MU-ESTIMATING AND SMOOTHING

Running Title: μ-estimation

Z.J. Liu and C.R. Rao

Technical Report 99-01

March 1999

Center for Multivariate Analysis
417 Thomas Building
Penn State University
University Park, PA 16802

Research work of authors was supported by the Army Research Office under Grant DAAHO4-96-1-0082. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	3-26-99	Technical - March 1999

4. TITLE AND SUBTITLE
Mu-estimating and Smoothing

5. FUNDING NUMBERS
DAAH04-96-1-0082

6. AUTHOR(S)
Z.J. Liu and C.R. Rao

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Center for Multivariate Analysis
417 Thomas Building
Department of Statistics
Penn State University
University Park, PA 16802

8. PERFORMING ORGANIZATION REPORT NUMBER
99-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 35518.49-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release: distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT

In the traditional M-estimation theory developed by Huber (1964), the parameter under estimation is the value of θ which minimizes the expectation of what is called a discrepancy measure (DISM) $\delta(x, \theta)$ which is a function of θ and the underlying random variable X. Such a setting does not cover the estimation of parameters such as multivariate median defined by Oja (1983) and Liu (1990), as the value of θ which minimizes the expectation of a DISM of the type $\delta(X_1, \ldots, X_m, \theta)$ where $X_1, \ldots, X_m$ are independent copies of the underlying random variable X. Arcones et al (1994) studied the estimation of such parameters. We call the associated M-estimation MU-estimation (or μ-estimation for convenience). When a DISM is not a differentiable function of θ, some complexities arise in studying the properties of estimations as well as in their computation. In such a case, we introduce a new method of smoothing the DISM with a kernel function and using it in estimation. It is seen that smoothing allows up to develop an elegant approach to the study of asymptotic properties and computation of estimations.

14. SUBJECT TERMS	15. NUMBER IF PAGES
Asymptotic, data depth, discrepancy, estimating equation, kernel, multivariate median, M-estimation, smooth, U-statistic	21
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

MU-estimation and Smoothing

(Running Title: μ-**estimation**)

by

Z. J. Liu and C. R. Rao

Department of Mathematics and Statistics, Mississippi State University, Mississippi State, MS 39762

Department of Statistics, Penn State University, University Park, PA 16802

Abstract

In the traditional M-estimation theory developed by Huber (1964), the parameter under estimation is the value of θ which minimizes the expectation of what is called a discrepancy measure (DISM) $\delta(x, \theta)$ which is a function of θ and the underlying random variable X. Such a setting does not cover the estimation of parameters such

Key Words and Phrases: Asymptotic, data depth, discrepancy, estimating equation, kernel, multivariate median, M-estimation, smooth, U-statistic.

AMS 1991 Subject Classifications. Primary: 62F35; **Secondary:** 62F12, 62H12.

as multivariate median defined by Oja (1983) and Liu (1990), as the value of θ which minimizes the expectation of a DISM of the type $\delta(X_1, ..., X_m, \theta)$ where $X_1, ..., X_m$ are independent copies of the underlying random variable X. Arcones et al (1994) studied the estimation of such parameters. We call the associated M-estimation MU-estimation (or μ-estimation for convenience). When a DISM is not a differentiable function of θ, some complexities arise in studying the properties of estimations as well as in their computation. In such a case, we introduce a new method of smoothing the DISM with a kernel function and using it in estimation. It is seen that smoothing allows us to develop an elegant approach to the study of asymptotic properties and computation of estimations.

1 Introduction

Let us consider a parameter $\theta_F \in R^q$ of a p-variate distribution F that

$$\theta_F = \arg\min \{\Delta(\mu, F)\}, \tag{1}$$

where

$$\Delta(\theta, F) = E_F \delta(X_1, ...X_m, \mu), \tag{2}$$

where $E_F(.)$ denotes the expectation with respect to F, $\delta(x_1, ..., x_m, \theta) : R^{mp} \times R^q \to R$ is a symmetric function of $x_1, ..., x_m$ for every fixed θ, called discrepancy measure (DISM), and $X_1, ..., X_m$ are independent copies of a random variable X distributed

as F. Let $X_1, ..., X_n$ be a sample of n independent observations on $X \sim F$, and define F_n as the empirical distribution function based on $X_1, ..., X_n$. The μ-estimate of $\theta = (\theta_1, ..., \theta_q)$ is defined as

$$\widehat{\theta} = \arg\min\{\Delta(\theta, F_n)\}, \tag{3}$$

where we define

$$\Delta(\theta, F_n) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < i_2 ... < i_m \le n} \delta(X_{i_1}, ...X_{i_m}, \theta). \tag{4}$$

