CONSISTENCY OF JACKKNIFE VARIANCE ESTIMATORS

by

Jun Shao
Purdue University

Technical Report #88-58

Department of Statistics
Purdue University

November 1988

◇

CONSISTENCY OF JACKKNIFE VARIANCE ESTIMATORS

Jun Shao
Purdue University

ABSTRACT

A class of delete-d jackknife estimators of the asymptotic variances of point estimators are shown to be consistent when d, the number of observations removed from the original sample, is a fraction of n and the point estimators are generated from a statistical functional which possesses a weak differentiability property. The computation of the delete-d jackknife estimators is almost as easy as the traditional delete-1 jackknife estimator. The results are applied to problems in robust M-estimation.

Keywords: Jackknife; Weak differentiability; M-estimators.

1. Introduction

The traditional jackknife estimator of the asymptotic variance of a point estimator $\hat{\theta}$ (which is based on n i.i.d. observations from a population and is used to estimate an unknown population parameter θ) is obtained by averaging the squared differences between point estimators calculated after removing one observation from the original sample and their average (Tukey, 1958). Although it is asymptotically valid in many cases, there are situations where the jackknife variance estimator is *inconsistent* (see Efron, 1982, Chapter 3). Shao and Wu (1989) studies the general delete-d jackknife, which is obtained by removing d observations at a time with d diverging to infinity at a certain rate, and explains why jackknife works or fails. In particular, they proved that when $\hat{\theta}$ is a sample quantile, the delete-d jackknife variance estimator is consistent while the traditional delete-1 jackknife estimator is not. Wu (1987) shows that the delete-d jackknife histograms are consistent estimators of the sampling distribution of $\hat{\theta}$ if and only if d diverges to infinity. It seems that the delete-d jackknife with a larger d requires less stringent smoothness condition on $\hat{\theta}$.

Shao and Wu (1989) established the consistency of delete-d jackknife variance estimator for a class of point estimators under several regularity conditions. One of these conditions is that the variance of $\hat{\theta}$ converges to the variance of the asymptotic distribution of $\hat{\theta}$. This excludes the situation where the variance of $\hat{\theta}$ does not exist (the asymptotic variance of $\hat{\theta}$ exists under weak conditions). Furthermore, even if the variance of $\hat{\theta}$ exists, it may be hard to check whether it converges to the asymptotic variance. In this paper, we show that for a class of point estimators $\hat{\theta}$ generated by a statistical functional admitting a certain differential, the delete-d jackknife variance estimator with d being a fraction of n is consistent without any moment condition on $\hat{\theta}$ (Section 2). A heuristic argument shows that the traditional delete-1 jackknife variance estimator may not work in this case, although its asymptotic property is still unknown. The result is applied to the situation where $\hat{\theta}$ is the commonly used M-estimator (Section 3).

2. The main results

Let F be an unknown population distribution and $X_1, \ldots, X_n$ be i.i.d. observations from F. In many application problems the parameter of interest is $\theta = T(F)$, where T is a functional defined on $\mathbf{F}$, a convex set of distribution functions containing F and all degenerate distributions. A point estimator of θ is $\hat{\theta} = T(F_n)$, where F_n is the empirical distribution function of $X_1, \ldots, X_n$. Under some regularity condition on T, $\hat{\theta} - \theta$ is asymptotically normal, i.e.,

$$n^{1/2}(\hat{\theta} - \theta) \to N(0, \sigma^2) \quad \text{in distribution}, \tag{2.1}$$

where σ^2 is usually unknown and σ^2/n is called the asymptotic variance of $\hat{\theta}$.

For various purposes in statistical analysis, one needs a consistent estimator of σ^2. The delete-d jackknife estimator of σ^2/n is defined as follows. Let $d = d_n < n$ be an integer and $\mathbf{S}_d$ be the collection of all the subsets of $\{1, \ldots, n\}$ which have size $n-d$. For $s = \{i_1, \ldots, i_{n-d}\} \in \mathbf{S}_d$, let F_{ns} be the empirical distribution function of $X_{i_1}, \ldots, X_{i_{n-d}}$ and $\hat{\theta}_s = T(F_{ns})$. The delete-$d$ jackknife estimator of σ^2/n is

$$s_d^2 = \frac{(n-d)}{dN} \sum_{s \in \mathbf{S}_d} (\hat{\theta}_s - \frac{1}{N} \sum_{s \in \mathbf{S}_d} \hat{\theta}_s)^2,$$

