Statistical Functional, Differentiability and Jackknife

by

Jun Shao
Purdue University

Technical Report #88-56

Department of Statistics
Purdue University

November, 1988

△

STATISTICAL FUNCTIONAL, DIFFERENTIABILITY AND JACKKNIFE

by

Jun Shao
Purdue University

Abstract

In statistical applications the unknown parameter of interest can frequently be defined as a functional $\theta = T(F)$, where F is the unknown population. Large sample statistical inference about θ usually relies on the degree of smoothness of the functional T. This paper studies the smoothness of T through the differentiability of T. Several versions of differential of T are discussed. Asymptotic properties of the point estimator of θ obtained by evaluating T at the empirical distribution function and consistency of the jackknife variance estimator, which provide procedures of making statistical inferences, are established for differentiable functionals. The results are applied to some examples including the situations where the point estimators are commonly used L- and M-estimators.

Keywords: Statistical inference, influence curve, asymptotic normality, law of iterated logarithm, Fréchet differentiability, weak differentiability, L- and M-estimators.

1. Introduction

Statistical inferences about an unknown parameter θ are usually based on a point estimator $\hat{\theta}$ of θ and the asymptotic behavior of $\hat{\theta}-\theta$. 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 corresponding to i.i.d. samples $X_1, \ldots, X_n$ from F. Often the statistical functional T possesses differentiability properties which provide information about the asymptotic behavior 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; Huber, 1981; Clarke, 1983, 1986). Serfling (1980, Chapter 6) provides an excellent review for this approach.

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

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

where ϕ_F is a real-valued function defined on $\mathbf{R}$ and satisfies $E\phi_F(X_1) = 0$. Note that $n^{-1}\sum_{i=1}^{n}\phi_F(X_i)$ is a linear statistic. If $E\phi_F^2(X_1) = \sigma^2$ is finite and

$$n^{1/2}R(F_n, F) \to_p 0,$$

then the asymptotic distribution of $n^{1/2}(\hat{\theta}-\theta)$ is normal, i.e.,

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

where $\to_p$ and $\to_d$ denote convergence in probability and in distribution, respectively. Since σ^2 is unknown in general, a consistent estimator of σ^2 is required for the purposes of statistical inferences. The jackknife provides a nonparametric method of estimating σ^2 (see Tukey, 1958; Shao and Wu, 1989). Let F_{ni} be the empirical distribution corresponding to the samples $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$. The jackknife estimator of σ^2 is

$$s_J^2 = (n-1)\sum_{i=1}^{n}[T(F_{ni}) - n^{-1}\sum_{i=1}^{n}T(F_{ni})]^2.$$

One advantage of using s_J^2 is that it does not require knowing the form of the function ϕ_F and therefore avoids the theoretical derivation of ϕ_F. It is desired to establish

$$s_J^2 \to \sigma^2 \ a.s. \tag{1.2}$$

Under (1.1) and (1.2), statistical inferences can be made based on the result

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

It was shown (e.g., Serfling, 1980) that (1.1) holds if T is $\| \|_\infty$-Fréchet differentiable (see Definition 2.1(i)). Parr (1985) proved (1.2) for uniformly $\| \|_\infty$-Fréchet differentiable T (see Definition 2.1(ii)). However, $\| \|_\infty$-Fréchet differentiability is too strong a requirement since some frequently used statistical functionals are not $\| \|_\infty$-Fréchet differentiable. The purpose of this paper is to study weaker versions of differential of T and to establish (1.1)-(1.2) using a unified approach. The results obtained are more general than the existing results in the literature.

In Section 2, several versions of differential of T are studied. The ρ^*-weak differentiability of T (see Definition 2.2) seems to be a reasonably weak requirement, since commonly encountered statistical functionals, including those corresponding to L- and M- estimators, are shown to be ρ^*-weakly differentiable. The results are applied to establishing the asymptotic normality of the point estimators $T(F_n)$ (e.g., L- and M-estimators and Cramér-von Mises test statistic). The existence of a differential of T asserts more, for

(a) It provides the influence curve, a measure of "influence" toward the estimation error $T(F_n)-T(F)$ (Hampel, 1974). The influence curve is a useful tool of robust statistics and sensitivity analysis (see Huber, 1981; Serfling, 1980, Section 6.6).

(b) The differential approach establishes the law of the iterated logarithm (LIL), i.e.,

$$\limsup_{n\to\infty} \frac{n^{1/2}[T(F_n)-T(F)]}{\sigma(2\log\log n)^{1/2}} = 1 \quad a.s., \tag{1.3}$$

which characterizes the extreme fluctuations of $T(F_n)-T(F)$.

(c) It provides a powerful tool for studying the asymptotic behavior of the jackknife variance estimator s_J^2.

The consistency of the jackknife variance estimators is studied in Section 3.

2. Differentiability and asymptotic properties of $T(F_n)$

Let T be a functional defined on a metric space $(\mathbf{F}, \rho)$, where $\mathbf{F}$ is a convex set of distribution functions containing the unknown population F and all degenerate distribution functions and ρ is a metric on $\mathbf{F}$. Note that $F_n \in \mathbf{F}$.

A functional T is said to be ρ-continuous at F if $T(G) \to T(F)$ as $\rho(G,F) \to 0$ ($G \in \mathbf{F}$). For the differentiability of T, we first consider the Fréchet differentiability.

Definition 2.1. (i) A functional T on $(\mathbf{F}, \rho)$ is ρ-Fréchet differentiable at F if there is a real-valued function ϕ_F on $\mathbf{R}$ such that $\int \phi_F(x) dF(x) = 0$ and

$$\frac{|T(G) - T(F) - \int \phi_F(x) d[G(x) - F(x)]|}{\rho(G, F)} \to 0 \qquad (2.1)$$

as $\rho(G, F) \to 0$, $G \in \mathbf{F}$.

