The Annals of Statistics
1980, Vol. 8, No. 1, 106–115

ESTIMATES DERIVED FROM ROBUST TESTS[1]

By Helmut Rieder

University of Freiburg and University of California, Berkeley

In this paper, an asymptotic minimax theory for robust estimation of a one-dimensional parameter is derived, which is an asymptotic counterpart, and generalization to an arbitrary parameter, of Huber's finite sample minimax theory for the location case. A particular variability measure and results from robust asymptotic testing are employed. The results show a relationship of this approach to Hampel's local theory of robustness.

0. Introduction. The problem studied in this paper is the estimation of the one-dimensional parameter θ of a parametric family of probability measures $\{P_\theta\}$, when the laws of the N independent observations need not exactly coincide with some member P_θ, but are only known to lie in a neighborhood $\mathcal{P}_\theta$ of P_θ that is defined in terms of ε-contamination and total variation. The variability of an estimate shall be assessed by the maximum probability that the true parameter falls below, or exceeds, an interval of prescribed width, $2\tau_\theta$, that is laid around the estimate. In this generality, only asymptotic approximations, as $N \to \infty$, will be available. So we assume the family $\{P_\theta\}$ to be locally asymptotically normal and let the neighborhoods and the intervals shrink at the rate $N^{-\frac{1}{2}}$. Furthermore, the estimates under consideration are supposed to be regular in a certain sense (Definition 1.1).

Our first aim is the explicit determination of the risk of an estimate (Theorem 2.2), in order to obtain a quantitative expression for the influence of outliers, upon which, for example, a measure of asymptotic relative efficiency may be based. Second, a minimax result is derived (Theorem 3.1). We come out with an (M)-estimate defined by a truncated likelihood function.

The minimax result may be viewed as an asymptotic analog, and generalization to an arbitrary parameter and arbitrary sample space, of Huber's (1968) finite sample minimax result for the location case and unimodal Lebesgue densities. The idea he employs there is to carry out a minimax test between the shifted neighborhoods $\mathcal{P}_{\theta-\tau_\theta}$, $\mathcal{P}_{\theta+\tau_\theta}$, and then to derive an estimate from this test in the manner of Hodges and Lehmann (1963). However, for a general parameter, this method does not work any longer. Nevertheless, as Huber (1972), pages 1060, 1061, conjectures, the (M)-estimate based upon the likelihood ratio of least favorable pairs is still a candidate for a minimax property in the general case. A tentative asymptotic study has been made by Huber-Carol (1970) who, under heavy assumptions and by an

Received January 1978; revised November 1978.

[1]Research performed while holding a scholarship of the Deutsche Forschungsgemeinschaft.

AMS 1970 *subject classifications.* Primary 62G35; secondary 62E20, 62G15.

Key words and phrases. Regular estimates, (M)-estimates, influence curve, ε-contamination, total variation, contiguity.

apparently different technique, shows that the asymptotic confidence level of the interval associated with the above mentioned estimate is bounded away from zero. An explicit expression for this bound is derived there only in special cases, but can be obtained in general as a by-product of our results. Moreover, it should be noted that she neither formulates nor proves any optimality of the estimate to which her study is restricted.

In deriving an asymptotic minimax result some care is needed. First, even in the location case, it would be objectionable merely to translate Huber's finite sample result into asymptotic terms, for his result requires the formation of the supremum over the whole parameter space of the pointwise risk, whereas, as Hajek (1972), page 176, points out, an asymptotic study has to be local. One should then note that even in the location case our asymptotic minimax result is not implied by Huber's finite one. Second, in the noninvariant case, the risk of an estimate can no longer be transformed into the risk of a test between $\mathcal{P}_{\theta-\tau_\theta}$ and $\mathcal{P}_{\theta+\tau_\theta}$.

It is remarkable, however, that for regular estimates the connection with testing shows up at least asymptotically and formally in the expression for the risk. Furthermore, in deriving this expression and in solving the resulting optimization problem, the asymptotic testing results of Rieder (1978) can be used.

A slight modification of these results becomes necessary due to our present restriction to contiguous sequences, which bears some mathematical convenience and may be justified by the fact that an estimate attains, or comes arbitrarily close to, its worst possible behavior even under contiguous sequences (Lemma 2.1).

A relationship of this approach and Hampel's local robustness theory is implied by the fact that the risk of an estimate turns out to depend on both its asymptotic variance and the extreme values of its influence curve at the model $\{P_\theta\}$. Accordingly, the minimax estimate also minimizes the asymptotic variance subject to bounds on the influence curve (Theorem 3.2).

