The Annals of Statistics
1991, Vol. 19, No. 1, 178–204

ON DIFFERENTIABLE FUNCTIONALS[1]

BY AAD VAN DER VAART

Free University Amsterdam

Given a sample of size n from a distribution P_λ, one wants to estimate a functional $\psi(\lambda)$ of the (typically infinite-dimensional) parameter λ. Lower bounds on the performance of estimators can be based on the concept of a differentiable functional $P_\lambda \to \psi(\lambda)$. In this paper we relate a suitable definition of differentiable functional to differentiability of $\lambda \to dP_\lambda^{1/2}$ and $\lambda \to \psi(\lambda)$. Moreover, we show that regular estimability of a functional implies its differentiability.

1. Introduction. Let $\mathscr{P}$ be a class of probability distributions on a measurable space $(\mathscr{X}, \mathscr{B})$ and let $X_1, \dots, X_n$ be i.i.d. random elements distributed according to an unknown $P \in \mathscr{P}$. The statistical problem is to estimate the value at P of a functional $\kappa: \mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ taking its values in a normed linear space.

The question how well κ can be estimated asymptotically as $n \to \infty$ has been studied by many authors [e.g., Hájek (1970, 1972), Le Cam (1972, 1986), Beran (1977), Levit (1978), Wellner (1983), Begun, Hall, Huang and Wellner (1983) and Millar (1983, 1985)]. The best-known results are the convolution and local asymptotic minimax (LAM) theorem in the case of locally asymptotic normal (LAN) models. For the case that $\mathbf{B}$ equals $\mathbb{R}^k$ Pfanzagl (1982), following ideas of Koshevnik and Levit (1976), works with a particular notion of *differentiable functional* κ, which makes a unified and relatively simple treatment of the asymptotic lower bound theory possible. As shown in van der Vaart (1988), this approach can be extended to general normed spaces.

In their study of semiparametric models, Begun, Hall, Huang and Wellner (1983) have introduced models P indexed by and differentiable in an infinitely dimensional parameter. From the derivative A of the density with respect to the parameter, they construct an *information operator* A^*A and obtain a convolution and LAM theorem under the assumption that this is one-to-one and onto. One can always parametrize a model by the roots of the probability measures P themselves. Then the derivative A is just the identity operator and the preceding condition is trivially satisfied. The main attraction of the Begun, Hall, Huang and Wellner (1983) approach, though, is the possibility to apply it to a "natural" parametrization, directly suggested by the underlying probability mechanism. Unfortunately, it has become increasingly clear that in

Received July 1988; revised May 1989.

[1]Research partially supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) and ONR contract N00014-80-C-0163.

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

Key words and phrases. Convolution theorem, asymptotic efficiency, semi-parametric model, information operator, efficient information, mixture model, censoring, truncation.

this case the previously mentioned assumption is too strong. Not only is the information operator often not one-to-one. More seriously, its range is also often not closed, which implies that it cannot be onto. Still, useful convolution and LAM theorems may be obtained by the approach based on *differentiable functionals*.

In this paper we describe the set of functionals which are differentiable in the sense of Pfanzagl (1982) and van der Vaart (1988) in terms of the derivative A of the density. This generalizes the approach of Begun, Hall, Huang and Wellner (1983). Moreover, we show that the existence of reasonable estimators of a functional implies its differentiability. The combination of these results allows us to investigate in several examples whether certain functionals are estimable at $\sqrt{n}$-rate.

A more precise outline of the paper is as follows. In a model P indexed by a Hilbert space valued parameter λ, we consider functionals $\kappa(P_\lambda)$ which are functionals $\psi(\lambda)$ of λ. We assume that both the maps $\lambda \to dP_\lambda^{1/2}$ and $\lambda \to \psi(\lambda)$ are differentiable. As a necessary and sufficient condition for $\kappa(P_\lambda)$ to be *differentiable* on P, we obtain that the gradients of ψ must be contained in the range of the adjoint operator of A (Section 3). We explain how this extends the approach of Begun, Hall, Huang and Wellner (1983). A further extension turns out to be impossible as it can be shown that the condition is necessary for the existence of regular estimator sequences (Section 2). For a real-valued functional the condition has an important statistical interpretation: It is equivalent to the condition that the *efficient information* is positive and finite (Section 4). In Sections 5 and 6, we translate the more general result of Section 3 to parametric and semiparametric models, respectively. Finally, we give several examples: mixture models, the censoring model, random truncation and incomplete censored observations (Sections 7–10). Technical proofs have been gathered in the appendix.

2. Differentiable functionals. In this section we state the definition of a differentiable functional that is used in this paper. It refers to a set of paths in $\mathscr{P}$ which start at a fixed point $P \in \mathscr{P}$ and is closely related to the definitions in Pfanzagl (1982) and van der Vaart (1988). Let $L_2(P)$ be the set of (equivalence classes of) measurable, P-square integrable functions, with inner product $\langle g_1, g_2\rangle_P = \int g_1 g_2\, dP$ and norm $\|g\|_P = \{\int g^2\, dP\}^{1/2}$; let $L_{2*}(P)$ be the subset of $g \in L_2(P)$ with $\int g\, dP = 0$.

Let $\mathscr{P}(P)$ be a collection of maps $t \to P_t$ from an interval $(0, \varepsilon) \subset \mathbb{R}$ to $\mathscr{P}$, satisfying for some $g \in L_2(P)$,

$$\int \left[t^{-1}\left(dP_t^{1/2} - dP^{1/2}\right) - \tfrac{1}{2} g\, dP^{1/2}\right]^2 \to 0 \quad \text{as } t \downarrow 0. \tag{2.1}$$

[Here dP_t and dP are densities of P_t and P, respectively, with respect to an arbitrary σ-finite measure dominating both P_t and P. The left-hand side of (2.1) is the same for every choice of such a dominating measure.] Note that g is a "quadratic mean version" of the "score function" $\partial[\log dP_t(x)]/\partial t|_{t=0}$.

To $\mathscr{P}(P)$ corresponds a *tangent space* $T(P)$, consisting of all elements g as in (2.1). As the name suggests, we assume that [$\mathscr{P}(P)$ can be and is chosen such that] $T(P)$ is a linear space. Moreover, we assume that $t \to P_{th}$ is in $\mathscr{P}(P)$ for every $h \in \mathbb{R}^+$, whenever $t \to P_t$ is.

Now $\kappa\colon \mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ is *differentiable at P relative to $\mathscr{P}(P)$*, if there exists a continuous, linear map $\kappa_P'\colon (T(P), \|\cdot\|_P) \to (\mathbf{B}, \|\cdot\|)$ such that

$$t^{-1}(\kappa(P_t) - \kappa(P)) \to \kappa_P'(g), \tag{2.2}$$

for every path $t \to P_t$ in $\mathscr{P}(P)$. [If $\mathscr{P}(P)$ is the set of all paths satisfying (2.1), then this could be formulated in terms of Hadamard differentiability. However, usually we do not take all paths.]

For differentiable functionals it is possible to derive a theory of bounds on the asymptotic performance of estimators, in particular, convolution and LAM theorems. Here the derivative κ_P' determines the optimal limiting measure, the form of which can be conveniently expressed in terms of the (*efficient*) *influence functions* $\tilde{\kappa}_{P, b^*}$ of κ.

These are defined as follows: For every b^* from the dual space $\mathbf{B}^*$ (the set of continuous, linear, real maps on $\mathbf{B}$), the map $b^* \circ \kappa_P'\colon (T(P), \|\cdot\|_P) \to \mathbb{R}$ is continuous and linear. Hence, by the Riesz representation theorem for Hilbert spaces, there is a unique element $\tilde{\kappa}_{P, b^*} \in \overline{T(P)}$, with

$$b^* \circ \kappa_P'(g) = \langle g, \tilde{\kappa}_{P, b^*}\rangle_P \quad \text{for every } g \in T(P). \tag{2.3}$$

Under the assumption that an optimal limiting measure N exists as a tight measure on the Borel σ-field of $(\mathbf{B}, \|\cdot\|)$, it can be uniquely determined by

$$N \circ b^{*-1} = N\big(0, \|\tilde{\kappa}_{P, b^*}\|_P^2\big), \qquad b^* \in \mathbf{B}^*. \tag{2.4}$$

Another way to express (2.4) is to write $N = \mathscr{L}(G)$ as the distribution of a Borel measurable random element G in $\mathbf{B}$. Then (2.4) is equivalent to saying that, for every finite set $b_1^*, \dots, b_k^*$ in $\mathbf{B}^*$, the Euclidean valued variable $(b_1^*(G), \dots, b_k^*(G))$ has a zero-mean multivariate Gaussian distribution with covariances determined by

$$Eb_i^*(G) b_j^*(G) = \langle \tilde{\kappa}_{P, b_i^*}, \tilde{\kappa}_{P, b_j^*}\rangle_P. \tag{2.4$'$}$$

Frequently, one considers a space $\mathbf{B}$ of uniformly bounded functions $b\colon T \to \mathbb{R}$ on some arbitrary index set T. Then the coordinate projections $\pi_t\colon b \to b(t)$ are contained in the dual space $\mathbf{B}^*$ and (2.4′) implies for this special situation,

$$EG(s)G(t) = \langle \tilde{\kappa}_{P, \pi_s}, \tilde{\kappa}_{P, \pi_t}\rangle_P.$$

In a sense, this definition is the "right" definition for use in connection with asymptotic bounds, as existence of reasonable estimators for a functional κ implies its differentiability. To state a result in this direction, we consider estimators $T_n = t_n(X_1, \dots, X_n)$ generated by maps $t_n\colon \chi^n \to \mathbf{B}$. Next, to avoid measurability problems, we formulate weak convergence in terms of outer integrals, following Hoffman-Jørgensen (1984). Precisely, we do *not* assume that the maps t_n are measurable and write

$$\sqrt{n}(T_n - b_n) \quad \Rightarrow_{P_{h_n/\sqrt{n}}} \quad L, \tag{2.5}$$

if and only if

$$E^*_{P_{h_n/\sqrt{n}}} f\left(\sqrt{n}(T_n - b_n)\right) \to \int f dL,$$

for every bounded, continuous $f\colon \mathbf{B} \to \mathbb{R}$ and a tight Borel measure L on $\mathbf{B}$. In the case that the maps t_n are measurable from $\mathscr{B}^n$ into the Borel σ-field, (2.5) is just weak convergence as usual. The minor extension using outer integral E^* is useful, since in the case of nonseparable $\mathbf{B}$, many estimators may correspond to maps t_n which are not Borel measurable. van der Vaart and Wellner (1989) contains a review of this notion of weak convergence suited to the proofs in the present paper.

THEOREM 2.1. *Suppose that, for every* $\{P_t\} \in \mathscr{P}(P)$ *and* $h_n \to h \in \mathbb{R}$,

$$\sqrt{n}\left(T_n - \kappa\left(P_{h_n/\sqrt{n}}\right)\right) \quad \Rightarrow_{P_{h_n/\sqrt{n}}} \quad L, \tag{2.6}$$

where L is a fixed tight Borel law on $\mathbf{B}$, *the same for every path* $\{P_t\}$. *Moreover, assume that, for every* $g \in T(P)$, $(\sqrt{n}(T_n - \kappa(P)), n^{-1/2}\sum_{j=1}^{n} g(X_j))$ *converges weakly under* P *on* $\mathbf{B} \times \mathbb{R}$ *[or, alternatively,* (2.7)*]. Then* $\kappa\colon \mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ *is differentiable at* $P \in \mathscr{P}$ *relative to* $\mathscr{P}(P)$. *Moreover, there exists a tight Borel measure N satisfying* (2.4) *and L is the convolution of N and some other Borel probability measure on* $\mathbf{B}$.

The last assertion of Theorem 2.1 is one form of the convolution theorem. The main reason to quote theorem 2.1 here, however, is the part asserting differentiability of κ. The proof of this assertion can be found in the appendix. For a proof of the present version of the convolution theorem we refer to van der Vaart and Wellner (1989), or Bickel, Klaassen, Ritov and Wellner (1990). Somewhat different versions of the convolution theorem (for the case that $\mathbf{B}$ is Euclidean space or a separable Banach space) can be found in Pfanzagl (1982) and Millar (1985).

