The Annals of Statistics
1992, Vol. 20, No. 3, 1534–1547

THE METRICALLY TRIMMED MEAN AS A ROBUST ESTIMATOR OF LOCATION

BY SEONG-JU KIM

Sung Kyun Kwan University and University of California, Berkeley

The metrically trimmed mean is defined as the average of observations remaining after a fixed number of outlying observations have been removed. A metric, the distance from the median, is used to determine which points are outlying. The influence curve and the asymptotic normality of the metrically trimmed mean are derived using von Mises expansions. The relative merits of the median, the trimmed mean and the metrically trimmed mean are discussed in neighborhoods of nonparametric models with natural parameters. It is observed that the metrically trimmed mean works well for the center of symmetry of a symmetric distribution function with asymmetric contamination. A multivariate extension of the metrically trimmed mean is discussed.

1. Introduction. Let $X_1, \dots, X_n$ be a random sample from a population with distribution function F. We will evaluate the performance of various estimates of the model $F = (1 - \varepsilon)G + \varepsilon H$, where G is an unknown symmetric distribution function, H is an unknown contaminating distribution function and ε is a contamination proportion. This model was briefly discussed by Bickel and Lehmann (1975) as neighborhoods of nonparametric models with natural parameters. In this model we are interested in estimating the center of symmetry of G when our data are contaminated by $100\varepsilon\%$ observations from H. As an alternative to the trimmed mean by Tukey, we will consider the metrically trimmed mean which seems capable of extension to the multivariate case.

In Section 2 the metrically trimmed mean is defined as the average of observations remaining after a fixed number of outlying observations have been removed. A metric, the distance from the median, is used to determine which points are outlying. The influence curve and the asymptotic normality of the metrically trimmed mean are derived using von Mises expansions in Section 3.

In Section 4 the influence curve and the breakdown point of the metrically trimmed mean are discussed. The metrically trimmed mean is compared with the median and the trimmed mean in terms of the asymptotic bias, variance and mean square error (MSE) in neighborhoods of nonparametric models with natural parameters. Monte Carlo results are presented for the same distribution functions. In Section 5 the asymptotic relative efficiency (ARE) of the metrically trimmed mean to the median and this same to the trimmed mean

Received October 1990; revised July 1991.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62F35.

Key words and phrases. Asymptotic normality, influence curve, metrically trimmed mean, neighborhoods of nonparametric models with natural parameters, von Mises expansions.

are discussed for symmetric distribution functions. Also the lower bounds are derived for all symmetric unimodal distribution functions.

In Section 6 the metrically trimmed mean is extended to the multivariate case preserving affine equivariance. To determine if a point is outlying, we use, as our distance of the point from the generalized median, the average volume of all p-dimensional simplexes formed by $p-1$ of the sample points, the point itself and the generalized median.

2. Definition of the metrically trimmed mean. Suppose we observe a random sample $X_1, \dots, X_n$ from a population with distribution function F. We will denote the empirical distribution function of $X_1, \dots, X_n$ by F_n, the sample median by $\tilde{X}$ and the empirical distribution function of $|X_1 - \tilde{X}|, \dots, |X_n - \tilde{X}|$ by K_n. The qth quantile of F is defined as $F^{-1}(q) = \inf\{x;\, F(x) \geq q\}$. Finally $1\{\cdot\}$ denotes the indicator function and $[\cdot]$ denotes the greatest integer function.

For $0 \leq \alpha < 1$, the α-metrically trimmed mean from the median is defined as the average of the $n - [\alpha n]$ observations remaining after the $[\alpha n]$ observations with the largest distance from the median have been removed. We will call it, for simplicity, the metrically trimmed mean and denote it by $\tilde{X}_\alpha$. That is,

$$
\begin{aligned}
\tilde{X}_\alpha &= \frac{1}{n-[\alpha n]} \sum_{i=1}^{n} X_i 1\left\{\left|X_i - \tilde{X}\right| \leq K_n^{-1}(1-\alpha)\right\} \\
&= \frac{1}{n-[\alpha n]} \sum_{i=1}^{n} X_i 1\left\{\tilde{X} - K_n^{-1}(1-\alpha) \leq X_i \leq \tilde{X} + K_n^{-1}(1-\alpha)\right\}.
\end{aligned} \tag{2.1}
$$

For an alternative functional representation of $\tilde{X}_\alpha$ in terms of F_n, K_n is expressed as

$$
K_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1\left\{\left|X_i - \tilde{X}\right| \leq x\right\} = F_n(\tilde{X} + x) - F_n^-(\tilde{X} - x), \tag{2.2}
$$

where F_n^- denotes the left limit of F_n. Now we will define two functionals $\xi(F)$ and $\lambda(F)$. The dependence on F will be suppressed for notational convenience if it is clear in the context. Let $\xi(F)$ be the median functional defined as

$$
\xi(F) = F^{-1}(1/2). \tag{2.3}
$$

Let $\lambda(F)$ be the functional defined implicitly as the root of

$$
F(\xi + \lambda) - F(\xi - \lambda) = 1 - \alpha. \tag{2.4}
$$

It is noticed that $K_n^{-1}(1-\alpha)$ in (2.1) is asymptotically equivalent to the functional $\lambda(F_n)$ from (2.2) and (2.4). Therefore the metrically trimmed mean $\tilde{X}_\alpha$ is asymptotically equivalent to the functional $T(F_n)$, where

$$
T(F) = (1-\alpha)^{-1} \int_{\xi-\lambda}^{\xi+\lambda} x\, dF(x). \tag{2.5}
$$

To understand the metrically trimmed mean, we will compare it with the trimmed mean by Tukey when the same number of observations is removed. We will denote it by $\bar{X}_\alpha$ when $[\alpha n/2]$ observations are removed from each side. Then $\bar{X}_\alpha$ is asymptotically equivalent to the functional $T^*(F_n)$, where

$$T^*(F) = (1-\alpha)^{-1} \int_{F^{-1}(\alpha/2)}^{F^{-1}(1-\alpha/2)} x \, dF(x). \tag{2.6}$$

Comparing (2.5) with (2.6), it is noticed that $T(F) = T^*(F)$ for a symmetric distribution function F. That is, $\tilde{X}_\alpha$ and $\bar{X}_\alpha$ are estimating the same quantity, namely the center of symmetry of a symmetric distribution function F. However, $T(F) \neq T^*(F)$ for an asymmetric distribution function F.

