The Annals of Statistics
1996, Vol. 24, No. 6, 2631–2654

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 $n_{(m)} = n!/(n-m)!$ 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

$$\sum_{i=1}^{N} c_{Ni} h_{i:N}. \tag{1.1}$$

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.
AMS 1991 *subject classifications*. Primary 62G30; secondary 62G20, 62F35.
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 $h(X_1, \dots, X_m)$, and consider estimating $T(H_F)$ by

$$T(H_n) = \int_0^1 J(t) H_n^{-1}(t)\, dt + \sum_{j=1}^{d} a_j H_n^{-1}(p_j), \tag{1.2}$$

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

$$c_{Ni} = \int_{(i-1)/N}^{i/N} J(t)\, dt + \sum_{j=1}^{d} a_j 1_{\{i = [Np_j] + 1\}}, \tag{1.3}$$

where $[x]$ 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)]; $D_n = S_n(m)$, $J \equiv 1$, $d = 0$, so that
$$T(H_n) = \frac{1}{n_{(m)}} \sum_{I \in S_n(m)} h(X_I).$$
2. Incomplete U-statistics [Blom (1976)]; $J \equiv 1$, $d = 0$, 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 $J \equiv 0$, $d = 1$, $p_1 = p$ 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 $J \equiv 0$, $d = 1$ 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 Szemeredi (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 $T(H_F)$.

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 $\sqrt{n}(H_n - H_F)$ 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 $1 \le i \le n$ occurs in equally many $I \in D_n$, *symmetric* if $I \in D_n \Rightarrow \pi I \in D_n$ for all $m!$ permutations π of the elements in I and *asymmetric* if $I \in D_n \Rightarrow \pi I \notin D_n$ for all π 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 $1/n_{(m)}, \dots, 1/n_{(m)}$.

D2. *Designs based on cyclic permutations*. Given a positive integer γ_n, define vectors $\mathbf{i} = (i, \dots, i)$, $\mathbf{d}_1 = (d_{11}, \dots, d_{1m}), \dots, \mathbf{d}_{\gamma_n} = (d_{\gamma_n 1}, \dots, d_{\gamma_n m})$ of length m, so that all $d_{jk} \ominus d_{jk'}$, $k \ne k'$, are different, $0 \le d_{j1} < \cdots < d_{jm}$ and $d_{1m} < \cdots < d_{\gamma_n m}$. Then put $D_n = \{\mathbf{i} \oplus \mathbf{d}_j;\ 1 \le i \le n,\ 1 \le j \le \gamma_n\}$. Notice that $\gamma_n = N(n)/n$. Examples are:

$m = 2$, $\mathbf{d}_k = (0, k)$, $k = 1, \dots, \gamma_n$;
$m = 3$, $\gamma_n = 1$ and $\mathbf{d}_1 = (0, 1, 3)$;
$m = 4$, $\gamma_n = 1$ and $\mathbf{d}_1 = (0, 1, 4, 6)$;
$m = 3$, $\gamma_n = 2$, $\mathbf{d}_1 = (0, 1, 3)$ and $\mathbf{d}_2 = (0, 4, 9)$.

D3. *Balanced asymmetric designs*, $m = 2$. Let γ_n be a positive integer. Each $1 \le i \le n$ occurs in $2\gamma_n$ distinct $I \in D_n$. Given $(i, j) \in D_n$, we have

$(j,i) \notin D_n$ and for each i there are γ_n distinct elements in D_n of the form (i,k) and γ_n distinct elements of the form (k,i). Notice that $\gamma_n = N(n)/n$. As a special case, we have the design

a. $D_n = \{i, j); 0 < j \ominus i \leq \gamma_n\}$.

D4. *Recursive design based on cyclic permutations*. Here $D_n = \{\mathbf{i} + \mathbf{d}_j;\ 1 \leq j \leq \tilde{\gamma}_n, 1 \leq i \leq n - d_{jm}\}$.

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

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 D1 will in general be unbalanced, whereas both D2 and D3 are balanced and asymmetric (if $n > 2d_{\gamma_n m}$). Actually, design D2 is a special case of design D3 when $m = 2$. The reason for restricting ourselves to asymmetric designs in D2 and D3 is that they have fewer elements than the corresponding symmetric design (obtained by given any $I \in D_n$ adding all other πI). Another important feature of D2 is that each pair (i,j) with $i < j$ is contained in at most one $I \in D_n$.

Designs D4 and D5 have several important properties. They are obtained by removing some multi-indices from D2 and D3a, respectively. This means that $\gamma_n := N(n)/n \neq \tilde{\gamma}_n$, but $\lim \gamma_n = \lim \tilde{\gamma}_n = \gamma$, whenever either limit exists. If $\tilde{\gamma}_n$ is a nondecreasing function of n, both D4 and D5 are *recursive*, which means that $D_n \subset D_{n+1}$. If, in addition, $\tilde{\gamma}_n \equiv \gamma$, designs D4 and D5 are also *on line*, in that $|D_{n+1} \setminus D_n| = O(1)$, so that the number of multi-indices added for each new observation does not increase with n. Actually, $|D_{n+1} \setminus D_n| = \gamma$ 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

$$\sup\left\{|i_j - i_{j'}|;\ j \neq j',\ i_j, i_{j'} \in I \text{ for some } I \in D_n\right\} \ll n.$$

Notice that design D5 is local if $\tilde{\gamma}_n \ll n$. Similarly, design D4 is local as soon as $d_{\tilde{\gamma}_n m} \ll n$. We could have dropped the assumption $i < j$ in D5, 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 D5 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

$$W_n(y) = \sqrt{n}\,(H_n(y) - H_F(y)) \tag{3.1}$$

on $(D[-\infty,\infty], \mathscr{B}_0)$, the space of right-continuous functions on $[-\infty,\infty]$ with left-hand limits, endowed with supremum norm $\|\cdot\|_\infty$ and the σ-algebra $\mathscr{B}_0$ 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

$$N(n)/n = \gamma_n \to \gamma \quad \text{as } n \to \infty, \qquad 0 < \gamma \le \infty. \tag{3.2}$$

If $\gamma < \infty$, γ can be any positive real number for the random design, but it has to be a positive integer for designs D2 and D3. [However, when $h(\cdot)$ is symmetric, it suffices that 2γ is a positive integer for design D3; cf. Section 5.] The case $\gamma = 0$ could also be treated, but is less interesting from the statistical point of view, since it results in estimators with efficiency 0.