Assumption (2.6) may [following Hájek (1970)] be called *regularity* of the estimator sequence $\{T_n\}$. In addition to this, Theorem 2.1 requires joint weak convergence of the standardized estimator and the log likelihood ratio (along the whole sequence $\{n\}$, not just along subsequences). We have not been able to prove the differentiability part of Theorem 2.1 without this second condition. Note that regular estimator sequences of which the marginals are asymptotically linear certainly do have the property of joint convergence. It is shown in the appendix that [under (2.6)] the joint convergence condition is equivalent to

$$\lim_{t \downarrow 0} t^{-1}(\kappa(P_t) - \kappa(P)) \quad \text{exists.} \tag{2.7}$$

Hence, existence of an "ordinary" regular estimator (2.6) and "ordinary" differentiability along paths, (2.7), also imply differentiability in the above sense.

3. Models parametrized by a subset of a Hilbert space. Let $(\mathbf{H}, \langle \cdot , \cdot \rangle_{\mathbf{H}})$ be a Hilbert space and suppose that $\mathscr{P} = \{P_\lambda \colon \lambda \in \Lambda\}$ for some subset Λ of $\mathbf{H}$. Fix $\lambda \in \Lambda$ and let $\Lambda(\lambda)$ be a set of paths $t \to \lambda_t$ such that

$$t^{-1}(\lambda_t - \lambda) \to \alpha \quad \text{as } t \downarrow 0, \tag{3.1}$$

for elements $\alpha \in \mathbf{H}$. Let $T(\lambda, \Lambda)$ be the set of all α thus obtained. Assume that it is a closed, linear subspace and that $t \to \lambda_{ht}$ is in $\Lambda(\lambda)$ for every $h \in \mathbb{R}^+$ if $t \to \lambda_t$ is. Furthermore, assume the existence of a continuous, linear operator $A = A_\lambda \colon T(\lambda, \Lambda) \to L_2(P_\lambda)$ such that

$$\int \left[t^{-1}\left(dP_{\lambda_t}^{1/2} - dP_\lambda^{1/2} \right) - \tfrac{1}{2} A\alpha \, dP_\lambda^{1/2} \right]^2 \to 0 \quad \text{as } t \downarrow 0, \tag{3.2}$$

for every path λ_t in $\Lambda(\lambda)$ [satisfying (3.1)]. This assumption is related to but weaker than the assumption of Hellinger differentiability in Begun, Hall, Huang and Wellner (1983). [Indeed, if $\Lambda(\lambda)$ is the set of all paths as in (3.1), then (3.2) is precisely Hadamard differentiability of $\lambda \to dP_\lambda^{1/2}$, whereas the condition in Begun, Hall, Huang and Wellner is Fréchet differentiability; for this terminology cf. Averbukh and Smolyanov (1967).]

As a collection of paths $\mathscr{P}(P_\lambda)$ in the sense of Section 2 we take all paths $t \to P_{\lambda_t}$ where $t \to \lambda_t$ is in $\Lambda(\lambda)$. Then the tangent space is the range of A: $T(P_\lambda) = AT(\lambda, \Lambda) = R(A)$. We consider a functional κ of the form

$$\kappa(P_\lambda) = \psi(\lambda). \tag{3.3}$$

Of course, we have to assume that this is well defined, i.e., that the functional ψ is identifiable over P. We also assume that $\psi \colon \Lambda \to \mathbf{B}$ is differentiable at λ in the sense that

$$t^{-1}(\psi(\lambda_t) - \psi(\lambda)) \to \psi_\lambda'(\alpha), \tag{3.4}$$

for some continuous, linear map $\psi_\lambda' \colon (T(\lambda, \Lambda), \| \cdot \|_P) \to (\mathbf{B}, \| \cdot \|)$ and every path λ_t in $\Lambda(\lambda)$ satisfying (3.1).

We now obtain a necessary and sufficient condition on A and ψ for differentiability of the functional κ in the sense of Section 2. First, we introduce the adjoint of A and gradients of ψ. The adjoint of A is the map A^*: $L_2(P_\lambda) \to T(\lambda, \Lambda)$ which satisfies $\langle A\alpha, g \rangle_{P_\lambda} = \langle \alpha, A^* g \rangle_{\mathbf{H}}$ for every $\alpha \in T(\lambda, \Lambda)$ and $g \in L_2(P_\lambda)$.

If $b^* \in \mathbf{B}^*$, then the map $b^* \circ \psi_\lambda'$ is a continuous, linear map from $T(\lambda, \Lambda)$ to $\mathbb{R}$. Hence, it has a representation as an inner product,

$$b^* \circ \psi_\lambda'(\alpha) = \langle \alpha, \tilde{\psi}_{\lambda, b^*} \rangle_{\mathbf{H}}, \tag{3.5}$$

for a unique $\tilde{\psi}_{\lambda, b^*} \in T(\lambda, \Lambda)$.

We shall refer to $\tilde{\psi}_{\lambda, b^*}$ as a *gradient* of ψ. Gradients are similar to influence functions as defined in (2.3). Different authors use different terminology here. In this paper we have gradients/influence functions on two levels, and (as we shall see) different parametrizations lead to influence functions/gradients which may differ by $\frac{1}{2}$ or a root density. We hope that this will not lead to misunderstanding.

We are ready to state the main result of this section.

THEOREM 3.1. *κ: $\mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ as in (3.3) is differentiable relative to $\mathscr{P}(P_\lambda)$ at P_λ if and only if*

$$\tilde{\psi}_{\lambda, b^*} \in R(A^*) \quad \text{for every } b^* \in \mathbf{B}^*. \tag{3.6}$$

The efficient influence functions of κ are related to the gradients of ψ by

$$\tilde{\psi}_{\lambda, b^*} = A^* \tilde{\kappa}_{P_\lambda, b^*}, \qquad \tilde{\psi}_{\lambda, b^*} \in T(\lambda, \Lambda), \ \tilde{\kappa}_{P_\lambda, b^*} \in \overline{R(A)}. \tag{3.7}$$

Relation (3.5) defines a map $(\psi_\lambda')^*$: $\mathbf{B}^* \to \mathbf{H}$ through $(\psi_\lambda')^*(b^*) = \tilde{\psi}_{\lambda, b^*}$. Identifying $\mathbf{H}$ and its dual $\mathbf{H}^*$ as usual, it is just the adjoint of ψ_λ'. Thus, we can restate (3.6) as

$$R\big((\psi_\lambda')^*\big) \subset R(A^*). \tag{3.6$'$}$$

With this notation, the above theorem remains true if the Hilbert space $\mathbf{H}$ is replaced by an arbitrary Banach space, provided the adjoints are given the right interpretation. This is not true for several results in the sequel, though, and the Hilbert space case appears to be by far the most interesting case.

PROOF OF THEOREM 3.1. Suppose κ is differentiable. Then, by (2.2), (3.3) and (3.4),

$$\kappa_{P_\lambda}'(A\alpha) = \lim_{t \downarrow 0} t^{-1}(\kappa(P_t) - \kappa(P)) = \psi_\lambda'(\alpha),$$

for every $\alpha \in T(\lambda, \Lambda)$. Thus, for every $b^* \in \mathbf{B}^*$ and $\alpha \in T(\lambda, \Lambda)$,

$$b^* \circ \kappa_{P_\lambda}'(A\alpha) = b^* \circ \psi_\lambda'(\alpha) = \langle \alpha, \tilde{\psi}_{\lambda, b^*} \rangle_{\mathbf{H}}.$$

On the other hand, by definition [cf. (2.3)],

$$b^* \circ \kappa_{P_\lambda}'(A\alpha) = \langle A\alpha, \tilde{\kappa}_{P_\lambda, b^*} \rangle_{P_\lambda} = \langle \alpha, A^* \tilde{\kappa}_{P_\lambda, b^*} \rangle_{\mathbf{H}}.$$

Combination implies (3.7) and hence (3.6).

Conversely, suppose that (3.6) holds. Define a "derivative" of κ on $R(A)$ by

$$\kappa_{P_\lambda}'(A\alpha) = \psi_\lambda'(\alpha), \qquad \alpha \in T(\lambda, \Lambda).$$

By (3.6′) this is well defined, for if $A\alpha_1 = A\alpha_2$, then $\alpha_1 - \alpha_2 \in N(A) = R(A^*)^\perp \subset R((\psi_\lambda')^*)^\perp = N(\psi_\lambda')$. Thus $\psi_\lambda'(\alpha_1) = \psi_\lambda'(\alpha_2)$. We only have to show that κ_{P_λ}' is continuous and linear. For every $b^* \in \mathbf{B}^*$,

$$b^* \circ \kappa_{P_\lambda}'(A\alpha) = b^* \circ \psi_\lambda'(\alpha) = \langle \alpha, \tilde{\psi}_{\lambda, b^*} \rangle_{\mathbf{H}}.$$

By (3.6), this is equal to

$$\langle \alpha, A^* \tilde{\kappa}_{P_\lambda, b^*} \rangle_{\mathbf{H}} = \langle A\alpha, \tilde{\kappa}_{P_\lambda, b^*} \rangle_{P_\lambda},$$

for some $\tilde{\kappa}_{P_\lambda, b^*} \in L_2(P_\lambda)$. But this shows that $b^* \circ \kappa_{P_\lambda}'$ is a continuous, linear real map on $R(A)$ for every $b^* \in \mathbf{B}^*$. By Lemma A.2 in the appendix, applied with $X = R(A)$, $Y = \mathbf{B}$ and $Y' = \mathbf{B}^*$, this is sufficient to conclude continuity and linearity of κ_{P_λ}': $R(A) \to \mathbf{B}$. □

Since we assume throughout that ψ is differentiable, (2.7) is trivially satisfied for the corresponding κ. Thus, Theorem 3.1 may be combined with Theorem 2.1 to say that (3.6) is necessary for the existence of regular estimator sequences. We return to this in the next section.

By taking orthocomplements we see that (3.6′) implies

$$N(A) \subset N(\psi_\lambda'). \tag{3.8}$$

Condition (3.8) has an easily understandable intuitive meaning as a (local) identifiability condition. Failure of the condition implies the existence of a path $t \to \lambda_t$ such that $\psi(\lambda_t) = \psi(\lambda) + t\psi_\lambda'(h) + o(t)$, $\psi_\lambda'(h) \neq 0$, whereas $P_{\lambda_t} = P_\lambda + o(t)$. Thus, P_{λ_t} is much closer to P_λ than $\psi(\lambda_t)$ to $\psi(\lambda)$ for t small, and we cannot hope to discriminate $\psi(\lambda_t)$ accurately from $\psi(\lambda)$. The examples support this interpretation of (3.8) as an identifiability condition.

The precise mathematical form of (3.6) is somewhat puzzling, though the condition can be translated in simple statistical terms (Section 4). In the rest of this section we show that (3.6) reduces to (3.8) if $R(A)$ is closed, and we show how one can obtain a formula reminiscent of some formulas of Begun, Hall, Huang and Wellner (1983) if (3.6) is strengthened to the condition that the gradients of ψ are contained in the range of A^*A. We start with the following corollary.

COROLLARY 3.2. *If* (3.8) *holds and* $R(A)$ *is closed, then* $\kappa\colon \mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ *is differentiable relative to* $\mathscr{P}(P_\lambda)$ *at* P_λ.

PROOF. $R(A)$ is closed if and only if $R(A^*)$ is closed [Rudin (1973), 4.14]. Then, by the well-known duality formulas and (3.8),

$$R(A^*) = \overline{R(A)} = N(A)^\perp \supset N(\psi_\lambda')^\perp = \overline{R((\psi_\lambda')^*)} \supset R((\psi_\lambda')^*). \qquad \square$$

Thus, if the identifiability condition (3.8) is satisfied, differentiability of κ can fail only if $R(A)$ is not closed. The condition that $R(A)$ is closed is a crude type of condition, as it does not refer to the particular functional ψ at all, whereas (3.6) does. We can go still one step further by also assuming that A is one-to-one. Then *any* functional of the form (3.3) is differentiable, since local identifiability (3.8) is trivially satisfied.

