Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

C.A.J. Klaassen, W.R. van Zwet

On estimating a parameter and its score function

Department of Mathematical Statistics Report MS-R8402 January

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

ON ESTIMATING A PARAMETER AND ITS SCORE FUNCTION

C.A.J. KLAASSEN, W.R. VAN ZWET
University of Leyden

We consider the problem of estimating a real-valued parameter θ in the presence of an abstract nuisance parameter η, such as an unknown distributional shape. Attention is restricted to the case where the "score functions" for θ and η are orthogonal, so that fully asymptotically efficient estimation is not a priori impossible. For fixed sample size we provide a bound of Cramér-Rao type. The bound differs from the classical one for known η by a term involving the integrated mean square error of an estimator of a multiple of the score function for θ for the case where θ is known. This implies that an estimator of θ can only perform well over a class of shapes η if it is possible to estimate the score function for θ accurately over this class.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 62F11, 62F35, 62G05.
KEY WORDS & PHRASES: adaptation, score function, Cramér-Rao inequality, semi-parametric models.
NOTE: Research of W.R. van Zwet supported in part by Office of Naval Research Contract N00014-80-c-0163. This paper will appear in *Proceedings of the Neyman-Kiefer Conference, Berkeley 1983.*

Report MS-R8402
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. AN INEQUALITY OF CRAMÉR-RAO TYPE.

Let $X_1,\ldots,X_N$ be independent and identically distributed (i.i.d.) random variables with a common density $f(\cdot;\eta,\theta)$ with respect to a σ - finite measure μ on $\mathbb{R}$. The parameter of interest θ belongs to an open subset Θ of $\mathbb{R}$ and the nuisance parameter η ranges over an arbitrary set H. For unknown η and θ, it is required to estimate θ and this is done by means of an estimator $T_N = T_N(X_1,\ldots,X_N)$ for some measurable function $T_N : \mathbb{R}^N \to \mathbb{R}$. We are interested in the variance of T_N under $f(\cdot;\eta,\theta)$. We shall write $P_{\eta\theta}$, $E_{\eta\theta}$ and $\sigma^2_{\eta\theta}$ for probabilities, expectations and variances under this model. The indicator function of a set B will be denoted by 1_B.

Throughout, we shall make the following regularity assumptions on the model and on the estimators to be considered. The first set of assumptions concerns differentiability in quadratic mean of the square root of the density with respect to θ. We assume that for every (η,θ) there exists a function $\tau(\cdot;\eta,\theta)$ such that

$$(1.1)\qquad E_{\eta\theta}\,\tau^2(X_1;\eta,\theta) > 0\ ,$$

$$(1.2)\qquad \lim_{\theta'\to\theta} E_{\eta\theta}\left[\frac{f^{\frac12}(X_1;\eta,\theta') - f^{\frac12}(X_1;\eta,\theta)}{(\theta'-\theta)f^{\frac12}(X_1;\eta,\theta)} - \tfrac12\tau(X_1;\eta,\theta)\right]^2 = 0\ ,$$

$$(1.3)\qquad \lim_{\theta'\to\theta} \frac{P_{\eta\theta'}(\ f(X_1;\eta,\theta) = 0\ )}{(\theta'-\theta)^2} = 0\ .$$

Clearly this defines $\tau(\,\cdot\,;\eta,\theta)$ a.e. $[P_{\eta\theta}]$ and ensures that $E_{\eta\theta}\tau^2(X_1;\eta,\theta) < \infty$. We complete the definition of τ by requiring arbitrarily that

$$(1.4)\qquad \tau(x;\eta,\theta) = 0 \quad\text{if}\quad f(x;\eta,\theta) = 0\ .$$

Note that an equivalent formulation of (1.1) - (1.3) is

$$(1.5)\qquad \int\rho^2(x;\eta,\theta)d\mu(x) > 0\ ,$$

$$(1.6)\qquad \lim_{\theta'\to\theta} \int\left[\frac{f^{\frac12}(x;\eta,\theta') - f^{\frac12}(x;\eta,\theta)}{(\theta'-\theta)} - \tfrac12\rho(x;\eta,\theta)\right]^2 d\mu(x) = 0\ ,$$

where ρ is of the form

$$(1.7) \quad \rho(x;\eta,\theta) = \tau(x;\eta,\theta) f^{\frac{1}{2}}(x;\eta,\theta) ,$$

i.e. $\rho = 0$ if $f = 0$.