1. The model. Let $(\Omega, \mathcal{B})$ be a measurable space and denote by $\mathfrak{M}$ the set of probability measures on $\mathcal{B}$. Let a family $\{P_\theta\} \subset \mathfrak{M}$ be given, which is parametrized by an open subset Θ of the real line, and parameter functions ε, δ and $\tau : \Theta \to [0, \infty)$ satisfying $\varepsilon_\theta + \delta_\theta > 0$ and $\tau_\theta > 0$ for all $\theta \in \Theta$. The family $\{P_\theta\}$ is assumed to be of the following regular kind.

(1.1) For each $\theta \in \Theta$ there exists a neighborhood $U_\theta \subset \Theta$ of θ, such that $P_\zeta \ll P_\theta$ for all $\zeta \in U_\theta$.

Denote by p_θ the density of P_θ with respect to some σ-finite measure that dominates the family $\{P_\theta\}$.

(1.2) For each $\theta \in \Theta$ there exists a nondegenerate function $\Lambda_\theta \in L^2(dP_\theta)$, such that

$$2\frac{p_\zeta^{\frac{1}{2}} - p_\theta^{\frac{1}{2}}}{(\zeta - \theta)p_\theta^{\frac{1}{2}}} \to \Lambda_\theta \quad \text{in } L^2(dP_\theta) \quad \text{as } \zeta \to \theta.$$

The parameters are subject to the following boundedness condition.

$$\varepsilon_\theta + 2\delta_\theta < 2\tau_\theta \int \Lambda_\theta^+ dP_\theta \quad \text{for all} \quad \theta \in \Theta. \tag{1.3}$$

Let $\mathbb{N}$ denote the set of positive integers, and let the symbol $\otimes$ denote stochastic product. For $N \in \mathbb{N}$, which always tends to ∞, let the Cartesian product Ω^N be endowed with the product σ-field $\mathfrak{B}^N$. Let the real line $\mathbb{R}$ be endowed with its Borel σ-field. Then the following definitions complete the distributional framework.

$$\begin{aligned}
&\varepsilon_{N,\theta} = N^{-\frac{1}{2}}\varepsilon_\theta, \qquad \delta_{N,\theta} = N^{-\frac{1}{2}}\delta_\theta, \qquad \tau_{N,\theta} = N^{-\frac{1}{2}}\tau_\theta \\
&\mathfrak{P}_{N,\theta} = \{Q \in \mathfrak{M} : Q \geqslant (1 - \varepsilon_{N,\theta})P_\theta - \delta_{N,\theta} \text{ on } \mathfrak{B}\} \\
&\mathfrak{P}_{N,\theta}^N = \{\otimes_{i=1}^N Q_{Ni} : Q_{Ni} \in \mathfrak{P}_{N,\theta} \quad \text{for } i = 1, \cdots, N\} \\
&H_\theta = \{(W_{N,\theta}) : W_{N,\theta} \in \mathfrak{P}_{N,\theta}^N \quad \text{for all } N, (W_{N,\theta}) \text{ contiguous to } (P_\theta^N)\},
\end{aligned} \tag{1.4}$$

where $P_\theta^N = \otimes_{i=1}^N P_\theta$. The objects of this study are estimates of the parameter θ that are regular in the sense of the following definition.

Definition 1.1. A sequence (T_N) of measurable functions $T_N : \Omega^N \to \mathbb{R}$ is called a regular estimate, iff for each $\theta \in \Theta$ there exists a function $IC_\theta \in L^2(dP_\theta)$, such that $\int IC_\theta dP_\theta = 0$, $\int IC_\theta \Lambda_\theta dP_\theta = 1$, and

$$N^{\frac{1}{2}}(T_N - \theta) = N^{-\frac{1}{2}}\Sigma_{i=1}^N IC_\theta(x_i) + o_{P_\theta^N}(1). \tag{1.5}$$

For convenience, call IC_θ IC-function of (T_N) at θ, and write $(T_N) = T(IC)$.

Remarks. For Fréchet differentiable von Mises functionals, Huber (1977), pages 10, 23, derives these properties with IC_θ equal to the influence curve at P_θ, thereby summarizing the variety of asymptotic normality proofs, which, under weaker assumptions and in special cases, are nevertheless concerned with the verification of these properties. The condition $\int IC_\theta \Lambda_\theta dP_\theta = 1$ is related to Fisher consistency, but can also be interpreted as the regularity condition used by Hajek (1970) in order to rule out super-efficient estimates, as can easily be seen by an application of LeCam's third lemma, in view of (1.5), and the log-likelihood expansion implied by (1.2).

2. The risk of an estimate. Given a regular estimate $(T_N) = T(IC)$, we assess its variability in the following way. Consider a partition of $2\tau_\theta$ by means of two functions $\tau', \tau'' : \Theta \to \mathbb{R}$, i.e.

$$\tau'_\theta + \tau''_\theta = 2\tau_\theta \quad \text{for all} \quad \theta \in \Theta, \tag{2.1}$$

and assume that $\tau'_\theta, \tau''_\theta$ can be estimated consistently, i.e.