As an example, suppose $\alpha = 0.1$ and $F = 0.9N(0,1) + 0.1N(4,9)$, where $N(\mu, \sigma^2)$ denotes the normal distribution function with mean μ and variance σ^2. Here we are estimating 0, which is the center of symmetry of $N(0,1)$. The trimmed mean functional $T^*(F)$ chooses two trimming points $F^{-1}(0.05) = -1.624$ and $F^{-1}(0.95) = 4.002$ so that $T^*(F) = 0.21$. The metrically trimmed mean functional $T(F)$ finds the interval with midpoint equal to the median, defined by (2.4), such that 90% of the total mass of distribution function F is contained in this interval. We evaluate $\xi = 0.112$, $\lambda = 2.192$ so that $T(F) = 0.04$. In this example, the metrically trimmed mean functional $T(F)$ is closer to the center of symmetry than the trimmed mean functional $T^*(F)$ as well as the median functional $\xi(F)$ in (2.3).

Of course, if F was known to be $0.9N(0,1) + 0.1N(4,9)$, then an asymmetrically trimmed mean should be used instead of the trimmed mean. What is important is that the metrically trimmed mean works well for the center of symmetry of a symmetric distribution function with symmetric contamination as well as asymmetric contamination.

If we replace $K_n^{-1}(1-\alpha)$ in (2.1) by some constant, then the metrically trimmed mean defined by (2.1) is an estimator proposed by Bickel (1965). It was a difficult problem in the 1960s to study the properties of the metrically trimmed mean without the results on von Mises expansions. There are some good reasons for using $K_n^{-1}(1-\alpha)$ instead of some constant. From the theoretical point of view, we can get affine equivariance. This fact should not be passed over lightly if we recall how many complications arise for Huber's estimator [Huber (1964)] to be affine equivariant when the scale is unknown. All the more, we can exploit results on the empirical distribution function. From the practical point of view, to decide how many observations are to be removed has more intuitive interpretation than to choose some constant.

3. Influence curve and asymptotic normality. We will derive the influence curve and the asymptotic normality of the metrically trimmed mean.

PROPOSITION 3.1. *Let F possess a density f which is absolutely continuous and positive on its support. Then the influence curve of the metrically trimmed* *mean*, $T(F_n)$, *is*

$$\text{(3.1)}\qquad \mathrm{IC}(x;F,T) = (1-\alpha)^{-1}\begin{cases} C_1(F,\alpha) - C_2(F,\alpha) - C_3(F,\alpha), & \text{if } x < \xi - \lambda, \\ x - C_2(F,\alpha) - C_3(F,\alpha), & \text{if } \xi - \lambda < x < \xi, \\ x + C_2(F,\alpha) - C_3(F,\alpha), & \text{if } \xi < x < \xi + \lambda, \\ C_1(F,\alpha) + C_2(F,\alpha) - C_3(F,\alpha), & \text{if } x > \xi + \lambda, \end{cases}$$

where $C_1(F,\alpha)$, $C_2(F,\alpha)$ *and* $C_3(F,\alpha)$ *are defined in* (3.9), (3.10) *and* (3.11) *respectively. Moreover,*

$$\sqrt{n}\,(T(F_n) - T(F)) \to N(0, \tilde{\sigma}_\alpha^2) \quad \text{in distribution},$$

where

$$\text{(3.2)}\qquad \tilde{\sigma}_\alpha^2 = E(\mathrm{IC}(X;F,T))^2.$$

PROOF. (i) Influence curve. Let $F_t = F + t(\delta_x - F)$, where δ_x denotes point mass 1 at x. From (2.5), we have

$$\text{(3.3)}\qquad (1-\alpha)T(F_t) = \int x\,dF + t\int x\,d(\delta_x - F),$$

where the integrations are taken from $\xi(F_t) - \lambda(F_t)$ to $\xi(F_t) + \lambda(F_t)$,

$$\text{(3.4)}\qquad \xi(F_t) = F_t^{-1}(1/2)$$

and $\lambda(F_t)$ satisfies

$$\text{(3.5)}\qquad F_t(\xi(F_t) + \lambda(F_t)) - F_t(\xi(F_t) - \lambda(F_t)) = 1 - \alpha.$$

Since

$$\mathrm{IC}(x;F,T) = \frac{d}{dt}T(F_t)\bigg|_{t=0},$$

we have from (3.3)

$$\text{(3.6)}\qquad \begin{aligned}(1-\alpha)\,\mathrm{IC}(x;F,T) &= (\xi+\lambda)\,f(\xi+\lambda)(\xi'+\lambda') \\ &\quad -(\xi-\lambda)\,f(\xi-\lambda)(\xi'-\lambda') + \int_{\xi-\lambda}^{\xi+\lambda} x\,d(\delta_x - F),\end{aligned}$$

where ξ' and λ' denote $(d/dt)\xi(F_t)|_{t=0}$ and $(d/dt)\lambda(F_t)|_{t=0}$ respectively. Similarly, we have from (3.4) and (3.5)

$$\text{(3.7)}\qquad \xi' = (1\{x > \xi\} - 1/2)/f(\xi)$$

and

$$(3.8)\quad \lambda' = \frac{(1-\alpha) - \xi'(f(\xi+\lambda) - f(\xi-\lambda)) - 1\{\xi - \lambda < x < \xi + \lambda\}}{f(\xi-\lambda) + f(\xi+\lambda)}.$$

Substituting (3.7) and (3.8) in (3.6), the proof is complete by defining $C_1(F, \alpha)$, $C_2(F, \alpha)$ and $C_3(F, \alpha)$ as follows:

$$(3.9)\qquad C_1(F,\alpha) = \xi + \lambda \frac{f(\xi+\lambda) - f(\xi-\lambda)}{f(\xi+\lambda) + f(\xi-\lambda)},$$

$$(3.10)\qquad C_2(F,\alpha) = \lambda \frac{f(\xi+\lambda)}{f(\xi)} \cdot \frac{2f(\xi-\lambda)}{f(\xi+\lambda) + f(\xi-\lambda)}$$

and

$$(3.11)\qquad C_3(F,\alpha) = \alpha C_1(F,\alpha) + \int_{\xi-\lambda}^{\xi+\lambda} x\,dF.$$

(ii) Asymptotic normality. Since $X_1, \dots, X_n$ are iid random variables from a distribution function F, then $F(X_1), \dots, F(X_n)$ are iid random variables from the uniform distribution function U on [0, 1]. If we denote by U_n the empirical distribution function corresponding to $F(X_1), \dots, F(X_n)$, it follows that $F_n = U_n \circ F$. Here "$\circ$" is used to denote a composition of two functions.