where $N = \binom{n}{d}$. When $d \equiv 1$, s_1^2 becomes the traditional jackknife variance estimator. In this paper, we focus on the following choice of d:

$$d = \text{the integer part of } \lambda n \tag{2.2}$$

for a fixed $\lambda \in (0,1)$. If d satisfies (2.2), N is very large for large n and therefore the computation of s_d^2 is impractical. We consider the following alternative. Draw a simple random sample (with or without replacement) $\{s_1, \ldots, s_m\}$ from $\mathbf{S}_d$, where m satisfies $m/n^{\delta} \to 1$ for a constant $\delta > 0$. Then define the delete-d jackknife estimator of σ^2/n to be

$$s_d^2(m) = \frac{(n-d)}{dm} \sum_{v=1}^{m} (\hat{\theta}_{s_v} - \frac{1}{m} \sum_{v=1}^{m} \hat{\theta}_{s_v})^2. \tag{2.3}$$

Note that when δ is nearly one, the amount of computation required for $s_d^2(m)$ is almost the same as that for s_1^2.

For an estimator s^2 of σ^2/n, s^2 is said to be consistent if

$$ns^2 \to_p \sigma^2,$$

where $\to_p$ denotes convergence in probability. To establish the consistency of $s_d^2(m)$, we need to assume some smoothness condition on T.

Definition. Let $\| \ \|$ be the sup-norm on $\mathbf{F}$.

(i) A functional T defined on $\mathbf{F}$ is Fréchet differentiable at $F \in \mathbf{F}$ if there is a real-valued function ϕ_F on $\mathbf{R}$ such that $E\phi_F(X_1)=0$, $E\phi_F^2(X_1)<\infty$ and

$$\frac{|T(G) - T(F) - \int \phi_F(x) d[G(x)-F(x)]|}{\|G-F\|} \to 0 \tag{2.4}$$

as $\|G-F\| \to 0$, $G \in \mathbf{F}$.

(ii) A functional T is weakly differentiable at F if (2.4) holds as $\|G-F\| \to 0$ and $|T(G)-T(F)| \to 0$.

The weak differentiability of T is substantially weaker than the Fréchet differentiability of T, since it requires (2.4) holds only for G satisfying both $T(G) \to T(F)$ and $\|G-F\| \to 0$. Note that the differentiability of T does not imply the continuity of T ($T(G) \to T(F)$ as $\|G-F\| \to 0$). Many commonly used estimators, such as L- and M-estimators, are generated by differentiable functionals (Boos, 1979; Clarke, 1983; Shao, 1988). See also Section 3. The function ϕ_F is called the influence curve, which is a measure of influence toward the estimation error $\hat{\theta}-\theta$ (Hampel, 1974).

A direct consequence from the differentiability of T at F is the asymptotic normality of $T(F_n)$. However, the existence of a differential asserts more, since it provides a useful tool of robust statistics through an analysis of the influence curve (see Huber, 1981). Furthermore, it ensures the consistency of the jackknife estimator $s_d^2(m)$.

Proposition 1. (i) If T is Fréchet differentiable at F, then (2.1) holds with $\sigma^2 = E\phi_F^2(X_1)$.

(ii) The same conclusion as in (i) holds if T is weakly differentiable at F and $T(F_n) \to_p T(F)$.

We now establish the main result.

Theorem 1. Let $\sigma^2 = E\phi_F^2(X_1)$ and $s_d^2(m)$ be defined as in (2.3) with d satisfying (2.2) and $m/n^\delta \to 1$.

(i) If T is Fréchet differentiable at F, then

$$ns_d^2(m) \to_p \sigma^2. \tag{2.5}$$

(ii) If T is weakly differentiable at F and

$$\max_{v \le m} |T(F_{ns_v}) - T(F)| \to_p 0, \tag{2.6}$$

then (2.5) holds.

Compared with the results in Shao and Wu (1989), Theorem 1 does not require any condition on the moment of $\hat{\theta}$ but requires more on the smoothness of $\hat{\theta}$. We prove the following lemmas first.