(2.2) there exist estimates $\tau'_N, \tau''_N : \Omega^N \to \mathbb{R}$, such that for all $\theta \in \Theta$, $\tau'_N = \tau'_\theta + o_{P_\theta^N}(1)$, $\tau''_N = \tau''_\theta + o_{P_\theta^N}(1)$.

Then look at the intervals $[T_N - N^{-\frac{1}{2}}\tau'_N, T_N + N^{-\frac{1}{2}}\tau''_N]$, which are laid around T_N and may be used as confidence intervals for the parameter θ that are of prescribed stochastic width $2\tau_{N,\theta}$. It is only reasonable to let the width shrink at an appropriate rate, since with increasing N even the contaminated sample provides more and more information to estimate θ. The possible dependence of the width on θ allows us to cover certain values of θ with higher precision than others. The partition of the width is to compensate possible asymmetries of T_N, which, in general, render the use of T_N as the strict midpoint unoptimal, in the sense defined subsequently. For $\theta \in \Theta$ and $W_\theta = (W_{N,\theta}) \in H_\theta$, consider the limiting error probabilities

$$\text{(2.3)} \qquad \begin{aligned} \alpha_\theta'(W_\theta) &= \limsup_N W_{N,\theta}\big(\theta < T_N - N^{-\frac{1}{2}}\tau_N'\big) \\ \alpha_\theta''(W_\theta) &= \limsup_N W_{N,\theta}\big(\theta > T_N + N^{-\frac{1}{2}}\tau_N''\big). \end{aligned}$$

Define the risk at θ, which still depends on the particular choice of τ_θ', τ_θ'', to be

$$\text{(2.4)} \qquad R(T(IC), \tau', \tau''; \theta) = \sup\{\alpha_\theta'(W_\theta) \vee \alpha_\theta''(W_\theta) : W_\theta \in H_\theta\}.$$

Subject to (2.1), (2.2), the functions τ', τ'' will be chosen in favor of the estimate, i.e., we define the risk $R(T(IC); \theta)$ of $T(IC)$ at θ to be

$$\text{(2.5)} \quad R(T(IC); \theta) = \inf\{R(T(IC), \tau', \tau''; \theta) : \tau', \tau'' \ \text{ subject to (2.1), (2.2)}\}.$$

For the explicit determination of this risk, fix $\theta \in \Theta$. Note that, corresponding to Proposition 3.1 of [12], the following asymptotic normality holds for every $W_\theta = (W_{N,\theta}) \in H_\theta$.

$$\text{(2.6)} \qquad \mathcal{L}_{W_{N,\theta}}\big(N^{\frac{1}{2}}(T_N - \theta) - S_N(IC_\theta; W_\theta)\big) \Rightarrow \mathfrak{N}(0, \sigma_\theta^2(IC_\theta)),$$

where $\sigma_\theta^2(IC_\theta) = \int IC_\theta^2 dP_\theta$. If $W_{N,\theta} = \otimes_{i=1}^N Q_{Ni}$, then the centering constants are of the form

$$\text{(2.7)} \qquad S_N(IC_\theta; W_\theta) = N^{-\frac{1}{2}}\Sigma_{i=1}^N \int IC_{N,\theta} dQ_{Ni}$$

for any sequence $(IC_{N,\theta}) \subset L^2(dP_\theta)$ satisfying $\int (IC_\theta - IC_{N,\theta})^2 dP_\theta \to 0$, $\|IC_{N,\theta}\|_\infty = o(N^{-\frac{1}{4}})$, $\int IC_{N,\theta} dP_\theta = 0$, cf. Behnen and Neuhaus (1975). Let $\inf_{[P_\theta]} IC_\theta$ denote the essential infimum and $\sup_{[P_\theta]} IC_\theta$ the essential supremum of IC_θ with respect to P_θ, and define the quantities $s_\theta'(IC_\theta)$, $s_\theta''(IC_\theta)$ by

$$\text{(2.8)} \quad s_\theta'(IC_\theta) = -\frac{\delta_\theta}{\tau_\theta}\inf_{[P_\theta]} IC_\theta + \frac{\varepsilon_\theta + \delta_\theta}{\tau_\theta}\sup_{[P_\theta]} IC_\theta, \qquad s_\theta''(IC_\theta) = s_\theta'(-IC_\theta).$$

Then the following lemma, which corresponds to Lemma 3.2, Lemma 3.3 of [12], yields the extreme shifts of the limiting normals occurring in (2.6).