COROLLARY 3.3. *If* $N(A) = 0$ *and* $R(A)$ *is closed, then* $\kappa\colon \mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ *is differentiable relative to* $\mathscr{P}(P_\lambda)$ *at* P_λ.

The condition of Corollary (3.3) is the condition used by Begun, Hall, Huang and Wellner (1983) when obtaining bounds for estimating the distribution function of the nuisance parameter in semiparametric models. We can see this by rephrasing it in terms of the *information operator* $A^*A\colon T(\lambda, \Lambda) \to T(\lambda, \Lambda)$.

LEMMA 3.4.

(i) A^*A *is one-to-one and onto if and only if* $N(A) = 0$ *and* $R(A)$ *is closed.*
(ii) $R(A^*A) \subset R(A^*)$ *with equality if and only if* $R(A)$ *is closed.*

PROOF. (i) If A^*A is one-to-one, then A is one-to-one too. If A^*A is onto, then A^* is onto too and $R(A^*) = T(\lambda, \Lambda)$ is closed by assumption. Then $R(A)$ is closed too [Rudin (1973), 4.14]. Conversely, suppose that $N(A) = 0$ and $R(A)$ is closed. If $A^*A\alpha = 0$, then $A\alpha \in N(A^*) = R(A)^\perp$. Of course, $A\alpha \in R(A)$ too, so that it must be zero. Hence, $\alpha \in N(A) = 0$. That A^*A is onto follows from (ii) and the equality $R(A^*A) = R(A^*) = N(A)^\perp = 0^\perp$.

(ii) The inclusion is obvious. Suppose $R(A)$ is closed and $\alpha = A^*g$. Then g can be decomposed in $g = g_1 + g_2$, where $g_1 \in R(A)$ and $g_2 \in R(A)^\perp = N(A^*)$. We conclude that $\alpha = A^*g_1 = A^*A\beta$ for some β. Conversely, suppose that $R(A)$ is not closed. Let $g \in \overline{R(A)} - R(A)$. Then $\alpha = A^*g \in R(A^*) - R(A^*A)$; for suppose it is not. Then $\alpha = A^*A\beta$ for some β. Then $g - A\beta \in N(A^*) = R(A)^\perp$. Since trivially $g - A\beta \in \overline{R(A)}$, we obtain $g - A\beta = 0$. This contradicts the assumption. □

In the case of parametric models (cf. Section 5) $R(A)$ is finite dimensional and hence automatically closed. Unfortunately, in the case of infinite-dimensional parameter spaces, $R(A)$ fails to be closed much more often than we would want it to. (See the examples in Sections 7–10.) In a way, these situations are the more interesting ones, though both the statistical and mathematical problems are more complicated. Then there are smooth functionals ψ such that the corresponding κ given by (3.3) is not differentiable. Thus, there exist no regular estimators in the sense of Section 2, which roughly means that estimation of ψ at $\sqrt{n}$-rate is impossible. Preliminary results show that, in principle, any rate slower than $\sqrt{n}$ occurs for some ψ, depending (possibly) on the precise position of $\tilde{\psi}_{\lambda, b^*}$ outside $R(A^*)$.

If $R(A)$ is not closed, we have to check (3.6) for the particular functional at hand to see whether ψ is estimable at $\sqrt{n}$-rate. By Lemma 3.4(ii), it may be replaced as a sufficient condition for differentiability of κ by the stronger condition

$$\tilde{\psi}_{\lambda, b^*} \in R(A^*A) \quad \text{for every } b^* \in \mathbf{B}^*. \tag{3.9}$$

Under (3.9) an alternative method for obtaining the influence functions of κ in terms of the gradients of ψ is as follows: Let $\alpha_0 \in T(\lambda, \Lambda)$ be a solution to

$$\tilde{\psi}_{\lambda, b^*} = A^*A\alpha.$$

A^*A need not be one-to-one. However, use the suggestive notation $\alpha_0 = (A^*A)^-\tilde{\psi}_{\lambda, b^*}$. Then,

$$\tilde{\kappa}_{P_\lambda, b^*} = A(A^*A)^-\tilde{\psi}_{\lambda, b^*}. \tag{3.10}$$

This function clearly satisfies (3.7) and, in fact, is the unique solution in $\overline{R(A)}$ of the equation $\tilde{\psi}_{\lambda, b^*} = A^*g$, $g \in L_2(P_\lambda)$. [If $A^*g_1 = A^*g_2$, then $g_1 - g_2 \in N(A^*) = R(A)^\perp$. Then, if moreover both g_i are in $\overline{R(A)}$, $g_1 - g_2$ is too, so that $g_1 - g_2 = 0$.]

By Lemma 3.4(ii), condition (3.9) can be really stronger than (3.6). In fact, (3.10) shows that (3.9) is a sufficient condition for the influence function of κ

to be not only contained in $\overline{R(A)}$ (which is true by definition), but also in $R(A)$. A closer look reveals that this may also be reversed. Thus, there are two cases. If $\tilde{\psi}_{\lambda, b^*}$ is contained in $R(A^*A)$, then the influence function of κ is contained in $R(A)$ and can be found from (3.10). On the other hand, when $\tilde{\psi}_{\lambda, b^*}$ is contained in $R(A^*) - R(A^*A)$, then the influence function of κ is contained in $\overline{R(A)} - R(A)$ and we must use (3.7).

Here we note that the equation $A^*g = \tilde{\psi}_{\lambda, b^*}$ has multiple solutions g in $L_2(P_\lambda)$. To find the unique solution $g = \tilde{\kappa}_{P_\lambda, b^*} \in \overline{R(A)}$, we could first find a solution g_0 of $A^*g = \tilde{\psi}_{\lambda, b^*}$ in $L_2(P_\lambda)$ and next, project g_0 onto $\overline{R(A)}$ to find $\tilde{\kappa}_{P_\lambda, b^*}$.

4. Differentiability and efficient information. In Section 3 we saw that (3.6) is necessary and sufficient for differentiability of a functional κ of the form (3.3) (given that ψ is differentiable). Since differentiability of ψ clearly implies (2.7), it follows by combination with Theorem 2.1 that (3.6) is also necessary for the existence of a regular estimator sequence for κ.

In this section we provide a further expression of this in terms of the *efficient Fisher information*. We restrict ourselves to real-valued functionals. (It is possible to state similar results for abstract functionals by reducing these to the set of functionals $b^* \circ \kappa$.) It will be shown that if (3.6) fails, then there exists a submodel $t \to P_{\lambda_t}$ such that the information about κ in the submodel is arbitrarily close to zero.

To make this precise, define the information about κ in $t \to \mathscr{P}_{\lambda_t}$, where $t \to \lambda_t$ satisfies (3.1), at P_λ by

$$i_\alpha = \frac{\|A\alpha\|_{P_\lambda}^2}{\langle \tilde{\psi}_\lambda, \alpha \rangle_{\mathbf{H}}^2}. \tag{4.1}$$

Here $\tilde{\psi}_\lambda$ is the gradient of ψ in the direction $1 \in \mathbb{R}$ [i.e., $\psi_\lambda'(\alpha) = \langle \alpha, \tilde{\psi}_\lambda \rangle_{\mathbf{H}}$]. The inverse of i_α is precisely the familiar lower bound given by the Cramér–Rao theorem: $\|A\alpha\|_{P_\lambda}^2$ is the information about t, while $\langle \tilde{\psi}_\lambda, \alpha \rangle_{\mathbf{H}} = \partial \psi(\lambda_t)/\partial t|_{t=0}$. Hence, the inverse of the information number defined in (4.1) is a lower bound for the variance of an estimator which is unbiased for $\psi(\lambda_t)$ over the submodel $t \to P_{\lambda_t}$. It can also be given an asymptotic interpretation, for instance, as a lower bound for the local asymptotic minimax risk over $t \to P_{\lambda_t}$ in the sense of Hájek (1972). This follows (almost) directly from Hájek (1972) and the fact that (3.2) implies local asymptotic normality:

$$\log \bigotimes_{j=1}^n \frac{dP_{\lambda_{h/\sqrt{n}}}}{dP_\lambda}(X_j) = hn^{-1/2} \sum_{j=1}^n A\alpha(X_j) - \frac{1}{2} h^2 \|A\alpha\|_{P_\lambda}^2 + o_{P_\lambda}(1).$$

Now, by "efficient Fisher information" we mean the infimum over all one-dimensional submodels of the expression in (4.1). We have that the efficient information is positive if and only if (3.6) holds.

THEOREM 4.1. *In the situation of Section* 3, $\inf_{\alpha \in T(\lambda, \Lambda)} i_\alpha > 0$ *if and only if* (3.6) *holds.*

PROOF. If (3.6) holds, then for every $\alpha \in T(\lambda, \Lambda)$,

$$i_\alpha^{-1} = \frac{\langle A^*\tilde{\kappa}_{P_\lambda, b^*}, \alpha \rangle_{\mathbf{H}}^2}{\|A\alpha\|_{P_\lambda}^2} = \frac{\langle \tilde{\kappa}_{P_\lambda, b^*}, A\alpha \rangle_{P_\lambda}^2}{\|A\alpha\|_{P_\lambda}^2} \le \|\tilde{\kappa}_{P_\lambda, b^*}\|_{P_\lambda}^2.$$

Our proof of the converse is somewhat involved and is deferred to the Appendix. □

5. Parametric models. The results of this section are not surprising. However, it may be helpful to have the more abstract situation of Section 3 translated back to the familiar situation of a smooth finite-dimensional model $\mathscr{P} = \{P_\theta: \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$ is open. Suppose

$$\int \left[t^{-1}\left(dP_{\theta + th_t}^{1/2} - dP_\theta^{1/2} \right) - \tfrac{1}{2} h' \dot{\ell}_\theta \, dP_\theta^{1/2} \right]^2 \to 0 \quad \text{as } t \to 0,\ h_t \to h,$$

for some $\dot{\ell}_\theta \in L_2(P_\theta)^k$. Let $I_\theta = \int \dot{\ell}_\theta \dot{\ell}_\theta' \, dP_\theta$ be the Fisher information matrix.

In the notation of Section 3 we have $Ah = h'\dot{\ell}_\theta$, $A^*g = \langle \dot{\ell}_\theta, g \rangle_{P_\theta}$ and $N(A) = \{h: h' I_\theta h = 0\} = N(I_\theta)$.

Let $\kappa(P_\theta) = \psi(\theta)$, where $\psi: \Theta \to \mathbb{R}^m$ is differentiable in the ordinary sense with derivative ψ_θ'. Because any finite-dimensional linear space is closed, we may use Corollary 3.2 to see that relation (3.6) reduces to the present form of (3.8),

$$N(I_\theta) \subset N(\psi_\theta').$$

In particular, to estimate $\psi(\theta) = \theta$, the Fisher information matrix must be nonsingular.

6. Semiparametric models. This section treats a special case of the general model of Section 3. The parameter λ is split into a Euclidean part and a probability density. The results are merely a translation of the results obtained before to this special situation.

