g(y - c\theta^T), \tag{4.1}$$

and

$$\dot{\ell}(c, y, \theta, G) = c\frac{g'}{g}(y - c\theta^T). \tag{4.2}$$

Then

$$E_{(\theta,G_0)}\dot{\ell}(C, Y, \theta, G) = E_{(\theta,G_0)}\left\{C\frac{g'(\varepsilon)}{g(\varepsilon)}\right\} = E(C)E_{G_0}\left\{\frac{g'(\varepsilon)}{g(\varepsilon)}\right\} = 0,$$

since g'/g is antisymmetric and G_0 is symmetric about 0. Thus, Condition S* is satisfied and by our argument, $\mathscr{H} \supset \mathscr{H}_0$ where $h \in \mathscr{H}_0$ if and only if

$$h(c, y, \theta) = c\psi(y - c\theta^T) \tag{4.3}$$

for ψ bounded and antisymmetric, i.e.

$$\psi(y) = -\psi(-y). \tag{4.4}$$

So we will use score function estimates of the form (4.3).

Step B. Let $\psi: R \to R$ be such that ψ is twice continuously differentiable, with ψ and its derivatives bounded. Suppose, moreover, that $\psi' > 0$ and that ψ is antisymmetric. Let $\{\tilde{\theta}_n\}$ be the M-estimates corresponding to ψ, i.e., the unique solutions of

$$\sum_{i=1}^{n} C_{ij}\psi(Y_i - C_i\tilde{\theta}_n^T) = 0, \quad j = 1, \cdots, p, \tag{4.5}$$

where $X_i = (C_i, Y_i)$, $C_i = (C_{i1}, \cdots, C_{ip})$. Then by Huber's theorem (Huber, 1973), $\{\tilde{\theta}_n\}$ are $\sqrt{n}$-consistent. (This is just the construction suggested in the previous section.)

Step C. By modifying the arguments of Hájek (1972) it is easy to see that (θ, G) is regular if g is absolutely continuous with derivative g' and if $I(G)$, the Fisher information for location given in Section 2, is finite. The converse is also true (proof available from author).

By (4.2) we calculate

$$\dot{\ell}(c, y, \theta, G) = c\frac{g'}{g}(y - c\theta^T)\{E(C^TC)I(G)\}^{-1}, \tag{4.6}$$

where the last term is just $I^{-1}(\theta, G)$. To apply Condition H or H′ we need to estimate g'/g and $I(G)$. This is achieved by the following lemma whose proof is given in Section 6.1.

LEMMA 4.1. *Let $\varepsilon_1, \varepsilon_2, \cdots$ be* i.i.d. *random variables. There exists a sequence of function estimates $q_m: R \times R^m \to R$, $m = 1, 2, \cdots$, such that q_m is bounded for each m and such that as $m \to \infty$*

$$\int \left\{q_m(y; \varepsilon_1, \cdots, \varepsilon_m) - \frac{g'(y)}{g(y)}\right\}^2 g(y)\, dy \to 0 \tag{4.7}$$

in probability whenever the common d.f. *of the ε_i is G with density g and $I(G) < \infty$.*

We proceed to show how to estimate $\dot{\ell}$ and $I(G)$ separately and verify Condition H′. Let

$$\hat{\varepsilon}_i = Y_i - C_i\bar{\theta}_m^T(X_1, \cdots, X_m), \quad i = 1, \cdots, m, \tag{4.8}$$

be the residuals with respect to the "discretized" estimate based on the first m observations. Define

$$\psi_m(y; X_1, \cdots, X_m) = ½\{q_m(y; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m) - q_m(-y; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m)\} \tag{4.9}$$

and

$$\hat{\dot{\ell}}_m(c, y, \theta; X_1, \cdots, X_m) = c\psi_m(y - c\theta^T; X_1, \cdots, X_m). \tag{4.10}$$

Clearly $\hat{\dot{\ell}}(\cdot; X_1, \cdots, X_m) \in \mathscr{H}_0$ and

$$\begin{aligned} &\int |\hat{\dot{\ell}}_m(c, y, \theta_m; X_1, \cdots, X_m) - \dot{\ell}(c, y, \theta_m, G)|^2 f(c, y, \theta_m, G)\, dy\, \nu(dc) \\ &= \int c\left|\psi_m(y - c\theta_m^T; X_1, \cdots, X_m) - \frac{g'}{g}(y - c\theta_m^T)\right|^2 c^T g(y - c\theta_m^T)\, dy\, \nu(dc) \\ &\le \left[\left\{\int \left|q_m(y; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m) - \frac{g'}{g}(y)\right|^2 g(y)\, dy\right\} ECC^T\right]. \end{aligned} \tag{4.11}$$

Now let $\theta_m = \theta + t_m$, where t_m and $c_1, \cdots, c_m$ are p-dimensional vectors such that $|t_m| = O(m^{-1/2})$ and $\sum_{i=1}^m c_i t_m^T t_m c_i^T$ is bounded independent of m. Then the sequence of m-dimensional product measures induced by $\varepsilon_1, \cdots, \varepsilon_m$ and $\varepsilon_1 - c_1 t_m^T, \cdots, \varepsilon_m - c_m t_m^T$ are contiguous if $I(G) < \infty$ (Hájek and Sidák, 1967, page 211). Since ECC^T is finite, if $|t_m| = O(m^{-1/2})$, $\sum_{i=1}^m c_i t_m^T t_m c_i^T = O_{P_\theta}(1)$. Thus, by Lemma 4.1,

$$\int \left|q_m(y; \varepsilon_1 - C_1 t_m^T, \cdots, \varepsilon_m - C_m t_m^T) - \frac{g'}{g}(y)\right|^2 g(y)\, dy \to_{P_\theta} 0. \tag{4.12}$$

But, as usual, by the structure of $\bar{\theta}_m$ and its $m^{1/2}$-consistency, this result is enough to establish

$$\int \left\{q_m(y; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m) - \frac{g'}{g}(y)\right\}^2 g(y)\, dy \to_{P_\theta} 0. \tag{4.13}$$

Substituting in (4.11), we see that $\hat{\dot{\ell}}_m$ is a consistent estimate of $\dot{\ell}$ in the sense of part (a) of Condition H′, in (3.11).

There are various ways to construct $\hat{I}_n$. For instance, we can verify (3.15) in this case as follows:

$$m^{-1}\sum_{i=1}^m \dot{\ell}^T\dot{\ell}(X_i,\theta_m,G) = m^{-1}\sum_{i=1}^m C_i^T C_i\left(\frac{g'}{g}\right)^2(Y_i - C_i\theta_m^T) \tag{4.14}$$

$$\to_{P_{\theta_m}} E(C^TC)I(G) = I(\theta, G)$$

by the weak law of large numbers. By contiguity we can replace θ_m by θ in P_{θ_m}. This yields as the consistent estimate of (3.14),

$$\hat{I}_n^{(1)} = \bar{n}^{-1}\sum_{i=m+1}^n C_i^T C_i \psi_m^2(Y_i - C_i\bar{\theta}_n^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m). \tag{4.15}$$

A more familiar alternative, which may similarly be shown to work, is

$$\hat{I}_n^{(2)} = \left(n^{-1}\sum_{i=1}^n C_i^T C_i\right)\bar{n}^{-1}\sum_{i=m+1}^n \psi_m^2(Y_i - C_i\bar{\theta}_m^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m). \tag{4.16}$$

We have proved the following result.

THEOREM 4.1. *Let $\bar{\theta}_n$ be defined as in* (4.5), *ψ_m as in* (4.9). *Let*

$$\hat{\theta}_n = \bar{\theta}_n + \bar{n}^{-1}\sum_{i=m+1}^n C_i\psi_m(Y_i - C_i\bar{\theta}_m^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m) \tag{4.17}$$

where $\hat{I}_n$ is given by (4.15) *or* (4.16). *Then $\{\hat{\theta}_n\}$ is adaptive in Example* 2.

