The Annals of Statistics
1983, Vol. 11, No. 2, 432–452

INFORMATION AND ASYMPTOTIC EFFICIENCY IN PARAMETRIC—NONPARAMETRIC MODELS[1]

By Janet M. Begun, W. J. Hall,
Wei-Min Huang, and Jon A. Wellner

University of North Carolina at Chapel Hill, University of Rochester, Lehigh University, and University of Rochester

Asymptotic lower bounds for estimation of the parameters of models with both parametric and nonparametric components are given in the form of representation theorems (for regular estimates) and asymptotic minimax bounds. The methods used involve: (i) the notion of a "Hellinger-differentiable (root-) density", where part of the differentiation is with respect to the nonparametric part of the model, to obtain appropriate scores; and (ii) calculation of the "effective score" for the real or vector (finite-dimensional) parameter of interest as that component of the score function orthogonal to all nuisance parameter "scores" (perhaps infinite-dimensional). The resulting asymptotic information for estimation of the parametric component of the model is just (4 times) the squared L^2-norm of the "effective score". A corollary of these results is a simple necessary condition for "adaptive estimation": adaptation is possible only if the scores for the parameter of interest are orthogonal to the scores for the nuisance function or nonparametric part of the model. Examples considered include the one-sample location model with and without symmetry, mixture models, the two-sample shift model, and Cox's proportional hazards model.

1. Introduction. Lower bounds for estimation have a long history and play an important role in statistical theory. In parametric problems the representation theorem of Hájek (1970), and the closely related asymptotic minimax theorems of Hájek (1972) and LeCam (1972), provide a rather complete description of optimal estimation in the case of large sample sizes. More recently, representation theorems and asymptotic minimax theorems have been established for a variety of nonparametric problems including: (i) estimation of a differentiable functional of an unknown density or distribution function (Beran, 1977a, 1977b; Levit, 1978); (ii) estimation of a distribution function (Beran, 1977; Koshevnik and Levit, 1976; Millar, 1979; Wellner, 1982); and (iii) estimation of a spectral distribution (Levit and Samarov, 1978).

However, many common statistical models are of "mixed" type, incorporating both parametric and nonparametric components; see Oakes (1981), who uses the term "semiparametric" for such models, and the examples below. The study of lower bounds for such "mixed" or "semiparametric" models was initiated by Stein (1956), but has been relatively neglected since Stein's original investigation; exceptions are the work of Lindsay (1978, 1980) who studied the interesting subclass of mixture models (see Example 2 below), and Bickel (1982) who examined the possibility of "adaptive estimation" generally. A related but different set of problems, involving robust estimation in parametric models, has been recently investigated by Beran (1980, 1981), Millar (1981), and Rieder (1981).

Received March 1982; revised November 1982.

[1] Research supported in part by NSF Grant MCS 81-00748 (Begun), ONR/CNA Grant 5-29155 (Hall, Huang, Wellner), the Alexander von Humboldt Foundation (Wellner), and NSF Grant MCS 81-02731 (Wellner).

AMS 1980 *subject classifications.* 62E20, 62G05, 62G20.

Key words and phrases. Adaptation, asymptotic minimax bounds, Hellinger differentiable likelihood, projection, representation theorem.

Our aim in this paper is to fill this gap by providing asymptotic lower bounds for estimation of the parameters of models with both parametric and nonparametric components in the form of representation theorems and asymptotic minimax bounds. Section 3 contains our results for the parametric component; Section 4 gives bounds for estimates of the nonparametric component, and remarks on joint bounds. Our theorems in Section 3 also provide fairly general *necessary* conditions for "adaptive estimation"; these conditions are related to conditions given by Bickel (1982). Our primary aim, however, is to provide lower bounds for the (more usual) cases in which adaptation is *not* possible.

The key ideas involve the notion of a "Hellinger-differentiable likelihood" to obtain appropriate scores for both the parametric and nonparametric components of the model, and then orthogonal projection of the score function for the parametric, finite-dimensional component of the model onto the infinite-dimensional space generated by the nonparametric or nuisance-parameter scores. The L^2 norm of the component of the parametric scores that is orthogonal to the infinite-dimensional subspace of nuisance-parameter scores is shown to yield the "information" for the parametric component in such a model. These are natural extensions to mixed parametric-nonparametric models of their more familiar parametric counterparts; see Wilks (1962), page 377, Efron (1977), page 564, and Neyman (1958) for examples of the projection of scores of the parameters of interest on nuisance parameter scores in classical parametric estimation and testing contexts.

The following examples illustrate the types of problems that we have in mind. In each case, the primary question which concerns us is: How well can we estimate the (real) parameter θ? In Section 4 we also give asymptotic lower bounds for estimates of the distribution G associated with the density g.

EXAMPLE 1. *One-sample location model.* Suppose that $X_1, \cdots, X_n$ are iid with density $f = f(\cdot; \theta, g) = g(\cdot - \theta)$ with respect to Lebesgue measure μ on R^1 where $\theta \in R^1$ and g belongs to a class of densities $\mathscr{G}$ sufficiently small that θ is identifiable, and $I_g \equiv \int (\dot{g}^2/g)\, d\mu < \infty$. Here $\theta \in R^1$, the "location", is the parametric component of the model, while $g \in \mathscr{G}$ is the nonparametric component.

EXAMPLE 1a. $\mathscr{G} = \mathscr{G}_s \equiv \{\text{all densities } g \text{ on } R^1 \text{ symmetric about } 0\}$.

EXAMPLE 1b. $\mathscr{G} = \mathscr{G}_T \equiv \{\text{all densities } g \text{ on } R^1 \text{ with } T(g) = 0\}$ where T is a specified location functional in the sense of Bickel and Lehmann (1975), e.g. $T(g) = \text{median } (g)$. Note that $\mathscr{G}_s \subset \mathscr{G}_T$, since $T(g) = 0$ for all $g \in \mathscr{G}_s$: see Theorem 1 of Bickel and Lehmann (1975), and $\mathscr{G}_T$ is substantially larger than $\mathscr{G}_s$.

EXAMPLE 2. *Mixture models.* Suppose that $X_1, \cdots, X_n$ are iid with density $f = f(\cdot; \theta, g)$ on $\mathscr{X}$ given by $f(x; \theta, g) = \int M(x; \theta, \phi) g(\phi)\, d\phi$ where M is a fixed density function with parameters $(\theta, \phi) \in R^1 \times R^k$, and $g \in \mathscr{G}$, a class of densities sufficiently small that θ and g are identifiable. The model is identifiable if $(\theta, g) \neq (\theta^*, g^*)$ implies $P_{\theta,g} \neq P_{\theta^*,g^*}$ where $P_{\theta,g}(A) \equiv \int_A f(x; \theta, g)\, d\mu(x)$; see Kiefer and Wolfowitz (1956, page 891), Teicher (1961), and Section 6 for further discussion. See Kiefer and Wolfowitz (1956) and Lindsay (1980) for many interesting special cases of this model. Here θ is the parametric component of the model, and the "mixing density" g is the nonparametric component.

EXAMPLE 3. *Two-sample shift model.* Suppose that $X_{11}, \cdots, X_{1n_1}$ are iid $f_1 = f_1(\cdot; \theta, \eta, g) = g(\cdot - \eta)$, and $X_{21}, \cdots, X_{2n_2}$ are iid $f_2 = f_2(\cdot; \theta, \eta, g) = g(\cdot - \eta - \theta)$ where $\theta \in R^1$, $\eta \in R^1$, and $g \in \mathscr{G}$, a class of densities sufficiently small that θ and η are identifiable; e.g. $\mathscr{G} = \mathscr{G}_T$ of Example 1b. In this model θ, the "shift parameter", is the parametric part of the model of primary interest; η is a nuisance or incidental parametric part of the model; and g is the nonparametric part.

EXAMPLE 4. *Cox's regression model.* Cox (1972) introduced a convenient regression model for censored survival analysis based on the assumption

$$\lambda(t \mid z) = \exp(\theta z)\lambda(t) \tag{1.1}$$

where $\lambda(t \mid z) = g(t \mid z)/\bar{G}(t \mid z)$ is the hazard function, and $g(t \mid z)$ the density function, governing the survival time of an individual with regression variable z, and where $\lambda(t) = g(t)/\bar{G}(t)$ is an underlying hazard function with $g \in \mathscr{G}^+ \equiv$ {all densities with respect to Lebesgue measure ν on R^+}. (In this example, for any density function g, denote the corresponding survival function by $\bar{G} \equiv 1 - G \equiv \int_t^\infty g\, d\nu$.) The regression parameter $\theta \in R^1$ (or R^k) is the parametric part of the model, while g, or equivalently λ, is the nonparametric part. For a more complete description of this model with censored observations, see Section 6.

An important feature of mixed parametric-nonparametric or "semiparametric" models such as these is that many functionals of the density f (or associated df F) agree: in Example 1a, the median, mean (when it exists), symmetric trimmed means, and many other functionals all yield θ, the center of symmetry; in Example 3, the difference between any measure of location applied to f_1 and f_2 yields the shift θ; in the two-sample version of Example 4, Begun (1981) has given a large class of functionals which all equal the regression parameter θ. Hence the completely nonparametric approach to representation theorems and asymptotic minimax bounds for estimating functionals of f or F taken by Beran (1977a, 1977b) and Levit (1978) fails for these models: it is simply not clear *which* functional should be analyzed. Our approach here is based instead on analysis of the local asymptotic behavior of the appropriate likelihoods.

A referee has suggested the following program for carrying through the functional approach even when many functionals agree: calculate the asymptotic minimax bound for every such (differentiable) functional, and then minimize this bound over all such functionals subject to the linear constraints on derivatives imposed by agreement of the functionals at the model. It may be possible to carry through this program in particular cases, but it seems to us to be a cumbersome method. We prefer the present approach and methods which seem to be more consistent with the philosophy that "the likelihoods tell the story."

2. Preliminaries; differentiable likelihoods. For notational and expositional ease, we will concentrate on the simplest but most important type of mixed parametric-nonparametric model with a single real-valued parametric component $\theta \in R^1$ and a single nonparametric component $g \in \mathscr{G}$ where $\mathscr{G}$ is a specified set of density functions. Our treatment of this case extends without difficulty to problems involving a finite-dimensional parametric component $\boldsymbol{\theta} \in R^r$ and a parametric nuisance component $\boldsymbol{\eta} \in R^s$, as is the case in Example 3. These extensions are sketched in the remarks at the end of this and the following section; further details and extensions, investigation of other differentiability conditions, and more examples are given in Huang (1982).

Suppose that $X_1, \cdots, X_n$ are iid $\mathscr{X}$-valued random variables with density function $f = f(\cdot\,; \theta, g)$ with respect to a sigma-finite measure μ on the measurable space $(\mathscr{X}, \mathscr{C})$ where $\theta \in R^1$ and $g \in \mathscr{G} \subset$ the collection of all densities with respect to a sigma-finite measure ν on some measurable space $(\mathscr{Y}, \mathscr{D})$. In many applications $\mathscr{X}$ and $\mathscr{Y}$ are some Euclidean space R^k and μ and ν are Lebesgue measure on R^k. Let $L^2(\mu) = L^2(\mathscr{X}, \mu)$ and $L^2(\nu) = L^2(\mathscr{Y}, \nu)$ denote the usual L^2-spaces of square-integrable functions and let $\langle \cdot, \cdot \rangle_\mu (\| \cdot \|_\mu)$ and $\langle \cdot, \cdot \rangle_\nu (\| \cdot \|_\nu)$ denote the usual inner products (and norms) in $L^2(\mu)$ and $L^2(\nu)$ respectively. Thus $f^{1/2} \in L^2(\mu)$, $g^{1/2} \in L^2(\nu)$, and $\| f^{1/2} \|_\mu = 1 = \| g^{1/2} \|_\nu$.

Our analysis of this mixed model will be framed, for convenience of comparison with the more familiar parametric case, in terms of the differentiability of f with respect to its parameters θ and g. An alternative (but less transparent) treatment would involve only the assumption that the local likelihood ratios behave as in Lemma 2.1 below with α as given in Proposition 2.1.