(ii) A functional T on $(\mathbf{F}, \rho)$ is uniformly ρ-Fréchet differentiable at F if

$$\frac{|T(G) - T(H) - \int \phi_F(x) d[G(x) - H(x)]|}{\rho(G, H)} \to 0 \qquad (2.2)$$

as $\rho(G, F) + \rho(H, F) \to 0$, $G, H \in \mathbf{F}$.

The function ϕ_F is the influence curve of T (Hampel, 1974). Throughout the paper we assume that $0 < \sigma^2 = \int \phi_F^2(x) dF(x) < \infty$. $\int \phi_F(x) d[G(x) - F(x)]$ is the differential of T at F and is a linear functional presenting the linear component of $T(G) - T(F)$. Let

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

Then $R(G, F)$ is the nonlinear component of $T(G) - T(F)$ and (2.1) becomes $|R(G, F)| / \rho(G, F) \to 0$. Note that a ρ-Fréchet differentiable T is not necessarily ρ-continuous. Huber (1981) showed that for a ρ-Fréchet differentiable T, (1.1) holds if $\rho(F_n, F) = O_p(n^{-1/2})$. It is also true that if $\rho(F_n, F) = O(n^{-1/2}(\log\log n)^{1/2})$ *a.s.*, then (1.3) holds for ρ-Fréchet differentiable T (Serfling, 1980, p.219).

The most commonly used metric on $\mathbf{F}$ is $\rho(G, F) = \|G - F\|_\infty$, where $\| \ \|_\infty$ is the sup norm ($\|h(x)\|_\infty = \sup_x |h(x)|$ for any bounded function h). Hence T is $\| \ \|_\infty$-Fréchet differentiable at F if (2.1) holds with the metric corresponding to $\| \ \|_\infty$. Since $\|F_n - F\|_\infty = O_p(n^{-1/2})$ and $\|F_n - F\|_\infty = O(n^{-1/2}(\log\log n)^{1/2})$ *a.s.*, $\| \ \|_\infty$-Fréchet differentiability of T at F implies (1.1) and (1.3).

However, there are some commonly used statistical functionals which are not $\| \ \|_\infty$-Fréchet differentiable. For example, the variance functional

$$T(F) = 2^{-1}\iint (x-y)^2 \mathrm{d}F(x)\mathrm{d}F(y) \tag{2.3}$$

is not $\| \ \|_\infty$-Fréchet differentiable (Serfling, 1980, p.220). Hence we need to seek weaker versions of differential of T.

Consider a subset of $\mathbf{F}$: $\mathbf{F}_1 = \{ G \in \mathbf{F} : \int |x| \mathrm{d}G(x) < \infty \}$, and the L_1 norm on $\mathbf{F}_1$ defined to be $\|G-F\|_1 = \int |G(x)-F(x)| \mathrm{d}x$. Then the variance functional (2.3) is uniformly $\| \ \|_1$-Fréchet differentiable, since for the variance functional, $R(G, H) = -(\int [G(x)-H(x)]\mathrm{d}x)^2$ satisfies $|R(G, H)|/\|G-H\|_1 \le \|G-H\|_1 \to 0$ as $\|G-F\|_1 + \|H-F\|_1 \to 0$.

Note that if ρ_1 and ρ_2 are two metrics on $\mathbf{F}_1$ satisfying $\rho_2(G, H) \le c\rho_1(G, H)$ for a constant c and all $G, H \in \mathbf{F}_1$, then (uniform) ρ_2-Fréchet differentiability implies (uniform) ρ_1-Fréchet differentiability. This suggests the use of the metric

$$\rho^*(G, H) = \|G-H\|_1 + \|G-H\|_\infty, \quad G, H \in \mathbf{F}_1.$$

Note that $E[n^{1/2}\|F_n - F\|_1] \le \int [F(x)(1-F(x))]^{1/2} \mathrm{d}x$. Hence $\rho^*(F_n, F) = O_p(n^{-1/2})$ if

$$\int [F(x)(1-F(x))]^{1/2} \mathrm{d}x < \infty. \tag{2.4}$$

Thus, (1.1) holds if T is ρ^*-Fréchet differentiable and (2.4) holds. The class of functionals differentiable in the metric ρ^* is substantially larger than the class of functionals differentiable in the metric $\| \ \|_\infty$. However, for the asymptotic normality, we need to assume a moment condition (2.4), which is almost the same as $EX_1^2 < \infty$ (Serfling, 1980, p.276).

Example 2.1. L-estimators. Consider the functional

$$T(F) = \int_0^1 F^{-1}(t)J(t)\mathrm{d}t, \tag{2.5}$$

where $F^{-1}(t) = \inf\{ x : F(x) \ge t \}$ and $J(t)$ 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), for $G, H \in \mathbf{F}_1$,

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

with

$$
\begin{aligned}
\phi_F(x) &= -\int [I(y \geq x) - F(y)] J[F(y)] \mathrm{d}y \\
R(G, H) &= \int W[G(x), H(x)][H(x) - G(x)] \mathrm{d}x,
\end{aligned}
\tag{2.6}
$$

where $I(A)$ is the indicator function of the set A, $W[G(x),H(x)] = 0$ if $G(x)=H(x)$ and $= [G(x)-H(x)]^{-1}\int_{H(x)}^{G(x)} J(t)\mathrm{d}t - J[F(x)]$ 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 uniformly $\| \ \|_\infty$-Fréchet differentiable. It is not clear whether T is still $\| \ \|_\infty$-Fréchet differentiable if J is untrimmed. The following result (the proof is in Section 4) shows that under weaker conditions on J, T is uniformly ρ^*-Fréchet differentiable. Note that this result holds, in particular, if J is continuous on [0,1]. Examples of continuous J include (i) $J(t) \equiv 1$ (the sample mean); (ii) $J(t) = 4t-2$ (Gini's mean difference); (iii) $J(t) = 6t(1-t)$ (the asymptotically efficient L-estimator for location for the logistic family). Boos (1979) proved that T is q-norm-Fréchet differentiable at F. Compared with his Theorem 2, the following theorem requires slightly more on J but less on the tails of F and provides a stronger result (the uniform Fréchet differentiability). Also, the q-norm is not easy to handle for establishing the consistency of the jackknife variance estimator (see Section 3).