Define $\eta_y(\cdot) = 1_{\{h(\cdot) \le y\}} - H_F(y)$ and notice that

$$W_n(y) = \frac{\sqrt{n}}{N} \sum_{I \in D_n} \eta_y(X_I)$$

is an incomplete U-statistic. We will also introduce

$$\sigma_{1ij}^2(x, y) = E\big(\eta_x(X_{I_i})\eta_y(X_{J_j})\big), \qquad 1 \le i, j \le m, \tag{3.3}$$

where $I_i = (i, i+1, \dots, m, 1, \dots, i-1)$ is an $i-1$ times shifted cyclic rearrangement of $(1, \dots, m)$ and J_j a $j-1$ times shifted cyclic rearrangement of $(1, m+1, \dots, 2m-1)$. Put also

$$\sigma_m^2(x, y) = E\big(\eta_x(X_I)\eta_y(X_I)\big) = H_F(x) \wedge H_F(y) - H_F(x)H_F(y) \tag{3.4}$$

and

$$\sigma^2(x, y) = \sum_{i,j=1}^{m} \sigma_{1ij}^2(x, y). \tag{3.5}$$

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*

$$\begin{aligned} C(x, y) &= \sigma^2(x, y) + \frac{\sigma_m^2(x, y)}{\gamma} && \textit{design } \mathrm{D1} \\ &= \sigma^2(x, y) + \frac{1}{\gamma}\left(\sigma_m^2(x, y) - \sum_{i=1}^{m} \sigma_{1ii}^2(x, y)\right) && \textit{design } \mathrm{D2} \\ &= \sigma^2(x, y) + \frac{1}{\gamma}\left(\sigma_2^2(x, y) - \sum_{i=1}^{2} \sigma_{1ii}^2(x, y)\right) && \textit{design } \mathrm{D3} \end{aligned} \tag{3.6}$$

when $\gamma < \infty$. The weak convergence also holds for designs $D4$ and $D5$ with the same covariance function $C(\cdot,\cdot)$ as for $D2$ and $D3$, respectively. In particular,

$$\|H_n - H_F\|_\infty = O_p(n^{-1/2}) \tag{3.7}$$

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 D1.

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

$$\begin{aligned} d_1T(F;G-F) &= \frac{d}{d\varepsilon}T(F+\varepsilon(G-F))|_{\varepsilon=0} \\ &= -\int_{-\infty}^{\infty}(G(y)-F(y))J(F(y))\,dy \\ &\quad + \sum_{j=1}^{d} a_j \frac{p_j - G\big(F^{-1}(p_j)\big)}{f\big(F^{-1}(p_j)\big)}; \end{aligned}$$

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

$$\begin{aligned} T(H_n) &= T(H_F) + d_1T(H_F, H_n - H_F) + \Delta_n \\ &= T(H_F) + \frac{1}{N}\sum_{I\in D_n} A(X_I) + \Delta_n, \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} A(x_1,\dots,x_m) &= d_1T\big(H_F, \delta_{h(x_1,\dots,x_m)} - H_F\big) \\ &= -\int_{-\infty}^{\infty}\big(1_{\{h(x_1,\dots,x_m)\le y\}} - H_F(y)\big)J\big(H_F(y)\big)\,dy \\ &\quad + \sum_{j=1}^{d} a_j \frac{p_j - 1_{\{h(x_1,\dots,x_m)\le H_F^{-1}(p_j)\}}}{h_F\big(H_F^{-1}(p_j)\big)}, \end{aligned}$$

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

$$I(x,T,F) = EA(x, X_2,\dots,X_m) + \cdots + EA(X_1,\dots,X_{m-1},x)$$

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),

$$\sigma_{1ij}^2 = E\left(A(X_{I_i})A(X_{J_j})\right),$$

where I_i and J_j are defined as in Section 3, and

$$\sigma_m^2 = EA(X_1, \dots, X_m)^2.$$

Note that $\sigma^2 = \sum_{i,j=1}^m \sigma_{1ij}^2$. We will assume that

$$\sigma^2 > 0, \tag{4.2}$$

$$\sigma_m^2 < \infty. \tag{4.3}$$

For H_F and the score function J, introduce the following conditions:

(A) H_F has positive derivative at $H_F^{-1}(p_j)$, $j = 1, \dots, d$.

(B) $\int (H_F(y)(1 - H_F(y)))^{1/2} \, dy < \infty$.

(C) $J \equiv 0$ outside $[\alpha, \beta]$, where $0 < \alpha < \beta < 1$, and J is bounded and continuous a.e. Lebesgue and H_F^{-1}.

(D) J 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*

$$\sqrt{n}\,(T(H_n) - T(H_F)) \to_{\mathscr{L}} N(0, \sigma^2(\gamma)),$$

with $\sigma^2(\infty) = \sigma^2$ *for all three designs* $D1$–$D3$ *defined in Section* 2, *and*

$$\begin{aligned} \sigma^2(\gamma) &= \sigma^2 + \frac{\sigma_m^2}{\gamma} && \textit{design } \mathrm{D1} \\ &= \sigma^2 + \frac{\sigma_m^2 - \sum_{i=1}^m \sigma_{1ii}^2}{\gamma} && \textit{design } \mathrm{D2} \\ &= \sigma^2 + \frac{\sigma_2^2 - \sum_{i=1}^2 \sigma_{1ii}^2}{\gamma} && \textit{design } \mathrm{D3} \end{aligned}$$

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 γ, the random design has higher asymptotic variance than designs D2 and D3.

