The Annals of Statistics
1993, Vol. 21, No. 1, 61–75

DIFFERENTIABILITY OF STATISTICAL FUNCTIONALS AND CONSISTENCY OF THE JACKKNIFE

BY JUN SHAO[1]

University of Ottawa

In statistical applications the unknown parameter of interest can frequently be defined as a functional $\theta = T(F)$, where F is an unknown population. Statistical inferences about θ are usually made based on the statistic $T(F_n)$, where F_n is the empirical distribution. Assessing $T(F_n)$ (as an estimator of θ) or making large sample inferences usually requires a consistent estimator of the asymptotic variance of $T(F_n)$. Asymptotic behaviour of the jackknife variance estimator is closely related to the smoothness of the functional T. This paper studies the smoothness of T through the differentiability of T and establishes some general results for the consistency of the jackknife variance estimators. The results are applied to some examples in which the statistics $T(F_n)$ are L-, M-estimators and some test statistics.

1. Introduction. Statistical inferences about an unknown parameter θ are usually based on a point estimator $\hat{\theta}$ of θ. Frequently θ can be considered as $T(F)$, where F is the unknown population distribution and T is a functional on a space of distribution functions containing F, and the estimate $\hat{\theta}$ is then obtained by evaluating T at the empirical distribution function F_n based on i.i.d. samples $X_1, \dots, X_n$ from F. Often the statistical functional T possesses differentiability properties which provide information about the asymptotic behaviour of $\hat{\theta} - \theta = T(F_n) - T(F)$ as well as methods for statistical inferences. These ideas were first introduced by von Mises (1947) and studied by many other authors [e.g., Reeds (1976), Boos (1979), Boos and Serfling (1980), Serfling (1980), Huber (1981), Clarke (1983, 1986), Gill (1989) and van der Vaart (1991)].

When T has an appropriately defined differential at F, $T(F_n)$ can be expressed as

$$T(F_n) = T(F) + n^{-1} \sum_{i=1}^{n} \phi_F(X_i) + o_p(n^{-1/2}),$$

where ϕ_F is a real-valued function satisfying $E\phi_F(X_1) = 0$. If $E\phi_F^2(X_1) = \sigma^2$ is finite, then

$$n^{1/2}[T(F_n) - T(F)] \to_d N(0, \sigma^2). \tag{1.1}$$

Received September 1989; revised February 1992.
[1]Partially supported by Natural Sciences and Engineering Research Council of Canada.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62E99.
Key words and phrases. Gâteaux derivative, Fréchet differentiability, continuous differentiability, L-estimators, M-estimators, linear rank statistics, variance estimation.

Of course, the existence of a differential of T asserts more. It provides a useful tool to study the behaviour of the jackknife estimator of σ^2, which is the main topic of this paper. See also the discussion in Serfling (1980) and Clarke (1983).

Since σ^2 in (1.1) is unknown in general, a consistent estimator of σ^2 is required for the purposes of assessing $\hat{\theta}$ and making statistical inferences. The jackknife provides a nonparametric method of estimating σ^2 [see Tukey (1958) and Shao and Wu (1989)]. Let F_{ni} be the empirical distribution based on the samples $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$. The jackknife estimator of σ^2 is

$$v_J = (n-1) \sum_{i=1}^{n} \left[T(F_{ni}) - n^{-1} \sum_{j=1}^{n} T(F_{nj}) \right]^2.$$

Using v_J does not require knowing the form of the function ϕ_F and therefore avoids the theoretical derivation of ϕ_F. An essential asymptotic requirement for v_J is its consistency:

$$v_J \to \sigma^2 \quad \text{a.s.} \tag{1.2}$$

Result (1.2) has been established for some particular types of estimators [e.g., Miller (1964, 1968), Arvesen (1969), Reeds (1978) and Parr and Schucany (1982)]. Using a differential approach, Parr (1985) proved (1.2) for continuously Fréchet differentiable T with respect to (w.r.t.) ρ_∞ (see Definition 2.3), where ρ_∞ is the metric generated by the sup-norm. However, Fréchet differentiability w.r.t. ρ_∞ is too strong a requirement since some frequently used statistical functionals are not Fréchet differentiable w.r.t. ρ_∞. Beran (1984) and Sen (1988) considered other types of differentiability in studying asymptotic behavior of the jackknife estimator.

The purpose of this paper is to establish (1.2) using a unified method, a differential approach. The differential approach adopted here is different from those in Beran (1984), Parr (1985) and Sen (1988) in the following sense:

1. We are not limited to the use of the sup-norm metric ρ_∞. By considering a metric other than ρ_∞, we obtain a much larger class of differentiable functionals.
2. We weaken the differentiability condition required by Parr (1985) so that the results are obtained under very minimal requirement of the differentiability of T.

Our results are more general than the existing results in the literature. Applications of the general theory to commonly encountered statistical functionals, including those corresponding to functions of the sample mean, L- and M-estimators and some test statistics, are discussed throughout the paper.

2. Differentiability of statistical functions. Let T be a functional on $\mathscr{F}$, a convex set of p-dimensional distributions containing the unknown population F and all degenerate distributions. Let $\mathscr{D}$ be the linear space generated by $\mathscr{F}$.

