The Annals of Statistics
1984, Vol. 12, No. 1, 76–86

GENERALIZED *L*-, *M*-, AND *R*-STATISTICS[1]

BY ROBERT J. SERFLING

The Johns Hopkins University

A class of statistics generalizing U-statistics and L-statistics, and containing other varieties of statistic as well, such as trimmed U-statistics, is studied. Using the differentiable statistical function approach, differential approximations are obtained and the influence curves of these generalized L-statistics are derived. These results are employed to establish asymptotic normality for such statistics. Parallel generalizations of M- and R-statistics are noted. Strong convergence, Berry-Esséen rates, and computational aspects are discussed.

1. Introduction. We consider a new class of statistics, which usefully generalizes the classes of U-statistics and L-statistics and contains other varieties of statistic as well. Let $X_1, \cdots, X_n$ be independent random variables having common probability distribution F. (More generally, the X_i's could be random elements of an arbitrary space.) Let a "kernel" $h(x_1, \cdots, x_m)$ be given, and denote by

$$W_{n,1} \leq \cdots \leq W_{n,n_{(m)}} \tag{1.1}$$

the ordered values of $h(X_{i_1}, \cdots, X_{i_m})$ taken over the $n_{(m)} = n(n-1) \cdots (n-m+1)$ m-tuples $(i_1, \cdots, i_m)$ of distinct elements from $\{1, \cdots, n\}$. Consider the statistics given by

$$\textstyle\sum_{i=1}^{n_{(m)}} c_{n,i} W_{n,i}, \tag{1.2}$$

where $c_{n,i}$, $1 \leq i \leq n_{(m)}$, are arbitrary constants. The form (1.2) is quite general. It includs the *U-statistic corresponding to the kernel h*, which is given by (1.2) with $c_{n,i} = 1/n_{(m)}$, all i. And it includes the class of *L-statistics* (linear functions of order statistics), given by (1.2) for the particular kernel $h(x) = x$. Moreover, it includes statistics such as

$$\text{median} \quad \{\tfrac{1}{2}(X_i + X_j),\ i \neq j\}, \tag{1.3}$$

which is a standard version of the well-known *Hodges-Lehmann location estimator*, but which is neither a U-statistic nor an L-statistic. Thus, for example, the sample mean, the sample median (a particular L-statistic), the sample variance (a particular U-statistic), the Hodges-Lehmann location estimator, and the 5% trimmed mean (an L-statistic)—a group of statistics which traditionally

Received September 1981; revised July 1983.
[1] Research supported by the U.S. Department of Navy under Office of Naval Research Contract No. N00014-79-C-0801. Reproduction in whole or in part is permitted for any purpose of the United States Government.
AMS 1970 *subject classifications*. Primary 62E20; secondary 60F05.
Key words and phrases. Order statistics, L-statistics, M-statistics, R-statistics, Hodges-Lehmann estimator, trimmed U-statistics, asymptotic normality.

have been viewed and analyzed as quite different types—may in fact be viewed from a single standpoint. In this way the form (1.2) provides a unifying concept relative to various familiar statistics. But (1.2) also embraces important new varieties of statistic. For example, "trimmed U-statistics" and "Winsorized U-statistics" fall in this class. In particular, a "trimmed variance" is defined by trimming the U-statistic corresponding to the kernel $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$. This provides a competitor to a somewhat similar nonparametric dispersion measure of Bickel and Lehmann (1976). Their measure is a trimmed variance which is simpler in form than a trimmed U-statistic but which is constructed assuming that the population median is known and incorporating its value into the measure. Likewise, one may consider trimmed versions of the pth power measures (of spread) introduced by Bickel and Lehmann (1979). Finally we note an important generalization of (1.3), namely the form

$$\text{median}\ \{m^{-1}(X_{i_1} + \cdots + X_{i_m})\}, \tag{1.4}$$

defined for a sample $X_1, \cdots, X_n$ and a fixed choice of m (a comparative study of the cases $m = 2, 3, 4, 5$ is in progress). Thus (1.2) represents a timely and effective generalization.