EXAMPLE 3.

Step A. If $c = (c^\circ, 1)$, $q(\theta) = (\theta_1, \cdots, \theta_{p-1})$ and $\tilde{\ell}$ is defined by (3.18), we get

$$\tilde{\ell}(c, y, \theta, G) = (c^\circ - EC^\circ)(\text{Var } C^\circ)^{-1}\frac{g'}{g}(y - c\theta^T)I^{-1}(G). \tag{4.18}$$

Thus, formally

$$E_{(\theta,G_0)}\tilde{\ell}(X,\theta,G) = E(C^\circ - EC^\circ)(\text{Var } C^\circ)^{-1}E\frac{g'}{g}(\varepsilon)I^{-1}(G) = 0$$

and Condition S* is satisfied. In view of (4.18) it is natural to choose

$$\mathscr{H}_0 = \{h: h(c, y, \theta) = (c^\circ - EC^\circ)(\text{Var } C^\circ)^{-1}\psi(y - c\theta^T),\ \psi \text{ bounded}\}. \tag{4.19}$$

Step B. Let ψ be as in Step B of Example 2 and define

$$\mathscr{G}_0 = \left\{G: \int \psi(y)G(dy) = 0\right\}. \tag{4.20}$$

Evidently $\mathscr{G}_0$ is an identifying subset and, by Huber's theorem, $\{\tilde{\theta}_n\}$ corresponding to ψ are $\sqrt{n}$-consistent when G is restricted to $\mathscr{G}_0$.

Step C. A possible definition of $\hat{\hat{\ell}}$ is just

$$\hat{\hat{\ell}}_m(c, y, \theta; X_1, \cdots, X_m) = (c^\circ - EC^\circ)(\text{Var } C^\circ)^{-1}q_m(y - c\theta^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m)\hat{I}^{-1}, \tag{4.21}$$

where

$$\hat{I} = \bar{n}^{-1}\sum_{i=m+1}^n q_m^2(Y_i - C_i\bar{\theta}_n^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m), \tag{4.22}$$

q_m is given in Lemma 4.1 and the $\hat{\varepsilon}_i$ are defined by (4.8). That $\hat{\hat{\ell}}$ works is evident by the same argument as we gave for Theorem 4.1, since regular (θ, G) again correspond to $I(G) < \infty$. This is not satisfactory, however, because the resultant estimates depend on the first and second moments of the unknown distribution of C°. We claim that estimating these does just as well. Here is one way of proceeding. Define

(4.23)
$$\bar{C}_n^\circ = n^{-1}\textstyle\sum_{i=1}^n C_i^\circ$$
$$\hat{\mathrm{V}}\mathrm{ar}\, C^\circ = n^{-1}\textstyle\sum_{i=1}^n (C_i^\circ - \bar{C}_n^\circ)^T(C_i^\circ - \bar{C}_n^\circ).$$

Let

(4.24)
$$\hat{q}_n = \bar{\theta}_n^{(p-1)} + \bar{n}^{-1}\textstyle\sum_{i=m+1}^n (C_i^\circ - \bar{C}_n^\circ)(\hat{\mathrm{V}}\mathrm{ar}\, C_n^\circ)^- q_m(Y_i - C_i\bar{\theta}_n^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m)\hat{I}^{-1}$$

where $\bar{\theta}_n^{(p-1)}$ is the vector of the initial $p-1$ elements of $\bar{\theta}_n$.

THEOREM 4.2. *The estimates $\hat{q}_n$ defined by* (4.24) *adaptively estimate* $(\theta_1, \cdots, \theta_{p-1})$ *in Example* 3.

PROOF. We know that

(4.25)
$$\bar{n}^{-1}\textstyle\sum_{i=m+1}^n (C_i^\circ - EC^\circ)q_m(Y_i - C_i\bar{\theta}_n^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m)(\mathrm{Var}\, C^\circ)^- = o_P(n^{-1/2})$$

and

(4.26)
$$\hat{\mathrm{V}}\mathrm{ar}\, C^\circ = \mathrm{Var}\, C^\circ + o_P(1).$$

Therefore, replacing Var C° by $\hat{\mathrm{V}}\mathrm{ar}\, C^\circ$ in (4.21) will still lead to adaptive estimates. Thus to establish that the estimates given by (4.24) are adaptive it suffices to prove that

(4.27)
$$\bar{n}^{-1}\textstyle\sum_{i=m+1}^n (\bar{C}_n^\circ - EC^\circ)(\hat{\mathrm{V}}\mathrm{ar}\, C^\circ)^- q_m(Y_i - C_i\bar{\theta}_n^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m) = o_P(n^{-1/2})$$

or, since $\bar{C}_n^\circ - EC^\circ = O_P(n^{-1/2})$, that

(4.28)
$$\bar{n}^{-1}\textstyle\sum_{i=m+1}^n q_m(Y_i - C_i\bar{\theta}_n^T; \hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m) = o_P(1).$$

To prove (4.28) we show that we can replace q_m by g'/g and $Y_i - c_i\bar{\theta}_n^T$ by ε_i and then apply the law of large numbers. Details are given in Section 6.2. □

EXAMPLE 4. *Step A.* In this case if $\theta = (\mu, [V])$, then

(4.29)
$$f(x, \theta, G) = \{\det(V)\}^{1/2}\gamma(\{(x-\mu)V(x-\mu)^T\}^{1/2})$$

where det denotes determinant, and γ maps R^+ into itself. Of course, $\gamma(|x|)$ is the density of G. We want to estimate

(4.30)
$$q(\mu, [V]) = (\mu, q_0([V]))$$

where q_0 is any homogeneous function of $[V]$. A "most general" choice is $q_0([V]) = [V]/\mathrm{tr}(V)$. We can write, for (θ, G_0) regular,

$$\dot{\ell}(x, \mu, G_0)I^{-1}(\theta, G_0) = (\psi^\circ(x, \mu, V), [\chi^\circ(x, \mu, V)])$$

where ψ° is $1 \times k$, χ° is $k \times k$ symmetric, and $[\chi]$ denotes the $k(k+1)/2$ dimensional vector of the lower half of χ. It is shown in Section 6.3 that

(4.31)
$$\psi^\circ(x, \mu, V) = \psi^\circ((x-\mu)V^{1/2}, 0, J)V^{-1/2}$$

(4.32)
$$\chi^\circ(x, \mu, V) = V^{1/2}\chi^\circ((x-\mu)V^{1/2}, 0, J)V^{1/2},$$

where J is the $k \times k$ identity matrix. We further show in Section 6.3 that, if $|\cdot|$ is the Euclidean norm and $\gamma_0(|x|)$ is the density of G_0, then

(4.33)
$$\psi^\circ(x, 0, J) = -\frac{x}{|x|}\frac{\gamma_0'}{\gamma_0}(|x|)kI_1^{-1}(G_0)$$

and

$$\chi^{\circ}_{ij}(x, 0, J) = \begin{cases} I_2^{-1}(G_0)k(k+2)\dfrac{x_i x_j}{|x|}\dfrac{\gamma_0'}{\gamma_0}(|x|), & i \neq j, \\ 2\left\{I_2(G_0)\dfrac{3}{k(k+2)} - 1\right\}^{-1}\left\{\dfrac{x_i^2}{|x|}\dfrac{\gamma_0'}{\gamma_0}(|x|) + 1\right\}, & i = j, \end{cases} \tag{4.34}$$

where

$$I_1(G) = c_k \int_0^\infty r^{k-1}\frac{[\gamma']^2}{\gamma}(r)\,dr \tag{4.35}$$

$$I_2(G) = c_k \int_0^\infty r^{k+1}\frac{[\gamma']^2}{\gamma}(r)\,dr \tag{4.36}$$

and c_k is the surface area of the unit sphere in R^k. Then by (4.31) and (4.32),

$$\begin{aligned} &E_{(\theta,G)}\{\psi^{\circ}(X, \mu, V), [\chi^{\circ}(X, \mu, V)]\}\dot{q}^T(\theta) \\ &\quad = E_{(0,[J],G)}\{\psi^{\circ}(X, 0, J)V^{-1/2}, [V^{1/2}\chi^{\circ}(X, 0, J)V^{1/2}]\}\dot{q}^T(\theta). \end{aligned} \tag{4.37}$$

