INCOMPLETE GENERALIZED $L$-STATISTICS

BY OLA HöSSJER

Lund University

Given data $X_1, \ldots , X_n$ and a kernel $h$ with $m$ arguments, Serfling introduced the class of generalized $L$-statistics (GL-statistics), which is defined by taking linear combinations of the ordered $h(X_{i_1}, \ldots , X_{i_m})$, where $(i_1, \ldots , i_m)$ ranges over all $n!/(n-m)!$ distinct $m$-tuples of $(1, \ldots , n)$. In this paper we derive a class of incomplete generalized $L$-statistics (IGL-statistics) by taking linear combinations of the ordered elements from a subset of $(h(X_{i_1}, \ldots , X_{i_m}))$ with size $N(n)$. A special case is the class of incomplete $U$-statistics, introduced by Blom. Under very general conditions, the IGL-statistic is asymptotically equivalent to the GL-statistic as soon as $N(n)/n \to \infty$ as $n \to \infty$, which makes the IGL much more computationally feasible. We also discuss various ways of selecting the subset of $(h(X_{i_1}, \ldots , X_{i_m}))$. Several examples are discussed. In particular, some new estimates of the scale parameter in nonparametric regression are introduced. It is shown that these estimates are asymptotically equivalent to an IGL-statistic. Some extensions, for example, functionals other than $L$ and multivariate kernels, are also addressed.

1. Introduction. Let $X_1, \ldots , X_n$ be independent random variables taking values in $\mathbb{R}^q$, with a common probability distribution $F$, and let $h: \mathbb{R}^{mq} \to \mathbb{R}$ be a Borel measurable function. For each $I = (i_1, \ldots , i_m)$ with $i_j \neq i_j$ if $j \neq j'$, define $h(X_I) = h(X_{i_1}, \ldots , X_{i_m})$. We will assume that $I$ takes values in $S_n(m)$, the set of all $n_1(m) = n!/(n-m)! m$-tuples with distinct elements. Let $D_n = \{I_1, \ldots , I_N\}$ be a subset of $S_n(m)$, called the design of the experiment, where some $I \in S_n(m)$ may occur several times in $D_n$ and $N = N(n) = |D_n|$. Assume now that $h_{1:N} \leq \cdots \leq h_{N:N}$ are the order statistics of $(h(X_I); I \in D_n)$. Given a triangular array of weights $\{c_{Ni}; N \geq 1, 1 \leq i \leq N\}$, we will consider statistics of the form

$$ \sum_{i=1}^{N} c_{Ni} h_{i:N}. $$

(1.1)

It will be convenient to put (a subclass of) the statistics in (1.1) into a functional form. In doing so, consider first the statistical $L$-functional

$$ T(G) = \int_0^1 J(t) G^{-1}(t) \, dt + \sum_{j=1}^{d} a_j G^{-1}(p_j), $$

Received April 1994; revised December 1995.


Key words and phrases. Asymptotic normality, incomplete $U$-statistics, invariance principle, order statistics, nonparametric regression, $L$-statistics, scale estimation, $U$-statistics.
with \( G^{-1}(t) = \inf \{ x; G(x) > t \} \) the right-continuous inverse of \( G \) and \( 0 < p_j < 1 \). Let \( H_F \) be the distribution of \( h(X_1, \ldots, X_m) \), and consider estimating \( T(H_F) \) by

\[
T(H_n) = \int_0^1 J(t) H_n^{-1}(t) \, dt + \sum_{j=1}^d a_j H_n^{-1}(p_j),
\]

where \( H_n \) is the empirical distribution formed by \( \{ h(X_i); \, I \in D_n \} \). Note that \( T(H_n) \) may be put into the form (1.1), with

\[
c_{N_i} = \int_{(i-1)/N}^{i/N} J(t) \, dt + \sum_{j=1}^d a_j \mathbf{1}_{\{ i \in [Np_j]+1 \}},
\]

where \( \lfloor x \rfloor \) is the largest integer smaller than or equal to \( x \) and \( 1_A \) is the indicator function for the event \( A \). Special cases of (1.2) include the following:

1. \( U \)-statistics [Hoeffding (1948)]; \( D_n = S_n(m), \, J = 1, \, d = 0 \), so that

\[
T(H_n) = \frac{1}{n^{(m)}} \sum_{I \in S_n(m)} h(X_I).
\]

2. Incomplete \( U \)-statistics [Blom (1976)]; \( J = 1, \, d = 0, \, D_n \) arbitrary, so that

\[
T(H_n) = \frac{1}{N} \sum_{I \in D_n} h(X_I).
\]

3. Generalized \( L \)-statistics (GL-statistics) introduced by Serfling (1984). [See also Janssen, Serfling and Veraverbeke (1984) and Choudhury and Serfling (1988).] Here \( D_n = S_n(m) \), and \( J, \, d, \, \{ p_j \}_1^d \) and \( \{ a_j \}_1^d \) are arbitrary. For the special case when \( J = 0, \, d = 1, \, p_1 = p \) and \( a_1 = 1 \), we have \( T(H_n) = H_n^{-1}(p) \). This statistic is frequently referred to as a \( U \)-quantile in the literature.

We will call the statistics in (1.1) and (1.2) incomplete generalized \( L \)-statistics (IGL-statistics), in that they combine cases 2 and 3 as generalizations of \( U \)-statistics. Further, we say incomplete \( U \)-quantile (IUQ) when \( J = 0, \, d = 1 \) and \( a_1 = 1 \). Actually, an IUQ was considered by Brown and Kildea (1978) for the well-known Hodges–Lehmann kernel \( [q = 1, \, m = 2, \, h(x_1, x_2) = (x_1 + x_2)/2] \).

One motivation for introducing IGL-statistics is to obtain estimators which are faster to compute than the corresponding GL-statistic. Note that the general GL-statistic requires \( O(n^m \log n) \) operations [due to the sorting of all \( h(X_I) \)], whereas the \( U \)-quantiles can be computed in \( O(n^m) \) time, since order statistics of \( M \) observations can be found in \( O(M) \) steps [Blum, Floyd, Pratt, Rivest and Tarjan (1973)]. The corresponding numbers for the IGL-estimators and IUQs are \( O(N(n) \log N(n)) \) and \( O(N(n)) \), respectively. If \( N(n) \ll n^m \), we have obtained a more easily computed estimator. In fact, we will show in Section 3 that it is possible to have \( N(n) = O(n) \) and efficiency arbitrarily close to 1 for the IGL-estimator w.r.t. the corresponding GL-estimator. Also, as soon as \( N(n)/n \to \infty \), the efficiency becomes 1.
Having said this, we must also mention that, for some kernels of special form, faster $O(n \log n)$ algorithms have been found for the corresponding $U$-quantile. For instance, Shamos (1976) and Johnson and Mizoguchi (1978) consider such algorithms for the Hodges–Lehmann location kernel (Example 1), Croux and Rousseeuw (1992) for the spread kernel (Example 2) and Cole, Salowe, Steiger and Szemerèdi (1987), Matousek (1991) and Dillencourt, Mount and Netanyahu (1992) treat the Theil–Sen kernel for the slope in simple linear regression (Example 3). However, each kernel requires a separate algorithm. The IGL-statistics provide an alternative and general way of finding easily computable estimators of $T(H_p)$.

The paper is organized as follows. In Section 2 we define three types of designs that will be used in the following discussion. An invariance principle for the process $\sqrt{n} (H_n - H_p)$ is proved in Section 3 and asymptotic normality for $T(H_n)$ in Section 4. In Section 5 we discuss how the results in Sections 2 to 4 reduce for symmetric kernels. Some examples are given in Section 6. In particular, we give a semiparametric example: robust estimation of the scale parameter in nonparametric regression with homoscedastic errors. The finite sample efficiencies of the proposed scale estimators are compared with theoretical asymptotic limits in a Monte Carlo simulation (Section 7). Generalizations to functionals other than $L$, multivariate kernels, robustness of designs, dependent data and recursive estimation are discussed in Section 8. Finally, the proofs are gathered in the Appendix.

2. Designs. We call a design balanced if each $1 \leq i \leq n$ occurs in equally many $I \in D_n$, symmetric if $I \in D_n \Rightarrow \pi I \in D_n$ for all $m!$ permutations $\pi$ of the elements in $I$ and asymmetric if $I \in D_n \Rightarrow \pi I \not\in D_n$ for all $\pi$ different from the identity permutation. (Note that a design may be neither symmetric nor asymmetric.) For scalars in $\mathbb{N}$, the symbols $\oplus$ and $\ominus$ will denote addition and subtraction modulus $n$, and for vectors, these operations are interpreted elementwise. The following five designs will be considered in the paper.

D1. Random sampling with replacement. This means that the number of occurrences of different $I \in S_n(m)$ follows a multinomial distribution with parameters $N$ and $1/n(m), \ldots, 1/n(m)$.

D2. Designs based on cyclic permutations. Given a positive integer $\gamma_n$, define vectors $i = (i, \ldots, i)$, $d_1 = (d_1, \ldots, d_m)$, $d_\gamma = (d_1, \ldots, d_\gamma, d_{\gamma+1}, \ldots, d_m)$ of length $m$, so that all $d_j \oplus d_{j\prime}$, $j \neq j\prime$, are different, $0 \leq d_{j1} < \cdots < d_{jm}$ and $d_{1m} < \cdots < d_{\gamma m}$. Then put $D_n = \{i \oplus d_j; 1 \leq i \leq n, 1 \leq j \leq \gamma_n\}$. Notice that $\gamma_n = N(n)/n$. Examples are:

$m = 2$, $d_k = (0, k)$, $k = 1, \ldots, \gamma_n$;
$m = 3$, $\gamma_n = 1$ and $d_1 = (0, 1, 3)$;
$m = 4$, $\gamma_n = 1$ and $d_1 = (0, 1, 4, 6)$;
$m = 3$, $\gamma_n = 2$, $d_1 = (0, 1, 3)$ and $d_2 = (0, 4, 9)$.

