Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.1 pp. 47 53

A New Redescending M-Estimating Function

Ro Jin Pak[1)]

Abstract

A new redescending M-estimating function is introduced. The estimators by this new redescending function attain the same level of robustness as the existing redescending M-estimators, but have less asymptotic variances than others except few cases. We have focused on estimating a location parameter, but the method can be extended for a scale estimation.

Keywords : Minimum L_2 distance; M-estimating function

1. Introduction

Redescending M-estimators have ψ functions which are nondecreasing near the origin but then decrease toward the axis as they go far from the origin. They usually satisfy $\psi(x) = 0$ for all x with $|x| \geq r$, r is a finite number which may be considered as the minimum rejection point. They were very successful in the Princeton Robustness Study (Andrews, et al., 1972). There are many representative redescending functions: *three-part redescending function* (Andrews, et al., 1972), *sine function* (Andrews, et al, 1972), *biweight function* (Beaton and Tukey, 1974), *Tanh function* (Andrews, et al., 1972).

New redescending function is based on minimization of L_2 distance of a model density and its density estimator. Let g_θ be a family of probability densities indexed by θ. The minimum distance estimator $\hat{\theta}$ is defined by a statistical quantity minimizing L_2 distance, which is a solution to L_2

$$\nabla_\theta \int (p(x) - g_\theta(x))^2 \, dx, \tag{1}$$

where we assume $p(x), g_\theta(x) \in L_2$ and ∇_θ represents a derivative with respect to

1. Associate Professor, Department of Computer Sciences & Statistics, Dankook University, Seoul, Korea 140-714. The research was conducted by the research fund of Dankook University in 2001.

$$\int (p(x) - g_\theta(x)) \nabla_\theta g_\theta(x)\, dx = 0. \tag{2}$$

θ. The equation (1) can be written as

Since we have $\int g_\theta(x) \nabla_\theta g_\theta(x) dx = (1/2) \nabla_\theta \int g_\theta^2(x)\, dx = 0$, if θ is a location parameter, the equation (2) becomes

$$\int p(x) \nabla_\theta g_\theta(x)\, dx = \nabla_\theta \int p(x) g_\theta(x) dx = 0. \tag{3}$$

Given a random sample, $X_1, X_2, \cdots, X_n$, having a density $g_\theta(x)$, let $p(x)$ be a density estimator for $g_\theta(x)$ such as $p(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$ where $K(\cdot)$ is a kernel and h is the window width (Silverman, 1986). The equation (3) can be written as

$$\sum_{i=1}^{n} \nabla_\theta \int \frac{1}{h} K\left(\frac{x - X_i}{h}\right) g_\theta(x)\, dx = 0.$$

If we follow Huber (1981), and if we denote T_n as an estimate of θ, we have

$$\psi(X_i; T_n) = \nabla_\theta \int \frac{1}{h} K\left(\frac{x - X_i}{h}\right) g_\theta(x) dx \Big|_{\theta = T_n}.$$

Suppose a kernel is a Gaussian and a model is the normal with mean μ and variance σ^2, then $\int (1/h) K\{(x - X_i)/h\} g_\mu(x) dx$, a convolution of a Gaussian kernel and a normal density, is the normal with mean μ and variance $h^2 + \sigma^2$. After dropping unnecessary constants, we have

$$\psi(X_i; T_n) = (X_i - \mu) \exp[-(X_i - \mu)^2 / 2(h^2 + \sigma^2)] \Big|_{\mu = T_n}.$$

Proposed Function: Redefine ψ-function as

$$\psi_r(t) = t \exp[-t^2/2r^2] \quad \text{for} \quad t \in (-\infty, \infty), \tag{4}$$

where r is a tuning constant.

We propose for generality to redefine the $\psi_r(t)$ with r as a tuning constant. The $\psi_r(t)$ is originally derived based on normal density, but in the section we will show that it works quite well for data sampled from not only normal distribution but also other distributions. One may consider r as a function of h and σ rather than just a constant, and try to find a proper value of r by replacing h and σ by proper estimators. There are a lot of things to talk about this idea, but this time we would like to have a function similar to the existing functions, which are controlled by constants like tuning constant, trimming constant or bending constant.