DEFINITION 2.1. (Hellinger differentiability of $f^{1/2}$). The root density $f^{1/2} =$

$f^{1/2}(\cdot\,;\theta, g)$ is said to be *Hellinger-differentiable* at $(\theta, g) \in R^1 \times \mathscr{G}$ if there exists a function $\rho_\theta \in L^2(\mu)$ and a bounded linear operator A: $L^2(\nu) \to L^2(\mu)$ such that, with $f_n \equiv f(\cdot\,;\theta_n, g_n)$,

$$\frac{\|f_n^{1/2} - f^{1/2} - \{\rho_\theta \cdot (\theta_n - \theta) + A(g_n^{1/2} - g^{1/2})\}\|_\mu}{|\theta_n - \theta| + \|g_n^{1/2} - g^{1/2}\|_\nu} \to 0 \quad \text{as} \quad n \to \infty \tag{2.1}$$

for all sequences $\theta_n \to \theta$ and $g_n^{1/2} \to g^{1/2}$ in $L^2(\nu)$, where $g_n \in \mathscr{G}$ for all $n \geq 1$; the expression inside the brackets { } in (2.1) is the *differential* at (θ, g).

If g is regarded as fixed and "known", then ρ_θ is typically just the usual parametric score function for θ, i.e. $(\partial/\partial\theta)\log f(\cdot\,;\theta, g)$, times $\frac{1}{2} f^{1/2}$; here, as in LeCam (1970), we only require differentiability in the L^2-sense. The operator A can be regarded as yielding a "score for g".

Now let $\Theta(h)$ denote the collection of all sequences $\{\theta_n\}_{n\geq 1}$ such that $|n^{1/2}(\theta_n - \theta) - h| \to 0$ as $n \to \infty$ where $h \in R^1$, and let $\Theta \equiv \cup\,\{\Theta(h) : h \in R^1\}$. Similarly, let $\mathscr{C}(g, \beta)$ denote the collection of all sequences $\{g_n\}_{n\geq 1}$ with each $g_n \in \mathscr{G}$ such that $\|n^{1/2}(g_n^{1/2} - g^{1/2}) - \beta\|_\nu \to 0$ as $n \to \infty$ where $\beta \in L^2(\nu)$ (and $\beta \perp g^{1/2}$ necessarily), let

$$\begin{aligned}\mathscr{B} \equiv \{\beta \in L^2(\nu) : \|n^{1/2}(g_n^{1/2} - g^{1/2}) - \beta\|_\nu \to 0 \quad &\text{as} \quad n \to \infty \\ &\text{for some sequence } g_n \text{ with all } g_n \in \mathscr{G}\},\end{aligned} \tag{2.2}$$

and let $\mathscr{C}(g) \equiv \cup\{\mathscr{C}(g, \beta) : \beta \in \mathscr{B}\}$; $\mathscr{C}(g)$ is the set of sequences converging to g in any of the possible "directions" $\beta \in \mathscr{B}$. Note that the subset $\mathscr{B}$ of $L^2(\nu)$ is closed (by an easy argument) and that it depends heavily on $\mathscr{G}$: we have insisted that all of the g_n's converging to g belong to $\mathscr{G}$. Our subsequent development will rely on the following Assumption about $\mathscr{B}$:

ASSUMPTION S. The set $\mathscr{B}$ defined in (2.2) is a subspace of $L^2(\nu)$.

This assumption amounts to saying that we can parameterize $\mathscr{G}$ locally by a subspace of $L^2(\nu)$. Assumption S is related to the convexity assumption (C) of Bickel (1982), and to the "extensiveness" hypothesis of Levit (1978, Condition 5). It is also related to the considerations on pages 246–247 of Millar (1979). Although Assumption S imposes some restrictions on our methods, it seems to be weaker than Bickel's convexity hypothesis and to be satisfied in many cases of interest.

The following proposition is an immediate consequence of Hellinger differentiability of $f^{1/2}$:

PROPOSITION 2.1. *Suppose $f^{1/2}$ is Hellinger-differentiable at $(\theta, g) \in R^1 \times \mathscr{G}$, and that $\{(\theta_n, g_n)\}_{n\geq 1} \in \Theta(h) \times \mathscr{C}(g, \beta)$ for some $h \in R^1$, $\beta \in L^2(\nu)$. Then, with $f_n \equiv f(\cdot\,;\theta_n, g_n)$ and $f \equiv f(\cdot\,;\theta, g)$,*

$$\|n^{1/2}(f_n^{1/2} - f^{1/2}) - \alpha\|_\mu \to 0 \quad \text{as} \quad n \to \infty \tag{2.3}$$

where $\alpha \in L^2(\mu)$ is given by

$$\alpha = h\rho_\theta + A\beta. \tag{2.4}$$

Note that $\alpha \perp f^{1/2}$ since $\|f^{1/2}\|_\mu = 1 = \|f_n^{1/2}\|_\mu$ for all $n \geq 1$. Let

$$H \equiv \{\alpha \in L^2(\mu) : \alpha = h\rho_\theta + A\beta \quad \text{for some} \quad h \in R^1, \beta \in \mathscr{B}\}. \tag{2.5}$$

Under Assumption S, H is a subspace of $L^2(\mu)$ since it is the image of a subspace (of $R^1 \times L^2(\nu)$) under a bounded linear transformation. For $\alpha \in H$ let $\mathscr{F}(f, \alpha)$ denote the collection of all sequences $\{f_n\}$ such that (2.3) holds and let $\mathscr{F}(f)$ denote the union of the $\mathscr{F}(f, \alpha)$'s over $\alpha \in H$.

For $\{f_n\}_{n\geq 1}$ and f as in Proposition 2.1, define the local log likelihood ratio L_n, whenever the right side is finite, by

$$L_n = 2\log\{\textstyle\prod_{i=1}^n [f_n^{1/2}(X_i)/f^{1/2}(X_i)]\}. \tag{2.6}$$

The following Lemma describing the asymptotic behavior of L_n is a consequence of arguments of LeCam (1969); it has been used repeatedly by Beran (1977a, 1977b).

LEMMA 2.1. (*Local asymptotic normality*). *If f_n and f as in Proposition* 2.1 *satisfy* (2.3), *then, for every $\varepsilon > 0$,*

$$P_f\{|L_n - 2n^{-1/2}\textstyle\sum_{i=1}^n \alpha(X_i)f^{-1/2}(X_i) + \tfrac{1}{2}\sigma^2| > \varepsilon\} \to 0 \quad as \quad n \to \infty \tag{2.7}$$

where $\sigma^2 = 4\|\alpha\|_\mu^2$. Thus, under P_f,

$$L_n \to_d N(-\tfrac{1}{2}\sigma^2, \sigma^2) \quad as \quad n \to \infty \tag{2.8}$$

and the sequences $\{\prod_{i=1}^n f_n(x_i)\}$ and $\{\prod_{i=1}^n f(x_i)\}$ are contiguous.

REMARK 2.1. When the model is given by a density $f = f(\cdot\,; \boldsymbol{\theta}, \boldsymbol{\eta}, g)$ where $\boldsymbol{\theta} \in R^r$, $\boldsymbol{\eta} \in R^s$, and $g \in \mathscr{G}$, Definition 2.1 and Proposition 2.1 must be altered as follows: with $f_n \equiv f(\cdot\,; \boldsymbol{\theta}_n, \boldsymbol{\eta}_n, g_n)$ and $|\boldsymbol{\theta}_n - \boldsymbol{\theta}| \to 0$, $|\boldsymbol{\eta}_n - \boldsymbol{\eta}| \to 0$, $\|g_n^{1/2} - g^{1/2}\|_\nu \to 0$ as $n \to \infty$ where $|\cdot|$ denotes the usual Euclidean norm, instead of (2.1) require that

$$\frac{\|f_n^{1/2} - f^{1/2} - [\boldsymbol{\rho}_\theta\cdot(\boldsymbol{\theta}_n - \boldsymbol{\theta}) + \boldsymbol{\rho}_\eta\cdot(\boldsymbol{\eta}_n - \boldsymbol{\eta}) + A(g_n^{1/2} - g^{1/2})]\|_\mu}{|\boldsymbol{\theta}_n - \boldsymbol{\theta}| + |\boldsymbol{\eta}_n - \boldsymbol{\eta}| + \|g_n^{1/2} - g^{1/2}\|_\nu} \to 0 \tag{2.1$'$}$$

as $n \to \infty$ for some r-vector $\boldsymbol{\rho}_\theta$ of functions in $L^2(\mu)$ (the scores for $\boldsymbol{\theta}$), s-vector $\boldsymbol{\rho}_\eta$ of functions in $L^2(\mu)$ (the scores for $\boldsymbol{\eta}$), and A a bounded linear transformation as in Definition 2.1. The required change in Proposition 2.1 is that (2.4) must be replaced by

$$\alpha = \boldsymbol{h}\cdot\boldsymbol{\rho}_\theta + \boldsymbol{k}\cdot\boldsymbol{\rho}_\eta + A\beta \tag{2.4$'$}$$

where $|n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}) - \boldsymbol{h}| \to 0$, $|n^{1/2}(\boldsymbol{\eta}_n - \boldsymbol{\eta}) - \boldsymbol{k}| \to 0$, and $\|n^{1/2}(g_n^{1/2} - g^{1/2}) - \beta\|_\nu \to 0$ as $n \to \infty$ for some $(\boldsymbol{h}, \boldsymbol{k}, \beta) \in R^r \times R^s \times \mathscr{B}$.

3. Main results: parametric component. Now we turn to asymptotic lower bounds for estimation of θ in the presence of the unknown nuisance parameter g. If the density g were known, $g = g_0$ say, then the results of Hájek (1970, 1972) together with Lemma 1 guarantee that any (regular) estimator of θ has a limiting distribution at least as dispersed as $N(0, 1/I_0)$ where $I_0 = 4\|\rho_\theta\|_\mu^2$ is the usual parametric Fisher information for θ. When g is unknown, the "information for θ" will, in general, be smaller than when g is known, and it will be harder to estimate θ. The questions are: How much smaller is the information? And how much larger will the asymptotic variance of a "best" estimate be? In terms of adaptive estimation (to be defined carefully below), is "adaptation" possible?

The main thesis of this paper is that asymptotic lower bounds for estimation of θ when g is unknown are determined by the geometry of the scores as given by (2.4): Finding the "information" for estimation of θ in the presence of nuisance parameters requires (orthogonal) projection of the score for the parameter of interest onto the space of nuisance parameter scores $\{A\beta : \beta \in \mathscr{B}\}$, thereby yielding the "effective" component of ρ_θ orthogonal to the nuisance parameter scores.

We now make this brief description more precise. By Assumption S, $\{A\beta : \beta \in \mathscr{B}\}$ is a closed subspace of H defined in (2.5). Hence, by the classical projection theorem (see e.g. Luenberger, 1969, page 51) there exists a $\beta^* \in \mathscr{B}$ minimizing $\|\rho_\theta - A\beta\|_\mu^2$; this β^* satisfies

$$(\rho_\theta - A\beta^*) \perp A\beta \quad \text{for all} \quad \beta \in \mathscr{B}. \tag{3.1}$$

Thus α of (2.4) can be decomposed as

$$\alpha = h(\rho_\theta - A\beta^*) + A(h\beta^* + \beta) \tag{3.2}$$

and hence

$$\begin{aligned}(3.3)\qquad \|\alpha\|_\mu^2 &= \|h\rho_\theta - A(-\beta)\|_\mu^2 = h^2\|\rho_\theta - A\beta^*\|_\mu^2 + \|A(h\beta^* + \beta)\|_\mu^2 \quad \text{by (3.1)} \\ &\geq h^2\|\rho_\theta - A\beta^*\|_\mu^2 \quad \text{for all} \quad \beta \in \mathscr{B}\end{aligned}$$

with equality if and only if $\beta = -h\beta^*$. The "geometric picture" should now be clear: $A\beta^*$ is the projection of ρ_θ onto $\{-A\beta : \beta \in \mathscr{B}\}$; therefore $\beta = -h\beta^*$ minimizes $\|\alpha\|_\mu^2 = \|h\rho_\theta - A(-\beta)\|_\mu^2$ and represents a "least favorable" or worst possible direction of approach to g for the problem of estimating θ. In the language of Stein (1956), β^* yields the most difficult one-dimensional sub-problem. The "effective score for θ" is now $\rho_\theta - A\beta^*$, and four times the square of its norm in $L^2(\mu)$ is the asymptotic "information" I_* for estimation of θ in the presence of g:

$$(3.4)\qquad I_* = 4\|\rho_\theta - A\beta^*\|_\mu^2.$$

The following representation theorem (Theorem 3.1) and asymptotic minimax lower bounds (Theorem 3.2) for estimates of θ make this precise.