Computationally, statistics requiring ordering such as the sample median and the Hodges-Lehmann have been deemed less satisfactory than statistics computed by averaging (such as the sample mean, the U-statistics) or by solving equations (e.g., M-estimates). It has appeared, and been asserted, that computation of the sample median required $O(n \log n)$ operations and that computation of the Hodges-Lehmann required $O(n^2)$ operations. However, with the advent of computer science and the development of ingenious algorithms for machine computation, this misunderstanding has been corrected. Indeed, the sample median requires only $O(n)$ operations (see Blum et al., 1973, and Floyd and Rivest, 1975) and the Hodges-Lehmann only $O(n \log n)$ operations (see Shamos, 1976). Therefore, statistics of form (1.2) are not necessarily more formidable for machine computation than simpler types of statistic.

Despite the complexity and generality of the form (1.2), the usual asymptotic normality and convergence properties hold and can be expressed in explicit form. It turns out that for theoretical study of the class (1.2), it is appropriate to view the class as a generalization of L-statistics—hence the terminology "generalized L-statistics." This will become evident from the developments of Sections 2 and 3, where statistics of the form (1.2) will be represented as "statistical functions" in the spirit of von Mises (1947), corresponding differential approximations will be derived, leading also to the influence curves of Hampel (1974), and results on asymptotic normality will be obtained. By analogy with the development of Section 2, whereby "generalized L-statistics" are formulated as statistical functions, one can formulate generalized M-statistics and R-statistics. These and other complements are discussed in Section 4.

An interesting feature of the treatment of *generalized* L-, M- and R-statistics is that the role played by the usual sample distribution function in the treatment of *simple* L-, M- and R-statistics is given over to a more complicated type of empirical distribution function, one having the structure of a general U-statistic.

Accordingly, interesting generalizations of the well-developed theory of the usual empirical process become needed as a fundamental tool.

2. Generalized L-statistics: formulation, differential approximations, and influence curves. The representation of a statistic as a functional, evaluated at a sample distribution function which estimates the underlying actual distribution function, helps to identify what parameter the statistic in question is actually estimating. It also sets the stage for application of differentiation methodology and influence curve analysis. Let us examine generalized L-statistics relative to these aims. We proceed by analogy with the treatment of *simple* L-statistics.

As before, we consider a sample $X_1, \cdots, X_n$ of independent observations having distribution F. Denote by F_n the usual sample distribution function,

$$F_n(x) = (1/n) \textstyle\sum_{i=1}^n I(X_i \leq x), \quad -\infty < x < \infty,$$

where $I(A) = 1$ or 0 according as the event A holds or not. The class of (simple) L-statistics may be represented in the form

$$\textstyle\sum_{i=1}^n c_{n,i} F_n^{-1}(i/n), \tag{2.1}$$

of which a suitably wide subclass can be represented as $T(F_n)$ for a functional $T(\cdot)$ of the form

$$T(F) = \int_0^1 F^{-1}(t)J(t)\,dt + \textstyle\sum_{j=1}^d a_j F^{-1}(p_j). \tag{2.2}$$

Such a functional weights the quantiles $F^{-1}(t)$, $0 < t < 1$, of F according to a specified function $J(\cdot)$ for smooth weighting and/or specified weights $a_1, \cdots, a_d$ for discrete weighting. A particular *L-functional* is thus determined by specifying $J(\cdot)$, d, $p_1, \cdots, p_d$ and $a_1, \cdots, a_d$. The corresponding *L-statistic* is then simply $T(F_n)$. Note that $T(F_n)$ may be written in the form

$$T(F_n) = \textstyle\sum_{i=1}^n \left[\displaystyle\int_{(i-1)/n}^{i/n} J(t)\,dt\right] F_n^{-1}\left(\frac{i}{n}\right) + \textstyle\sum_{j=1}^d a_j F_n^{-1}(p_j), \tag{2.3}$$

which exhibits the statistic explicitly as a linear function of the order statistics $F_n^{-1}(i/n)$, $1 \leq i \leq n$.

