FUNCTIONAL CALCULUS AND ASYMPTOTIC THEORY FOR STATISTICAL ANALYSIS

by

Jun Shao
Purdue University

Technical Report #88-57

Department of Statistics
Purdue University

November 1988

FUNCTIONAL CALCULUS AND ASYMPTOTIC THEORY FOR STATISTICAL ANALYSIS

Jun Shao
Purdue University, West Lafayette, IN, USA

ABSTRACT

Using a functional calculus approach, we study asymptotic properties of statistics obtained by evaluating some functionals at the empirical distribution function. Asymptotic properties of the corresponding bootstrap statistics are also obtained. The results are applied to robust M-estimation and L-estimation problems.

Keywords: Bahadur representation, bootstrap, convergence rate, Lipschitz continuity and differentiability, L- and M-estimators.

1. Introduction

Let $\mathbf{F}$ be a convex class of distribution functions on $\mathbf{R}$ containing all degenerate distributions and T be a functional defined on $\mathbf{F}$. The unknown quantity of interest is $T(F)$, where $F \in \mathbf{F}$ is an unknown population distribution function. Statistical inferences for $T(F)$ are based on n independent and identically distributed (i.i.d.) samples $X_1, \ldots, X_n$ from F and the statistic $T(F_n)$, where $F_n(x) = n^{-1}\sum_{i=1}^n I_{(X_i \le x)}$ is the empirical distribution function. Asymptotic properties of $T(F_n)$ depend on the smoothness of the functional T. Von Mises (1947) introduced a functional calculus approach for studying properties of $T(F_n)$. Consistency, asymptotic normality and law of iterated logarithm (LIL) for $T(F_n)$ were established in many situations by using this approach (e.g., Reeds, 1976; Boos, 1979; Boos and Serfling, 1980; Serfling, 1980; Huber, 1981; Clarke, 1983). Note that the functional calculus approach not only establishes asymptotic normality and LIL for $T(F_n)$, but also provides a useful tool for robust statistical and sensitivity analyses (Huber, 1981) and for studying other asymptotic properties of $T(F_n)$ (which is the objective of this paper) such as

(1) exponential-rate bounds for $P(\ |T(F_n)-T(F)|>t\ )$, which gives estimates of deviation probabilities;

(2) Bahadur representations;

(3) convergence rate of $T(F_n)-T(F)$ to the normal distribution; and

(4) asymptotic representations for bootstrap (Efron, 1979) type of statistics.

In this paper, we study (1)-(4) for two classes of functionals corresponding to commonly used estimators of $T(F)$ in robust estimation (Huber, 1981): the M- and L-estimators. In Section 2, we introduce basic tools and prove some general results. The M- and L-estimators are studied in Sections 3 and 4, respectively, as applications of the general results obtained in Section 2.

2. General Results

Let r be a metric on $\mathbf{F}$, e.g., the Lévy distance, the Prohorov distance, the Kolmogorov distance and L_p-distance (Huber, 1981). The following definitions are with respect to (w.r.t.) the metric r.

Definition. A functional T on $\mathbf{F}$ is

(i) continuous at F if

$$T(G)-T(F)=o(1) \text{ as } r(G,F)\to 0, G\in\mathbf{F};$$

(ii) locally Lipschitz continuous at F if

$$T(G)-T(F)=O[r(G,F)] \text{ as } r(G,F)\to 0, G\in\mathbf{F};$$

(iii) differentiable at F if there is a real-valued function ϕ_F on $\mathbf{R}$ such that

$$T(G)-T(F)-\int\phi_F(x)\mathrm{d}[G(x)-F(x)]=o[r(G,F)] \text{ as } r(G,F)\to 0, G\in\mathbf{F};$$

(iv) locally Lipschitz differentiable at F if

$$T(G)-T(F)-\int\phi_F(x)\mathrm{d}[G(x)-F(x)]=O[r^2(G,F)] \text{ as } r(G,F)\to 0, G\in\mathbf{F}.$$

The differentiability of T in the above definition is referred to as Fréchet differentiability in the literature. The function ϕ_F is called the influence curve of T, 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). Without loss of generality, we assume that $\int\phi_F(x)\mathrm{d}F(x)=0$. If we write

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

$\int\phi_F\mathrm{d}(G-F)$ is then the differential of T at F, a linear functional presenting the linear component of $T(G)-T(F)$, and $R(G,F)$ is the remainder.