For fixed η , the function $\tau(\cdot;\eta,\theta)$ is called the score function for θ and if f is differentiable in the ordinary sense, it coincides with $\partial \log f(\cdot;\eta,\theta)/\partial\theta$ a.e. $[P_{\eta\theta}]$. For known η , the Fisher information concerning θ that is contained in a single observation X_1 is defined by

$$(1.8) \quad I_\eta(\theta) = E_{\eta\theta}\tau^2(X_1;\eta,\theta) = \int \rho^2(x;\eta,\theta)d\mu(x) \in (0,\infty) .$$

Our second set of assumptions concerns the estimator T_N . We assume that for every (η,θ)

$$(1.9) \quad E_{\eta\theta}\, T_N = \chi(\eta,\theta) \in (-\infty,\infty)$$

and that if $E_{\eta\theta}\, T_N^2 < \infty$ for a certain (η,θ) , then T_N^2 is uniformly integrable with respect to $P_{\eta\theta'}$ for all θ' in a neighborhood of θ . Thus, for some $\varepsilon > 0$,

$$(1.10) \quad \lim_{C\to\infty} \sup_{|\theta'-\theta|<\varepsilon} E_{\eta\theta'}\, T_N^2\, 1_{\{|T_N|\geq C\}} = 0 .$$

Under the assumptions made so far, the Cramér-Rao inequality for known η is valid for T_N , so

$$(1.11) \quad \sigma^2_{\eta\theta}(T_N) \geq \frac{\{\dot\chi(\eta,\theta)\}^2}{NI_\eta(\theta)}$$

where $\dot\chi(\eta,\theta) = \partial\chi(\eta,\theta)/\partial\theta$. Define the function $J(\cdot;\eta,\theta)$ by

$$(1.12) \quad J(x;\eta,\theta) = \frac{\dot\chi(\eta,\theta)}{I_\eta(\theta)}\, \tau(x;\eta,\theta)$$

and let

$$(1.13) \quad S_N(\eta,\theta) = \frac{1}{N}\sum_{i=1}^{N} J(X_i;\eta,\theta) .$$

We note that (1.11) is a consequence of the orthogonality of $S_N(\eta,\theta)$ and $T_N - S_N(\eta,\theta)$ which yields

$$\sigma^2_{\eta\theta}(T_N) = \sigma^2_{\eta\theta}(S_N(\eta,\theta)) + \sigma^2_{\eta\theta}(T_N - S_N(\eta,\theta)) =$$

$$= \frac{\{\dot{\chi}(\eta,\theta)\}^2}{NI_\eta(\theta)} + \sigma^2_{\eta\theta}(T_N - S_N(\eta,\theta)) \ . \tag{1.14}$$

But this implies, in addition, that $\sigma^2_{\eta\theta}(T_N)$ can only come close to the Cramér-Rao bound (1.11) if $T_N - \chi(\eta,\theta)$ is close to $S_N(\eta,\theta)$ under $P_{\eta\theta}$. However, if H is a large set in some function space, say, then $T_N - \chi(\eta,\theta)$ can obviously not mimic $S_N(\eta,\theta)$ arbitrarily well for all $\eta \in H$ and consequently, $\sigma^2_{\eta\theta}(T_N)$ can't come arbitrarily close to the Cramér-Rao bound for all $\eta \in H$ simultaneously. Since we are considering the case where $\eta \in H$ is unknown, it should be possible to improve on (1.11).