Theorem 2.1. Let T be defined as in (2.5) and $F \in \mathbf{F}_1$. 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 uniformly ρ^*-Fréchet differentiable.

However, ρ^*-Fréchet differentiability is still somewhat too strong. The requirement (2.1) or (2.2) can be relaxed. In fact, for statistical analysis, we only need (2.1) or (2.2) holds for those sequences of G which satisfy $T(G) \to T(F)$ as $\rho(G, F) \to 0$. This leads to the following weaker notion of differential of T.

Definition 2.2. (i) A functional T on $(\mathbf{F}, \rho)$ is ρ-weakly differentiable at F if (2.1) holds as $\rho(G, F) + |T(G) - T(F)| \to 0$.

(ii) A functional T on $(\mathbf{F}, \rho)$ is uniformly ρ-weakly differentiable at F if (2.2) holds as $\rho(G, F) + \rho(H, F) + |T(G) - T(F)| + |T(H) - T(F)| \to 0$.

Apparently, (uniform) ρ-Fréchet differentiability implies (uniform) ρ-weak differentiability and they are equivalent if T is ρ-continuous. Note that many commonly used statistical functionals are not ρ^*-continuous (e.g., the variance functional (2.3)) but satisfy $T(F_n) \to T(F)$. Hence the ρ^*-weak differentiability is a much weaker requirement than the ρ^*-Fréchet differentiability and provides a useful tool for statistical analysis.

The asymptotic normality and LIL for a ρ-weakly differentiable T can be established if $\rho(F_n, F)$ has certain stochastic order. We show the following result for ρ^*-weakly differentiable functional.

Theorem 2.2. Assume that T is ρ^*-weakly differentiable at F.

(i) Asymptotic Normality. If (2.4) holds and $T(F_n) \to_p T(F)$, then (1.1) holds.

(ii) LIL. If there is a constant $\delta>0$ such that

$$\int [F(x)(1-F(x))]^{1/2-\delta} dx < \infty \tag{2.7}$$

and $T(F_n) \to T(F)$ *a.s.*, then (1.3) holds.

Remarks. (i) Condition (2.7) is equivalent to that $E\,|X_1|^{2+\varepsilon} < \infty$ for an $\varepsilon>0$.

(ii) If T is $\|\ \|_\infty$-weakly differentiable at F, conditions (2.4) in (i) and (2.7) in (ii) are not required.

Proof. (i) It suffices to show that $R(F_n, F) = o_p(n^{-1/2})$. From the differentiability of T, for any $\tau>0$ and $\varepsilon>0$, there is a $\delta_\varepsilon>0$ such that

$$P\{\, n^{1/2}|R(F_n, F)|>\tau \,\} \le P\{\, n^{1/2}\rho^*(F_n, F)>\tau/\varepsilon \,\} + P\{\, \rho^*(F_n, F)+|T(F_n)-T(F)|>\delta_\varepsilon \,\}.$$

The result follows from $\rho^*(F_n, F) = O_p(n^{-1/2})$ and $\rho^*(F_n, F)+|T(F_n)-T(F)| \to_p 0$.

(ii) It suffices to show $\int |F_n(x)-F(x)|dx = O(n^{-1/2}(\log\log n)^{1/2})$ *a.s.* From the result in James (1975), $\| [F_n(x)-F(x)]w[F(x)] \|_\infty = O(n^{-1/2}(\log\log n)^{1/2})$ *a.s.*, where $w(t)=[t(1-t)]^{\delta-1/2}$. Then the result follows from $\int [w(F(x))]^{-1}dx < \infty$. □

In statistical applications, often the parameter of interest is $g[T(F)]$ with a real-valued function g. Hence we need to consider the functional $g \circ T$.

Proposition 2.1. Suppose that T is ρ-weakly differentiable at F and g is a function on **R** and differentiable at $T(F)$. Then $g \circ T$ is ρ-weakly differentiable at F if the following regularity condition holds:

$$T(G) - T(F) = O[\rho(G, F)] \quad \text{as} \quad |T(G) - T(F)| \to 0. \tag{2.8}$$

Proof. If T is ρ-weakly differentiable at F, $T(G) - T(F) = \int \phi_F \mathrm{d}G + R(G, F)$ with $R(G, F)/\rho(G, F) \to 0$ as $\rho(G, F) + |T(G) - T(F)| \to 0$. From the differentiability of g,

$$g[T(G)] - g[T(F)] = g'[T(F)] \int \phi_F \mathrm{d}G + g'[T(F)] R(G, F) + o(|T(G) - T(F)|).$$

From (2.8),

$$\frac{o(|T(G)-T(F)|)}{\rho(G, F)} = \frac{o(|T(G)-T(F)|)}{|T(G)-T(F)|} \frac{|T(G)-T(F)|}{\rho(G, F)} \to 0$$

as $|T(G) - T(F)| \to 0$. Hence $g \circ T$ is ρ-weakly differentiable at F. $\square$

Remarks. (i) Condition (2.8) is weaker than $T(G) - T(F) = O[\rho(G, F)]$ as $\rho(G, F) \to 0$. In fact, (2.8) does not imply ρ-continuity of T at F. See Example 2.3.

(ii) If T is ρ-weakly differentiable at F with differential $\int \phi_F \mathrm{d}G$, (2.8) is equivalent to

$$\int \phi_F \mathrm{d}(G - F) = O[\rho(G, F)] \quad \text{as} \quad |T(G) - T(F)| \to 0. \tag{2.9}$$

We now study some examples of application of Theorem 2.2. First, consider the L-estimator in Example 2.1.