We now establish some asymptotic properties of $T(F_n)$ using the Kolmogorov distance:

$$r(G,F)=\|G-F\|_\infty=\sup_x|G(x)-F(x)|.$$

If T is continuous at F, Huber (1981) showed that $T(F_n)$ is consistent ($T(F_n) \to T(F)$ *a.s.*) and robust in the sense that if G_n is the empirical distribution corresponding to i.i.d. samples from $G \in \mathbf{F}$, then for any $\varepsilon>0$, there is a $\delta>0$ such that $r(H_{T(G_n)}, H_{T(F_n)}) < \varepsilon$ for sufficiently large n whenever $r(G,F) < \delta$, where $H_{T(G_n)}$ is the distribution function of $T(G_n)$. The following result gives an exponential-rate bound for the error in approximating $T(F)$ by $T(F_n)$.

Theorem 2.1. If T is locally Lipschitz continuous at F w.r.t. $\| \ \|_\infty$, then for any $t>0$,

$$P(\ |T(F_n) - T(F)| > t\ ) \le C(e^{-2nK^{-2}t^2} + e^{-2n\delta^2}),$$

where K and δ are constants depending only on T and F (K is the Lipschitz constant) and C is a constant independent of n, T and F.

Remarks. (i) The result in Theorem 2.1 implies that $T(F_n)$ converges to $T(F)$ completely (Hsu and Robbins, 1947).

(ii) The value of t in the above inequality may depend on n if desired. For example, an asymptotic estimate of the moderate deviation probability $P(\ n^{1/2}[T(F_n) - T(F)] \ge (a \log n)^{1/2}\ )$ is $O(n^{-2aK^{-2}})$.

Proof. From Definition (ii), there are constants $\delta>0$ and $K>0$ such that

$$|T(G) - T(F)| \le K \|G-F\|_\infty \quad \text{when} \quad \|G-F\|_\infty \le \delta.$$

From Dvoretzky, Kiefer and Wolfowitz inequality, $P(\ K\|F_n - F\|_\infty > t\ ) \le Ce^{-2nK^{-2}t^2}$ and $P(\ \|F_n - F\|_\infty > \delta\ ) \le Ce^{-2n\delta^2}$. Thus, the result follows from

$$P(\ |T(F_n) - T(F)| > t\ ) \le P(\ K\|F_n - F\|_\infty > t\ ) + P(\ \|F_n - F\|_\infty > \delta\ ). \ \square$$

It is well known that $\|F_n - F\|_\infty = O_p(n^{-1/2})$ and $\|F_n - F\|_\infty = O[n^{-1/2}(\log\log n)^{1/2}]$ *a.s.* Note that $\int \phi_F(x) d[F_n(x) - F(x)] = n^{-1}\sum_{i=1}^n \phi_F(X_i)$ is an average of i.i.d. random variables. Hence if T is differentiable at F w.r.t. $\| \ \|_\infty$ and

$$0 < \sigma^2 = \int \phi_F^2(x) dF(x) < \infty, \tag{2.1}$$

then the asymptotic normality and LIL hold for $T(F_n)$, i.e.,

$$n^{1/2}[T(F_n) - T(F)] \to N(0, \sigma^2) \quad \text{in distribution}$$

and

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

(see Serfling, 1980). If T is locally Lipschitz differentiable, more asymptotic properties of $T(F_n)$ can be obtained.

Theorem 2.2. (i) Bahadur representation. Suppose that T is locally Lipschitz differentiable at F w.r.t. $\| \ \|_\infty$. Then

$$T(F_n) - T(F) - n^{-1}\sum_{i=1}^n \phi_F(X_i) = O(n^{-1}\log\log n) \quad a.s. \tag{2.2}$$

and

$$T(F_n) - T(F) - n^{-1}\sum_{i=1}^n \phi_F(X_i) = O_p(n^{-1}). \tag{2.3}$$

(ii) Convergence rate. Suppose that T is locally Lipschitz differentiable at F w.r.t. $\| \ \|_\infty$, $\sigma^2 > 0$ and $\int |\phi_F(x)|^3 dF(x) < \infty$. Then

$$\sup_x |P(\, n^{1/2}[T(F_n) - T(F)]/\sigma < x \,) - \Phi(x)| = O(n^{-1/2}\log n),$$

where Φ is the standard normal distribution function.