Let us turn this argument around for a moment. If T_N performs well as an estimator of θ - or rather of $\chi(\eta,\theta)$ which may include a bias term - for all $\eta \in H$, then $T_N - \chi(\eta,\theta)$ must resemble $S_N(\eta,\theta)$ under $P_{\eta\theta}$ for every $\eta \in H$ and $\theta \in \Theta$. It would seem therefore that $T_N - \chi(\eta,\theta)$ must contain information about the unknown function $J(\cdot;\eta,\theta)$. Let us try to extract this information. For every fixed $\theta \in \Theta$, let $\psi(X_1;\theta)$ be a sufficient statistic for X_1 with respect to the remaining parameter $\eta \in H$. According to the factorization theorem this means that

$$f(x;\eta,\theta) = g(\psi(x;\theta);\eta,\theta)\cdot h(x;\theta) \quad \text{a.e.} \quad [\mu] \tag{1.15}$$

for appropriately chosen g and h. Suppose, moreover, that for all (η,θ)

$$E_{\eta\theta}\Big(\tau(X_1;\eta,\theta)\,|\,\psi(X_1;\theta)\Big) = 0 \quad \text{a.e.} \quad [P_{\eta\theta}] \ . \tag{1.16}$$

Then, for $i = 1,\ldots,N$, we have

$$\begin{aligned} &E_{\eta\theta}\Big(S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i;\ X_i=x\Big) + \\ &- E_{\eta\theta}\Big(S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i;\ \psi(X_i;\theta) = \psi(x;\theta)\Big) = \frac{1}{N}\, J(x;\eta,\theta) \ \text{a.e.}\ [P_{\eta\theta}] \ . \end{aligned} \tag{1.17}$$

Since $T_N - \chi(\eta,\theta)$ resembles $S_N(\eta,\theta)$ under $P_{\eta\theta}$, we may hope that
$N\, E_{\eta\theta}(T_N|\psi(X_j;\theta) \text{ for } j\neq i;\ X_i=x) - N\, E_{\eta\theta}(T_N|\psi(X_j;\theta) \text{ for } j\neq i;\ \psi(X_i;\theta) = \psi(x;\theta))$
or rather its symmetrized version

$$
(1.18)\quad J_N(x;\theta) = \sum_{i=1}^{N} \{E_{\eta\theta}(T_N | \psi(X_j;\theta) \text{ for } j \neq i;\ X_i = x) + \\
- E_{\eta\theta}(T_N | \psi(X_j;\theta) \text{ for } j \neq i;\ \psi(X_i;\theta) = \psi(x;\theta))\}
$$

can serve as an estimator of $J(x;\eta,\theta)$. Note that since for each j, $\psi(X_j;\theta)$ is sufficient for X_j for fixed θ, J_N is indeed independent of η. For known θ it is therefore a legitimate estimator.

We shall prove the following result.

THEOREM 1.1.
Suppose that for every (η,θ) *assumptions* (1.1) - (1.3), (1.9) *and* (1.10) *are satisfied. For every fixed* θ, *let* $\psi(X_1;\theta)$ *be sufficient for* X_1 *with respect to* η *and let* (1.16) *hold for all* (η,θ). *Then, for every* (η,θ),

$$
(1.19)\quad \sigma^2_{\eta\theta}(T_N) \geq \frac{\{\dot{\chi}(\eta,\theta)\}^2}{N I_\eta(\theta)} + \frac{1}{N} E_{\eta\theta} \int \{J_N(x;\theta) - J(x;\eta,\theta)\}^2 f(x;\eta,\theta)\,d\mu(x) .
$$

The theorem asserts that the Cramér-Rao bound may be improved by adding N^{-1} times the integrated mean square error (MSE) of the estimator J_N of the function J, which is an unknown multiple of the score function τ. It is unsatisfactory that the right-hand side of (1.19) depends on the choice of T_N. However, one may obviously rephrase the theorem to assert only the existence of an estimator J_N such that (1.19) holds. The message of the theorem is then clear: the accuracy with which one can estimate θ for unknown η is delimited by how well one can do for known η on the one hand and how well one can estimate $J(\cdot;\eta,\theta)$ for known θ on the other. Clearly the latter depends strongly on the class H. If $\tau(\cdot;\eta,\theta)$ runs through a large class of score functions as η ranges over H, then the integrated MSE of any estimator of J may be quite large, especially for some particularly irregular choices of τ. If τ is restricted to a smaller class of nicely behaved score functions as $\eta \in H$, then the integrated MSE may be much smaller. Finally, if η is known so that H consists of a single element, then $J(\cdot;\eta,\theta)$ can serve as an estimator of itself and (1.19) reduces to the Cramér-Rao inequality.