REMARK 4.4. Let $I = (1, \dots, m)$ and put $A_i(x) = E(A(X_I)|X_i = x)$, $i = 1, \dots, m$, and $\tilde{A}(x_I) = A(x_I) - \sum_{i=1}^n A_i(x_i)$. Then $EA_i(X_i)^2 = \sigma_{1ii}^2$ and

$$\sigma_m^2 = \sum_{i=1}^m \sigma_{1ii}^2 + E\tilde{A}(X_I)^2 + 2\sum_{i=1}^m E\big(A_i(X_i)\tilde{A}(X_I)\big) \geq \sum_{i=1}^m \sigma_{1ii}^2,$$

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 $\tilde{A}(x_I) = 0$ a.e. $F \times \cdots \times F$.

REMARK 4.5. Note that σ_m^2 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,

$$\sigma_m^2 = \int_{-\infty}^{\infty} \bar{A}(y)^2 \, dH_F(y),$$

with

$$\bar{A}(y) = -\int_{-\infty}^{\infty} \big(1_{\{y \leq z\}} - H_F(z)\big) J\big(H_F(z)\big) dz + \sum_{j=1}^d a_j \frac{p_j - 1_{\{y \leq H_F^{-1}(p_j)\}}}{h_F\big(H_F^{-1}(p_j)\big)}$$

the "influence function" for the L-statistic. For U-quantiles,

$$\sigma_m^2 = p(1-p)\big(h_F\big(H_F^{-1}(p)\big)\big)^{-2}.$$

The other quantities σ^2 and $\{\sigma_{1ij}^2\}$ 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

$$I(x, T, F) = mEA(x, X_2, \dots, X_m)$$

and $\sigma_{1ij}^2 \equiv \sigma_1^2 = \sigma^2/m^2$. For symmetric kernels, it is more convenient to consider the reduced class of multi-indices

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

with

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

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

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

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. $\tilde{D}_n = \{(I);\, I = \mathbf{i} \oplus \mathbf{d}_j,\, 1 \le i \le n,\, 1 \le j \le \gamma_n\}$.
3. Each index i occurs in $2\gamma_n$ different $I \in \tilde{D}_n$; that is, $\tilde{D}_n$ is balanced. It suffices that $2\gamma_n$ is a positive integer in this case, and hence that 2γ is a positive integer if $\gamma < \infty$.

The asymptotic variances for designs D1–D3 become

$$\begin{aligned} \sigma^2(\gamma) &= \sigma^2 + \frac{\sigma_m^2}{\gamma} && \text{design D1} \\ &= \sigma^2 + \frac{\sigma_m^2 - \sigma^2/m}{\gamma} && \text{design D2} \\ &= \sigma^2 + \frac{\sigma_2^2 - \sigma^2/2}{\gamma} && \text{design D3.} \end{aligned} \tag{5.1}$$

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

$$\begin{aligned} C(x, y) &= \sigma^2(x, y) + \frac{\sigma_m^2(x, y)}{\gamma} && \text{design D1} \\ &= \sigma^2(x, y) + \frac{1}{\gamma}\left(\sigma_m^2(x, y) - \frac{\sigma^2(x, y)}{m}\right) && \text{design D2} \\ &= \sigma^2(x, y) + \frac{1}{\gamma}\left(\sigma_2^2(x, y) - \frac{\sigma^2(x, y)}{2}\right) && \text{design D3,} \end{aligned}$$

with $\sigma_m^2(x,y) = H_F(x) \wedge H_F(y) - H_F(x)H_F(y)$ (as before) and $\sigma^2(x,y) = \mathrm{Cov}(\eta_x(X_I), \eta_y(X_J))$, with $|I \cap J| = 1$.

There is a nice interpretation of W^* for the random design D1. 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 $\sigma_m^2(\cdot,\cdot)$ [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,

$$E(W_1^*(y) - W_1^*(x))^2 \le C(y-x)^2,$$
$$E(W_2^*(y) - W_2^*(x))^2 \le C|y-x|$$

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 μ and τ are unknown location and scale parameters and F_0 is a known distribution. Here $q = 1$, and $h(x_1,\dots,x_m) = \sum_{i=1}^m \lambda_i x_i$ can be used as a location kernel for any nonnegative λ_i with $\sum_i \lambda_i = 1$. The choice $\lambda_i = 1/m$ seems most natural and results in the sample mean if $J \equiv 1$, $d = 0$, and the generalized Hodges–Lehmann estimator [Serfling (1984) and Choudhury and Serfling (1988)] when $J \equiv 0$, $d = 1$ 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 ($q = 1$)

$$h(x_1,\dots,x_m) = c\left|\sum_{i=1}^m \lambda_i x_i\right|^r, \tag{6.1}$$

with $\sum_i \lambda_i = 0$, $r \ge 1$ [cf. also Example (ii) in Choudhury and Serfling (1988)]. Here c is a multiplicative constant chosen so that $T(H_F) = \tau^r$. (For instance, if F_0 is the standard normal distribution, τ 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 $Y_i = \alpha + \beta U_i + e_i$, $i = 1, \dots, n$, with α and β 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 $(h_1(x_1, \dots, x_m), h_2(x_1, \dots, x_m))$ be the LS-estimate of (α, β) computed from $x_i = (u_i, y_i)$, $i = 1, \dots, m$. 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 $U_1 \le U_2 \le U_3$ 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

$$\begin{aligned} \varepsilon(x_1, x_2, x_3) &= \frac{u_3 - u_2}{u_3 - u_1} y_1 + \frac{u_2 - u_1}{u_3 - u_1} y_3 - y_2 \\ &= \frac{u_3 - u_2}{u_3 - u_1} e_1 + \frac{u_2 - u_1}{u_3 - u_1} e_3 - e_2, \end{aligned} \tag{6.2}$$