We say that an estimator $\hat{\theta}_n$ of θ is *regular* (or $\Theta \times \mathscr{G}$-regular) *at* $f = f(\cdot\,; \theta, g)$ if, for every sequence $\{f_n = f(\cdot\,; \theta_n, g_n)\}_{n\geq 1}$ with $\{(\theta_n, g_n)\}_{n\geq 1} \in \Theta \times \mathscr{C}(g)$, the distribution of $n^{1/2}(\hat{\theta}_n - \theta_n)$ (under f_n) converges weakly to a law $\mathscr{L} = \mathscr{L}(f)$ which depends on f (and hence θ and g) but not on the particular sequence $\{(\theta_n, g_n)\}$. Thus $\mathscr{L} = \mathscr{L}(f)$ does not depend on h or β. Regularity of an estimator $\hat{\theta}_n$ in this sense is a local stability property in the spirit of Hájek (1970) and Beran (1977a, 1977b).

THEOREM 3.1. *Suppose that $\hat{\theta}_n$ is a regular estimator of θ in the model $f = f(\cdot\,; \theta, g)$ with limit law $\mathscr{L} = \mathscr{L}(f)$, that the conclusion of Proposition* 2.1 *holds with α given by* (2.4), *and that Assumption S holds. Then $\mathscr{L}$ may be represented as the convolution of a $N(0, 1/I_*)$ distribution with $\mathscr{L}_1 = \mathscr{L}_1(f)$, a distribution depending only on $f = f(\cdot\,; \theta, g)$, where I_* is given by* (3.4).

Equivalently, if S, Z_*, and W denote random variables with laws $\mathscr{L}$, $N(0, 1/I_*)$, and $\mathscr{L}_1$ respectively,

$$(3.5)\qquad S =_d Z_* + W,$$

where Z_* and W are independent.

To state our asymptotic minimax bound, which is a special case of the general Hájek-LeCam asymptotic minimax theorem (see Proposition 2.1 of Millar (1979) for a nice restatement), we introduce a loss function ℓ: $R^1 \to R^+$ which will be assumed to be subconvex (i.e. $\{x : \ell(x) \leq y\}$ is closed, convex, and symmetric for every $y \geq 0$) and to satisfy

$$(3.6)\qquad \int_{-\infty}^{\infty} \ell(z)\phi(\lambda z)\, dz < \infty \quad \text{for all} \quad \lambda > 0,$$

where ϕ denotes the standard normal density function.

THEOREM 3.2. *Suppose that the conclusion of Proposition* 2.1 *holds, that Assumption S holds, and that ℓ is subconvex and satisfies* (3.6). *Then, with $B_n(c) \equiv \{f_n \in \mathscr{F}(f) : n^{1/2}\|f_n^{1/2} - f^{1/2}\|_\mu \leq c\}$,*

$$(3.7)\qquad \lim_{c\to\infty}\lim_{n\to\infty}\inf_{\hat{\theta}_n}\sup_{f_n\in B_n(c)} E_{f_n}\ell(n^{1/2}(\hat{\theta}_n - \theta_n)) \geq E\ell(Z_*)$$

where $Z_* \sim N(0, 1/I_*)$ and I_* is given by (3.4).

Here the infimum over estimates $\hat{\theta}_n$ is taken over the class of "generalized procedures," the closure of the class of randomized (Markov kernel) procedures, as explained in Millar (1979, page 235).

REMARK 3.1. Although (3.1) does not provide a concrete recipe for finding the important "least favorable" $\beta^* \in \mathscr{B}$ and hence I_*, it follows from classical L^2 theory (e.g. Theorem 1, page 160, Luenberger, 1969) that β^* satisfies the "normal equation"

$$A^*A\beta^* = A^*\rho_\theta \tag{3.8}$$

where A^* denotes the adjoint of the bounded linear operator A. This is precisely analogous to the familiar (finite-dimensional) normal equations encountered in standard linear model theory where the role of A is played by a "design matrix" X, and that of A^* by the transpose X^T. Hence $\beta^* = (A^*A)^{-1}A^*\rho_\theta$ whenever A^*A is invertible. In the examples with which we are familiar, determination of β^*, and hence I_*, has previously proceeded by guesswork, calculus of variations techniques, or finite-dimensional approximations (see Efron (1977) together with Begun and Wellner (1982) for an example of the latter), followed by direct verification of the key orthogonality relationship (3.1). See the examples given in Section 6.

Before proceeding to the proofs of Theorems 3.1 and 3.2, we want to briefly outline their implications for *adaptive estimation.* Recall that

$$I_0 = 4\|\rho_\theta\|_\mu^2 \tag{3.9}$$

is the usual parametric information for θ when g is *known.*

DEFINITION 3.1. A sequence of estimates $\{\hat{\theta}_n\}$ of θ is said to be *$\mathscr{G}$-adaptive* (or simply adaptive) if, under $f_n = f(\cdot\,; \theta_n, g_n)$,

$$n^{1/2}(\hat{\theta}_n - \theta_n) \to_d N(0, 1/I_0) \quad \text{as} \quad n \to \infty \tag{3.10}$$

for all $\{(\theta_n, g_n)\}_{n\geq 1} \in \Theta \times \mathscr{C}(g)$ and all (θ, g) with $0 < I_0 < \infty$.

Thus an adaptive estimator, *without knowledge of* $g \in \mathscr{G}$, estimates θ as well asymptotically as if g were *known*, uniformly in local (shrinking) neighborhoods of each $g \in \mathscr{G}$. This is a more stringent definition of an adaptive estimator than the definitions which seem to be accepted in the current literature; see also Fabian and Hannan (1982). For example, Bickel (1982) only requires that (3.10) hold for all $f_n = f(\cdot\,; \theta_n, g)$, $\{\theta_n\} \in \Theta$ with g fixed; this is in keeping with most work on adaptive estimation in the cases of Examples 1a or 3 (e.g. Stone, 1975). Beran (1978) has shown in the context of Example 1a however, that estimates which are adaptive in the sense of Definition 3.1 exist in this case. Beran's result bolsters our feeling that Definition 3.1 captures the local uniformity that should reasonably be required of adaptive estimates when $g \in \mathscr{G}$ is unknown. In this connection, see Section 5(e) of Bickel (1982), and the work of Klaassen (1981) which indicates that we can *not* require the convergence in (3.10) to be uniform in *fixed* neighborhoods of $g \in \mathscr{G}$.

Comparison of (3.4) and (3.9) shows immediately that

$$I_* \leq I_0; \tag{3.11}$$

and from the definition of β^* it is clear that equality holds in (3.11) if and only if

$$\rho_\theta \perp A\beta \quad \text{for all} \quad \beta \in \mathscr{B}. \tag{3.12}$$

Thus (3.12) provides a necessary condition for adaptive estimation:

COROLLARY 3.1. *If Assumption S holds, and the conclusion of Proposition* 2.1 *holds with* α *given by* (2.4), *then a necessary condition for the existence of a sequence of adaptive estimates is that* (3.12) *hold for all* (θ, g) *such that* $0 < I_0 < \infty$.

PROOF. This follows immediately from Theorem 3.1 and Definition 3.1 upon noting that $I_* < I_0$ if (3.12) fails. □

REMARK 3.2. When the model is given by a density $f(\cdot; \boldsymbol{\theta}, \boldsymbol{\eta}, g)$, where $\boldsymbol{\theta} \in R^r$, $\boldsymbol{\eta} \in R^s$, $g \in \mathscr{G}$ as in Remark 2.1, which is differentiable in the sense that (2.3) holds with α given by (2.4)′ of Remark 2.1, then the arguments and Theorem statements of this section require slight modifications: the scores for $\boldsymbol{\theta}$, the parameters of interest, must first be projected onto the scores $\boldsymbol{\rho}_\eta$ for the (parametric) nuisance parameters $\boldsymbol{\eta}$. This part is easy; introducing the $r \times s$ matrix B and the $s \times s$ matrix D defined by

$$B = \langle \boldsymbol{\rho}_\theta, \boldsymbol{\rho}_\eta^T \rangle_\mu, \quad D = \langle \boldsymbol{\rho}_\eta, \boldsymbol{\rho}_\eta^T \rangle_\mu,$$

then, supposing D nonsingular with inverse D^{-1}, the part of $\boldsymbol{\rho}_\theta$ orthogonal to the space generated by $\boldsymbol{\rho}_\eta$ is just $\boldsymbol{\rho}_{\theta\cdot\eta} \equiv (\boldsymbol{\rho}_\theta - BD^{-1}\boldsymbol{\rho}_\eta) \perp \boldsymbol{\rho}_\eta$; thus $\boldsymbol{\rho}_{\theta\cdot\eta} = (\boldsymbol{\rho}_\theta - BD^{-1}\boldsymbol{\rho}_\eta)$ is the "effective score for $\boldsymbol{\theta}$" in the presence of the nuisance parameter $\boldsymbol{\eta}$ in the parametric problem with g fixed and known. The argument then proceeds as before by projecting each of the r components of $\boldsymbol{\rho}_{\theta\cdot\eta}$ onto $\{A\beta: \beta \in \mathscr{B}\}$: i.e. find $\boldsymbol{\beta}^* = (\beta_1^*, \cdots, \beta_r^*)$, $\beta_i^* \in \mathscr{B}$, such that

$$(\boldsymbol{\rho}_{\theta\cdot\eta} - A\boldsymbol{\beta}^*) \perp A\beta \quad \text{for all} \quad \beta \in \mathscr{B}. \tag{3.1$'$}$$

Then, defining the $r \times r$ matrix I_* by

$$I_* = 4\langle \boldsymbol{\rho}_{\theta\cdot\eta} - A\boldsymbol{\beta}^*, (\boldsymbol{\rho}_{\theta\cdot\eta} - A\boldsymbol{\beta}^*)^T \rangle_\mu, \tag{3.4$'$}$$

the vector versions of Theorems 3.1 and 3.2 have statements with the number $1/I_*$ replaced by the matrix I_*^{-1} (assuming the inverse exists) with I_* given by (3.4)′, and the random variable Z_* replaced by a random vector $\mathbf{Z}_* \sim N_r(\mathbf{0}, I_*^{-1})$. Paralleling the discussion of adaptation for the simple model $f(\cdot; \theta, g)$, a necessary condition for adaptation to $g \in \mathscr{G}$ for the model $f(\cdot; \boldsymbol{\theta}, \boldsymbol{\eta}, g)$ is that the "effective scores for $\boldsymbol{\theta}$", $\boldsymbol{\rho}_{\theta\cdot\eta}$, satisfy

$$\boldsymbol{\rho}_{\theta\cdot\eta} = \boldsymbol{\rho}_\theta - BD^{-1}\boldsymbol{\rho}_\eta \perp A\beta \quad \text{for all} \quad \beta \in \mathscr{B}. \tag{3.12$'$}$$

Of course the analogue of the number I_0 in the vector version of the problem is the $r \times r$ matrix

$$I_0 \equiv 4\langle \boldsymbol{\rho}_{\theta\cdot\eta}, \boldsymbol{\rho}_{\theta\cdot\eta}^T \rangle_\mu. \tag{3.9$'$}$$

In this connection, note that Stein's (1956) condition for adaptation $C = BD^{-1}E$ can be rewritten as $E(\mathbf{S}_\theta - BD^{-1}\mathbf{S}_\eta)S_\gamma = 0$ $(r \times 1)$ where $\mathbf{S}_\theta$, $\mathbf{S}_\eta$, and S_γ denote the classical scores (derivatives of log density with respect to the parameters); i.e. $\mathbf{S}_\theta - BD^{-1}\mathbf{S}_\eta \perp S_\gamma$.