Let $\mathscr{P} = \{P_{\theta\eta}: \theta \in \Theta,\ \eta \in \mathscr{H}\}$, where $\Theta \subset \mathbb{R}^k$ is open and $\mathscr{H}$ is the class of all probability densities with respect to a σ-finite measure ν on a measurable space $(\mathscr{Y}, \mathscr{A})$. Assume the existence of $\dot{\ell}_\theta \in L_2(P_{\theta\eta})^k$ and of a continuous, linear operator $\dot{\ell}_\eta: L_{2*}(\eta) \to L_2(P_{\theta\eta})$ such that for every $h \in \mathbb{R}^k$

$$(6.1) \qquad \int \left[t^{-1}\left(dP_{\theta+th, \eta_t}^{1/2} - dP_{\theta\eta}^{1/2} \right) - \tfrac{1}{2}\left(\dot{\ell}_\theta' h + \dot{\ell}_\eta \beta \right) dP_{\theta\eta}^{1/2} \right]^2 \to 0,$$

whenever $t \downarrow 0$, and

$$(6.2) \qquad \int \left[t^{-1}\left(\eta_t^{1/2} - \eta^{1/2} \right) - \tfrac{1}{2}\beta \eta^{1/2} \right]^2 d\nu \to 0.$$

[Of course, the derivatives $\dot{\ell}_\theta'$ and $\dot{\ell}_\eta$ do depend on (θ, η). Following the convention introduced for A in Section 3, we don't let this show up in the notation.] This can be accommodated in the setup of Section 3 by setting $\Lambda = \{(\theta, \eta^{1/2}): \theta \in \Theta,\ \eta \in \mathscr{H}\}$, which is contained in the Hilbert space

$\mathbb{R}^k \times L_2(\nu)$. We consider the set of paths $t \to (\theta + th, \eta_t^{1/2})$, where $\eta_t^{1/2}$ satisfies (6.2). Then $T((\theta, \eta^{1/2}), \Lambda) = \mathbb{R}^k \times L_{2*}(\eta)\eta^{1/2}$ and A: $T((\theta, \eta^{1/2}), \Lambda) \to L_2(P_{\theta\eta})$ satisfies

$$A\left(h, \tfrac{1}{2}\beta\eta^{1/2}\right) = \dot{\ell}_\theta' h + \dot{\ell}_\eta \beta. \tag{6.3}$$

Since $\langle A(h, \frac{1}{2}\beta\eta^{1/2}), g\rangle_{P_{\theta\eta}} = h'\langle \dot{\ell}_\theta, g\rangle_{P_{\theta\eta}} + \langle \frac{1}{2}\beta\eta^{1/2}, 2\dot{\ell}_\eta^* g\eta^{1/2}\rangle_\nu$, we have

$$A^* g = \left(\langle \dot{\ell}_\theta, g\rangle_{P_{\theta\eta}}, 2\dot{\ell}_\eta^* g\eta^{1/2}\right). \tag{6.4}$$

Let us first consider estimation of a functional χ: $\Theta \to \mathbb{R}^m$, which is differentiable in the ordinary sense, with derivative the $(m \times k)$ matrix χ_θ'. To state the result, we need the concept of *efficient Fisher information matrix* for θ. This is defined as follows: Decompose every one of the k elements of $\dot{\ell}_\theta$ in a component in $\overline{R(\dot{\ell}_\eta)}$ and a component orthogonal to $R(\dot{\ell}_\eta)$. The k-vector of the latter is called the *efficient score function for* θ. Denote it by $\tilde{\ell}_\theta$. Thus,

$$\dot{\ell}_\theta = \tilde{\ell}_\theta + \left(\dot{\ell}_\theta - \tilde{\ell}_\theta\right),$$

where

$$\tilde{\ell}_\theta \in \overline{R\left(\dot{\ell}_\eta\right)}, \qquad \dot{\ell}_\theta - \tilde{\ell}_\theta \perp R\left(\dot{\ell}_\eta\right).$$

The efficient information matrix is defined by $\tilde{I}_{\theta\eta} = \int \tilde{\ell}_\theta \tilde{\ell}_\theta' \, dP_{\theta\eta}$.

COROLLARY 6.1. *The functional κ: $\mathscr{P} \to \mathbb{R}^m$ given through $\kappa(P_{\theta\eta}) = \chi(\theta)$ is differentiable at $P_{\theta\eta}$ if and only if $N(\tilde{I}_{\theta\eta}) \subset N(\chi_\theta')$.*

PROOF. Define ψ: $\Lambda \to \mathbb{R}^m$ by $\psi(\theta, \eta^{1/2}) = \chi(\theta)$. Then, for every $h \in \mathbb{R}^k$ and $\beta \in L_{2*}(\eta)$,

$$\begin{aligned}\psi_\lambda'\left(h, \tfrac{1}{2}\beta\eta^{1/2}\right) &= \lim_{t\downarrow 0} t^{-1}\left(\psi\left(\theta + th, \eta_t^{1/2}\right) - \psi\left(\theta, \eta^{1/2}\right)\right)\\ &= \lim_{t\downarrow 0} t^{-1}\left(\chi(\theta + th) - \chi(\theta)\right) = \chi_\theta'(h).\end{aligned}$$

Thus, the gradient of ψ in the direction $f \in \mathbb{R}^m$ is given by $\tilde{\psi}_{\lambda, f} = (f' \circ \chi_\theta', 0)$. This is contained in $R(A^*)$ if and only if

$$(f' \circ \chi_\theta')' \in \left\{\left\langle \dot{\ell}_\theta, g\right\rangle_{P_{\theta\eta}} : g \in N\left(\dot{\ell}_\eta^*\right) = R\left(\dot{\ell}_\eta\right)^\perp\right\} = R\left(\tilde{I}_{\theta\eta}\right).$$

This is equivalent to the condition of the corollary. □

Next consider estimation of a functional of η. Identifying the set of η's with a set of probability measures, we speak of differentiability in the sense of Section 2 of χ: $\mathscr{H} \to (\mathbf{B}, \|\cdot\|)$, relative to the set of all paths $t \to \eta_t$ in $\mathscr{H}$ which satisfy (6.2), and accordingly define its influence functions. [We prefer to view χ as a functional in this way rather than as a functional on a subset of $L_2(\nu)$, as this will lead to simpler influence functions/gradients. For instance, the influence function of $\eta \to \int_C \eta \, d\nu$ will be $1_C - \int_C \eta \, d\nu$ rather than $2(1_C -$

$\int_C \eta\, d\nu)\eta^{1/2}$.] Thus, we assume the existence of a continuous, linear operator χ'_η: $L_{2*}(\eta) \to (\mathbf{B}, \|\cdot\|)$ with

$$t^{-1}(\chi(\eta_t) - \chi(\eta)) \to \chi'_\eta(\beta), \tag{6.5}$$

for every path $t \to \eta_t$ satisfying (6.2). To avoid technicalities, we assume that θ and η are not locally confounded, i.e., that the efficient Fisher information matrix for θ is nonsingular.

COROLLARY 6.2. *Suppose that $\tilde{I}_{\theta\eta}$ is nonsingular. Then κ: $\mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ given by $\kappa(P_{\theta\eta}) = \chi(\eta)$ is differentiable at $P_{\theta\eta}$ if and only if*

$$\tilde{\chi}_{\eta, b^*} \in R\left(\dot{\ell}_\eta^*\right) \quad \textit{for every } b^* \in \mathbf{B}^*.$$

Influence functions of κ and χ are related by

$$\begin{aligned} 0 &= \left\langle \dot{\ell}_\theta, \tilde{\kappa}_{P_{\theta\eta}, b^*} \right\rangle_{P_{\theta\eta}}, \\ \tilde{\chi}_{\eta, b^*} &= \dot{\ell}_\eta^* \tilde{\kappa}_{P_{\theta\eta}, b^*}. \end{aligned} \tag{6.6}$$

Under the stronger condition that $\tilde{\chi}_{\eta, b^*} \in R(\dot{\ell}_\eta^* \dot{\ell}_\eta)$, the solution of (6.6) *is given by*

$$\tilde{\kappa}_{P_{\theta\eta}, b^*} = \dot{\ell}_\eta\left(\dot{\ell}_\eta^* \dot{\ell}_\eta\right)^{-} \tilde{\chi}_{\eta, b^*} - \left\langle \dot{\ell}_\eta\left(\dot{\ell}_\eta^* \dot{\ell}_\eta\right)^{-} \tilde{\chi}_{\eta, b^*}, \tilde{\ell}_\theta \right\rangle'_{P_{\theta\eta}} \tilde{I}_{\theta\eta}^{-1} \tilde{\ell}_\theta. \tag{6.7}$$

Here $(\dot{\ell}_\eta^* \dot{\ell}_\eta)^- \tilde{\chi}_{\eta, b^*}$ is a solution in $L_{2*}(\eta)$ of $\dot{\ell}_\eta^* \dot{\ell}_\eta \beta = \tilde{\chi}_{\eta, b^*}$.

PROOF. Define ψ: $\Lambda \to \mathbf{B}$ by $\psi(\theta, \eta^{1/2}) = \chi(\eta)$. Then, for every $h \in \mathbb{R}^k$ and $\beta \in L_{2*}(\eta)$,

$$\begin{aligned} \psi'_\lambda\left(h, \tfrac{1}{2}\beta\eta^{1/2}\right) &= \lim_{t \downarrow 0} t^{-1}\left(\psi(\theta + th, \eta_t^{1/2}) - \psi(\theta, \eta^{1/2})\right) \\ &= \lim_{t \downarrow 0} t^{-1}\left(\chi(\eta_t^{1/2}) - \chi(\eta^{1/2})\right) = \chi'_\eta(\beta). \end{aligned}$$

Thus, the gradient of ψ in the direction b^* is given by $\tilde{\psi}_{\lambda, b^*} = (0, 2\tilde{\chi}_{\eta, b^*}\eta^{1/2})$. The first part of the corollary follows from Theorem 3.1 if we can show that $R(A^*) = \mathbb{R}^k \times R(\dot{\ell}_\eta^*)\eta^{1/2}$. Now $R(A^*)$ is certainly not larger than the expression on the right. Furthermore, it contains the set

$$\left\{\left(\langle \dot{\ell}_\theta, g \rangle_{P_{\theta\eta}}, 2\dot{\ell}_\eta^* g \eta^{1/2}\right) : g \in \operatorname{lin}\left(\tilde{\ell}_\theta\right)\right\} = \mathbb{R}^k \times \{0\},$$

because $\operatorname{lin}(\tilde{\ell}_\theta) \subset R(\dot{\ell}_\eta^*)^\perp = N(\dot{\ell}_\eta^*)$. But then it also contains

$$\mathbb{R}^k \times \left\{\dot{\ell}_\eta^* g \eta^{1/2}\right\} = \left\{\left(\langle \dot{\ell}_\theta, g \rangle_{P_{\theta\eta}}, \dot{\ell}_\eta^* g \eta^{1/2}\right)\right\} + \mathbb{R}^k \times \{0\},$$

for every g.

Finally (6.6) is the translation of (3.7) and direct substitution shows that (6.7) gives a solution to (6.6), which is clearly contained in $\overline{R(A)}$. □

7. Example: Estimating a functional of a mixing distribution. Let $p(\cdot, z)$ be the density with respect to Lebesgue measure on $\mathbb{R}$ of a one-dimensional exponential family member, i.e.,

$$p(x,z) = h(x)c(z)e^{xz}, \qquad x \in \mathscr{X} \subset \mathbb{R}. \tag{7.1}$$

Let $\mathscr{H}$ be the set of continuous probability densities with respect to Lebesgue measure on the natural parameter space $\mathscr{Z} = \{z\colon \int h(x)e^{xz}\,dx < \infty\}$ of the family.

Set $\mathscr{P} = \{P_\eta\colon \eta \in \mathscr{H}\}$, where P_η has density

$$p(\cdot,\eta) = \int p(\cdot,z)\eta(z)\,dz. \tag{7.2}$$

[We abuse notation in using both $p(\cdot,\eta)$ and $p(\cdot,z)$.] This example fits in the setup of Section 3. We find it convenient to make translations similar to those in Section 6. Hence, we refer below to Section 6 with the Euclidean parameter left out.