Let Ω denote the space of right-continuous functions with left limits on [0, 1]. Then the metrically trimmed mean functional T induces a functional τ on Ω by

$$\tau(U_n) = T(U_n \circ F) = T(F_n)$$

and

$$\tau(U) = T(U \circ F) = T(F).$$

To prove asymptotic normality of $T(F_n)$, we will show that the functional τ defined by

$$\tau(D) = T(D \circ F) \quad \text{for } D \in \Omega$$

is Hadamard differentiable at the uniform distribution function U and apply Fernholz (1983), Theorem 4.4.2.

First, we notice from (2.4) that the functional λ induces a functional κ on Ω defined as a root $\kappa(D) = \theta$ of

$$\gamma_2(D,\theta) - \gamma_1(D,\theta) = 1 - \alpha \quad \text{for } D \in \Omega,$$

where

$$\gamma_1(D,\theta) = D \circ F(F^{-1} \circ D^{-1}(1/2) - \theta)$$

and

$$\gamma_2(D,\theta) = D \circ F(F^{-1} \circ D^{-1}(1/2) + \theta).$$

Thus the functional T defined by (2.5) induces a functional τ on Ω as

$$\tau(D) = (1-\alpha)^{-1}\int_{\psi_1(D)}^{\psi_2(D)} F^{-1}\circ D^{-1}(x)\,dx, \tag{3.12}$$

where

$$\psi_j(D) = \gamma_j(D, \kappa(D)) \quad \text{for } j = 1, 2.$$

From the chain rule and Fernholz (1983), Proposition 6.1.8, $\tau(D)$ is Hadamard differentiable at U provided:

1. $F^{-1}\circ D^{-1}$ is Hadamard differentiable at U,
2. ψ_j is Hadamard differentiable at U,
3. $F^{-1}\circ D^{-1}$ is continuous at $\psi_j(U)$.

Here we recall that the trimmed mean functional T^* defined by (2.6) induces a functional τ^* on Ω as

$$\tau^*(D) = (1-\alpha)^{-1}\int_{\alpha/2}^{1-\alpha/2} F^{-1}\circ D^{-1}(x)\,dx. \tag{3.13}$$

Also the so-called Hodges–Lehmann estimator induces a functional τ_H on Ω defined as a root $\tau_H(D) = \theta$ of

$$\int_0^1 D\circ F(2\theta - F^{-1}\circ D^{-1}(x))\,dx = 1/2. \tag{3.14}$$

It is noticed from (3.12) that the integrand, $F^{-1}\circ D^{-1}$, is the same as the integrand of (3.13) and the integration limit ψ_j is the same form of integrand as in (3.14). Therefore both $F^{-1}\circ D^{-1}$ and ψ_j are Hadamard differentiable at U with modification of D and truncation of F as discussed by Fernholz (1983). With these modifications of D, then $F^{-1}\circ D^{-1}$ is continuous at $\psi_j(U)$. Therefore $\tau(D)$ is Hadamard differentiable at U and the proof is complete. □

REMARK. If F is a symmetric distribution function about 0, then the influence curve of the metrically trimmed mean in (3.1) simplifies to

$$\mathrm{IC}(x; F, T) = (1-\alpha)^{-1}\begin{cases} -C(F,\alpha), & \text{if } x < F^{-1}(\alpha/2),\\ x - C(F,\alpha), & \text{if } F^{-1}(\alpha/2) < x < 0,\\ x + C(F,\alpha), & \text{if } 0 < x < F^{-1}(1-\alpha/2),\\ C(F,\alpha), & \text{if } x > F^{-1}(1-\alpha/2),\end{cases}$$

where

$$C(F,\alpha) = F^{-1}(1-\alpha/2)\,f(F^{-1}(1-\alpha/2))/f(0).$$

4. Comparisons. In this section we discuss the relative merits of the median $\tilde{X}$, the trimmed mean $\overline{X}_\alpha$ and the metrically trimmed mean $\tilde{X}_\alpha$.

The breakdown points of $\tilde{X}$ and $\overline{X}_\alpha$ are well known to be $1/2$ and $\alpha/2$ respectively. It is easy to see that the breakdown point of $\tilde{X}_\alpha$ is α. That is, the breakdown point of $\tilde{X}_\alpha$ is double that of $\overline{X}_\alpha$ when the same number of observations is removed.

The influence curves of $\tilde{X}_{0.1}$ are shown in Figure 1 for some distribution functions with those of $\tilde{X}$ and $\overline{X}_{0.1}$. The upper figures refer to symmetric distribution functions and the lower figures refer to asymmetric distribution functions. It is observed that the influence curve of $\tilde{X}_\alpha$ has jumps at $\xi - \lambda$, ξ and $\xi + \lambda$. Also it has relatively small absolute value at $\pm\infty$. It is surprising to observe that the influence curve of $\tilde{X}_\alpha$ has positive value at $-\infty$ in the lower right figure.

The discontinuity of the influence curve of the metrically trimmed mean can be relieved by considering smooth weighting instead of trimming. It is noticed that the metrically trimmed mean is the weighted average where the weight W_i is 1 if $|X_i - \tilde{X}|$ is small and 0 otherwise. As an alternative to the 1 and 0 weight, the weight W_i may be given inversely proportional to $|X_i - \tilde{X}|$ or a studentized version of $|X_i - \tilde{X}|$. A successful example of weighted averages is the biweighted mean discussed by Mosteller and Tukey (1977).