REMARK 3.3. Although we have not attempted to construct a general class of estimators which achieve the asymptotic bounds given in Theorems 3.1, 3.2, and the following section, it seems very likely that "generalized" or nonparametric maximum likelihood estimators as defined by Kiefer and Wolfowitz (1956) (see Scholz, 1980) will typically be asymptotically fully efficient and attain our bounds under conditions only slightly stronger than the differentiability condition of Definition 2.1. This would seem to be especially the case when the necessary condition (3.12) for adaptive estimation fails; i.e. when adaptation to $g \in \mathscr{G}$ is *not* possible. When (3.12) holds, so that the possibility of adapation is not ruled out, the method of nonparametric maximum likelihood estimation apparently breaks down, and construction of estimators which achieve our bounds typically involves estimation of the score function ρ_θ or $\rho_\theta/f^{1/2}$; this is certainly the case for the asymptotically efficient adaptive estimators of Stone (1975) and Beran (1974) for Examples 1a and 3, and also, more generally, for Bickel's (1982) adaptive estimators, which rely upon his condition (H) that the scores can be estimated sufficiently well. Our point is that estimation in situations in which the necessary condition (3.12) for adaptation holds seems to be qualitatively different (and more difficult) than estimation in non-adaptive situations where (3.12) fails; we feel that *both* cases are of interest and importance.

Of course other families of nonparametric estimators, such as minimum distance estimators as in Beran (1978) and Millar (1981a), or the sieve estimators of Grenander (1981) (see also Geman and Hwang, 1982), may achieve the bounds given here and hence be proven to be asymptotically fully efficient. The asymptotic efficiency of these and other nonparametric estimators deserves further study.

A promising method of estimation of θ proposed by Huang (1982) which generalizes the approach of Stone (1975), is the following: let $\psi^*(\cdot; \theta, g) \equiv (\rho_\theta - A\beta^*)/f^{1/2} \in L^2(F)$ denote the effective score for θ (divided by $f^{1/2}$), and consider finding $\hat{\theta}$ such that the "effective score equation"

$$\int \psi^*(x; \hat{\theta}, \hat{g})\, d\,\mathbb{F}_n(x) = n^{-1} \textstyle\sum_{i=1}^n \psi^*(X_i; \hat{\theta}, \hat{g}) = 0$$

is (asymptotically) satisfied, where $\mathbb{F}_n$ is the empirical measure of X's iid f and $\hat{g}$ is some suitable initial estimator of g. Preliminary results of Huang (1982) indicate that $\hat{\theta}$ attains the bounds given in Section 3 in great generality. This will be explored fully elsewhere.

4. Main results: nonparametric component. Our object in this section is to give asymptotic lower bounds for estimation of $G = \int_{-\infty}^{\cdot} g d\nu$, the continuous distribution function corresponding to $g \in \mathscr{G}$, in the special case when g is a density function on $\mathscr{Y} = R^1$. We will also briefly indicate lower bounds for estimation of (θ, G) jointly, and for estimates of functions of (θ, G). For simplicity, we suppose that the support of g is contained in the unit interval $[0, 1]$. This can usually be accomplished by means of a fixed strictly continuous mapping of R^1 into $[0, 1]$, such as a probability integral transformation.

In our initial formulation of the lower bounds, we will also suppose that the subset $\mathscr{B}$ of $L^2(\nu)$ defined in (2.2) is exactly the subspace

$$\mathscr{B}_0 \equiv \{\beta \in L^2(\nu) \colon \beta \perp g^{1/2}\}. \tag{4.1}$$

When $\mathscr{B}$ is a proper subspace of $\mathscr{B}_0$, a corresponding projection operator is required; these modifications will be indicated later in this section.

Recall that $A^*\colon L^2(\mu) \to L^2(\nu)$ denotes the adjoint of the linear operator A. Our lower bounds for estimation of G will require the following invertibility assumption:

ASSUMPTION I. The linear operator $A^*A\colon L^2(\nu) \to L^2(\nu)$ is invertible with bounded inverse $(A^*A)^{-1}$.

Of course $(A^*A)^{-1}$ is necessarily a linear operator; see e.g. Luenberger (1969, page 147).

Suppose that Assumption I holds, and define $C\colon L^2(\nu) \to L^2(\mu)$ by

$$C \equiv A(A^*A)^{-1}. \tag{4.2}$$

Note that $C^* = (A^*A)^{-1}A^*$ and $C^*C = (A^*A)^{-1}$. Let $G_s \equiv (1_{[0,s]} - G(s))g^{1/2}$, and define the covariance function K on $[0, 1] \times [0, 1]$ by

$$K(s, t) \equiv \langle CG_s, CG_t \rangle_\mu = \langle G_s, (A^*A)^{-1}G_t \rangle_\nu. \tag{4.3}$$

Now let $\mathbb{Z}$ be a zero-mean Gaussian process on $[0, 1]$ with covariance function K given by (4.3), let $Z_* \sim N(0, 1/I_*)$ be independent of $\mathbb{Z}$, and let $\mathbb{Z}_*$ be the zero-mean Gaussian process on $[0, 1]$ defined by

$$\mathbb{Z}_*(t) \equiv \mathbb{Z}(t) - Z_* \int_0^t 2\beta^* g^{1/2}\, d\nu \quad \text{for} \quad 0 \le t \le 1 \tag{4.4}$$

where β^* satisfies (3.1); $\mathbb{Z}_*$ has covariance function $K_*(s, t) = K(s, t) + 4I_*^{-1} \int_0^s \beta^* g^{1/2} d\nu \int_0^t \beta^* g^{1/2}\, d\nu$. Since $\mathbb{Z}$ is Gaussian and $(A^*A)^{-1}$ is a bounded operator by Assumption I, for any $0 \le s, t \le 1$ we have

$$E|\mathbb{Z}(t) - \mathbb{Z}(s)|^4 = 3\{E|\mathbb{Z}(t) - \mathbb{Z}(s)|^2\}^2 = 3\{\langle (G_t - G_s), (A^*A)^{-1}(G_t - G_s)\rangle_\nu\}^2$$
$$\le 3\|(A^*A)^{-1}\|^2\{G(t) - G(s)\}^2.$$

Hence, by Theorem 12.4 of Billingsley (1968), $\mathbb{Z}$ and $\mathbb{Z}_*$ have continuous sample paths.

In parallel to Section 3, a (continuous) estimator $\hat{\mathbb{G}}_n$ of G is said to be **regular at $f = f(\cdot;\, \theta, g)$** if, for every sequence $\{f_n\} = \{f(\cdot;\, \theta_n, g_n)\}$ with $\{(\theta_n, g_n)\} \in \Theta \times \mathscr{C}(g)$, the process $n^{1/2}(\hat{\mathbb{G}}_n - G_n)$, with $G_n \equiv \int_0^{\cdot} g_n \, d\nu$, converges weakly on $C[0, 1]$ to the **same** limit process $\mathbb{S}$:

$$n^{1/2}(\hat{\mathbb{G}}_n - G_n) \Rightarrow \mathbb{S} \text{ on } C[0, 1]$$

(under f_n) where the law of $\mathbb{S}$ on $C[0, 1]$ does not depend on h or β.

THEOREM 4.1. *Suppose that $\hat{\mathbb{G}}_n$ is a regular estimator of $G = \int_0^{\cdot} g d\nu$ in the model $f = f(\cdot;\, \theta, g)$ with limit process $\mathbb{S}$,* that the conclusion of Proposition 2.1 holds with α given by (2.4), that Assumption S holds with $\beta = \beta_0$ of (4.1), and that Assumption I holds. Then

$$\mathbb{S} =_d \mathbb{Z}_* + \mathbb{W}, \tag{4.5}$$

where the process $\mathbb{Z}_*$ is defined in (4.4) *and the process $\mathbb{W}$ is independent of $\mathbb{Z}_*$.*

To state a local asymptotic minimax bound, we let ℓ: $C[0, 1] \to R^+$ be a subconvex loss function such as $\ell(x) \equiv \| x \| \equiv \sup_t | x(t) |$, $\ell(x) = \int | x(t) |^2 \, dt$, or $\ell(x) = 1\{x : \| x \| \geq c\}$.

THEOREM 4.2. *Suppose that the conclusion of Proposition* 2.1 *holds with α given by* (2.4), *that Assumption S holds with $\mathscr{B} = \mathscr{B}_0$ of* (4.1), *that Assumption* I *holds, and that ℓ is subconvex. Then, with $B_n(c) = \{f_n \in \mathscr{F}(f) : n^{1/2} \| f_n^{1/2} - f^{1/2} \|_\mu \leq c\}$,*

$$\lim_{c \to \infty} \lim_{n \to \infty} \inf_{\hat{\mathbb{G}}_n} \sup_{f_n \in B_n(c)} E_{f_n} \ell(n^{1/2}(\hat{\mathbb{G}}_n - G_n)) \geq E\ell(\mathbb{Z}_*) \tag{4.6}$$

where $\mathbb{Z}_*$ is the zero-mean Gaussian process defined in (4.4).

The infimum over estimates $\hat{\mathbb{G}}_n$ in (4.6) is taken over the class of "generalized procedures," the closure of the class of randomized procedures as in Millar (1979, page 235).

REMARK 4.1. Asymptotic lower bounds for joint estimation of the pair (θ, G) can also be easily formulated: under the hypotheses of Theorem 4.1 or Theorem 4.2, bounds for joint estimates of (θ, G) are determined by the joint distribution of $(Z_*, \mathbb{Z}_*)$ on $R^1 \times C[0, 1]$ (see (4.4) and above); note that

$$\operatorname{Cov}[Z_*, \mathbb{Z}_*(t)] = -\frac{1}{I_*} \int_0^t 2\beta^* g^{1/2} \, d\nu \quad \text{for} \quad 0 \leq t \leq 1. \tag{4.7}$$

REMARK 4.2. When the necessary condition (3.12) for adaptation holds, $\beta^* = 0$ and $\mathbb{Z}_* = \mathbb{Z}$ in (4.4). In this case the coordinates of the pair $(Z_*, \mathbb{Z}_*) = (Z_*, \mathbb{Z})$ are independent.

REMARK 4.3. Explicit calculation of the inverse operator $(A^*A)^{-1}$ may be difficult in many problems. For example, we do not yet know'$(A^*A)^{-1}$ for the class of mixture models given in Example 2; see also Section 6. But once $(A^*A)^{-1}$ is known, everything can be easily calculated, including β^* satisfying (3.1) in view of Remark 3.1.

REMARK 4.4 When the subspace $\mathscr{B}$ of (2.2) is a proper subspace of $\mathscr{B}_0$, $\mathscr{B} \subsetneq \mathscr{B}_0$, a projection operator π, defined as follows, is required: let $\tau : H \to C[0, 1] \equiv B$ be defined by $\tau\alpha(t) \equiv 2 \int_0^t \beta g^{1/2} \, d\nu$ as in the proofs of Theorems 4.1 and 4.2. Let π be a continuous projection of $B = C[0, 1]$ to the closure in B of τH, $\tau\bar{H}$. (If π_0 is a projection of $\mathscr{B}_0$ to $\mathscr{B}$, then π is easily found explicitly in particular cases by composition of the maps π_0, α, and τ.) Then Theorems 4.1 and 4.2 continue to hold with the process $\mathbb{Z}_*$ replaced by the process $\pi\mathbb{Z}_*$ in (4.5) and (4.6). This procedure is closely related to that of Millar's (1979) Proposition 5.2 and Example 6e.

REMARK 4.5. Asymptotic minimax lower bounds and representation theorems for differentiable functionals $\Psi(\theta, G)$ ($\Psi : R^1 \times C[0, 1] \to$ a normed linear space S with norm $\|\cdot\|_S$) are also easily formulated. Suppose that Ψ is differentiable in the following sense: for $\{(\theta_n, g_n)\} \in \Theta \times \mathscr{C}(g)$ let $\Psi_n \equiv \Psi(\theta_n, G_n)$, $\Psi \equiv \Psi(\theta, G)$, and suppose that there exists a continuous linear function $\Psi' : R^1 \times C[0, 1] \to S$ such that $\|\Psi_n - \Psi - \Psi'(\theta_n - \theta, G_n - G)\|_S = o(n^{-1/2})$. Then regular estimates of Ψ have limit laws with $\Psi'(Z_*, \mathbb{Z}_*)$ playing the role of Z_* in (3.5), and $\mathbb{Z}_*$ in (4.5); and asymptotic lower bounds for estimation of $\Psi = \Psi(\theta, g)$ have the quantity $E\ell(\Psi'(Z_*, \mathbb{Z}_*))$ (where ℓ, subconvex on S, is a loss function) appearing on the right side of the bound with $(Z_*, \mathbb{Z}_*)$ as in Remark 4.1. For example, when $\mathscr{X} = R^1$ this approach will easily yield bounds for estimation of $F = F(\cdot;\, \theta, g) = \int_{-\infty}^{\cdot} f\, d\mu$ whenever F can be written as a function of (θ, G).