2. Properties of the proposed function

1. Shape: For a given finite constant $c>0$, the class of redescending functions Ψ_c consists of all mappings ψ: $\mathbf{R} \to \mathbf{R}$ satisfying

- ψ is continuous on $\mathbf{R}$, $\psi(-t) = -\psi(t)$ for all t, and $\psi(t) \geq 0$ for $t \geq 0$;
- the set $D(\psi)$ of points in which ψ' is not defined or not continuous is finite;
- $\psi(t) = 0$ for $|t| \geq c$.

The ψ_r in (4) satisfies the condition 1 and 2. Though it does not fully satisfy the condition 3, it is basically redescending (Figure 1). Since the propose ψ_r function is slowly decaying but never becomes 0, we can avoid some computational problems mentioned by Huber (1981, p103) and Hampel, et al. (1986, p152), while it can effectively handle extreme outliers like the other redescending M-estimating functions.

Note that the estimates we got in actual calculations are so called one-step M-estimates
which are defined by

$$T_n = T_n^{(0)} + \frac{S_n \sum_{i=1}^{n} \psi((x_i - T_n^{(0)})/S_n)}{\sum_{i=1}^{n} \psi'((x_i - T_n^{(0)}/S_n)},$$

where the initial estimates of location $T_n^{(0)}$ is the median of the observations $x_1, \cdots, x_n$, and $S_n = 1.483 \text{ med}_i \{|x_i - \text{med}_j (x_j)\}$. In the process of getting one-step M-estimates, we are not going to have '0' in denominator, which could cause 'overflow' during computation, so that the proposed $\psi_r(t)$ produces very stable estimates compare to the other redescending M-estimating functions. The function ψ_r deals various M-estimating functions from a redescending function to a non-decreasing $\psi(t) = t$ according to r. When r is infinite, we have $\psi(t) = t$ for all t, which produces (non-robust, but most efficient) least-squares estimators. Of course, if r is moderate, the propose $\psi_r(t)$ plays like a redescending M-estimating function. With a proper value of r, we can have the $\psi_r(t)$ which produces the estimator as efficient as least squares estimator while we keep reasonable level of robustness.

2. Robustness: Figure 1 displays ψ_r with $r = 2$ and 5 (top), Influence Functions (IF) at $F = \Phi$ (middle) and Change-of-variance functions (CVF) at $F = \Phi$ (bottom). For each r, IF and CVF are bounded, and CVF with $r = 5$ looks similar to the CVF of the logistic likelihood estimator. Figure 2 displays γ^* (gross-error-sensitivity), χ^* (change-of-variance sensitivity) and *efficiency* (asymptotic efficiency) of ψ_r for $r \in (0, 5]$ at $F = \Phi$. It can be said that when r is moderate, an M-estimator by ψ_r is both V-robust and B-robust. When r is 1.6, 3.3, *efficiency* is about 0.9, 0.99, respectively, and it will converge to 1 as r increases. Both γ^* and χ^* are going upward to ∞, but minimized near $r = 1.5$, which would be a good choice for r.

3. Efficiency : We compare the proposed ψ_r with some of the well-known redescending M-estimating functions (Table 1) as Hampel, et al. (pp. 166 - 167, 1986). The asymptotic efficiency of the proposed estimator at the standard normal distribution is little higher than or approximately equal to those of the other estimators except Huber-estimator. The asymptotic variances of the proposed estimator under various distributions are smaller than those for the other estimators except the variance of Huber-estimator under 5%3N. We have simulated sets of observations and calculated the estimates for two representative M-estimation functions; Huber's, Biweight; and the estimates by the function proposed in this article. 500 samples of 20 and 40 observations are simulated for various distributions listed in Table 1, asymptotic efficiencies and variances are recorded. The simulated statistics displays the similar pattern to the theoretical values in Table 1, except the case of Cauchy distribution.