DEFINITION 2.1. Let $\mathscr{S}$ be a class of subsets of $\mathscr{D}$. T is $\mathscr{S}$-differentiable at $G \in \mathscr{F}$ if there is a linear functional L_G on $\mathscr{D}$ such that for any $\mathscr{C} \in \mathscr{S}$,

$$\lim_{t \to 0} \frac{T(G + tD) - T(G) - L_G(tD)}{t} = 0,$$

uniformly in $D \in \mathscr{C}$ and $G + tD \in \mathscr{F}$.

If $\mathscr{S}$ is the collection of all single point subsets of $\mathscr{D}$, then the differentiability in Definition 2.1 is Gâteaux differentiability. Suppose that $\mathscr{D}$ is a topological space. If $\mathscr{S}$ is the collection of all bounded subsets of $\mathscr{D}$, then the differentiability in Definition 2.1 is Fréchet differentiability. If $\mathscr{S}$ is the collection of all compact subsets of $\mathscr{D}$, then the differentiability in Definition 2.1 is Hadamard differentiability. From the definition, Fréchet differentiability is stronger than Hadamard differentiability and Hadamard differentiability is stronger than Gâteaux differentiability. It is known that Gâteaux differentiability is too weak to be useful for establishing (1.1), the asymptotic normality of $T(F_n)$. By choosing a suitable topology on $\mathscr{D}$, Hadamard differentiability of T at F ensures (1.1).

Let ρ be a metric on $\mathscr{D}$. Then a natural topology on $\mathscr{D}$ is the one generated by ρ and the corresponding differentiability is called differentiability w.r.t. ρ. The most commonly used metric on $\mathscr{D}$ is the one generated by the sup-norm: for D_1 and $D_2 \in \mathscr{D}$, $\rho_\infty(D_1, D_2) = \|D_1 - D_2\|_\infty = \sup_x |D_1(x) - D_2(x)|$. However, it is necessary to consider other metrics. For example, let $T(G) = g(\int x\, dG(x))$, where g is a real-valued differentiable function. Then T is not necessarily Hadamard differentiable w.r.t. ρ_∞, but is Fréchet differentiable w.r.t. the metric generated by the L_1-norm: $\rho_1(D_1, D_2) = \|D_1 - D_2\|_1 = \int |D_1(x) - D_2(x)| dx$.

Note that if ρ_a and ρ_b are two metrics on $\mathscr{D}$ satisfying $\rho_b(D_1, D_2) \le c\rho_a(D_1, D_2)$ for a constant c and all $D_1, D_2 \in \mathscr{D}$, then differentiability w.r.t. ρ_b implies differentiability w.r.t. ρ_a. This suggests use of the metric

$$\rho_*(D_1, D_2) = \rho_1(D_1, D_2) + \rho_\infty(D_1, D_2), \qquad D_1, D_2 \in \mathscr{D}. \tag{2.1}$$

The class of functionals differentiable w.r.t. ρ_* is substantially larger than the class of functionals differentiable w.r.t. ρ_∞ or ρ_1.

However, even Fréchet differentiability does not ensure the consistency of the jackknife estimator v_J. For example, $T(G) = g(\int x\, dG)$ is Fréchet differentiable at F w.r.t. ρ_1 if g is differentiable at $\mu = \int x\, dF$. But if the derivative of g is not continuous at μ, v_J is not necessarily consistent. If g is continuously differentiable at μ, then v_J is consistent [Miller (1964)]. This indicates that the consistency of v_J requires a more stringent smoothness condition on T than the asymptotic normality of $T(F_n)$. It requires that T is differentiable continuously in some sense. Despite this requirement on smoothness of T and the fact that there are other data-resampling methods for variance estimation (e.g., the delete-d jackknife and the bootstrap interquartile range) which may require fewer conditions for their asymptotic validity, the jackknife estimator is widely used in practice because of its simplicity in computation (relative to other resampling methods).

DEFINITION 2.2. A functional T is continuously Gâteaux differentiable at $G \in \mathscr{F}$ if T is Gâteaux differentiable at G and for any sequences of numbers $t_k \to 0$ and $G_k \in \mathscr{F}$ satisfying $\rho_\infty(G_k, G) \to 0$,

$$\lim_{k \to \infty} \left[\frac{T(G_k + t_k(\delta_x - G_k)) - T(G_k)}{t_k} - L_G(\delta_x - G_k) \right] = 0,$$

uniformly in x, where δ_x is the distribution degenerated at the point x.

Note that if T is a function on the real line, then $\mathscr{S}$-differentiability in Definition 2.1 is the same as the ordinary differentiability of T and the differentiability in Definition 2.2 corresponds to that T is continuously differentiable in the ordinary sense.

When the metric ρ_∞ is used, continuous Gâteaux differentiability is just enough for establishing the consistency of v_J. If a metric other than ρ_∞ is used, we may need to consider a stronger differentiability, which is a generalization of Definition 1 in Parr (1985).

DEFINITION 2.3. A functional T is continuously Fréchet differentiable at $G \in \mathscr{F}$ w.r.t. ρ if T is Fréchet differentiable at G w.r.t. ρ and $\rho(G_k, G) \to 0$ and $\rho(H_k, G) \to 0$ imply

$$\lim_{k \to \infty} \frac{T(H_k) - T(G_k) - L_G(H_k - G_k)}{\rho(H_k, G_k)} = 0.$$

It can be verified that continuous Fréchet differentiability w.r.t. ρ_∞ implies continuous Gâteaux differentiability.

We now study some examples of continuously differentiable functionals.