Moreover, if $i \neq j$, X°_{ij} changes sign if all the coordinates of x other than x_i are left unchanged while $x_i \to -x_i$. Since if $\theta = (0, [J])$, all the X_i are identically distributed and the distributions of $(X_1, \cdots, X_k)$ and $(\pm X_1, \cdots, \pm X_k)$ are the same, we conclude that

$$E_{(0,[J],G)}\psi^{\circ}(X, 0, J) = 0 \tag{4.38}$$

$$E_{(0,[J],G)}\chi^{\circ}(X, 0, J) = cJ, \tag{4.39}$$

where c depends on G and G_0. Therefore

$$E_{(0,[J],G)}[V^{1/2}\chi^{\circ}(X, 0, J)V^{1/2}] = c[V]. \tag{4.40}$$

Substituting (4.38) and (4.40) back into (4.37) we find that all components of (4.37) vanish either by (4.38) or by Euler's equation $\sum_{k \geq \ell} v_{k\ell}\, \partial q_0/\partial v_{k\ell} = 0$.

The orthogonality Condition S* follows and our argument makes it clear that $\mathscr{H}$ defined in (3.3), contains the set $\mathscr{H}_0$ of $h(x, \theta)$ defined by

$$h(x, \theta) = (\psi((x-\mu)V^{1/2})V^{-1/2}, [V^{1/2}\chi((x-\mu)V^{1/2})V^{1/2}])\dot{q}^T(\theta), \tag{4.41}$$

where ψ is $1 \times k$ and χ is symmetric $k \times k$ with forms

$$\psi(x) = \omega(|x|)\frac{x}{|x|}a_1 \tag{4.42}$$

$$\chi_{ij}(x) = \begin{cases} \omega(|x|)\dfrac{x_i x_j}{|x|}a_2, & i \neq j, \\ \left\{\omega(|x|)\dfrac{x_i^2}{|x|} + 1\right\}a_3, & i = j, \end{cases} \tag{4.43}$$

where ω is bounded and a_1, a_2, a_3 are constant. Clearly $\mathscr{H}$ is much bigger than $\mathscr{H}_0$, but $\mathscr{H}_0$ is the space of natural estimates of ℓ.

Step B. Thanks to Maronna (1976) we can find an identifying subset $\mathscr{G}_0$ and corresponding $\sqrt{n}$ − consistent $\tilde{\theta}_n$ as follows. Let u_1 and u_2 be functions on R^+. Define the M-estimate $(\tilde{\mu}_n, \tilde{V}_n)$ corresponding to u_1 and u_2 to be any solution of

$$\begin{aligned} &n^{-1}\textstyle\sum_{i=1}^n u_1(\{(X_i - \tilde{\mu}_n)\tilde{V}_n(X_i - \tilde{\mu}_n)^T\}^{1/2}) = 0 \\ &n^{-1}\textstyle\sum_{i=1}^n u_2(\{(X_i - \tilde{\mu}_n)\tilde{V}_n(X_i - \tilde{\mu}_n)^T\})(X_i - \tilde{\mu}_n)^T(X_i - \tilde{\mu}_n) = [\tilde{V}_n]^{-1} \end{aligned} \tag{4.44}$$

if one exists, and arbitrarily otherwise.

It is easy to see that the maximum likelihood estimates for a particular G are of this type. Let u_1, u_2 satisfy conditions (A) – (D) on page 53 of Maronna (1976). In addition, if $\psi_i(s) = su_i(s)$, $i = 1, 2$, suppose that $s\psi_j'(s)$ are bounded, $j = 1, 2$, and $\psi_1' > 0$. By Theorem 5.6 of Maronna, under these conditions $n^{1/2}(\tilde{\mu}_n - \tilde{\mu}, \tilde{V}_n - \tilde{V}) = O_P(1)$ for all $F \in \mathscr{F}$ where $\tilde{\mu}(\mu, V, G)$, $\tilde{V}(\mu, V, G)$ satisfy *uniquely*

$$\int u_1(\{(x-\tilde{\mu})\tilde{V}(x-\tilde{\mu})^T\}^{1/2})(x-\tilde{\mu})f(x,\theta,G)\,dx = 0 \tag{4.45}$$

$$\int u_2(\{(x-\tilde{\mu})\tilde{V}^T(x-\tilde{\mu})^T\})(x-\tilde{\mu})^T(x-\tilde{\mu})f(x,\theta,G)\,dx = [\tilde{V}]^{-1}. \tag{4.46}$$

It is clear by the unicity of $\tilde{\mu}$, $\tilde{V}$ that

$$\tilde{\mu}(\mu, V, G) = \mu, \tag{4.47}$$

$$\tilde{V}(\mu, V, G) = c(G)V, \tag{4.48}$$

where $c(G)$ is that measure of scale which is the unique solution of the equation

$$E\{u_2(c\,\varepsilon\,\varepsilon^T)\} = \frac{1}{c};$$

existence is guaranteed by the monotonicity of u_2. Clearly we can take as an identifying subset

$$\mathscr{G}_0 = \{G : c(G) = 1\} \tag{4.49}$$

and $\tilde{\theta}_n = (\tilde{\mu}_n, \tilde{V}_n)$ defined by (4.44).

Step C. It may be shown that regularity of (θ, G) is equivalent to absolute continuity of γ on $(0, \infty)$ and finiteness of $I_1(G)$ and $I_2(G)$. (Proof available from author.) We will show how to construct adaptive estimates of $q_0(V)$ in a simple fashion and then discuss the simultaneous adaptive estimation of μ.

Note that if X has density given by (4.29), then $\log|(X-\mu)V^{1/2}|$ has density j given by

$$j(z) = c_k e^{kz}\gamma(e^z). \tag{4.50}$$

Thus

$$\frac{\gamma'}{\gamma}(y) = y^{-1}\left\{\frac{j'}{j}(\log y) - k\right\}, \quad y > 0, \tag{4.51}$$

and this leads to the following construction of an estimate of γ'/γ.

Let $\bar{\mu}_m$ be obtained by discretizing $\tilde{\mu}_m$ as usual while $[\tilde{V}_m]$ is the closest member of the $m^{-1/2}$ lattice to $\tilde{V}_m$ which itself corresponds to a positive definite matrix. Let

$$z_{im} = \log|(X_i - \bar{\mu}_m)\bar{V}_m^{1/2}|, \quad i = 1, \cdots, m,$$

and define

$$\omega_m(y; X_1, \cdots, X_m) = y^{-1}\{q_m'(\log y; z_{1m}, \cdots, z_{mm}) - k\}. \tag{4.52}$$

We claim that

$$\int |x|^2 \left|\omega_m(|x|; X_1, \cdots, X_m) - \frac{\gamma'}{\gamma}(|x|)\right|^2 \gamma(|x|)\,dx \to 0 \tag{4.53}$$

in P_θ probability if (θ, G) is regular. The proof follows the usual lines. By construction of $\bar{\mu}_m$, $\bar{V}_m$ it is possible to treat them as deterministic sequences such that $|\bar{\mu}_m - \mu|$ and $|\bar{V}_m - V| = O(m^{-1/2})$. Since (θ, G) is regular the m-dimensional product measures induced by $\varepsilon_1, \cdots, \varepsilon_m$ and $(X_1 - \bar{\mu}_m)\bar{V}_m^{1/2}, \cdots, (X_m - \bar{\mu}_m)\bar{V}_m^{1/2}$ are contiguous. If we also use (4.51) we can conclude that (4.53) is equivalent to

$$\int \left| q_m(\log|x|; \log|\varepsilon_1|, \cdots, \log|\varepsilon_m|) - \frac{j'}{j}(\log|x|)\right|^2 \gamma(|x|)\,dx \to 0 \tag{4.54}$$

in probability whenever $\varepsilon_1, \cdots, \varepsilon_m$, are i.i.d. with common distribution G such that $I_1(G)$ and $I_2(G)$ are finite. But the integral in (4.54) equals

$$\int_{-\infty}^{\infty} \left| q_m(z; \log|\varepsilon_1|, \cdots, \log|\varepsilon_m|) - \frac{j'}{j}(z) \right|^2 g(z)\, dz. \tag{4.55}$$