D3. Balanced asymmetric designs, $m = 2$. Let $\gamma_n$ be a positive integer. Each $1 \leq i \leq n$ occurs in $2\gamma_n$ distinct $I \in D_n$. Given $(i, j) \in D_n$, we have
(\( j, i \) \( \in \mathbb{D}_n \) and for each \( i \) there are \( \gamma_n \) distinct elements in \( \mathbb{D}_n \) of the form \((i, k)\) and \( \gamma_n \) distinct elements of the form \((k, i)\). Notice that \( \gamma_n = N(n)/n \). As a special case, we have the design

\[ \mathbb{D}_n = \{i, j\}; 0 < j \otimes i \leq \gamma_n \} \]

D4. Recursive design based on cyclic permutations. Here \( \mathbb{D}_n = \{i + d_j; 1 \leq j \leq \tilde{\gamma}_n, 1 \leq i \leq n - d_{jm}\} \).

D5. Recursive design, \( m = 2 \). In this design \( \mathbb{D}_n = \{(i, j); 1 \leq i < j \leq n, j - i \leq \tilde{\gamma}_n\} \).

These three and many other designs have been considered in the \( U \)-statistics literature, for instance, random designs by Janson (1984) and balanced designs for \( m = 2 \) by Blom (1976) and Brown and Kildea (1978). See also Lee (1990), Section 4.3, for an overview and further references. However, the designs presented here slightly generalize those considered in the incomplete \( U \)-statistics literature, in that we let \( I \) range over all \( n_{(m)} \) possible \( m \)-tuples of elements, not only all \( \binom{n}{m} \) subsets with \( m \) elements. For instance, the distinction between symmetric and asymmetric designs is important in our setting.

The random design \( \mathbb{D}_1 \) will in general be unbalanced, whereas both \( \mathbb{D}_2 \) and \( \mathbb{D}_3 \) are balanced and asymmetric (if \( n > 2d_{\gamma(m)} \)). Actually, design \( \mathbb{D}_2 \) is a special case of design \( \mathbb{D}_3 \) when \( m = 2 \). The reason for restricting ourselves to asymmetric designs in \( \mathbb{D}_2 \) and \( \mathbb{D}_3 \) is that they have fewer elements than the corresponding symmetric design (obtained by given any \( I \in \mathbb{D}_n \) adding all other \( \pi I \)). Another important feature of \( \mathbb{D}_2 \) is that each pair \( (i, j) \) with \( i < j \) is contained in at most one \( I \in \mathbb{D}_n \).

Designs \( \mathbb{D}_4 \) and \( \mathbb{D}_5 \) have several important properties. They are obtained by removing some multi-indices from \( \mathbb{D}_2 \) and \( \mathbb{D}_3 \), respectively. This means that \( \gamma_n = N(n)/n \neq \tilde{\gamma}_n \), but \( \lim \gamma_n = \lim \tilde{\gamma}_n \equiv \gamma \), whenever either limit exists. If \( \tilde{\gamma}_n \) is a nondecreasing function of \( n \), both \( \mathbb{D}_4 \) and \( \mathbb{D}_5 \) are recursive, which means that \( \mathbb{D}_n \subset \mathbb{D}_{n+1} \). If, in addition, \( \tilde{\gamma}_n \equiv \gamma \), designs \( \mathbb{D}_4 \) and \( \mathbb{D}_5 \) are also on line, in that \( |\mathbb{D}_{n+1} \setminus \mathbb{D}_n| = O(1) \), so that the number of multi-indices added for each new observation does not increase with \( n \). Actually, \( |\mathbb{D}_{n+1} \setminus \mathbb{D}_n| = \gamma \) for all \( n \) sufficiently large. [Notice that \( S_n(m) \) is recursive but not on line, since \( |S_{n+1}(m) \setminus S_n(m)| \sim mn^{m-1} \).] We refer to a design as local if

\[ \sup\{i, j; i \neq j, i, j, i, j, i, j \in I \text{ for some } I \in \mathbb{D}_n\} \ll n. \]

Notice that design \( \mathbb{D}_5 \) is local if \( \tilde{\gamma}_n \ll n \). Similarly, design \( \mathbb{D}_4 \) is local as soon as \( d_{\tilde{\gamma}(m)} \ll n \). We could have dropped the assumption \( i < j \) in \( \mathbb{D}_5 \), still obtaining a recursive, on-line and local design. In fact, such a definition might be more natural in some applications. However, the present formulation makes \( \mathbb{D}_5 \) almost asymmetric (apart from those \( I \) with entries close to the boundaries 1 and \( n \)). In applications where \( i \) represents time, all three concepts, recursive, on-line and local, are very useful. For nonparametric applications, local designs are important. For further discussion on this, see Sections 6 and 8.
3. An invariance principle. In this section we will prove weak convergence of the process

\[ W_n(y) = \sqrt{n} (H_n(y) - H_F(y)) \]

on \((D[-\infty, \infty], \mathcal{B}_0)\), the space of right-continuous functions on \([-\infty, \infty]\) with left-hand limits, endowed with supremum norm \(\| \cdot \|_\infty\) and the \(\sigma\)-algebra \(\mathcal{B}_0\) is generated by all open balls. Notice that \(W_n\) is a \(U\)-process when \(D_n = S_n(m)\). This case has been treated by Silverman (1976, 1983) and Ruymgaart and van Zuijlen (1992). Weak convergence theory for more general \(U\)-processes indexed by functions instead of real numbers has been considered by Nolan and Pollard (1988) for \(m = 2\) and Arcones and Giné (1993) for arbitrary \(m\).

We will assume that

\[ N(n)/n = \gamma_n \to \gamma \text{ as } n \to \infty, \quad 0 < \gamma \leq \infty. \]

If \(\gamma < \infty\), \(\gamma\) can be any positive real number for the random design, but it has to be a positive integer for designs \(D_2\) and \(D_3\). [However, when \(h(\cdot)\) is symmetric, it suffices that \(2\gamma\) is a positive integer for design \(D_3\); cf. Section 5.] The case \(\gamma = 0\) could also be treated, but it is less interesting from the statistical point of view, since it results in estimators with efficiency 0.

Define \(\eta(Y) = \mathbb{1}_{\{h(Y) \leq y\}} - H_F(y)\) and notice that

\[ W_n(y) = \sqrt{n} \frac{1}{N} \sum_{i \in D_n} \eta(Y_i) \]

is an incomplete \(U\)-statistic. We will also introduce

\[ \sigma_{1i,j}^2(x, y) = \mathbb{E}(\eta_i(Y_i) \eta_j(Y_{ij})), \quad 1 \leq i, j \leq m, \]

where \(Y_i = (i, i + 1, \ldots, m, 1, \ldots, i - 1)\) is an \(i - 1\) times shifted cyclic rearrangement of \((1, \ldots, m)\) and \(Y_{ij}\) a \(j - 1\) times shifted cyclic rearrangement of \((1, m + 1, \ldots, 2m - 1)\). Put also

\[ \sigma_m^2(x, y) = \mathbb{E}(\eta(Y) \eta(Y)), \quad H_F(x) \wedge H_F(y) - H_F(x) H_F(y) \]

and

\[ \sigma_m^2(x, y) = \sum_{i, j = 1}^{m} \sigma_{1i,j}^2(x, y). \]

**Theorem 3.1.** The process \(W_n(\cdot)\) defined in (3.1) converges weakly on \(D[-\infty, \infty]\) to a zero-mean Gaussian process \(W^*(\cdot)\) with covariance function \(C(x, y) = \sigma^2(x, y)\) if \(\gamma = \infty\) in (3.2) and

\[ C(x, y) = \sigma^2(x, y) + \frac{\sigma_m^2(x, y)}{\gamma} \quad \text{design D1} \]

\[ = \sigma^2(x, y) + \frac{1}{\gamma} \left( \sigma_m^2(x, y) - \sum_{i = 1}^{m} \sigma_{1i,i}^2(x, y) \right) \quad \text{design D2} \]

\[ = \sigma^2(x, y) + \frac{1}{\gamma} \left( \sigma_m^2(x, y) - \sum_{i = 1}^{2} \sigma_{1i,i}^2(x, y) \right) \quad \text{design D3} \]
when $\gamma < \infty$. The weak convergence also holds for designs $D_4$ and $D_5$ with the same covariance function $C(\cdot, \cdot)$ as for $D_2$ and $D_3$, respectively. In particular,

\begin{equation}
\|H_n - H_F\|_\infty = O_p(n^{-1/2})
\end{equation}

for all designs.

Remark 3.1. Note that $W^* (\cdot)$ is a.s. continuous outside the countable set of discontinuities of $H_F$.

Remark 3.2. The proof of weak convergence of $W_n (\cdot)$ when $D_n = S_n (m)$ is based on writing $W_n (\cdot)$ as an average of (dependent) empirical processes, each of which is based on i.i.d. samples with distribution $H_F$ [cf. Hoeffding (1963)]. The proof of Theorem 3.1 does not make use of such a representation, which is difficult to define, at least for the random design $D_1$.