Another useful approach to bounds for estimation of F is as follows: let $\pi_0 : H_0 \equiv \{\alpha \in L^2(\mu) : \alpha \perp f^{1/2}\} \to H$ be the projection defined by

$$\pi_0\alpha \equiv \frac{4}{I_*} \langle \alpha, \rho_\theta - A\beta^* \rangle_\mu (\rho_\theta - A\beta^*) + A(A^*A)^{-1}A^*\alpha.$$

Hence with $\tau\alpha \equiv \int_{-\infty}^{\cdot} 2\alpha f^{1/2}\, d\mu$, and $\mathscr{B} = \mathscr{B}_0$, easy computations as in the proofs of Theorems 4.1 and 4.2 show that $(\tau\pi_0)^* = \pi_0\tau^*$, and that bounds for estimation of F are determined by the process

$$\mathscr{W}_* \equiv \mathscr{W} + Z_*\tau(\rho_\theta - A\beta^*)$$

where $Z_* \sim N(0, 1/I_*)$ and $\mathscr{W}$ is a mean zero Gaussian process independent of Z_* with covariance function

$$\langle D(1_{(-\infty,s]} - F(s))f^{1/2}, D(1_{(-\infty,t]} - F(t))f^{1/2} \rangle_\mu$$

where $D \equiv A(A^*A)^{-1}A^*$. When $\mathscr{B}$ is a proper subspace of $\mathscr{B}_0$, this approach must be combined with the projection introduced in Remark 4.4.

5. Proofs. To prepare for the proofs of Theorems 3.1 and 3.2, we now define $\tau : H \to R^1 = B$, using notation to agree with that of Millar (1979), by

$$\tau\alpha \equiv \langle \alpha, 4(\rho_\theta - A\beta^*)/I_* \rangle_\mu = h \quad \text{for all} \quad \alpha = h\rho_\theta + A\beta \in H. \tag{5.1}$$

The following Lemma, which gives the adjoint τ^* of τ, is a key step in the proofs to follow.

LEMMA 5.1. *The adjoint $\tau^* : R^1 \to H$ of τ defined by* (5.1) *is given by*

$$\tau^*h^* = 4h^*(\rho_\theta - A\beta^*)/I_* \quad \textit{for all} \quad h^* \in R^1 \tag{5.2}$$

where $\beta^* \in \mathscr{B}$ satisfies (3.1) *and I_* is given by* (3.4).

PROOF. Since $B = R^1$ and H are self-dual, the adjoint τ^* of τ must satisfy

$$\langle \tau\alpha, h^* \rangle_B = \langle \alpha, \tau^*h^* \rangle_\mu \tag{a}$$

for all $h^* \in R^1$ and $\alpha = h\rho_\theta + A\beta \in H$. By definition of τ, the left side in (a) is just hh^*. But, by the definition of τ^* (5.2), the decomposition of α given by (3.2), and the orthogonality relation (3.1), the right side in (a) also equals hh^*, and hence (a) holds. □

REMARK 5.1. The important role of the "derivative" mapping τ and its adjoint τ^* has become clear through the work of Millar (1979, pages 236 and 241) and Levit (1978, page 372). (Levit uses the terminology "conjugate mapping" rather than adjoint.) Note that the particular choices involved in Beran's (1977a, b) proofs of his representation theorems are easily understood in terms of adjoints.

PROOF OF THEOREM 3.1. Let $\hat{\theta}_n$ be a regular estimator of θ with limit law $\mathscr{L}$, and suppose that $\{f_n\} \in \mathscr{F}(f)$. The characteristic function of $n^{1/2}(\hat{\theta}_n - \theta_n)$ under f_n is

$$\text{(a)} \qquad E_{f_n}\exp(iun^{1/2}(\hat{\theta}_n - \theta_n)) = E_{f_n}\exp(iun^{1/2}(\hat{\theta}_n - \theta) - iuh) + o(1)$$

$$\text{(b)} \qquad = E_f\exp(iun^{1/2}(\hat{\theta}_n - \theta) + L_n - iuh) + o(1).$$

This holds for all $\alpha \in H$ and $h \in R^1$. We choose $\alpha = \frac{1}{4}I_*\tau^*h$ where τ^* is given in Lemma 5.1; then $4\|\alpha\|_\mu^2 = h^2 I_*$. Hence, under $f = f(\cdot;\, \theta,\, g)$, the random vectors $(n^{1/2}(\hat{\theta}_n - \theta)$, $2n^{-1/2}\sum_{i=1}^n \alpha(X_i)f^{-1/2}(X_i))$ converge weakly coordinatewise to a random vector (S, hZ) with $Z \sim N(0, I_*)$; and, by considering only a subsequence if necessary, they converge jointly as well. It then follows from Lemma 2.1 that the random vectors $(n^{1/2}(\hat{\theta}_n - \theta), L_n)$ converge weakly under f to $(S, hZ - \frac{1}{2}h^2 I_*)$.

Hence, by regularity of $\hat{\theta}_n$ the characteristic function (a) converges to $E\exp(iuS)$ while (b) converges, by Vitali's theorem and an almost surely convergent construction as in Beran (1977b), to

$$E\exp(iuS + hZ)\exp(-\tfrac{1}{2}h^2 I_* - iuh).$$

Therefore, letting $\phi(u, v) \equiv E\exp(iuS + ivZ)$ denote the characteristic function of (S, Z), we have from (a), (b), and the preceding that

$$\text{(c)} \qquad \phi(u, 0) = E\exp(iuS + hZ)\exp(-iuh - \tfrac{1}{2}h^2 I_*)$$

for all real h. The right hand side of (c) is analytic in h, constant for all real h, hence constant for all complex h. The choice $h = -iu/I_*$ yields

$$\text{(d)} \qquad \phi(u, 0) = \phi(u, -u/I_*)\exp(-u^2/2I_*)$$

for all real u, and this factorization into the characteristic functions of $W \equiv S - Z_*$ and $Z_* \equiv Z/I_*$ completes the proof. □

Proof of Theorem 3.2. The mapping τ defined in (5.1) is linear, bounded, and has dense range in $R^1 = B$; but it is not one-to-one and hence does not yet satisfy the requirements for τ of Section 3 of Millar (1979). However the restriction of τ to the (one-dimensional) subspace $H^* \equiv \{2h\alpha^* : h \in R^1\}$, where $\alpha^* \equiv \rho_\theta - A\beta^*$ satisfies (3.1), is one-to-one. (Take H in Millar's set-up to be twice our H.) Moreover, $\|\tau^*h^*\|_\mu^2 = 4h^{*2}/I_*$, so the image law P_0 of the unit normal on H^* is simply $N(0, 1/I_*)$. But, with $B_n^*(c) \equiv \{f_n \in \mathscr{F}(f, \alpha^*) : n^{1/2}\|f_n^{1/2} - f^{1/2}\|_\mu \le c\}$, the left side of (3.7) is no smaller than

$$\lim_{c\to\infty}\lim_{n\to\infty}\inf_{\hat{\theta}}\sup_{f_n\in B_n^*(c)} E_{f_n}\ell(n^{1/2}(\hat{\theta}_n - \theta_n)) \ge E\ell(Z_*)$$

where the inequality follows from Propositions 2.1 and 3.1 of Millar (1979). □

Now define $\tau : H \to B_0 \equiv \{x \in C[0, 1] : x(0) = x(1) = 0\}$ by

$$\tau\alpha(t) = \int_0^t 2\beta g^{1/2}\, dv = \langle \beta, 2g^{1/2}1_{[0,t]}\rangle_\nu \quad \text{for} \quad 0 \le t \le 1. \tag{5.3}$$

It is not hard to exhibit τ explicitly as a function of $\alpha \in$ H: since $A^*(\rho_\theta - A\beta^*) = 0$, it follows from (3.2) that $A^*\alpha = A^*A(h\beta^* + \beta)$, and hence, under Assumption I,

$$\beta = (A^*A)^{-1}A^*\alpha - h\beta^* = C^*\alpha - \langle \alpha, 4(\rho_\theta - A\beta^*)/I_*\rangle_\mu \beta^*. \tag{5.4}$$

Substitution of this expression for β into (5.3) yields

$$\begin{aligned} \tau\alpha(t) &= \langle C^*\alpha, 2g^{1/2}1_{[0,t]}\rangle_\nu - \langle \alpha, \frac{4}{I_*}(\rho_\theta - A\beta^*)\rangle_\mu \langle \beta^*, 2g^{1/2}1_{[0,t]}\rangle_\nu \\ &= \tau_0(C^*\alpha)(t) - \langle \alpha, \frac{4}{I_*}(\rho_\theta - A\beta^*)\rangle_\mu (\tau_0\beta^*)(t) \end{aligned} \tag{5.5}$$

where $\tau_0 : \mathscr{B} \to \mathscr{B}_0 \subset C[0, 1]$ is defined by

$$\tau_0\beta(t) = \langle \beta, 2g^{1/2}1_{[0,t]}\rangle_\nu, \qquad 0 \le t \le 1.$$

The crux of the proof of Theorems 4.1 and 4.2 is contained in the following lemma which gives the adjoint τ^* of the mapping τ.

LEMMA 5.2. *The adjoint* $\tau^* : BV = B_0^* \to H = H^*$ *of* τ *defined by* (5.3) *is given by*

$$\tau^* v = C\tau_0^* v - \frac{4}{I_*}(\rho_\theta - A\beta^*)\langle \tau_0\beta^*, v\rangle \quad \textit{for} \quad v \in BV \tag{5.6}$$

where β^* *satisfies* (3.1),

$$\tau_0^* v = -2\left(v - \int vg\, dv\right)g^{1/2} \perp g^{1/2}, \tag{5.7}$$

and

$$\langle \tau_0\beta^*, v\rangle \equiv \int_0^1 \tau_0\beta^*\, dv.$$

PROOF. With $\langle \tau\alpha, v\rangle \equiv \int_0^1 \tau\alpha\, dv$, we have, using (5.5) and the definition of the adjoints τ_0^* and C^* of τ_0 and C,

$$\begin{aligned}\langle \tau\alpha, v\rangle &= \langle \tau_0 C^*\alpha, v\rangle - \langle \alpha, 4(\rho_\theta - A\beta^*)/I_*\rangle_\mu\langle \tau_0\beta^*, v\rangle \\ &= \langle C^*\alpha, \tau_0^* v\rangle_\nu - \langle \alpha, 4(\rho_\theta - A\beta^*)/I_*\rangle_\mu\langle \tau_0\beta^*, v\rangle \\ &= \langle \alpha, C\tau_0^* v\rangle_\mu - \langle \alpha, 4(\rho_\theta - A\beta^*)/I_*\rangle_\mu\langle \tau_0\beta^*, v\rangle \\ &= \langle \alpha, C\tau_0^* v - \frac{4}{I_*}(\rho_\theta - A\beta^*)\langle \tau_0\beta^*, v\rangle\rangle_\mu \\ &= \langle \alpha, \tau^* v\rangle_\mu,\end{aligned}$$

where $\tau^* v$ is given in (5.6). □

REMARK 5.2. Note that by (3.1), (3.4), Remark 3.1, and the properties of adjoins, it follows that

$$\begin{aligned}\langle \tau^* v, \rho_\theta\rangle_\mu &= \langle C\tau_0^* v, \rho_\theta\rangle_\mu - \langle \tau_0\beta^*, v\rangle = \langle \tau_0^* v, C^*\rho_\theta\rangle_\nu - \langle \tau_0\beta^*, v\rangle \\ &= \langle \tau_0^* v, \beta^*\rangle_\nu - \langle \tau_0\beta^*, v\rangle = \langle \tau_0\beta^*, v\rangle - \langle \tau_0\beta^*, v\rangle = 0.\end{aligned} \tag{5.8}$$

This is the analogue of (3.1) in the present problem of estimating G: the "effective score" for G, $\tau^* v$, is orthogonal to the "nuisance parameter score" ρ_θ.