If (4) is differentiable, then $\widehat{\theta}$, an estimator of θ, is a solution of

$$D\Delta(\theta, F_n) = 0, \tag{5}$$

where D is the differential operator

$$D = \left(\frac{\partial}{\partial \mu_i}\right)_{1 \times q}, \quad D^2 = \left(\frac{\partial^2}{\partial \mu_i \partial \mu_j}\right)_{q \times q}. \tag{6}$$

If the second derivative exists, we could use an iterative procedure like Newton-Raphson to compute the solution of (5). Further the asymptotic properties of the estimator can be studied without heavy assumptions. If the DISM is not differentiable, as in the case of median estimations of Oja (1983) and Liu (1990), then the problems of computing the μ-estimate and deriving its asymptotic properties need complex approaches. To overcome the possible difficulties, we propose to smooth the DISM using a differentiable kernel function, which results in a differentiable function. We

work with such a smoothed DISM and derive the asymptotic properties of the resulting μ-estimators.

Let $k(u)$: $R^q \rightarrow R$ be a differentiable function, called a kernel function, such that

$$\int_{R^q} k(u)du = 1. \tag{7}$$

Using such a $k(u)$, we obtain the smoothed version of the DISM by convolution

$$\delta_n(x_1, ..., x_m, \theta) = \int_{R^q} \delta(x_1, ..., x_m, \mu - h_n u)k(u)du, \tag{8}$$

where h_n is a bandwidth such that $\{h_n, n = 1, 2, ...\}$ decreasing to zero as n tends to infinite.. The derivative of δ_n is

$$D\delta_n(x_1, ..., x_m, \theta) = \frac{1}{h_n^q} \int_{R^q} \delta(x_1, ..., x_m, u)Dk\left(\frac{\theta - u}{h_n}\right) du. \tag{9}$$

Let

$$\Delta_n(\theta, F_n) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < i_2 ... < i_m \leq n} \delta_n(X_{i_1}, ...X_{i_m}, \theta). \tag{10}$$

Then the μ-estimator obtained from the smoothed DISM is a solution of the equation

$$D\Delta_n(\theta, F_n) = 0. \tag{11}$$

Since we have the estimating equation in an explicit form, we can use standard computational techniques to obtain the estimator $\widehat{\theta}$. We now investigate the asymptotic stochastic properties of the smoothed μ-estimator. We establish the consistency, asymptotic normality and asymptotic representation of $\widehat{\theta}$ in Section 3 of the paper.

A discussion of smoothed μ-estimator of the median when $p = 1$ and $m = 1$, and multivariate median of Liu (1990) is given in Section 4, and the results are compared with those obtained by Arcones et al (1994) using the unsmoothed DISM.

For convenience and easy understanding of the paper, we list some symbols and statements in the following:

- $\{X_i, i = 1, ..., n\}$ is an *i.i.d.* sample from a p-variate distribution F. $f(x)$ denotes the density function, and F_n the empirical distribution.

- $\|.\|$ denotes the Euclidean norm. C denote a positive number which may take different values in different places. Denote $a_n = n^{-1} \log \log n$. A matrix may be denoted as $O(.)$ (or $o(.)$) if its norm is $O(.)$ (or $o(.)$). $E(.)$ and $cov(.)$ denote mean and variance-covariance matrix respectively.

- Conditions for changing the order between integrals, between derivatives, and between integral and derivative are assumed to be satisfied.

- $k(u) : R^q \to R$ is a function, called a kernel function, such that $\int_{R^p} k(u)du = 1$. $\{h_n, n = 1, 2...\}$ is a sequence of positive numbers called bandwidth.