We now designate an analogous subclass of the statistics of form (1.2). For a given kernel $h(x_1, \cdots, x_m)$, let H_n denote the empirical distribution function of the evaluations $h(X_{i_1}, \cdots, X_{i_m})$, i.e.,

$$H_n(y) = \frac{1}{n_{(m)}} \sum I[h(X_{i_1}, \cdots, X_{i_m}) \leq y], \quad -\infty < y < \infty,$$

where the sum is taken over the $n_{(m)}$ m-tuples $(i_1, \cdots, i_m)$ of distinct elements from $\{1, \cdots, n\}$. The statistics of form (1.2) may be represented in the form

$$\textstyle\sum_{i=1}^{n_{(m)}} c_{n,i} H_n^{-1}(i/n_{(m)}),$$

and, by analogy with (2.1) and (2.3), a wide and useful subclass of (1.2) is thus given in terms of the functional (2.2), by

$$T(H_n) = \sum_{i=1}^{n_{(m)}} \left[\int_{(i-1)/n_{(m)}}^{i/n_{(m)}} J(t)\, dt \right] H_n^{-1}\left(\frac{i}{n_{(m)}}\right) + \sum_{j=1}^{d} a_j H_n^{-1}(p_j). \tag{2.4}$$

The parameter estimated by this "generalized L-statistic" (GL-statistic) is given by $T(H_F)$, where H_F is the distribution function estimated by H_n, i.e.,

$$H_F(y) = P_F\{h(X_1, \cdots, X_m) \le y\}, \quad -\infty < y < \infty,$$

the distribution function of the random variable $h(X_1, \cdots, X_m)$.

This functional approach allows the *estimation error* $T(H_n) - T(H_F)$ to be approximated by a differential quantity, which in practice can be obtained as a certain Gâteaux differential. As in Serfling (1980), Chapter 6, let us in general define the kth *order Gâteaux differential* of a functional T at a distribution F in the direction of a distribution G to be

$$d_k T(F; G - F) = (d^k/d\lambda^k) T(F + \lambda(G - F))\,|_{\lambda=0+}, \tag{2.5}$$

provided that the given right-hand derivative exists. For the simple L-functional $T(\cdot)$ given by (2.2), we have the first-order Gâteaux differential

$$\begin{aligned} d_1 T(F; G - F) = &-\int_{-\infty}^{\infty} [G(y) - F(y)] J(F(y))\, dy \\ &+ \sum_{j=1}^{d} a_j \frac{p_j - G(F^{-1}(p_j))}{f(F^{-1}(p_j))}, \end{aligned} \tag{2.6}$$

assuming that F has a positive density f in neighborhoods of $p_1, \cdots, p_d$ (see Huber, 1977, or Serfling, 1980, for details). Accordingly, the estimation error $T(H_n) - T(H_F)$ of a GL-statistic becomes approximated by

$$\begin{aligned} d_1 T(H_F; H_n - H_F) = &-\int_{-\infty}^{\infty} [H_n(y) - H_F(y)] J(H_F(y))\, dy \\ &+ \sum_{j=1}^{d} a_j \frac{p_j - H_n(H_F^{-1}(p_j))}{h_F(H_F^{-1}(p_j))}, \end{aligned} \tag{2.7}$$

where h_f denotes the density of H_F, assumed to exist and be positive at $p_1, \cdots, p_d$.