In a sense, the result of theorem 1.1 is not at all surprising. Adaptive estimators of a parameter for an unknown distributional shape are always based on some kind of preliminary estimate of the unknown score function followed by a good estimate of θ for the distributional shape corresponding to the estimated score function. For such estimators it is to be expected that a bound on their accuracy should involve both the accuracy of estimating θ for known η and that of estimating η for known θ. The novel aspect of theorem 1.1, however, is that it is not assumed that the estimator T_N is based on a preliminary estimate of the score function, but that an estimate of J for known θ is derived from T_N. In effect we are saying that any successful adaptive estimation procedure must involve - either explicitly or implicitly - the estimation of the score function (or rather of J) and that because of this, the accuracy of estimating J enters into the lower bound for the variance of the adaptive estimator.

Though theorem 1.1 is purely a finite sample result, it obviously has asymptotic implications. As an example, it clearly provides a finite sample analogue of a conjecture of Bickel (1982) which states, loosely speaking, that asymptotically fully efficient adaptive estimation is possible only if a consistent estimator of the score function exists.

In this connection the role of assumption (1.16) is of interest. It is well-known (cf. Stein (1956), Bickel (1982) and Begun, Hall, Huang and Wellner (1983)) that a necessary condition for asymptotically fully efficient adaptive estimation to be possible, is that the two estimation problems - that of θ for known η and that of η for known θ - are, in a sense, asymptotically orthogonal. Since $\psi(X_1;\theta)$ is sufficient with respect to η for known θ and $\tau(X_1;\eta,\theta)$ contains the information about θ locally for known η, assumption (1.16) is indeed an asymptotic orthogonality condition of this kind. In making this assumption we are therefore restricting attention to the case where fully asymptotically efficient estimation is not a priori impossible. In a way, this restriction is a reasonable one because without it, the Cramér-Rao inequality (1.11) is no longer a logical point of departure. In a companion paper we intend to discuss the more general situation where orthogonality is not necessarily present.

Even though it serves the same purpose, assumption (1.16) looks a bit different from the orthogonality conditions employed by other authors. Stein (1956) and Begun et al. (1983) define a class of score functions for η as the class of all limits, in the ordinary sense or in L^2, of the form

$$\lim_{\nu\to\infty} \frac{\log f(\cdot;\eta_\nu,\theta) - \log f(\cdot;\eta,\theta)}{d(\eta_\nu,\eta)} ,$$

or

$$2 \lim_{\nu\to\infty} \frac{f^{\frac{1}{2}}(\cdot;\eta_\nu,\theta) - f^{\frac{1}{2}}(\cdot;\eta,\theta)}{d(\eta_\nu,\eta) f^{\frac{1}{2}}(\cdot;\eta,\theta)} ,$$

where d denotes an appropriately chosen distance and $\lim d(\eta_\nu,\eta) = 0$. For all such score functions $\sigma(\cdot;\eta,\theta)$ they require

$$(1.20)\quad E_{\eta\theta}\ \tau(X_1;\eta,\theta)\sigma(X_1;\eta,\theta) = 0 .$$

Bickel (1982) considers all "score functions" $\sigma(\cdot;\eta,\theta)$ of the form

$$\frac{f(\cdot;\eta',\theta) - f(\cdot;\eta,\theta)}{f(\cdot;\eta,\theta)}$$

for $\eta' \in H$ and again requires (1.20), which now reduces to

$$(1.21)\quad E_{\eta'\theta}\ \tau(X_1;\eta,\theta) = 0 \qquad \text{for all } \eta' \in H .$$