Lemma 1. Let $\{ s_1, \ldots, s_m \}$ be a simple random sample from $\mathbf{S}_d$ with $m/n^\delta \to 1$. Then

$$\max_{v \le m} \| F_{ns_v} - F \| \to_p 0.$$

Proof. Let $a > \min(1, 2\delta)$ be a constant and E_* be the expectation taken under the probability distribution corresponding to the random selection of s_v. Note that

$$\begin{aligned} E(\max_{v \le m} \| F_{ns_v} - F \|^a) &= E[E_*(\max_{v \le m} \| F_{ns_v} - F \|^a)] \\ &\le E(E_* \textstyle\sum_{v=1}^m \| F_{ns_v} - F \|^a) = mE(E_* \| F_{ns_1} - F \|^a) \\ &= mE(N^{-1} \textstyle\sum_{s \in \mathbf{S}_d} \| F_{ns} - F \|^a) = mE \| F_r - F \|^a, \end{aligned}$$

where $r = n - d$ and F_r is the empirical distribution of $X_1, \ldots, X_r$. From Dvoretzky, Kiefer and Wolfowitz inequality,

$$E(r^{1/2} \| F_r - F \|)^a = a \int t^{a-1} P(r^{1/2} \| F_r - F \| > t) dt \le c \int t^{a-1} e^{-2t^2} dt,$$

where c is a constant. Hence $mE\,\|F_r - F\|^a = O(mr^{-a/2}) = o(1)$, since $n^{\delta}/r^{a/2} \to 0$ under (2.2) and $a > 2\delta$. Thus, $\max_{v \le m} \|F_{ns_v} - F\|^a \to_p 0$ and the result follows. □

Lemma 2. Let z_i be i.i.d. with $Ez_i = 0$ and $Ez_i^2 = \alpha < \infty$, $\bar{z} = n^{-1}\sum_{i=1}^{n} z_i$ and $\bar{z}_s = (n-d)^{-1}\sum_{i \in s} z_i$. If $m \to \infty$ as $n \to \infty$ and d satisfies (2.2), then

$$\frac{n^2}{mN}\sum_{s \in \mathbf{S}_d}(\bar{z}_s - \bar{z})^4 \to_p 0.$$

Proof. Since $Ez_i = 0$ and $Ez_i^2 = \alpha$, $n\bar{z}^2 = O_p(1)$. Hence it suffices to show

$$\frac{n^2}{mN}\sum_{s \in \mathbf{S}_d}\bar{z}_s^4 \to_p 0.$$

Note that

$$\frac{n^2}{mN}\sum_{s \in \mathbf{S}_d}\bar{z}_s^4 \le \frac{2}{mn^2N}\sum_{s \in \mathbf{S}_d}\Big(\sum_{i \ne j,\, i,j \in s} z_i z_j\Big)^2 + \frac{2}{mn^2N}\sum_{s \in \mathbf{S}_d}\sum_{i \ne j,\, i,j \in s} z_i^2 z_j^2 + \frac{2}{mn^2N}\sum_{s \in \mathbf{S}_d}\sum_{i \in s} z_i^4. \tag{2.7}$$

The first and second terms on the right hand side of (2.7) are $o_p(1)$ since

$$E\Big[\frac{1}{n^2N}\sum_{s \in \mathbf{S}_d}\Big(\sum_{i \ne j,\, i,j \in s} z_i z_j\Big)^2\Big] = E\Big(\frac{1}{n^2N}\sum_{s \in \mathbf{S}_d}\sum_{i \ne j,\, i,j \in s} z_i^2 z_j^2\Big) = \frac{(n-d)(n-d-1)}{n^2}\alpha^2 \le \alpha^2.$$

The result follows from the fact that the third term on the right hand side of (2.7) is bounded by

$$\frac{1}{mn^2}\sum_{i=1}^{n} z_i^4,$$

which converges to zero almost surely by Marcinkiewicz strong law of large numbers. □

Lemma 3. Let z_i and $\bar{z}_s$ be the same as in Lemma 2 and $\{ s_1, \ldots, s_m \}$ be a simple random sample from $\mathbf{S}_d$ with m satisfying $m/n^{\delta} \to 1$. Then

$$\frac{n(n-d)}{dm}\sum_{v=1}^{m}\Big(\bar{z}_{s_v} - \frac{1}{m}\sum_{v=1}^{m}\bar{z}_{s_v}\Big)^2 \to_p \alpha.$$