PROOF OF THEOREMS 4.1 AND 4.2. These proofs proceed along the lines of the proofs given by Beran (1977a) and Millar (1979), so we will omit most of the details. The main chore remaining to complete the proof is the computation of $\|\tau^* v\|_\mu^2$ where τ^* is given in (5.6): From (3.1) it follows that

$$\rho_\theta - A\beta^* \perp C\tau_0^* v = A(A^*A)^{-1}\tau_0^* v,$$

and hence,

$$\|\tau^* v\|_\mu^2 = \|C\tau_0^* v\|_\mu^2 + \frac{4}{I_*}\langle \tau_0\beta^*, v\rangle^2. \tag{a}$$

But, by definition of τ_0 and the properties of adjoints,

$$\|C\tau_0^* v\|_\mu^2 = \langle C\tau_0^* v, C\tau_0^* v\rangle_\mu = \langle C^*C\tau_0^* v, \tau_0^* v\rangle_\nu = \langle \tau_0 (A^*A)^{-1}\tau_0^* v, v\rangle$$

$$= \int \langle (A^*A)^{-1}\tau_0^* v, 2g^{1/2} 1_{[0,t]}\rangle_\nu \, d\nu(t)$$

$$= \int \langle (A^*A)^{-1}\tau_0^* v, 2(1_{[0,t]} - G(t))g^{1/2}\rangle_\nu \, d\nu(t),$$

since $(A^*A)^{-1}\tau_0^* v \perp g^{1/2}$,

$$= \int \langle \tau_0^* 1_{[0,t]}, (A^*A)^{-1}\tau_0^* v\rangle_\nu \, d\nu(t) \quad \text{by (5.7)}$$

$$= \int \langle \tau_0 (A^*A)^{-1}\tau_0^* 1_{[0,t]}, v\rangle \, d\nu(t),$$

using self-adjointness of $(A^*A)^{-1}$,

$$= \int\int \langle (A^*A)^{-1}\tau_0^* 1_{[0,t]}, 2g^{1/2}{}_{[0,s]}\rangle_\nu \, d\nu(s)\, d\nu(t)$$

$$= \int\int \langle \tau_0^* 1_{[0,s]}, (A^*A)^{-1}\tau_0^* 1_{[0,t]}\rangle_\nu \, d\nu(s)\, d\nu(t)$$

$$= \int\int \langle C\tau_0^* 1_{[0,s]}, \mathrm{C}\tau_0^* 1_{[0,t]}\rangle_\mu \, d\nu(s)\, d\nu(t)$$

$$= 4\int\int K(s,t)\, d\nu(s)\, d\nu(t), \tag{b}$$

where K is as defined in (4.3). Combining (a) and (b) yields

$$\frac{1}{4}\|\tau^* v\|_\mu^2 = \int\int K(s,t)\, d\nu(s)\, d\nu(t) + \frac{1}{I_*}\langle \tau_0\beta^*, v\rangle^2 = E\left(\int_0^1 \mathbb{Z}_*\, d\nu\right)^2 \tag{c}$$

where $\mathbb{Z}_*$ is as defined in (4.4). □

REMARK 5.3. V. Fabian and J. Hannan have pointed out to us that the local asymptotic minimax theorem Proposition 2.1 of Millar (1979) is incorrect as stated. The difficulty is that a stronger definition of "convergence of experiments" than the one given by Millar (1979, page 235) is required. A corrected version of the theorem with additional detail and a variety of related material is given by Millar (1981b); or see Le Cam (1979). In our present context, with the differentiability condition of our Proposition 2.1 and Lemma 2.1 yielding nice expansions of local likelihood ratios, the stronger convergence of experiments necessary for (the corrected version of) Millar's Proposition 2.1 is in force, and the conclusion of the asymptotic minimax theorem holds.

6. Examples, continued.

Now we return to the examples introduced in Section 1, and show how they can be treated in the framework of Sections 2, 3, 4.

EXAMPLE 1*a*. *One-sample location model with symmetry.* If $f(x; \theta, g) = g(x - \theta)$ for symmetric $g(g \in \mathscr{G}_s)$ having finite Fisher information I_g, then (2.3) and (2.4) hold with $\rho_\theta(x) = \rho(x - \theta)$ where $\rho \equiv -\frac{1}{2}\dot{g}g^{-1/2} \in L^2(\mu)$; $(A\beta)(x) = \beta(x - \theta)$; and $\mathscr{B} = \{\beta \in L^2(\mu): \beta \perp g^{1/2}, \beta \text{ symmetric about } 0\}$ since $g_n, g \in \mathscr{G}_s$ and $\|n^{1/2}(g_n^{1/2} - g^{1/2}) - \beta\|_\mu \to 0$ imply that β is symmetric. Thus Assumption S holds, and

$$\langle \rho_\theta, A\beta\rangle_\mu = \langle \rho, \beta\rangle_\mu = 0 \quad \text{for all } \beta \in \mathscr{B} \tag{6.1}$$

since $\rho = -\frac{1}{2}\dot{g}g^{-1/2}$ is odd (by symmetry of g) and $\beta \in \mathscr{B}$ is even (symmetric). Hence the necessary condition (3.12) for adaptation is satisfied, and our Theorems 3.1 and 3.2 show that the best possible asymptotic variance for estimates of θ is $1/I_0 = 1/I_g$ where $I_0 = I_g = 4\|\rho\|_\mu^2$, the usual parametric information for θ as if g were known. The adaptive estimators of Stone (1975), Beran (1978), and others achieve this minimum variance. As noted previously, Beran's (1978) estimators even achieve our bounds "locally uniformly" in g.

To describe the lower bounds for estimation of the df G corresponding to $g \in \mathscr{G}_s$, first note that $A^*\beta = \beta(\cdot + \theta)$, so that $A^*A = I =$ the identity, and hence Assumption I holds with $(A^*A)^{-1} = I$. After an application of the probability integral transformation we can assume that $g = 1_{[0,1]}$, and that $\beta \in \mathscr{B} = \{\beta \in L^2[0, 1] : \beta \perp 1$ and β symmetric about $\frac{1}{2}\}$. Thus $\pi_0 : \mathscr{B}_0 \to \mathscr{B} \subset \mathscr{B}_0$ may be defined by $\pi_0\beta(t) = \frac{1}{2}(\beta(t) + \beta(1-t))$, $0 \le t \le 1$, and easy computation then shows that $\tau\pi_0\beta(t) = \pi(\tau\beta)(t)$ where $\pi x(t) = \frac{1}{2}(x(t) - x(1-t))$. Straightforward computation shows that $K(s, t) = s \wedge t - st$, so the process $\mathbb{Z}$ is Brownian bridge on $[0, 1]$, and $\pi\mathbb{Z}$ is the symmetrized Brownian bridge $\mathbb{Z}_s(t) \equiv \frac{1}{2}(\mathbb{Z}(t) - \mathbb{Z}(1-t))$. In view of Remark 4.2 and the preceding, $\pi\mathbb{Z}_* = \pi\mathbb{Z} \equiv \mathbb{Z}_s$ is independent of Z_*.

This example has been treated in considerable detail in a Berkeley thesis by Shaw-Hwa Lo (1982); Lo (1982) also gives lower bounds for estimation of $F = \int_{-\infty}^{\cdot} f\,d\mu$ (described by $\mathbb{Z}_s(F) - Z_* f$ with Z_* and $\mathbb{Z}_s$ independent as above) and, following Stone (1975), constructs estimators which achieve the bounds asymptotically.

EXAMPLE 1b. *One sample location model without symmetry.* Let $f(x;\, \theta, g) = g(x - \theta)$ for T-centered $g(g \in \mathscr{G}_T)$ having finite Fisher information I_g; here T is a fixed Hellinger-differentiable location functional (such as the median) defined for all $g \in \mathscr{G}_s$ with derivative ρ_T; see Beran (1977b). Since $T(-X) = -T(X)$ for a location functional, $\mathscr{G}_s \subset \mathscr{G}_T$.

Now (2.3) and (2.4) hold with $\rho_\theta = \rho(\cdot - \theta)$, $\rho = -\frac{1}{2}\dot{g}g^{-1/2}$, and $A\beta = \beta(\cdot - \theta)$ as in Example 1a, but $\beta = \{\beta \in L^2(\mu) : \beta \perp g^{1/2},\ \beta \perp \rho_T\}$. Again Assumption S holds, but the condition for adaptation (3.12) fails; the subspace $\{A\beta = \beta(\cdot - \theta) : \beta \in \mathscr{B}\}$ is "too big." Let $\beta^* \equiv (\rho - \rho_T\|\rho_T\|_\mu^{-2})$; then $\beta^* \perp g^{1/2}$ and $\beta^* \perp \rho_T$ since $\langle \beta^*, \rho_T\rangle_\mu = \langle \rho, \rho_T\rangle_\mu - 1 = 0$ by Theorem 2(ii) of Beran (1977b). Hence $\beta^* \in \mathscr{B}$. Furthermore, $\rho - \beta^* = \rho_T/\|\rho_T\|_\mu^2 \perp \beta$ for all $\beta \in \mathscr{B}$, so β^* satisfies (3.1). Thus $I_* = 4\|\rho - \beta^*\|_\mu^2 = 4/\|\rho_T\|_\mu^2$, and the best possible asymptotic variance for estimating $\theta = T(f)$ is $\|\rho_T\|_\mu^2/4$. Thus our Theorem 3.1 gives precisely the same result as Theorem 6 of Beran (1977b) in this case, but our method allows for comparison with Example 1a where adaptation is possible. In the present example, "analogue estimates" achieve the lower bounds: e.g. if $T(g) =$ median of g, then $\rho_T(x) = g(0)^{-1}\operatorname{sign}(x)g^{1/2}(x)$ so $\|\rho_T\|_\mu^2/4 = 1/\{4g^2(0)\}$, and this asymptotic variance is achieved by the sample median.

To compute the bounds for estimation of G, note that $A^*A = I$ as in Example 1a, so that Assumption I holds. By introducing the projection map $\pi_0 : \mathscr{B}_0 \to \mathscr{B}$ defined by $\pi_0\beta = \beta - \langle \beta, \rho_T\rangle\rho_T/\|\rho_T\|^2$ and noting that $(\tau_0\pi_0)^* = \pi_0\tau_0^*$, it is easily found that the process $\mathbb{Z}$ has covariance function

$$K(s, t) = \langle \pi_0\tau_0^* 1_{(-\infty,s]},\ \pi_0\tau_0^* 1_{(-\infty,t]}\rangle_\mu .$$

Letting B^0 denote a Brownian bridge process on $[0, 1]$, it is easily verified that K is the covariance function of

$$\mathbb{Z} \equiv B^0(G) - Z_* \int_{-\infty}^{\cdot} 2(\rho_T/\|\rho_T\|^2)g^{1/2}\,d\mu$$

where $Z_* \equiv \int \frac{1}{2}\rho_T\,dB^0(G) \sim N(0, \|\rho_T\|^2/4)$; note that Z_* and $\mathbb{Z}$ are independent. Hence, recalling the definition of Z_* and that $\beta^* = \rho - \rho_T/\|\rho_T\|^2$ in the present case,

$$\mathbb{Z}_* = B^0(G) - Z_* \int_{-\infty}^{\cdot} 2\rho g^{1/2}\,d\mu = B^0(G) + Z_* g$$

where $\operatorname{Cov}\{B^0(G(t)), Z_*\} = \frac{1}{2} \int_{-\infty}^{t} \rho_T g^{1/2}\, d\mu$. Under mild regularity conditions this bound is achieved asymptotically by the obvious estimator $\mathbb{F}_n(\cdot + \hat{\theta}_n)$ where $\mathbb{F}_n$ denotes the empirical df of X's iid F and $\hat{\theta}_n = T(\mathbb{F}_n)$ is the natural analogue estimator of $\theta = T(f)$.

Related work on a robust version of this example, with emphasis on construction of asymptotically optimal estimates, is contained in a Berkeley thesis by Neng Hsin Chen (1980).

EXAMPLE 2. *Mixture models.* Let $M = M(\cdot\,; \theta, \phi)$ be a family of density functions (with respect to the dominating measure μ on R^1) indexed by $(\theta, \phi) \in R^2$, and let $f(x; \theta, g) = \int M(x; \theta, \phi) g(\phi)\, d\phi$ where $g \in \mathcal{G}$, a class of densities small enough that θ is identifiable. Suppose that $M_\theta \equiv (\partial/\partial\theta) M(\cdot\,; \theta, \phi)$ exists. Then, under some additional regularity conditions on M and its derivatives, it seems reasonable to expect that (2.2) and (2.3) will hold with

$$\rho_\theta(x) = \int M_\theta(x; \theta, \phi) g(\phi)\, d\phi / 2f^{1/2}(x)$$

and

$$A\beta(x) = \int M(x; \theta, \phi) \beta(\phi) g^{1/2}(\phi)\, d\phi / f^{1/2}(x);$$

precise conditions under which this is true would be interesting and important. For this model we do not yet know β^*, $(A^*A)^{-1}$, or I_*. The identifiability issue is important here since smaller $\mathcal{G}$'s will result in smaller subspaces $\mathcal{B}$. For many M's, including scale mixtures of normal densities, or mixtures of paired exponentials with common hazard ratio, $\mathcal{G}$ may be taken to be all densities on $R^+ = [0, \infty)$; see Teicher (1961) and Lindsay (1980). Note that this class of models is of interest even when the parameter θ is known or vacuous and attention focuses entirely on estimation of G. See Laird (1978) and Jewell (1982) for some recent work on computational and consistency aspects (but not asymptotic efficiency) of generalized maximum likelihood estimators of G in this setting. This class of models deserves further study.