- $\delta(x, \theta) = \delta(x_1, ..., x_m, \theta) : R^{mp} \times R^q \to R$ is a function called a DISM (discrepancy measure) which is symmetric with respect to permutations of $(x_1, ..., x_m)$.

A kernel smoothed DISM is

$$\delta_n(x,\theta) = \int_{R^q} \delta(x_1,...,x_m,\theta - h_n u)k(u)du.$$

- Denote $A_{i,j}$ the element in the ith row and jth column of matrix A. $A^{'}$ denotes the transpose of matrix A. And

$$D = \left(\frac{\partial}{\partial\theta_i}\right)_{1\times q}, \quad D^2 = D\otimes D^{'}, \quad D^3 = D\otimes\left(D\otimes D^{'}\right),$$

$$D_i = \frac{\partial}{\partial\theta_i}, \quad D^2_{i,j} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}, \quad D^3_{i,j,l} = \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_l},$$

$$k_1(u) = Dk(u), \quad k_2 = D^2k(u),$$

$$\Delta(\theta,F) = E_F\delta(X_1,...,X_m,\theta), \quad \Delta(\theta,F_n) = \binom{n}{m}^{-1}\sum_{1\le i_1<i_2...<i_m\le n}\delta(X_{i_1},...X_{i_m},\theta),$$

$$\Delta_n(\theta,F) = E_F\delta_n(X_1,...,X_k,\theta), \quad \Delta_n(\theta,F_n) = \binom{n}{m}^{-1}\sum_{1\le i_1<i_2...<i_m\le n}\delta_n(X_{i_1},...X_{i_m},\theta).$$

$$G(\theta,F) = D\Delta(\theta,F), \quad G(\theta,F_n) = D\Delta(\theta,F_n),$$

$$G_n(\theta,F) = D\Delta_n(\theta,F), \quad G_n(\theta,F_n) = D\Delta_n(\theta,F_n).$$

$$S(\theta,F) = D^2\Delta(\theta,F), \quad S(0,F) = D^2\Delta(\theta,F)|_{\theta=0}.$$

$$\Sigma = cov\left(E\left(D\delta(X_1,...,X_m,0)\,|X_1\right)\right), \ \Sigma_n = cov\left(E\left(D\delta_n(X_1,...,X_m,0)\,|X_1\right)\right)$$

2 Existence and consistency

Under some regularity conditions, the strong law of large numbers of U-statistic implies that $G_n(\theta,F_n)$ will converge to $G(\theta,F)$ for every fixed θ, and therefore when

$G(\theta, F) = 0$ is solvable, the existence of solution of (11) depends on the uniform closeness of $G_n(\theta, F_n)$ and $G(\theta, F)$.

Theorem 1 *Assume:*

A1. $\Delta(\theta, F)$ *has a unique minimum value, reached at* $\theta = \theta_0$. *Without loss of generality, assume* $\theta_0 = 0$. $G(\theta, F)$ *and* $S(\theta, F)$ *exist for* $\theta \in B$, *an open neighborhood of 0, and* $S(\theta, F)$ *is continuous and positive definite for every* $\theta \in B$.

A2. the kernel function chosen has a compact support, is three times differentiable. $D^2\delta_n(x_1, ..., x_n, \theta)$ *is continuous and has full rank for every* $\theta \in B$.

A3. the bandwidth sequence is such that

$$\lim_{n\to\infty} h_n = 0 \tag{12}$$

A4.

$$\lim_{n\to\infty} \sup_{\theta\in B} \|G_n(\theta, F_n) - G_n(\theta, F)\| = 0. \quad w.p.1. \tag{13}$$

Then for large n, (11) has a unique solution $\widehat{\theta}$, $w.p.1$ *and*

$$\lim_{n\to\infty} \widehat{\theta} = 0, \quad w.p.1. \tag{14}$$

Remark 1 *(1) It would have been more satisfactory if the conditions under which A4 holds could be stated. We hope to study this problem in a future communication. It may be easy to verify A4 in particular cases.*

(2) When the DISM δ *is twice differentiable, no smoothing is needed. In such a case the bandwidth can be chosen to be* 0.