Moreover,

$$\int_{-\infty}^{\infty} \frac{(j')^2}{j}(z)\, dz = \int_{-\infty}^{\infty} \left\{ e^z \frac{\gamma'}{\gamma}(e^z) + k \right\}^2 g(z)\, dz = I_2(G) - k^2 \tag{4.56}$$

using integration by parts. Thus the integral in (4.55) tends to 0 whenever $I_2(G) < \infty$ by Lemma 4.1 and (4.54) and hence (4.53) holds. Now that we have an estimate $\omega_m(\cdot; X_1, \cdots, X_m)$ of γ'/γ we can estimate $I_2(G)$ by, for instance, splitting our preliminary sample of m, taking $m = 2\ell$ and letting

$$\hat{I}_2 = \ell^{-1} \textstyle\sum_{i=\ell+1}^{m} q_m^2(z_{im}; z_{1m}, \cdots, z_{\ell m}) + k^2. \tag{4.57}$$

Evidently $\hat{I}_2$ depends only on $X_1, \cdots, X_m$. Moreover, we can argue as for (4.28) that, whenever (θ, G) is regular,

$$\hat{I}_2 \to I_2(G) \text{ in probability.} \tag{4.58}$$

Now define $\hat{\chi}_0(\cdot, O, J)$ by substituting $\hat{I}_2$ for $I_2(G_0)$ and $\omega_m(\cdot; X_1, \cdots, X_m)$ for γ_0'/γ_0 in (4.34) and let

$$\hat{\hat{\ell}}_m(x, \theta; X_1, \cdots, X_m) = [V_m^{1/2}\, \hat{\chi}_0((x-\mu)V^{1/2}, O, J)V_m^{1/2}]\dot{q}_0^T([V]). \tag{4.59}$$

This is the natural estimate of $\tilde{\ell}$ corresponding to $q_0([V])$. Now after some algebra, if $\theta_m = (\mu_m, [V_m])$,

$$\begin{aligned}\int | \hat{\hat{\ell}}_m(x, \theta_m; X_1, \cdots, X_m) - \dot{\ell}(x, \theta_m, G) I^{-1}(\theta_m, G)(0, \dot{q}_0([V]))^T |^2 f(x, \theta_m, G)\, dx \\ = O_P\Big(\int |(x-\mu_m)V_m^{1/2}|^2 \Big| \omega_m((x-\mu_m)V_m^{1/2}; X_1, \cdots, X_m) \\ - \frac{\gamma'}{\gamma}(|(x-\mu_m)V_m^{1/2}|) \Big|^2 f(x, \theta_m, G)\, dx \Big) + O_P(\hat{I}_2 - I_2).\end{aligned} \tag{4.60}$$

But the right-hand side of (4.60) is $o_p(1)$ by (4.53) and (4.58). From (4.60) and the structure of $\hat{\hat{\ell}}$ we see that $\hat{\hat{\ell}}$ falls in $\mathscr{H}_0$ given by (4.41) and is appropriately consistent. We have proved the following result.

THEOREM 4.3. *In Example* 4, *if we define*

$$\hat{q}_{on} = q_0([\bar{V}_n]) + \bar{n}^{-1} \textstyle\sum_{i=m+1}^{n} \hat{\hat{\ell}}_m(X_i, \bar{\theta}_n; X_1, \cdots, X_m), \tag{4.61}$$

then $\{\hat{q}_{on}\}$ *is an adaptive estimate of* $q_0([V])$.

To estimate μ simultaneously and adaptively using the estimate of γ'/γ we need to show that

$$\int \left| \omega_m(|x|; X_1, \cdots, X_m) - \frac{\gamma'}{\gamma}(|x|) \right|^2 \gamma(|x|)\, dx \to 0 \tag{4.62}$$

in probability, or equivalently that

$$\int_{-\infty}^{\infty} e^{-2z} \left| q_m(z; \log|\varepsilon_1|, \cdots, \log|\varepsilon_m|) - \frac{j'}{j}(z) \right|^2 g(z)\, dz \tag{4.63}$$

in probability. Unfortunately, to show (4.63) we need

$$\int_{-\infty}^{\infty} e^{-2z} \frac{(j')^2}{j}(z)\,dz = c_k \int_0^{\infty} y^{k-1}\left\{\frac{\gamma'}{\gamma}(y) + ky^{-1}\right\}^2 \gamma(y)\,dy < \infty \tag{4.64}$$

and this happens if $I_1(G) < \infty$ and

$$\int_0^{\infty} y^{k-3}\gamma(y)\,dy < \infty, \tag{4.65}$$

a superfluous condition.

To get rid of (4.65) we need to estimate γ'/γ differently by smoothing the multivariate empirical distribution of $(X_i - \bar{\mu}_m)\bar{V}_m^{1/2}$ and constructing an estimate of γ'/γ out of the first partial derivatives of the smoothed empirical distribution. This can be done but we omit the tedious and rather technical definition of the estimate and the necessary argument.

5. Questions raised by this work and other issues in adaptive estimation.

5.1 *When is adaptation not possible*? We have seen heuristically the necessity of the $\sqrt{n}$-consistency condition GR(iv) and the orthogonality Condition S when there are no nuisance parameters. In parametric models $\sqrt{n}$-consistency is available under mild smoothness and identifiability conditions while orthogonality is special. Orthogonality seems special in these nonparametric nuisance parameter models as well. We illustrate with a famous example of Neyman and Scott. The failure of adaptation in this case was already noted by Wolfowitz (1953).

EXAMPLE 5. *Estimation in Model II.* Suppose $X_i = (X_{i1}, X_{i2})$, $i = 1, \cdots, n$, such that

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, 2, \tag{5.1}$$

where the ε_{ij} are independent identically distributed $\mathcal{N}(0, \theta)$, and the μ_i are independent and identically distributed with common distribution G. Let $\Theta = R^+$, $\mathcal{G} = \{$all distributions on $R\}$. It is easy to see that all (θ, G) are regular, and there is a natural $\sqrt{n}$-consistent estimate, the best unbiased estimate when the μ_i are treated as constants,

$$\bar{\theta}_n = \frac{1}{2n} \sum_{i=1}^n (X_{i1} - X_{i2})^2. \tag{5.2}$$

Thus Condition GR(iv) holds. But Condition H does not. For instance, take G_0 to be point mass at 0. Then

$$\dot{\ell}(x_1, x_2, \theta, G_0) = \frac{1}{\theta}\left\{\frac{(x_1^2 + x_2^2)}{2\theta} - 1\right\} \tag{5.3}$$

and

$$E_{(\theta,G)}\dot{\ell}(X, \theta, G_0) = \frac{1}{\theta^2}\int \mu^2 dG(\mu) > 0 \tag{5.4}$$

unless $G = G_0$. Thus adaptation in the sense we have discussed is not possible. Note that the natural estimate $\bar{\theta}_n$ has asymptotic variance $2\theta^2/n$ in this case while $I^{-1}(\theta, G_0) = \theta^2/n$. Lindsay (1978, 1980) and Hammerstrom (1978) have independently studied situations such as this one (which are the rule rather than the exception) where adaptation is not possible. They have obtained what may be viewed as a minimax optimality property of $\bar{\theta}_n$ in Example 5 and analogous results in other problems of this type. We are investigating the natural extension of adaptation in this context.