and it is independent of α and β. [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 $U_1 \le \cdots \le U_n$. Rousseeuw and Hubert (1993) consider an IUQ based on the kernel $h(x_i, x_j, x_k) = c|\varepsilon(x_i, x_j, x_k)|$ for defining regression-free estimators of τ. The constant c is chosen so that $T(H_F) = \tau$. Note that $h(X_1, X_2, X_3) = \tilde{h}(e_1, e_2, e_3)$, where $\tilde{h}(\cdot, \cdot, \cdot)$ has the form (6.1), with $r = 1$, $m = 3$ and constants λ_i depending on (U_1, U_2, U_3). Several designs are considered by Rousseeuw and Hubert (1993), among others $D_n = \tilde{S}_n(m)$ and

$$D_n = \{(1, 2, 3), (2, 3, 4), \dots, (n-2, n-1, n)\}, \tag{6.3}$$

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 $U_1 \le \cdots \le U_n$, 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

$$D_n = \{(1,2,4),(2,3,5),\dots,(n-3,n-2,n)\} \tag{6.4}$$

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

$$\sqrt{n}\,(T(H_n) - \tau) \to_{\mathscr{L}} N\left(0, \sigma^2 + \sigma_3^2 - \sum_{i=1}^{3} \sigma_{1ii}^2\right).$$

EXAMPLE 5 (Scale estimation in nonparametric regression). Suppose that $q = 1$ and $X_i = m(U_i) + e_i$, $i = 1,\dots,n$, where $\{e_i\}$ are i.i.d. with common distribution $F(\cdot) = F_0(\cdot/\tau)$, $0 \le U_1 \le \cdots \le U_n \le 1$, are design points (nonrandom or random) and m: $[0,1] \to \mathbb{R}$ is an unknown regression function. Rice (1984) considers estimating τ through

$$\tilde{\tau}_n = \sqrt{\frac{1}{(n-1)} \sum_{i=2}^{n} \frac{(X_i - X_{i-1})^2}{2}}\,,$$

which is actually an incomplete version of the sample standard deviation using design D5, with $\tilde{\gamma}_n = \gamma = 1$ [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 γ, 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 τ 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 $T(H_F) = \tau$. 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 $\tilde{H}_n$ 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 $T(\tilde{H}_n)$. Under certain assumptions, $h(X_I)$ is close to $h(e_I)$. Put

$$M_n = \max_{2 \le i \le n} (U_i - U_{i-1})$$

for the maximal spacing of the design points. Suppose also that $\|m\|_\varepsilon < \infty$ for some $0.5 < \varepsilon \le 1$, where

$$\|m\|_\varepsilon = \sup_{0 \le s < t \le 1} \frac{|m(t) - m(s)|}{|t - s|^\varepsilon}.$$

Then $|h(X_i, X_j) - h(e_i, e_j)| \le c|m(U_j) - m(U_i)| \le c\|m\|_\varepsilon(\tilde{\gamma}_n M_n)^\varepsilon$ for any $(i, j) \in D_n$. Suppose also that the (nonnegative) weights c_{Ni} sum to 1 in (1.3) [i.e. $\int J(t)\,dt + \sum_1^d a_j = 1$]. This implies

$$|\hat{\tau}_n - T(\tilde{H}_n)| \le c\|m\|_\varepsilon(\tilde{\gamma}_n M_n)^\varepsilon = o(n^{-1/2}) \quad \text{a.s.} \tag{6.5}$$

as long as

$$\tilde{\gamma}_n M_n = o(n^{-1/(2\varepsilon)}) \quad \text{a.s.} \tag{6.6}$$

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

$$\sqrt{n}\,(\hat{\tau}_n - \tau) \to_{\mathscr{L}} N\left(0, \sigma^2 + \frac{\sigma_2^2 - \sigma^2/2}{\gamma}\right)$$

[cf. (5.1)], with $\gamma = \lim \gamma_n = \lim \tilde{\gamma}_n$, $0 < \gamma \le \infty$. When $\gamma = \infty$, the asymptotic variance σ^2 is the same as for the corresponding GL-estimator of τ in the location–scale model. For instance suppose F_0 is the standard normal distribution and τ 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 $\tilde{\gamma}_n = o(n^{1-1/(2\varepsilon)})$. 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 $\inf_t w(t) = \underline{w} > 0$, then $P(M_n > (1 + \delta)(\log n)/(\underline{w}n) \text{ i.o.}) = 0$ for any $\delta > 0$, under some regularity conditions on w [cf. Deheuvels (1984)].

An estimator of τ 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 $h(x_1, x_2, x_3) = c|2x_1/3 + x_3/3 - x_2|$. Let H_n and $\tilde{H}_n$ be defined as above, and put

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

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

$$\begin{aligned}|h(X_i, X_{i+1}, X_{i+3}) - h(e_i, e_{i+1}, e_{i+3})| &\le c \max_i |2m_i/3 + m_{i+3}/3 - m_{i+1}| \\ &= O(M_n^\varepsilon) = O(n^{-\varepsilon}) = o(n^{-1/2}).\end{aligned}$$

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

$$\sqrt{n}\,(\check{\tau}_n - \tau) \to_{\mathscr{L}} N\left(0, \sigma^2 + \sigma_3^2 - \sum_{i=1}^{3} \sigma_{1ii}^2\right).$$

Notice that the bias is much smaller if m is twice differentiable with $\|m''\|_{L_\infty[0,1]} < \infty$. Then $|\check{\tau}_n - T(\tilde{H}_n)| \le c \max_i |2m_i/3 + m_{i+3}/3 - m_{i+1}| = O(n^{-2})$. Further, $\check{\tau}_n$ 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 $\tilde{\tau}_n$ and $\hat{\tau}_n(p)$ with design D5 and $\tilde{\gamma}_n \equiv \gamma$. Two regression functions were used: $m_1(\cdot) \equiv 0$ (which corresponds to Example 2) and

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

having a jump of height 4 at $u = 1/2$. The error distribution F_0 was the standard normal distribution Φ (so that $\tau = 1$). The kernel of $\hat{\tau}_n(p)$ 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 $\tilde{\tau}_n$, $\hat{\tau}_n(0.5)$ and $\hat{\tau}_n(0.9)$ for sample sizes $n = 30, 100$ and $\gamma = 3, 10, \infty$. Since the Cramér–Rao lower bound for the asymptotic efficiency is $1/2$,