To compare the three estimates in terms of the bias, variance and MSE, we consider neighborhoods of nonparametric models with natural parameters where $F = (1 - \varepsilon)G + \varepsilon H$. Without loss of generality, the center of symmetry of G is set at 0. For given F and α, we evaluate the asymptotic bias and the asymptotic variance of $\tilde{X}_\alpha$ as $T(F)$ in (2.5) and $\tilde{\sigma}_\alpha^2$ in (3.2) respectively. In view of the asymptotic bias and variance, the question again arises of the applicability of the results to finite sample size. For this reason, we present Monte Carlo results with sample size $n = 20$ for the same distribution functions.

In this simulation study, $\tilde{X}$, $\overline{X}_\alpha$ and $\tilde{X}_\alpha$ are computed for 20 pseudo-random variates from F and this process is replicated 3000 times. For $\tilde{X}_\alpha$, let AVE denote the average of 3000 realizations of $\tilde{X}_\alpha$ and let VAR denote their variance multiplied by the sample size. We then evaluate the Monte Carlo bias and the Monte Carlo variance of $\tilde{X}_\alpha$ as AVE and VAR respectively. Finally, the asymptotic MSE and the Monte Carlo MSE of $\tilde{X}_\alpha$ are, normalized by the sample size, given below. These remarks apply in a similar way to $\tilde{X}$ and $\overline{X}_\alpha$:

$$\text{Asymptotic MSE} = 20(T(F))^2 + \tilde{\sigma}_\alpha^2,$$

$$\text{Monte Carlo MSE} = 20\text{AVE}^2 + \text{VAR}.$$

Table 1 represents the asymptotic results and Table 2 represents the Monte Carlo results corresponding to Table 1. From Table 1 it is noticed that the asymptotic bias of $\tilde{X}_\alpha$ is smaller than that of $\overline{X}_\alpha$. The difference between the two biases grows fast as the center of H gets farther from the center of G. The asymptotic bias of $\tilde{X}_\alpha$ is smaller than that of $\tilde{X}$ if $\alpha \geq \varepsilon$. As the center of H gets farther from the center of G, the asymptotic variance of $\tilde{X}$ grows slow, the asymptotic variance of $\overline{X}_\alpha$ grows fast and the asymptotic variance of $\tilde{X}_\alpha$ fluctuates depending on G and H. The asymptotic MSE of $\tilde{X}_\alpha$ gets smaller than that of $\overline{X}_\alpha$ as the center of H gets farther from the center of G or as G

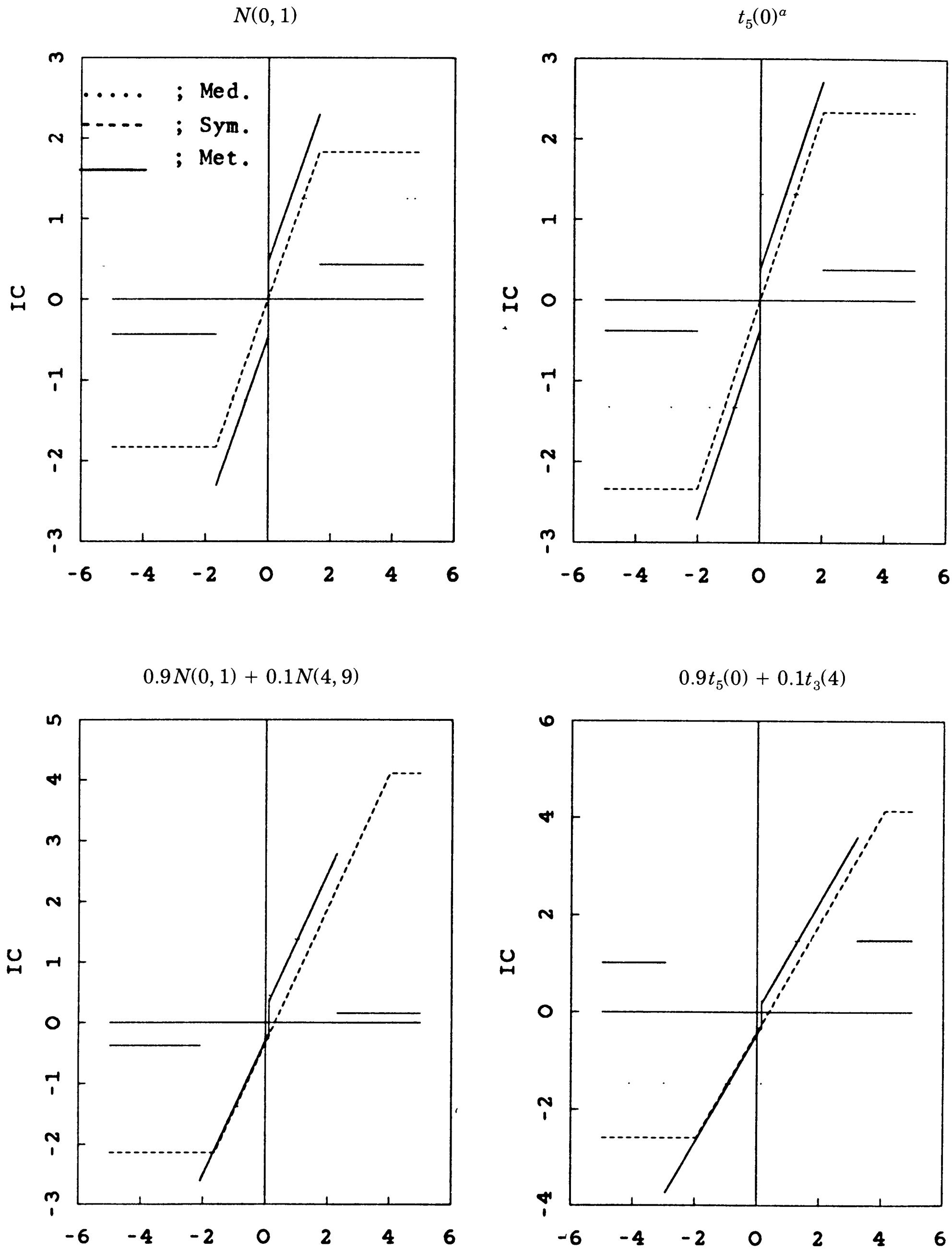

FIG. 1. *The influence curve of* $\tilde{X}$ *(dotted line),* $\overline{X}_{0.1}$ *(broken line) and* $\tilde{X}_{0.1}$ *(solid line).* ${}^a t_\nu(\phi)$ *denotes the t-distribution with* ν *degrees of freedom centered at* ϕ.