5.2 *Better estimates.* The estimates we construct in Examples 2–4 have some serious failings: (i) the estimate of $\dot{\ell}$ is based on a small subsample rather than all the data; (ii) the estimates do not have natural invariance properties possessed by reasonable estimates in these problems, primarily because of the discretization of $\bar{\theta}_n$; and (iii) the behavior of the estimates when $I(\theta, G)$ is singular is not analyzed.

We believe that analogues of Stone's procedures in the location problem (which meet all these criticisms) can be constructed using the special structures of our examples. We have not pursued this since our interest lies primarily in illustrating the applicability of the general Condition H.

5.3 *Extensions to other asymptotic structures.* The theory we have developed extends naturally to cases where the observations are independent but not identically distributed, e.g., the usual linear model context. It can be applied, we believe, to the linear model and, as Stein's calculations and Wolfowitz (1974) indicate, to multiple regression models where both the location and the scale of the dependent variable are functions (possibly nonlinear) of the independent variables. Other extensions to non-independent situations, such as that treated in part in Beran (1976), should also be possible.

5.4 *Efficient estimation of functionals.* Levitt (followed by Ibragimov and Khazminski and others), in a series of papers starting with Levitt (1974), has studied how best to estimate functions $\theta(F)$ in nonparametric models, basing this work in part on Stein (1956). In some sense our problem can be viewed as the estimation of the solution $\theta(F)$ of $\int \dot{\ell}\,(x, \theta, G)\, dF_{(\theta,G)}(x) = 0$ which is meaningful (though possibly nonexistent) for $F \in \mathscr{F}$. Beyond this formal connection there seems to be no real link between our studies.

5.5 *Uniformity of adaptation.* Beran (1978) notes in the location problem (Example 1) that adaptive estimates converge to their limiting distributions uniformly on (shrinking n-dependent) "contiguous" neighborhoods of each G. This property can, we believe, be suitably re-expressed to apply generally. However, the weakness of this property is pointed out by Klaassen (1980) who shows (in Example 1, his Theorems 3.2.1 and 3.3.2) that for reasonable fixed neighborhoods the convergence is far from uniform. Thus from a practical point of view adaptive estimates may not work nearly as well for moderate samples as we might expect.

5.6 *Practical questions.* The difficulty of nonparametric estimation of score functions suggests that a more practical goal is partial adaptation, the construction of estimates which are (i) always $\sqrt{n}$-consistent, and (ii) efficient over a large parametric subfamily of $\mathscr{F}$. Our results indicate that when the orthogonality Condition S* and $\sqrt{n}$-consistency Condition GR(iv) hold, this goal should be achievable by using a one-step Newton approximation to the maximum likelihood estimate for the parametric subfamily by starting with an estimate which is $\sqrt{n}$-consistent for all of $\mathscr{F}$. Partial adaptation in Example 2 is discussed in Hogg (1980). This highlights an important practical and theoretical question in problems of this type, how to construct $\sqrt{n}$-consistent estimates. When there are no nuisance parameters present and adaptation is possible, maximum likelihood estimates for fixed shapes are natural candidates. In general, this question deserves further study. The constructions of Birgé (1980) may prove useful.

6. Theoretical Details.

6.1 *Proof of Lemma* 4.1. We use Stone's (1975) approach. Let ϕ_σ be the $\mathscr{N}(0, \sigma^2)$ density, g be any density, and define the convolution of g and ϕ_σ

$$g_\sigma = g * \phi_\sigma \tag{6.1}$$

and the convolution of the empirical d.f. and ϕ_σ

$$\hat{g}_\sigma(y) = m^{-1} \textstyle\sum_{i=1}^{m} \phi_\sigma(y - \varepsilon_i). \tag{6.2}$$

We suppress dependence on $\varepsilon_1, \cdots, \varepsilon_m$ in what follows.

For given $\sigma_m, c_m, d_m, e_m > 0$ define

$$q_m(y) = \begin{cases} \dfrac{\hat{g}'_{\sigma_m}}{\hat{g}}(y) & \text{if } \hat{g}_{\sigma_m}(y) \geq d_m, \quad |y| \leq e_m \text{ and } |\hat{g}'_{\sigma_m}(y)| \leq c_m \hat{g}_{\sigma_m}(y), \\ 0 & \text{otherwise.} \end{cases} \tag{6.3}$$

We claim that if $c_m \to \infty$, $e_m \to \infty$, $\sigma_m \to 0$ and $d_m \to 0$ in such a way that

$$\sigma_m c_m \to 0, \tag{6.4}$$

$$e_m \sigma_m^{-3} = o(m), \tag{6.5}$$

then q_m satisfies the conclusions of Lemma 4.1. The argument proceeds by

LEMMA 6.1. *If the conditions of Lemma* 4.1 *hold and* q_m *satisfies* (6.3)–(6.5), *then*

$$\int_{[g>0]} \left\{ q_m(y) - \frac{g'_{\sigma_m}}{g_{\sigma_m}}(y) \right\}^2 g_{\sigma_m}(y)\, dy \to_P 0. \tag{6.6}$$

PROOF. We use the elementary estimates noted in Stone. For κ_i universal constants and all y,

$$\operatorname{Var} \hat{g}_\sigma^{(i)}(y) \leq \kappa_i \sigma^{-(2i+1)} m^{-1} g_\sigma(y), \qquad i = 0, 1, \cdots. \tag{6.7}$$

Denote the conditions in (6.3) by A, B, C and the left-hand side of (6.6) by $I_1 + I_2$, where

$$I_1 = \int_{ABC} \left\{ \frac{\hat{g}'_{\sigma_m}}{\hat{g}_{\sigma_m}}(y) - \frac{g'_{\sigma_m}}{g_{\sigma_m}}(y) \right\}^2 g_{\sigma_m}(y)\, dy \tag{6.8}$$

$$I_2 = \int_{[ABC]^c} \frac{[g'_{\sigma_m}]^2}{g_{\sigma_m}}(y)\, dy. \tag{6.9}$$

Bound $E(I_1)$ by

$$2\left[\int_{ABC} g_{\sigma_m}^{-1}(y) E\{\hat{g}'_{\sigma_m}(y) - g'_{\sigma_m}(y)\}^2\, dy + \int_{ABC} c_m^2 g_{\sigma_m}^{-1}(y) E\{\hat{g}_{\sigma_m}(y) - g_{\sigma_m}(y)\}^2\, dy \right] = o(1) \tag{6.10}$$

by (6.7), (6.4) and (6.5). Bound

$$E(I_2) \leq \int \frac{[g'_{\sigma_m}]^2}{g_{\sigma_m}}(y)[P\{|\hat{g}'_{\sigma_m}(y)| > c_m \hat{g}_{\sigma_m}(y)\} + P\{\hat{g}_{\sigma_m}(y) < d_m, g(y) > 0\} + I(|y| > e_m)]\, dy. \tag{6.11}$$

We claim that

$$\hat{g}_{\sigma_m}(y) \to g(y) \quad \text{in probability for all } y \quad \text{if} \quad m\sigma_m \to \infty, \tag{6.12}$$

$$\hat{g}'_{\sigma_m}(y) \to g'(y) \quad \text{in probability a.e. } y \quad \text{if} \quad m\sigma_m^3 \to \infty, \tag{6.13}$$

$$\int \frac{{g'_{\sigma_m}}^2}{g_{\sigma_m}}(y)\, dy \leq \int \frac{g'^2}{g}(y)\, dy \quad \text{for all } m. \tag{6.14}$$

Evidently (6.12) and (6.13) imply that if $c_m \to \infty$ and $d_m \to 0$, then the two probabilities in (6.11) tend to 0 a.e. y, while (6.12)–(6.14) imply uniform integrability of ${g'_{\sigma_m}}^2/g_{\sigma_m}(y)$ and hence that

$$EI_2 \to 0. \tag{6.15}$$

Together (6.10) and (6.15) will establish Lemma 6.1. It remains to prove (6.12)-(6.14). Now by (6.7), for all y,

$$\hat{g}_{\sigma_m}(y) - g_{\sigma_m}(y) \to 0 \quad \text{in probability if} \quad m\sigma_m \to \infty, \tag{6.16}$$

$$\hat{g}'_{\sigma_m}(y) - g'_{\sigma_m}(y) \to 0 \quad \text{in probability if} \quad m\sigma_m^3 \to \infty. \tag{6.17}$$