Proof. We prove this theorem for $q = 2$ and $h_n \neq 0$. Similarly, one can prove the theorem for other cases. Note that

$$G_n(\theta, F) = \int_{R^p} G(\theta - h_n u, F)k(u)du. \tag{15}$$

Then by conditions A2, A3, A4 and (15) we have

$$\lim_{n\to\infty} \sup_{\theta\in B} \|G_n(\theta, F) - G(\theta, F)\| = 0. \tag{16}$$

By condition A1 and the implicit function theorem, there exist continuous curves $C_1(t) = (t, c_1(t))$ and $C_2(t) = (t, c_2(t))$ such that

$$C_i(t) \subset B, \quad C_i(0) = (0, 0), \quad G(C_i(t), F) = 0, \quad i = 1, 2. \tag{17}$$

By condition A4 and (16), it is obvious that

$$\lim_{n\to\infty} \sup_{\theta\in B} \|G_n(\theta, F_n) - G(\theta, F)\| = 0, \quad w.p.1. \tag{18}$$

By condition A2, (18), and the implicit function theorem, for every $\varepsilon > 0$ and large n, there exist continuous curves $C_1'(t) = (t, c_1'(t))$ and $C_2'(t) = (t, c_2'(t))$ such that

$$C_i'(t) \subset B, \quad G_n(C_i'(t), F_n) = 0, \quad i = 1, 2, \tag{19}$$

and

$$C_i(t) - \varepsilon < C_i'(t) < C_i(t) + \varepsilon, \quad w.p.1, \quad i = 1.2. \tag{20}$$

Condition A1 implies that $C_1(t)$ and $C_2(t)$ cross with each other at a unique point which is $(0, 0)$. (20) implies that $C_1'(t)$ and $C_2'(t)$ cross with each other at a unique

point in B when ε is small enough. This cross point is $\widehat{\theta}$. Let $\varepsilon \to 0$, which requires $n \to \infty$. Then $\widehat{\theta} \to 0$, $w.p.1$, which establish the desired result.

3 Asymptotic Normality and Asymptotic Representation

We first consider a μ-estimator with a twice differentiable discrepancy function. Since we proved the existence and consistency of μ-estimates under certain conditions, we assume $\widehat{\theta}$ exists and is consistent in the following.

Theorem 2 *Assume that the DISM δ is twice differentiable with respect to θ, and $\widehat{\theta}$ exists and is strongly consistent. Further assume:*

B1. Both $D\delta(X_1, ..., X_m, \theta)$ and $D^2\delta(X_1, ..., X_m, \theta)$ have second order moments for every $\theta \in B$.

B2. $G(0, F_n)$ is not a degenerate U-statistic, i.e. Σ is positive definite.

Then

$$\widehat{\theta} = -S^{-1}(0, F)G(0, F_n) + O(a_n), \quad w.p.1, \tag{21}$$

and

$$n^{\frac{1}{2}} S(0, F)\widehat{\theta} \to N(0, m^2\Sigma), \quad \text{in dist., as } n \to \infty. \tag{22}$$

Proof. By Taylor's theorem and condition B1

$$G(0, F_n) = G(0, F_n) - G(\widehat{\theta}, F_n) \tag{23}$$

$$= -DG(0, F_n)\widehat{\theta} + O(||\widehat{\theta}||^2), \quad w.p.1.$$

By the law of iterated logarithm for U-statistics, we have

$$G(0, F_n) = O(a_n^{\frac{1}{2}}), \quad w.p.1, \tag{24}$$

$$DG(0, F_n) = S(0, F) + O\left(a_n^{\frac{1}{2}}\right), \quad w.p.1. \tag{25}$$

Therefore

$$\widehat{\theta} = -S^{-1}(0, F)G(0, F_n) + O(a_n^{\frac{1}{2}}||\widehat{\theta}||) + O\left(||\widehat{\theta}||^2\right), \quad w.p.1. \tag{26}$$

(21) is proved. Since $G(0, F_n)$ is a non-degenerated U-statistic, by Hoeffding's (1948) central limit theorem for U-statistics, (22) is true.