4. Asymptotics for $T(H_n)$. The $L$-statistic $T$ is first-order Gateaux differentiable with derivative

\[ d_1 T(F; G - F) = \frac{d}{d\varepsilon} T(F + \varepsilon(G - F))|_{\varepsilon = 0} \]

\[ = - \int_{-\infty}^{\infty} (G(y) - F(y)) J(F(y)) \, dy \]

\[ + \sum_{j=1}^{d} \frac{a_j p_j - G(F^{-1}(p_j))}{f(F^{-1}(p_j))}, \]

see Serfling (1984) for details. A Taylor expansion of $T$ around $H_F$ yields

\begin{equation}
T(H_n) = T(H_F) + d_1 T(H_F, H_n - H_F) + \Delta_n
\end{equation}

where

\[ A(x_1, \ldots, x_m) = d_1 T(H_F, \delta_{h(x_1, \ldots, x_n)} - H_F) \]

\[ = - \int_{-\infty}^{\infty} \left( 1_{\{h(x_1, \ldots, x_n) \leq y\}} H_F(y) \right) J(H_F(y)) \, dy \]

\[ + \sum_{j=1}^{d} \frac{a_j p_j - 1_{\{h(x_1, \ldots, x_n) \leq h_F^{-1}(p_j)\}}}{h_F(H_F^{-1}(p_j))}, \]

$h_F = H_F'$ and $\delta_x$ is the one-point distribution at $x$. Note that $T(H_n) - T(H_F)$ is an incomplete $U$-statistic plus $\Delta_n$. In Theorems 4.1 and 4.2, we will show that $\Delta_n$ is asymptotically negligible. Following Serfling (1984), we define the influence function

\[ I(x, T, F) = EA(x, X_2, \ldots, X_m) + \cdots + EA(X_1, \ldots, X_{m-1}, x) \]
and the variance
\[ \sigma^2 = EI(X_1, T, F)^2 = \int I(x, T, F)^2 \, dF(x). \]

Put also, in analogy with (3.3) and (3.4),
\[ \sigma^2_{ij} = E\{A(X_i) A(X_j)\}, \]
where \( I_i \) and \( J_j \) are defined as in Section 3, and
\[ \sigma_m^2 = EA(X_1, \ldots, X_m)^2. \]

Note that \( \sigma^2 = \sum_{i,j=1}^m \sigma_{ij}^2 \). We will assume that
\[ (4.2) \quad \sigma^2 > 0, \]
\[ (4.3) \quad \sigma_m^2 < \infty. \]

For \( H_F \) and the score function \( J \), introduce the following conditions:
\[(A) \quad H_F \text{ has positive derivative at } H_F^{-1}(p_j), \ j = 1, \ldots, d. \]
\[(B) \quad j(H_F(y)(1 - H_F(y)))^{1/2} \, dy < \infty. \]
\[(C) \quad J = 0 \text{ outside } [\alpha, \beta], \text{ where } 0 < \alpha < \beta < 1, \text{ and } J \text{ is bounded and continuous a.e. Lebesgue and } H_F^{-1}. \]
\[(D) \quad J \text{ is continuous on } [0, 1]. \]

The asymptotic behavior of \( T(H_n) \) is resolved by the following two theorems.

**Theorem 4.1.** Assume that \((3.2), (4.2), (4.3), (A) \) and \((C) \) hold. Then \( \Delta_n = o_p(n^{-1/2}) \) in \((4.1) \) and \( T(H_n) \) is an asymptotically normal estimator of \( T(H_F) \), in the sense that
\[ \sqrt{n} \left( T(H_n) - T(H_F) \right) \Rightarrow N(0, \sigma^2(\gamma)), \]
with \( \sigma^2(\infty) = \sigma^2 \) for all three designs D1–D3 defined in Section 2, and
\[ \sigma^2(\gamma) = \sigma^2 + \frac{\sigma_m^2}{\gamma} \quad \text{ design D1} \]
\[ = \sigma^2 + \frac{\sigma_m^2 - \sum_{i=1}^m \sigma_{ii}^2}{\gamma} \quad \text{ design D2} \]
\[ = \sigma^2 + \frac{\sigma_2^2 - \sum_{i=1}^2 \sigma_{ii}^2}{\gamma} \quad \text{ design D3} \]
for \( \gamma < \infty. \) Finally, the recursive designs D4 and D5 have the same asymptotics as designs D2 and D3, respectively.
**Theorem 4.2.** The conclusions of Theorem 4.1 hold under (3.2), (4.2), (4.3), (A), (B) and (D).

**Remark 4.1.** The conditions of Theorem 4.1 (Theorem 4.2) are very similar to those of Theorem 3.1 (Theorem 3.2) in Serfling (1984). We have replaced the requirement \(0 < \sigma^2 < \infty\) by the slightly stronger (4.2)—(4.3) and also added (3.2).

**Remark 4.2.** As \(\gamma \to \infty\), \(\sigma^2(\gamma) \to \sigma^2(\infty) = \sigma^2\), the asymptotic variance of the corresponding GL-statistic with \(D_n = S_n(m)\). Hence, it suffices that \(N(n)/n \to \infty\) arbitrarily slowly, in order for the IGL-statistic to have relative efficiency 1 w.r.t. the corresponding GL-statistic. Also, by choosing \(\gamma_n \equiv \gamma\) sufficiently large, we may have \(N(n) = O(n)\) and relative efficiency arbitrarily close to 1.

**Remark 4.3.** Given \(\gamma\), the random design has higher asymptotic variance than designs D2 and D3.

**Remark 4.4.** Let \(I = (1, \ldots, m)\) and put \(A_i(x) = E(A(X_i)|X_i = x), i = 1, \ldots, m\), and \(\tilde{A}(x_i) = A(x_i) - \sum_{i=1}^{m} A_i(x_i)\). Then \(E(A_i(X_i))^2 = \sigma^2_{1i}\) and

\[
\sigma^2_m = \sum_{i=1}^{m} \sigma^2_{1i} + E\tilde{A}(X_i)^2 + 2 \sum_{i=1}^{m} E(A_i(X_i) \tilde{A}(X_i)) \geq \sum_{i=1}^{m} \sigma^2_{1i},
\]

since \(E(\tilde{A}(X_i)|X_i = x_i) = 0\) a.e. (\(F\)). However, this implies that \(\sigma^2(\gamma) \geq \sigma^2\), with equality iff \(\tilde{A}(x_1) = 0\) a.e. \(F \times \cdots \times F\).

**Remark 4.5.** Note that \(\sigma^2_m\) is the same as the asymptotic variance of an ordinary “complete” \(L\)-statistic (with \(m = 1\)) based on an i.i.d. sample with marginal distribution \(H_F\). Hence,

\[
\sigma^2_m = \int_{-\infty}^{\infty} \tilde{A}(y)^2 \ dH_F(y),
\]

with

\[
\tilde{A}(y) = -\int_{-\infty}^{y} (1_{y \leq z} - H_F(z)) J(H_F(z)) \ dz + \sum_{j=1}^{d} \frac{p_j - 1_{(y \leq H_F^{-1}(p_j))}}{h_F(H_F^{-1}(p_j))}
\]

the “influence function” for the \(L\)-statistic. For \(U\)-quantiles,

\[
\sigma^2_m = p(1 - p)(h_F(H_F^{-1}(p)))^{-2}.
\]

The other quantities \(\sigma^2\) and \(\sigma^2_{1i}\) are more difficult to compute.

**5. Symmetric kernels.** Many kernels are symmetric; that is, \(h(x) = h(xP)\) for all \(m \times m\) permutation matrices \(P\). For \(U\)-statistics we may always assume this, since otherwise \(h\) may be replaced by the symmetric kernel

\[
\tilde{h}(x) = \frac{1}{m!} \sum_P h(xP),
\]

without changing \(T(H_n)\). For incomplete \(U\)-statistics, \(h\) may be replaced by \(\tilde{h}\) as soon as the design \(D_n\) is symmetric.
Typically, the IGL-statistic changes as we replace $h$ by $\tilde{h}$, even for symmetric designs. However, when the kernel is symmetric, much of the previous simplifies. First note that $A(\cdot)$ is symmetric if $h(\cdot)$ is, so the influence function becomes

$$I(x, T, F) = mA(x, X_2, \ldots, X_m)$$

and $\sigma^2_{1i} = \sigma^2_i = \sigma^2/m^2$. For symmetric kernels, it is more convenient to consider the reduced class of multi-indices

$$\tilde{S}_n(m) = \{I = (i_1, \ldots, i_n) \in S_n(m); i_1 < \cdots < i_n\},$$

with

$$|\tilde{S}_n(m)| = \binom{n}{m}.$$ 

Notice that, for any design $D_n = \{I_1, \ldots, I_n\} \subset S_n(m)$, there exists a corresponding design $\tilde{D}_n \subset \tilde{S}_n(m)$ defined through

$$\tilde{D}_n = \{(I_1), \ldots, (I_N)\},$$

where $(I)$ has the components of $I$ listed in ascending order. Observe that $|D_n| = |\tilde{D}_n|$ for any asymmetric design $D_n$, and the empirical distributions formed by $\{h(X_I); I \in D_n\}$ and $\{h(X_I); I \in \tilde{D}_n\}$ are the same. For the three designs in Section 2 we obtain the following characterizations:

1. The design $\tilde{D}_n$ is obtained by random sampling with replacement from $\tilde{S}_n(m)$.
2. $\tilde{D}_n = \{(I); I = i \oplus d_j, 1 \leq i \leq n, 1 \leq j \leq \gamma\}$.
3. Each index $i$ occurs in $2\gamma$ different $I \in \tilde{D}_n$; that is, $\tilde{D}_n$ is balanced. It suffices that $2\gamma$ is a positive integer in this case, and hence that $2\gamma$ is a positive integer if $\gamma < \infty$.