EXAMPLE 3. *Two-sample shift model.* Our treatment of this model fits it into a one-sample mold by introducing a (random) indicator variable and letting the two sample sizes be binomial rv's as in Bickel (1982). An alternative treatment would simply involve an appropriate two-sample generalization of Proposition 2.1 and Lemma 2.1. Suppose that $(X_1, Z_1), \cdots (X_n, Z_n)$ are iid with density function $f(x, z) = f(x, z; \theta, \eta, g)$ with respect to the product μ of Lebesgue measure ν and counting measure on $R^1 \times \{0, 1\}$ given by

$$f(x, z) = \begin{cases} f(x, 0) = \lambda g(x - \eta), \\ f(x, 1) = \bar{\lambda} g(x - \eta - \theta), \end{cases}$$

where $0 < \lambda < 1$, $\bar{\lambda} = 1 - \lambda$, $\theta \in R^1$, $\eta \in R^1$, and $g \in \mathcal{G}_T$ as defined in Example 1b. In this model η is a one-dimensional or parametric nuisance parameter (which could have been absorbed into g, but separating it out gives a more realistic two-sample problem when g is known), $g \in \mathcal{G}_T$ is an infinite dimensional nuisance parameter or nonparametric component of the model, and θ, the "shift parameter", is the parameter of primary interest. Thus our calculations will follow the outline given in Remarks 2.1 and 3.2: For this model it is easily seen that (2.3) and (2.4)′ hold (assuming $I_g < \infty$) with

$$\rho_\theta(x, z) = \begin{cases} \rho_\theta(x, 0) = 0, \\ \rho_\theta(x, 1) = \bar{\lambda}^{1/2} \rho(x - \eta - \theta), \end{cases}$$

where $\rho \equiv -\frac{1}{2} \dot{g}\, g^{-1/2}$ as before;

$$\rho_\eta(x, z) = \begin{cases} \rho_\eta(x, 0) = \lambda^{1/2} \rho(x - \eta) \\ \rho_\eta(x, 1) = \bar{\lambda}^{1/2} \rho(x - \eta - \theta); \end{cases}$$

and

$$(A\beta)(x, z) = \begin{cases} (A\beta)(x, 0) = \lambda^{1/2}\beta(x - \eta) \\ (A\beta)(x, 1) = \bar{\lambda}^{1/2}\beta(x - \eta - \theta). \end{cases}$$

Also $\mathscr{B} = \{\beta \in L^2(\mu): \beta \perp g^{1/2}, \beta \perp \rho_T\}$ where $\rho_T \in L^2(\mu)$ is the derivative of the Hellinger-differentiable functional T, so Assumption S holds.

By straightforward calculation $B = \langle \rho_\theta, \rho_\eta \rangle_\mu = \bar{\lambda} \| \rho \|_\nu^2$, $D = \| \rho_\eta \| \|_\mu^2 = \| \rho \|_\nu^2$, so $BD^{-1} = \bar{\lambda}$ and $\rho_{\theta \cdot \eta} = \rho_\theta - BD^{-1}\rho_\eta$ is given by

$$\rho_{\theta\cdot\eta}(x, z) = \begin{cases} \rho_{\theta\cdot\eta}(x, 0) = -\lambda^{1/2}\bar{\lambda}\, \rho(x - \eta) \\ \rho_{\theta\cdot\eta}(x, 1) = \lambda\bar{\lambda}^{1/2}\rho(x - \eta - \theta). \end{cases}$$

Thus $\langle \rho_{\theta\cdot\eta}, A\beta \rangle_\mu = -\lambda\bar{\lambda}\ \langle \rho, \beta \rangle_\nu + \lambda\bar{\lambda}\ \langle \rho, \beta \rangle_\nu = 0$ *for all* $\beta \in \mathscr{B}$, *the* condition (3.12)′ for adaptation to $g \in \mathscr{G}_T$ is satisfied, and the best possible asymptotic variance for estimators of θ is given by $1/I_0 = 1/(\lambda\bar{\lambda}\, I_g)$ since $I_0 = 4 \| \rho_{\theta\cdot\eta} \|_\mu = 4\, \lambda\bar{\lambda} \,| \rho \|_\nu^2 = \lambda\bar{\lambda}\, I_g$ where $I_g = 4 \int \rho^2 d\nu = \int (\dot{g}^2/g)d\nu$. The adaptive estimators of Beran (1974) attain this bound.

Easy computations using Remark 3.2 also show that the effective score for η is $(\rho_{\eta\cdot\theta} - A\beta^*)(x, z) = \lambda^{1/2}(\bar{\lambda}\rho + \lambda\rho_T / \| \rho_T \|_\nu^2)(x - \eta)$, $z = 0$, $= -\bar{\lambda}^{1/2}\lambda(\rho - \rho_T / \| \rho_T \|_\nu^2)(x - \eta - \theta)$, $z = 1$, with $\beta^* = \lambda(\rho - \rho_T / \| \rho_T \|_\nu^2)$. Thus the information for estimation of η is $4 \| \rho_{\eta\cdot\theta} - A\beta^* \|_\mu^2 = \lambda\bar{\lambda} I_g + 4\lambda^2 / \| \rho_T \|_\nu^2$.

To find the bounds for estimation of $G = \int_{-\infty}^{\cdot} g\, d\nu$ corresponding to $g \in \mathscr{G}_T$, we first compute $A^*\alpha = \lambda^{1/2}\alpha(\cdot\ + \eta, 0) + \bar{\lambda}^{1/2}\alpha(\cdot\ + \eta + \theta, 1)$ for $\alpha \in L^2(\mu)$, so that $A^*A = I$ and Assumption I holds. Since $\beta \perp \rho_T$ for all $\beta \in \mathscr{B}$, the same projection π_0 introduced in Example lb is required, and the process $\mathbb{Z}$ is exactly the same as was given there: $\mathbb{Z} = B^0(G) - Z \int_{-\infty}^{\cdot} 2(\rho_T / \| \rho_T \|^2)g^{1/2}d\nu$ with $Z \equiv \int \tfrac{1}{2}\, \rho_T dB^0(G)$. Since $\beta^* = 0$, $\mathbb{Z}_* = \mathbb{Z}$ is independent of $Z_* \sim N(0, 1/I_*)$ with $I_* = I_0 = \lambda\bar{\lambda} I_g$; note that $Z \neq Z_*$ in the present example, however.

EXAMPLE 4. *Cox's regression model.* To illustrate our methods, we follow Tsiatis's (1981) formulation of Cox's model and treat the covariate Z as a random variable. Suppose that the covariates $Z_1, \cdots, Z_n$ are iid with density $h(z)$, the survival time X_i^0 and the censoring time Y_ι are conditionally independent given Z_i, and the triples (X_i^0, Y_i, Z_i) are mutually independent for $i = 1, \cdots, n$. Further suppose that, given $Z = z$, X^0 has density $g(\cdot \,|\, z)$ determined by the hazard function $\lambda(\cdot \,|\, z)$ given in (1.1), and Y has density $c(\cdot \,|\, z)$.

In this censored survival data problem, we observe the random vectors $\mathbf{X}_i = (T_\iota, \Delta_i, Z_i)$, $i = 1, \cdots, n$, where $T_i = \min\{X_i^0, Y_\iota\}$ and $\Delta_i = 1$ or 0 according as $T_\iota = X_i^0$ or not. Thus $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are iid with density function $f(\mathbf{x}) = f(\mathbf{x}; \theta, g)$ with respect to the product measure $\mu = \nu \times$ (counting measure) $\times\ \nu$ on $\mathscr{X} = R^+ \times \{0, 1\} \times R$ given by

$$f(\mathbf{x}) = \{g(t \,|\, z)\bar{C}(t \,|\, z)h(z)\}^\Delta \{c(t \,|\, z)\bar{G}(t \,|\, z)h(z)\}^{1-\Delta}, \quad \mathbf{x} = (t, \Delta, z) \in \mathscr{X}.$$

This model is slightly unstable for certain values of θ and densities g. For example, when Z has a Bernoulli distribution, g has bounded support, $\bar{C}(t \,|\, z) \equiv 1$, and $g_n \in \mathscr{G}^+$ satisfy $\| n^{1/2}(g_n^{1/2} - g^{1/2}) - \beta \|_\nu \to 0$ where β has support outside that of g, then (2.3) fails. We restrict attention to g_n's which are absolutely continuous with respect to g so that the support of the resulting β's is contained in that of g, and we assume that $E\{Z^2\exp(\theta Z)\}$ is bounded uniformly in a neighborhood of θ as in Tsiatis (1981, page 95). Then (2.3) holds with

$$\begin{aligned} \rho_\theta(t, \Delta, z) &= \frac{1}{2} z \left\{ \Delta \cdot 1 - \exp(\theta z) \int_0^t 1 \frac{dG}{1 - G} \right\} f^{1/2}(t, \Delta, z) \\ &= \frac{1}{2} z \{\Delta + \log \bar{G}(t \,|\, z)\} f^{1/2}(t, \Delta, z) \end{aligned} \tag{6.2}$$

and

$$A\beta(t, \Delta, z) = \left\{\Delta R\beta(t) - \exp(\theta z) \int_0^t R\beta \frac{dG}{1-G}\right\} f^{1/2}(t, \Delta, z)$$

$$(6.3) \qquad = \left\{\Delta R\beta(t) + \exp(\theta z) \int_t^\infty \beta g^{1/2} d\nu / \bar{G}(t)\right\} f^{1/2}(t, \Delta, z),$$

where $R\beta(t) \equiv \beta(t) g^{-1/2}(t) - \int_t^\infty \beta g^{1/2} d\nu / \bar{G}(t)$, and $\mathscr{B} = \{\beta \in L^2(\nu)\colon \beta \perp g^{1/2}, \text{support}(\beta) \subset \text{support}(g)\}$, so that Assumption S holds. For verification of (2.3) in the two-sample case of Cox's model without censoring, see Lemma 1 of Begun and Wellner (1982); the details for the general case are similar to those of the two-sample case.