Proof. Let P_* be the probability corresponding to the random selection of s_v and E_* and V_* be the expectation and variance taken under P_*, respectively. Note that

$$E_*[\frac{n(n-d)}{d}(\bar{z} - \frac{1}{m}\sum_{v=1}^{m}\bar{z}_{s_v})^2] = \frac{n(n-d)}{dmN}\sum_{s\in S_d}(\bar{z}_s - \bar{z})^2$$
$$= \frac{1}{m(n-1)}\sum_{i=1}^{n}(z_i - \bar{z})^2 \to 0 \ \ a.s.$$

by the strong law of large numbers. From

$$\frac{n(n-d)}{dm}\sum_{v=1}^{m}(\bar{z}_{s_v} - \frac{1}{m}\sum_{v=1}^{m}\bar{z}_{s_v})^2 = \frac{n(n-d)}{dm}\sum_{v=1}^{m}(\bar{z}_{s_v} - \bar{z})^2$$
$$+ \frac{n(n-d)}{d}(\bar{z} - \frac{1}{m}\sum_{v=1}^{m}\bar{z}_{s_v})^2 + \frac{2n(n-d)}{dm}\sum_{v=1}^{m}(\bar{z}_{s_v} - \frac{1}{m}\sum_{v=1}^{m}\bar{z}_{s_v})(\bar{z} - \frac{1}{m}\sum_{v=1}^{m}\bar{z}_{s_v})$$

and the Cauchy-Schwartz inequality, the result follows from

$$A_n = \frac{n(n-d)}{dm}\sum_{v=1}^{m}(\bar{z}_{s_v} - \bar{z})^2 \to_p \alpha.$$

From the sampling theory and (2.2), there is a constant $c>0$ such that

$$V_*(A_n) \le c\frac{n^2}{mN}\sum_{s\in S_d}(\bar{z}_s - \bar{z})^4.$$

From Lemma 2, $V_*(A_n) \to_p 0$. Then for any $\varepsilon>0$,

$$P_*(\ |A_n - E_*(A_n)|>\varepsilon\ ) \le \varepsilon^{-2}V_*(A_n) \to_p 0$$

and therefore

$$P(\ |A_n - E_*(A_n)|>\varepsilon\ ) = E[P_*(\ |A_n - E_*(A_n)|>\varepsilon\ )] \to 0.$$

That is, $A_n - E_*(A_n) = o_p(1)$. The result follows from

$$E_*(A_n) = \frac{n(n-d)}{dN}\sum_{s\in S_d}(\bar{z}_s - \bar{z})^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(z_i - \bar{z})^2 \to_p \alpha. \ \square$$

Proof of Theorem 1. (i) From the Fréchet differentiability of T at F, for any $\varepsilon>0$, there is a $\tau>0$ such that $R(G,F) = T(G) - T(F) - \int\phi_F dG$ satisfies $|R(G,F)| < \varepsilon\|G-F\|$ for any $G\in\mathbf{F}$ satisfying $\|G-F\|<\tau$. Then for any $\varepsilon_0>0$,

$$P(\ \frac{n(n-d)}{dm}\sum_{v=1}^{m}R^2(F_{ns_v},F)>\varepsilon_0\ )$$
$$\le P(\ \frac{n(n-d)}{dm}\sum_{v=1}^{m}\|F_{ns_v}-F\|^2>\varepsilon_0/\varepsilon^2\ ) + P(\ \max_{v\le m}\|F_{ns_v}-F\|>\tau\ ).$$

From Lemma 1, $P(\ \max_{v\le m}\|F_{ns_v}-F\|>\tau\ )\to 0$. Also,

$$\frac{n(n-d)}{dm}\sum_{v=1}^{m}\|F_{ns_v}-F\|^2 = O_p(1)$$

since

$$E\left[\frac{n(n-d)}{dm}\sum_{v=1}^{m}\|F_{ns_v}-F\|^2\right]=O(1)$$

from the proof of Lemma 1. Thus,

$$\frac{n(n-d)}{dm}\sum_{v=1}^{m}R^2(F_{ns_v},F)=o_p(1). \tag{2.8}$$

Let $z_i = \phi_F(X_i)$. Then

$$\hat{\theta}_{s_v}=\theta+(n-1)^{-1}\sum_{i\in s_v}z_i+R(F_{ns_v},F)$$

and from (2.8), Lemma 3 and Cauchy-Schwartz inequality,

$$ns_d^2(m)=\frac{n(n-d)}{dm}\sum_{v=1}^{m}(\bar{z}_{s_v}-\frac{1}{m}\sum_{v=1}^{m}\bar{z}_{s_v})^2+o_p(1).$$

Then (2.5) follows from Lemma 3.