The asymptotic variances for designs D1–D3 become

$$\sigma^2(\gamma) = \sigma^2 + \frac{\sigma^2_{m'}}{\gamma} \quad \text{design D1}$$

$$= \sigma^2 + \frac{\sigma^2_m - \sigma^2/m}{\gamma} \quad \text{design D2}$$

$$= \sigma^2 + \frac{\sigma^2_{2m} - \sigma^2/2}{\gamma} \quad \text{design D3.}$$

Similarly, the covariance function of the limiting Gaussian process $W^*$ in Theorem 3.1 simplifies to

$$C(x, y) = \sigma^2(x, y) + \frac{\sigma^2_{m'}(x, y)}{\gamma} \quad \text{design D1}$$

$$= \sigma^2(x, y) + \frac{1}{\gamma} \left( \sigma^2_m(x, y) - \frac{\sigma^2(x, y)}{m} \right) \quad \text{design D2}$$

$$= \sigma^2(x, y) + \frac{1}{\gamma} \left( \sigma^2_{2m}(x, y) - \frac{\sigma^2(x, y)}{2} \right) \quad \text{design D3,}$$
with $\sigma^2_m(x, y) = H_p(x) \wedge H_p(y) - H_p(x)H_p(y)$ (as before) and $\sigma^2(x, y) = \text{Cov}(\eta_i(X_i), \eta_j(X_j))$, with $|I \cap J| = 1$.

There is a nice interpretation of $W^*$ for the random design $D_1$. Then

$$W^*(y) = W_1^*(y) + \frac{1}{\sqrt{\gamma}} W_2^*(y),$$

where $W_1^*$ and $W_2^*$ are independent Gaussian processes with covariance functions $\sigma^2(\cdot, \cdot)$ and $\sigma^2_m(\cdot, \cdot)$ [i.e., $W_2^*$ is a (transformed) Brownian bridge]. Here $W_1^*$ is the weak limit of $W_n$ when $\gamma = \infty$, for example, the $U$-process case $D_n = S_n(m)$. On the other hand, the Brownian bridge $W_2^*$ is the weak limit when $\gamma = 0$. Then $W_n$ is essentially an empirical process [which corresponds to all $I \in D_n$ being disjoint and the quantities $h(X_i)$ i.i.d.]. Typically, for “well-behaved” $h$ and $F$,

$$E(W_1^*(y) - W_1^*(x))^2 \leq C(y - x)^2,$$

$$E(W_2^*(y) - W_2^*(x))^2 \leq C|y - x|$$

for some constant $C$. This means that $W_1^*$ is smoother than $W_2^*$. Intuitively, this phenomenon can be explained by the fact that $W_1^*$ is the weak limit of $W_n$ when the latter process has a large number $N(n) \gg n$ of jumps of size $1/N$. By contrast, $W_2^*$ is the limit when $W_n$ has few ($\ll n$) jumps of larger order.

6. Examples. The following examples treat estimation of location, scale and regression functions. There are many other kernels of statistical interest; see Lee (1990).

Example 1 (Extended Hodges–Lehmann location estimator in the location–scale model). Assume that the common distribution of $X_i$ is $F_0((\cdot - \mu)/\tau)$, where $\mu$ and $\tau$ are unknown location and scale parameters and $F_0$ is a known distribution. Here $q = 1$, and $h(x_1, \ldots, x_m) = \sum_{i=1}^m \lambda_i x_i$ can be used as a location kernel for any nonnegative $\lambda_i$ with $\sum_1^m \lambda_i = 1$. The choice $\lambda_i = 1/m$ seems most natural and results in the sample mean if $J = 1$, $d = 0$, and the generalized Hodges–Lehmann estimator [Serfling (1984) and Choudhury and Serfling (1988)] when $J = 0$, $d = 1$ and $p = 0.5$.

Example 2 (Scale estimation in location–scale models). A quite general class of kernels for estimating the scale in the previous example has the form ($q = 1$)

$$h(x_1, \ldots, x_m) = c \left| \sum_{i=1}^m \lambda_i x_i \right|^r,$$

with $\sum_1^m \lambda_i = 0$, $r \geq 1$ [cf. also Example (ii) in Choudhury and Serfling (1988)]. Here $c$ is a multiplicative constant chosen so that $T(H_p) = \tau^r$. (For instance, if $F_0$ is the standard normal distribution, $\tau$ equals the standard deviation.)
The $U$-statistic based on the symmetric kernel $(x_1 - x_2)^2/2$ gives the sample variance. Generalized $L$-statistics based on this kernel and $c|x_1 - x_2|$ are discussed by Bickel and Lehmann (1979) and Rousseeuw and Croux (1993). Other choices of kernels for $m > 2$ are treated by Rousseeuw and Croux (1992).

**Example 3** (Slope and intercept estimation in simple linear regression). Assume that $Y_i = \alpha + \beta U_i + e_i$, $i = 1, \ldots, n$, with $\alpha$ and $\beta$ the intercept and slope, $\{U_i\}$ i.i.d. explanatory variables, $\{Y_i\}$ response variables and $\{e_i\}$ i.i.d. error terms with common distribution $F_0(\cdot; \tau)$. We put $X_i = (U_i, Y_i)$, so that $q = 2$. Let $(h_1(x_1, \ldots, x_m), h_2(x_1, \ldots, x_m))$ be the LS-estimate of $(\alpha, \beta)$ computed from $x_i = (u_i, y_i)$, $i = 1, \ldots, m$. When $m = 2$ we have

$$(h_1(x_1, x_2), h_2(x_1, x_2)) = \left( \frac{u_1 y_2 - u_2 y_1}{u_1 - u_2}, \frac{y_1 - y_2}{u_1 - u_2} \right).$$

The median of all $(Y_i - Y_j)/(U_i - U_j)$, $i < j$ is the Theil–Sen estimator [Theil (1950) and Sen (1968)]. Various kinds of trimmed means based on $h_2(\cdot, \cdot)$ are reviewed and analyzed by Frees (1991).

**Example 4** (Scale estimation in simple linear regression). Given three observations $X_1, X_2, X_3$ with $U_1 \leq U_2 \leq U_3$ and $U_1 < U_3$. Define $\varepsilon(X_1, X_2, X_3)$ as the difference between $Y_2$ and the point at $U_2$ on the straight line joining $X_1$ and $X_3$. One finds that $\varepsilon(x_1, x_2, x_3)$ has the form

$$\varepsilon(x_1, x_2, x_3) = \frac{u_3 - u_2}{u_3 - u_1} y_1 + \frac{u_2 - u_1}{u_3 - u_1} y_3 - y_2$$

$$= \frac{u_3 - u_2}{u_3 - u_1} e_1 + \frac{u_2 - u_1}{u_3 - u_1} e_3 - e_2,$$

and it is independent of $\alpha$ and $\beta$. [If $u_1 = u_2$ we can put $\varepsilon(x_1, x_2, x_3) = y_1/2 + y_3/2 - y_2$.] Assume that the data are relabeled so that $U_1 \leq \cdots \leq U_n$. Rousseeuw and Hubert (1993) consider an IUP based on the kernel $h(x_1, x_j, x_k) = c|\varepsilon(x_i, x_j, x_k)|$ for defining regression-free estimators of $\tau$. The constant $c$ is chosen so that $T(H_{\varepsilon}) = \tau$. Note that $h(X_1, X_2, X_3) = h(e_1, e_2, e_3)$, where $h(\cdot, \cdot, \cdot)$ has the form (6.1), with $r = 1$, $m = 3$ and constants $\lambda_i$ depending on $(U_1, U_2, U_3)$. Several designs are considered by Rousseeuw and Hubert (1993), among others $D_n = \hat{S}_n(m)$ and

$$D_n = \{(1, 2, 3), (2, 3, 4), \ldots, (n - 2, n - 1, n)\},$$

which is related to design D4. The asymptotic theory of Section 4 is not directly applicable to these designs, since the explanatory variables are ordered, and therefore $D_n$ depends on $(X_i)$.

By contrast, for fixed explanatory variables satisfying $U_1 \leq \cdots \leq U_n$, we put $X_i = Y_i$, and then $D_n$ is independent of $(X_i)$. Now, the kernel $h(X_i)$ does not depend on $X_i$ alone, but also on $U_i$, in all but special cases. One exception
is equispaced \( \{U_i\} \) and a design
\[(6.4) \quad D_n = \{(1, 2, 4), (2, 3, 5), \ldots, (n - 3, n - 2, n)\}\]
of type D4 with \( \gamma = 1 \). Then \( h(X_i, X_{i+1}, X_{i+3}) = c[2e_i/3 + e_{i+3}/3 - e_{i+1}] \).

Application of Theorems 4.1 and 4.2 gives
\[\sqrt{n}(T(H_n) - \tau) \rightarrow \mathcal{N} \left(0, \sigma^2 + \sigma^2_3 - \sum_{i=1}^{n} \sigma^2_{iii} \right),\]

**EXAMPLE 5** (Scale estimation in nonparametric regression). Suppose that \( q = 1 \) and \( X_i = m(U_i) + e_i, \ i = 1, \ldots, n, \) where \( \{e_i\} \) are i.i.d. with common distribution \( F(\cdot) = F_0(\cdot/\tau), \ 0 \leq U_1 \leq \cdots \leq U_n \leq 1, \) are design points (non-random or random) and \( m: [0, 1] \rightarrow \mathbb{R} \) is an unknown regression function. Rice (1984) considers estimating \( \tau \) through
\[\hat{\tau}_n = \sqrt{n} \left(1 \over (n - 1) \sum_{i=2}^{n} (X_i - X_{i-1})^2 \right),\]
which is actually an incomplete version of the sample standard deviation using design D5, with \( \gamma_n = \gamma = 1 \) [and \( \gamma_n = (n - 1)/n \)]. It has asymptotic efficiency 0.667 compared to the sample standard deviation in the location–scale model for normal errors. By increasing \( \gamma \), the efficiency quickly approaches 1. Note that a local design is appropriate here, since we want to include only points \( h(X_i, X_j) \) for which \( U_i \) and \( U_j \) are close; otherwise the regression function \( m \) may cause a large bias.