Corollary 2.1. Let T be an L-functional and g be a real-valued function and differentiable at $t_0 = T(F)$. Under the conditions in Theorem 2.1 and (2.4), we have

$$n^{1/2}\{g[T(F_n)] - g[T(F)]\} \to_d N(0, [g'(t_0)]^2 \sigma^2), \tag{2.10}$$

where $\sigma^2 = E\phi_F^2(X_1)$ and ϕ_F is given in (2.6). If in addition, (2.7) holds and $g'(t_0) \neq 0$, then

$$\limsup_{n \to \infty} \frac{n^{1/2}[T(F_n) - T(F)]}{|g'(t_0)| \sigma (2 \log\log n)^{1/2}} = 1 \quad a.s. \tag{2.11}$$

Proof. From Theorem 2.1, T is ρ^*-weakly differentiable at F. From (2.6),

$$\left|\int \phi_F \mathrm{d}(G - F)\right| = \left|\int [G(x) - F(x)] J[F(x)] \mathrm{d}x\right| \le \|J\|_\infty \|G - F\|_1,$$

which implies (2.9) and the ρ^*-continuity of T. Hence $T(F_n) \to T(F)$ *a.s.* and the result follows from Proposition 2.1 and Theorem 2.2. □

Example 2.2. V-statistics. Let T be the functional

$$T(F) = \int \cdots \int h(x_1,...,x_k) dF(x_1) \cdots dF(x_k)$$

with a symmetric kernel $h(x_1,...,x_k)$. A V-statistic is $T(F_n)$. For this functional, $T(G) - T(F)$ can be written as (e.g., Serfling, 1980, Chapter 6)

$$\sum_{v=1}^{k} \frac{1}{v!} D_v(G, F)$$

with

$$D_v(G, F) = k(k-1) \cdots (k-v+1) \int \cdots \int h_v(x_1,...,x_v) \prod_{j=1}^{v} d[G(x_j)-F(x_j)],$$

where

$$h_v(x_1,...,x_v) = \int \cdots \int h(x_1,...,x_v, x_{v+1},...,x_k) \prod_{j=v+1}^{k} dF(x_j).$$

Then $\int \phi_F(x) d[G(x)-F(x)] = D_1(G, F)$. Suppose that $q(x_1,x_2) = \partial^2 h_2(x_1,x_2)/\partial x_1 \partial x_2$ exists. Under suitable conditions,

$$D_2(G, F) = \iint [G(x_1)-F(x_1)][G(x_2)-F(x_2)] q(x_1,x_2) dx_1 dx_2.$$

Then $|D_2(G, F)| \le [\rho^*(G, F)]^2$ if $q(x_1,x_2)$ is either bounded or integrable. The other terms $D_v(G, F)$, $v=3,...,k$, can be handled similarly. Hence V-statistics are ρ^*-Fréchet differentiable at F under certain conditions.

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

$$\int r(x, T(F)) dF(x) = \min_t \int r(x, t) dF(x), \tag{2.12}$$

where $r(x, t)$ is a real-valued function on $\mathbf{R}^2$. Examples of M-functionals can be found in Serfling (1980, Chapter 7). Let $t_0 = T(F)$. $T(F_n)$ is called the M-estimator of t_0. 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_F is differentiable at t_0 with $\lambda_F'(t_0) \neq 0$. Define $h_F(s) = [\lambda_F(s) - \lambda_F(t_0)]/(s-t_0)$ if $s \neq t_0$ and $= \lambda_F'(t_0)$ if $s = t_0$. Then

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

with $\phi_F(x) = -\psi(x, t_0)/\lambda_F'(t_0)$ and

$$R(G, H) = [\lambda_F(T(G)) - \lambda_F(T(H))]/h_F[T(G)] + [\lambda_G(t_0) - \lambda_H(t_0)]/\lambda_F'(t_0). \tag{2.13}$$

The following result shows that T is ρ^*-differentiable under weak conditions. Its proof is given in Section 4.

Theorem 2.3. Let T be an M-functional defined in (2.12).

(i) Assume that $\|\psi(\cdot, t_0)\|_V < \infty$ and $\|\psi(\cdot, t) - \psi(\cdot, t_0)\|_V \to 0$ as $t \to t_0$, where $\| \ \|_V$ is the total variation norm (see Natanson, 1961). Then T is $\| \ \|_\infty$-weakly differentiable at F.

(ii) Assume that there is a neighborhood $\mathbf{N}_{t_0}$ of t_0 such that for $t \in \mathbf{N}_{t_0}$, $\|\psi(\cdot, t)\|_V < \infty$ and $\eta(x, t) = \partial\psi(x, t)/\partial t$ is bounded and continuous at t_0. Assume further that either $\eta(x, t_0)$ is continuous in x or $\|\eta(\cdot, t_0)\|_V < \infty$. Then T is $\| \ \|_\infty$-weakly differentiable at F.

(iii) Assume that there is a neighborhood $\mathbf{N}_{t_0}$ of t_0 such that for $t \in \mathbf{N}_{t_0}$, $q(x, t) = \partial\psi(x, t)/\partial x$ is bounded. Assume further that for each $t \in \mathbf{N}_{t_0}$ there is a set D_t such that as $t \to t_0$, $m(D_t) \to 0$ and $\sup_{x \in D_t} |q(x, t) - q(x, t_0)| \to 0$ (m is the Lebesgue measure). Then T is ρ^*-weakly differentiable at F.

Under more conditions (e.g., T is continuous), one can show that T is Fréchet differentiable. Note that an M-functional may not be continuous. However, the consistency of $T(F_n)$ can be established under weaker conditions (see Proposition 3.2 in Section 3).

Corollary 2.3. Let T be an M-functional and g be a real-valued function and differentiable at $t_0 = T(F)$.

(i) Assume the condition in Theorem 2.3(i) or (ii). If $T(F_n) \to_p t_0$, then (2.10) holds with $\sigma^2 = \int \psi^2(x, t_0) dF(x)/[\lambda_F'(t_0)]^2$. If $T(F_n) \to t_0$ *a.s.* and $g'(t_0) \neq 0$, then (2.11) holds.