TABLE 1

The asymptotic bias, variance and MSE[a] of $\tilde{X}$, $\bar{X}_\alpha$ and $\tilde{X}_\alpha$ for distribution functions in neighborhoods of nonparametric models with natural parameters

		$0.9N(0,1) + 0.1N(\mu,\sigma^2)$					**$0.8N(0,1) + 0.2N(\mu,\sigma^2)$**				
		$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$	$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$
	BIAS	0.13	0.17	0.10	0.16	0.10	0.29	0.37	0.26	0.35	0.23
$\mu = 2, \sigma = 1$	VAR	1.90	1.36	1.78	1.37	2.10	2.37	1.70	2.28	1.75	2.66
	MSE	2.24	1.96	1.99	1.87	2.29	4.06	4.46	3.63	4.19	3.76
	BIAS	0.07	0.10	0.03	0.09	0.04	0.15	0.27	0.08	0.21	0.08
$\mu = 2, \sigma = 3$	VAR	1.84	1.46	1.52	1.36	1.92	2.19	2.54	1.70	1.96	2.01
	MSE	1.93	1.68	1.54	1.51	1.95	2.62	3.96	1.82	2.80	2.13
	BIAS	0.14	0.29	0.04	0.21	0.06	0.32	0.71	0.36	0.61	0.09
$\mu = 4, \sigma = 1$	VAR	1.98	2.47	1.43	1.94	1.76	2.71	3.93	3.64	4.24	2.05
	MSE	2.37	4.15	1.46	2.80	1.83	4.74	13.88	6.21	11.59	2.21
	BIAS	0.11	0.21	0.04	0.15	0.06	0.25	0.57	0.19	0.42	0.10
$\mu = 4, \sigma = 3$	VAR	1.90	2.23	1.43	1.53	1.86	2.42	4.69	3.06	3.77	1.87
	MSE	2.15	3.07	1.46	2.00	1.93	3.67	11.24	3.79	7.29	2.08
		$0.9N(0,1) + 0.1t_\nu(\phi)$[b]					**$0.8N(0,1) + 0.2t_\nu(\phi)$**				
		$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$	$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$
	BIAS	0.12	0.17	0.09	0.15	0.09	0.27	0.36	0.24	0.33	0.21
$\phi = 2, \nu = 5$	VAR	1.89	1.37	1.76	1.37	2.08	2.35	1.73	2.22	1.76	2.61
	MSE	2.19	1.92	1.93	1.83	2.24	3.82	4.29	3.36	3.99	3.52
	BIAS	0.12	0.16	0.09	0.15	0.08	0.26	0.35	0.23	0.32	0.20
$\phi = 2, \nu = 3$	VAR	1.88	1.37	1.75	1.37	2.07	2.34	1.76	2.18	1.78	2.58
	MSE	2.16	1.89	1.90	1.80	2.21	3.69	4.18	3.22	3.86	3.39
	BIAS	0.14	0.28	0.04	0.20	0.06	0.31	0.69	0.34	0.59	0.09
$\phi = 4, \nu = 5$	VAR	1.97	2.44	1.40	1.86	1.78	2.68	3.92	3.59	4.18	1.93
	MSE	2.35	4.03	1.43	2.67	1.85	4.65	13.45	5.93	11.10	2.10
	BIAS	0.14	0.27	0.04	0.20	0.06	0.31	0.68	0.33	0.57	0.09
$\phi = 4, \nu = 3$	VAR	1.96	2.43	1.39	1.81	1.79	2.66	3.93	3.58	4.16	1.87
	MSE	2.33	3.94	1.41	2.58	1.86	4.54	13.10	5.76	10.72	2.05
		$0.9t_5(0) + 0.1t_\nu(\phi)$					**$0.8t_5(0) + 0.2t_\nu(\phi)$**				
		$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$	$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$
	BIAS	0.13	0.18	0.12	0,17	0.10	0.29	0.38	0.30	0.36	0.25
$\phi = 2, \nu = 5$	VAR	2.09	1.78	2.01	1.73	2.19	2.63	2.11	2.55	2.12	2.89
	MSE	2.42	2.46	2.31	2.30	2.38	4.26	5.04	4.31	4.76	4.10
	BIAS	0.12	0.18	0.12	0.16	0.09	0.27	0.37	0.28	0.35	0.23
$\phi = 2, \nu = 3$	VAR	2.08	1.80	1.99	1.74	2.18	2.62	2.16	2.52	2.15	2.86
	MSE	2.39	2.44	2.27	2.27	2.35	4.11	4.94	4.13	4.63	3.94
	BIAS	0.15	0.32	0.08	0.24	0.05	0.33	0.73	0.40	0.64	0.15
$\phi = 4, \nu = 5$	VAR	2.19	2.93	2.21	2.57	1.88	3.06	4.32	4.28	4.59	3.08
	MSE	2.61	4.97	2.34	3.74	1.94	5.26	14.90	7.52	12.80	3.53
	BIAS	0.14	0.31	0.08	0.24	0.05	0.32	0.71	0.39	0.63	0.14
$\phi = 4, \nu = 3$	VAR	2.18	2.92	2.16	2.53	1.89	3.02	4.33	4.24	4.57	2.94
	MSE	2.59	4.87	2.28	3.63	1.95	5.13	14.54	7.28	12.40	3.35

[a]Normalized by the sample size $n = 20$.

[b]$t_\nu(\phi)$ denotes the t-distribution with ν degrees of freedom centered at ϕ.