EXAMPLE 2.1 (Functions of means). Let g be a function on $\mathbb{R}^p$ and $T(G) = g(\int x\,dG)$. Then $T(F_n)$ is $g(\bar{X})$, where $\bar{X}$ is the sample mean. If g is differentiable at $\mu = \int x\,dF$ and the derivative ∇g is continuous at μ, then T is continuously Fréchet differentiable at F w.r.t. ρ_1.

EXAMPLE 2.2 (L-estimators). Consider the functional

$$T(G) = \int xJ(G(x))\,dG(x), \tag{2.2}$$

where $J(t)$ is a function defined on $[0, 1]$. Examples of T can be found in Serfling [(1980), Chapter 8]. $T(F_n)$ is called the L-estimator of $T(F)$. From Parr (1985),

$$T(G) - T(H) = \int \phi_F(x)\,d[G(x) - H(x)] + R(G, H)$$

with

$$\phi_F(x) = -\int [I(y \geq x) - F(y)] J[F(y)]\, dy, \tag{2.3}$$
$$R(G,H) = \int W[G(x), H(x)][H(x) - G(x)]\, dx,$$

where $I(A)$ is the indicator function of the set A, $W[G(x), H(x)] = 0$ if $G(x) = H(x)$ and $= [G(x) - H(x)]^{-1}\int_{H(x)}^{G(x)} J(t)\, dt - J[F(x)]$ if $G(x) \neq H(x)$. Parr (1985) showed that if J is bounded, continuous a.e. Lebesgue and a.e. F^{-1}, and 0 outside of $[\alpha, 1-\alpha]$ for a constant $\alpha > 0$, then T is continuously Fréchet differentiable w.r.t. ρ_∞. However, T may not be Hadamard differentiable w.r.t. ρ_∞ if J is untrimmed, that is, $J(t) \neq 0$ when t is near 0 or 1. In general, T is also not necessarily Hadamard differentiable w.r.t. ρ_1. Using the metric ρ_* in (2.1), we have the following result.

THEOREM 2.1. *Let T be given by* (2.2). *If J is bounded, continuous a.e. Lebesgue and a.e. F^{-1}, and continuous on* $[0, \alpha) \cup (1-\alpha, 1]$ *for a constant* $\alpha > 0$, *then T is continuously Fréchet differentiable at F w.r.t. ρ_*.*

EXAMPLE 2.3 (M-estimators). The M-functional $T(G)$ is defined to be a solution of

$$\int r(x, T(G))\, dG(x) = \min_t \int r(x,t)\, dG(x), \tag{2.4}$$

where $r(x,t)$ is a real-valued function on $\mathbb{R}^{p+1}$. Examples of M-functionals can be found in Serfling [(1980), Chapter 7]. Let $\theta = T(F)$. $T(F_n)$ is called the M-estimator of θ. Assume that $\psi(x,t) = \partial r(x,t)/\partial t$ exists and $\lambda_G(t) = \int \psi(x,t)\, dG(x)$ is well defined. Consequently, $\lambda_G(T(G)) = 0$. Assume further that λ_G is differentiable at $T(G)$ with $\lambda'_G(T(G)) \neq 0$. Define $h_G(t,s) = [\lambda_G(t) - \lambda_G(s)]/(t-s)$ if $s \neq t$ and $= \lambda'_G(s)$ if $s = t$. Then

$$T(H) - T(G) = \int \phi_F(x)\, d[H(x) - G(x)] + R_1 + R_2 + R_3$$

with $\phi_F(x) = -\psi(x,\theta)/\lambda'_F(\theta)$,

$$R_1 = [\lambda'_F(\theta)]^{-1} \int \psi(x,\theta)\, d[H(x) - G(x)] \tag{2.5a}$$
$$- [\lambda'_G(\theta)]^{-1} \int \psi(x, T(G))\, d[H(x) - G(x)],$$

$$R_2 = \left\{ [\lambda'_G(\theta)]^{-1} - [h_G(T(H), T(G))]^{-1} \right\} \tag{2.5b}$$
$$\times \int \psi(x, T(G))\, d[H(x) - G(x)]$$

and

$$R_3 = [h_G(T(H), T(G))]^{-1} \tag{2.5c}$$
$$\times \int [\psi(x, T(G)) - \psi(x, T(H))]\, d[H(x) - G(x)].$$

Clarke (1983, 1986) proved that T is Fréchet differentiable w.r.t. ρ_∞ under some conditions on the function ψ. The following result shows continuous Fréchet differentiability of T.

THEOREM 2.2. *Let T be given by* (2.4). *Suppose that as $k \to \infty$,*

$$T(G_k) \to \theta \quad \text{for } \rho_\infty(G_k, F) \to 0 \tag{2.6}$$

and

$$\lambda'_{G_k}(\xi_k) \to \lambda'_F(\theta) \quad \text{for } \rho_\infty(G_k, F) \to 0 \text{ and } \xi_k \to \theta. \tag{2.7}$$

(i) *Assume that $\|\psi(\cdot, t) - \psi(\cdot, \theta)\|_v \to 0$ as $t \to \theta$, where $\| \ \|_v$ is the total variation norm* [*Natanson* (1961)] *and that*

$$\lambda_H(T(G)) = O(\rho_\infty(H, G)). \tag{2.8}$$

Then T is continuously Fréchet differentiable at F w.r.t. ρ_∞.