Under an additional completeness assumption on the sufficient statistic $\psi(X_1;\theta)$, condition (1.16) in theorem 1.1 may be replaced by a condition of the form (1.20) for an appropriate class of "score functions" σ . Since we are not concerned with asymptotics where only local properties count, there seems to be no need to introduce differentiation with respect to η in order to define our score functions. Bickel's definition, however, has the drawback that the expectation in (1.21) need not exist. To remedy this we consider all score functions of the form

$$\frac{f^{\frac{1}{2}}(\cdot;\eta',\theta) - f^{\frac{1}{2}}(\cdot;\eta,\theta)}{f^{\frac{1}{2}}(\cdot;\eta,\theta)}$$

for $\eta' \in H$ and require (1.20), which reduces to

$$(1.22)\quad \int \tau(x;\eta,\theta)f^{\frac{1}{2}}(x;\eta,\theta)f^{\frac{1}{2}}(x;\eta',\theta)d\mu(x) = 0 \qquad \text{for all } \eta' \in H .$$

Of course we have to tailor the completeness assumption on $\psi(X_1;\theta)$ to this particular choice of score functions. Define densities

$$f(\cdot;\eta,\eta',\theta) = A(\eta,\eta',\theta)f^{\frac{1}{2}}(\cdot;\eta,\theta)f^{\frac{1}{2}}(\cdot;\eta',\theta)$$

for all η' in the set $H_{\eta\theta}$ where

$$A^{-1}(\eta,\eta',\theta) = \int f^{\frac{1}{2}}(x;\eta,\theta)f^{\frac{1}{2}}(x;\eta',\theta)d\mu(x) > 0 .$$

We shall write $P_{\eta\eta'\theta}$ and $E_{\eta\eta'\theta}$ for probabilities and expectations under this model. For every fixed η and θ we assume that $\psi(X_1;\theta)$ is complete with respect to $\eta' \in H_{\eta\theta}$ under this model, i.e. if for some (η,θ) and for some measurable function m, $E_{\eta\eta'\theta}\, m(\psi(X_1;\theta)) = 0$ for all $\eta' \in H_{\eta\theta}$, then $P_{\eta\eta'\theta}(m(\psi(X_1;\theta)) = 0) = 1$ for all $\eta' \in H_{\eta\theta}$.

THEOREM 1.2.

Suppose that for every (η,θ) *assumptions* (1.1) - (1.3), (1.9) *and* (1.10) *are satisfied. For every fixed* θ, *let* $\psi(X_1;\theta)$ *be sufficient for* X_1 *with respect to* η *; for every fixed* (η,θ), *let* $\psi(X_1;\theta)$ *be complete with respect to* η' *under the model* $P_{\eta\eta'\theta}$. *Suppose finally that* (1.22) *holds for all* (η,η',θ). *Then, for every* (η,θ), *inequality* (1.19) *holds.*

In section 2 we provide the proofs of theorems 1.1 and 1.2. The most obvious example, i.e. the estimation of location for an unknown symmetric density, is briefly discussed in section 3.

2. PROOF OF THE THEOREMS.

Let

$$f_N(x;\eta,\theta) = \prod_{i=1}^{N} f(x_i;\eta,\theta)$$

denote the density of $X = (X_1,\ldots,X_N)$ with respect to the N - fold product measure μ_N taken at the point $x = (x_1,\ldots,x_N)$. Since N is fixed, a standard argument shows that (1.6) and (1.7) - or equivalently (1.2) and (1.3) - imply

$$\lim_{\theta'\to\theta} \int \left[\frac{f_N^{\frac{1}{2}}(x;\eta,\theta') - f_N^{\frac{1}{2}}(x;\eta,\theta)}{(\theta'-\theta)} - \tfrac{1}{2}\rho_N(x;\eta,\theta)\right]^2 d\mu_N(x) = 0 , \tag{2.1}$$

where

$$\rho_N(x;\eta,\theta) = f_N^{\frac{1}{2}}(x;\eta,\theta) \sum_{i=1}^{N} \tau(x_i;\eta,\theta) . \tag{2.2}$$