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

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

As expected, $\tilde{\tau}_n$ [and also $\hat{\tau}_n(0.9)$] 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 $\hat{\tau}_n(0.5)$ 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 $\tilde{\tau}_n$ to break down, whereas the more robust IUQ $\hat{\tau}_n(p)$ 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* $\tilde{\gamma}_n = n - 1$ *for regression function* $m_1[D_n = S_n(2)]$ *and* $\tilde{\gamma}_n = [5n^{1/3}]$ *for* m_2 *(local design)*

		$\tilde{\tau}_n$			$\hat{\tau}_n(0.5)$			$\hat{\tau}_n(0.9)$		
m	**n**	$\gamma = 3$	$\gamma = 10$	$\gamma = \infty$	$\gamma = 3$	$\gamma = 10$	$\gamma = \infty$	$\gamma = 3$	$\gamma = 10$	$\gamma = \infty$
m_1	30	0.835	0.914	0.974	0.524	0.679	0.774	0.727	0.850	0.940
	100	0.849	0.935	0.991	0.533	0.712	0.836	0.732	0.887	0.967
	∞	0.857	0.952	1.000	0.534	0.729	0.864	0.735	0.898	0.992
m_2	30	0.190	0.028	0.012	0.368	0.087	0.024	0.234	0.018	0.010
	100	0.414	0.095	0.021	0.479	0.357	0.102	0.522	0.145	0.016
	∞	0.857	0.952	1.000	0.534	0.729	0.864	0.735	0.898	0.992

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 $(D[-\infty, \infty], \mathscr{B}_0)$ 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

$$\begin{aligned}\sqrt{n}\,(T(H_n) - T(H_F)) &= \sqrt{n}\, d_1 T(H_F, H_n - H_F) + o_p(n^{-1/2}) \\ &= \frac{\sqrt{n}}{N} \sum_{I \in D_n} d_1 T\big(H_F, \delta_{h(X_I)} - H_F\big) + o_p(n^{-1/2}).\end{aligned}$$

Here $\delta_{h(X_I)}$ is the one-point distribution at $h(X_I)$. 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

$$\sqrt{n}\, d_1 T(H_F, H_n - H_F) \to_{\mathscr{L}} d_1 T(H_F, W^*),$$

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 $X_1, \dots, X_n \in \mathbb{R}^1$, let $h\colon \mathbb{R}^{mq} \to \mathbb{R}^q$ be defined by $h(X_I) = \sum_{i \in I} X_i / m$, and let $|\cdot|$ denote the Euclidean norm in $\mathbb{R}^q$. Then the statistic

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

is an estimator of $\boldsymbol{\mu} = \arg\min_{\boldsymbol{\mu}' \in \mathbb{R}^q} E|h(X_I) - \boldsymbol{\mu}'|$. It reduces to (the m-argument extension of) the Hodges–Lehmann estimator for $q = 1$, and it has been shown that $\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}$ is asymptotically equivalent to a U-statistic having a certain multivariate kernel A [Theorem 3.2 in Chaudhuri (1992a)]. Replacing $\tilde{S}_n(m)$ by some design $\tilde{D}_n$ 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 $\tilde{D}_n$. 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 $\mathbf{U}_i$ of dimension p, response variables Y_i and an unknown vector of regression parameters $\boldsymbol{\beta}$. 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

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

w.r.t. $\boldsymbol{\beta}'$. Here $r_i(\boldsymbol{\beta}) = Y_i - \boldsymbol{\beta}\mathbf{U}_i^T$ is the ith residual using $\boldsymbol{\beta}$ as the regression parameter and $1 \le \lambda_n \le \binom{n}{2}$. ($T$ 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

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

w.r.t. $\boldsymbol{\beta}'$.

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 ε^* (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 $\tilde{\gamma} \equiv \gamma = 1$, it may be shown that $\varepsilon^* = 0.25$. (However, $\varepsilon^* \to 1 - 1/\sqrt{2}$ as $\gamma \to \infty$.) Random design D1 gives $\varepsilon^* = 0$.

The LTD-estimator has $\varepsilon^* = \min(\sqrt{\lambda}, 1 - \sqrt{\lambda})$ if $\lambda_n/\binom{n}{2} \to \lambda$ as $n \to \infty$, 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 $\tilde{\gamma}_n \equiv \gamma = 1$ and $\lambda_n/N(n) \to \lambda$, 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 α, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ and τ. A regression-free IGL-estimator of τ may be constructed, using $X_i = (Y_{i-p}, \dots, Y_i)$, $m = p + 2$ 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 $\{X_i\}$, 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 $\tilde{\gamma}_n \equiv \gamma$ may be used to construct on-line estimators of $\theta = T(H_F)$. Notice that $|D_n \setminus D_{n-1}| = \gamma$ for n large enough. Hence, we propose the recursive scheme

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

for estimating θ, with b_n a recursive estimator of $\int \psi'(y - \theta) dH_F(y)$. 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 $\psi(x) = p - 1_{\{x \le 0\}}$].