LEMMA 2.1. *We have*

$$\text{(2.9)} \qquad \begin{aligned} &\sup\{\limsup_N S_N(IC_\theta; W_\theta) : W_\theta \in H_\theta\} \geqslant \tau_\theta s_\theta'(IC_\theta) \\ &\inf\Big\{\liminf_N S_N(IC_\theta; W_\theta) : W_\theta \in H_\theta\Big\} \leqslant -\tau_\theta s_\theta''(IC_\theta), \end{aligned}$$

with equalities holding in (2.9), *if* $\inf_{[P_\theta]} IC_\theta$ *and* $\sup_{[P_\theta]} IC_\theta$ *are finite.*

PROOF. Because of symmetry it suffices to prove the assertions concerning the sup. Drop θ as an index in this proof. Choose numbers a_1, a_2, such that $\inf_{[P]}IC < a_1 < a_2 < \sup_{[P]}IC$. Introduce the sets $A_1 = \{IC < a_1\}$, $A_2 = \{IC > a_2\}$ and put $\pi_j = P(A_j)$, $j = 1, 2$. Then, with $\Xi = \pi_2^{-1}I_{A_2}$, define probability measures Q_N by $dQ_N = (1 + \varepsilon_N(\Xi - 1))dP$, for N sufficiently large. Again, put $\pi_{jN} = Q_N(A_j)$, $j = 1, 2$, define the functions $\Delta_N = -\pi_{1N}^{-1}I_{A_1} + \pi_{2N}^{-1}I_{A_2}$ and then the probability measure R_N by $dR_N = (1 + \delta_N\Delta_N)dQ_N$, for N sufficiently large. Note that $R_N \in \mathcal{P}_N$. Furthermore, Linderberg's condition can be verified, so as to obtain that

$$\mathcal{L}_{P^N}\left(\Sigma_{i=1}^N \log\frac{dQ_N}{dP}(x_i)\right) \Rightarrow \mathcal{N}\left(-\tfrac{1}{2}\varepsilon^2\sigma^2(\Xi), \varepsilon^2\sigma^2(\Xi)\right)$$

$$\mathcal{L}_{Q_N^N}\left(\Sigma_{i=1}^N \log\frac{dR_N}{dQ_N}(x_i)\right) \Rightarrow \mathcal{N}\left(-\tfrac{1}{2}\delta^2\sigma^2(\Delta), \delta^2\sigma^2(\Delta)\right),$$

where $\Delta = -\pi_1^{-1}I_{A_1} + \pi_2^{-1}I_{A_2}$. Hence $(R_N^N) \in H$. It is now easy to see that $\lim_N N^{\frac{1}{2}}\int IC_N dR_N = \int IC(\varepsilon\Xi + \delta\Delta)dP$, which is greater than $-\delta a_1 + (\varepsilon + \delta)a_2$. Then let a_1 tend to $\inf_{[P]}IC$, and a_2 to $\sup_{[P]}IC$. In order to prove the equality assertion, if $\inf_{[P]}IC$ and $\sup_{[P]}IC$ are finite, we apply Lemma 3.2 and Lemma 3.3 of [12] to the function $IC^0 = \inf_{[P]}IC \vee IC \wedge \sup_{[P]}IC$, bearing in mind that the statistics $N^{-\frac{1}{2}}\Sigma_{i=1}^N IC^0(x_i)$ and $N^{-\frac{1}{2}}\Sigma_{i=1}^N IC(x_i)$ are asymptotically equivalent over H. The proof also shows in this case that, if $P(IC = \inf_{[P]}IC) > 0$ and $P(IC = \sup_{[P]}IC) > 0$, then the sup is a max and the inf is a min in (2.9). $\square$

Let Φ be the standard normal cdf. The preceding lemma implies that

$$R(T(IC), \tau', \tau''; \theta) = \Phi\left(\frac{(\tau_\theta s_\theta'(IC_\theta) - \tau_\theta') \vee (\tau_\theta s_\theta''(IC_\theta) - \tau_\theta'')}{\sigma_\theta(IC_\theta)}\right). \tag{2.10}$$

This quantity is minimized by the following choice of the functions τ', τ'' :

$$\begin{aligned} \tau_\theta'(IC_\theta) &= \tau_\theta\left(1 + \tfrac{1}{2}(s_\theta'(IC_\theta) - s_\theta''(IC_\theta))\right) \\ \tau_\theta''(IC_\theta) &= \tau_\theta\left(1 - \tfrac{1}{2}(s_\theta'(IC_\theta) - s_\theta''(IC_\theta))\right). \end{aligned} \tag{2.11}$$

Therefore, the following theorem has been proved.