Suppose that $E_{\eta\theta}\, T_N^2 < \infty$ for a certain (η,θ). Take $\varepsilon > 0$ as in (1.10) and $|\theta'-\theta| < \varepsilon$. In view of (1.9) and (2.1),

$$\frac{\chi(\eta,\theta')-\chi(\eta,\theta)}{(\theta'-\theta)} = \int T_N(x)\cdot\frac{f_N^{\frac12}(x;\eta,\theta')-f_N^{\frac12}(x;\eta,\theta)}{(\theta'-\theta)}\cdot\{f_N^{\frac12}(x;\eta,\theta')+f_N^{\frac12}(x;\eta,\theta)\}d\mu_N(x) =$$

(2.3)

$$= \int T_N(x)\{\tfrac12\rho_N(x;\eta,\theta)+\Delta_N(x;\eta,\theta,\theta')\}\{f_N^{\frac12}(x;\eta,\theta')+f_N^{\frac12}(x;\eta,\theta)\}d\mu_N(x)$$

with

$$\text{(2.4)}\qquad \lim_{\theta'\to\theta}\int\Delta_N^2(x;\eta,\theta,\theta')d\mu_N(x) = 0\ .$$

Because of (1.10), $E_{\eta\theta'}\, T_N^2$ is bounded for $|\theta'-\theta| < \varepsilon$ and by the Cauchy-Schwarz inequality

$$\text{(2.5)}\qquad \lim_{\theta'\to\theta}\int T_N(x)\Delta_N(x;\eta,\theta,\theta')\{f_N^{\frac12}(x;\eta,\theta') + f_N^{\frac12}(x;\eta,\theta)\}d\mu_N(x) = 0\ .$$

By another application of the Cauchy-Schwarz inequality combined with (1.6), (2.1) and (1.10) ,

$$\text{(2.6)}\qquad \lim_{\theta'\to\theta}\int T_N(x)\rho_N(x;\eta,\theta)\{f_N^{\frac12}(x;\eta,\theta') - f_N^{\frac12}(x;\eta,\theta)\}d\mu_N(x) = 0\ .$$

Together, (2.3), (2.5), (2.6) and (2.2) imply the existence of $\dot\chi(\eta,\theta)$ as well as

$$\text{(2.7)}\qquad \dot\chi(\eta,\theta) = E_{\eta\theta}\, T_N \sum_{i=1}^{N} \tau(X_i;\eta,\theta)\ .$$

Repeating this argument with both T_N and χ replaced by 1 , we find

$$\text{(2.8)}\qquad E_{\eta\theta}\, \tau(X_1;\eta,\theta) = 0\ .$$

Combining (2.7) and (2.8) we arrive at the decomposition (1.14).

To prove theorem 1.1 it remains to study $\sigma^2_{\eta\theta}(T_N-S_N(\eta,\theta))$ for $S_N(\eta,\theta)$ as defined by (1.12) and (1.13). We begin by noting that

$$\text{(2.9)}\qquad \sigma^2_{\eta\theta}(T_N-S_N(\eta,\theta)) \ge E\ \sigma^2_{\eta\theta}\Big(T_N-S_N(\eta,\theta)\,|\,\psi(X_1;\theta),\ldots,\psi(X_N;\theta)\Big)\ .$$

Consider the conditional distribution of $X = (X_1,\ldots,X_N)$ given $\psi(X_1;\theta),\ldots,\psi(X_N;\theta)$. Under this conditional probability model, $X_1,\ldots,X_N$ are still i.i.d. and an application of Hājěk's projection lemma (cf. Hājek (1968)) to this conditional setup yields

$$\sigma^2_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_1;\theta),\ldots,\psi(X_N;\theta)\Big) \geq \tag{2.10}$$
$$\geq \sum_{i=1}^N \sigma^2_{\eta\theta}\Big\{E_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i; X_i\Big)\Big|\,\psi(X_1;\theta),\ldots,\psi(X_N;\theta)\Big\} .$$