We may think of the present model as arising in the following manner: From an unobservable pair (X, Z) with density $p(x,z)\eta(z)$ we observe the measurable transformation $t(X,Z) = X$. Thus, it follows from general theorems on the preservation of LAN and differentiability in quadratic mean [cf. Le Cam and Yang (1988), van der Vaart (1988), Appendix A.3, and Bickel, Klaassen, Ritov and Wellner (1990)], that in the present situation (6.2) implies (6.1). Here, $\dot{\ell}_\eta\colon T(\eta,\mathscr{H}) = L_{2*}(\eta) \to L_2(P_\eta)$ is defined by

$$\dot{\ell}_\eta\beta(x) = E_\eta(\beta(Z)\,|\,X = x) = \frac{\int \beta(z)p(x,z)\eta(z)\,dz}{p(x,\eta)}. \tag{7.3}$$

Using Fubini's theorem, it is then easily checked that

$$\begin{aligned}\dot{\ell}_\eta^* g(z) &= E_\eta(g(X)\,|\,Z = z) - E_\eta g(X)\\ &= \int g(x)p(x,z)\,dx - \int g\,dP_\eta.\end{aligned} \tag{7.4}$$

Completeness of the exponential family implies

$$N\big(\dot{\ell}_\eta\big) = 0 \quad \text{and} \quad N\big(\dot{\ell}_\eta^*\big) = \{\text{constants}\}, \tag{7.5}$$

whence by taking orthocomplements,

$$\overline{R\big(\dot{\ell}_\eta\big)} = L_{2*}(P_\eta) \quad \text{and} \quad \overline{R\big(\dot{\ell}_\eta^*\big)} = L_{2*}(\eta). \tag{7.6}$$

Consider a differentiable functional $\chi\colon \mathscr{H} \to \mathbb{R}$ [cf. (6.5)] and let $\tilde{\chi}_\eta$ be its influence function in the direction $1 \in \mathbb{R}^*$ [i.e., $\chi'_\eta(\beta) = \int \tilde{\chi}_\eta \beta\eta\,dz$]. By Corollary 6.1, $\kappa(P_\eta) = \chi(\eta)$ is differentiable at P_η if and only if

$$\tilde{\chi}_\eta \in R\big(\dot{\ell}_\eta^*\big). \tag{7.7}$$

The influence function of κ is then given by $\tilde{\kappa}_{P_\eta} = (\dot{\ell}_\eta^*)^- \tilde{\chi}_\eta$.

Unfortunately, $R(\dot{\ell}_\eta^*)$ is not closed. If it was, then, by the second part of (7.6), it would contain, e.g., the functions

$$1_{[-\infty,u]} - \int_{-\infty}^{u} \eta(z)\,dz, \qquad u \in \mathbb{R}. \tag{7.8}$$

However, from the completeness of the exponential family it follows readily that $\dot{\ell}_\eta^* g(z)$ cannot be constant in an open interval, unless it is constant everywhere. Hence the functions (7.8) are not contained in $R(\dot{\ell}_\eta^*)$.

The function in (7.8) is just the influence function for estimating the distribution function $\chi(\eta) = \int_{-\infty}^{u} \eta(z)\,dz$ of η. Theorem 4.1 tells us that the efficient information for estimating this functional equals zero. Thus there exist no estimators with a $\sqrt{n}$-rate (more precisely, with $\sqrt{n}$-rate "uniformly along one-dimensional paths in shrinking neighborhoods").

The study of which rates are attainable has only just started. We mention Ritov (1987), Carroll and Hall (1988) and Fan (1988), who consider "deconvolution" problems.

By the same argument we see that many other functionals are also not estimable at $\sqrt{n}$-rate. For instance, quantiles of η or M-functionals with an influence function which is constant on an interval. This should not make us blind to the fact that many functionals are regularly estimable. Let us show how our result can be used in a positive manner to characterize all the linear functionals of the form $\eta \to \int c\eta\,dz$ which are regularly estimable (c fixed, known). Each such functional has gradient $c - \int c\eta\,dz$. Hence, if it is regularly estimable, we must have $c - \int c\eta\,dz \in R(\dot{\ell}_\eta^*)$. From (7.4) we see that $c(z) = \int g_\eta(x)p(x,z)\,dx$, for some $g_\eta \in L_2(P_\eta)$. Since c itself does not depend on η, we must have

$$c(z) = \int g(x)p(x,z)\,dx,$$

where $g \in L_2(P_\eta)$ for every η. A regular estimator for this functional is given by $T_n = n^{-1}\sum_{j=1}^{n} g(X_j)$.

8. Example: Univariate right censoring. Let Y and C be independent, nonnegative random variables with distribution functions F and G. Let $\mathscr{P}$ be the set of all distributions P_{FG} of the pair $(Y \wedge C, 1_{Y \le C})$, when F and G range over all continuous distribution functions on $(0,\infty)$. This is again a model where one "looses information," because one only observes a measurable transformation of the underlying random element (Y, C). Therefore, we can again refer to Le Cam and Yang (1988), van der Vaart (1988) or Bickel, Klaassen, Ritov and Wellner (1990) to establish an implication of the form (3.1) $\Rightarrow$ (3.2). In the present case this can be described as follows:

Define maps from $L_{2*}(F)$ and $L_{2*}(G)$, respectively, to $L_2(P_{FG})$ by

$$\dot{\ell}_1 a(x,\delta) = E_{FG}\big(a(Y)\,|\,Y \wedge C = x, 1_{Y\le C} = \delta\big) = \delta a(x) + (1-\delta)\frac{\int_{(x,\infty)} a\,dF}{1-F(x)},$$

$$\dot{\ell}_2 b(x,\delta) = E_{FG}\big(b(C)\,|\,Y \wedge C = x, 1_{Y\le C} = \delta\big) = \delta\frac{\int_{[x,\infty)} b\,dG}{1-G(x)} + (1-\delta)b(x).$$

Define A: $L_{2*}(F) \times L_{2*}(G) \to L_2(P_{FG})$ by $A(a,b) = \dot{\ell}_1 a + \dot{\ell}_2 b$. Now, if

$$\int\left[t^{-1}\left(dF_t^{1/2} - dF^{1/2}\right) - \tfrac{1}{2} a\,dF^{1/2}\right]^2 \to 0 \tag{8.1a}$$

and

$$\int \left[t^{-1}\left(dG_t^{1/2} - dG^{1/2} \right) - \tfrac{1}{2} b \, dG^{1/2} \right]^2 \to 0, \tag{8.1b}$$

then

$$\int \left[t^{-1}\left(dP_{F_t G_t}^{1/2} - dP_{FG}^{1/2} \right) - \tfrac{1}{2} A(a, b) \, dP_{FG}^{1/2} \right]^2 \to 0. \tag{8.2}$$

In this model the distribution F is usually the parameter of interest. So, consider a functional of the type $\kappa(P_{FG}) = \chi(F)$, where $\chi\colon \mathscr{F} \to (\mathbf{B}, \|\cdot\|)$ is differentiable relative to the set of paths $t \to F_t$ satisfying (8.1) [in the sense of Section 2 and $\mathscr{F}$ the set of all distributions on $(0, \infty)$].

COROLLARY 8.1. *$\kappa\colon \mathscr{P} \to (\mathbf{B}, \|\cdot\|)$ is differentiable at P_{FG} relative to the paths $t \to P_{F_t G_t}$ generated by (8.1) if and only if*

$$\tilde{\chi}_{F, b^*} \in R\left(\dot{\ell}_1^* \right) \quad \textit{for every } b^* \in \mathbf{B}^*.$$

Influence functions of κ and χ are related by $\dot{\ell}_1^* \tilde{\kappa}_{P_{FG}, b^*} = \tilde{\chi}_{F, b^*}$.

PROOF. It is straightforward to establish that $A^* g = (\dot{\ell}_1^* g, \dot{\ell}_2^* g)$ [where $g \in L_2(P_{FG})$] and, using orthogonality of $R(\dot{\ell}_1)$ and $R(\dot{\ell}_2)$, that $R(A^*) = R(\dot{\ell}_1^*) \times R(\dot{\ell}_2^*)$. The corollary can then be derived from Theorem 3.1 in the same manner as Corollary 6.1. □

Let $\tau_G = \sup\{s\colon G(s) < 1\}$ be the right end-point of G and define τ_F similarly. Then,

$$N\left(\dot{\ell}_1 \right) = \left\{ a \in L_{2*}(F) \colon a(x) = 0 \text{ for } F\text{-almost all } x < \tau_G \right\}. \tag{8.3}$$

The present form of (3.8) is $N(\dot{\ell}_1) \subset N(\chi_F')$. This is satisfied if χ depends on F only through the values of F on $[0, \tau_G)$. This supports the interpretation of (3.8) as an identifiability condition: It is clearly impossible to estimate F to the right of τ_G, as all Y's larger than τ_G will be censored.

Unfortunately, again $R(\dot{\ell}_1^*)$ need not be closed, so that the question whether a functional is differentiable needs careful analysis. A good starting point is the identity

$$\dot{\ell}_1^* \dot{\ell}_1 = R^{-1} S R, \tag{8.4}$$

where R is an isometry of $L_{2*}(F)$ onto $L_2(F)$, satisfying

$$Ra(x) = a(x) - \frac{\int_{(x, \infty)} a \, dF}{1 - F(x)} \quad \text{and} \quad R^{-1} a(x) = a(x) - \int_{[0, x]} a \frac{dF}{1 - F}, \tag{8.5}$$

and $S\colon L_2(F) \to L_2(F)$ is given by

$$Sa(x) = a(x)(1 - G(x)). \tag{8.6}$$

Formula (8.4) is a form of the spectral decomposition of the positive, self-adjoint operator $\dot{\ell}_1^* \dot{\ell}_1$ [cf. Reed and Simon (1980), page 221]. [In the finite-dimensional analogue, R would be the transformation to an orthonormal base of eigenvectors and the multiplication (8.6) would be coordinatewise multiplication with the vector of eigenvalues.] The key relations (8.4)–(8.6) were all communicated to us by Wellner and were obtained by martingale calculations along the lines of Ritov and Wellner (1988). We do not prove (8.4)–(8.6) here, but refer to Bickel, Klaassen, Ritov and Wellner (1990).

Since $a1_{[0,\tau_G-\varepsilon]} \in R(S)$ for every $a \in L_2(F)$ and $\varepsilon > 0$, we have $\overline{R(S)} = \{a \in L_2(F)\colon a(x) = 0$ for F-almost all $x > \tau_G\}$. However, $R(S)$ is typically not closed. Closedness would imply that every $a \in L_2(F)$ satisfies $\int_{[0,\tau_G]} a^2/(1-G)\,dF < \infty$, which may be true for a particular F and G, but fails, for instance, if F has a density which is bounded away from zero near τ_G. Since R is an isometry, $R(\dot{\ell}_1^*\dot{\ell}_1)$ [and hence, by Lemma 3.4, $R(\dot{\ell}_1)$] is closed if and only if $R(S)$ is closed. Hence, in the important case that $\tau_G < \tau_F$ one often has that $R(\dot{\ell}_1)$ is not closed.

The polar decomposition of $\dot{\ell}_1^*$ (see Lemma A.3 in the appendix) yields

$$R\left(\dot{\ell}_1^*\right) = R\left(\left(\dot{\ell}_1^*\dot{\ell}_1\right)^{1/2}\right) = R\left(R^{-1}S^{1/2}R\right), \tag{8.7}$$

where $S^{1/2}a = a(1-G)^{1/2}$.