(ii) The same conclusions as in (i) hold if we assume the condition in Theorem 2.3(iii) and condition (2.4) (for asymptotic normality) or (2.7) (for LIL).

Proof. From Theorems 2.2 and 2.3 and Proposition 2.1, it suffices to show (2.9). Since the differential of T is $-\lambda_G(t_0)/\lambda_F'(t_0)$, (2.9) is implied by the condition in any of Theorem 2.3(i)-

(iii) (see the proof of Theorem 2.3). □

Example 2.4. Cramér-von Mises test statistic. Let F_0 be a specified hypothetical distribution. Define a functional

$$T(G) = \int [G(x) - F_0(x)]^2 dF_0(x). \tag{2.14}$$

$T(F_n)$ is the Cramér-von Mises test statistic for the test problem: H_0: $F = F_0$ vs H_1: $F \neq F_0$. Consider the asymptotic distribution of $T(F_n)$ under the alternative hypothesis H_1.

Theorem 2.4. The functional defined by (2.14) is uniformly $\| \ \|_\infty$-Fréchet differentiable at F with the influence function

$$\phi_F(x) = 2\int [I(x \le y) - F(y)][F(y) - F_0(y)] dF_0(y).$$

If F_0 is continuous, then T is $\| \ \|_\infty$-continuous.

The proof is in Section 4. A direct consequence of this result is the following.

Corollary 2.4. Under the alternative hypothesis $F \neq F_0$, the Cramér-von Mises test statistic satisfies (1.1) and (1.3) with

$$\sigma^2 = 4\int\int [F(\min(y,z)) - F(y)F(z)][F(y) - F_0(y)][F(z) - F_0(z)] dF_0(y) dF_0(z). \tag{2.15}$$

Note that under the null hypothesis, $F = F_0$, $T(F) = 0$ and $\phi_F(x) \equiv 0$ and therefore $n^{1/2} T(F_n) \to_p 0$. In fact, under the null hypothesis $nT(F_n) \to_d$ a weighted sum of chi-squared variates (e.g., Serfling, 1980, Theorem 2.1.7B).

3. The jackknife

We now establish the consistency of the jackknife estimator s_J^2 defined in Section 1, which is desired for making statistical inference. Let ρ be a metric on $\mathbf{F}$ or $\mathbf{F}_1$.

Theorem 3.1. Assume that ρ satisfies

$$\rho(F_n, F) \to 0 \ \ a.s. \quad \text{and} \quad (n-1)\sum_{i=1}^{n}[\rho(F_{ni}, F_n)]^2 = O(1) \ \ a.s. \tag{3.1}$$

(i) If T is uniformly ρ-Fréchet differentiable at F (with $\sigma^2=\int\phi_F^2 dF$), then

$$s_J^2 \to \sigma^2 \ \ a.s.$$

(ii) The result in (i) holds if T is uniformly ρ-weakly differentiable at F and

$$\max_{i\le n} |T(F_{ni})-T(F)| \to 0 \ \ a.s. \tag{3.2}$$

Proof. Let $Z_i=\phi_F(X_i)$ and $\bar{Z}=n^{-1}\sum_{i=1}^{n}Z_i$. From the differentiability of T,

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

where $R_{ni}=R(F_{ni}, F_n)$. Let $\bar{R} = n^{-1}\sum_{i=1}^{n}R_{ni}$. Then

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

From the strong law of large numbers (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 (3.1),

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

If T is uniformly ρ-Fréchet differentiable at F, 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 (3.1). This proves (i). The proof for (ii) is similar by using (3.2). □

If ρ is the metric corresponding to $\|\ \|_\infty$, then (3.1) is satisfied since $\|F_{ni}-F_n\|_\infty \le n^{-1}$ for all i. For the metric ρ^*, the following result shows that (3.1) holds (and therefore the results in Theorem 3.1 hold) if the second moment of F exists. The proof of this result is in Section 4.

Proposition 3.1. Assume that $EX_1^2 < \infty$. Then

$$\int|F_n(x)-F(x)|dx \to 0 \ \ a.s. \quad \text{and} \quad (n-1)\sum_{i=1}^{n}[\int|F_{ni}(x)-F_n(x)|dx]^2 = O(1) \ \ a.s.$$

In some cases we need to consider a function of several functionals: $g\circ T$, where T is a k-vector whose jth component is a functional $T_j(F)$ on $\mathbf{F}$ and g is a real-valued function on

$\mathbf{R}^k$. Let ∇g be the gradient of g. We have the following result.

Theorem 3.2. Let ρ be a metric satisfying (3.1). Assume that T_j is uniformly ρ-weakly differentiable at F with influence function ϕ_j, $j=1,...,k$, ∇g is continuous at $t_0=T(F)$ and (3.2) holds for each T_j. Then the jackknife estimator

$$s_{gJ}^2 = (n-1)\sum_{i=1}^n \{g[T(F_{ni})]-n^{-1}\sum_{i=1}^n g[T(F_{ni})]\}^2 \to [\nabla g(t_0)]^\tau V[\nabla g(t_0)] \quad a.s.,$$

where V is a $k\times k$ matrix whose (p,q)th element is $E[\phi_p(X_1)\phi_q(X_1)]$.