It follows from (2.9), (2.10) and the inequality $\sum_{i=1}^N a_i^2 \geq N^{-1}(\sum_{i=1}^N a_i)^2$ that

$$\sigma^2_{\eta\theta}(T_N - S_N(\eta,\theta)) \geq \sum_{i=1}^N E_{\eta\theta}\Big\{E_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i; X_i\Big) +$$
$$- E_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_1;\theta),\ldots,\psi(X_N;\theta)\Big)\Big\}^2 =$$
$$= \sum_{i=1}^N E_{\eta\theta} \int\Big\{E_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i; X_i = x\Big) + \tag{2.11}$$
$$- E_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i; \psi(X_i;\theta) = \psi(x;\theta)\Big)\Big\}^2 f(x;\eta,\theta)d\mu(x) \geq$$
$$\geq N^{-1} E_{\eta\theta} \int\Big\{\sum_{i=1}^N \Big[E_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i; X_i = x\Big) +$$
$$- E_{\eta\theta}\Big(T_N - S_N(\eta,\theta)\,|\,\psi(X_j;\theta) \text{ for } j\neq i; \psi(X_i;\theta) = \psi(x;\theta)\Big)\Big]\Big\}^2 f(x;\eta,\theta)d\mu(x) .$$

But since (1.16) implies (1.17) and in view of definition (1.18), we may write (2.11) as

$$\sigma^2_{\eta\theta}(T_N - S_N(\eta,\theta)) \geq N^{-1} E_{\eta\theta}\int \{J_N(x;\theta) - J(x;\eta,\theta)\}^2 f(x;\eta,\theta)d\mu(x) . \tag{2.12}$$

Theorem 1.1 now follows from (1.14) and (2.12).

To prove theorem 1.2, we note that the factorization theorem (cf. (1.15)) ensures that for fixed (η,θ), $\psi(X_1;\theta)$ is sufficient for X_1 with respect to $\eta' \in H_{\eta\theta}$ under the model $P_{\eta\eta'\theta}$. It follows that

$$E_{\eta\eta'\theta}\Big(\tau(X_1;\eta,\theta)\,|\,\psi(X_1;\theta)\Big) \tag{2.13}$$

is independent of $\eta' \in H_{\eta\theta}$. However, according to (1.22),

$$E_{\eta\eta'\theta}\Big\{E_{\eta\eta'\theta}\Big(\tau(X_1;\eta,\theta)\,|\,\psi(X_1;\theta)\Big)\Big\} = 0 \tag{2.14}$$

for all $\eta' \in H_{\eta\theta}$ and the completeness assumption implies that the conditional expectation in (2.13) vanishes a.s. under $P_{\eta\eta'\theta}$ for every $\eta' \in H_{\eta\theta}$. Since (2.13) is independent of η', we may take $\eta' = \eta$ and (1.16) follows. Theorem 1.2 is now a consequence of theorem 1.1.

3. ESTIMATING LOCATION UNDER SYMMETRY.

Let H be the class of probability densities η with respect to Lebesgue measure on $\mathbb{R}$, which are symmetric about 0 and absolutely continuous with derivative η', and which possess a finite Fisher information

$$(3.1) \qquad I_\eta = \int \left\{\frac{\eta'(x)}{\eta(x)}\right\}^2 \eta(x)dx < \infty .$$

Let $X_1,\ldots,X_N$ be i.i.d. with a common density $f(\cdot;\eta,\theta) = \eta(\cdot-\theta)$, where $\eta \in H$ and $\theta \in \mathbb{R}$ are both unknown. Under this model it is reasonable to estimate θ by a location equivariant estimator $T_N = T_N(X_1,\ldots,X_N)$, i.e. an estimator satisfying

$$(3.2) \qquad T_N(x_1+a,\ldots,x_N+a) = T_N(x_1,\ldots,x_N) + a$$

for all $x = (x_1,\ldots,x_N) \in \mathbb{R}^N$ and $a \in \mathbb{R}$. If we assume that $E_{\eta\theta}\, T_N^2 < \infty$, then

$$(3.3) \qquad \chi(\eta,\theta) = E_{\eta\theta}\, T_N = \phi(\eta) + \theta ,$$

so that $\dot{\chi}(\eta,\theta) \equiv 1$.