(ii) *Assume that there is a neighborhood N_θ of θ such that for $t \in N_\theta$, $q(x,t) = \partial\psi(x,t)/\partial x$ is bounded. Assume further that for each $t \in N_\theta$ there is a set D_t such that as $t \to \theta$, $m(D_t) \to 0$ and $\sup_{x \in D_t^c}|q(x,t) - q(x,\theta)| \to 0$, where m is the Lebesgue measure and D_t^c is the complement of D_t. Then T is continuously Fréchet differentiable at F w.r.t. ρ_*.*

REMARKS. (i) If condition (2.8) in (i) is replaced by

$$\lambda_H(T(G)) = O(\rho_*(H, G)),$$

then T is continuously Fréchet differentiable w.r.t. ρ_*.

(ii) Clarke (1983, 1986) established some results for the continuity of T which implies (2.6). Condition (2.7) is implied by A_4 in Clarke (1983) or A'_4 in Clarke (1986). In particular, (2.7) is satisfied if both $\psi(x,t)$ and $\partial\psi(x,t)/\partial t$ are bounded and continuous.

(iii) A sufficient condition for (2.8) is $\|\psi(\cdot, \theta)\|_v < \infty$, since

$$|\lambda_H(T(G))| \le \|\psi(\cdot, T(G))\|_v \rho_\infty(H, G).$$

EXAMPLE 2.4 (Linear rank statistics). Let $\mathscr{F} =$ {all distributions on $\mathbb{R}$} and for $G \in \mathscr{F}$,

$$T(G) = \int_0^\infty J(\tilde{G}(x))\, dG(x), \tag{2.9}$$

where J is a differentiable function on $[0, 1]$ and satisfies $J(1 - t) = -J(t)$

and

$$\tilde{G}(x) = G(x) - G((-x)-), \qquad x \geq 0.$$

$T(F_n)$ is then a linear rank statistic. Note that Wilcoxon signed rank statistic and Winsorized signed rank statistic are special cases of $T(F_n)$. For any $G \in \mathscr{F}$, T in (2.9) is Hadamard differentiable at G w.r.t. ρ_∞ and

$$L_G(D) = \int_0^\infty J'(\tilde{G}(x))\,\tilde{D}(x)\,dG(x) + \int_0^\infty J(\tilde{G}(x))\,dD(x), \qquad D \in \mathscr{D}.$$

T is not Fréchet differentiable w.r.t. ρ_∞, but the following result shows that T is continuously Gâteaux differentiable at F.

THEOREM 2.3. *Suppose that J' is continuous on* $[0,1]$ *and* $\|J'\|_v < \infty$. *Then T in* (2.9) *is continuously Gâteaux differentiable at F.*

EXAMPLE 2.5 (Cramér–von Mises test statistic). Let F_0 be a specified hypothetical distribution and

$$T(G) = \int [G(x) - F_0(x)]^2\,dF_0(x).$$

$T(F_n)$ is then the Cramér–von Mises test statistic for the test problem: H_0: $F = F_0$ versus H_1: $F \neq F_0$. Let $L_F(D) = 2\int D(x)[F(x) - F_0(x)]\,dF_0(x)$. Then

$$\begin{aligned}|T(H) - T(G) - L_F(H-G)| &= \left|\int [H(x) - G(x)][H(x) + G(x) - 2F(x)]\,dF_0(x)\right| \\ &\leq \rho_\infty(H,G)[\rho_\infty(H,F) + \rho_\infty(G,F)].\end{aligned}$$

Hence T is continuously Fréchet differentiable at F w.r.t. ρ_∞.

EXAMPLE 2.6 (Two-sample Wilcoxon statistics). Let $\mathscr{F}_1 = \mathscr{F}_2 = \{$all q-dimensional distributions$\}$, $\mathscr{F} = \{$all distributions of the product measure $F \times G$, $F \in \mathscr{F}_1$, $G \in \mathscr{F}_2\}$, and

$$T(F,G) = \int F(x)\,dG(x), \qquad (F,G) \in \mathscr{F}.$$

This functional T plays an important role in many applications [Gill (1989)] and if F_n and G_m are the empirical distributions based on the data from F and G, respectively, then $T(F_n, G_m)$ is the two-sample Wilcoxon statistic. Gill (1989) showed that T is Hadamard differentiable w.r.t. ρ_∞ with $L_{F,G}(D_1, D_2) = \int F(x)\,dD_2(x) + \int D_1(x)\,dG(x)$. Let t_k and s_k be two sequences satisfying $t_k \to 0$, $s_k = O(t_k)$ and $t_k = O(s_k)$. Let $r_k = t_k s_k/(t_k + s_k)$ and $F_k \in \mathscr{F}_1$, $G_k \in \mathscr{F}_2$ satisfying $\rho_\infty(F_k, F) \to 0$ and $\rho_\infty(G_k, G) \to 0$. Then a straightforward calculation shows that

$$\begin{aligned}&\big|T\big(F_k + t_k(\delta_x - F_k), G_k + s_k(\delta_y - G_k)\big)\\&-T(F_k, G_k) - L_{F,G}\big(t_k(\delta_x - F_k), s_k(\delta_y - G_k)\big)\big|\\&\quad = \bigg|t_k s_k \int (\delta_x - F_k)(u)\, d(\delta_y - G_k)(u)\\&\qquad + s_k \int (F_k - F)(u)\, d(\delta_y - G_k)(u) + t_k \int (\delta_x - F_k)(u)\, d(G_k - G)(u)\bigg|\\&\quad \le 2|t_k s_k| + |s_k| \rho_\infty(F_k, F) + |t_k| \rho_\infty(G_k, G) = o(r_k).\end{aligned}$$

Hence T is continuously Gâteaux differentiable at (F, G).

