*Journal of the Korean*
*Data & Information Science Society*
*2000, Vol. 11, No. 2. pp. 181 ~ 188*

# A Method of Choosing a Value of the Bending Constant in Huber's M-Estimation Function [1]

**Ro Jin Park** [2]

**Abstract**

The shape of an M-estimation function is generally determined in the sense of either/both maximizing efficiency of an M-estimator at the model or/and bounding the influence function of an M-estimator. We propose an empirical method of choosing a value of the bending constant in Huber's $\psi$-function, which is the most widely used M-estimation function when estimating the location parameter.

*Key Words and Phrases*: Influential Observation; M-estimation; Robustness

## 1. Motivation and Algorithm

We assume that the observations $X_1, X_2, \ldots, X_n$ are i.i.d., each according to the distribution $F_\theta(x) = F(x-\theta)$, where $F$ is known and $\theta$ is unknown. Any estimator $\hat{\theta}$, if $\sigma$ is known, defined by a minimum problem of the form

$$\sum_{i=1}^{n} \rho\left(\frac{X_i - \hat{\theta}}{\sigma}\right) = \min!,$$

where $\rho$ is an arbitrary differentiable function, is called an M-estimator. In practice, $\sigma$ should be estimated by a proper estimator, say $\hat{\sigma}$.

We would like to measure the influence of a particular observation, say the $i$th observation, by comparing $\hat{\theta}$ by $\hat{\theta}_{-i}$, where $\hat{\theta}_{-i}$ is the M-estimate of $\theta$ computed without the $i$th observation. By a second order Taylor series, we have the followings;

$$\sum_{i=1}^{n} \rho\left(\frac{X_i - \hat{\theta}_{-i}}{\hat{\sigma}}\right) - \sum_{i=1}^{n} \rho\left(\frac{X_i - \hat{\theta}}{\hat{\sigma}}\right) \approx$$
$$- \left(\frac{\hat{\theta}_{-i} - \hat{\theta}}{\hat{\sigma}}\right) \sum_{i=1}^{n} \rho'\left(\frac{X_i - \theta}{\hat{\sigma}}\right)\Big|_{\theta=\hat{\theta}} + \frac{1}{2}\left(\frac{\hat{\theta}_{-i} - \hat{\theta}}{\hat{\sigma}}\right)^2 \sum_{i=1}^{n} \rho''\left(\frac{X_i - \theta}{\hat{\sigma}}\right)\Big|_{\theta=\hat{\theta}}. \quad (1)$$

[1]This research was supported by BSRI-98-015-D00049.

[2]Assistant Professor, Department of Information and Statistics, Taejon University, Taejon, Korea

Let us define twice of the left-hand side of (1) as the measure of *Influence of ith observation for an M-estimation function in estimating a location parameter,* and denote it as $IM_i$. The first term on right-hand side of the equation (1) is in fact zero, so that we have

$$\begin{aligned} IM_i &= 2\left\{\sum_{i=1}^{n}\rho\left(\frac{X_i-\hat{\theta}_{-i}}{\hat{\sigma}}\right)-\sum_{i=1}^{n}\rho\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}}\right)\right\} \\ &\approx \left(\frac{\hat{\theta}_{-i}-\hat{\theta}}{\hat{\sigma}}\right)^2\sum_{i=1}^{n}\rho''\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}}\right). \end{aligned} \quad (2)$$

For Huber's proposal (Huber, 1981) such as $\rho'(t)=\psi_k(t)$, where

$$\psi_k(t)=\min\{k,\max\{t,-k\}\}=t\cdot\min\left\{1,\frac{k}{|t|}\right\}$$

for $0<k<\infty$, the sign ' $\approx$ ' in (2) can be replaced by ' $=$ ', so that we have

$$IM_i(k)=\left(\frac{\hat{\theta}_{-i}-\hat{\theta}}{\hat{\sigma}}\right)^2\sum_{i=1}^{n}\mathrm{I}_{[-k,k]}\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}}\right).$$

We can notice that there is a relationship between $IM_i$ and Cook's distance (Cook , 1977a; Cook and Weisberg, 1982), defined as

$$D_i\equiv D_i(\mathbf{X}^T\mathbf{X},p\hat{\sigma}_n^2)=\frac{(\hat{\mathbf{b}}_{(i)}-\hat{\mathbf{b}})^T(\mathbf{X}^T\mathbf{X})(\hat{\mathbf{b}}_{(i)}-\hat{\mathbf{b}})}{p\hat{\sigma}_n^2},$$

where

$\mathbf{X}$ is an $n\times p$ full $-$ rank matrix of known constants,
$\mathbf{Y}$ is an $n-$ vector of observable responses,
$\hat{\sigma}_n^2$ is a sample variance of residuals and $\hat{\mathbf{b}}$, and
$\hat{\mathbf{b}}_{(i)}$ are the vectors of the least squares estimates of the vector of regression parameters, $\mathbf{b}$, with all observations and observation except the $i$th observation, respectively.