We will now treat \( L \)-functionals other than the mean. Consider an IGL-estimate of \( \tau \) based on the kernel \( h(x_1, x_2) = c|x_1 - x_2| \). Let \( H_n \) be the e.d.f. of \( (h(X_i); \ i \in D_n) \), and define
\[\hat{\tau}_n = T(H_n) = \sum_{i=1}^{n} c_{Ni}; h_{i:N},\]
with weights \( c_{Ni} \) as defined in (1.3). The constant \( c = c(T, F) \) is chosen so that \( T(H_n) = \tau \). For instance, choosing \( T(G) = G^{-1}(p) \), we obtain the IUQ:
\[\hat{\tau}_n = \hat{\tau}_n(p) = H_n^{-1}(p) = h_{[Np],1:N};.\]

What about the asymptotic properties of \( \hat{\tau}_n \)? Since \( \{X_i\} \) are not identically distributed, we cannot apply Theorems 4.1 and 4.2 directly to \( \hat{\tau}_n = T(H_n) \).

Introduce \( H_n \) as the e.d.f. of \( (h(e_i); \ i \in D_n) \). Since the error terms \( \{e_i\} \) are i.i.d., the asymptotics from Section 4 is applicable to \( T(H_n) \). Under certain assumptions, \( h(X_i) \) is close to \( h(e_i) \). Put
\[M_n = \max_{2 \leq i \leq n} (U_i - U_{i-1}),\]
for the maximal spacing of the design points. Suppose also that \( \|m\|_{\varepsilon} < \infty \) for some \( 0.5 < \varepsilon \leq 1 \), where
\[\|m\|_{\varepsilon} = \sup_{0 \leq s < t \leq 1} \frac{|m(t) - m(s)|}{|t - s|^{\varepsilon}}.\]
Then $|h(X_i, X_j) - h(e_i, e_j)| \leq c |m(U_i) - m(U_j)| \leq c \|m\|_{\infty}(\hat{\gamma}_n M_n)^\varepsilon$ for any $(i, j) \in D_n$. Suppose also that the (nonnegative) weights $c_N$, sum to 1 in (1.3) [i.e. $\int f(t) \, dt + \sum_j a_j = 1$]. This implies

$$|\hat{\tau}_n - T(\tilde{H}_n)| \leq c \|m\|_{\infty}(\hat{\gamma}_n M_n)^\varepsilon = o(n^{-1/2}) \text{ a.s.}$$

as long as

$$\hat{\gamma}_n M_n = o(n^{-1/(2\varepsilon)}) \text{ a.s.}$$

In other words, if (6.6) holds, $\hat{\tau}_n$ is asymptotically equivalent to $T(\tilde{H}_n)$, so that

$$\sqrt{n} (\hat{\tau}_n - \tau) \to \mathcal{N}(0, \sigma^2 + \frac{\sigma_2^2 - \sigma_2^2/2}{\gamma})$$

[cf. (5.1)], with $\gamma = \lim \gamma_n = \lim \hat{\gamma}_n$, $0 < \gamma \leq \infty$. When $\gamma = \infty$, the asymptotic variance $\sigma^2$ is the same as for the corresponding GL-estimator of $\tau$ in the location–scale model. For instance suppose $F_0$ is the standard normal distribution and $\tau$ the standard deviation of $e_i$. If $\gamma = \infty$, $\hat{\tau}_n(0.25)$ has asymptotic efficiency $0.827$ compared to the sample standard deviation, and $\hat{\tau}_n(0.91)$ has efficiency $0.99$ [cf. Rousseeuw and Croux (1993)]. Another advantage of $\hat{\tau}_n(p)$ is that it is robust for low values of $p$, not only against large error terms $e_i$ but also against discontinuities in $m(\cdot)$.

For equispaced points, $U_i = (i - 1/2)/n$, we have $M_n = 1/n$, so (6.6) is valid provided $\hat{\gamma}_n = o(n^{1-1/(2\varepsilon)})$. If $(U_i)$ are the order statistics of an i.i.d. sample from a distribution on $[0, 1]$ that has a density $w(\cdot)$ with $\inf_j w(t) = w > 0$, then $P(M_n > (1 + \delta)(\log n)/(\log n)) i.o. = 0$ for any $\delta > 0$, under some regularity conditions on $w$ [cf. Deheuvels (1984)].

An estimator of $\tau$ based on the residual kernel $\varepsilon(\cdot, \cdot, \cdot)$ in Example 4 and design (6.3) is considered by Gasser, Stroka and Jenner (1986). As mentioned in Example 4, this kernel in general depends on $(U_i)$. However, assume $U_i = (i - 1/2)/n$, design (6.4) and kernel $h(x_1, x_2, x_3) = c|x_1/3 + x_3/3 - x_2/2$. Let $H_n$ and $\tilde{H}_n$ be defined as above, and put

$$\tilde{\tau}_n = T(H_n)$$

and $m_i = m(U_i)$. Since $(U_i)$ are equispaced and $\varepsilon > 0.5$,

$$|h(X_i, X_{i+1}, X_{i+3}) - h(e_i, e_{i+1}, e_{i+3})|$$

$$\leq c \max_i |2m_i/3 + m_{i+3}/3 - m_{i+1}|$$

$$= O(M_n) = O(n^{-\varepsilon}) = o(n^{-1/2}).$$

For an $L$-functional $T$ with weights summing to 1, we therefore have a bias $\tilde{\tau}_n - T(\tilde{H}_n) = o(n^{-1/2})$. Again, the asymptotic behavior of $\tilde{\tau}_n$ now follows from Theorems 4.1 and 4.2:

$$\sqrt{n} (\tilde{\tau}_n - \tau) \to \mathcal{N}(0, \sigma^2 + \sum_{i=1}^3 \sigma_{3i}^2).$$

Notice that the bias is much smaller if $m$ is twice differentiable with $\|m''\|_{L(0,1)} < \infty$. Then $|\tilde{\tau}_n - T(\tilde{H}_n)| \leq c \max_i |2m_i/3 + m_{i+3}/3 - m_{i+1}| = O(n^{-3/2})$. Further, $\tilde{\tau}_n$ is unchanged if any linear function is added to $m$. 
7. A simulation example. The performance of IGL-statistics was tested for scale estimators from Examples 2 and 5 of the previous section. We considered $\hat{r}_n$ and $\hat{r}_n(p)$ with design D5 and $\gamma_{\tau} = \gamma$. Two regression functions were used: $m_1(\cdot) = 0$ (which corresponds to Example 2) and

$$m_2(u) = \begin{cases} u, & 0 \leq u < 1/2, \\ 9/2 + (u - 1/2) + 2(u - 1/2)^2, & 1/2 \leq u \leq 1, \end{cases}$$

having a jump of height 4 at $u = 1/2$. The error distribution $F_0$ was the standard normal distribution $\Phi$ (so that $\tau = 1$). The kernel of $\hat{r}_n(p)$ then takes the form $h(x_1, x_2) = c(p)|x_1 - x_2|$ with $c(p) = (\sqrt{2} \Phi^{-1}(1 + p)/2)^{-1}$.

Table 1 displays Monte Carlo estimates of finite sample efficiencies for $\hat{r}_n$, $\hat{r}_n(0.5)$ and $\hat{r}_n(0.9)$ for sample sizes $n = 30, 100$ and $\gamma = 3, 10, \infty$. Since the Cramér– Rao lower bound for the asymptotic efficiency is $1/2$,

$$\text{eff}_n = \left( \frac{2}{N_{MC}} \sum_{i=1}^{N_{MC}} (\tau_{ni} - 1)^2 \right)^{-1}$$

was used as an estimate of finite sample efficiency. Here $N_{MC}$ is the number of Monte Carlo iterates ($= 100,000$) and $\tau_{ni} [= \hat{r}_n$ or $\hat{r}_n(p)]$ is an estimate of $\tau = 1$ in the $i$th Monte Carlo trial. The theoretical asymptotic limits ($n = \infty$) are also included in Table 1, for comparison.

As expected, $\hat{r}_n$ [and also $\hat{r}_n(0.9)$] has high efficiency when $m = m_1$, because of the Gaussian error distribution. However, both of these two estimators are very sensitive to the jump in $m_2$, whereas $\hat{r}_n(0.5)$ is more robust. The finite sample efficiencies converge quickly to the asymptotic limits for $m_1$. The convergence for $m_2$ is much slower, mainly because of the jump, but also since $m_2$ is nonconstant outside the jump.

The simulations could also be extended to cover more heavy-tailed distributions. In this case we expect the good efficiency properties of $\hat{r}_n$ to break down, whereas the more robust IUQ $\hat{r}_n(p)$ is less affected by outliers.