3. Consistency of the jackknife.

Throughout this section we assume that (1.1) holds with an unknown σ^2. We now establish the consistency of the jackknife estimator v_J.

THEOREM 3.1. *Assume that T is continuously Gâteaux differentiable at F and*

$$E\phi_F(X_1) = 0 \quad \text{and} \quad E\phi_F^2(X_1) = \sigma^2 < \infty, \tag{3.1}$$

where $\phi_F(x) = L_F(\delta_x - F)$. Then

$$v_J \to \sigma^2 \quad a.s. \tag{3.2}$$

PROOF. Note that $F_{ni} = F_n + t_n(\delta_{X_i} - F_n)$ with $t_n = -1/(n-1)$. Then the continuous differentiability of T and the fact that $\rho_\infty(F_n, F) \to 0$ a.s. imply

$$\frac{T(F_{ni}) - T(F_n)}{t_n} - L_F\big(\delta_{X_i} - F_n\big) \to 0,$$

uniformly in i, a.s. Hence

$$\max_{i \le n} \big|(n-1)[T(F_n) - T(F_{ni})] - \big(Z_i - \bar{Z}\big)\big| \to 0 \quad \text{a.s.},$$

where $Z_i = \phi_F(X_i)$ and $\bar{Z} = n^{-1}\sum_{i=1}^n Z_i$. Then

$$v_J = \frac{1}{n-1} \sum_{i=1}^n \big(Z_i - \bar{Z}\big)^2 + o(1) \quad \text{a.s.},$$

and the result follows from the strong law of large numbers (SLLN) under condition (3.1). □

THEOREM 3.2. *Let ρ be a metric on $\mathscr{F}$. Assume that ρ satisfies*

$$\rho(F_n, F) \to 0 \quad a.s. \quad \text{and} \quad \sum_{i=1}^n [\rho(F_{ni}, F_n)]^2 = O(n^{-1}) \quad a.s. \tag{3.3}$$

If T is continuously Fréchet differentiable at F w.r.t. ρ and (3.1) *holds, then* (3.2) *holds.*

PROOF. Let Z_i be defined as in the proof of Theorem 3.1,

$$R_{ni} = T(F_{ni}) - T(F_n) - (n-1)^{-1} \sum_{j \neq i} Z_j + \bar{Z}$$

and $\bar{R} = n^{-1}\sum_{i=1}^n R_{ni}$. Then

$$v_J = (n-1)^{-1} \sum_{i=1}^{n} \left(Z_i - \bar{Z}\right)^2 + (n-1) \sum_{i=1}^{n} \left(R_{ni} - \bar{R}\right)^2$$
$$+ 2(n-1) \sum_{i=1}^{n} R_{ni} \left[(n-1)^{-1} \sum_{j \neq i} Z_i - \bar{Z}\right].$$

From the SLLN, $(n-1)^{-1}\sum_{i=1}^n (Z_i - \bar{Z})^2 \to \sigma^2$ a.s. It remains to show that $(n-1)\sum_{i=1}^n R_{ni}^2 \to 0$ a.s. From (3.3),

$$\max_{i \le n} \rho(F_{ni}, F) \le \rho(F_n, F) + \max_{i \le n} \rho(F_{ni}, F_n) \to 0 \quad \text{a.s.}$$

If T is continuously Fréchet differentiable at F w.r.t. ρ, then for any $\varepsilon > 0$,

$$R_{ni}^2 \le \varepsilon^2 [\rho(F_{ni}, F_n)]^2 \quad \text{for all } i \le n \text{ and sufficiently large } n.$$

Thus, $(n-1)\sum_{i=1}^n R_{ni}^2 \le \varepsilon^2 (n-1)\sum_{i=1}^n [\rho(F_{ni}, F_n)]^2$ and $(n-1)\sum_{i=1}^n R_{ni}^2 \to 0$ a.s. follows from (3.3). This proves (3.2). □

Note that Theorem 3.2 is mainly for the situations where Theorem 3.1 is not applicable, that is, T is not continuously Gâteaux differentiable and a metric ρ other than ρ_∞ has to be considered. The metric ρ, however, has to have property (3.3). The following lemma shows that (3.3) holds for ρ_1 if the second moment of F exists. Since $\rho_\infty(F_{ni}, F_n) \le n^{-1}$, (3.3) also holds for the metric ρ_* in (2.1).

LEMMA 3.1. *Assume that $E|X_1|^2 < \infty$. Then*

$$\int |F_n(x) - F(x)|\,dx \to 0 \quad a.s.,$$

and

$$\sum_{i=1}^{n} \left[\int |F_{ni}(x) - F_n(x)|\,dx\right]^2 = O(n^{-1}) \quad a.s.$$

Thus, v_J is a consistent estimator of σ^2 when T is either continuously Gâteaux differentiable at F or continuously Fréchet differentiable at F w.r.t. ρ_*. Theorem 3.2 is applicable to Example 2.1 (functions of sample means), Example 2.2 (L-estimators) and Example 2.5 (Cramér–von Mises test statistic). For the L-estimators, the condition on J we required in Theorem 2.1 is much weaker than that in Parr and Schucany [(1982), Theorem 2]. Theorem 3.1 is applicable to Example 2.4 (linear rank statistic) and Example 2.6 (two sample Wilcoxon statistic), since T are continuously Gâteaux differentiable in these examples. Theorem 3.2 is also applicable to Example 2.3 (M-estimators), since by Theorem 2.2, T in (2.4) is continuously Fréchet differentiable w.r.t. ρ_∞ or ρ_*.