Proof. From the Lipschitz differentiability of T at F and $\|F_n - F\|_\infty \to 0$ $a.s.$, almost surely,

$$|T(F_n) - T(F) - n^{-1}\sum_{i=1}^n \phi_F(X_i)| \le K \|F_n - F\|_\infty^2$$

for sufficiently large n, where K is a constant. Thus, (2.2) follows from $\|F_n - F\|_\infty^2 = O(n^{-1}\log\log n)$ $a.s.$ The proof for (2.3) is similar.

For (ii), using Dvoretzky, Kiefer and Wolfowitz inequality, we obtain

$$P(\, \|F_n - F\|_\infty^2 > n^{-1}\log n \,) = O(n^{-2}).$$

Let $R(F_n,F) = T(F_n) - T(F) - n^{-1}\sum_{i=1}^n \phi_F(X_i)$. From the Lipschitz differentiability of T at F, there is a $\delta>0$ such that

$$P(\ |R(F_n,F)|>Kn^{-1}\log n\ ) = P(\ \|F_n-F\|_\infty^2>n^{-1}\log n\ )$$
$$+P(\ \|F_n-F\|_\infty>\delta\ ) = O(n^{-2}).$$

Note that (see Serfling, 1980, p.229)

$$\sup_x |P(\ n^{1/2}[T(F_n)-T(F)]/\sigma < x\ ) - \Phi(x)| \le P(\ |R(F_n,F)|>Kn^{-1}\log n\ )$$
$$+ O(n^{-1/2}\log n) + \sup_x |P(\ n^{-1/2}\sum_{i=1}^n \phi_F(X_i)/\sigma < x\ ) - \Phi(x)|.$$

The result follows from Berry-Esséen theorem for $n^{-1}\sum_{i=1}^n \phi_F(X_i)$. □

Under more smoothness conditions on T (e.g., T is second order differentiable), we can obtain the best convergence rate $O(n^{-1/2})$. See Serfling (1980, Theorem 6.4.3).

Let $X_1^*,\dots,X_n^*$ be i.i.d. samples from F_n and $F_n^*(x) = n^{-1}\sum_{i=1}^n I_{(X_i^*\le x)}$. X_i^* are bootstrap samples (see Efron, 1979) and $T(F_n^*)$ is a bootstrap analog of $T(F_n)$. The following result gives a representation for the bootstrap statistic $T(F_n^*) - T(F_n)$.

Theorem 2.3. If T is locally Lipschitz differentiable at F w.r.t. $\|\ \|_\infty$, then

$$T(F_n^*) - T(F_n) - n^{-1}[\sum_{i=1}^n \phi_F(X_i^*) - \sum_{i=1}^n \phi_F(X_i)] = O(n^{-1}\log n)\ \ a.s. \tag{2.4}$$

and

$$T(F_n^*) - T(F_n) - n^{-1}[\sum_{i=1}^n \phi_F(X_i^*) - \sum_{i=1}^n \phi_F(X_i)] = O_p(n^{-1}). \tag{2.5}$$

Proof. For the first assertion, it suffices to show that

$$R(F_n^*,F) = T(F_n^*) - T(F) - n^{-1}\sum_{i=1}^n \phi_F(X_i^*) = O(n^{-1}\log n)\ \ a.s.$$

Let P_* be the probability corresponding to the bootstrap sampling. From Dvoretzky, Kiefer and Wolfowitz inequality,

$$P(\ \|F_n^*-F_n\|_\infty>n^{-1/2}(\log n)^{1/2}\ ) = E[P_*(\ \|F_n^*-F_n\|_\infty>n^{-1/2}(\log n)^{1/2}\ )] \le Cn^{-2}.$$

Hence $\|F_n^* - F_n\|_\infty^2 = O(n^{-1}\log n)$ *a.s.* and $\|F_n^* - F_n\|_\infty \to 0$ *a.s.* Then from the Lipschitz differentiability of T at F, almost surely,

$$|R(F_n^*,F)| \le K\,\|F_n^*-F\|_\infty^2$$

for sufficiently large n. The result follows from $\|F_n^*-F\|_\infty^2 \le 2(\|F_n^*-F_n\|_\infty^2 + \|F_n-F\|_\infty^2)$.