Corollary 8.1 and (8.7) show that a real-valued functional $\chi\colon \mathscr{F} \to \mathbb{R}$ is differentiable at P_{FG} if and only if its influence function satisfies $\tilde{\chi}_F = R^{-1}S^{1/2}Ra$ for some $a \in L_{2*}(F)$. Since R^{-1} is an isometry, this reduces to

$$\int \frac{(R\tilde{\chi}F)^2}{1-G}\,dF < \infty. \tag{8.8}$$

For instance, for the mean of F to be regularly estimable it is necessary that

$$\int \left(x - \frac{\int_{(x,\infty)} s\,dF(s)}{1-F(x)}\right)^2 \frac{dF(x)}{1-G(x)} < \infty.$$

The spectral decomposition greatly simplifies obtaining the lower bound $\|\tilde{\kappa}_{P_{FG}}\|^2_{P_{FG}}$ for the asymptotic variance. By Lemma A.3, this equals $\|\beta\|_F^2$, where $\beta \in N(\dot{\ell}_1)^\perp$ solves $(\dot{\ell}_1^*\dot{\ell}_1)^{1/2}\beta = \tilde{\chi}_F = \dot{\ell}_1^*\tilde{\kappa}_{P_{FG}}$. By the spectral decomposition,

$$(R\beta)1_{[0,\tau_G]} = \frac{R\tilde{\chi}_F}{(1-G)^{1/2}}1_{[0,\tau_G]}.$$

Next, by (8.3), β must be constant on $[\tau_G,\infty)$, so that $(R\beta)1_{[\tau_G,\infty)} \equiv 0$ by (8.5). Hence,

$$\|\tilde{\kappa}_{P_{FG}}\|^2_{P_{FG}} = \|\beta\|_F^2 = \|R\beta\|_F^2 = \int_{[0,\tau_G]} \frac{(R\tilde{\chi}_F)^2}{1-G}\,dF.$$

Perhaps the most interesting functional is the distribution function F. It is well known that the product-limit estimator is a regular estimator for the restriction of F to $[0,\tau]$, seen as an element of the space $\mathbf{D}[0,\tau]$, if $\tau < \tau_G$. Hence, by Theorem 2.1, this functional is differentiable. Let us compute the influence function of $P_{FG} \to F(u)$ where $u < \tau_G$. The gradient of $F \to F(u)$ is $1_{[0,u]} - F(u)$. By (8.5),

$$R\left(1_{[0,u]} - F(u)\right) = 1_{[0,u]}\frac{1 - F(u)}{1 - F}.$$

Under the condition that $u < \tau_G$, this function is contained in the range of S. By (8.4),

$$\dot{\ell}_1^* \dot{\ell}_1 R^{-1} 1_{[0,u]}\frac{1 - F(u)}{(1 - F)(1 - G)} = 1_{[0,u]} - F(u),$$

so that by (3.10) and (8.5) an influence function of $P_{FG} \to F(u)$ is given by

$$\dot{\ell}_1\left(1_{[0,u]}\frac{1 - F(u)}{(1 - F)(1 - G)} - \int_{[0,\cdot]} 1_{[0,u]}(1 - F(u))\frac{dF}{(1 - F)^2(1 - G)}\right).$$

9. Example: Random truncation. Let U and V be independent positive random variables with unknown continuous distribution functions F and G, respectively. Let the observations be an i.i.d. sample from $P_{FG} = \mathscr{L}_{FG}((U, V) | V < U)$. We show that the functional $P_{FG} \to F(u)$ (where $0 < u < \infty$) is regularly estimable only if

$$\int_0^\infty \frac{dF}{G} < \infty. \tag{9.1}$$

Moreover, we address differentiability of $P_{FG} \to F$.

This model is treated by Woodroofe (1985) and Keiding and Gill (1990), who show that, given identifiability, under (9.1) the maximum likelihood estimator yields an asymptotically normal estimator for $F(u)$. Furthermore, Woodroofe establishes that (9.1) is necessary for the asymptotic variance of the maximum likelihood estimator of $F(t)$ to remain finite as $t \to 0$, and that the maximum likelihood estimator yields an asymptotically normal estimator for the quotient $F(u)/F(t)$, $0 < t < u < \infty$, under only an identifiability condition. This leads to the following interpretation of (9.1). If it fails, then the mass that G puts near zero is small relative to F, so that most of the smaller U's will be taken away by the truncation mechanism. This precludes the possibility of establishing the amount of mass F puts near zero, and hence makes estimation of $F(u)$ hard for any u, though accurate estimation of the conditional distribution $F(u)/F(t)$, $u > 0$, is still possible, even without (9.1).

It is straightforward to show for this model that (8.1) implies (8.2) with

$$\begin{aligned} A(a,b)(x,y) &= a(x) - \int a\,dF^* + b(y) - \int b\,dG^* \\ &= \dot{\ell}_1 a(x) + \dot{\ell}_2 b(y), \qquad (a,b) \in L_{2*}(F) \times L_{2*}(G), \end{aligned}$$

where F^* and G^* are the marginals of P_{FG}:

$$dF^* = \alpha^{-1} G\,dF \quad \text{and} \quad dG^* = \alpha^{-1}(1 - F)\,dG, \quad \text{where} \quad \alpha = \int G\,dF.$$

Then $A^*g = (\dot{\ell}_1^* g, \dot{\ell}_2^* g)$. Now view $L_2(F^*)$ as a subspace of $L_2(P_{FG})$ through $g(x,y) = a(x)$, $a \in L_2(F^*)$. Since $R(\dot{\ell}_1) \subset L_{2*}(F^*)$, we have $\dot{\ell}^*_{1|L_{2*}(F^*)^\perp} \equiv 0$, while by direct calculation,

$$\dot{\ell}_1^* a(x) = \alpha^{-1} a(x) G(x), \qquad a \in L_{2*}(F^*).$$

For the existence of a regular estimator of $F(u)$ it is necessary that $1_{[0,u]} - F(u) = \dot{\ell}_1 a$ for some $a \in L_{2*}(F^*)$. Thus,

$$\int \left(\frac{1_{[0,u]} - F(u)}{G} \right)^2 G\, dF < \infty.$$

This is equivalent to (9.1).

The argument can be reversed to show that (9.1) is also sufficient for differentiability of $P_{FG} \to F(u)$ with respect to the set of paths $t \to P_{F_t G}$, where G is fixed and $t \to F_t$ satisfies (8.1). To have differentiability of this functional with respect to the set of all paths $t \to P_{F_t G_t}$ generated by (8.1) and (8.2), more seems to be needed. We now show that (9.1) together with its dual,

$$\text{(9.2)} \qquad \int_0^\infty \frac{dG}{1-F} < \infty,$$

is sufficient for differentiability of $\kappa(P_{FG}) = F \in (\mathbf{D}[0,\infty], \|\cdot\|_\infty)$.

It is easily seen that (8.1) implies

$$t^{-1}\big(\kappa(P_{F_t G_t}) - \kappa(P_{FG})\big) \to \int_{[0,\cdot]} a\, dF, \quad \text{in } (\mathbf{D}[0,\infty], \|\cdot\|_\infty).$$

Thus, it suffices to show that the derivative $\kappa'_{P_{FG}}$: $T(P_{FG}) \to (\mathbf{D}[0,\infty], \|\cdot\|_\infty)$ given by $A(a,b) \to \int_{[0,\cdot]} a\, dF$ is continuous. By the Cauchy–Schwarz inequality,

$$\left\| \kappa'_{P_{FG}}(A(a,b)) \right\|_\infty^2 \le \sup_{0 \le t \le \infty} \int_{[0,t]} a^2 G\, dF \int_{[0,t]} \frac{dF}{G}$$
$$\le \alpha \|a\|^2_{P_{FG}} \int_{[0,\infty]} \frac{dF}{G}.$$

To complete the proof, we show the existence of a constant such that

$$\|a\|^2_{P_{FG}} \le (\text{constant}) \|A(a,b)\|^2_{P_{FG}}.$$

As explained to us by Jon Wellner, (9.1) and (9.2) imply that the sumspace $\mathbf{L} := \{a(x) + b(y)\colon a \in L_{2*}(F^*),\ b \in L_{2*}(G^*)\}$ is closed in $L_2(P_{FG})$. This follows from Proposition 2 in Appendix 4 of Bickel, Klaassen, Ritov and Wellner (1990) and

$$\iint \frac{(dP_{FG}(x,y))^2}{dF^*(x)\, dG^*(y)} = \iint_{y<x} \frac{dF(x)^2\, dG(y)^2}{G(x)\, dF(x)(1-F(y))\, dG(y)}$$
$$\le \int \frac{dF}{G} \int \frac{dG}{1-F}.$$

Hence, the map T: $L_{2*}(F^*) \times L_{2*}(G^*) \to \mathbf{L}$ given by $T(a,b) = A(a,b)$ is a one-to-one, linear and continuous map from a Banach space onto a Banach space. (Note: $T \neq A$ in the sense that its domain space is different.) By the bounded inverse theorem [Rudin (1973), 2.12], it has a continuous inverse. Thus,

$$\|a\|_{P_{FG}} \vee \|b\|_{P_{FG}} = \|(a,b)\|_{L_{2*}(F^*)\times L_{2*}(G^*)} \leq \|T^{-1}\| \, \|A(a,b)\|_{P_{FG}}.$$

This concludes the proof. □

It can be shown that, under (9.1) and (9.2), the maximum likelihood estimator is efficient for the estimation of F seen as an element of $(\mathbf{D}[0,\infty], \|\cdot\|_\infty)$ [van der Vaart (1990)].

10. Example: Incomplete censored observations. Let Y and C be independent positive random variables with unknown absolutely continuous cumulative distribution functions F and G, respectively. Let the model $\mathscr{P}$ consist of the set of distributions P_{FG} of $(C, 1_{Y\leq C})$. Thus, P_{FG} has density

$$\delta g(x)F(x) + (1-\delta)g(x)(1-F(x)), \qquad x \in (0,\infty),\ \delta \in \{0,1\},$$

with respect to the product of Lebesgue and counting measure. Since the observation is a measurable transformation of (Y, C), we obtain as in Sections 7 and 8 the implication (8.1) ⇒ (8.2), where, presently, $A(a,b) = \dot{\ell}_1 a + \dot{\ell}_2 b$, with

$$\begin{aligned}\dot{\ell}_1 a(x,\delta) &= E_{FG}\big(a(Y)\,|\,C = x, 1_{Y\leq C} = \delta\big)\\ &= \delta\frac{\int_0^x a\,dF}{F(x)} + (1-\delta)\frac{\int_x^\infty a\,dF}{1-F(x)},\\ \dot{\ell}_2 b(x,\delta) &= E_{FG}\big(b(C)\,|\,C = x, 1_{Y\leq C} = \delta\big) = b(x).\end{aligned}$$

Using Fubini's theorem, we obtain that $R(\dot{\ell}_1) \perp R(\dot{\ell}_2)$. Moreover, on $L_{2*}(P_{FG})$,

$$\dot{\ell}_1^* h(y) = \int_y^\infty h(x,1)\,dG(x) + \int_0^y h(x,0)\,dG(x), \tag{10.1}$$

while $\dot{\ell}_1^*(\text{constant}) = 0$.

When h is an element of $L_2(P_{FG})$, then

$$\int h^2(x,1)F(x)\,dG(x) < \infty \quad\text{and}\quad \int h^2(x,0)(1-F(x))\,dG(x) < \infty.$$

Let $[a,b]$ be an interval on which both F and $1-F$ are bounded away from zero. Then both $h(\cdot,0)$ and $h(\cdot,1)$ are elements of $L_2(G)$, so that $\dot{\ell}_1^* h$ is an absolutely continuous function on $[a,b]$ for every h. Thus, real functionals $F \to \chi(F)$ which have a gradient which is discontinuous somewhere in the interval (a,b) (more precisely, a gradient which is not a.e. equal to a continuous function) are not differentiable functionals of P_{FG}. An example is the functional $F \to F(c)$, where $c \in (a, b)$. Groeneboom (1987) shows that the maximum likelihood estimator for this functional has (the best possible) rate $n^{1/3}$.

However, also in this model there are many interesting functionals which are differentiable. Let us consider the mean of F. This has gradient $y - \int y\, dF$. For the mean of F to be regularly estimable, it is necessary that

$$y = \int_y^\infty h(x,1)\, dG(x) + \int_0^y h(x,0)\, dG(x),$$

for some $h \in L_2(P_{FG})$. This implies $(h(y,0) - h(y,1))g(y) = 1$ so that

$$(10.2)\qquad \int_0^{(1/2)\tau_G} \frac{F}{g} = \int_0^{(1/2)\tau_G} (h(y,0) - h(y,1))^2 F(y)\, dG(y) < \infty,$$

$$(10.3)\qquad \int_{(1/2)\tau_G}^\infty \frac{1-F}{g} = \int_{(1/2)\tau_G}^\infty (h(y,0) - h(y,1))^2 (1 - F(y))\, dG(y) < \infty.$$

Under these conditions we can set $h(y,0) = g^{-1}(y)1_{[(1/2)\tau_G,\infty)}(y) + \frac{1}{2}\tau_G$ and $h(y,1) = -g^{-1}(y)1_{(0,(1/2)\tau_G]}(y) + \frac{1}{2}\tau_G$.