For M-estimators, the consistency of v_J can be established if condition (2.6) in Theorem 2.2 is replaced by a weaker condition

$$\max_{i \le n} |T(F_{ni}) - \theta| \to 0 \quad \text{a.s.,} \tag{3.4}$$

which is a necessary condition for the consistency of v_J if $T(F_n) \to \theta$ a.s. For example, (3.4) holds if ψ is nondecreasing in t and there is a neighbourhood N_θ of θ such that for each fixed x, $\psi(x, t)$ is continuous on N_θ, $|\psi(x, t)| \le M(x)$ for $t \in N_\theta$ and $EM(X_1) < \infty$.

In some cases we need to consider a function of several functionals: $g \circ T$, where T is a d-vector whose components are functionals on $\mathscr{F}$ and g is a real-valued function on R^d.

THEOREM 3.3. *Suppose that the gradient ∇g is continuous at $\theta = T(F)$ and that the components of T satisfy the conditions in either Theorem 3.1 or Theorem 3.2. Then the jackknife estimator for $g \circ T(F_n)$ is consistent, that is,*

$$v_{gJ} = (n-1) \sum_{i=1}^{n} \left[g \circ T(F_{ni}) - n^{-1} \sum_{j=1}^{n} g \circ T(F_{nj}) \right]^2$$
$$\to \nabla g(\theta) V [\nabla g(\theta)]^\tau \quad a.s.,$$

where V is the asymptotic covariance matrix of $n^{1/2}[T(F_n) - T(F)]$.

PROOF. The conditions on ∇g and T imply that

$$(n-1) \sum_{i=1}^{n} \left[\nabla g(T(F_n)) \left(T(F_{ni}) - n^{-1} \sum_{j=1}^{n} T(F_{nj}) \right) \right]^2 \to \nabla g(\theta) V [\nabla g(\theta)]^\tau \quad \text{a.s.,} \tag{3.5}$$

and

$$(n-1) \sum_{j=1}^{n} \| T(F_{ni}) - T(F_n) \|^2 \to \operatorname{trace}(V) \quad \text{a.s.,} \tag{3.6}$$

where $\| \; \|$ is the Euclidean norm. From the mean-value theorem,

$$v_{gJ} = (n-1) \sum_{i=1}^{n} \left[\nabla g(T(F_n)) \left(T(F_{ni}) - n^{-1} \sum_{j=1}^{n} T(F_{nj}) \right) \right]^2$$
$$+ (n-1) \sum_{i=1}^{n} \left(U_{ni} - n^{-1} \sum_{j=1}^{n} U_{nj} \right)^2$$
$$+ 2(n-1) \sum_{i=1}^{n} \nabla g(T(F_n)) T(F_{ni}) \left(U_{ni} - n^{-1} \sum_{j=1}^{n} U_{nj} \right),$$

where $U_{ni} = [\nabla g(\xi_{ni}) - \nabla g(T(F_n))][T(F_{ni}) - T(F_n)]$ and ξ_{ni} lies between $T(F_{ni})$ and $T(F_n)$. From (3.6) and the continuity of ∇g at θ,

$$\max_{i \le n} \|\nabla g(\xi_{ni}) - \nabla g(T(F_n))\| \to 0 \quad \text{a.s.}$$

Hence

$$(3.7)\quad (n-1)\sum_{i=1}^{n} U_{ni}^2 \le o(1)(n-1)\sum_{i=1}^{n} \|T(F_{ni}) - T(F_n)\|^2 = o(1) \quad \text{a.s.},$$

and the result follows from (3.5), (3.7) and the Cauchy–Schwarz inequality. □

As a final remark, we indicate that the differential approach for establishing consistency of the jackknife estimators can be handily applied to some situations where the observations $X_1, \dots, X_n$ are non-i.i.d. For example, for independent but not necessarily identically distributed $X_1, \dots, X_n$, the jackknife estimator v_J is still consistent as an estimator of the asymptotic variance of $n^{1/2}[T(F_n) - T(\bar{F})]$, where $\bar{F}$ is the average of the distributions of $X_1, \dots, X_n$, as long as T is continuously differentiable, (3.3) holds,

$$E\phi_F(X_t) = 0, \qquad \sup_i E\phi_F^2(X_i) < \infty$$

and

$$\frac{1}{n-1}\sum_{i=1}^{n}\left[\phi_F(X_i) - n^{-1}\sum_{j=1}^{n}\phi_F(X_i)\right]^2 - \frac{1}{n}\sum_{i=1}^{n} E\phi_F^2(X_i) \to 0 \quad \text{a.s.}$$

In fact, (3.3) holds for ρ_∞ and for ρ_1 under an additional condition $\sup_i E|X_i|^{2+\delta} < \infty$ for a constant $\delta > 0$.