(ii) The result in this part can be proved using condition (2.6) and the same argument as in (i). □

In many application problems the parameter of interest is a function of several population parameters: $\theta=g[T(F)]$, where $T=(T_1,\ldots,T_k)^\tau$, T_j are functionals on $\mathbf{F}$ and g is a real-valued differentiable function on $\mathbf{R}^k$. Let $\hat{\theta}=g[T(F_n)]$ and $\hat{\theta}_s=g[T(F_{ns})]$ for $s\in\mathbf{S}_d$. If T_j are differentiable at F with influence functions ϕ_j, $j=1,\ldots,k$, then (2.1) holds with

$$\sigma^2=[\nabla g(t_0)]^\tau\Sigma[\nabla g(t_0)], \tag{2.9}$$

where ∇g is the gradient of g, $t_0=T(F)$ and Σ is a $k\times k$ matrix whose (i,j)th element is $E\phi_i(X_1)\phi_j(X_1)$. The delete-d jackknife estimator of σ^2/n is defined the same as in (2.3). The following result shows the consistency of $s_d^2(m)$ in this case.

Theorem 2. Assume that T_j, $j=1,\ldots,k$, are weakly differentiable at F and satisfy condition (2.6) and ∇g is continuous at $T(F)$. Then (2.5) holds with σ^2 given by (2.9).

Proof. Let $t_s=T(F_{ns})$ and $t_m=m^{-1}\sum_{v=1}^{m}T(F_{ns_v})$. From the mean-value theorem,

$$\hat{\theta}_{s_v}=g(t_m)+\nabla g(t_m)(t_{s_v}-t_m)+W_{s_v},$$

where $W_s = [\nabla g(\xi_s) - \nabla g(t_m)]^\tau (t_s - t_m)$ and ξ_s is a point on the line segment between t_s and t_m. Let $\bar{W} = m^{-1}\sum_{v=1}^m W_{s_v}$. Then

$$
\begin{aligned}
ns_d^2(m) &= \frac{n(n-d)}{dm}[\nabla g(t_m)]^\tau \sum_{v=1}^m (t_{s_v}-t_m)(t_{s_v}-t_m)^\tau [\nabla g(t_m)] \\
&+ \frac{n(n-d)}{dm}\sum_{v=1}^m (W_{s_v}-\bar{W})^2 + \frac{2n(n-d)}{dm}[\nabla g(t_m)]^\tau \sum_{v=1}^m (t_{s_v}-t_m)(W_{s_v}-\bar{W}),
\end{aligned}
\tag{2.10}
$$

Under (2.6), $t_m \to_p T(F)$ and therefore the first term on the right hand side of (2.10) converges to σ^2 in probability. It remains to show that the second term on the right hand side of (2.10) is $o_p(1)$. But this follows from (2.6) and the continuity of ∇g at $T(F)$. □

We end this section with a remark about consistency of the traditional delete-1 jackknife variance estimator. It was shown that if T is differentiable uniformly in a neighborhood of F, then the delete-1 jackknife estimator s_1^2 is consistent (Parr, 1985; Shao, 1988). There are functionals which are differentiable at F but not uniformly differentiable. The following heuristic argument shows that Fréchet differentiability is not enough to guarantee the consistency of s_1^2. Let $F_{ni}=F_s$ with $s=\{1, \dots, i-1, i+1, \dots, n\}$, $R_{ni}=T(F_{ni})-T(F)-(n-1)^{-1}\sum_{j\neq i}\phi_F(X_j)$ and $\bar{R}=R_{ni}$. Following the proof of Theorem 1, we need

$$(n-1)\sum_{i=1}^n (R_{ni}-\bar{R})^2 = o_p(1)$$

for the consistency of s_1^2. This requires $(R_{ni}-\bar{R})^2 = o_p(n^{-2})$. However, Fréchet differentiability of T only guarantees that $R_{ni}^2 = o_p(n^{-1})$.

3. Applications to M-estimators

We apply the results in Section 2 to the problems involving commonly used M-estimators. Results for another type of commonly used estimators, the L-estimators, can be found in Parr and Schucany (1982) and Shao (1988).