Continuity of g and (6.16) imply (6.12). To prove (6.13) write (using the absolute continuity of g),

$$\begin{aligned}\int_{-\infty}^{\infty} | g'_{\sigma m}(y) - g'(y)| \, dy &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} (g'(y - \sigma_m x) - g'(y))\phi(x) \, dx \right| dy \\ &\le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} | g'(y - \sigma_m x) - g'(y)| \, dy\phi(x) \, dx.\end{aligned} \tag{6.18}$$

Note that $I(G) < \infty$ implies $\int_{-\infty}^{\infty} |g'(y)| \, dy < \infty$. Thus we can apply the L_1 continuity theorem and the dominated convergence theorem to conclude that the right-hand side of (6.18) tends to 0 as $\sigma_m \to 0$ and (6.13) follows from (6.17) and (6.18). Finally, (6.14) is a well known inequality (see Hájek and Šidák, 1967, page 17). The lemma is proved. □

Next we need

LEMMA 6.2. *If* $\sigma \to 0$,

$$\int_{[g>0]} \left\{ \frac{g'_\sigma}{\sqrt{g_\sigma}}(y) - \frac{g'}{\sqrt{g}}(y) \right\}^2 dy \to 0. \tag{6.19}$$

PROOF. Apply (6.12)-(6.14).

LEMMA 6.3. *If* $\sigma_m c_m \to 0$,

$$\int_{[g>0]} q_m^2(y)(\sqrt{g_{\sigma_m}(y)} - \sqrt{g(y)}\,)^2 \, dy \to_P 0. \tag{6.20}$$

PROOF. Write, using Cauchy's form of Taylor's theorem,

$$\begin{aligned}\sqrt{g_\sigma(y)} - \sqrt{g(y)} &= \sigma \int_0^1 \left\{ \frac{\partial}{\partial \sigma} g_{\sigma\lambda}(y)/2g_{\sigma\lambda}^{1/2}(y) \right\} d\lambda \\ &= -\frac{\sigma}{2} \int_0^1 g_{\sigma\lambda}^{-1/2}(y) \int_{-\infty}^{\infty} zg'(y - \lambda\sigma z)\phi(z) \, dz \, d\lambda.\end{aligned} \tag{6.21}$$

Thus we can bound the square in the integrand of (6.20) by

$$\begin{aligned}&\frac{\sigma_m^2}{4} \int_0^1 g_{\lambda\sigma_m}^{-1}(y) \left\{ \int_{-\infty}^{\infty} zg'(y - \lambda\sigma_m z)\phi(z) \, dz \right\}^2 d\lambda \\ &\le \frac{\sigma_m^2}{4} \int_0^1 \int_{-\infty}^{\infty} \frac{\{zg'(y - \lambda\sigma_m z)\}^2}{g(y - \lambda\sigma_m z)} \phi(z) \, dz \, d\lambda\end{aligned} \tag{6.22}$$

by convexity of $(u, v) \to u^2/v$. Substitute (6.22) in (6.20) and use $| q_m | \le c_m$ to bound (6.20) by

$$\frac{c_m^2\sigma_m^2}{4}\int_0^1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{g'^2}{g}(v)\,z^2\phi(z)\,dz\,dv\,d\lambda.$$

Since the integrals stay bounded independent of m, the result follows. □

Lemma 4.1 follows from Lemmas 6.1 and 6.3 since

$$\begin{aligned}(6.23)\qquad &\int\left\{q_m(y)-\frac{g'}{g}(y)\right\}^2 g(y)\,dy\\ &\le 3\Bigg[\int_{[g>0]}\left\{q_m(y)-q_m\left(\frac{g_{\sigma_m}}{g}\right)^{1/2}(y)\right\}^2 g(y)\,dy\\ &\quad+\int_{[g>0]}\left\{q_m\left(\frac{g_{\sigma_m}}{g}\right)^{1/2}(y)-\left(\frac{g'_{\sigma_m}}{g_{\sigma_m}}\right)\left(\frac{g_{\sigma_m}}{g}\right)^{1/2}(y)\right\}^2 g(y)\,dy\\ &\quad+\int_{[g>0]}\left\{\left(\frac{g'_{\sigma_m}}{g_{\sigma_m}}\right)\left(\frac{g_{\sigma_m}}{g}\right)^{1/2}(y)-\frac{g'}{g}(y)\right\}^2 g(y)\,dy\Bigg],\end{aligned}$$

and the first term tends to 0 by Lemma 6.3, the second by Lemma 6.1, and the last by Lemma 6.2. □

6.2 *Consistency Proofs.*

(i) *Consistency of $\hat{I}_n$ in* (3.14). As usual, we can take $\bar{\theta}_n$ to be deterministic, and in view of (3.15) we need only check that

$$(6.24)\qquad \Delta_n=\bar{n}^{-1}\textstyle\sum_{i=m+1}^{n}\{\hat{\dot{\ell}}^T\hat{\dot{\ell}}(X_i,\theta_n;X_1,\cdots,X_m)-\dot{\ell}^T\dot{\ell}(X_i,\theta_n,G)\}\to_{P_{\theta_n}}0$$

whenever $|\theta_n-\theta|=O(n^{-1/2})$. But by (3.11),

$$\begin{aligned}&E_{\theta_n}\{|\Delta_n|\,|X_1,\cdots,X_m\}\\ (6.25)\qquad &\le E\{|\hat{\dot{\ell}}^T\hat{\dot{\ell}}(X_{m+1},\theta_n;X_1;X_1,\cdots,X_m)-\dot{\ell}^T\dot{\ell}(X_{m+1},\theta_n,G)|\,X_1,\cdots,X_m\}\\ &=o_{P_{\theta_n}}(1)\end{aligned}$$

and the result follows.

(ii) *Consistency in Theorem* 4.2. Again we can treat $\bar{\theta}_n$ as deterministic. Define measures $\{Q_n\}$ on $(R^{p+1})^n$ with densities

$$\textstyle\prod_{i=1}^m r(c_i)g(y_i-c_i\theta)^T\prod_{i=m+1}^n r(c_i)g(y_i-c_i(\theta-\bar{\theta}_n)^T).$$

We can argue as in the proof of (4.12) that the measures $\{Q_n\}$ are contiguous to the product measures specifying the distribution of the observations when θ is true. It follows that (4.28) is equivalent to

$$(6.26)\qquad \bar{n}^{-1}\textstyle\sum_{i=m+1}^n q_m(\varepsilon_i;\hat{\varepsilon}_1,\cdots,\hat{\varepsilon}_m)=o_P(1).$$

By the usual calculation, conditioning on the first m observations,

$$\begin{aligned}&E\left(\left[\bar{n}^{-1}\textstyle\sum_{i=m+1}^n\left\{q_m(\varepsilon_i;\hat{\varepsilon}_1,\cdots,\hat{\varepsilon}_m)-\dfrac{g'}{g}(\varepsilon_i)\right\}\right]^2\,\middle|\,\hat{\varepsilon}_1,\cdots,\hat{\varepsilon}_m\right)\\ &\quad=\int\left\{q_m(y;\hat{\varepsilon}_1,\cdots,\hat{\varepsilon}_m)-\frac{g'}{g}(y)\right\}^2 g(y)\,dy=o_P(1)\end{aligned}$$

by (4.13) and we can substitute g'/g for q_m in (6.26). With this final substitution, (4.28) follows from the WLLN. □

6.3 *Identities of Example* 4.