4. Proofs.

PROOF OF THEOREM 2.1. Let $A = \{x\colon F(x) \le c\}$ and $B = \{x\colon c \le F(x) \le 1 - c\}$ with $c = \alpha/2$. If $F(x) \in A$ and $\rho_\infty(G, F) + \rho_\infty(H, F) \le \delta$, then $G(x), H(x) \in [0, c + \delta]$. Let $\delta \le \alpha/2$. Since $J(t)$ is uniformly continuous on $[0, c + \delta]$,

$$(4.1)\quad \begin{aligned} &\left|\int_A W[G(x), H(x)][G(x) - H(x)]\,dx\right| \Big/ \rho_*(G, H) \\ &\quad \le \sup_{x \in A} |W[G(x), H(x)]| \to 0 \end{aligned}$$

as $\rho_*(G, F) + \rho_*(H, F) \to 0$. Similarly, (4.1) holds with A replaced by $\{x\colon F(x) \ge 1 - c\}$. Note that there are constants a and b such that $B \subset [a, b]$. Then

$$\begin{aligned} &\left|\int_B W[G(x), H(x)][G(x) - H(x)]\,dx\right| \Big/ \rho_*(G, H) \\ &\quad \le \int_a^b |W[G(x)], H(x)]|\,dx \to 0 \end{aligned}$$

as $\rho_*(G, F) + \rho_*(H, F) \to 0$, since $|W[G(x), H(x)]| \to 0$ if $J \circ F$ is continuous at x and $\|W(G, H)\|_\infty \le 2\|J\|_\infty < \infty$. This shows that $R(G, H)$ in (2.3) is of order $o(\rho_*(G, H))$. □

Proof of Theorem 2.2. (i) From conditions (2.6)–(2.8) and the fact that

$$\left|\int [\psi(x,\theta) - \psi(x,T(G))]\, d(H-G)(x)\right| \le \|\psi(\cdot,T(G)) - \psi(\cdot,\theta)\|_v \rho_\infty(G,H),$$

R_1 in (2.5a) is of the order $o(\rho_\infty(G, H))$. Also from (2.6)–(2.8), R_2 in (2.5b) satisfies

$$|R_2| \le o(1)\rho_\infty(H,G) = o(\rho_\infty(H,G)).$$

The result follows since R_3 in (2.5c) satisfies,

$$\begin{aligned}|R_3| &\le O(1)\left|\int [\psi(x,T(G)) - \psi(x,T(H))]\, d(H-G)(x)\right|\\ &\le O(1)\|\psi(\cdot,T(G)) - \psi(\cdot,T(H))\|_v \rho_\infty(H,G)\\ &= o(\rho_\infty(G,H)).\end{aligned}$$

(ii) Under (2.6)–(2.7),

$$\begin{aligned}|R_2| &\le o(1)\left|\int \psi(x,T(G))\, d(H-G)(x)\right|\\ &= o(1)\left|\int q(x,T(G))(H-G)(x)\, dx\right| = o(\rho_1(G,H)),\end{aligned}$$

since $q(x, T(G))$ is bounded on N_θ and $T(G) \to \theta$ as $\rho_\infty(G, F) \to 0$. From the proof of part (i), R_1 and R_3 are of the order $o(\rho_*(G, H))$ if

$$w_{H,G} = \int [\psi(x,\theta) - \psi(x,T(G))]\, d(H-G)(x) = o(\rho_*(G,H)).$$

Under the conditions in (ii),

$$\begin{aligned}|w_{H,G}| &\le O(1)\rho_\infty(G,H)m(D_{T(G)}) + \rho_1(G,H) \sup_{x \in D^c_{T(G)}} |q(x,T(G)) - q(x,\theta)|\\ &= o(\rho_*(G,H)).\end{aligned}$$

This completes the proof. □

Proof of Theorem 2.3. Suppose that $t_k \to 0$, and $\rho_\infty(G_k, F) \to 0$. Let $H_k = G_k + t_k(\delta_x - G_k)$. Note that

$$\frac{T(H_k) - T(G_k)}{t_k} - L_F(\delta_x - G_k) = R_{1k}(x) + R_{2k}(x) + R_{3k}(x),$$

where

$$R_{1k}(x) = \int_0^\infty J'(\tilde{G}_k(u))(\tilde{\delta}_x - \tilde{G}_k)(u)\, dG_k(u)$$

$$- \int_0^\infty J'(\tilde{F}(u))(\tilde{\delta}_x - \tilde{F})(u)\, dF(u)$$

$$+ J(\tilde{G}_k(x)) - J(\tilde{F}(x)) + T(F) - T(G_k),$$

$$R_{2k}(x) = \int_0^\infty [J(\tilde{H}_k(u)) - J(\tilde{G}_k(u))]\, d(\delta_x - G_k)(u),$$

and

$$R_{3k}(x) = \int_0^\infty \left[\frac{J(\tilde{H}_k(u)) - J(\tilde{G}_k(u))}{t_k} - J'(\tilde{G}_k(u))(\tilde{\delta}_x - \tilde{G}_k)(u) \right] dG_k(u).$$