**Table 1**

Estimated finite sample efficiencies using 100,000 Monte Carlo replicates for IGL scale estimates with design D5. The last column $\gamma = \infty$ means $\gamma_n = n - 1$ for regression function $m_1[D_n = S_n(2)]$ and $\gamma_n = [5n^{1/3}]$ for $m_2$ (local design)

<table>
<thead>
<tr>
<th>$m$</th>
<th>$n$</th>
<th>$\gamma = 3$</th>
<th>$\gamma = 10$</th>
<th>$\gamma = \infty$</th>
<th>$\gamma = 3$</th>
<th>$\gamma = 10$</th>
<th>$\gamma = \infty$</th>
<th>$\gamma = 3$</th>
<th>$\gamma = 10$</th>
<th>$\gamma = \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1$</td>
<td>30</td>
<td>0.835</td>
<td>0.914</td>
<td>0.974</td>
<td>0.524</td>
<td>0.679</td>
<td>0.774</td>
<td>0.727</td>
<td>0.850</td>
<td>0.940</td>
</tr>
<tr>
<td>100</td>
<td>0.849</td>
<td>0.935</td>
<td>0.991</td>
<td>0.533</td>
<td>0.712</td>
<td>0.836</td>
<td>0.732</td>
<td>0.887</td>
<td>0.967</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.857</td>
<td>0.952</td>
<td>1.000</td>
<td>0.534</td>
<td>0.729</td>
<td>0.864</td>
<td>0.735</td>
<td>0.898</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>$m_2$</td>
<td>30</td>
<td>0.190</td>
<td>0.028</td>
<td>0.012</td>
<td>0.368</td>
<td>0.087</td>
<td>0.024</td>
<td>0.234</td>
<td>0.018</td>
<td>0.010</td>
</tr>
<tr>
<td>100</td>
<td>0.414</td>
<td>0.095</td>
<td>0.021</td>
<td>0.479</td>
<td>0.357</td>
<td>0.102</td>
<td>0.522</td>
<td>0.145</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.857</td>
<td>0.952</td>
<td>1.000</td>
<td>0.534</td>
<td>0.729</td>
<td>0.864</td>
<td>0.735</td>
<td>0.898</td>
<td>0.992</td>
<td></td>
</tr>
</tbody>
</table>
8. Outlook.

8.1. Other functionals. In this paper we have only treated $L$-functionals. It is clear that other functionals (e.g., $M$- and $R$-functionals) could also be investigated in a similar way. Suppose now that $T$ is a measurable functional on $(D, -\infty, \infty)$ which is compactly differentiable [cf. Rieder (1991), page 2]. From the proof of Theorem 3.1 it follows that the sequence $\{\sqrt{n}(H_n - H_F)\}$ is tight, so Theorem 1.3.3 of Rieder (1991) implies

$$\sqrt{n} (T(H_n) - T(H_F)) = \sqrt{n} \sum_{i=1}^{d_1} T(H_F, H_n - H_F) + o_p(n^{-1/2})$$

Here $\delta_{h(X)}$ is the one-point distribution at $h(X)$. Hence, $T(H_n)$ is asymptotically equivalent to an incomplete $U$-statistic for any compactly differentiable functional, and asymptotic normality may be deduced from the asymptotic theory of incomplete $U$-statistics. Alternatively, we may use Theorem 3.1 and the continuous mapping theorem to deduce

$$\sqrt{n} \sum_{i=1}^{d_1} T(H_F, H_n - H_F) \to_d \sum_{i=1}^{d_1} T(H_F, W^*)$$

since $d_1 T(H_F, \cdot)$ is a continuous linear functional, by the definition of compact differentiability.

8.2. Multivariate kernels. Multivariate kernels are of interest, for example, for multivariate location estimation. Chaudhuri (1992a) has constructed a multivariate extension of the Hodges–Lehmann estimator. Given data $X_1, \ldots, X_n \in \mathbb{R}^q$, let $h: \mathbb{R}^{mq} \to \mathbb{R}^q$ be defined by $h(X) = \sum_{i=1}^{m} X_i / m$, and let $|\cdot|$ denote the Euclidean norm in $\mathbb{R}^q$. Then the statistic

$$(8.1) \quad \hat{\mu}_n = \arg \min_{\mu \in \mathbb{R}^q} \sum_{I \in \mathcal{S}_n^{m/q}} |h(X_I) - \mu'|$$

is an estimator of $\mu = \arg \min_{\mu' \in \mathbb{R}^q} E|h(X_I) - \mu'|$. It reduces to (the $m$-argument extension of) the Hodges–Lehmann estimator for $q = 1$, and it has been shown that $\hat{\mu}_n - \mu$ is asymptotically equivalent to a $U$-statistic having a certain multivariate kernel $A$ [Theorem 3.2 in Chaudhuri (1992a)]. Replacing $\mathcal{S}_n^{m/q}$ by some design $\mathcal{D}_n$ in (8.1) results in a more tractable estimator that we conjecture is asymptotically equivalent to an incomplete $U$-statistic with kernel $A$ and design $\mathcal{D}_n$. The Oja median [Oja (1983)] and the simplicial median [Liu (1990)] are two other multivariate location estimates that are asymptotically equivalent to $U$-statistics; see Arcones, Chen and Giné (1993). Again, we conjecture that the corresponding incomplete versions of these estimators are asymptotically equivalent to incomplete $U$-statistics with the same kernels as the “complete” ones.

Consider next a multiple linear regression model with explanatory vectors $U_i$ of dimension $p$, response variables $Y_i$ and an unknown vector of regression parameters $\beta$. Chaudhuri (1992b) considers certain estimators of the
regression parameters that generalize the Hodges–Lehmann estimator. Croux, Rousseeuw and Hössjer (1994) and Hössjer, Croux and Rousseeuw (1994) introduce the class of generalized S-estimators (GS-estimators) and Stromberg, Hawkins and Hössjer (1995) a related least trimmed differences estimator (LTD-estimator). For instance, the LTD-estimator is defined by minimizing the objective function

\[
\sum_{k=1}^{\lambda_n} \left\{ \left( r_i(\beta') - r_j(\beta') \right)^2; i < j \right\}_{k: \binom{n}{2}}
\]

w.r.t. \( \beta' \). Here \( r_i(\beta) = Y_i - \beta' U_i \) is the \( i \)th residual using \( \beta \) as the regression parameter and \( 1 \leq \lambda_n \leq \binom{n}{2} \). (\( T \) is the transpose of a vector.) Both the GS-, LTD- and Chaudhuri (1992b) estimators are asymptotically equivalent to multivariate \( U \)-statistics, and reduced versions of them can be obtained by considering incomplete designs. For the LTD-estimator, this corresponds to minimizing

\[
\sum_{k=1}^{\lambda_n} \left\{ \left( r_i(\beta') - r_j(\beta') \right)^2; (i, j) \in D_n \right\}_{k: N(n)}
\]

w.r.t. \( \beta' \).

8.3. Robustness. As seen in Sections 6 and 7.2, incomplete designs can be used to construct computationally more tractable robust estimators. A word of caution though; the (asymptotic) breakdown point \( \varepsilon^* \) (the smallest proportion of data that is able to carry the estimate over all bounds) may decrease using a reduced design. Consider, for instance, the Hodges–Lehmann estimator, for which \( \varepsilon^* = 1 - 1/\sqrt{2} \approx 0.293 \) [Hodges (1967)]. If design D5 is used with \( \tilde{\gamma} = \gamma = 1 \), it may be shown that \( \varepsilon^* = 0.25 \). (However, \( \varepsilon^* \rightarrow 1 - 1/\sqrt{2} \) as \( \gamma \rightarrow \infty \).) Random design D1 gives \( \varepsilon^* = 0 \).

The LTD-estimator has \( \varepsilon^* = \min(\sqrt{\lambda}, 1 - \sqrt{\lambda}) \) if \( \lambda_n/\binom{n}{2} \rightarrow \lambda \) as \( n \rightarrow \infty \), so the maximal breakdown point 0.50 is achieved with \( \lambda = 0.25 \) [cf. Croux, Rousseeuw and Hössjer (1994), Theorem 2, and Stromberg, Hawkins and Hössjer (1997), Theorem 2.1]. If design D5 is used with \( \tilde{\gamma}_n = \gamma = 1 \) and \( \lambda_n/N(n) \rightarrow \lambda \), a similar proof shows that \( \varepsilon^* = \min(\lambda, (1 - \lambda)/2) \) which attains the maximal value \( \varepsilon^* = 1/3 \) for \( \lambda = 1/3 \). Similar results hold for reduced versions of GS-estimators.

A more systematic study of reduced designs \( D_n \) that result in good robustness and efficiency is an interesting topic for future research.

8.4. Time series. Consider the AR(\( p \)) model

\[
Y_i = \alpha + \beta_1 Y_{i-1} + \cdots + \beta_p Y_{i-p} + e_i,
\]
where $e_i$ are i.i.d. innovations with distribution $F_0(\cdot/\tau)$. The kernels of Examples 1–4 (and their multivariate extensions) may be used for estimating $\alpha, \beta = (\beta_1, \ldots, \beta_p)$ and $\tau$. A regression-free IGL-estimator of $\tau$ may be constructed, using $X_i = (Y_{i-p}, \ldots, Y_i)$, $m = p + 2$ and

$$h(X_i) = \text{vertical height of simplex formed by } \{X_i; i \in I\}.$$ 

This kernel was proposed by Rousseeuw and Hubert (1993) in the multivariate regression setup. The design for the IGL-estimator does not have to be local in terms of $X$, even though each $X_i$ is a local block of $Y_i$'s. It would be interesting with an asymptotic theory for dependent data, which is needed for this example.