Proof. Let $t_n=n^{-1}\sum_{i=1}^n T(F_{ni})$. From the mean-value theorem,

$$\begin{aligned} s_{gJ}^2 &= (n-1)[\nabla g(t_n)]^\tau \sum_{i=1}^n [T(F_{ni})-t_n][T(F_{ni})-t_n]^\tau[\nabla g(t_n)] \\ &+ (n-1)\sum_{i=1}^n (U_{ni}-n^{-1}\sum_{i=1}^n U_{ni})^2 + 2(n-1)[\nabla g(t_n)]^\tau \sum_{i=1}^n [T(F_{ni})-t_n](U_{ni}-n^{-1}\sum_{i=1}^n U_{ni}), \end{aligned} \tag{3.3}$$

where $U_{ni} = [T(F_{ni})-t_n]^\tau[\nabla g(\xi_i)-\nabla g(t_n)]$ and ξ_i is a point on the line segment between t_n and $T(F_{ni})$. From the proof of Theorem 3.1 and the continuity of ∇g at t_0, the first term on the right hand side of (3.3) $\to [\nabla g(t_0)]^\tau V[\nabla g(t_0)]$ *a.s.* It remains to show that the second term on the right hand side of (3.3) $\to 0$ *a.s.* From (3.2) and the continuity of ∇g at t_0, for any $\varepsilon>0$ and almost all $X_1, X_2,...$, there is an N such that $\max_{i\le n} \|\nabla g(\xi_i)-\nabla g(t_n)\| < \varepsilon$, where $\| \ \|$ is the Euclidean norm. Hence

$$(n-1)\sum_{i=1}^n U_{ni}^2 \le \varepsilon^2(n-1)\sum_{i=1}^n \|T(F_{ni}) - t_n\|^2.$$

This completes the proof. □

Example 3.1. Functions of central moments. Let m_l be positive integers, $l=1,...,L$, $k=\max_l m_l$. Assume that $E|X_1|^k<\infty$. Let $S_l(G)=\int[x-\int x\,dG(x)]^{m_l}dG(x)$ (when $m_l=1$, $S_l(G)$ is defined to be $\int x\,dG(x)$). Consider the functional $T=h(S_1,...,S_L)$ with a continuously differentiable function h. Note that for each l, S_l is a continuously differentiable function of functionals $T_1,...,T_{m_l}$, where $T_p(G)=\int x^p dG(x)$. Hence $T=g(T_1,...,T_k)$ with a continuously differentiable g. For each p, T_p is $\| \ \|_\infty$- Fréchet differentiable at F with the influence function $\phi_p(x)=x^p-\int x^p dF(x)$. In this case, (3.2) can be established by using Lemma 3 of Ghosh et al. (1984). Hence Theorem 3.2 applies. Special cases of this result include (1) functions of mean ($L=1$, $m_l=1$); and (2) functions of variance ($L=1$, $m_1=2$), which were treated by Miller (1964, 68) who established weak consistency of the jackknife estimators.

Example 3.2. (Continuation of Example 2.4). Under the alternative hypothesis, the Cramér-von Mises test statistic $T(F_n)$ is asymptotically normal with asymptotic variance σ^2/n, where σ^2 is given in (2.15). Apply Theorems 2.4 and 3.1, the jackknife estimator s_J^2 is consistent for σ^2. In this case, another estimator of σ^2 can be obtained by replacing F in (2.15) by F_n. It is not hard to see that this estimator also satisfies (1.2). The jackknife method gives an alternative in this case.

For the L-estimators, a direct application of Theorems 2.1 and 3.1 and Proposition 3.1 gives the following result. Note that the conditions we assumed are weaker than those in Parr and Schucany (1982, Theorem 2). These conditions are necessary for the consistency of s_J^2 when $J \equiv 1$ ($T(F_n)$ is the sample mean).

Corollary 3.1. Let T be an L-functional satisfying the conditions in Theorem 2.1. Assume that $EX_1^2 < \infty$. Then $s_J^2 \to \sigma^2$ *a.s.*, where σ^2 is given in (2.10).

The uniform ρ^*-Fréchet differentiability of T plays an important role for the consistency of s_J^2. It is natural that the consistency of s_J^2 requires more smoothness condition on T than the asymptotic normality of $T(F_n)$, since variance estimation is a "second order" operation. For the M-estimators, it was shown in Section 2 that the M-functionals are ρ^*-weakly differentiable. Unfortunately, it is not clear whether the M-functionals are uniformly ρ^*-weakly differentiable. We establish the consistency of s_J^2 by using the following results.

Theorem 3.3. Assume that ρ satisfies (3.1) and that T satisfies (3.2) and is ρ-weakly differentiable at F and F_n for all n. If

$$\max_{i \le n} \frac{|T(F_{ni}) - T(F_n) - \int \phi_{F_n}(x) dF_{ni}(x)|}{\rho(F_{ni}, F_n)} \to 0 \quad a.s. \tag{3.4}$$

and

$$\| \phi_{F_n} - \phi_F \|_V \to 0 \quad a.s., \tag{3.5}$$

then $s_J^2 \to \sigma^2 = \int \phi_F^2(x) dF(x)$ *a.s.*

Proof. Let $W(F_{ni}, F_n) = T(F_{ni}) - T(F_n) - \int \phi_{F_n}(x) \mathrm{d} F_{ni}(x)$. Then

$$R_{ni} = \int [\phi_{F_n}(x) - \phi_F(x)] \mathrm{d} [F_{ni}(x) - F_n(x)] + W(F_{ni}, F_n).$$

From (3.1) and (3.4), $(n-1)\sum_{i=1}^n W^2(F_{ni}, F_n) \to 0$ *a.s.* Since

$$\left| \int (\phi_{F_n} - \phi_F) \mathrm{d}(F_{ni} - F_n) \right| \le \| F_{ni} - F_n \|_\infty \| \phi_{F_n} - \phi_F \|_V,$$

we have $(n-1)\sum_{i=1}^n R_{ni}^2 \to 0$ *a.s.* and thus the result. □

Proposition 3.2. Let T be an M-functional given in Example 2.3.

(i) If ψ is nondecreasing in t and there is a neighborhood $\mathbf{N}_{t_0}$ of $t_0 = T(F)$ such that for each fixed x, $\psi(x, t)$ is continuous on $\mathbf{N}_{t_0}$, $|\psi(x, t)| \le M(x)$ for $t \in \mathbf{N}_{t_0}$ and $\int M(x) \mathrm{d} F(x) < \infty$, then (3.2) holds.