TABLE 2

The Monte Carlo bias, variance and MSE[a] *of* $\tilde{X}$, $\bar{X}_\alpha$ *and* $\tilde{X}_\alpha$ *for distribution functions in neighborhoods of nonparametric models with natural parameters*

		$0.9N(0,1) + 0.1N(\mu,\sigma^2)$					**$0.8N(0,1) + 0.2N(\mu,\sigma^2)$**				
		$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$	$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$
	BIAS	0.14	0.18	0.12	0.17	0.11	0.30	0.37	0.28	0.35	0.25
$\mu = 2, \sigma = 1$	VAR	1.87	1.42	1.82	1.44	2.09	2.30	1.72	2.22	1.76	2.63
	MSE	2.25	2.09	2.10	2.01	2.33	4.04	4.43	3.77	4.19	3.84
	BIAS	0.07	0.13	0.04	0.10	0.03	0.15	0.30	0.12	0.24	0.08
$\mu = 2, \sigma = 3$	VAR	1.84	1.77	1.63	1.55	1.92	2.06	2.49	1.99	2.13	2.01
	MSE	1.94	2.10	1.65	1.74	1.95	2.52	4.23	2.27	3.23	2.14
	BIAS	0.15	0.31	0.12	0.25	0.07	0.34	0.70	0.40	0.62	0.19
$\mu = 4, \sigma = 1$	VAR	1.85	2.27	1.77	2.04	1.73	2.75	3.64	3.30	3.77	2.80
	MSE	2.33	4.25	2.04	3.31	1.82	5.04	13.56	6.53	11.44	3.54
	BIAS	0.12	0.25	0.08	0.19	0.06	0.26	0.62	0.29	0.49	0.13
$\mu = 4, \sigma = 3$	VAR	1.80	2.40	1.75	1.93	1.85	2.55	5.04	3.48	4.31	2.33
	MSE	2.10	3.69	1.87	2.64	1.92	3.94	12.69	5.19	9.09	2.69
		$0.9N(0,1) + 0.1t_\nu(\phi)$[b]					**$0.8N(0,1) + 0.2t_\nu(\phi)$**				
		$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$	$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$
	BIAS	0.12	0.17	0.10	0.15	0.09	0.29	0.38	0.27	0.35	0.24
$\phi = 2, \nu = 5$	VAR	1.75	1.33	1.65	1.33	1.99	2.48	1.97	2.43	1.99	2.79
	MSE	2.03	1.90	1.86	1.80	2.14	4.19	4.78	3.94	4.47	3.98
	BIAS	0.12	0.16	0.10	0.15	0.09	0.28	0.37	0.27	0.34	0.23
$\phi = 2, \nu = 3$	VAR	1.81	1.40	1.69	1.39	2.05	2.31	1.76	2.17	1.78	2.66
	MSE	2.08	1.94	1.90	1.84	2.21	3.92	4.44	3.58	4.12	3.72
	BIAS	0.15	0.31	0.11	0.25	0.06	0.33	0.69	0.39	0.60	0.18
$\phi = 4, \nu = 5$	VAR	1.87	2.30	1.81	2.06	1.78	2.95	3.71	3.36	3.80	2.93
	MSE	2.34	4.22	2.06	3.27	1.86	5.14	13.26	6.40	11.05	3.60
	BIAS	0.14	0.30	0.10	0.23	0.05	0.33	0.69	0.38	0.59	0.19
$\phi = 4, \nu = 3$	VAR	1.90	2.39	1.85	2.11	1.83	2.90	3.86	3.39	3.91	2.81
	MSE	2.31	4.13	2.07	3.20	1.89	5.07	13.26	6.34	10.97	3.50
		$0.9t_5(0) + 0.1t_\nu(\phi)$					**$0.8t_5(0) + 0.2t_\nu(\phi)$**				
		$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$	$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$
	BIAS	0.14	0.19	0.14	0.18	0.12	0.30	0.38	0.31	0.36	0.26
$\phi = 2, \nu = 5$	VAR	1.97	1.74	1.98	1.69	2.24	2.58	2.10	2.48	2.09	2.92
	MSE	2.36	2.45	2.36	2.30	2.52	4.34	5.00	4.41	4.73	4.31
	BIAS	0.14	0.19	0.14	0.18	0.11	0.28	0.36	0.29	0.34	0.25
$\phi = 2, \nu = 3$	VAR	2.15	1.90	2.10	1.83	2.36	2.53	2.18	2.56	2.12	2.89
	MSE	2.54	2.61	2.47	2.45	2.62	4.11	4.80	4.26	4.50	4.18
	BIAS	0.16	0.32	0.14	0.27	0.07	0.37	0.74	0.46	0.66	0.26
$\phi = 4, \nu = 5$	VAR	2.27	2.90	2.50	2.64	2.22	3.71	4.33	4.42	4.44	4.12
	MSE	2.76	5.00	2.86	4.06	2.32	6.49	15.17	8.65	13.20	5.46
	BIAS	0.16	0.34	0.15	0.28	0.08	0.38	0.74	0.46	0.66	0.26
$\phi = 4, \nu = 3$	VAR	2.15	2.82	2.39	2.55	2.15	3.39	4.28	4.22	4.36	3.80
	MSE	2.67	5.14	2.83	4.13	2.27	6.25	15.20	8.44	13.11	5.12

[a] Normalized by the sample size $n = 20$.

[b] $t_\nu(\phi)$ denotes the t-distribution with ν degrees of freedom centered at ϕ.

TABLE 3
Normalized variance of $\tilde{X}$, $\bar{X}_{0.1}$ and $\bar{X}_{0.2}$ from $N(0,1)$ $(n = 20)$

	$\tilde{X}$	$\bar{X}_{0.1}$	$\bar{X}_{0.2}$
Exact value	1.47	1.02	1.06
Princeton study	1.50	1.02	1.06
Duplication I	1.46	1.02	1.05
Duplication II	1.42	1.00	1.02

has heavier tails. The asymptotic MSE of $\tilde{X}_\alpha$ tends to be smaller than that of $\tilde{X}$ unless $\alpha < \varepsilon$.