THEOREM 2.2. *Let $T(IC)$ be a regular estimate. Then, at each $\theta \in \Theta$, we have*

$$R(T(IC); \theta) \geqslant \Phi\left[-\frac{\left(\tau_\theta 1 - \frac{1}{2}(s_\theta'(IC_\theta) + s_\theta''(IC_\theta))\right)}{\sigma_\theta(IC_\theta)}\right]. \tag{2.12}$$

Equality holds true in (2.12), *if $\tau_\theta'(IC_\theta)$, $\tau_\theta''(IC_\theta)$ can be estimated consistently.*

REMARKS. (1) The side condition will be fulfilled in most practical cases. Then the risk equals $\Phi\left(-\frac{\tau_\theta(1 - \eta_\theta \mathrm{Var}_{[P_\theta]}IC_\theta)}{\sigma_\theta(IC_\theta)}\right)$, where $2\tau_\theta\eta_\theta = \varepsilon_\theta + 2\delta_\theta$ and $\mathrm{Var}_{[P_\theta]}IC_\theta = \sup_{[P_\theta]}IC_\theta - \inf_{[P_\theta]}IC_\theta$.

(2) Let two regular estimates $T(IC')$, $T(IC'')$ be given. Apply, at stage N, $T(IC'')$ to N and $T(IC')$ to M_N observations, where M_N is chosen in such a way that, as $N \to \infty$, $T(IC')$ matches $T(IC'')$ with respect to the above risk. Then the sample size ratios $\frac{M_N}{N}$ necessarily tend to the following limit:

$$ARE_{mx}(T(IC'') : T(IC'); \theta) = \frac{\sigma_\theta^2(IC_\theta')}{\sigma_\theta^2(IC_\theta'')} \cdot \frac{(1 - \eta_\theta \mathrm{Var}_{[P_\theta]} IC_\theta'')^2}{(1 - \eta_\theta \mathrm{Var}_{[P_\theta]} IC_\theta')^2}.$$

Thus we rediscover the asymptotic relative efficiency of tests, introduced by Definition 5.2 of [12]. This quantity may be an alternative to the measure considered by Sievers (1978) and is not based upon such negligible probabilities. It should be added that, in order to be admitted to the comparison, an estimate $T(IC)$ must satisfy $R(T(IC); \theta) \leq \frac{1}{2}$, i.e., $T(IC)$ must not be worse than choosing $-\infty$ and $+\infty$, with probability $\frac{1}{2}$ each, without looking at the observations.

3. The Minimax Result. In minimizing the risk, we assume equality in (2.12) and make the substitution $IC_\theta = \frac{\psi_\theta}{\int \psi_\theta \Lambda_\theta dP_\theta}$. Thus, as if a test had been carried out in the testing model of [12], based upon the statistics $N^{-\frac{1}{2}}\Sigma_{i=1}^N \psi_\theta(x_i)$, between $\varepsilon_{N,\theta}$, $\delta_{N,\theta}$-neighborhoods of $P_{\theta-\tau_{N,\theta}}$ and $P_{\theta+\tau_{N,\theta}}$, we arrive at the problem: maximize

$$\frac{2\int \psi_\theta \Lambda_\theta dP_\theta - (s_\theta'(\psi_\theta) + s_\theta''(\psi_\theta))}{\sigma_\theta(\psi_\theta)} \tag{3.1}$$

with respect to $\psi_\theta \in L^2(dP_\theta)$, subject to $\int \psi_\theta dP_\theta = 0$, $\int \psi_\theta \Lambda_\theta dP_\theta \neq 0$.

This problem is solved by Theorem 3.7 of [12]. We have only to notice that the quantity $s(IC)$, defined by (3.6) of [12], does not decrease, when IC is truncated at its essential extrema. Therefore, the P_θ-a.e. unique solution is of the following form:

$$\psi_\theta^* = d_\theta' \vee \Lambda_\theta \wedge d_\theta'', \tag{3.2}$$

where the truncation points d_θ', d_θ'' are uniquely determined by the equations

$$\int (d_\theta' - \Lambda_\theta)^+ \, dP_\theta = \frac{\varepsilon_\theta + 2\delta_\theta}{2\tau_\theta} = \int (\Lambda_\theta - d_\theta'')^+ \, dP_\theta. \tag{3.3}$$

The corresponding IC-function at θ is given by

$$IC_\theta^* = \frac{\psi_\theta^*}{\psi_\theta^* \Lambda_\theta dP_\theta}. \tag{3.4}$$

So we have the following minimax result.

THEOREM 3.1. *Let $T(IC^*)$ be a regular estimate whose IC-function at $\theta \in \Theta$ is of the form* (3.4), *and assume that $\tau_\theta'(IC_\theta^*)$, $\tau_\theta''(IC_\theta^*)$ can be estimated consistently. Then $T(IC^*)$ minimizes $R(T(IC); \theta)$ among all regular estimates $T(IC)$, at each $\theta \in \Theta$, and $R(T(IC^*); \theta) = \Phi(-\tau_\theta \sigma_\theta(\psi_\theta^*))$.*

The optimality of $T(IC^*)$ can be interpreted in still another way. As already remarked, the IC-function will in most cases be supplied by the influence curve, which measures the influence of additional observations. It is a desirable robustness property that this function be bounded. (Let $\inf IC_\theta$ and $\sup IC_\theta$ denote the pointwise extrema of IC_θ.)