4. Stableness : We also simulated 500 sets of 20 and 40 observations from $0.5N(0,1) + 0.25N(-5,1) + 0.25N(5,1)$, that is, almost half of simulated observations below -4 or above +4. We calculated Huber's estimates, the estimates by using the Biweight and the estimates by the proposed function. Those three functions are the same functions as we used in the above simulation; Huber's with $b = 1.4088$, Biweight with $r = 4$, and the proposed function with $r = 1.9388$. The variances for Huber' estimates are 1.3730 (n=20) and 1.4827 (n=40); those for the proposed estimates are 1.0869 (n=20) and 1.2981(n=40), while 3.6291 (n=20) and 2.4543(n=40) for the Biweight. The values for Biweight are almost double compare to those for the proposed function. Since the Biweight function becomes 0 outside (-4, 4), while the proposed function is slowly decay but never becomes 0, the proposed function produced more stable estimates than Biweight function.

3. Conclusions

We proposed a new redescending type of M-estimating function induced by minimizing

L_2 distance between a normal density and a Gaussian kernel density estimator. This newly proposed M-estimating function is everywhere differentiable while it is redescending, and it has been shown that estimators by the new M-estimating function perform better than existing M-estimators in terms of robustness, efficiency and computational stableness under various distributions.

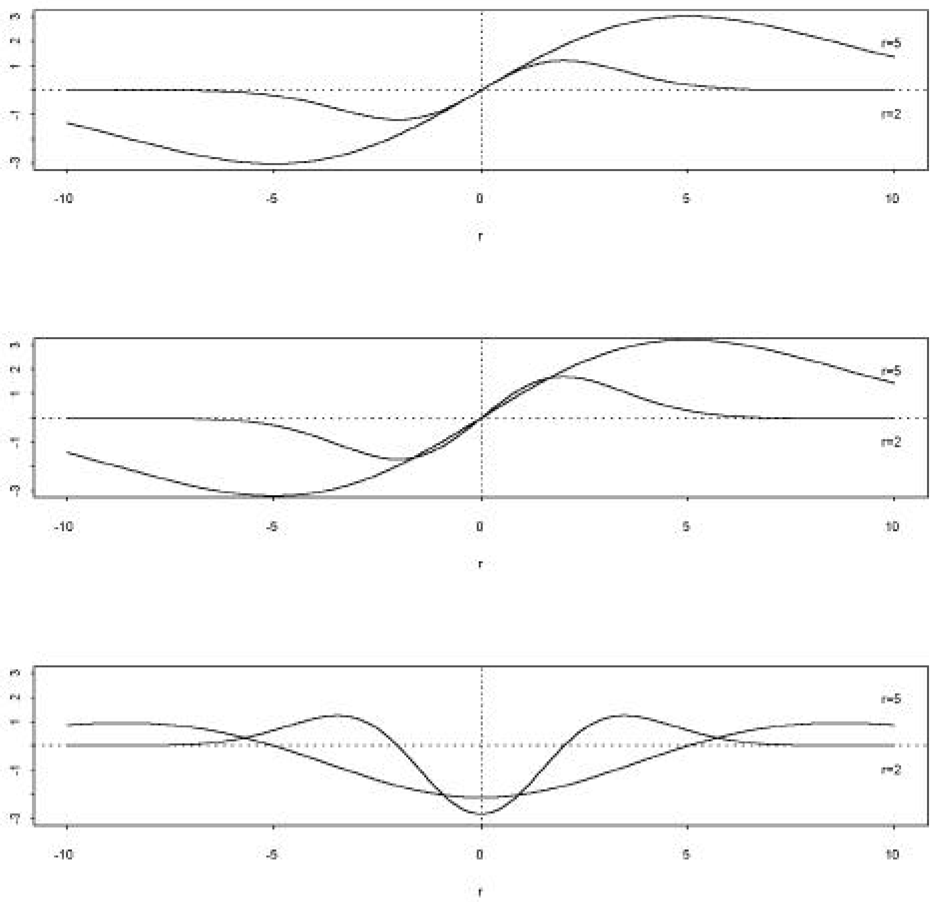

Figure 1: ψ_r (top), IF (middle) and CVF (bottom)