From Dvoretzky, Kiefer and Wolfowitz inequality, $\|F_n^*-F_n\|_\infty^2 = O_p(n^{-1})$. Hence the proof of the second assertion is similar. □

A direct application of the representation (2.4) gives

$$\sup_x |P(\, n^{1/2}[T(F_n^*)-T(F_n)] < x\,) - P(\, n^{1/2}[T(F_n)-T(F)] < x\,)| \to 0 \;\; a.s. \qquad (2.6)$$

under (2.1). That is, the bootstrap estimator of the sampling distribution of $T(F_n)-T(F)$ is strongly consistent.

It is possible to establish some of the preceding results using other metrics. For example, the L_2 distance given by

$$r(G,F) = \|G-F\|_2 = \{\int [G(x)-F(x)]^2 dx\}^{1/2}$$

(assume that $\mathbf{F} = \{\, G: \int |x|\, dG(x) < \infty \,\}$ and $F \in \mathbf{F}$). We prove the following results for L_2 distance. They will be applied to L-estimation problems in Section 4.

Theorem 2.4. Suppose that $E|X_1| < \infty$ and T is locally Lipschitz differentiable at F w.r.t. $\|\ \|_2$. Then

(i) (2.3) holds;

(ii) (2.2) holds if $E|X_1|^{1+\delta} < \infty$ for a $\delta>0$;

(iii) if $E|X_1|^k < \infty$ for a positive integer k,

$$\sup_x |P(\, n^{1/2}[T(F_n) - T(F)]/\sigma < x\,) - \Phi(x)| = O(n^{-1/2+\tau}),$$

where $\tau=1/2(k+1)$.

Proof. From the proof of Theorem 2.2, (2.3) follows since

$$E\,\|F_n-F\|_2^2 = n^{-1}\int F(x)[1-F(x)]\mathrm{d}x = O(n^{-1}).$$

If $E\,|X_1|^{1+\delta} < \infty$ for a $\delta>0$, then there is an $\varepsilon>0$ such that $\int[F(x)(1-F(x)]^{1-2\varepsilon}\mathrm{d}x < \infty$. From James (1975), $W_n = \|(F_n-F)/[F(1-F)]^{1/2-\varepsilon}\|_\infty = O[n^{-1/2}(\log\log n)^{1/2}]$ *a.s.* Then (2.2) follows from

$$\|F_n-F\|_2 \le W_n^2\int[F(x)(1-F(x))]^{1-2\varepsilon}\mathrm{d}x = O(n^{-1}\log\log n) \ \ a.s.$$

If $E\,|X_1|^k < \infty$, then from Lemma 8.2.5B in Serfling (1980), $E\,\|F_n-F\|_2^{2k} = O(n^{-k})$. Then the result in part (iii) follows from

$$P(\ \|F_n-F\|_2^2 > n^{-1+\tau}\ ) \le n^{k-k\tau}O(n^{-k}) = O(n^{-k\tau}) = O(n^{-1/2+\tau}). \ \Box$$

For the bootstrap statistic $T(F_n^*)-T(F_n)$, (2.5) still holds if T is locally Lipschitz differentiable w.r.t. $\|\ \|_2$. It is not clear whether (2.4) holds in this case. However, we establish (2.6) under some moment condition.

Theorem 2.5. Suppose that $E\,|X_1| < \infty$ and T is locally Lipschitz differentiable at F w.r.t. $\|\ \|_2$. Then (2.5) holds. If in addition, $E\,|X_1|^{2+\delta} < \infty$ for a $\delta>0$, then (2.6) holds.