THEOREM 3.2. *Let $T(IC^*)$ be a regular estimate whose IC-function at $\theta \in \Theta$ is given by* (3.4). *Then $T(IC^*)$ minimizes the asymptotic variance $\int IC_\theta^2 dP_\theta$, subject to the bounds* $\inf IC_\theta \geqslant \inf IC_\theta^*$, $\sup IC_\theta \leqslant \sup IC_\theta^*$, *among all regular estimates $T(IC)$, at each $\theta \in \Theta$.*

PROOF. We can equivalently minimize $\int (IC_\theta - \Lambda_\theta)^2 dP_\theta$, because $\int IC_\theta^2 dP_\theta = \int (IC_\theta - \Lambda_\theta)^2 dP_\theta + 2 - \int \Lambda_\theta^2 dP_\theta$. It is obvious then that $\int (IC_\theta - \Lambda_\theta)^2 dP_\theta \geqslant \int (IC_\theta^* - \Lambda_\theta)^2 dP_\theta$ (with strict inequality unless $IC_\theta = IC_\theta^*$ a.e. P_θ). $\square$

This is just a reformulation of Hampel's (1968) Lemma 5, cf. Huber (1977), pages 32, 33. Indeed, if the laws $\mathcal{L}_{P_\theta}(\Lambda_\theta)$ are symmetric about zero, both statements coincide. In general, Hampel forces the absolute values of the upper and lower bounds to be the same, b_θ say, by determining a_θ, such that $\int (-b_\theta) \vee (\Lambda_\theta - a_\theta) \wedge b_\theta dP_\theta = 0$, and, accordingly, finds $\tilde{\psi}_\theta = (-b_\theta) \vee (\Lambda_\theta - a_\theta) \wedge b_\theta$ to be optimal.

4. Construction of a minimax estimate. This section provides the construction of a regular estimate $T(IC^*)$, IC^* given by (3.4), which is then optimal in the sense of Theorem 3.3, Theorem 3.4. Additional assumptions are required for this construction. The kind of estimates we shall look at are (M)-estimates, because they can easily be reconstructed from their local, resp. asymptotic, properties. The tools for the following one-step construction are borrowed from LeCam (1969), pages 79–81, 101–107.

Certainly, it must be assumed that the family is not overparametrized, i.e.,

$$\theta_1, \theta_2 \in \Theta, P_{\theta_1} = P_{\theta_2} \quad \text{implies that} \quad \theta_1 = \theta_2. \tag{4.1}$$

Then, in view of assumptions (1.1), (1.2) and since Θ is open, the conditions are satisfied under which LeCam constructs $N^{\frac{1}{2}}$-consistent estimates, i.e., $\tilde{\theta}_N : \Omega^N \to \Theta$, such that $\{\mathcal{L}_{P_\theta^N}(N^{\frac{1}{2}}(\tilde{\theta}_N - \theta)) : N \in \mathbb{N}\}$ is tight, for each $\theta \in \Theta$. It turns out to be technically more convenient to use as an initial estimate, instead of $\tilde{\theta}_N$, a discretized version $\hat{\theta}_N$ of $\tilde{\theta}_N$, which may be obtained in the following way. Fix $b \in (0, \infty)$ and cover Θ, for each N, by disjoint intervals of length $N^{-\frac{1}{2}}b$. For each such interval, fix a point of its intersection with Θ. Assign this value to $\hat{\theta}_N$, if $\tilde{\theta}_N$ falls into the interval. Note that $\hat{\theta}_N$ is still $N^{\frac{1}{2}}$-consistent. (In practice, such initial estimates can be obtained by more direct methods.)

Now let a function $\psi : \Theta \times \Omega \to \mathbb{R}$ be given, such that $\psi_\theta \in L^2(dP_\theta)$ for all $\theta \in \Theta$. Introduce the functions $\lambda_\theta(\zeta) = \int \psi_\zeta dP_\theta$, $\dot{\lambda}(\theta) = -\int \psi_\theta \Lambda_\theta dP_\theta$, θ, $\zeta \in \Theta$. Assume that

$$\dot{\lambda} \text{ maps } \Theta \text{ continuously into } (-\infty, 0), \tag{4.2}$$

and assume that, at each $\theta \in \Theta$:

(4.3) $$\lambda_\theta(\theta) = 0.$$

(4.4) λ_θ is differentiable at $\zeta = \theta$, with derivative $\dot{\lambda}(\theta)$.