A basic difference between the treatment of simple and generalized L-statistics, even though the same functional $T(\cdot)$ is involved in both cases, is that the quantity in (2.7) is a U-statistic in the more general case, but simply an average of IID's in the simple case. This stems from the fact that H_n, which assumes in the general treatment the role played by F_n in the simple case, is a U-statistic. That is, for each fixed y, $H_n(y)$ is the U-statistic corresponding to the kernel $I[h(x_1, \cdots, x_m) \le y]$. Consequently, $d_1 T(H_F{:}H_n - H_F)$ is seen to be the U-statistic corresponding to the kernel

$$
\begin{aligned}
A(x_1, \cdots, x_m) = &-\int_{-\infty}^{\infty} \{I[h(x_1, \cdots, x_m) \le y] - H_F(y)\}J(H_F(y))\, dy \\
&+ \textstyle\sum_{j=1}^{d} a_j \dfrac{p_j - I[h(x_1, \cdots, x_m) \le H_F^{-1}(p_j)]}{h_F(H_F^{-1}(p_j))}.
\end{aligned} \tag{2.8}
$$

The formulas (2.7) and (2.8) will be relevant in treating the convergence theory of $T(H_n)$ in Section 3.

Also, formula (2.8) may be interpreted as an analogue of the usual influence curve. In the special case of a *simple* L-statistic, the "influence curve" associated with the statistic $T(H_n) = T(F_n)$ is obtained by putting $G = \delta_x$ (the distribution placing mass 1 at x) in the formula (2.6), which then yields the function $A(x)$ given by (2.8) with $h(x) = x$. In this case $A(X_i)$ represents the approximate "influence" of the observation X_i on the estimation error when $T(F)$ is estimated by $T(F_n)$. (This interpretation, due to Hampel (1968), has become a standard concept in robustness considerations; see also Hampel, 1974, Huber, 1977.) Proceeding now to the *generalized* L-statistic, we see that $A(X_{i_1}, \cdots, X_{i_m})$ may be interpreted as the approximate influence of the combination of observations $X_{i_1}, \cdots, X_{i_m}$ on the estimation error when $T(H_F)$ is estimated by $T(H_n)$.

When the parameter of interest is represented by $T(H_F)$, for some functional T evaluated at a distribution H_F related to the distribution F of the observations, it is natural to use the estimator $T(H_n)$ based on an estimator H_n of H_F. As we have seen, however, the fact that H_n is in general a U-statistic introduces complications not present in the case of simple L-statistics, $T(F_n)$. Therefore, it is of some interest to view the parameter $T(H_F)$ as also, equivalently, the evaluation of some functional $\tilde{T}$ at the basic distribution F. That is, $\tilde{T}(\cdot)$ is defined by

$$
\tilde{T}(F) = T(H_F). \tag{2.9}
$$

From this standpoint, a natural estimator is $\tilde{T}(F_n)$, or equivalently $T(H_{F_n})$, where by definition

$$
\begin{aligned}
H_{F_n}(y) &= \int \cdots \int I[h(x_1, \cdots, x_m) \le y]\, dF_n(x_1) \cdots dF_n(x_m) \\
&= \frac{1}{n^m} \textstyle\sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} I[h(X_{i_1}, \cdots, X_{i_m}) \le y].
\end{aligned} \tag{2.10}
$$

Note that $H_n(y)$ and $H_{F_n}(y)$ are somewhat different, although closely related, estimators. Thus $T(H_n)$ and $\tilde{T}(F_n) = T(H_{F_n})$ are two somewhat different estimators of the single parameter expressed in two ways by (2.9). Although H_{F_n} is less straightforward than H_n for estimation of H_F, the estimator $T(H_{F_n}) = \tilde{T}(F_n)$ lends itself more straightforwardly to a standard influence curve analysis. Therefore, we derive the Gâteaux derivative of the functional $\tilde{T}$, as follows.