(ii) Assume that (2.12) has a unique solution. Then (3.2) holds if $|r(x, t)| \le M(x)$ for all t with $\int M(x) \mathrm{d} F(x) < \infty$ and for any $a > 0$,

$$\lim_{t \to \infty} r(x, t) = \alpha \quad \text{uniformly for } |x| \le a, \tag{3.6}$$

where $\alpha > \int r(x, t_0) \mathrm{d} F(x)$ (α can be infinity). If in addition, r is bounded (or ψ is bounded and continuous in t and $\lambda_F(t)$ has a unique root), then T is $\| \; \|_\infty$-continuous.

This result implies the strong consistency of $T(F_n)$. Examples of functions satisfying (3.6) include Huber's ($r(x, t) = (x-t)^2$ if $|x-t| \le K$ and $= K^2$ if $|x-t| > K$) and those given in Examples 7.1.2F and 7.1.2G in Serfling (1980). Note that the conditions in Proposition 3.2 do not imply that T is ρ^*-continuous. We now establish the consistency of s_J^2 for M-estimators.

Theorem 3.4. Let T be an M-functional satisfying the conditions in either (i) or (ii) of Proposition 3.2. Assume further that

(a) there is a neighborhood $\mathbf{N}_{t_0}$ of $t_0 = T(F)$ such that $\eta(x, t) = \partial \psi(x, t)/\partial t$ exists and is continuous on $\mathbf{N}_{t_0}$ and $|\eta(x, t)| \le M_1(x)$, where M_1 satisfies $\int M_1(x) \mathrm{d} F(x) < \infty$

and either

(b) the conditions in Theorem 2.3(i) or (ii) or

(b′) $EX_1^2 < \infty$ and the conditions in Theorem 2.3(iii).

Then $s_J^2 \to \sigma^2$ *a.s.*, where σ^2 is given in Corollary 2.3.

The proofs of Proposition 3.2 and Theorem 3.4 are in Section 4. Reeds (1978) proved the consistency of s_J^2 for M-estimators under his condition L. The conditions in our result and Reeds' condition L are not comparable.

4. Proofs

Proof of Theorem 2.1. Let $A = \{\, x : F(x) \le c \,\}$ and $B = \{\, x : c \le F(x) \le 1-c \,\}$ with $c = \alpha/2$. If $F(x) \in A$ and $\|G-F\|_\infty + \|H-F\|_\infty < \delta$, $G(x), H(x) \in [0, c+\delta]$. Let $\delta < \alpha/2$. Since $J(t)$ is uniformly continuous on $[0, c+\delta]$,

$$\Big|\int_A W[G(x), H(x)][G(x)-H(x)]dx\Big| / \rho^*(G,H) \le \sup_{x\in A} |W[G(x), H(x)]| \to 0 \quad (4.1)$$

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

$$\Big|\int_B W[G(x), H(x)][G(x)-H(x)]dx\Big| / \rho^*(G,H) \le \int_a^b |W[G(x), H(x)]|\, dx \to 0$$

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 completes the proof. □

Proof of Theorem 2.3. (i) From (2.13), $R(G,F) = R_1(G,F) + R_2(G,F)$, where

$$R_1(G,F) = \int[\psi(x, T(G)) - \psi(x, t_0)]d[G(x)-F(x)]/h_F[T(G)]$$

and

$$R_2(G,F) = \lambda_G(t_0)\{1/\lambda_F'(t_0) - 1/h_F[T(G)]\}.$$

Note that $h_F[T(G)] \to \lambda_F'(t_0)$ as $T(G) \to T(F)$ and

$$\Big|\int[\psi(x, T(G)) - \psi(x, t_0)]d[G(x)-F(x)]\Big| \le \|\psi(\,\cdot\,, T(G)) - \psi(\,\cdot\,, t_0)\|_V \|G-F\|_\infty.$$

Hence $R_1(G,F)/\|G-F\|_\infty \to 0$ as $\|G-F\|_\infty + |T(G)-T(F)| \to 0$. This is also true with R_1 replaced by R_2 since

$$|\lambda_G(t_0)| \le \|\psi(\,\cdot\,, t_0)\|_V \|G-F\|_\infty.$$

(ii) Assume $T(G) \to t_0$. From the proof of part (i), $T(G)-t_0 = O(\|G-F\|_\infty)$. Using this fact and the continuity of $\eta(x, t)$ at t_0, we have

$$\int [\psi(x, T(G)) - \psi(x, t_0)] \mathrm{d}[G(x)-F(x)] = [T(G)-t_0] \int \eta(x, t_0) \mathrm{d}[G(x)-F(x)] + o(\|G-F\|_\infty).$$

The rest of the proof is similar to that of part (i).

(iii) Assume that $T(G) \to T(F)$. Let $t=T(G)$. Under the conditions in part (ii), there is a constant c such that

$$|\lambda_G(t_0)| \le c \|G-F\|_1$$

and

$$|\int [\psi(x, T(G)) - \psi(x, t_0)] \mathrm{d}[G(x)-F(x)]| \le c \|G-F\|_\infty m(D_t)$$
$$+ \|G-F\|_1 \sup_{x \in D_t} |q(x, T(G)) - q(x, t_0)|.$$

Hence the result follows. □

Proof of Theorem 2.4. For $G, H \in \mathbf{F}$,

$$T(G)-T(H)-\int \phi_F(x) \mathrm{d}[G(x)-H(x)] = \int \{[G(y)-F_0(y)]^2 - [H(y)-F_0(y)]^2\} \mathrm{d}F_0(y)$$
$$-2\int\int [I(x \le y)-F(y)][F(y)-F_0(y)] \mathrm{d}[G(x)-H(x)] \mathrm{d}F_0(y)$$
$$= \int [G(y)-H(y)][G(y)+H(y)-2F_0(y)] \mathrm{d}F_0(y) - 2\int [G(y)-H(y)][F(y)-F_0(y)] \mathrm{d}F_0(y)$$
$$= \int [G(y)-H(y)][G(y)+H(y)-2F(y)] \mathrm{d}F_0(y).$$

Thus,

$$|R(G, H)| \le \|G-H\|_\infty (\|G-F\|_\infty + \|H-F\|_\infty)$$

and T is uniformly $\|\ \|_\infty$-Fréchet differentiable at F. The continuity of T follows from the fact that $\phi_F(x)$ is bounded and continuous if F_0 is continuous.