The following results are on smoothed estimates, i.e. those obtained by a smoothed DISM.

Theorem 3 *Assume $\widehat{\theta}$ exists and is strongly consistent and*

C1. There exist $\alpha > 0$ such that

$$cov\left(D^2\delta_n(X_1, .., X_m, 0)\right) = O(h_n^{-\alpha}), \ therefore \tag{27}$$

$$cov\left(D^3\delta_n(X_1, ..., X_m, 0)\right) = O(h_n^{-\alpha-2}).$$

C2. There is an integer $l > 2$ such that C4 holds, $\Delta(\theta, F)$ is l times continuously differentiable, and $\lim_{n\to\infty} \Sigma_n = \Sigma$ exists and is positive definite.

C3. $k(u)$ has a compact support, is twice differentiable, and there is integer $r \geq 1$ such that C4 holds and

$$\int_{R^q} u_1^{l_1} ... u_m^{l_m} k(u) du = 0, \quad for\ l_j \geq 0\ and\ 0 < \sum_{j=1}^{m} l_j \leq r \leq l - 2, \tag{28}$$

where $u = (u_1, ...u_m)$.

C4. Choose $h_n = O\left(a_n^{\frac{1}{2r+2}}\right)$..

Then

$$\widehat{\theta} = -S^{-1}(0,F)G_n(0,F_n) + O\left(a_n^{\frac{4r+\alpha}{4r+2\alpha}}\right), \quad w.p.1, \tag{29}$$

and

$$n^{\frac{1}{2}}S(0,F)\left(\widehat{\theta} - E\left(D\delta_n(X_1,...,X_k,0)\right)\right) \to N(0,k^2\Sigma), \text{ in dist. as } n \to \infty. \tag{30}$$

Remark 2 *1. In condition C1, the value of* α *depends on the smoothness of the underlying density and the DISM. In fact* $0 \leq \alpha \leq 4$.

2. In the worst case where $\alpha = 4$, *(29) becomes*

$$\widehat{\theta} = -S^{-1}(0,F)G_n(0,F_n) + o\left(a_n^{\frac{r+1}{r+4}}\right), \ w.p.1. \tag{31}$$

The remainder can be as close to $O(a_n)$ *as possible if* r *is large enough. In the best case where* $\alpha = 0$, *(29) becomes*

$$\widehat{\theta} = -S^{-1}(0,F)G_n(0,F_n) + O(a_n), \ w.p.1. \tag{32}$$

3. In fact

$$EG_n(0,F_n) = \int_{R^q} D\Delta(-h_n u, F)k(u)du = O(h_n^{r+1}). \tag{33}$$

Condition C4 implies

$$\lim_{n\to\infty} n^{\frac{1}{2}}EG_n(0,F_n) = 0. \tag{34}$$

If we replace condition C4 with $\lim_{n\to\infty} n^{1/2}h_n^{r+1}$ *exists and is greater than* 0, *then* $\widehat{\theta}$ *converges to a non-central normal distribution.*

Proof. By Taylor's theorem, we have

$$G_n(0,F_n) = G_n(0,F_n) - G_n(\widehat{\theta},F_n) \tag{35}$$

$$= -D'G_n(0,F_n)\widehat{\theta} + O\left(||\widehat{\theta}||^2 \left\|D^3\Delta_n(0,F_n)\right\|\right)$$

By condition C1 and the law of iterated logarithms for U-statistics, we have

$$D'G_n(0,F_n) = D'G_n(0,F) + O\left(h_n^{-\frac{\alpha}{2}}a_n^{\frac{1}{2}}\right), \quad w.p.1. \tag{36}$$

and

$$D^3\Delta_n(0,F_n) = D^3\Delta_n(0,F) + O\left(h_n^{-\frac{\alpha}{2}-1}a_n^{\frac{1}{2}}\right), \quad w.p.1. \tag{37}$$