8.5. Recursive estimation of $M$-functionals. Consider an $M$-functional, implicitly defined by

$$\int \psi(y - T(G)) \, dG(y) = 0,$$

given some odd function $\psi : \mathbb{R} \to \mathbb{R}$. The recursive designs D4 and D5 with $\tilde{\gamma}_n = \gamma$ may be used to construct on-line estimators of $\theta = T(H_p)$. Notice that $|D_n \setminus D_{n-1}| = \gamma$ for $n$ large enough. Hence, we propose the recursive scheme

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \frac{1}{n\gamma b_n} \sum_{I \in D_n \setminus D_{n-1}} \psi(h(X_i) - \hat{\theta}_{n-1})$$

for estimating $\theta$, with $b_n$ a recursive estimator of $|\psi(y - \theta)dH_p(y)$. Under the appropriate assumptions, we conjecture that this estimator is asymptotically equivalent to $T(H_n)$ with $H_n$ computed from the same design D4 or D5. In fact, this is proved in Hössjer (1997) for the special case of quantiles $[T(G) = G^{-1}(p)$ and $\psi(x) = p - 1_{\{x \leq 0\}}].$

APPENDIX

We now give the proofs of the results from Sections 3 and 4.

PROOF OF THEOREM 3.1. We first treat designs D1–D3. The convergence of finite-dimensional distributions of $(W_n(\cdot))$ follows from the Cramer–Wold device and the asymptotic theory for incomplete $U$-statistics; see Janson (1984) for design D1, Lee (1990), pages 212–213, for D2 and Brown and Kildea (1978) for D3. Even though these results are formulated for symmetric kernels only, the generalization to nonsymmetric kernels is straightforward. [Notice that $\sum_{i} \lambda_i W_i(y_{\cdot})$ is an incomplete $U$-statistic.]

For proving weak convergence, it remains to establish tightness. For any function $Z(\cdot) \in D[\infty, \infty]$ we define the modulus of continuity:

$$\omega_Z(\delta) = \sup_{x, y; |H_p(y) - H_p(x)| \leq \delta} |Z(y) - Z(x)|.$$ 

We will prove that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P(\omega_{W_n(\cdot)}(\delta) \geq 3\epsilon) = 0.$$ 

(A.1)
Tightness then follows from (A.1); see the proof of Theorem A in Silverman (1983). Let $\Xi_n = \{\xi_0, \xi_1, \ldots, \xi_{L_n}\}$ be a grid in $[-\infty, \infty]$ with $\xi_0 = -\infty, \xi_{L_n} = \infty$ and

\[
H_F(\xi_i) - H_F(\xi_{i-1}) \leq \frac{C_1}{n},
\]

for $i = 1, \ldots, L_n$. (The dependence of $\xi_i$ on $n$ is suppressed in the notation.) It follows that $L_n \leq n$. Introduce next

\[
H_{in} = H_n(\xi_i - ) - H_n(\xi_{i-1}), \quad i = 1, \ldots, L_n.
\]

Given any $x, y \in \mathbb{R}$, we may find $\xi_{i-1}, \xi_{j-1} \in \Xi_n$ such that $0 \leq H_F(x) - H_F(\xi_{i-1}) \leq 1/n$, $0 \leq H_n(x) - H_n(\xi_{i-1}) \leq H_{in}$, $0 \leq H_F(y) - H_F(\xi_{j-1}) \leq 1/n$ and $0 \leq H_n(y) - H_n(\xi_{j-1}) \leq H_{jn}$. Hence,

\[
|W_n(y) - W_n(x)| \leq |W_n(\xi_{j-1}) - W_n(\xi_{i-1})| + \sqrt{n} \left( H_{jn} + H_{in} \right) + \frac{2}{\sqrt{n}}.
\]

Observe also that $|H_F(y) - H_F(x)| \leq \delta$ implies $|H_F(\xi_{j-1}) - H_F(\xi_{i-1})| \leq 2\delta$, provided $n$ is so large that $1/n \leq \delta$. This implies

\[
\omega_{\tilde{w}}(\delta) \leq \tilde{\omega}_{\tilde{w}}(2\delta) + 2\sqrt{n} \max_{1 \leq i \leq L_n} H_{in} + \frac{2}{\sqrt{n}},
\]

where

\[
\tilde{\omega}_x(\delta) = \sup_{x, y; |H_F(y) - H_F(x)| \leq \delta} |Z(y) - Z(x)|.
\]

We will prove below that

\[
E(W_n(y) - W_n(x))^4 \leq \frac{C_1}{n} |H_F(y) - H_F(x)| + C_2 (H_F(y) - H_F(x))^2
\]

for some constants $C_1$ and $C_2$ (not depending on $x$, $y$ or $n$), which in turn implies

\[
E(W_n(y) - W_n(x))^4 \leq C (H_F(y) - H_F(x))^2, |H_F(y) - H_F(x)| \geq \frac{1}{n},
\]

with $C = C_1 + C_2$. Put $W_{n1} = W_n(\xi_i - ) - W_n(\xi_{i-1})$. Then, if $n$ is so large that $2/\sqrt{n} \leq \epsilon/2$, 

\[
P\left( 2\sqrt{n} \max_{1 \leq i \leq L_n} H_{ni} \geq \epsilon \right) \leq \sum_{i=1}^{L_n} P\left( 2\sqrt{n} H_{ni} \geq \epsilon \right)
= \sum_{i=1}^{L_n} \left[ 2W_{ni} \geq \epsilon - 2\sqrt{n} \left( H_F(\xi_i) - H_F(\xi_{i-1}) \right) \right]
\leq \sum_{i=1}^{L_n} \left[ 2W_{ni} \geq \epsilon - \frac{2}{\sqrt{n}} \right]
\leq \sum_{i=1}^{L_n} \left[ W_{ni} \geq \frac{\epsilon}{4} \right] \leq n \frac{C(1/n)^2}{(\epsilon/4)^4} \to 0
\]

as \( n \to \infty \), where \( C \) is the same constant as in (A.5). Next, we will prove that
\[
\lim_{\delta \to 0} \lim_{n \to \infty} P\left( \tilde{\omega}_{\lambda_n}(2\delta) \geq \epsilon \right) = 0.
\]

To this end, let \( \Theta_{n \delta} = \{\theta_0, \theta_1, \ldots, \theta_{L_n}\} \) be a subset of \( \Xi_n \) with \( \theta_0 = -\infty \), \( \theta_{L_n} = \infty \) and
\[
H_F(\theta_i) - H_F(\theta_{i-1}) \leq 3\delta,
H_F(\theta_i) - H_F(\theta_{i-1}) \geq 3\delta
\]
for \( i = 1, \ldots, L_n \). Observe that \( L_n \leq 1/(3\delta) \). Then define sets \( K_i, \ldots, K_{L_n} \) by \( K_i = [\theta_{i-1}, \theta_i] \cap \Xi_n \) if \( H_F(\theta_i) - H_F(\theta_{i-1}) \leq 2\delta \) and \( K_i = [\theta_{i-1}, \theta_i] \cap \Xi_n \) if \( H_F(\theta_i) - H_F(\theta_{i-1}) > 2\delta \). Suppose \( x \neq y \in \Xi_n \) and \( |H_F(y) - H_F(x)| \leq 2\delta \). Then \( x \in K_i \) and \( y \in K_j \) for some \( i, j \) with \( |i - j| \leq 1 \). Moreover, if \( |i - j| = 1 \), then \( K_i \cap K_j \neq \emptyset \). It follows that
\[
P\left( \tilde{\omega}_{\lambda_n}(2\delta) \geq \epsilon \right) \leq \sum_{i=1}^{L_n} \left[ \sup_{x, y \in K_i} |W_n(y) - W_n(x)| \geq \frac{\epsilon}{3} \right].
\]

If \( x, y \in K_i \), then \( |H_F(y) - H_F(x)| \leq 5\delta \). Therefore, (A.2), (A.5) and Theorem 12.2 in Billingsley (1968) [with \( \gamma = 4 \), \( \alpha = 2 \) and \( u_i = H_F(\xi_i) \) in that theorem] imply that, for some constant \( \bar{C} \),
\[
P\left( \sup_{x, y \in K_i} |W_n(y) - W_n(x)| \geq \frac{\epsilon}{3} \right) \leq \frac{\bar{C}(5\delta)^2}{(\epsilon/3)^4},
\]
which implies (A.7), since
\[
P\left( \tilde{\omega}_{\lambda_n}(2\delta) \geq \epsilon \right) \leq L_n \frac{\bar{C}(5\delta)^2}{(\epsilon/3)^4} \leq \frac{\bar{C}25\delta}{3(\epsilon/3)^4}.
\]

Formula (A.1) now follows from (A.3), (A.6) and (A.7). It remains to prove (A.4) to establish tightness. We will indicate how this is done for the random design D1, the other cases being similar. Assume \( x < y \) and put, for ease of notation, \( \eta(x_i) = \eta(x_i) - \eta(x_i) \). Assume that \( I_1, \ldots, I_N \) are i.i.d. multi-
indices, uniformly distributed over \( S_n(m) \), and write

\[
W_n(y) - W_n(x) = \frac{\sqrt{n}}{N} \sum_{j=1}^{N} \eta(X_{i_j}).
\]

Then

\[
E(W_n(y) - W_n(x))^4 = \frac{n^2}{N^4} \sum_{j_1, \ldots, j_4=1}^{N} E\left( \prod_{k=1}^{4} \eta(X_{i_{j_k}}) \right)
\]