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 symmetic kernels only, the generalization to nonsymmetric kernels is straightforward. [Notice that $\sum_1^l \lambda_i W_n(y_i)$ 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 contunuity:

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

We will prove that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P\big(\omega_{W_n}(\delta) \ge 3\varepsilon\big) = 0. \tag{A.1}$$

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

$$\begin{aligned} H_F(\xi_i -) - H_F(\xi_{i-1}) &\le \frac{1}{n}, \\ H_F(\xi_i) - H_F(\xi_{i-1}) &\ge \frac{1}{n} \end{aligned} \tag{A.2}$$

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

$$H_{in} = H_n(\xi_i -) - H_n(\xi_{i-1}), \qquad i = 1, \dots, L_n.$$

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

$$|W_n(y) - W_n(x)| \le |W_n(\xi_{j-1}) - W_n(\xi_{i-1})| + \sqrt{n}\,(H_{jn} + H_{in}) + \frac{2}{\sqrt{n}}.$$

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

$$\omega_{W_n}(\delta) \le \tilde{\omega}_{W_n}(2\delta) + 2\sqrt{n} \max_{1 \le i \le L_n} H_{in} + \frac{2}{\sqrt{n}}, \tag{A.3}$$

where

$$\tilde{\omega}_Z(\delta) = \sup_{\substack{x, y;\, |H_F(y) - H_F(x)| \le \delta \\ x, y \in \Xi_n}} |Z(y) - Z(x)|.$$

We will prove below that

$$\begin{aligned} E(W_n(y) - W_n(x))^4 \le{}& \frac{C_1}{n} |H_F(y) - H_F(x)| \\ &+ C_2 (H_F(y) - H_F(x))^2 \end{aligned} \tag{A.4}$$

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

$$E(W_n(y) - W_n(x))^4 \le C(H_F(y) - H_F(x))^2, |H_F(y) - H_F(x)| \ge \frac{1}{n}, \tag{A.5}$$

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

$$
\begin{aligned}
P\Big(2\sqrt{n} \max_{1\le i\le L_n} H_{ni} \ge \varepsilon\Big) &\le \sum_{i=1}^{L_n} P\big(2\sqrt{n}\, H_{ni} \ge \varepsilon\big) \\
&= \sum_{i=1}^{L_n} P\big(2W_{ni} \ge \varepsilon - 2\sqrt{n}\,(H_F(\xi_i -) - H_F(\xi_{i-1}))\big) \\
&\le \sum_{i=1}^{L_n} P\bigg(2W_{ni} \ge \varepsilon - \frac{2}{\sqrt{n}}\bigg) \\
&\le \sum_{i=1}^{L_n} P\bigg(W_{ni} \ge \frac{\varepsilon}{4}\bigg) \le n\frac{C(1/n)^2}{(\varepsilon/4)^4} \to 0
\end{aligned} \tag{A.6}
$$

as $n \to \infty$, where C is the same constant as in (A.5). Next, we will prove that

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\big(\tilde{\omega}_{W_n}(2\delta) \ge \varepsilon\big) = 0. \tag{A.7}$$

To this end, let $\Theta_{n\delta} = \{\theta_0, \theta_1, \dots, \theta_{L_{n\delta}}\}$ be a subset of Ξ_n with $\theta_0 = -\infty$, $\theta_{L_{n\delta}} = \infty$ and

$$
\begin{aligned}
H_F(\theta_i -) - H_F(\theta_{i-1}) &\le 3\delta, \\
H_F(\theta_i) - H_F(\theta_{i-1}) &\ge 3\delta
\end{aligned}
$$

for $i = 1, \dots, L_{n\delta}$. Observe that $L_{n\delta} \le 1/(3\delta)$. Then define sets $K_1, \dots, K_{L_{n\delta}}$ by $K_i = [\theta_{i-1}, \theta_i] \cap \Xi_n$ if $H_F(\theta_i) - H_F(\theta_i -) \le 2\delta$ and $K_i = [\theta_{i-1}, \theta_i) \cap \Xi_n$ if $H_F(\theta_i) - H_F(\theta_i -) > 2\delta$. Suppose $x \ne y \in \Xi_n$ and $|H_F(y) - H_F(x)| \le 2\delta$. Then $x \in K_i$ and $y \in K_j$ for some i, j with $|i-j| \le 1$. Moreover, if $|i-j| = 1$, then $K_i \cap K_j \ne \emptyset$. It follows that

$$P\big(\tilde{\omega}_{W_n}(2\delta) \ge \varepsilon\big) \le \sum_{i=1}^{L_{n\delta}} P\bigg(\sup_{x,y\in K_i} |W_n(y) - W_n(x)| \ge \frac{\varepsilon}{3}\bigg).$$

If $x, y \in K_i$, then $|H_F(y) - H_F(x)| \le 5\delta$. 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 $\overline{C}$,

$$P\bigg(\sup_{x,y\in K_i} |W_n(y) - W_n(x)| \ge \frac{\varepsilon}{3}\bigg) \le \frac{\overline{C}(5\delta)^2}{(\varepsilon/3)^4},$$

which implies (A.7), since

$$P\big(\tilde{\omega}_{W_n}(2\delta) \ge \varepsilon\big) \le L_{n\delta}\frac{\overline{C}(5\delta)^2}{(\varepsilon/3)^4} \le \frac{\overline{C}25\delta}{3(\varepsilon/3)^4}.$$

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, $\eta(x_I) = \eta_y(x_I) - \eta_x(x_I)$. Assume that $I_1, \dots, I_N$ 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

$$\begin{aligned} E(W_n(y) - W_n(x))^4 &= \frac{n^2}{N^4} \sum_{j_1,\dots,j_4=1}^{N} E\left(\prod_{k=1}^{4} \eta(X_{I_{j_k}}) \right) \\ &= \frac{n^2}{N^4} \big(N\alpha_1 + 3N(N-1)\alpha_2 \\ &\qquad + 4N(N-1)\alpha_3 + 6N(N-1)(N-2)\alpha_4 \\ &\qquad\qquad + N(N-1)(N-2)(N-3)\alpha_5 \big), \end{aligned} \tag{A.8}$$

where

$$\alpha_1 = E\big(\eta(X_{I_1})\big)^4 \le H_F(y) - H_F(x), \tag{A.9}$$

$$\begin{aligned} \alpha_2 &= E\big(\eta(X_{I_1})^2 \eta(X_{I_2})^2\big) \\ &\le \frac{C_{21}}{n}(H_F(y) - H_F(x)) + (H_F(y) - H_F(x))^2, \end{aligned} \tag{A.10}$$

$$\begin{aligned} \alpha_3 &= \left| E\big(\eta(X_{I_1})^3 \eta(X_{I_2})\big) \right| \\ &\le \frac{C_{31}}{n}(H_F(y) - H_F(x)), \end{aligned} \tag{A.11}$$

$$\begin{aligned} \alpha_4 &= \left| E\big(\eta(X_{I_1})^2 \eta(X_{I_2}) \eta(X_{I_3})\big) \right| \\ &\le \frac{C_{41}}{n^2}(H_F(y) - H_F(x)) + \frac{C_{42}}{n}(H_F(y) - H_F(x))^2 \end{aligned} \tag{A.12}$$

and

$$\begin{aligned} \alpha_5 &= \left| E\big(\eta(X_{I_1}) \eta(X_{I_2}) \eta(X_{I_3}) \eta(X_{I_3})\big) \right| \\ &\le \frac{C_{51}}{n^3}(H_F(y) - H_F(x)) + \frac{C_{52}}{n^2}(H_F(y) - H_F(x))^2. \end{aligned} \tag{A.13}$$