It is easy to see that for this model the regularity conditions (1.5) - (1.7) (or equivalently (1.1) - (1.3)) are satisfied with $\tau = -\eta'(\cdot-\theta)/\eta(\cdot-\theta)$. Clearly assumptions (1.9) and (1.10) on T_N also hold. Choosing

$$(3.4) \qquad \psi(x;\theta) = |x-\theta| ,$$

we see that for fixed θ, $\psi(X_1;\theta)$ is sufficient for X_1 with respect to $\eta \in H$. Since η'/η is an odd function and η is symmetric, we have

$$(3.5) \qquad E_{\eta\theta}\left(\frac{\eta'(X_1-\theta)}{\eta(X_1-\theta)}\,\middle|\,|X_1-\theta|\right) = E_{\eta 0}\left(\frac{\eta'(X_1)}{\eta(X_1)}\,\middle|\,|X_1|\right) = 0 \quad \text{a.s.}$$

in view of (3.1). Hence the assumptions of theorem 1.1 are satisfied and

$$(3.6) \qquad \sigma^2_{\eta\theta}(T_N) \geq \frac{1}{NI_\eta} + \frac{1}{N} E_{\eta\theta}\int \{J_N(x;\theta)-J(x;\eta,\theta)\}^2 \eta(x-\theta)dx ,$$

where

$$(3.7) \qquad J(x;\eta,\theta) = -I_\eta^{-1} \frac{\eta'(x-\theta)}{\eta(x-\theta)} ,$$

$$J_N(x;\theta) = \sum_{i=1}^{N} \left\{E_{\eta\theta}\left(T_N \middle| |X_j-\theta| \text{ for } j\neq i; X_i=x\right) + \right.$$

$$\left. - E_{\eta\theta}\left(T_N \middle| |X_j-\theta| \text{ for } j\neq i; |X_i-\theta|=|x-\theta|\right)\right\} =$$

$$(3.8) \qquad = \tfrac{1}{2} \sum_{i=1}^{N} \left\{E_{\eta\theta}\left(T_N \middle| |X_j-\theta| \text{ for } j\neq i; X_i-\theta=x-\theta\right) + \right.$$

$$\left. - E_{\eta\theta}\left(T_N \middle| |X_j-\theta| \text{ for } j\neq i; X_i-\theta=-(x-\theta)\right)\right\} .$$

Obviously, neither side of (3.6) depends on θ and we may therefore simplify (3.6) to

$$(3.9) \qquad \sigma^2_{\eta 0}(T_N) \geq \frac{1}{NI_\eta} + \frac{1}{N} E_{\eta 0}\int \{J_N(x)-J(x;\eta)\}^2 \eta(x)dx ,$$

where

$$(3.10) \qquad J(x;\eta) = - I_\eta^{-1} \frac{\eta'(x)}{\eta(x)} ,$$

$$(3.11) \qquad J_N(x) = \tfrac{1}{2}\sum_{i=1}^{N} \left\{E_{\eta 0}\left(T_N \middle| |X_j| \text{ for } j\neq i; X_i=x\right) - E_{\eta 0}\left(T_N \middle| |X_j| \text{ for } j\neq i; X_i=-x\right)\right\} .$$

This result is given in Klaassen (1981), which also contains a discussion of the implications of inequality (3.9).

REFERENCES.

[1] 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.

[2] Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.* 10, 647-671.

[3] Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* 39, 325-346.

[4] Klaassen, C.A.J. (1981). *Statistical Performance of Location Estimators.* Mathematical Centre Tract 133, Mathematisch Centrum, Amsterdam.

[5] Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. and Probability* 1, 187-195. University of California Press.