First, consider the special case of the functional

$$\tilde{T}_0(F) = H_F^{-1}(p),$$

where $0 < p < 1$, and assume that H_F has a density h_F in the neighborhood of $H_F^{-1}(p)$, with $h_F(H_F^{-1}(p)) > 0$. By standard methods, we obtain

$$d_1\tilde{T}_0(F; G - F) = m\,\frac{p - \int \cdots \int I[h(x_1, \cdots, x_m) \le H_F^{-1}(p)] \prod_{i=1}^{m-1} dF(x_i)\, dG(x_m)}{h_F(H_F^{-1}(p))}. \tag{2.11}$$

In particular, putting $G = \delta_x$ (the distribution placing mass 1 at x), we obtain the influence curve of the functional $\tilde{T}_0$,

$$IC(x; \tilde{T}_0, F) = m\,\frac{p - \int \cdots \int I[h(x_1, \cdots, x_{m-1}, x) \le H_F^{-1}(p)] \prod_{i=1}^{m-1} dF(x_i)}{h_F(H_F^{-1}(p))}. \tag{2.12}$$

Let us now consider the more general functional $\tilde{T}$ given by (2.9), i.e.,

$$\tilde{T}(F) = \int_0^1 H_F^{-1}(t)J(t)\, dt + \textstyle\sum_{j=1}^d a_j H_F^{-1}(p_j). \tag{2.13}$$

It follows in straightforward fashion, from the results for $\tilde{T}_0(\cdot)$, that

$$\begin{aligned} d_1T(F; G - F) &= -m \int_{-\infty}^{\infty} \left\{\int \cdots \int I[h(x_1, \cdots, x_m) \le y] \prod_{i=1}^{m-1} dF(x_i)\, dG(x_m) - H_F(y)\right\} \\ &\qquad \cdot J(H_F(y))\, dy \\ &\quad + m \sum_{j=1}^d a_j \frac{p_j - \int \cdots \int I[h(x_1, \cdots, x_m) \le H_F^{-1}(p_j)] \prod_{i=1}^{m-1} dF(x_i)\, dG(x_m)}{h_F(H_F^{-1}(p_j))}. \end{aligned} \tag{2.14}$$

Accordingly, the *influence curve* of the functional $\tilde{T}(\cdot)$ in (2.13) is

$$\begin{aligned} \mathrm{IC}(x; T, F) &= -m \int_{-\infty}^{\infty} \left\{\int \cdots \int I[h(x_1, \cdots, x_{m-1}, x) \le y] \prod_{i=1}^{m-1} dF(x_i) - H_F(y)\right\} \\ &\qquad \cdot J(H_F(y))\, dy \\ &\quad + m \sum_{j=1}^d a_j \frac{p_j - \int \cdots \int I[h(x_1, \cdots, x_{m-1}, x) \le H_F^{-1}(p_j)] \prod_{i=1}^{m-1} dF(p_i)}{h_F(H_F^{-1}(p_j))}. \end{aligned} \tag{2.15}$$

Since $\tilde{T}(F_n) = \tilde{T}(F)$ is approximately (under appropriate conditions)

$$d_1\tilde{T}(F; F_n - F) = (1/n) \textstyle\sum_{i=1}^n \mathrm{IC}(X_i; \tilde{T}, F), \tag{2.16}$$

the IC represents the approximate "influence" of the observation X_i on the estimation error, when $\tilde{T}(F)$ is estimated by $\tilde{T}(F_n)$.

3. Asymptotic normality of *GL*-statistics. Under appropriate conditions the GL-statistics $T(H_n)$and $\tilde{T}(F_n)$ are asymptotically normal in distribution:

$$n^{1/2}[T(H_n) - T(H_F)] \to_d N(0, \sigma^2(T, H_F)), \tag{3.1}$$

and

$$n^{1/2}[\tilde{T}(F_n) - \tilde{T}(F)] \to_d N(0, \sigma^2(\tilde{T}, F)), \tag{3.2}$$