Since $\rho_\infty(G_k, F) \to 0$ and J and F are continuous, $T(G_k) \to T(F)$ and

$$J(\tilde{G}_k(x)) - J(\tilde{F}(x)) \to 0 \quad \text{uniformly in } x.$$

Since J' is continuous,

$$\left| \int_0^\infty J'(\tilde{G}_k(u))(\tilde{\delta}_x - \tilde{G}_k)(u)\, dG_k(u) - \int_0^\infty J'(\tilde{F}(u))(\tilde{\delta}_x - \tilde{F}_k)(u)\, dG_k(u) \right|$$

$$\le \|J' \circ \tilde{G}_k - J' \circ \tilde{F}\|_\infty + \|J'\|_\infty \rho_\infty(G_k, F) \to 0.$$

Also,

$$\left| \int_0^\infty J'(\tilde{F}(u))(\tilde{\delta}_x - F)(u)\, d(G_k - F)(u) \right|$$

$$\le \rho_\infty(G_k, F) \left\| (J' \circ \tilde{F})(\tilde{\delta}_x - \tilde{F}) \right\|_v$$

$$\le \rho_\infty(G_k, F) \left(\|J'\|_v \rho_\infty(\tilde{\delta}_x, \tilde{F}) + \|J'\|_\infty \|\tilde{\delta} - \tilde{F}\|_v \right)$$

$$\le 2\|J'\|_v \rho_\infty(G_k, F).$$

Hence

$$R_{1k}(x) \to 0 \quad \text{uniformly in } x.$$

From the continuity of J and J',

$$R_{2k}(x) \to 0 \quad \text{and} \quad R_{3k}(x) \to 0 \quad \text{uniformly in } x.$$

The result follows. □

PROOF OF LEMMA 3.1. Let $I_i(x)$ be the indicator function of the set $\{X_i \le x\}$ and $W_i = \int_{-\infty}^0 [I_i(x) - F(x)]\, dx$. Note that

$$E|W_i| \le \int E|I_i(x) - F(x)|\, dx = 2\int F(x)[1 - F(x)]\, dx < \infty.$$

Thus, from the SLLN,

$$\int_{-\infty}^{0}[F_n(x) - F(x)]\,dx = n^{-1}\sum_{i=1}^{n} W_i \to EW_i = 0 \quad \text{a.s.} \tag{4.2}$$

Since $[F_n(x) - F(x)]^- \le F(x)$ and $\int_{-\infty}^{0} F(x)\,dx < \infty$, we have

$$\int_{-\infty}^{0}[F_n(x) - F(x)]^-\,dx \to 0 \quad \text{a.s.},$$

which and (4.2) imply $\int_{-\infty}^{0}|F_n(x) - F(x)|\,dx \to 0$ a.s. Similarly we can show that $\int_0^{\infty}|F_n(x) - F(x)|\,dx \to 0$ a.s. Hence the first assertion follows.

For the second assertion, note that

$$\begin{aligned}(n-1)\sum_{i=1}^{n}\left[\int|F_{ni}(x) - F_n(x)|\,dx\right]^2 &= (n-1)^{-1}\sum_{i=1}^{n}\left[\int|F_n(x) - I_i(x)|\,dx\right]^2\\ &\le 2n(n-1)^{-1}\left[\int|F_n(x) - F(x)|\,dx\right]^2\\ &\quad + 2(n-1)^{-1}\sum_{i=1}^{n}\left[\int|I_i(x) - F(x)|\,dx\right]^2.\end{aligned}$$

Since $E[\int|I_1(x) - F(x)|\,dx]^2 \le E(|X_1| + E|X_1|)^2 < \infty$,

$$n^{-1}\sum_{i=1}^{n}\left[\int|I_i(x) - F(x)|\,dx\right]^2 \to E\left[\int|I_1(x) - F(x)|\,dx\right]^2 \quad \text{a.s.},$$

by the SLLN. The result then follows from $\int|F_n(x) - F(x)|\,dx \to 0$ a.s. □

Acknowledgment. I would like to thank the referees for their helpful comments and suggestions.

REFERENCES

ARVESEN, J. N. (1969). Jackknifing *U*-statistics. *Ann. Math. Statist.* **40** 2076–2100.

BERAN, R. (1984). Jackknife approximations to bootstrap estimates. *Ann. Statist.* **12** 101–118.

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

BOOS, D. D. and SERFLING, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions with applications to *M*-estimators. *Ann. Statist.* **8** 618–624.

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

CLARKE, B. R. (1986). Nonsmooth analysis and Fréchet differentiability of *M*-functionals. *Probab. Theory Related Fields* **73** 197–209.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scand. J. Statist.* **16** 97–128.

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

MILLER, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.

MILLER, R. G. (1968). Jackknifing variances. *Ann. Math. Statist.* **39** 567–582.

NATANSON, I. P. (1961). *Theory of Functions of a Real Variable* **1**, rev. ed. Ungar, New York.

Parr, W. C. (1985). Jackknifing differentiable statistical functionals. *J. Roy. Statist. Soc. Ser. 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.

Reeds, J. A. (1976). On the definition of von Mises functionals. Ph.D. dissertation, Harvard Univ.

Reeds, J. A. (1978). Jackknifing maximum likelihood estimates. *Ann. Statist.* **6** 727–739.

Sen, P. K. (1988). Functional jackknifing: Rationality and general asymptotics. *Ann. Statist.* **16** 450–469.

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

Shao, J. and Wu, C. F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17** 1176–1197.

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

van der Vaart, A. W. (1991). On the differentiable functionals. *Ann. Statist.* **19** 178–204.

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

Department of Mathematics
University of Ottawa
585 King Edward
Ottawa, Ontario
Canada K1N 6N5