Further, by condition C3

$$D'G_n(0,F) = ED^2\int_{R^q}\delta(X_1,...,X_k,\theta-h_nu)k(u)du|_{\theta=0} \tag{38}$$

$$= D^2\int_{R^q}\Delta(\theta-h_nu,F)k(u)du|_{\theta=0} = \int_{R^q}D^2\Delta(-h_nu,F)k(u)du$$

$$= S(0,F) + o(h_n^r).$$

Consequently

$$G_n(0,F_n) = -S(0,F)\widehat{\theta} + o(h_n^r||\widehat{\theta}||) + O\left(h_n^{-\frac{\alpha}{2}}a_n^{\frac{1}{2}}||\widehat{\theta}||\right) \tag{39}$$

$$+O(||\widehat{\theta}||^2) + O\left(h_n^{-\frac{\alpha}{2}-1} a_n^{\frac{1}{2}} ||\widehat{\theta}||^2\right), \quad w.p.1.$$

By the law of iterated logarithms, we have

$$S(0,F)\widehat{\theta} = -EG_n(0,F_n) + O\left(a_n^{\frac{1}{2}}\right), \ w.p.1. \tag{40}$$

On the other hand

$$EG_n(0,F_n) = D\int_{R^q} E\delta(X_1,...,X_m,\theta - h_n u)k(u)du|_{\theta=0} \tag{41}$$

$$= \int_{R^q} D\Delta(-h_n u, F)k(u)du = O(h_n^{r+1}).$$

Therefore

$$\widehat{\theta} = O(h_n^{r+1}) + O\left(a_n^{\frac{1}{2}}\right) = O(a_n^{\frac{1}{2}}), \quad w.p.1. \tag{42}$$

(29) is proved. By central limit theorem of U-statistics, we have

$$n^{\frac{1}{2}}\Sigma_n^{-\frac{1}{2}}\left(G_n(0,F_n) - EG_n(0,F_n)\right) \to N(0, m^2 I_{q\times q}), \text{ in dist.} \tag{43}$$

(30) is proved.

4 Applications

We apply the asymptotic theory of μ-estimation established in previous sections to the median where $p = 1$, $m = 1$, and to the median of Liu (1994) using a smoothed version of the DISM.

4.1 Smoothed median estimator

It is well known that the L_1-norm estimator that minimizes

$$\sum_{i=1}^{n} |X_i - \theta| \tag{44}$$

is a sample median. Asymptotic properties of a sample median have been well studied. We use this example to demonstrate how well the kernel smoothing method performs. In this case, the observations are *i.i.d.* univariate random variables and $m = 1$. Without loss of generality, let us assume that the median is 0. Note that the smoothed discrepancy function becomes

$$\delta_n(X, \theta) = \int_R |X - \theta + h_n u| k(u) du. \tag{45}$$

Denote $K(x) = \int_{-\infty}^{x} k(x)dx$. Then we have

$$D\delta_n(X, \theta) = 2K\left(\frac{\theta - X}{h_n}\right) - 1, \tag{46}$$

and

$$G_n(\theta, F_n) = \sum_{i=1}^{n} \left(2K\left(\frac{\theta - X_i}{h_n}\right) - 1\right). \tag{47}$$

Theorem 4 *Assume that*

D1. $f(0) > 0$ *and* $f(x)$ *is differentiable in a neighborhood of* 0.

D2. $k(u)$ *is twice differentiable and has a compact support, and*

$$\int_R uk(u)du = 0. \tag{48}$$

D3. Choose the bandwidth as

$$h_n = O\left(n^{-\frac{1}{3}}(\log\log n)^{\frac{1}{3}}\right). \tag{49}$$

Then

1. $\widehat{\theta}$ is strongly consistent, i.e.

$$\lim_{n\to\infty}\widehat{\theta} = 0,\ w.p.1. \tag{50}$$

2. $\widehat{\theta}$ is asymptotically normal, i.e.

$$2f(0)n^{\frac{1}{2}}\widehat{\theta} \to N(0,1),\ in\ dist.,\ as\ n\to\infty. \tag{51}$$

3. $\widehat{\theta}$ can be represented as

$$\widehat{\theta} = \frac{1}{2f(0)n}\sum_{i=1}^{n}\left(1-2K\left(\frac{-X_i}{h_n}\right)\right) + O\left(n^{-\frac{5}{6}}(\log\log n)^{\frac{5}{6}}\right),\ w.p.1. \tag{52}$$