where $\sigma^2(T, H_F) = \sigma^2(\tilde{T}, F) = \sigma^2$ is given by

$$\sigma^2 = \mathrm{Var}\{\mathrm{IC}(X; \tilde{T}, F)\}. \tag{3.3}$$

(Here $T(\cdot)$, $\tilde{T}(\cdot)$ and $\mathrm{IC}(\cdot)$ are as defined in (2.2), (2.9) and (2.15), respectively.) The asymptotic normality of $T(H_n)$ is established by making rigorous the approximation of $T(H_n) - T(H_F)$ by $d_1T(H_F{:}H_n - H_F)$ as given in (2.7), in which case (3.1) follows immediately from U-statistic theory (e.g., Serfling (1980), Chapter 5) and the asymptotic variance $\sigma^2(T, H_F)$ is given by $m^2\mathrm{Var}\{A_1(X)\}$, where $A_1(x) = E\{A(x, X_1, \cdots, X_{m-1})\}$ and $A(x_1, \cdots, x_m)$ is the function in (2.8). However, it is readily seen that $mA_1(x) = \mathrm{IC}(x; \tilde{T}, F)$, so that (3.3) is valid. Likewise, the asymptotic normality of $\tilde{T}(F_n)$ is established by approximating $\tilde{T}(F_n) - \tilde{T}(F)$ by $d_1\tilde{T}(F; F_n - F)$ and utilizing (2.16), in which case (3.2) follows directly from classical central limit theory and the appropriate asymptotic variance is given immediately by (3.3). Specifically, these assertions are formalized in the following results (we shall deal explicitly only with $T(H_n)$ and discuss $\tilde{T}(F_n)$ in Remark 3.2(ii) at the end of this section).

THEOREM 3.1. *Let H_F have positive derivatives at its p_j-quantiles, $1 \le j \le m$. Let $J(t)$ vanish for t outside $[\alpha, \beta]$, where $0 < \alpha < \beta < 1$, and suppose that on $[\alpha, \beta]$ J is bounded and continuous* a.e. *Lebesgue and* a.e. *H_F^{-1}. Assume that $0 < \sigma^2(T, H_F) < \infty$. Then* (3.1) *holds.*

The foregoing result applies to examples such as trimmed and Winsorized U-statistics. The following result applies to untrimmed J functions.

THEOREM 3.2. *Let H_F satisfy $\int [H_F(y)(1 - H_F(y))]^{1/2}\, dy < \infty$ and have positive derivatives at its p_j-quantiles, $1 \le j \le m$. Let J be continuous on* $[0, 1]$. *Assume that $0 < \sigma^2(T, H_F) < \infty$. Then* (3.1) *holds.*

To prove these results, the functional $T(H_F)$ is treated in two parts, $T(H_F) = T_1(H_F) + T_2(H_F)$, where $T_1(H_F) = \int_0^1 J(t)H_F^{-1}(t)\, dt$ and $T_2(H_F) = \sum_{j=1}^d a_jH_F^{-1}(p_j)$. We follow in part the treatment of simple L-functionals in Serfling (1980), Section 8.2.4. Define $\Delta_{in} = T_i(H_n) - T_i(H_F) - d_1T_i(H_F; H_n - H_F)$. Then

$$\Delta_{1n} = -\int_{-\infty}^{\infty} W_{H_n,H_F}(y)[H_n(y) - H_F(y)]\, dy, \tag{3.4}$$

where we define

$$W_{G_1,G_2}(y) = \frac{K(G_1(y)) - K(G_2(y))}{G_1(y) - G_2(y)} - J(G_2(y)), \quad G_1(y) \neq G_2(y),$$
$$= 0, \qquad G_1(y) = G_2(y),$$

and $K(u) = \int_0^u J(t)\,dt$. From (3.4) we obtain two inequalities,

$$|\Delta_{1n}| \leq \| W_{H_n,H_F} \|_{L_1} \cdot \| H_n - H_F \|_\infty, \tag{3.5}$$

and

$$|\Delta_{1n}| \leq \| W_{H_n,H_F} \|_\infty \cdot \| H_n - H_F \|_{L_1}, \tag{3.6}$$

where $\| g \|_\infty = \sup_x |g(x)|$ and $\| g \|_{L_1} = \int |g(x)|\,dx$. We seek to establish

$$\sqrt{n}\Delta_{1n} \to_p 0 \tag{3.7}$$

by analyzing the factors on the right-hand sides of (3.5) and (3.6).