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,

$$\begin{aligned} \alpha_2 &= \frac{1}{n_{(m)}^2} \sum_{I, J \in S_n(m)} E\big(\eta(X_I)^2 \eta(X_J)^2\big) \\ &\le \frac{n_{(m)}^2 - n_{(m)}(n-m)_{(m)}}{n_{(m)}^2} E\big(\eta(X_I)^4\big) + \frac{n_{(m)}(n-m)_{(m)}}{n_{(m)}^2} \big(E\eta(X_I)^2\big)^2. \end{aligned}$$

Here $n_{(m)}(n-m)_{(m)}$ is the number of pairs (I, J) with $I \cap J = \varnothing$ and $n_{(m)} - n_{(m)}(n-m)_{(m)} = O(n^{2m-1})$ the number of pairs with $I \cap J \neq \varnothing$. Formula (A.10) follows since $E(\eta(X_I))^2 \le H_F(y) - H_F(x)$ and $E(\eta(X_I))^4 \le H_F(y) - H_F(x)$.

Consider now designs D4 and D5. Observe first that $\|W_n^{(\mathrm{D2})} - W_n^{(\mathrm{D4})}\|_\infty$ and $\|W_n^{(\mathrm{D3a})} - W_n^{(\mathrm{D5})}\|_\infty$ 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 $y \in \mathbb{R}$, $E(W_n^{(\mathrm{D4})}(y) - W_n^{S_n(m)}(y))^2 \to 0$ and $E(W_n^{(\mathrm{D5})}(y) - W_n^{S_n(m)}(y))^2 \to 0$, where $W_n^{S_n(m)}(\cdot)$ 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 $W_n^{S_n(m)}(\cdot)$. The inequality (A.4) is established analogously for $W_n^{(\mathrm{D4})}(\cdot)$ and $W_n^{(\mathrm{D5})}(\cdot)$, and this proves tightness.

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

$$\sqrt{n}\,\|H_n - H_F\|_\infty = \|W_n\|_\infty \to_{\mathscr{L}} \|W^*\|_\infty = O_p(1). \qquad \square$$

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 $T_1(H_F) = \int_0^1 J(t) H_F^{-1}(t)\,dt$, $T_2(H_F) = \sum_{j=1}^d a_j H_F^{-1}(p_j)$. Decompose the remainder into two terms, $\Delta_n = \Delta_{1n} + \Delta_{2n}$, where $\Delta_{in} = T_i(H_n) - T_i(H_F) - d_1 T_i(H_F; H_n - H_F)$. Then

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

where

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

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

$$|\Delta_{1n}| \le \|V_{H_n, H_F}\|_1 \|H_n - H_F\|_\infty \tag{A.15}$$

and

$$|\Delta_{1n}| \le \|V_{H_n, H_F}\|_\infty \|H_n - H_F\|_1, \tag{A.16}$$

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

$$|\Delta_{1n}| = o_p(n^{-1/2}) \tag{A.17}$$

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 $\|V_{H_n, H_F}\|_\infty = o_p(1)$. 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* $E(\|H_n - H_F\|_1) = O(n^{-1/2})$ *for all designs in Section* 2.

PROOF. By Hölder's inequality,

$$\sqrt{n}\, E(\|H_n - H_F\|_1) = \int E|W_n(y)|\, dy \le \int \left(EW_n(y)^2\right)^{1/2} dy,$$

where $W_n(\cdot)$ is defined in (3.1). Let $\tilde{W}_n(t) = \sqrt{n}\sum_{I \in S_n(m)} \eta_y(X_I)/n_{(m)}$ be the (complete) U-statistic corresponding to W_n. After some calculations one finds, for each of the three designs of Section 2,

$$\begin{aligned}
E\left(W_n(y)^2\right) &= \operatorname{Var}(W_n(y)) \\
&= \left(1 - \frac{1}{N}\right)\operatorname{Var}\left(\tilde{W}_n(y)\right) + \frac{\sigma_m^2(y, y)}{\gamma} && \text{design D1} \\
&= \sigma^2(y, y) + \frac{1}{\gamma}\left(\sigma_m^2(y, y) - \sum_{i=1}^{m} \sigma_{1ii}^2(y, y)\right) && \text{design D2} \\
&= \sigma^2(y, y) + \frac{1}{\gamma}\left(\sigma_2^2(y, y) - \sum_{i=1}^{2} \sigma_{1ii}^2(y, y)\right) && \text{design D3,}
\end{aligned}$$

with $\sigma^2(\cdot,\cdot)$, $\sigma_{1ij}^2(\cdot,\cdot)$ and $\sigma_m^2(\cdot,\cdot)$ defined in (3.3)–(3.5). Observe that $\sigma_{1ij}^2(y, y) \le (\sigma_{1ii}^2(y, y)) + \sigma_{1jj}^2(y, y))/2$. In conjunction with Remark 4.4, this implies $\sigma^2(y, y) = \sum_{i,j} \sigma_{1ij}^2(y, y) \le m\sigma_m^2(y, y)$. Note further that $\operatorname{Var}(\tilde{W}_n(y)) \le m\sigma_m^2(y, y)$. The lemma now follows for designs D1–D3 because of (B) and the identity $\sigma_m^2(y, y) = H_F(y)(1 - H_F(y))$. 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

ARCONES, M. A., CHEN, Z. and GINÉ, E. (1993). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. Unpublished manuscript.