Proof. Let E_* be the expectation taken under the bootstrap probability P_*. Then

$$E[E_*\,\|F_n^* - F_n\|_2^2] = n^{-1}E[\int F_n(x)(1-F_n(x))\mathrm{d}x] = n^{-1}(1-n^{-1})\int F(x)(1-F(x))\mathrm{d}x.$$

Hence (2.5) follows since $\|F_n^*-F_n\|_2^2 = O_p(n^{-1})$ under $E\,|X_1| < \infty$. For the second assertion, let $X_{(1)} = \min_{i\le n}X_i$ and $X_{(n)} = \max_{i\le n}X_i$. Then $\|F_n^*-F_n\|_2^2 \le \|F_n^*-F_n\|_\infty^2(X_{(n)}-X_{(1)})$. From the proof of Theorem 2.3, $\|F_n^*-F_n\|_\infty^2 = O(n^{-1}\log n)$ *a.s.* Then $n^{1/2}\|F_n^*-F_n\|_2^2 \to 0$ *a.s.* follows from $(X_{(n)}-X_{(1)})/n^{1/(2+\delta)} \to 0$ *a.s.* Hence (2.6) holds. $\Box$

3. M-estimators

We apply the results in Section 2 to robust M-estimation problems. Let $\rho(x, t)$ be a real-valued function on $\mathbf{R}^2$. For $G \in \mathbf{F}$, define $T(G)$ to be a solution of

$$\int \rho(x, T(G)) \mathrm{d}G(x) = \min_t \int \rho(x, t) \mathrm{d}G(x).$$

T is called the M-functional and $T(F_n)$ is the M-estimator of $T(F)$. Examples of M-estimators can be found in Serfling (1980) and Lehmann (1983). Assume that for each t, $\psi(x, t) = \partial \rho(x, t)/\partial t$ exists a.e. Lebesgue and $\partial \int \rho(x, t) \mathrm{d}G(x)/\partial t = \int \psi(x, t) \mathrm{d}G(x)$. Thus, $\int \psi(x, T(G)) \mathrm{d}G(x) = 0$. Assume also that the function $\lambda_F(t) = \int \psi(x, t) \mathrm{d}F(x)$ has positive derivative at $T(F)$, i.e., $\lambda_F'(T(F))>0$. The continuity and differentiability of T are studied for a class of M-functionals corresponding to robust M-estimators.

Theorem 3.1. Let T be an M-functional and assume $T(F)$ is the unique minimum of $\int \rho(x, t) \mathrm{d}F(x)$.

(i) Assume that there is an α (may be infinity) such that $\alpha > \int \rho(x, T(F)) \mathrm{d}F(x)$ and for any $c>0$,

$$\lim_{t \to \infty} \rho(x, t) = \alpha \quad \text{uniformly in } x \in \{ x : |x| \le c \}. \tag{3.1}$$

Assume further that either ρ is bounded or ψ is bounded and continuous in t and $\lambda_F(t)$ has a unique root. Then T is continuous at F w.r.t. $\| \ \|_\infty$.

(ii) Let $\| \ \|_V$ be the total variation norm (see Natanson, 1961). Assume the conditions in (i) and there is a neighborhood $\mathbf{N}$ of $T(F)$ such that $\| \psi(\cdot, t) \|_V < \infty$ for $t \in \mathbf{N}$. Then T is locally Lipschitz continuous at F w.r.t. $\| \ \|_\infty$.

(iii) Assume the conditions in (ii) and $\| \psi(\cdot, t) - \psi(\cdot, T(F)) \|_V \to 0$ as $t \to T(F)$. Then T is differentiable at F w.r.t. $\| \ \|_\infty$. The influence function is $\phi_F = -\psi(x, T(F))/\lambda_F'(T(F))$.

(iv) Assume the conditions in (ii) and $\zeta(x, t) = \partial \psi(x, t)/\partial t$ exists for $t \in \mathbf{N}$ and satisfies $\| \zeta(\cdot, T(F)) \|_V < \infty$ and

$$|\zeta(x,t)-\zeta(x,T(F))| \le M(x)|t-T(F)| \quad \text{for } t\in\mathbf{N}, \tag{3.2}$$

where $M(x)$ satisfies $\int M(x)\mathrm{d}F(x) < \infty$. Then T is locally Lipschitz differentiable at F w.r.t. $\|\ \|_\infty$.

Remark. If ψ is increasing in t, condition (3.1) is not required.