Verification of (4.31) *and* (4.32). Write $\dot{\ell} = (\dot{\ell}_1, \dot{\ell}_2)$ where

$$\dot{\ell}_1 = \left(\frac{\partial \ell}{\partial \mu_1}, \cdots, \frac{\partial \ell}{\partial \mu_k}\right), \qquad \dot{\ell}_2 = \left\{\frac{\partial \ell}{\partial v_{ij}} : i \geq j\right\}.$$

Evidently

$$\dot{\ell}_1(x, \theta, G_0) = -\,|(x-\mu)V^{1/2}|^{-1}\frac{\gamma_0'}{\gamma_0}(|(x-\mu)V^{1/2}|)(x-\mu)V \tag{6.27}$$

$$= \dot{\ell}_1((x-\mu)V^{1/2}, 0, [J], G_0)V^{1/2},$$

$$\dot{\ell}_2(x, \theta, G_0) = \left\{\left(\frac{(x_i-\mu_i)(x_j-\mu_j)}{|(x-\mu)V^{1/2}|}\frac{\gamma_0'}{\gamma_0}(|(x-\mu)V^{1/2}|) - v^{ij}\right)\left(1 - \frac{\delta_{ij}}{2}\right)\right\}, \tag{6.28}$$

where $V^{-1} = \| v^{ij} \|$ and $x = (x_1, \cdots, x_k)$.

Define a linear operator L_B on $R^{k(k+1)/2}$, corresponding to a $k \times k$ matrix $B = \| b_{ij} \|$, by the $\frac{k(k+1)}{2} \times \frac{k(k+1)}{2}$ matrix

$$L_B = \left\|(b_{ir}b_{sj} + b_{jr}b_{is})\left(1 - \frac{\delta_{ij}}{2}\right)\right\|, \qquad r \geq s,\ i \geq j,$$

where (r, s) indexes rows and (i, j) columns. It is easy to verify that

$$\dot{\ell}_2(x, \theta, G_0) = \dot{\ell}_2((x-\mu)V^{1/2}, 0, [J], G_0)L_V^{-1/2}. \tag{6.29}$$

By (6.27) and (6.29) we have

$$I(\theta, G_0) = \begin{pmatrix} V^{1/2} & 0 \\ 0 & L_{V^{-1/2}} \end{pmatrix}^T I(0, [J], G_0)\begin{pmatrix} V^{1/2} & 0 \\ 0 & L_{V^{-1/2}} \end{pmatrix} \tag{6.30}$$

and, finally,

$$\dot{\ell}(x, \theta, G_0)I^{-1}(\theta, G_0) = \dot{\ell}((x-\mu)V^{1/2}, 0, [J], G_0)I^{-1}(0, [J], G_0) \times \begin{pmatrix} V^{1/2} & 0 \\ 0 & L^T_{V^{-1/2}} \end{pmatrix}^{-1}$$

Since $V^{1/2}$ is symmetric, (4.31) follows. To get (4.32) it is enough to verify that

$$L_B^{-1} = L_{B^{-1}} \quad \text{for any} \quad B, \tag{6.31}$$

and that if x is a triangular array

$$xL_B^T = [BQ(x)B^T], \tag{6.32}$$

where $Q(x)$ is the symmetric matrix whose ij-th entry is x_{ij} if $i \geq j$, or x_{ji} if $i < j$. The verifications of (6.31) and (6.32) are straightforward exercises in matrix multiplication.

Verification of (4.33) *and* (4.34). In this case $V^{1/2} = J$. For convenience suppress $(0, [J], G_0)$ in the arguments of functions for this discussion. We have

$$\dot{\ell}_1(x) = -\frac{x}{|x|}\frac{\gamma_0'}{\gamma_0}(|x|), \tag{6.33}$$

$$E\dot{\ell}_1^T\dot{\ell}_1(X) = E\left(\frac{\gamma_0'}{\gamma_0}\right)^2(|X|)\frac{X^TX}{|X|^2} = \frac{1}{k}\left\{E\left(\frac{\gamma_0'}{\gamma_0}\right)^2(|X|)\right\}J \tag{6.34}$$

by symmetry. Next, note that

$$\dot{\ell}_2(x) = \left\{\left(\frac{x_ix_j}{|x|}\frac{\gamma_0'}{\gamma_0}(|x|) - \delta_{ij}\right)(1 - \delta_{ij}/2)\right\}_{i \geq j} \tag{6.35}$$

and by symmetry

(6.36) $$E\dot{\ell}_2^T\dot{\ell}_1(X) = 0$$

(6.37) $$E\dot{\ell}_2^T\dot{\ell}_2(X) = \| a_{rs,ij} \|_{r\geq s, i\geq j},$$

where $X = (X_1, \cdots, X_k)$

(6.38)
$$a_{rs,ij} = 0, \quad \text{unless } r = i,\ s = j,$$
$$a_{rs,rs} = E\left\{\frac{X_1^2X_2^2}{|X|^2}\left(\frac{\gamma_0'}{\gamma_0}\right)(|X|)\right\}, \quad r \neq s,$$
$$a_{rr,rr} = E\left\{\frac{X_1^2}{|X|}\frac{\gamma_0'}{\gamma_0}(|X|) + 1\right\}^2.$$

(6.39) $$E\left\{\frac{X_1^2X_2^2}{|X|^2}\left(\frac{\gamma_0'}{\gamma_0}\right)^2(|X|)\right\} = E\left\{\frac{X_1^2X_2^2}{|X|^4}\right\}E\left\{|X|^2\left(\frac{\gamma_0'}{\gamma_0}\right)^2(|X|)\right\}$$

by spherical symmetry of G_0. The second term in (6.39) is just $I_2(G_0)$, while the first term is independent of G_0 and may be shown to equal $k^{-1}(k+2)^{-1}$ by taking G_0 to be the spherical normal distribution. Thus

(6.40) $$a_{rs,rs} = k^{-1}(k+2)^{-1}I_2(G_0), \quad r \neq s.$$

A similar computation gives

(6.41) $$a_{rr,rr} = \frac{1}{4}E\left\{\frac{X_1^4}{|X|^2}\left(\frac{\gamma_0'}{\gamma_0}\right)^2(|X|)\right\} - 1 = \frac{1}{4}3k^{-1}(k+2)^{-1}I_2(G_0) - 1.$$

We see from (6.37) that $I(0, [J], G_0)$ is a diagonal matrix with entries given by (6.40) and (6.41). Upon inverting it and substituting (6.40) and (6.41) in $\dot{\ell}(x, 0, [J], G_0)$, we obtain (4.33) and (4.34).

6.4 *Two Theorems on efficient estimates.*

THEOREM 6.1. *Under R suppose $\{\hat{\theta}_n\}$ are such that, for a given θ, $\mathscr{L}_{\theta_n}\{n^{1/2}(\hat{\theta}_n - \theta_n)\} \to \mathscr{N}(0, I^{-1}(\theta))$ whenever $n^{1/2}|\theta_n - \theta| \leq M$ for all n, $M < \infty$. Then,*

(6.42) $$n^{1/2}(\hat{\theta}_n - \theta) = n^{-1/2}\textstyle\sum_{i=1}^n \dot{\ell}(X_i, \theta)I^{-1}(\theta) + o_{p_\theta}(1).$$

NOTE. This claim is in fact valid in great generality if the local asymptotic normality (LAN) condition of Hájek (1972) holds with $\Delta_n(\theta)$ replacing $n^{-1/2}\sum_{i=1}^n \dot{\ell}(X_i, \theta)$. Moreover it is clear that everything is local so that the condition and conclusion need only hold at a point θ on which $\hat{\theta}_n$ can depend.