LEMMA 3.1. *Let J be as in Theorem 3.1. Then*

$$\lim_{\| G_1 - G_2 \|_\infty \to 0} \| W_{G_1,G_2} \|_{L_1} = 0.$$

LEMMA 3.2. *Let J be as in Theorem 3.2. Then*

$$\lim_{\| G_1 - G_2 \|_\infty \to 0} \| W_{G_1,G_2} \|_\infty = 0.$$

(These are given as Lemmas 8.2.4A and 8.2.4E, respectively in Serfling, 1980.)

LEMMA 3.3. *If H_F is continuous, then*

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

PROOF. Silverman (1976) establishes that the empirical stochastic process of a U-statistic array (indeed, of a more general type of array),

$$n^{1/2}[H_n(H_F^{-1}(t)) - t], \quad 0 \leq t \leq 1,$$

converges weakly in the Skorohod topology to an a.s. continuous Gaussian process, say W^*. By continuity of the mapping $\| \cdot \|_\infty$ with respect to the Skorohod topology, it follows that $n^{1/2} \| H_n - H_F \|_\infty \to_d \| W^* \|_\infty$ and hence that $n^{1/2} \| H_n - H_F \|_\infty = O_p(1)$. □

LEMMA 3.4. *Let H_F satisfy $\int [H_F(1 - H_F)]^{1/2} < \infty$. Let J be as in Theorem 3.2. Then*

$$E\{\| H_n - H_F \|_{L_1}\} = O(n^{-1/2}).$$

PROOF. Adapting the proof of Lemma 8.2.4D of Serfling (1980), write

$$H_n(y) - H_F(y) = \binom{n}{m}^{-1} \sum \eta_y(X_{i_1}, \cdots, X_{i_m}),$$

with $\eta_y(\cdot) = I[h(\cdot) \le y] - H_F(y)$. Then

$$E\{\| H_n - H_F \|_{L_1}\} = \int E\{| n_{(m)}^{-1} \sum \eta_y(X_{i_1}, \cdots, X_{i_m}) |\} \, dy.$$

Now, by a result on U-statistics (Hoeffding, 1948; Serfling, 1980, page 183), and using Jensen's inequality, we have

$$E\{| n_{(m)}^{-1} \sum \eta_y(X_{i_1}, \cdots, X_{i_m}) |\} \le [(m/n) E\eta_y^2(X_1, \cdots, X_m)]^{1/2}.$$

Thus

$$E\{\| H_n - H_F \|_{L_1}\} \le m^{1/2} n^{-1/2} \int [H_F(1 - H_F)]^{1/2}. \quad \square$$

REMARK 3.1. In the proofs of Theorems 3.1 and 3.2, we will require (3.7). Note that this follows from the conditions of Lemmas 3.1 and 3.3 together, as well as from the conditions of Lemmas 3.2 and 3.4 together. □

Now, regarding Δ_{2n}, let us note that it may be written in the form

$$\Delta_{2n} = \textstyle\sum_{j=1}^{d} a_j \left[\hat{\xi}_{p_j,n} - \xi_{p_j} - \dfrac{p_j - H_n(p_j)}{h_F(\xi_{p_j})} \right], \tag{3.8}$$

where ξ_{p_j} denotes $H_F^{-1}(p_j)$ and $\hat{\xi}_{p_j,n}$ denotes $H_n^{-1}(p_j)$. In the case of simple L-statistics, the jth term above is recognized to be the remainder term R_n in the Bahadur representation for sample quantiles (see Bahadur, 1966, and Serfling, 1980, page 236). Bahadur (1966) showed, under second-order differentiability conditions on F, that $R_n =_{\text{wp1}} O(n^{-3/4}(\log n)^{3/4})$. Ghosh (1971) showed $R_n = O_p(n^{-1/2})$ under only first-order differentiability conditions on F. The extension of Ghosh's result to the more general situation involving the terms in (3.8) is straightforward (details omitted), and we have

LEMMA 3.5. *Let H_F have positive derivatives at its p_j-quantiles. Then* $n^{1/2}\Delta_{2n} \to_p 0$.