Proof. We only give partial proofs for (i) and (iv). Proofs for (ii) and (iii) are similar. Assume that ψ is bounded and continuous in t. Let $G_n \in \mathbf{F}$ be a sequence satisfying $\|G_n - F\|_\infty \to 0$. From Jennrich (1969), for any $C>0$,

$$\int\psi(x,t)\mathrm{d}G_n(x) \to \int\psi(x,t)\mathrm{d}F(x) \text{ uniformly in } t\in\{\, t:\ |t|\le C\ \}.$$

Let ξ be a limit point of $\{\ T(G_n),\ n=1,2,...\ \}$. Condition (3.1) implies $|\xi| < \infty$. Let $\{\ n_j\ \}$ be a subsequence such that $T(G_{n_j}) \to \xi$. Then

$$\int\psi(x,T(G_{n_j}))\mathrm{d}G_{n_j}(x) - \int\psi(x,T(G_{n_j}))\mathrm{d}F(x) = -\int\psi(x,T(G_{n_j}))\mathrm{d}F(x) \to 0.$$

But $\int\psi(x,T(G_{n_j}))\mathrm{d}F(x) \to \int\psi(x,\xi)\mathrm{d}F(x) = \lambda_F(\xi)$. From the uniqueness of the root of $\lambda_F(t)$, $\xi = T(F)$. This proves $T(G_n) \to T(F)$ and therefore T is continuous at F.

Assume the conditions in (iv). Note that

$$T(G) - T(F) = \int\phi_F(x)\mathrm{d}[G(x)-F(x)] + R_1(G,F) + R_2(G,F),$$

where

$$R_1(G,F) = \int[\psi(x,T(G))-\psi(x,T(F))]\mathrm{d}[G(x)-F(x)]/h(T(G)),$$

$$R_2(G,F) = [1/\lambda_F'(T(F)) - 1/h(T(G))]\int\psi(x,T(F))\mathrm{d}[G(x)-F(x)],$$

and $h(t) = \lambda_F(t)/(t-T(F))$ if $t\neq T(F)$ and $=0$ if $t=T(F)$. Under the conditions in (ii), $T(G)-T(F) = O(\|G-F\|_\infty)$. Then under (3.2),

$$\int[\psi(x,T(G))-\psi(x,T(F))]\mathrm{d}[G(x)-F(x)] = O(\|G-F\|_\infty^2)$$
$$+ [T(G)-T(F)]\int\zeta(x,T(F))\mathrm{d}[G(x)-F(x)].$$

Hence $R_1(G,F) = O(\|G-F\|_\infty^2)$ since

$$\left|\int\zeta(x,T(F))\mathrm{d}[G(x)-F(x)]\right| \le \|\zeta(\cdot,T(F))\|_V \|G-F\|_\infty.$$

Similarly, $1/\lambda_F'(T(F)) - 1/h(T(G)) = O(\|G-F\|_\infty)$. Hence $R_2(G,F) = O(\|G-F\|_\infty^2)$ since

$$\left|\int \psi(x, T(F)) \mathrm{d}[G(x)-F(x)]\right| \le \|\psi(\cdot, T(F))\|_V \|G-F\|_\infty .$$

This completes the proof of (iv). □

Therefore, the results in Theorems 2.1-2.3 apply to M-estimators under appropriate conditions. Carroll (1978) establishes Bahadur representation for location M-estimators with a remainder term $O(n^{-1}\log n)$ *a.s.* Under the conditions in Theorem 3.1(iv), our result (2.2) gives a sharper bound on the remainder term.

4. L-estimators

Another class of commonly used estimators is the class of L-estimators $T(F_n)$, where T is defined to be

$$T(G) = \int x J[G(x)] \mathrm{d}G(x), \quad G \in \mathbf{F} \tag{4.1}$$

with a function $J(t)$ defined on [0,1]. Serfling (1980) and Lehmann (1983) provide many examples of L-estimators. Let

$$\phi_F(x) = -\int [I_{(x \le y)} - F(y)] J[F(y)] \mathrm{d}y. \tag{4.2}$$

and

$$R(G,F) = \int W[G(x)][G(x)-F(x)] \mathrm{d}x, \tag{4.3}$$

where $W[G(x)] = [G(x)-F(x)]^{-1}\int_{F(x)}^{G(x)} J(t)\mathrm{d}t - J[F(x)]$ if $G(x) \ne F(x)$ and $=0$ if $G(x) = F(x)$. Then

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

We first consider a class of trimmed L-functionals which corresponds to robust L-estimation. An example of trimmed L-estimator is the trimmed mean (Serfling, 1980, p.236).