It is not hard to show in the present case that the condition (3.12) for adaptation fails, and hence we need to find the projection of ρ_θ onto $\{A\beta\colon \beta \in \mathscr{B}\}$. Standard calculations (as in Luenberger, 1969, pages 150–153), yield

$$A^*A\beta(t) = \left\{R\beta(t) \frac{M_0(t)}{1 - G(t)} - \int_0^t R\beta \frac{M_0}{1-G} \frac{dG}{1-G}\right\} g^{1/2}(t)$$

where $M_j(t) \equiv E\{Z^j \exp(\theta Z) 1_{(T>t)}\}$ for $t \in R^+$ and $j = 0, 1, 2$. It is straightforward to verify that

$$(6.4) \qquad (A^*A)^{-1}\beta(t) = \left\{R\beta(t) \frac{1 - G(t)}{M_0(t)} - \int_0^t R\beta \frac{1-G}{M_0} \frac{dG}{1-G}\right\} g^{1/2}(t).$$

Hence we find that

$$(6.5) \qquad \beta^*(t) = (A^*A)^{-1} A^* \rho_\theta(t) = \frac{1}{2}\left\{\frac{M_1(t)}{M_0(t)} - \int_0^t \frac{M_1}{M_0} \frac{dG}{1-G}\right\} g^{1/2}(t).$$

With $\mathbf{x} = (t, \Delta, z) \in \mathscr{X}$, this yields

$$A\beta^*(\mathbf{x}) = \frac{1}{2}\left\{\Delta \frac{M_1(t)}{M_0(t)} - \exp(\theta z) \int_0^t \frac{M_1}{M_0} \frac{dG}{1-G}\right\} f^{1/2}(\mathbf{x}),$$

and therefore, by (6.2),

$$(\rho_\theta - A\beta^*)(\mathbf{x}) = \frac{1}{2}\left\{\Delta\left(z - \frac{M_1(t)}{M_0(t)}\right) - \exp(\theta z) \int_0^t \left(z - \frac{M_1}{M_0}\right) \frac{dG}{1-G}\right\} f^{1/2}(\mathbf{x}).$$

Then, by straightforward computation,

$$I_* = 4\|\rho_\theta - A\beta^*\|_\mu^2 = \int_{-\infty}^\infty \int_0^\infty \left(z - \frac{M_1(t)}{M_0(t)}\right)^2$$

$$\cdot \exp(\theta z) \bar{G}(t \mid z) \bar{C}(t \mid z) \frac{dG(t)}{1 - G(t)} h(z)\, d\nu(z)$$

$$(6.6) \qquad = \int_0^\infty \left\{\frac{M_2(t)}{M_0(t)} - \frac{M_1(t)^2}{M_0(t)^2}\right\} M_0(t) \frac{dG(t)}{1 - G(t)}$$

$$= \int_0^\infty \left\{\frac{M_2(t)}{M_0(t)} - \frac{M_1(t)^2}{M_0(t)^2}\right\} dF(t, 1)$$

$$\le \int_0^\infty \frac{M_2(t)}{M_0(t)} dF(t, 1) = 4\|\rho_\theta\|_\mu^2 = I_0,$$

with $<$ if $\theta \neq 0$, where the third line follows from Fubini's theorem and the definition of $M_j(t)$, and where the fourth line follows from

$$
\begin{aligned}
M_0(t)g(t)/\bar{G}(t) &= \int_{-\infty}^{\infty} e^{\theta z}\bar{G}(t\,|\,z)\bar{C}(t\,|\,z)h(z)d\nu(z)g(t)/\bar{G}(t) \\
&= \int_{-\infty}^{\infty} g(t\,|\,z)\bar{C}(t\,|\,z)h(z)d\nu(z) \quad \text{by (1.1)} \\
&= \int_{-\infty}^{\infty} f(t, 1, z)d\nu(z) = F'(t, 1)
\end{aligned} \tag{6.7}
$$

where $F(t, 1) = P(T \leq t, \Delta = 1)$ is the subdistribution function of an uncensored observation. The smallest possible asymptotic variance $1/I_*$ for this model is attained by the Cox (1972, 1975) partial likelihood estimator; see also Efron (1977), Chapter 4 of Kalbfleisch and Prentice (1980), and Tsiatis (1981).

To describe the bounds for estimates of G (on an interval $[0, T_0]$ with $P(T > T_0) > 0$), we first compute the covariance function K of the process $\mathbb{Z}$ in (4.4). By direct calculation, $R(1_{[0,t]} - G(t))g^{1/2} = 1_{[0,t]}\bar{G}(t)/\bar{G}$, so that, by (6.4),

$$
(A^*A)^{-1}\{1_{[0,t]} - G(t)\}g^{1/2} = \bar{G}(t)\left\{\frac{1_{[0,t]}}{M_0} - \int_0^{\cdot} \frac{1_{[0,t]}}{M_0}\frac{dG}{1-G}\right\}g^{1/2}
$$

and hence that

$$
\begin{aligned}
K(s, t) &= \langle (1_{[0,s]} - G(s))g^{1/2}, (A^*A)^{-1}(1_{[0,t]} - G(t))g^{1/2}\rangle_\nu \\
&= \bar{G}(t)\left(\int_0^{\infty} 1_{[0,s]}1_{[0,t]}\frac{dG}{M_0} - \int_0^{\infty} 1_{[0,s]}\left\{\int_0^{\cdot} \frac{1_{[0,t]}}{M_0}\frac{dG}{1-G}\right\}dG\right) \\
&= \bar{G}(t)\left(\int_0^{s\wedge t} \frac{dG}{M_0} - \int_0^{s\wedge t} (\bar{G} - \bar{G}(s))\frac{1}{M_0}\frac{dG}{1-G}\right) \\
&= \bar{G}(s)\bar{G}(t)\int_0^{s\wedge t} \frac{1}{M_0(u)^2}\,dF(u, 1)
\end{aligned} \tag{6.8}
$$

by (6.7). Next, using (6.5), Fubini's theorem, and (6.7), we find that for $0 \leq t \leq T_0$

$$
\int_0^t 2\beta^* g^{1/2}d\nu = -\bar{G}(t)\int_0^t (M_1/M_0^2)dF(\cdot, 1).
$$

Thus the process $\mathbb{Z}_*$ of (4.4) becomes, in this case,

$$
\mathbb{Z}_* = \mathbb{Z} + Z_*\bar{G}\int_0^{\cdot} (M_1/M_0^2)dF(\cdot, 1)
$$

where $\mathbb{Z}$ is a mean zero Gaussian process on $[0, T_0]$ with covariance K given by (6.8) independent of $Z_* \sim N(0, 1/I_*)$ with I_* given by (6.6); this is precisely the limit process of the estimator of G derived by Breslow (1974) as the nonparametric maximum likelihood estimator of G under $\hat{\theta} =$ the Cox partial likelihood estimator; see Efron (1977) and Tsiatis (1981) Theorem 5.1 and Lemma 6.2.

To illustrate Remark 4.5, consider estimation of the survival function for an individual with covariate z_0: i.e. $\bar{G}(t\,|\,z_0) = \bar{G}(t)^{\exp(\theta z_0)} \equiv \Psi(\theta, G)(t)$ for $0 \leq t \leq T_0$. It is easily shown that $\Psi'(h, \Delta)(t) = \exp(\theta z_0)\bar{G}(t \mid z_0)\{z_0 h \log \bar{G}(t) - \Delta(t)/\bar{G}(t)\}$, so that

$$\begin{aligned}\Psi'(Z_*, \mathbb{Z}_*)(t) &= -\exp(\theta z_0)\bar{G}(t \mid z_0)\{\mathbb{Z}_*(t)/\bar{G}(t) - z_0 Z_* \log \bar{G}(t)\} \\ &= -\exp(\theta z_0)\bar{G}(t \mid z_0)\Bigg\{\mathbb{Z}(t)/\bar{G}(t) \\ &\qquad + Z_*\left(\int_0^t (M_1/M_0^2)\,dF(\cdot, 1) - z_0 \log \bar{G}(t)\right)\Bigg\},\end{aligned}$$

a mean zero Gaussian process with covariance function

$$\begin{aligned}K_{z_0}(s, t) = e^{2\theta z_0}\bar{G}(s \mid z_0)\bar{G}(t \mid z_0)\Bigg[&\int_0^{s\wedge t} \frac{1}{M_0^2}\,dF(\cdot, 1) + \frac{1}{I_*}\left\{\int_0^s \left(\frac{M_1}{M_0} - z_0\right)\frac{1}{M_0}\,dF(\cdot, 1)\right\} \\ &\cdot\left\{\int_0^t \left(\frac{M_1}{M_0} - z_0\right)\frac{1}{M_0}\,dF(\cdot, 1)\right\}\Bigg]\end{aligned}$$

by using (6.7), where I_* is given in (6.6). This agrees with Lemma 6.2 of Tsiatis (1981).

REFERENCES

Begun, Janet M. (1981). A class of rank estimators of relative risk. Institute of Statistics Mimeo Series #1381, Univ. of North Carolina at Chapel Hill.

Begun, Janet M. (1981). Estimates of relative risk. Institute of Statistics Mimeo Series #1382, Univ. of North Carolina at Chapel Hill.

Begun, Janet M. and Wellner, Jon A. (1983). Asymptotic efficiency of relative risk estimates. In *Contributions to Statistics: Essays in Honor of Norman L. Johnson*, ed. by P. K. Sen, to appear.

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

Beran, R. (1977a). Estimating a distribution function. *Ann. Statist.* **5** 400–404.

Beran, R. (1977b). Robust location estimates. *Ann. Statist.* **5** 431–444.

Beran, R. (1977c). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

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

Beran, R. (1980). Asymptotic lower bounds for risk in robust estimation. *Ann. Statist.* **8** 1252–1264.

Beran, R. (1981). Efficient robust estimates in parametric models. *Z. Wahrsch. verw. Gebiete* **55** 91–108.

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

Bickel, P. J. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models II. Location. *Ann. Statist.* **3** 1045–1069.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89–99.

Chen, Neng Hsin (1980). On the construction of a nonparametric efficient location estimator. Ph.D. Thesis, University of California at Berkeley.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–279.

Efron, Bradley (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

Fabian, Vaclav and Hannan, James (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. verw. Gebiete* **59** 459–478.

Geman, Stuart and Hwang, Chii-Ruey (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414.

Grenander, U. (1981). *Abstract Inference.* Wiley, New York.

Hájek, Jaroslav (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14** 323–330.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. of California Press, Berkeley.

Huang, Wei-Min (1982). Parameter estimation when there are nuisance functions. Ph.D. Thesis, University of Rochester.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

JEWELL, NICHOLAS P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479-484.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Statist.* **27** 887-906.

KALBFLEISCH, JOHN D. and PRENTICE, ROSS L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

KLAASSEN, C. A. J. (1981). Statistical performance of location estimators. *Mathematical Centre Tract* **133,** Mathematisch Centrum, Amsterdam.

KOSHEVNIK, YU. A. and LEVIT, B. YA. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738-753.

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Assoc.* **73** 805-811.

LO, SHAW-HWA (1982). Estimation of distribution functions and location parameters. Pre-print, Rutgers University.

LECAM, L. (1969). Theorie Asymptotique de la Decision Statistique. Les Presses de l'Universite de Montreal.

LECAM, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802-828.

LECAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. and Probab.* **1** 245-261. University of California Press, Berkeley.

LECAM, L. (1979). On a theorem of J. Hajek. In *Contributions to Statistics; Jaroslav Hajek Memorial Volume*, Ed. J. Jureckova, 119-135. Reidel, Dordrecht.

LEVIT, B. YA. (1975). On the efficiency of a class of non-parametric estimates. *Theory Probab. Appl.* **20** 723-740.

LEVIT, B. YA. (1978). Infinite-dimensional informational lower bounds. *Theory Probab. Appl.* **23** 388-394.

LEVIT, B. YA. and SAMAROV, A. M. (1978). Estimation of spectral functions. *Problems Inform. Transmission* **14** 120-124.

LINDSAY, B. (1978). Ph.D. Thesis, University of Washington, Seattle, Washington.

LINDSAY, BRUCE G. (1980). Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators. *Philos. Trans. Roy. Soc. London Ser. A* **296** 639-665.

LUENBERGER, DAVID G. (1969). *Optimization by Vector Space Methods.* Wiley, New York.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. verw. Gebiete* **48** 233-252.

MILLAR, P. W. (1981a). Robust estimation via minimum distance methods. *Z. Wahrsch. verw. Gebiete* **55** 72-89.

MILLAR, P. W. (1981b). The minimax principle in asymptotic statistical theory. *Proc. Ecole d'Ete St. Flour,* to appear.

NEYMAN, J. (1958). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics: The Cramér Volume*, 213-234. Almquist and Wiksells, Uppsala.

OAKES, DAVID (1981). Survival times: aspects of partial likelihood. *Internat. Statist. Rev.* **49** 235-264.

REIDER, HELMUT (1981). On local asymptotic minimaxity and admissibility in robust estimation. *Ann. Statist.* **9** 266-277.

SCHOLZ, F. W. (1980). Towards a unified definition of maximum likelihood. *Canad. J. Statist.* **8** 193-203.

STEIN, CHARLES (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187-195. University of California Press, Berkeley.

STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267-284.

TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244-248.

TSIATIS, ANASTASIOS A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93-108.

WELLNER, JON A. (1982). Asymptotic optimality of the product-limit estimator. *Ann. Statist.* **10** 595-602.

WILKS, SAMUEL S. (1962). *Mathematical Statistics.* Wiley, New York.

JANET M. BEGUN
DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL
321 PHILLIPS HALL 039A
CHAPEL HILL, NORTH CAROLINA 27514

WEI-MIN HUANG
DEPARTMENT OF MATHEMATICS
CHRISTMAS-SANCON HALL 14
LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA 18015

W. J. HALL
DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NEW YORK 14627

JON A. WELLNER
DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NEW YORK 14627