(4.5) $$N^{-\frac{1}{2}}\Sigma_{i=1}^N\left(\psi_{\zeta_N}(x_i) - \psi_\theta(x_i) - \lambda_\theta(\zeta_N)\right) = o_{P_\theta^N}(1)$$

for every $(\zeta_N) \subset \Theta$, such that $(N^{\frac{1}{2}}(\zeta_N - \theta))$ is bounded.

Then define the estimate $T_N : \Omega^N \to \mathbb{R}$ by

(4.6) $$T_N = \hat{\theta}_N - \frac{1}{\dot{\lambda}(\hat{\theta}_N)}\left(N^{-1}\Sigma_{i=1}^N\psi_{\hat{\theta}_N}(x_i)\right).$$

THEOREM 4.1. *Under assumptions* (1.1), (1.2), (4.1)-(4.5), (T_N) *is a regular estimate* $T(IC)$ *with* $IC_\theta = \dfrac{\psi_\theta}{\int \psi_\theta \Lambda_\theta dP_\theta}$.

PROOF. For each $\theta \in \Theta$ we have

$$\begin{aligned} N^{\frac{1}{2}}(T_N - \theta) &- N^{-\frac{1}{2}}\Sigma_{i=1}^N IC_\theta(x_i) \\ &= N^{\frac{1}{2}}(\hat{\theta}_N - \theta)\left(1 - \frac{1}{\dot{\lambda}(\hat{\theta}_N)}\cdot\frac{\lambda_\theta(\hat{\theta}_N) - \lambda_\theta(\theta)}{\hat{\theta}_N - \theta}\right) \\ &\quad - \left(\frac{1}{\dot{\lambda}(\hat{\theta}_N)} - \frac{1}{\dot{\lambda}(\theta)}\right)\left(N^{-\frac{1}{2}}\Sigma_{i=1}^N\psi_\theta(x_i)\right) \\ &\quad - \frac{1}{\dot{\lambda}(\hat{\theta}_N)}\left(N^{-\frac{1}{2}}\Sigma_{i=1}^N\left(\psi_{\hat{\theta}_N}(x_i) - \psi_\theta(x_i) - \lambda_\theta(\hat{\theta}_N)\right)\right). \end{aligned}$$

The first two terms obviously tend to zero in P_θ^N-probability. To show the same for the third term, let $\kappa > 0$ be given and pick a $c \in (0, \infty)$, such that $|N^{\frac{1}{2}}(\hat{\theta}_N - \theta)| > c$ with probability less than κ, for all N. Denote by θ_{Nl}, $l = 1, \cdots, L$, the possible values of $\hat{\theta}_N$ in the disjoint intervals of length $N^{-\frac{1}{2}}b$ that intersect $[\theta - N^{-\frac{1}{2}}c, \theta + N^{-\frac{1}{2}}c]$. Note that $L \leqslant \left[\dfrac{2c}{b}\right] + 2$, and observe that, with probability exceeding $1 - \kappa$, we have

$$\begin{aligned} &\left|\Sigma_{i=1}^N\left(\psi_{\hat{\theta}_N}(x_i) - \psi_\theta(x_i) - \lambda_\theta(\hat{\theta}_N)\right)\right| \\ &\qquad \leqslant \max_{l=1,\cdots,L}\left|\Sigma_{i=1}^N\left(\psi_{\theta_{Nl}}(x_i) - \psi_\theta(x_i) - \lambda_\theta(\theta_{Nl})\right)\right|. \end{aligned}$$

Since $|N^{\frac{1}{2}}(\theta_{Nl} - \theta)| \leqslant c$ for all N and l, the assertion follows. □

We sketch only briefly how to verify these conditions for the particular ψ_θ^* defined by (3.2). It satisfies $\int\psi_\theta^* dP_\theta = 0$, $\int\psi_\theta^*\Lambda_\theta dP_\theta > 0$ automatically. As for conditions (4.4), (4.5), it is sufficient to ensure that the truncation points are continuous and to assume that $\Lambda_\zeta(x)$ is continuous in $\zeta = \theta$ a.e. P_θ. Continuity of d', d'' may be inferred from continuity of ε, δ, τ and from continuity, with respect to θ, of the integrals $\int(d - \Lambda_\theta)^+ dP_\theta$, $\int(\Lambda_\theta - d)^+ dP_\theta$, which are strictly isotone, resp. antitone, with respect to $d \in (\inf_{[P_\theta]}\Lambda_\theta, \sup_{[P_\theta]}\Lambda_\theta)$. The continuity of these integrals, as well as that of $\dot{\lambda}$, seems to require conditions which, if spelled out further, amount to Cramér-type differentiability assumptions such as those used in [12]. Given the continuity of $\varepsilon, \tau, d', d'', \dot{\lambda}$, note that also $\tau_\theta'(IC_\theta^*)$, $\tau_\theta''(IC_\theta^*)$ are continuous, and hence can be estimated consistently.