(A.8)

\[
= \frac{n^2}{N^4} \left( N\alpha_1 + 3N(N - 1)\alpha_2 
+ 4N(N - 1)\alpha_3 + 6N(N - 1)(N - 2)\alpha_4 
+ N(N - 1)(N - 2)(N - 3)\alpha_5 \right),
\]

where

\[
\alpha_1 = E\left( \eta(X_{i_1}) \right)^4 \leq H_F(y) - H_F(x),
\]

\[
\alpha_2 = E\left( \eta(X_{i_1})^2 \eta(X_{i_2})^2 \right)
\leq \frac{C_{21}}{n}(H_F(y) - H_F(x)) + (H_F(y) - H_F(x))^2,
\]

\[
\alpha_3 = \left| E\left( \eta(X_{i_1})^3 \eta(X_{i_2}) \right) \right|
\leq \frac{C_{31}}{n}(H_F(y) - H_F(x)),
\]

\[
\alpha_4 = \left| E\left( \eta(X_{i_1})^2 \eta(X_{i_2}) \eta(X_{i_3}) \right) \right|
\leq \frac{C_{41}}{n^2}(H_F(y) - H_F(x)) + \frac{C_{42}}{n}(H_F(y) - H_F(x))^2
\]

and

\[
\alpha_5 = \left| E\left( \eta(X_{i_1}) \eta(X_{i_2}) \eta(X_{i_3}) \eta(X_{i_4}) \right) \right|
\leq \frac{C_{51}}{n^2}(H_F(y) - H_F(x)) + \frac{C_{52}}{n^2}(H_F(y) - H_F(x))^2.
\]

Formula (A.4) follows now from (A.8)–(A.13) and the fact that \( \gamma > 0 \) in (3.2). The estimates in (A.8)–(A.13) are deduced from the i.i.d. assumption on \( \{I_j\} \). For instance,

\[
\alpha_2 = \frac{1}{n_{(m)}} \sum_{I, J \in S_{n(m)}} E\left( \eta(X_{I})^2 \eta(X_{J})^2 \right)
\leq \frac{n_{(m)}^2 - n_{(m)}(n - m)_{(m)}}{n_{(m)}^2} E\left( \eta(X_{I})^4 \right) + \frac{n_{(m)}(n - m)_{(m)}}{n_{(m)}^2} \left( E\eta(X_{I})^2 \right)^2.
\]
Here \( n_{(m)}(n - m)_{(m)} \) is the number of pairs \((I, J)\) with \( I \cap J = \emptyset \) and \( n_{(m)}(n - m)_{(m)} = O(n^{2m-1}) \) the number of pairs with \( I \cap J \neq \emptyset \). Formula (A.10) follows since \( E(\eta(X_I))^2 \leq H_p(y) - H_p(x) \) and \( E(\eta(X_I))^3 \leq H_p(y) - H_p(x) \).

Consider now designs D4 and D5. Observe first that \( ||W_n^{(D4)} - W_n^{(D5)}||_\infty \) and \( ||W_n^{(D3a)} - W_n^{(D5)}||_\infty \) are both \( O(1/\sqrt{n}) \) if \( \gamma < \infty \), so the result then follows from what we have already proved. If \( \gamma = \infty \), it is easy to see that, for each \( y \in \mathbb{R} \), \( E(W_n^{(D4)}(y) - W_n^{S,\gamma}(m)(y))^2 \to 0 \) and \( E(W_n^{(D5)}(y) - W_n^{S,\gamma}(m)(y))^2 \to 0 \), where \( W_n^{S,\gamma}(\cdot) \) is the process corresponding to \( D_n = S_n(m) \). This proves convergence of finite-dimensional distributions, using the asymptotic theory of \( U \)-statistics and the Cramér–Wold device for the finite-dimensional distributions of \( W_n^{S,\gamma}(\cdot) \). The inequality (A.4) is established analogously for \( W_n^{(D4)}(\cdot) \) and \( W_n^{(D5)}(\cdot) \), and this proves tightness.

Finally, (3.7) follows from the weak convergence of \( W_n \) to \( W^* \) and the continuous mapping theorem:

\[
\sqrt{n} ||H_n - H_F||_\infty = ||W_n||_\infty \to ||W^*||_\infty = O_p(1). \]

**Proofs of Theorems 4.1 and 4.2.** We will show that \( \Delta_n = o_p(n^{-1/2}) \), by adapting the proof of Serfling (1984). This implies that \( T(H_n) - T(H_F) \) is asymptotically equivalent to an incomplete \( U \)-statistic according to (4.1), and the rest of the theorem(s) will follow by the asymptotic theory for incomplete \( U \)-statistics, as in the proof of Theorem 3.1.

Write \( T(H_F) = T_1(H_F) + T_2(H_F) \), where \( T_1(H_F) = \int_0^t J(t)H_F^{-1}(t) \, dt \), \( T_2(H_F) = \sum_{i=1}^d a_i H_F^{-1}(p_i) \). Decompose the remainder into two terms, \( \Delta_n = \Delta_{1n} + \Delta_{2n} \), where \( \Delta_{1n} = T_1(H_n) - T_1(H_F) = d_1 T_1(H_F; H_n - H_F) \). Then

(A.14) \[
\Delta_{1n} = -\int_{-\infty}^{\infty} V_{H_n, H_F}(y)(H_n(y) - H_F(y)) \, dy,
\]

where

\[
V_{G_1, G_2}(y) = \begin{cases} \frac{K(G_2(y)) - K(G_1(y))}{G_1(y) - G_2(y)} - J(G_2(y)), & G_1(y) \neq G_2(y), \\ 0, & G_1(y) = G_2(y), \end{cases}
\]

and \( K(u) = \int_0^u J(t) \, dt \). Now (A.14) implies

(A.15) \[
|\Delta_{1n}| \leq ||V_{H_n, H_F}||_1 ||H_n - H_F||_\infty
\]

and

(A.16) \[
|\Delta_{1n}| \leq ||V_{H_n, H_F}||_\infty ||H_n - H_F||_1,
\]

where \( || \cdot ||_1 \) denotes the \( L_1 \)-norm on \( \mathbb{R} \). It follows from (3.7) and Lemma 8.2.4A in Serfling (1980) that \( ||V_{H_n, H_F}||_1 = o_p(1) \) under the assumptions of Theorem 4.1. Another application of (3.7) proves that

(A.17) \[
|\Delta_{1n}| = o_p(n^{-1/2})
\]
under the assumptions of Theorem 4.1. Similarly, under the assumptions of Theorem 4.2, it is a consequence of (3.7) and Lemma 8.2.4E in Serfling (1980) that \( \|V_{H_n, H_p}\| = o_p(1) \). In conjunction with Lemma A.1, this proves (A.17) under the assumptions of Theorem 4.2. Finally, (A) implies
\[
|\Delta_{2n}| = o_p(n^{-1/2}),
\]
by extending the argument in Ghosh (1971) from i.i.d. sequences to the set \( \{h(X_i); I \in D_n\} \).

**Lemma A.1.** Assume (B). Then \( E(\|H_n - H_p\|_1) = O(n^{-1/2}) \) for all designs in Section 2.

**Proof.** By Hölder’s inequality,
\[
\sqrt{n} E(\|H_n - H_p\|_1) = \int E|W_n(y)| \, dy \leq \left( \int EW_n(y)^2 \right)^{1/2} \, dy,
\]
where \( W_n(\cdot) \) is defined in (3.1). Let \( \hat{W}_n(t) = \sqrt{n} \sum_{i \in S_{n,m}} \eta_i(X_i) / n_{(m)} \) be the (complete) \( U \)-statistic corresponding to \( W_n \). After some calculations one finds, for each of the three designs of Section 2,
\[
E(W_n(y)^2) = \text{Var}(W_n(y))
\]
\[
\begin{align*}
&= \left( 1 - \frac{1}{N} \right) \text{Var}(\hat{W}_n(y)) + \frac{\sigma^2_m(y, y)}{\gamma} \quad \text{design D1} \\
&= \sigma^2(y, y) + \frac{1}{\gamma} \left( \sigma^2_m(y, y) - \sum_{i=1}^m \sigma^2_{1i}(y, y) \right) \quad \text{design D2} \\
&= \sigma^2(y, y) + \frac{1}{\gamma} \left( \sigma^2_2(y, y) - \sum_{i=1}^2 \sigma^2_{1i}(y, y) \right) \quad \text{design D3},
\end{align*}
\]
with \( \sigma^2(\cdot, \cdot), \sigma^2_m(\cdot, \cdot), \) and \( \sigma^2_{1i}(\cdot, \cdot) \) defined in (3.3)–(3.5). Observe that \( \sigma^2_{1i}(y, y) \leq (\sigma^2_m(y, y)) + \sigma^2_{ij}(y, y) / 2 \). In conjunction with Remark 4.4, this implies \( \sigma^2(y, y) = \sum_{i,j} \sigma^2_{ij}(y, y) \leq m \sigma^2_m(y, y) \). Note further that \( \text{Var}(\hat{W}_n(y)) \leq m \sigma^2_m(y, y) \). The lemma now follows for designs D1–D3 because of (B) and the identity \( \sigma^2_m(y, y) = \text{Var}(\hat{W}(y))(1 - \text{Var}(\hat{W}(y))) \). Designs D4 and D5 are handled similarly. □

**Acknowledgments.** I would like to thank two anonymous referees, whose comments on an earlier version of this paper made the presentation much clearer.

**References**


**DEPARTMENT OF MATHEMATICAL STATISTICS**
**LUND UNIVERSITY**
**BOX 118**
**S-221 00 LUND**
**SWEDEN**

E-MAIL: ola@maths.lth.se