Proof: We need only to verify conditions in previous theorems. We first verify condition A4 in theorem 1. Other conditions are easy to be verified. Note that

$$\sup|G_n(\theta,F_n) - G_n(\theta,F)| \tag{53}$$

$$= 2\sup\left|\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{\theta-X_i}{h_n}\right) - \int_R K\left(\frac{\theta-x}{h_n}\right)f(x)dx\right|$$

$$= 2\sup\left|\int_R \left(F_n(\theta-h_nx) - F(\theta-h_nx)\right)k(x)dx\right|$$

$$\le C\sup_x|F_n(x)-F(x)| \le Cn^{-\frac{1}{2}},\ w.p.1.$$

The last inequality is from Dvoretzky-Kiefer-Wolfowitz theorem (1956) on Kolmogorov-Simirnov distance (Serfling, 1980). By Theorem 1, (50) is true. Note that

$$S(0,F) = \frac{d^2}{d\theta^2} E|X_1 - \theta| \,|_{\theta=0} \tag{54}$$

$$= \frac{d^2}{d\theta^2} \left(\int_{x>\theta} (x-\theta) f(x) dx - \int_{x<\theta} (x-\theta) f(x) dx \right) |_{\theta=0}$$

$$= \frac{d}{d\theta} \left(2F(\theta) - 1 \right) |_{\theta=0} = 2f(0),$$

and

$$\Sigma_n = var\left(D\delta_n(X_1, 0) \right) = 4var\left(K\left(\frac{-X_1}{h_n} \right) \right) \tag{55}$$

$$= 4\int_R K^2\left(\frac{-x}{h_n} \right) f(x) dx - 4\left(\int_R K\left(\frac{-x}{h_n} \right) f(x) dx \right)^2$$

$$= 8\int_R F(-h_n u) K(u) k(u) du - 4\left(\int_R F(-h_n u) k(u) du \right)^2$$

$$\to 4F(0) \int_R 2K(u)k(u)du - 4F^2(0) = 1, \text{ as } n \to \infty.$$

All other conditions in Theorem 3 are satisfied and note that $\alpha = 1$, and $r = 1$. Then (51) and (52) follow theorem 3.

Remark 3 *1. Bahadur (1966) and Kiefer (1967) found an asymptotic representation of a sample quantile. The Bahadur-Kiefer representation of a sample median is given as*

$$\xi_n = \xi + \frac{1}{2f(\xi)} \left(1 - 2F_n(\xi) \right) + O\left(n^{-\frac{3}{4}} (\log\log n)^{\frac{3}{4}} \right), \ w.p.1. \tag{56}$$

The remainder rate cannot be improved. However, under the same conditions as those of Bahadur-Kiefer representation, a smoothed median estimator can have a representation with a remainder rate $n^{-5/6}(\log\log n)^{5/6}$. *If* $f(x)$ *is* r *times differentiable, the remainder rate can be* $n^{-(4r+1)/(4r+2)}(\log\log n)^{(4r+1)/(4r+2)}$.

2. The smoothed median estimate provides an alternative computing method for L_1*-norm locations and linear models. We will carry on comparison study on computation of* L_1*-norm for linear programing based methods and the smoothed estimation method in future communication.*

4.2 Simplicial median estimator

By the notion of data depth (Liu, 1990), the deepest point which is called simplicial median of a multivariate distribution is a generalization of the median to a multivariate distribution. Estimating of a simplicial median falls within the μ-estimation framework. Let $A(x) = A(x_1, ..., x_{p+1})$ be a closed simplex in R^p with vertices $x_1, ..., x_{p+1}$, and define

$$I_B(y) = \begin{cases} 1 & if \quad y \in B \\ 0 & if \quad y \notin B \end{cases}. \tag{57}$$

A simplicial median DISM is

$$\delta(x, \theta) = -I_{A(x)}(\theta). \tag{58}$$

Applying kernel smooth method, we have a smoothed DISM.

$$\delta_n(x,\theta) = \int_{R^p} -I_{A(x)}(\theta - h_n u)k(u)du. \tag{59}$$

Theorem 5 *Assume:*

E1. $F(x)$ *has a unique simplicial median which is assumed to be 0 for convenience.* $f(x)$ *is differentiable, and* $S(0,F)$ *is positive definite.*

E2. $k(u)$ *has a compact support, twice differentiable and*

$$\int_{R^p} uk(u)du = 0 \tag{60}$$

E3. $h_n = O\left(n^{-\frac{1}{4}}\left(\log\log n\right)^{\frac{1}{4}}\right)$.