5. Complementary remarks. Over the full classes $H_\theta^0 = \{(W_{N,\theta}) : W_{N,\theta} \in \mathcal{P}_{N,\theta}^N$ for all $N\}$ essentially the same results obtain. Since contiguity is missing in this enlarged model, stronger assumptions must be imposed on the estimates under consideration, in order to make sure that convergence in probability under P_θ^N extends to all sequences in H_θ^0. Once conditions (1.5), (2.2) are strengthened in this way (i.e., $o_{P_\theta^N}(1)$ replaced by: $o_{W_{N,\theta}}(1)$ for all $(W_{N,\theta}) \in H_\theta^0$), we can appeal to the results of [12], thereby exchanging $s_\theta'(IC_\theta)$, $s_\theta''(IC_\theta)$ for the corresponding quantities defined by (3.5) of [12]. Then sections 2, 3 carry over.

In order to construct estimates that are consistent for θ, or that have an asymptotic expansion (1.5), over H_θ^0 for all $\theta \in \Theta$, one may try to modify existing proofs and assumptions for consistency and asymptotic normality of (M)-estimates. Let us look at the work by Huber (1967) and carry out the following modifications for each $\theta_0 \in \Theta$. Assume equality in (15), page 224. Take over his assumptions (B-1), (B-2), (B-3), (B-4), with expectations referring to P_{θ_0}. Replace "integrable" in (B-4)(i) by "bounded". Then the consistency results, Lemma 2 and Theorem 2, extend from $P_{\theta_0}^N$ to every $(W_{N,\theta_0}) \in H_{\theta_0}^0$. Similarly, assume this uniform consistency. Assume equality in (27), take over assumptions (N-1), (N-2), (N-3)(i), with expectations referring to P_{θ_0}. Assume that (N-3)(ii) holds true with the expectation sign replaced by the supremum over x. Assume that ψ_{θ_0} is bounded. Then the asymptotic normality results, Theorem 3 and Corollary, extend from $(P_{\theta_0}^N)$ to every $(W_{N,\theta_0}) \in H_{\theta_0}^0$.

It may also be possible to generalize LeCam's method of constructing $N^{\frac{1}{2}}$-consistent estimates, so as to obtain initial estimates $\tilde{\theta}_N$ that are $N^{\frac{1}{2}}$-consistent for θ under every $W_\theta \in H_\theta^0$, for all $\theta \in \Theta$. The one-step construction goes then through with only minor modifications.

It should also be noted that, since the members $W_{N,\theta}$ of a sequence $W_\theta \in H_\theta^0$ need not be connected over different sample sizes, the thus obtained approximations hold in fact uniformly over H_θ^0. This is not so in the contiguity-submodel, where one could have, for instance, tightness of $\{\mathcal{L}_{W_{N,\theta}}(N^{\frac{1}{2}}(\tilde{\theta}_N - \theta)) : N \in \mathbb{N}\}$ for every $W_\theta = (W_{N,\theta}) \in H_\theta$, without the family $\{\mathcal{L}_{W_{N,\theta}}(N^{\frac{1}{2}}(\tilde{\theta}_N - \theta)) : N \in \mathbb{N}, W_\theta \in H_\theta\}$ being tight. Clearly, such an initial estimate cannot be recommended for practical purposes.

REFERENCES

[1] Behnen, K. and Neuhaus, G. (1975). A central limit theorem under contiguous alternatives. *Ann. Statist.* **3** 1349–1353.

[2] Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 323–330.

[3] HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175–194. University of California Press.

[4] HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. Thesis, University of California, Berkeley.

[5] HODGES, J. L. AND LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

[6] HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 221–233.

[7] HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 269–278.

[8] HUBER, P. J. (1972). Robust statistics: a review. *Ann. Math. Statist.* **43** 1041–1067.

[9] HUBER, P. J. (1977). *Robust Statistical Procedures*. SIAM, Philadelphia.

[10] HUBER-CAROL, C. (1970). Etude asymptotique de tests robustes. Ph.D. Thesis, ETH Zürich.

[11] LECAM, L. (1969). Theorie Asymptotique de la Décision Statistique. Les Presses de l'Université de Montréal.

[12] RIEDER, H. (1978). A robust asymptotic testing model. *Ann. Statist.* **6** 1080–1094.

[13] SIEVERS, G. L. (1978). Estimates of location: a large deviation comparison. *Ann. Statist.* **6** 610–618.

UNIVERISITY OF CALIFORNIA
DEPARTMENT OF STATISTICS
EVANS HALL
BERKELEY, CA 94720