From Table 2 it is noticed that the asymptotic results give a reasonable approximation to the finite sample size Monte Carlo results. As the sample size increases, we can see that comparison of MSEs among $\tilde{X}$, $\bar{X}_\alpha$ and $\tilde{X}_\alpha$ will be one sided in favor of $\tilde{X}_\alpha$ because the bias increases faster than the variance. As a check of the simulation results, duplication of VAR is made for $\tilde{X}$, $\bar{X}_{0.1}$ and $\bar{X}_{0.2}$ from $N(0,1)$. They are shown in Table 3 with the exact values and the Princeton robustness study in Andrews, Bickel, Hampel, Huber, Rogers and Tukey (1972), page 69. Some idea of the precision for the Monte Carlo results can be gained from Table 3.

5. ARE for symmetric distribution functions. Once we observe that the metrically trimmed mean $\tilde{X}_\alpha$ works well for distribution functions in neighborhoods of nonparametric models with natural parameters, it is interesting to investigate the performance of $\tilde{X}_\alpha$ for symmetric distribution functions as well. Let $\tilde{\sigma}^2$ and $\bar{\sigma}_\alpha^2$ denote the asymptotic variance of $\tilde{X}$ and $\bar{X}_\alpha$ respectively. We will denote by $e_1(F, \alpha)$ the asymptotic relative efficiency (ARE) of $\tilde{X}_\alpha$ to $\tilde{X}$ and similarly by $e_2(F, \alpha)$ the ARE of $\tilde{X}_\alpha$ to $\bar{X}_\alpha$, where

$$e_1(F, \alpha) = \tilde{\sigma}^2/\tilde{\sigma}_\alpha^2 \quad \text{and} \quad e_2(F, \alpha) = \bar{\sigma}_\alpha^2/\tilde{\sigma}_\alpha^2.$$

The values of $\tilde{\sigma}^2$, $\bar{\sigma}_\alpha^2$, $\tilde{\sigma}_\alpha^2$, $e_1(F, \alpha)$ and $e_2(F, \alpha)$ are shown in Table 4 for some symmetric distribution functions. Table 4 shows that $\tilde{X}_{0.1}$ is more efficient than $\tilde{X}$ but $\tilde{X}_{0.1}$ is less efficient than $\bar{X}_{0.1}$. The efficiency of $\tilde{X}_\alpha$ decreases as α increases from 0.1 to 0.2. We will derive the lower bounds of $e_1(F, \alpha)$ and $e_2(F, \alpha)$ for all symmetric unimodal distribution functions in the following proposition.

PROPOSITION 5.1. *Let $\mathscr{F}$ be the family of all symmetric unimodal distribution functions which possess the regularity conditions of Proposition* 3.1. *Then*

$$\inf\{F \in \mathscr{F}; e_1(F, \alpha)\} = 3/(7 - 4\alpha) \tag{5.1}$$

and

$$\inf\{F \in \mathscr{F}; e_2(F, \alpha)\} = (1 + 2\alpha)/(7 - 4\alpha). \tag{5.2}$$

TABLE 4
The asymptotic variance of $\tilde{X}$, $\bar{X}_\alpha$ and $\tilde{X}_\alpha$ and the ARE for symmetric distribution functions

	$\tilde{X}$	$\bar{X}_{0.1}$	$\tilde{X}_{0.1}$	$\bar{X}_{0.2}$	$\tilde{X}_{0.2}$	$e_1(F, 0.1)$	$e_2(F, 0.1)$	$e_1(F, 0.2)$	$e_2(F, 0.2)$
$N(0, 1)$	1.57	1.03	1.54	1.06	1.83	1.02	0.67	0.86	0.58
$t_5(0)$[a]	1.73	1.39	1.59	1.35	1.82	1.09	0.87	0.95	0.74
$0.9N(0, 1) + 0.1N(0, 9)$	1.80	1.32	1.56	1.30	1.93	1.15	0.85	0.93	0.67
$0.9N(0, 1) + 0.1t_5(0)$	1.59	1.05	1.53	1.08	1.83	1.04	0.69	0.87	0.59
$0.9N(0, 1) + 0.1t_3(0)$	1.60	1.07	1.53	1.10	1.83	1.05	0.70	0.87	0.60
$0.9t_5(0) + 0.1t_3(0)$	1.75	1.43	1.59	1.37	1.82	1.10	0.90	0.96	0.75
$0.8N(0, 1) + 0.2N(0, 9)$	2.09	1.80	1.66	1.63	2.05	1.26	1.08	1.02	0.80
$0.8N(0, 1) + 0.2t_5(0)$	1.60	1.08	1.53	1.11	1.82	1.05	0.71	0.88	0.61
$0.8N(0, 1) + 0.2t_3(0)$	1.62	1.12	1.52	1.14	1.82	1.07	0.74	0.89	0.63
$0.8t_5(0) + 0.2t_3(0)$	1.76	1.46	1.60	1.39	1.82	1.10	0.91	0.97	0.76

[a] $t_\nu(\phi)$ denotes the t-distribution with ν degrees of freedom centered at ϕ.

PROOF. We will prove (5.2). In a similar way, (5.1) can be proved. Without loss of generality, let F be a symmetric distribution function about 0. Let $a = F^{-1}(1 - \alpha/2)$, $b = f(a)/f(0)$, $c = \int_{-a}^{a} x^2 f(x)\,dx$ and $d = \int_{-a}^{a} |x| f(x)\,dx$. Then

$$e_2(F, \alpha) = (\alpha a^2 + c)/(a^2 b^2 + 2abd + c). \tag{5.3}$$

The problem is now to maximize d subject to the given conditions on a, b and c. The maximum of d is clearly approached by a density concentrating its mass more and more closely inside of $-a$ and a. Since $f(x)$ is symmetric and unimodal, the maximum value of d is approached by the density $f^*(x)$, where

$$f^*(x) = \begin{cases} (1 - \alpha)/(2a), & \text{if } |x| \le a, \\ \beta^*, & \text{if } |x| > a, \end{cases}$$

and β^* is arbitrary, provided that $f^*(x)$ is continuous and unimodal. A formal proof for the density $f^*(x)$ can be found in Bickel (1965), Lemma 4.1. For this density $f^*(x)$, we have $b = 1$, $c = a^2(1 - \alpha)/3$ and $d = a(1 - \alpha)/2$. Substituting the values of b, c and d in (5.3), the proof is complete. □

6. Extension to multivariate location. The trimmed mean $\bar{X}_\alpha$ has been extended to the multivariate case in various directions such as convex peeling and depth trimming as discussed by Barnett (1976) and Donoho and Huber (1983). However, few theoretical results are available. This is due to the fact that there is no universal agreement on multivariate ordering and it is rarely possible to remove outliers without reliable robust estimates of multivariate location. For recent developments, we refer to Davies (1987), Lopuhaä (1989), Rousseeuw and van Zomeren (1990) and Lopuhaä and Rousseeuw (1991). On trimming and weighting for robust estimation in the linear model, we also refer to Koenker and Basset (1978), Ruppert and Carroll (1980), Portnoy (1987), Welsh (1987) and Portnoy and Koenker (1989).