To obtain the efficient influence function for the mean, it suffices to project h onto $\overline{R(\dot{\ell}_1)}$. From (10.1) we immediately see that

$$N(\dot{\ell}_1^*) = \{h \in L_2(P_{FG}) \colon h(x,0) = h(x,1), G\text{-a.e.}\}.$$

But this is precisely $R(\dot{\ell}_2)$, so that $R(\dot{\ell}_1)^\perp = R(\dot{\ell}_2)$. Since the two ranges are orthogonal, the sum of the orthogonal projections onto $\overline{R(\dot{\ell}_1)}$ and $R(\dot{\ell}_2)$ equals the identity in $L_{2*}(P_{FG})$. This is helpful, since the projection onto $R(\dot{\ell}_2)$ is conditional expectation with respect to C. Thus, we calculate

$$\begin{aligned} h(x,\delta) &- E_{FG}\big(h(C, 1_{Y \le C}) \,|\, C = x\big) \\ &= (h(x,1) - h(x,0))[\delta(1 - F(x)) - (1-\delta)F(x)]. \end{aligned}$$

For the h given previously this reduces to the influence function of the mean functional $P_{FG} \to \int y\, dF(y)$,

$$-\delta \frac{1 - F(x)}{g(x)} + (1-\delta)\frac{F(x)}{g(x)}.$$

We have not investigated in any detail whether (10.2) and (10.3) are sufficient for the existence of a regular estimator sequence, though it is easy to convince oneself of the existence of $\sqrt{n}$-consistent estimators under (for instance) the condition that the density of G is bounded and bounded away from zero on the compact support of F. It would be particularly interesting to obtain results on the mean of the maximum likelihood estimator.

APPENDIX

Proofs. This Appendix contains the proofs of Theorems 2.1 and 4.1. We start with a lemma that prepares for the Proof of Theorem 2.1 and actually gives some additional information: (2.6) implies that the map $t \to \kappa(P_t)$ is Lipschitz at $t = 0$.

LEMMA A.1. *Suppose that* $\mathbf{B} = \mathbb{R}$ *and that* (2.6) *holds. Then* $\lim_{t \downarrow 0} t^{-1}(\kappa(P_t) - \kappa(P))$ *exists if and only if*

$$\left(\sqrt{n}\,(T_n - \kappa(P)), n^{-1/2} \sum_{j=1}^{n} g(X_j)\right)$$

converges under P weakly to the law of some random vector (S, V). *In this case*

$$\lim_{t \downarrow 0} t^{-1}(\kappa(P_t) - \kappa(P)) = \frac{EVe^{ius}}{iuE\,e^{iuS}}, \tag{A.1}$$

for every u for which the right-hand side is defined. If $\mathbf{B} = \mathbb{R}$ *and* (2.6) *holds for a probability measure L with* $\int x^2\, dL(x) < \infty$, *then* $t^{-1}(\kappa(P_t) - \kappa(P)) = O(1)$ *as* $t \downarrow 0$. *If, moreover, the limit exists then it also equals* $E(SV)$.

PROOF. Fix $h > 0$. We first prove the second part of the lemma. Let $t_m \downarrow 0$ be arbitrary. Define a subsequence of $\{n\}$ by $(n_m + 1)^{-1/2} < t_m h \le n_m^{-1/2}$ and set $h_{n_m} = t_m h n_m^{1/2}$. Then

$$t_m^{-1}\left(\kappa(P_{t_m h}) - \kappa(P)\right) = (1 + o(1))\, n_m^{1/2}\left(\kappa\left(P_{h_{n_m}/\sqrt{n_m}}\right) - \kappa(P)\right).$$

There is a further subsequence of $\{n\}$ (abusing notation, denoted $\{n\}$) such that

$$\left(\sqrt{n}\,(T_n - \kappa(P)), n^{-1/2} \sum_{j=1}^{n} g(X_j)\right) \Rightarrow_P (S, V). \tag{A.2}$$

Here $\mathscr{L}(S) = L$ and $\mathscr{L}(V) = N(0, I)$. Let Λ_n be the log-likelihood ratio of the product measures corresponding to $P_{h_n/\sqrt{n}}$ and P. Then by the local asymptotic normality lemma and (A.2)

$$\left(\sqrt{n}\,(T_n - \kappa(P)), \Lambda_n\right) \Rightarrow_P \left(S, hV - \tfrac{1}{2}h^2 I\right). \tag{A.3}$$

By contiguity arguments we conclude that $\{\sqrt{n}\,(T_n - \kappa(P))\}$ is asymptotically tight under $P_{h_n/\sqrt{n}}$. But by (2.6), $\{\sqrt{n}\,(T_n - \kappa(P_{h_n/\sqrt{n}}))\}$ is asymptotically tight under $P_{h_n/\sqrt{n}}$ too. Hence, $\{\sqrt{n}\,(\kappa(P_{h_n/\sqrt{n}}) - \kappa(P))\}$ is asymptotically tight. Choose a further subsequence which converges to a limit (say) $a(h)$. Now, by (A.3) and a version of Le Cam's third lemma,

$$\sqrt{n}\left(T_n - \kappa\left(P_{h_n/\sqrt{n}}\right)\right) \Rightarrow_{P_{h_n/\sqrt{n}}} L_h,$$

where

$$L_h(B) = \int_{\mathbb{R}} \int_B e^{\lambda}\, d\mathscr{L}\left(S - a(h), hV - \tfrac{1}{2}h^2 I\right)(y, \lambda). \tag{A.4}$$

By (2.6), $L_h = L = \mathscr{L}(S)$. Comparing expectations we see $ES = E(S - a(h))\exp[hV - \frac{1}{2}h^2 I]$ or

$$a(h) = ES\left(\exp\left[hV - \tfrac{1}{2}h^2 I\right] - 1\right).$$

Thus, $|a(h)| \le \{ES^2 E(\exp[hV - \frac{1}{2}h^2 I] - 1)^2\}^{1/2}$. The latter expression is independent of any of the subsequences we have chosen.

Thus, we obtain that the limit points of $t^{-1}(\kappa(P_{th}) - \kappa(P))$ are contained in a compact interval, concluding the proof that $t^{-1}(\kappa(P_{th}) - \kappa(P)) = O(1)$.

Next, drop the assumption that $ES^2 < \infty$. Equation (A.4) is still valid. Comparing characteristic functions rather than expectations, we obtain

$$e^{iua(h)} E e^{iuS} = E \exp\left[iuS + hV - \tfrac{1}{2}h^2 I\right]. \tag{A.5}$$

Thus $a(h)$ depends on the joint law of S and V only. If we have joint convergence along the whole sequence of $\{n\}$ in (A.2), then we may conclude that every sequence $t_m^{-1}(\kappa(P_{t_m h}) - \kappa(P))$ with $t_m \downarrow 0$ has a subsequence converging to the fixed limit $a(h)$. Hence,

$$\lim_{t \downarrow 0} t^{-1}(\kappa(P_{th}) - \kappa(P)) = a(h).$$

Then,

$$a(h) = h \lim_{t \downarrow 0} (th)^{-1}(\kappa(P_{th}) - \kappa(P)) = ha(1). \tag{A.6}$$

Insert this in (A.5) and differentiate with respect to h at $h = 0$ to get (A.1). The last assertion of the lemma can be obtained in the same manner from the representation of $a(h)$ in terms of expectations.

Finally, we show that [under (2.6)] (A.2) holds along the whole sequence $\{n\}$ if

$$a(1) := \lim_{t \downarrow 0} t^{-1}(\kappa(P_t) - \kappa(P))$$

exists. In fact, this follows from the convolution theorem. For completeness we outline the proof. Take an arbitrary subsequence of $\{n\}$. There is a further subsequence such that (A.2) holds along the subsequence. We must show that $\mathscr{L}(S, V)$ is the same for each subsequence. By (2.6), (A.4) and (A.6),

$$\int e^{itx}\, dL(x) = E \exp\left[it(S - ha(1)) + hV - \tfrac{1}{2}h^2 I\right].$$

Choose $h = -ita(1)/I$ to get

$$Ee^{itS} = E \exp\left[it(S - a(1)VI^{-1})\right] \exp\left[-\tfrac{1}{2}t^2 a(1)^2 I^{-1}\right]. \tag{A.7}$$

Choose $h = -ita(1)/I + iu$ to get

$$\begin{aligned} E e^{itS} &= E \exp\left[it(S - a(1)VI^{-1}) + iuV\right] \\ &\quad \times \exp\left[-\tfrac{1}{2}t^2 a(1)^2 I^{-1}\right] \exp\left[\tfrac{1}{2}u^2 I\right]. \end{aligned} \tag{A.8}$$

Infer that $S - a(1)V/I$ and V are independent. Moreover, $\mathscr{L}(S - a(1)V/I)$ is completely determined by $\mathscr{L}(S) = L$ and (A.7), while V has a fixed normal distribution. Thus $\mathscr{L}(S - a(1)V/I, V)$ and hence $\mathscr{L}(S, V)$ is completely determined. □

PROOF OF THEOREM 2.1. (i) First consider the case that $\mathbf{B} = \mathbb{R}$. Given $g \in T(P)$, define

$$\kappa_P'(g) = \lim_{t \downarrow 0} t^{-1}(\kappa(P_t) - \kappa(P)),$$

where $t \to P_t$ satisfies (2.1). This is well defined. In fact, by the preceding lemma, if (S, V_g) is a weak limit in law under P of $(\sqrt{n}(T_n - \kappa(P)), n^{-1/2}\sum_{j=1}^n g(X_j))$, then

$$\kappa_P'(g) = \frac{EV_g\, e^{iuS}}{iuE\, e^{iuS}},$$

for any u for which the right-hand side is well defined. We have to show that κ_P' is linear and continuous. Let $g_0, g_1, \dots$ be a sequence in $T(P)$. Let $\mathscr{L}(S, V_0, V_1, \dots)$ be a weak limit point under P in $\mathbb{R} \times \mathbb{R}^\infty$ of

$$\left(\sqrt{n}(T_n - \kappa(P)), n^{-1/2} \sum_{j=1}^n g_0(X_j), n^{-1/2} \sum_{j=1}^n g_1(X_j), \dots\right).$$

Then, given $\alpha \in \mathbb{R}^2$, it is not hard to see from this representation applied to $g = \alpha_1 g_1 + \alpha_2 g_2$ and $g = g_i$ that

$$\kappa_P'(\alpha_1 g_1 + \alpha_2 g_2) = \alpha_1 \kappa_P'(g_1) + \alpha_2 \kappa_P'(g_2).$$

Furthermore, if $g_i \to g_0$ in $\|\cdot\|_P$ as $i \to \infty$, then $E(V_i - V_0)^2 = E_P(g_i(X_1) - g_0(X_1))^2 \to 0$, and again applying the representation we see that

$$|\kappa_P'(g_i) - \kappa_P'(g_0)|^2 \to 0.$$

Thus κ_P' is continuous and linear.

(ii) Now consider the case of a general $\mathbf{B}$. By (2.6) and tightness of L we have asymptotical tightness of the sequence $\{\sqrt{n}(T_n - \kappa(P))\}$ under P. Let $h_n \to h$. By contiguity arguments we can infer that $\{\sqrt{n}(T_n - \kappa(P))\}$ is relatively compact under $P_{h_n/\sqrt{n}}$ with tight limiting points [see, for instance, van der Vaart (1988), Lemma 4.6]. Combination with (2.6) shows that $\{\sqrt{n}(\kappa(P_{h_n/\sqrt{n}}) - \kappa(P))\}$ is relatively compact.

For any $b^* \in \mathbf{B}^*$, $b^* \circ T_n$ satisfies the conditions of Theorem 2.1 as an estimator for the real-valued functional $b^* \circ \kappa$. Thus, by part (i) of this proof we know that the functional $b^* \circ \kappa$ is differentiable. In particular, the limits

$$\lim_{t \downarrow 0} b^* \circ (\kappa(P_t) - \kappa(P))$$

exist. But this uniquely identifies the limit points of $\{\sqrt{n}(\kappa(P_{h_n/\sqrt{n}}) - \kappa(P))\}$. Hence, the latter sequence must be convergent to a limit $\kappa_P'(g)$. We only have to show that this defines a continuous, linear map from $T(P)$ in $\mathbf{B}$. Now, by (i), $b^* \circ \kappa_P'$ is a continuous, linear map for every b^* as above, because it is the derivative of the differentiable functional $b^* \circ \kappa$. The following well-known lemma, applied with $X = T(P)$, $Y = \mathbf{B}$ and $Y' = \mathbf{B}^*$ shows that this is enough to infer continuity and linearity of κ'_P. □

LEMMA A.2. *Let ϕ be a map from a normed space X into a normed space Y, such that $y' \circ \phi \in X^*$ for every y' in a closed subspace Y' of Y^* satisfying $\|y\| = \sup\{y'(y): \|y'\| \le 1\}$ for every $y \in Y$. Then ϕ is continuous and linear.*

PROOF. By the condition Y' separates points of Y. Linearity of ϕ follows easily.