Theorem 4.1. Let T be defined in (4.1) with a function J satisfying $J(t) = 0$ for $t<\alpha$ or $t>\beta$, where $0<\alpha<\beta<1$.

(i) If J is bounded, continuous a.e. Lebesgue and a.e. F^{-1}, then T is differentiable and locally Lipschitz continuous at F w.r.t. $\| \ \|_\infty$. The influence function is given by (4.2).

(ii) If J is Lipschitz continuous, i.e., $|J(t)-J(s)|\le K|t-s|$ for $t, s\in[\alpha,\beta]$ and a constant $K>0$, then T is locally Lipschitz differentiable at F w.r.t. $\| \ \|_\infty$.

Proof. (i) The differentiability of T was shown in Boos (1979). We show that T is locally Lipschitz continuous at F. Choose two constants a and b such that a and b are continuity points of F and $F(a)<\alpha$ and $F(b)>\beta$. Then

$$
\begin{aligned}
\left|\int \phi_F(x)\mathrm{d}[G(x)-F(x)]\right| &= \left|\int_a^b [F(x)-G(x)]J[F(x)]\mathrm{d}x\right| \\
&\le (b-a)\|J\|_\infty \|G-F\|_\infty .
\end{aligned}
$$

The result follows since $R(G,F) = o(\|G-F\|_\infty)$ by the differentiability of T at F.

(ii) Choose an $\varepsilon_0>0$ so that $F(a)+\varepsilon_0<\alpha$ and $F(b)-\varepsilon_0>\beta$. Then for G satisfying $\|G-F\|_\infty < \varepsilon_0$, $R(G,F) = \int_a^b W[G(x)][G(x)-F(x)]\mathrm{d}x$. From the Lipschitz continuity of J, $|W[G(x)]| \le K|G(x)-F(x)|$. Then $|R(G,F)| \le K(b-a)\|G-F\|_\infty^2$. □

Hence the results in Theorems 2.1-2.3 hold for trimmed L-estimators. For general untrimmed L-estimators, the following result shows that the corresponding L-functionals are locally Lipschitz differentiable at F w.r.t. the L_2 distance. Therefore the results in Theorems 2.4 and 2.5 hold for untrimmed L-estimators.

Theorem 4.2. Assume $E|X_1| < \infty$. Let T be given by (4.1) with J being Lipschitz continuous on [0,1].

(i) T is locally Lipschitz differentiable at F w.r.t. $\| \ \|_2$ and the influence function is given in (4.2).

(ii) If in addition, $\int J^2[F(x)]dx < \infty$, then T is locally Lipschitz continuous w.r.t. $\| \ \|_2$.

Proof. (i) Let $R(G,F)$ be defined in (4.3). The result follows from

$$|R(G,F)| \le K\int[G(x)-F(x)]^2 dx = K\|G-F\|_2^2.$$

(ii) From Cauchy-Schwarz inequality,

$$\left|\int \phi_F(x)d[G(x)-F(x)]\right| \le \|G-F\|_2 \{\int J^2[F(x)]dx\}^{1/2}.$$

The result follows from part (i). □

An example of J satisfying the condition in Theorem 4.2(i) is $J(t)=4t-2$ (the corresponding L-estimator is the Gini's mean difference). Examples of J satisfying $\int J^2[F(x)]dx < \infty$ are J satisfying $|J(t)| \le M[t(1-t)]^{1/2}$ with a constant $M>0$. See Serfling (1980, p.277).

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.

Carroll, R. J. (1978). On almost sure expansions for M-estimates. *Ann. Statist.* **8**, 314-318.

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

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

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

Hsu, P. L. and Robbins, H. (1947). Complete convergence and the law of large numbers. *Proc.*

Nat. Acad. Sci. U.S.A. **33**, 25-31.

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.

Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

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

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

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

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