The same difficulties arise in the extension of the metrically trimmed mean to the multivariate case except for spatial data. For spatial data, the univariate median is naturally extended to the spatial median as discussed by Brown (1983). Thus the univariate definition of the metrically trimmed mean in (2.1) still makes sense, interpreting $\tilde{X}$ as the spatial median and $|\cdot|$ as the Euclidean distance. It is noticed that the spatial median and the metrically trimmed mean from the spatial median are not affine equivariant if they are rotationally equivariant.

As an alternative to the spatial median, Oja (1983) proposed an affine equivariant estimator, called the generalized median, which is based on minimizing a U-statistic as follows: Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a p-dimensional random sample. Let $\boldsymbol{\Delta}(\mathbf{X}_{j_1}, \dots, \mathbf{X}_{j_p}; \boldsymbol{\theta})$ be the volume of the p-dimensional simplex formed by $\mathbf{X}_{j_1}, \dots, \mathbf{X}_{j_p}$ and $\boldsymbol{\theta} \in \mathbb{R}^p$. Then the generalized median, denoted by $\tilde{\mathbf{X}}$, is a solution of $\boldsymbol{\theta}$ which minimizes

$$(6.1) \qquad \binom{n}{p}^{-1} \sum \boldsymbol{\Delta}(\mathbf{X}_{j_1}, \dots, \mathbf{X}_{j_p}; \boldsymbol{\theta}),$$

where the summation in (6.1) is over the $\binom{n}{p}$ combinations of p distinct elements $\{j_1, \dots, j_p\}$ from $\{1, \dots, n\}$. Recently Lopuhaä and Rousseeuw (1991) indicate that the finite sample replacement breakdown point of $\tilde{\mathbf{X}}$ might be greater than $[(n-p+1)/2]/n$.

Let

$$(6.2) \qquad Y_i = \binom{n-1}{p-1}^{-1} \sum \boldsymbol{\Delta}\left(\mathbf{X}_{j_1}, \dots, \mathbf{X}_{j_{p-1}}; \mathbf{X}_i, \tilde{\mathbf{X}}\right) \quad \text{for } i = 1, \dots, n,$$

where the summation in (6.2) is over the $\binom{n-1}{p-1}$ combinations of $p-1$ distinct elements $\{j_1, \dots, j_{p-1}\}$ from $\{1, \dots, i-1, i+1, \dots, n\}$. That is, Y_i is the average volume of simplexes formed by $p-1$ of the sample points, $\mathbf{X}_i$ and $\tilde{\mathbf{X}}$. To determine if $\mathbf{X}_i$ is outlying, we use Y_i as our distance of $\mathbf{X}_i$ from $\tilde{\mathbf{X}}$. Let $\tilde{K}_n$ denote the empirical distribution function of $Y_1, \dots, Y_n$. Then the metrically trimmed mean in (2.1) is extended to the multivariate case as

$$(6.3) \qquad \tilde{\mathbf{X}}_\alpha = \frac{1}{n - [\alpha n]} \sum_{i=1}^{n} \mathbf{X}_i 1\left\{Y_i \le \tilde{K}_n^{-1}(1-\alpha)\right\}.$$

Since the volume of simplex is affine equivariant, $\tilde{\mathbf{X}}_\alpha$ in (6.3) is affine equivariant. The statistical properties of $\tilde{\mathbf{X}}_\alpha$ in (6.3) are not known yet.

Acknowledgments. This paper is a part of the author's Ph.D. dissertation completed at the University of California, Berkeley. The author would like to thank Professor Kjell A. Doksum for his guidance and Professor Peter J. Bickel for many valuable comments. Helpful comments from the Editor and the referee are also acknowledged.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton Univ. Press.

BARNETT, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A* **139** 318–344.

BICKEL, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858.

BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for nonparametric models. *Ann. Statist.* **3** 1038–1069.

BROWN, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30.

DAVIES, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, Calif.

FERNHOLZ, L. T. (1983). *von Mises Calculus for Statistical Functionals*. Springer, New York.

HUBER, P. J. (1964). Robust estimation of location parameter. *Ann. Math. Statist.* **35** 73–101.

KOENKER, R. W. and BASSET, G. W. (1978). Regression quantiles. *Econometrica* **46** 33–50.

LOPUHAÄ, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17** 1662–1683.

LOPUHAÄ, H. P. and ROUSSEEUW, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229–248.

MOSTELLER, F. and TUKEY, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, Reading, Mass.

OJA, H. (1983). Descriptive statistics for multivariate distribution. *Statist. Probab. Lett.* **1** 327–333.

PORTNOY, S. (1987). Using regression fractiles to identify outliers. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (Y. Dodge, ed.) 345–356. North-Holland, Amsterdam.

PORTNOY, S. and KOENKER, R. (1989). Adaptive L-estimation for linear models. *Ann. Statist.* **17** 362–381.

ROUSSEEUW, P. J. and VAN ZOMEREN, B. C. (1990). Unmasking multivariate outliers and leverage points. *J. Amer. Statist. Assoc.* **85** 633–639.

RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

WELSH, A. H. (1987). The trimmed mean in the linear model. *Ann. Statist.* **15** 20–36.

DEPARTMENT OF STATISTICS
SUNG KYUN KWAN UNIVERSITY
CHONGRO-KU, SEOUL, 110-745
KOREA