PROOF. Since the sequence of joint laws $\mathscr{L}_n$ of $n^{1/2}(\hat{\theta}_n - \theta)$ and $n^{-1/2}\sum_{i=1}^n \dot{\ell}(X_i, \theta)I^{-1}(\theta)$ is tight under P_θ it is enough to show that if $\mathscr{L}_{m_n}$ is any subsequence weakly convergent to $\mathscr{L}^*$ (say) then $\mathscr{L}^*$ must concentrate on the diagonal. by a contiguity and analyticity argument, see Roussas (1972, pages 136-141), we can show that the joint characteristic function $\phi^*(u, v)$ of $\mathscr{L}^*$ satisfies the equation

$$\phi^*(u, v) = \phi^*(u, 0)\exp\{-uI^{-1}(\theta)v^T\}\exp\{-\tfrac{1}{2}vI^{-1}(\theta)v^T\}$$

(Substitute $\Gamma = I(\theta)$, $h = vI^{-1}(\theta)$ in (3.11) of Roussas.) But, by hypothesis,

$$\phi^*(u, 0) = \exp\{-\tfrac{1}{2}uI^{-1}(\theta)u^T\}$$

so that

$$\phi^*(u, v) = \exp\{-\tfrac{1}{2}(u+v)I^{-1}(\theta)(u+v)^T\},$$

and the theorem follows. □

THEOREM 6.2. *If* R(i), R(ii) *and* UR(iii) *hold and if* $\bar{\theta}_n$ *is* $\sqrt{n}$*-consistent and discretized as in* (2.3) *and*

$$\hat{\theta}_n = \bar{\theta}_n + n^{-1} \sum_{j=1}^{n} \dot{\ell}(X_j, \bar{\theta}_n) I^{-}(\bar{\theta}_n),$$

then $\hat{\theta}_n$ *is efficient in the usual sense.*

PROOF. In view of the arguments leading to Theorem 4 of Le Cam (1968), it is enough to show that for θ regular and any sequence θ_n such that $n^{1/2} | \theta_n - \theta | \le M$ for all n

$$n^{-1/2} \sum_{i=1}^{n} \{\dot{\ell}(X_i, \theta_n) - \dot{\ell}(X_i, \theta)\} + n^{1/2}(\theta_n - \theta) I(\theta) = o_{P_\theta}(1). \tag{6.43}$$

We claim that (6.43) is implied by the fact that

$$\sum_{i=1}^{n} \{\ell(X_i, \theta_n + hn^{-1/2}) - \ell(X_i, \theta_n)\} = hn^{-1/2} \sum_{i=1}^{n} \dot{\ell}(X_i, \theta_n) - \tfrac{1}{2} h I(\theta_n) h^T + o_{P_\theta}(1) \tag{6.44}$$

for all h. To see this, note that from the usual LAN condition

$$\begin{aligned}\sum_{i=1}^{n} \{\ell(X_i, \theta_n + hn^{-1/2}) - \ell(X_i, \theta)\} &= n^{1/2}(\theta_n - \theta) + hn^{-1/2} \sum_{i=1}^{n} \dot{\ell}(X_i, \theta) \\ &\quad -\tfrac{1}{2}\{n^{1/2}(\theta_n - \theta) + h\} I(\theta) \{n^{1/2}(\theta_n - \theta) + h\}^T + o_{P_\theta}(1);\end{aligned} \tag{6.45}$$

$$\begin{aligned}\sum_{i=1}^{n} \{\ell(X_i, \theta_n) - \ell(X_i, \theta)\} &= n^{1/2}(\theta_n - \theta) n^{-1/2} \sum_{i=1}^{n} \dot{\ell}(X_i, \theta) \\ &\quad - \frac{n}{2} \{(\theta_n - \theta) I(\theta) (\theta_n - \theta)^T\} + o_{P_\theta}(1).\end{aligned} \tag{6.46}$$

Subtracting (6.46) from (6.45) and matching the coefficient of h in (6.44) yields (6.43).

Finally, (6.44) is just the usual statement of LAN with θ replaced by θ_n. It is argued in exactly the same way as the usual equivalence,—see pages 54–63 of Roussas (1972) for example,—but, of course, we use the uniformity in UR(iii). The theorem follows. □

Acknowledgement V. Fabian and J. Hannan corrected my mistaken impression that R(i) − R(iii) were sufficient to establish the efficiency of $\bar{\theta}_n + n^{-1/2} \sum \dot{\ell}(X_i, \bar{\theta}_n)$. I am grateful to them for prompting me to prove Theorems 6.1 and 6.2 as well as other valuable comments. I am also grateful to Chris A. J. Klaassen for a careful reading of the paper resulting in several substantial corrections and to J. Pfanzagl for the Ibragimov-Hasminskii reference.

REFERENCES

BERAN, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2** 63–74.

BERAN, R. (1975). Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **28** 77–89.

BERAN, R. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6** 292–313.

BICKEL, P. J. (1981). Lectures on robustness and adaptation, to appear in 1979 *St. Flour Conference Lecture Notes in Mathematics* 876 Springer, Berlin.

BIRGÉ, L. (1980). Approximation dans les espaces métrique et théorie de l'estimation. Unpublished thesis, University of Paris.

DIONNE, L. (1981). Efficient nonparametric estimators of parameters in the general linear hypothesis. *Ann. Statist.* **9** 457–460.

FABIAN V. and HANNAN J. (1980). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. verw. Gebiete.* To appear.

HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.

HÁJEK, J. and SIDÁK, Z. (1967). *Theory of Rank Tests*, Academic, New York and Academia, Prague.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175–194. University of California Press, Berkeley.
HAMMERSTROM, T. (1978). Ph.D. Thesis, University of California, Berkeley.
HASMINSKII, P. Z. and IBRAGIMOV I. A. (1978). On the nonparametric estimation of functionals. *Proc. Second Prague Symp. Asympt. Statistics and Probability*, J. Jureçkova Ed., Prague.
HOGG, R. (1980). On adaptive robust inference. Tech. Report, Univ. of Iowa.
HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings Fifth Berkeley Symp. Math. Statist. Prob.* **1** 221–233 University of California Press, Berkeley.
HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Math. Statist.* **1** 799–821.
HUBER, P. J. (1977). Robust covariances *Statistical Decision Theory and Related Topics*, II. S. S. Gupta and D. S. Moore, eds. 165–192 Academic, New York.
KLAASSEN, C. (1980). Statistical performance of location estimators. Thesis, University of Leiden, Netherlands.
LE CAM, L. (1969). *Théorie Asymptotique de la Décision Statistique*, Les Presses de l'Univérsite de Montreal.
LEVITT, B. (1974). On optimality of some statistical estimates, *Proc. Prague Symp. Asympt. Statist.* 215–238 J. Jureckova Ed. Prague.
LINDSAY, B. (1978). Information in the presence of nuisance parameters. Thesis, University of Washington, Seattle.
LINDSAY, B. (1980). Nuisance parameters, mixture models, and the efficiency of partial likelihood. *Phil. Trans. Roy. Soc. London* **296** 639–665.
MARONNA, R. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.
ROUSSAS, G. (1972). *Contiguity of Probability Measures: Applications in Statistics.* Cambridge University Press.
STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 187–196. University of California Press.
STONE, C. (1975). Adaptive maximum likelihood estimation of a location parameter. *Ann. Statist.* **3** 267–284.
TAKEUCHI, K. (1971). A uniformly asymptotically efficient estimator of a location parameter, *J. Amer. Statist. Assoc.* **66** 292–301.
VAN EEDEN, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.* **41** 172–181.
WEISS, L. and WOLFOWITZ, J. (1970). Asymptotically efficient nonparametric estimators of location and scale parameters. *Z. Wahrsch. verw. Gebiete* **16** 134–150.
WEISS, L. and WOLFOWITZ, J. (1971). Asymptotically efficient estimation of nonparametric regression coefficients, *Statistical Decision Theory and Related Topics*, S. Gupta and J. Yackel, 29–40 eds. Academic, New York.
WOLFOWITZ, J. (1953). The method of maximum likelihood and the Wald theory of decision functions. *Indag. Mathemat.* **56** 114–119.
WOLFOWITZ, J. (1974). Asymptotically efficient nonparametric estimators of location and scale parameters. II. *Z. Wahrsch. verw. Gebiete* **30** 117–128.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720