Proof of Proposition 3.1. Let $I_i(x)$ be the indicator function of the set $\{X_i \le x\}$ and $W_i = \int_{-\infty}^0 [I_i(x)-F(x)] \mathrm{d}x$. Note that

$$E|W_i| \le \int E|I_i(x)-F(x)| \mathrm{d}x = 2\int F(x)[1-F(x)] \mathrm{d}x < \infty.$$

Thus, from the SLLN,

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

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

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

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

For the second assertion, note that

$$\begin{aligned}(n-1)\textstyle\sum_{i=1}^{n}[\int |F_{ni}(x)-F_n(x)| \mathrm{d}x]^2 &= (n-1)^{-1}\textstyle\sum_{i=1}^{n}(\int |F_n(x)-I_i(x)| \mathrm{d}x)^2 \\ &\le 2(1-n^{-1})[\int |F_n(x)-F(x)| \mathrm{d}x]^2 + 2(n-1)^{-1}\textstyle\sum_{i=1}^{n}[\int |I_i(x)-F(x)| \mathrm{d}x]^2.\end{aligned}$$

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

$$n^{-1}\textstyle\sum_{i=1}^{n}[\int |I_i(x)-F(x)| \mathrm{d}x]^2 \to E[\int |I_1(x)-F(x)| \mathrm{d}x]^2 \ \ a.s.$$

by the SLLN. The result then follows from $\int |F_n(x)-F(x)| \mathrm{d}x \to 0$ *a.s.* □

Proof of Proposition 3.2. (i) Since $\lambda_F'(t_0)\neq 0$, $\lambda_F(t)$ is strictly decreasing and continuous in a neighborhood $\mathbf{N}\subset\mathbf{N}_{t_0}$. Assume that $\mathbf{N} = [t_0-\tau, t_0+\tau]$ with a $\tau>0$. An application of Theorem 2 of Jennrich (1969) yields that $\int\psi(x, t)\mathrm{d}[F_n(x)-F(x)] \to 0$ *a.s.* uniformly in $t\in\mathbf{N}$. Since

$$\max_{i\le n}\sup_{t\in\mathbf{N}} |\int\psi(x, t)\mathrm{d}[F_{ni}(x)-F_n(x)]| \le n^{-1}\max_{i\le n}M(X_i) \to 0 \ \ a.s.$$

by Lemma 3 of Ghosh et al. (1984), we have

$$\max_{i\le n}\sup_{t\in\mathbf{N}} |\int\psi(x, t)\mathrm{d}[F_{ni}(x)-F(x)]| \to 0 \ \ a.s. \tag{4.3}$$

Since $\lambda_F(t_0-\tau)$ and $\lambda_F(t_0+\tau)$ are nonzero and $\lambda_F(t)$ is decreasing in t, (4.3) implies

$$P\{\ T(F_{ni})\in\mathbf{N},\ i=1,\ldots,n, \text{ for sufficiently large } n\ \} = 1. \tag{4.4}$$

Then from (4.3) and $\int\psi(x, T(F_{ni}))\mathrm{d}F_{ni}(x) = 0$,

$$\max_{i\le n} |\int\psi(x, T(F_{ni}))\mathrm{d}F(x)| \to 0 \ \ a.s. \tag{4.5}$$

Thus, (3.2) follows from the fact that the inverse function of $\lambda_F(t)$ exists and is continuous at $\int\psi(x, t_0)\mathrm{d}F(x) = 0$.

(ii) Under the conditions in (ii), (4.3) and (4.5) hold with ψ replaced by the function r. We now show (4.4) also holds for a neighborhood $\mathbf{N}_{t_0}$. Suppose that there is a sequence $\{\ i_j, n_j\ \}$ such that $t_j=T(F_{n_j i_j})\to\infty$. From (3.6),

$$\int r(x, t_j)\mathrm{d}F_{n_j i_j}(x) \to \alpha.$$

Since $\int r(x, t_0)\mathrm{d}F_{n_j i_j}(x) \to \int r(x, t_0)\mathrm{d}F(x) < \alpha$,

$$\int r(x, t_j)\mathrm{d}F_{n_j i_j}(x) > \int r(x, t_0)\mathrm{d}F_{n_j i_j}(x) \quad \text{for large } j,$$

which is contrary to the definition of t_j. Hence (4.4) holds. The rest of the proof is similar to (i). □

Proof of Theorem 3.4. From Proposition 3.2, T satisfies (3.2). Under condition (a), $\lambda_F'(t)$ is continuous at t_0 and $\sup_{t\in N_{t_0}}|\lambda_{F_n}'(t)-\lambda_F'(t)| \to 0$ *a.s.* Then from the proof of Theorem 2.3, (3.4) and (3.5) hold under (b) or (b′) and

$$\max_{i\le n}\sup_{t\in N_{t_0}}|\lambda_{F_{ni}}'(t)-\lambda_{F_n}'(t)| = \max_{i\le n}\sup_{t\in N_{t_0}}|(n-1)^{-1}\textstyle\sum_{j\ne i}\eta(X_j, t)-n^{-1}\sum_{i=1}^{n}\eta(X_i, t)|$$

$$\le n^{-1}\max_{i\le n}M_1(X_i) \to 0 \ \ a.s.$$

Then the result follows from Theorem 3.3. □

References

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

Boos, D. D. and Serfling, R. L. (1980). A note no 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. *Prob. Theory and Related Fields* **73**, 197-209.

Ghosh, M., Parr, W. C., Singh, K. and Babu, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12**, 1130-1135.

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.

James, B. R. (1975). A functional law of the iterated logarithm for weighted empirical distributions. *Ann. Probability* **3**, 762-772.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.

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.* Vol. 1, rev. ed., Ungar, 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.

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

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

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**, to appear.

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

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