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542.

BICKEL, P. J. and LEHMANN, E. L. (1979). Descriptive statistics for nonparametric models. IV. Spread. In *Contributions to Statistics. Hajek Memorial Volume* (J. Jurečková, ed.) 33–40. Academia, Prague.

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

BLOM, G. (1976). Some properties of incomplete U-statistics. *Biometrika* **63** 573–580.

BLUM, M., FLOYD, R. W., PRATT, V., RIVEST, R. and TARJAN, R. E. (1973). Time bounds for selection. *J. Comput. System Sci.* **7** 448–461.

BROWN, B. M. and KILDEA, D. G. (1978). Reduced U-statistics and the Hodges–Lehmann estimator. *Ann. Statist.* **6** 828–835.

CHAUDHURI, P. (1992a). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Ann. Statist.* **20** 897–916.

CHAUDHURI, P. (1992b). Generalized regression quantiles: forming a useful toolkit for robust linear regression. In *L_1-Statistical Analysis and Related Methods* (Y. Dodge, ed.) 169–186. North-Holland, Amsterdam.

CHOUDHURY, J. and SERFLING, R. J. (1988). Generalized order statistics, Bahadur representations, and sequential nonparametric fixed-width confidence intervals. *J. Statist. Plann. Inference* **19** 269–282.

COLE, R., SALOWE, J. S., STEIGER, W. L. and SZEMEREDI, E. (1989). An optimal-time algorithm for slope selection. *SIAM J. Comput.* **18** 792–810.

CROUX, C. and ROUSSEEUW, P. J. (1992). Time-efficient algorithms for two highly robust estimators of scale. In *Computational Statistics* (Y. Dodge and J. Whittaker, eds.) **1** 411–428. Physika, Heidelberg.

CROUX, C., ROUSSEEUW, P. J. and HÖSSJER, O. (1994). Generalized S-estimators. *J. Amer. Statist. Assoc.* **89** 1271–1281.

DEHEUVELS, P. (1984). Strong limit theorems for maximal spacings from a general univariate distribution. *Ann. Probab.* **12** 1181–1193.

DILLENCOURT, M. B., MOUNT, D. M. and NETANYAHU, N. S. (1992). A randomized algorithm for slope selection. *Internat. J. Comput. Geom. Appl.* **2** 1–27.

FREES, E. W. (1991). Trimmed slope estimates for simple linear regression. *J. Statist. Plann. Inference* **27** 203–221.

GASSER, T., STROKA, L. and JENNER, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

HODGES, J. L., JR. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 163–186. Univ. California Press, Berkeley.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HÖSSJER, O. (1997). Recursive U-quantiles. *Sequential Analysis*. To appear.

HÖSSJER, O., CROUX, C. and ROUSSEEUW, P. J. (1994). Asymptotics of generalized S-estimators. *J. Multivariate Anal.* **51** 148–177.

JANSON, S. (1984). The asymptotic distribution of incomplete U-statistics. *Z. Wahrsch. Verw. Gebiete* **66** 495–505.

JANSSEN, P., SERFLING, R. J. and VERAVERBEKE, N. (1984). Asymptotic normality for a general class of statistical functions and applications to measures of spread. *Ann. Statist.* **12** 1369–1379.

JOHNSON, D. B. and MIZOGUCHI, T. (1978). Selecting the Kth element in $X + Y$ and $X_1 + X_2 + \cdots + X_m$. *Siam J. Comput.* **7** 147–153.

LEE, A. J. (1990). *U-Statistics, Theory and Practice. Statistics, Textbooks and Monographs* **110**. Marcel Dekker, New York.

LIU, R. V. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414.

MATOUSEK, J. (1991). Randomized optimal algorithm for slope selection. *Inform. Process. Lett.* **39** 183–187.

NOLAN, D. and POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** 1291–1298.

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

RIEDER, H. (1991). *Robust Statistics I. Asymptotic Statistics*. Univ. Bayreuth.

ROUSSEEUW, P. J. and CROUX, C. (1992). Explicit scale estimators with high breakdown point. In L_1-*Statistical Analysis and Related Methods* (Y. Dodge, ed.) 77–92. North-Holland, Amsterdam.

ROUSSEEUW, P. J. and CROUX, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88** 1273–1283.

ROUSSEEUW, P. J. and HUBERT, M. (1993). Regression-free and robust estimators of scale. Technical Report 93-23, Dept. Mathematics and Computer Science, Univ. Instelling Antwerpen.

RUYMGAART, F. H. and VAN ZUIJLEN, M. C. A. (1992). Empirical U-statistics processes. *J. Statist. Plann. Inference* **32** 259–269.

SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63** 1379–1389.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SERFLING, R. J. (1984). Generalized L-, M- and R-statistics. *Ann. Statist.* **12** 76–86.

SHAMOS, M. I. (1976). Geometry and statistics: problems at the interface. In *New Directions and Recent Results in Algorithms and Complexity* (J. F. Traub, ed.) 251–280. Academic Press, New York.

SILVERMAN, B. W. (1976). Limit theorems for dissociated random variables. *Adv. in Appl. Probab.* **8** 806–819.

SILVERMAN, B. W. (1983). Convergence of a class of empirical distribution functions of dependent random variables. *Ann. Probab.* **11** 745–751.

STROMBERG, A. J., HAWKINS, D. M. and HÖSSJER, O. (1995). The least trimmed differences regression estimator and alternatives. Technical Report 1995:26, Dept. Mathematical Statistics, Lund Univ. and Lund Institute of Technology.

THEIL, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I, II and III. *Koninklijke Nederlandse Akademie van Wetenschappen*, *Proceedings* **53** 386–392, 521–525, 1397–1412.

DEPARTMENT OF MATHEMATICAL STATISTICS
LUND UNIVERSITY
BOX 118
S-221 00 LUND
SWEDEN
E-MAIL: ola@maths.lth.se