REMARKS 3.2.

(i) The proofs of Theorems 3.1 and 3.2 are now straightforward, from Remark 3.1, Lemma 3.5, and the discussion at the beginning of this section.

(ii) To treat the alternative estimator $\tilde{T}(F_n)$, note that we need to deal with $\tilde{\Delta}_{in} = T_i(F_n) - \tilde{T}_i(F) - d_1\tilde{T}_i(F; F_n - F)$, $i = 1, 2$. Since $\tilde{T}_i(F) = T_i(H_F)$ and $\tilde{T}_i(F_n) = T_i(H_{F_n})$, we have the following analogues of (3.5) and (3.6):

$$| \tilde{\Delta}_{1n} | \le \| W_{H_{F_n}, H_F} \|_{L_1} \cdot \| H_{F_n} - H_F \|_\infty \tag{3.9}$$

and

$$| \tilde{\Delta}_{1n} | \le \| W_{H_{F_n}, H_F} \|_\infty \cdot \| H_{F_n} - H_F \|_{L_1}. \tag{3.10}$$

The proof (of analogues of Theorems 3.1 and 3.2) utilizes Lemmas 3.1, 3.2 and 3.5 without change, but requires analogues of Lemmas 3.3 and 3.4 with H_n replaced by H_{F_n}. Evidently these entail additional conditions on the kernel h. We shall not pursue these details here. □

4. Complements. An M-estimate (of location) may be defined in terms of the M-functional $T(\cdot)$ defined by

$$\int \psi(x - T(F))\, dF(x) = 0, \tag{4.1}$$

where ψ is a given function. (See Huber, 1977, for example.) Just as we defined *generalized* L-functionals by replacing $T(F)$ by $T(H_F)$ for a specified L-functional $T(\cdot)$, we may define a *generalized* M-functional by putting H_F for F in (4.1). Thus a generalized M-statistic is given by $T(H_n)$. The analysis of such statistics follows standard lines with appropriate modifications due to the structure of H_n as a U-statistic. Likewise we may define and analyze generalized R-statistics.

Recent further work on generalized L-statistics includes a Berry-Esséen Theorem (Helmers, Janssen and Serfling, 1983) and extension to censored samples (Akritas, 1982). Further work on the empirical process of U-statistic structure includes weak convergence in metrics stronger than the Skorohod topology on $D[0, 1]$ (Silverman, 1983), a strong approximation (Csörgő, S., Horváth and Serfling, 1983) and Glivenko-Cantelli theorems (Helmers, Janssen and Serfling, 1983).

REFERENCES

AKRITAS, M. G. (1982). Empirical processes associated with V-statistics and a class of estimators under random censoring. Technical report, University of Rochester.

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BICKEL, P. J. and LEHMANN, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Ann. Statist.* **4** 1139–1158.

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

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

CSÖRGŐ, S., HORVÁTH, L. and SERFLING, R. (1983). An approximation for the empirical process of U-statistic structure. Technical report, Johns Hopkins University.

FLOYD, R. W. and RIVEST, R. (1975). Expected time bounds for selection. *Commun. A.C.M.* **18** 165–172.

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

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Univ. of California–Berkeley.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–397.

HELMERS, R., JANSSEN, P. and SERFLING, R. (1983). Glivenko-Cantelli theorems for EDF of U-statistic structure and SLLN and Berry-Esséen results for generalized L-statistics. Technical report, Johns Hopkins University.

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

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

ROYDEN, H. L. (1968). *Real Analysis,* 2nd ed. Macmillan, New York.

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

SHAMOS, M. J. (1976). Geometry and Statistics: problems at the interface. In *New Directions and Recent Results in Algorithms and Complexity,* ed. by J. F. Truab, Academic, New York.

SILVERMAN, B. W. (1976). Limit theorems for dissociated random variables. *Adv. 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.

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

DEPARTMENT OF MATHEMATICAL SCIENCES
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218