Define a map $k: Y \to (Y')^*$ by

$$k(y)(y') = y'(y), \qquad y' \in Y'.$$

By assumption, this is an isometric embedding of Y in $(Y')^*$. Thus, a subset S of Y is bounded if and only if its image in $(Y')^*$ is bounded. By the Banach–Steinhaus theorem [Rudin (1973), 2.6], this is true if and only if $\{k(y)(y'): y \in S\}$ is bounded for every $y' \in Y'$ separately, i.e., every $y' \in Y'$ is bounded over S. Now, let U be the unit ball in X. Since ϕ is linear, it is continuous if and only if $\phi(U)$ is bounded in Y. By the above argument, this is true if and only if $y' \circ \phi(U)$ is bounded in $\mathbb{R}$ for every $y' \in Y'$, i.e., $y' \circ \phi$ is continuous for every $y' \in Y'$. □

As preparation for the Proof of Theorem 4.1, we have the following lemma.

LEMMA A.3. *Let A be a continuous, linear map from a Hilbert space $\mathbf{H}$ in a Hilbert space $\mathbf{L}$. Then there exists a unique self-adjoint positive definite operator $(A^*A)^{1/2}: \mathbf{H} \to \mathbf{H}$ such that $A^*A = (A^*A)^{1/2}(A^*A)^{1/2}$ (the square-root of A^*A). Moreover $\overline{R(A^*)} = \overline{R((A^*A)^{1/2})}$. Finally, if $g \in \overline{R(A)}$ and $\beta \in \overline{R(A^*)}$ satisfy $A^*g = (A^*A)^{1/2}\beta$, then $\|g\|^2_{\mathbf{L}} = \|\beta\|^2_{\mathbf{H}}$.*

PROOF. The first part of the lemma is a standard result from functional analysis [cf. Rudin (1973), 12.33]. Next we obtain a polar decomposition of A similarly as in Rudin [(1973), pages 315 and 316]. First, note that

$$\|A\beta\|^2_{\mathbf{L}} = \langle A\beta, A\beta\rangle_{\mathbf{L}} = \langle A^*A\beta, \beta\rangle_{\mathbf{H}} = \left\langle (A^*A)^{1/2}\beta, (A^*A)^{1/2}\beta\right\rangle_{\mathbf{H}}$$
$$= \left\|(A^*A)^{1/2}\beta\right\|^2_{\mathbf{H}}.$$

Thus, it is possible to define an isometry U of $R((A^*A)^{1/2})$ onto $R(A)$ by $A\beta = U(A^*A)^{1/2}\beta$. U can first be extended to an isometry of $\overline{R\left((A^*A)^{1/2}\right)}$ onto $\overline{R(A)}$ and subsequently to a "partial isometry" on $\mathbf{H}$ by setting it equal to zero on the orthocomplement of $\overline{R((A^*A)^{1/2})}$ and extending linearly.

Now $A^* = (A^*A)^{1/2}U^*$. Here U^* acts as the inverse of U on $\overline{R(A)}$, mapping this set onto $\overline{R\left((A^*A)^{1/2}\right)}$, and maps the orthocomplement of $R(A)$ into zero. Next, note that $(A^*A)^{1/2}$ is zero on $R((A^*A)^{1/2})^{\perp} = N((A^*A)^{1/2})$, so that

$$A^*\mathbf{L} = (A^*A)^{1/2}U^*\mathbf{L} = (A^*A)^{1/2}\,\overline{R\left((A^*A)^{1/2}\right)} = (A^*A)^{1/2}\mathbf{H}.$$

Finally, if $A^*g = (A^*A)^{1/2}\beta$, then $U^*g - \beta \in N((A^*A)^{1/2}) = R((A^*A)^{1/2})^{\perp}$. If, moreover, g and β satisfy the stated conditions, we also have $U^*g - \beta \in \overline{R((A^*A)^{1/2})}$. Thus $U^*g = \beta$. The result follows from the fact that U is an isometry on $\overline{R(A)}$. □

PROOF OF THEOREM 4.1. Suppose (3.6) fails. By Lemma A.3 we can assume that $\tilde{\psi}_\lambda \notin R((A^*A)^{1/2})$. Our proof uses this in combination with a characterization of $R((A^*A)^{1/2})$ in terms of the spectral resolution $u \to P_u$ of $A^*A = \int u\, dP_u$. By definition [see, e.g., Rudin (1974), pages 300–310 and 341–355, or Reed and Simon (1980), VII], this defines for every $\alpha \in T(\lambda, \Lambda)$ a finite Borel measure $B \to \mu_\alpha(B) = \int_B d\langle P_u\alpha, \alpha\rangle_{\mathbf{H}}$ on $[0, \|A^*A\|]$, called the spectral measure of α. Now,

$$R\big((A^*A)^{1/2}\big) = \left\{\alpha \in T(\lambda, \Lambda)\colon \int u^{-1}\, d\mu_\alpha(u) < \infty\right\}. \tag{A.8}$$

To see this, note first that $\alpha = (A^*A)^{1/2}\beta = \int u^{1/2}\, dP_u(\beta)$ implies $d\mu_\alpha(u) = u\, d\mu_\beta(u)$ [cf. Rudin (1973), 13.23]. Hence we have inclusion of the left-hand side in the right-hand side.

Next assume that $\int u^{-1}\, d\mu_\alpha(u) < \infty$. Then α is in the domain of the (possibly unbounded) operator $\int u^{-1/2}\, dP_u$ and we can define $\beta = \int u^{-1/2}\, dP_u(\alpha)$ [Rudin (1973), 13.23]. The proof of (A.6) will be complete if it is shown that $\alpha = (A^*A)^{1/2}\beta$. Since

$$\int \left(u^{-1/2} - u^{-1/2}1_{u \ge n^{-1}}\right)^2 d\mu_\alpha(u) \to 0,$$

as $n \to \infty$, we have that $\beta_n := \int u^{-1/2}1_{u \ge n^{-1}}\, dP_u(\alpha) \to \beta$ in $T(\lambda, \Lambda)$. Then by continuity $(A^*A)^{1/2}\beta_n \to (A^*A)^{1/2}\beta$, while on the other hand, by the "symbolic" calculus summarized on pages 309 and 310 of Rudin (1973),

$$\begin{aligned}(A^*A)^{1/2}\beta_n &= \int u^{1/2}\, dP_u \int u^{-1/2}1_{u \ge n^{-1}}\, dP_u(\alpha)\\ &= \int 1_{u \ge n^{-1}}\, dP_u(\alpha) \to \int dP_u(\alpha),\end{aligned}$$

since $\int (1_{u \ge n^{-1}} - 1)^2\, d\mu_\alpha(u) \to 0$. Thus, $(A^*A)^{1/2}\beta = \int dP_u(\alpha) = \alpha$.

Thus, if $\tilde{\psi}_\lambda \notin R((A^*A)^{1/2})$, then $\mu_{\tilde{\psi}_\lambda}(\{0\}) > 0$, or $\int u^{-1}1_{u > 0}\, d\mu_{\tilde{\psi}_\lambda}(u) = \infty$, or both.

In the first case, choose $\alpha = P_{\{0\}}(\tilde{\psi}_\lambda)$. Then, again by the symbolic calculus,

$$A^*A\alpha = \int u\, dP_u\, P_{\{0\}}(\tilde{\psi}_\lambda) = \int u 1_{\{0\}}\, dP_u(\tilde{\psi}_\lambda) = 0(0) = 0.$$

However, $\langle \tilde{\psi}_\lambda, \alpha\rangle_{\mathbf{H}}^2 = \int 1_{\{0\}}\, d\mu_{\tilde{\psi}_\lambda}(u) = \mu_{\tilde{\psi}_\lambda}(\{0\}) > 0$. But then $i_\alpha = 0$.

In the second case, set $\alpha_n = \int u^{-1}1_{u \ge n^{-1}}\, dP_u(\tilde{\psi}_\lambda)$. Then, $A^*A\alpha_n = \int 1_{u \ge n^{-1}}\, dP_u(\tilde{\psi}_\lambda)$ and

$$\|A\alpha_n\|_{P_\lambda}^2 = \langle A^*A\alpha_n, \alpha_n\rangle_{\mathbf{H}} = \int u^{-1}1_{u \ge n^{-1}}\, d\mu_{\tilde{\psi}_\lambda}(u).$$

Thus,

$$i_{\alpha_n}^{-1} = \frac{\langle \tilde{\psi}_\lambda, \alpha_n \rangle_{\mathbf{H}}^2}{\int u^{-1} 1_{u \geq n^{-1}} \, d\mu_{\tilde{\psi}_\lambda}(u)} = \int u^{-1} 1_{u \geq n^{-1}} \, d\mu_{\tilde{\psi}_\lambda}(u) \to \infty. \qquad \square$$

Acknowledgments. I thank Jon Wellner, Peter Bickel and Peter Sasieni for a number of helpful remarks.

REFERENCES

AVERBUKH, V. I. and SMOLYANOV, O. G. (1967). The theory of differentiation in linear topological spaces. *Russian Math. Surveys* **22** 201–258.

BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

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

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1990). Efficient and adaptive inference in semi-parametric models. Unpublished manuscript.

CARROLL, R. J. and HALL, P. (1988). Optimal rates for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

DUDLEY, R. M. (1966). Weak convergence of measures on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

DUDLEY, R. M. (1967). Measures on nonseparable metric spaces. *Illinois J. Math.* **11** 449–453.

FAN, J. (1988). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* To appear.

GAENSSLER, P. (1983). *Empirical Processes.* IMS, Hayward, Calif.

GILL, R. D. (1986). Non- and semi-parametric maximum likelihood estimators and the von Mises method. CWI Report MS-R8604, Centrum Wisk. Inform., Amsterdam.

GROENEBOOM, P. (1987). Asymptotics for incomplete censored observations. Unpublished manuscript.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimators. *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. California Press, Berkeley.

HOFFMAN-JØRGENSEN, J. (1984). Unpublished manuscript.

KEIDING, N. and GILL, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.* **18** 582–602.

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

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press, Berkeley.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

LE CAM, L. and YANG, G. L. (1988). On the preservation of local asymptotic normality under information loss. *Ann. Statist.* **16** 483–520.

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

MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. *Ecole d'Eté de Probabilités de Saint-Flour XI. Lecture Notes in Math.* **976** 76–265. Springer, New York.

MILLAR, P. W. (1985). Nonparametric applications of an infinite dimensional convolution theorem. *Z. Wahrsch. Verw. Gebiete* **68** 545–556.

PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statist.* **13**. Springer, New York.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

REED, M. and SIMON, B. (1980). *Functional Analysis*. Academic, Orlando.

RITOV, Y. (1987). On the deconvolution of a mixture of normal distributions. Unpublished manuscript.

RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales and the Cox model. *Contemp. Math.* **80** 191–219.

RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill, New York.

VAN DER VAART, A. W. (1988b). *Statistical Estimation in Large Parameter Spaces. CWI Tract* **44**. Centrum Wisk. Inform., Amsterdam.

VAN DER VAART, A. W. (1990). Efficiency and Hadamard differentiability. *Scand. J. Statist.* To appear.

VAN DER VAART, A. W. and WELLNER, J. A. (1989). Convolution and asymptotic minimax theorems for nonmeasurable estimators. Unpublished manuscript.

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

WOODROOFE, M. (1985). Estimating a distribution function with censored data. *Ann. Statist.* **13** 163–177.

SUBFACULTEIT WISKUNDE EN INFORMATICA
VRIJE UNIVERSITEIT AMSTERDAM
DE BOELELAAN 1081
1081 HV AMSTERDAM
THE NETHERLANDS