An M-functional $T(F)$ is defined to be a solution θ of

$$\int \rho(x, \theta) dF(x) = \min_t \int \rho(x, t) dF(x), \tag{3.1}$$

where $\rho(x, t)$ is a real-valued function on $\mathbf{R}^2$, and the corresponding estimator $T(F_n)$ is called the M-estimator of θ. Examples of M-estimators can be found in Serfling (1980, Chapter 7).

Assume that $\psi(x, t) = \partial\rho(x, t)/\partial t$ and $\lambda_G(t) = \int\psi(x, t)dG(x)$ exist and λ_F is differentiable at θ with $\lambda_F'(\theta)\neq 0$. Shao (1988) showed that T is weakly differentiable at F with $\phi_F(x) = -\psi(x, \theta)/\lambda_F'(\theta)$, if $\|\psi(\cdot, \theta)\|_V < \infty$ and $\|\psi(\cdot, t) - \psi(\cdot, \theta)\|_V \to 0$ as $t\to\theta$, where $\| \ \|_V$ is the total variation norm. To apply Theorems 1 and 2, we need to check condition (2.6). If T is continuous, (2.6) follows from Lemma 1. However, an M-functional need not to be continuous. Under some conditions, we may establish the continuity of T or (2.6) directly (Proposition 2) and therefore Theorems 1 and 2 apply for these M-estimators.

Proposition 2. Let T be an M-functional defined in (3.1).

(i) Suppose that ψ is nondecreasing in t and there is a neighborhood N of θ such that for each fixed x, $\psi(x, t)$ is continuous on N, $|\psi(x, t)|^a \le M(x)$ for $t\in$N, where $a>2$ and $M(x)$ satisfies $\int M(x)dF(x) < \infty$. If m is chosen so that $m/n^\delta\to 1$ with $\delta<a/2$, then (2.6) holds.

(ii) Assume that (3.1) has a unique solution. If ψ is continuous in t and bounded and ρ satisfies that for any $c > 0$,

$$\lim_{t\to\infty}\rho(x, t) = \alpha \quad \text{uniformly in } x \text{ satisfying } |x| \le c, \tag{3.2}$$

where $\int\rho(x, t_0)dF(x) < \alpha \le \infty$, then T is continuous.

Proof. We give a proof for (i) only. Under the conditions in (i) and $\lambda_F'(\theta)\neq 0$, there is a neighborhood $\mathrm{N}_0\subset\mathrm{N}$ such that $\int\psi(x, t)dF(x)$ is continuous and strictly decreasing on N_0. Then (2.6) follows from

$$\max_{v\le m} \left|\int\psi(x, T(F_{ns_v}))dF(x)\right| \to_p 0. \tag{3.3}$$

By using a similar proof to that of Lemma 1, we can show that

$$\max_{v\le m}\sup_{t\in\mathrm{N}_0}\left|\int\psi(x, t)d[F_{ns_v}(x)-F(x)]\right| \to_p 0.$$

This and the fact that $\int\psi(x, t)dF(x)$ is decreasing in t imply that

$$P(\,T(F_{ns_v})\in\mathrm{N}_0 \text{ for all } v\,) \to 0$$

and therefore (3.3) holds. □

Note that for certain types of M-estimators, the computation of $T(F_{ns})$ is faster than that of $T(F_{ni})$ when d is larger than one. Therefore, the computation of the delete-d jackknife estimator (2.3) is faster than the computation of the delete-1 jackknife estimator.

References

Boos, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7**, 955-959.

Clarke, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11**, 1196-1205.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. SIAM, Philadelphia.

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

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.

Parr, W. C. (1985). Jackknifing differentiable statistical functionals. *J. R. Statist. Soc. B* **47**, 56-66.

Parr, W. C. and Schucany, W. R. (1982). Jackknifing L-statistics with smooth weight functions. *J. Amer. Statist. Assoc.* **77**, 629-638.

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

Shao, J. (1988). Statistical functionals, differentiability and jackknife. Technical Report #88-56, Department of Statistics, Purdue University.

Shao, J. and Wu, C. F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17**, to appear.

Tukey, J. (1958). Bias and confidence in not quite large samples. *Ann. Math. Statist.* **29**, 614.

Wu, C. F. J. (1987). On the asymptotic properties of the jackknife histograms. Tech. Report, Department of Statistics, University of Wisconsin, Madison.