Then:

1. $\widehat{\theta}$ *is strongly consistent, i.e.*

$$\lim_{n\to\infty} \widehat{\theta} = 0, \quad w.p.1. \tag{61}$$

2. $\widehat{\theta}$ *can be represented as*

$$\widehat{\theta} = S^{-1}(0,F)G_n(0,F_n) + O\left(n^{-\frac{3}{4}}\left(\log\log n\right)^{\frac{3}{4}}\right), \quad w.p.1. \tag{62}$$

3. $\widehat{\theta}$ *is asymptotically normal, i.e.*

$$n^{\frac{1}{2}}S(0,F)\widehat{\theta} \to N\left(0,(p+1)^2\Sigma\right), \quad in\ dist., \quad n\to\infty, \tag{63}$$

where Σ *is the variance-covariance matrix of* $D(E(\delta(X,0)|X_1)$

Remark 4 *The asymptotic distribution we obtained is the same as Arcones et al's (1994) result. Our conditions are much milder, simpler, and easy to verify.*

Proof. We need to verify condition A1 in Theorem 1 and find α and r for the smoothed simplicial median estimator. By the conditions in the theorem, we know that $r = 1$. Note that

$$E\left(D^2_{i_1,j_1}\delta_n(X,0)D^2_{i_2,j_2}\delta_n(X,0)\right) \tag{64}$$

$$= h_n^{-4}\int_{R^p}\int_{R^p}\Pr\left(-h_nu,-h_nv\in A(X)\right)D^2_{i_1,j_1}k(u)D^2_{i_2,j_2}k(v)dudv = O(h_n^{-2}),$$

and

$$E\left(D^3_{i_1,j_1,l_1}\delta_n(X,0)D^3_{i_2,j_2,l_2}\delta_n(X,0)\right) \tag{65}$$

$$= h_n^{-6}\int_{R^p}\int_{R^p}\Pr\left(-h_nu,-h_nv\in A(X)\right)D^3_{i_1,j_1,l_1}k(u)D^3_{i_2,j_2,l_2}k(v)dudv = O(h_n^{-4}).$$

Therefore $\alpha = 2$. It is easy to show that

$$\lim_{n\to\infty}\sup_x|F_n(x)-F(x)| = 0, \quad w.p.1, \tag{66}$$

as in the multivariate case. It is seen that

$$\sup\|G_n(\theta,F_n)-G_n(\theta,F)\| \tag{67}$$

$$\le C\sup\left\|\int I\left(\theta\in A(X)\right)d\underbrace{(F_n\times\ldots\times F_n)}_{p+1}-\int I\left(\theta\in A(X)\right)d\underbrace{(F\times\ldots\times F)}_{p+1}\right\|$$

$$\to 0, \quad w.p.1, \quad n\to\infty.$$

Other conditions in Theorem 3 are easy to verify. This theorem is proved.

References

Arcones, M. A, Chen, Z. and Gine, E (1994). Estimations related to U-processes with application to multivariate median: Asymptotic normality. *Ann. Statist.*, 22, 1460-1477.

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

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function of the multinomial estimator. *Ann. Math. Statist.*, 27, 642-669.

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

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, 43, 1449-1458.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.*, 38, 1323-1342.

Liu, R. Y. (1990). On a notion of data depth based on random simplicies. *Ann. Statist.*, 18, 405-414.

Oja, H. (1983). Descriptive statistics for multivariate distribution.. *Statist. and Prob. Letters,* 1, 327-333.

Serfling, R. J. (1980). *Approximation theorems for mathematical statistics.* Wiley, New York