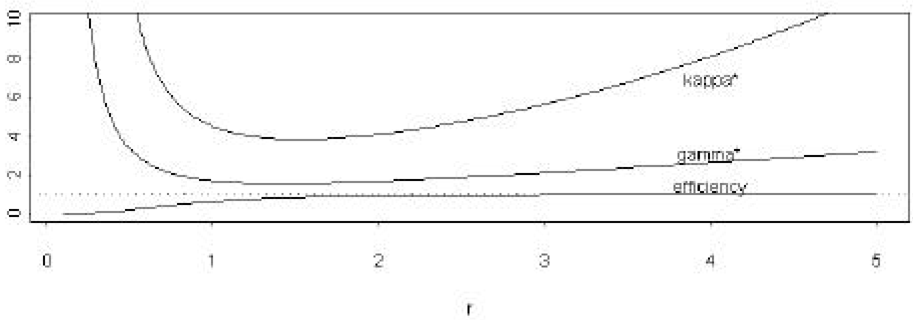

Figure 2: γ^*, κ^*, and *efficiency*

Table 1: Comparison of Some Redescending M-estimators

		Asymptotic Variances				
Estimator	efficiency	5%3N	10%10N	t_3	25%3N	Cauchy
Sine	0.9093	1.1991	1.2691	1.5769	1.7687	2.2688
Huber-Collins	0.9107	1.1966	1.2689	1.5581	1.7583	2.2591
Three-Part	0.9119	1.1954	1.2662	1.5783	1.7603	2.3306
Tanh	0.9205	1.1866	1.2590	1.5625	1.7579	2.2977
Scaled-logistic MLE	0.9344	1.1872	1.4624	1.5380	1.7989	2.6390
Huber	0.9563	1.1649	1.4385	1.5663	1.7877	2.7890
n=20	0.8879	1.2063	1.4129	1.5848	1.8338	3.9073
n=40	0.9514	1.1330	1.5215	1.5225	1.8216	3.7798
Biweight	0.9100	1.1978	1.2683	1.5708	1.7645	2.2593
n=20	0.7849	1.2660	1.3046	1.6716	1.7322	3.2951
n=40	0.8953	1.1834	1.3249	1.6160	1.8080	2.7647
Proposed	0.9344	1.1709	1.2491	1.5279	1.7360	2.2498
n=20	0.8502	1.2015	1.2721	1.5950	1.7488	3.3522
n=40	0.9341	1.1303	1.2931	1.5319	1.7961	2.8858

The figures in this table except the ones for `Proposed estimator' are as same as those in Table 3 on p.167 of Hampel et al. (1986). All estimators satisfy γ^* = 1.6749 at the standard normal distribution, where also the asymptotic efficiency is evaluated. The estimators under study are the following: sine, $\psi(t) = \sin(x/a)$ for $|x| < \pi a$ and zero otherwise, with a =1.142; biweight,

$\psi(x) = x(r^2 - x^2)^2$ for $|x|<r$ and zero otherwise, with r=4; Huber-collins, p=1.277, x_1=1.344, r=4; three-part redescending, ψ bends at 1.31, 2.039, 4; tanh-estimator, r=4, k=3.732, p=1.312, A=0.667, B=0.783; Huber-estimator; ψ bends at b=1.4088; nd scaled logistic *MLE*, $\psi(x) = [\exp(x/a) - 1]/[\exp(x/a) + 1]$ with a=1.036; the proposed estimator, r=1.9388. The abbreviation $\alpha\%\beta N$ stands for the distribution $(1-\alpha/100)\Phi(x) + (\alpha/100)\Phi(x/\beta)$ and t_3 is the t distribution with 3 degrees of freedom and Cauchy is the Cauchy distribution with 0 (location) and 1 (scale).

References

1. Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber P. J., Rogers W. H., & Tukey J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton NJ: Princeton Univ. Press.
2. Beaton, A. E. & Tukey, J. W. (1974). *Outlier in Statistical Data*. Wiley, New York.
3. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics, the Approach Based on Influence Functions*. John Wiley & Sons, New York.
4. Huber, P. J. (1981), *Robust Statistics*, John Wiley & Sons, New York.
5. Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.