The $D_i$ is a result of the second-order expansion of $\rho(\mathbf{Y}-\mathbf{X}\hat{\mathbf{b}})$ w.r.t $\hat{\mathbf{b}}_{(i)}$ if $\rho(\mathbf{Z})=\mathbf{Z}^T\mathbf{Z}$ for a matrix $\mathbf{Z}$. Furthermore, if $p=1$, $D_i$ turns out

$$D_i=n\left(\frac{\bar{X}_{-i}-\bar{X}}{\hat{\sigma}_n}\right)^2=2\left\{\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}_{-i}}{\hat{\sigma}_n/\sqrt{n}}\right)^2-\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\hat{\sigma}_n/\sqrt{n}}\right)^2\right\}, \quad (3)$$

where $\bar{X}=\sum X_i/n$, $\hat{\sigma}_n^2=\sum(X_i-\bar{X})^2/(n-1)$, and $\bar{X}_{-i}$ is a sample mean calculated without an $i$th observation. $D_i$ in (3) is in fact a location parameter version of Cook's distance. The above quantity is also a special case of $IM_i$ when M-estimating function, $\rho'(t) = \psi(t) = t$ for all $t$. The (3) is in fact also a special case of the likelihood displacement in Cook, et al. (1988). The values of $D_i(\mathbf{X}^T\mathbf{X}, p\hat{\sigma}_n^2)$ can be converted to a familiar probability scale by comparing computed values to the $F(p, n-p)$ distribution. For example, if $D_i(\mathbf{X}^T\mathbf{X}, p\hat{\sigma}_n^2)$ equals the 0.5 value of the corresponding $F$ distribution, then deletion of the $i$-th case would move the estimate of $\mathbf{b}$ to the edge of a 50% confidence ellipsoid relative to $\hat{\mathbf{b}}$, a potentially important change (Cook, 1977b). If the largest $D_i$ is substantially less than 1, deletion of a case will not change the estimate of $\mathbf{b}$ by much.

Consider how to figure out a proper value of the bending constant for a particular data set by using $IM_i$. We want Huber's M-estimating function effectively to handle influential observations, so that no observation is considered as an influential observation; that is, we want $IM_i$ for all $i$ to be less than or equal to a proper value from $F$ distribution. In order to attain our goal, we want the largest $k$, say $k^*$ satisfying that

$$\max\{IM_i(k)\} \leq F(1-\alpha, 1, n-1), \tag{4}$$

where $F(1-\alpha, 1, n-1)$ is the $(1-\alpha)\times 100\%$ point of the $F$ distribution with 1 and $n-1$ degrees of freedom, for a given $\alpha$. The observations bounded by $k^*$ and $-k^*$ are no longer influential observations in terms of Cook's distance, and in that sense Huber's estimate with that $k^*$ is a robust alternative to a sample mean. Let see how Huber's estimates behaves according to the various values of $\alpha$. We have generated 1000 samples of 20 random observations from the normal distributions in which samples are from $N(0,1)$ with probability 1.0, 0.9, 0.8, 0.7, 0.6, respectively, and otherwise from $N(5,1)$; that is, samples are simulated from $(1-\epsilon)N(0,1) + \epsilon N(5,1)$ for $\epsilon =$ 0.0, 0.1, 0.2, 0.3, and 0.4. For each sample, we picked up the largest $k$ satisfying; $\max\{IM_i(k)\} \leq F(1-\alpha, 1, 19)$ for $\alpha = 0.5, 0.625$, and 0.75, and calculated Huber's estimates for each level of $\alpha$ with bending constants obtained. The box-plots of the various estimates for a location parameter $\mu$ show how the estimates behave according to $\alpha$'s under various contamination (Figure 1). There are five sections separated by the vertical lines according to the levels of contamination, and in each section. The box-plots of sample means, Huber's estimates according to the upper limit in (4) with $\alpha$=0.5, 0.625, 0.75, the estimates with $k = 1.28$ (suggested by Wilcox, 1997) and $k = 1.36$ (suggested by Hampel, et al., 1986), and medians are plotted from the left to the right. Under the normal density Huber's M-estimator attains its asymptotic efficiency 0.94 and 0.95 when $k$=1.28 and 1.36, respectively. When $\alpha = 0.5$, the Huber's M-estimates are closer to sample means, but when $\alpha = 0.75$, the estimates are closer to medians. When a level of contamination is 30% and more, Huber's M-estimates with $k = 1.28, 1.36$ seem to break down while Huber's M-estimates with $\alpha = 0.75$ are robust as the medians. However, Huber's M-estimates with $\alpha = 0.75$ and the medians have larger standardized errors than

Table 1: Means and Standardized errors in parentheses of the various M-estimates; $\alpha = \alpha_0$ stands for the M-estimates based on $(1 - \alpha_0)$th quartile of F-distribution

| | 0% | 10% | 20% | 30% | 40% |
|---|---|---|---|---|---|
| sample mean | -0.0103 | 0.5216 | 0.9397 | 1.4921 | 1.9947 |
| | (0.2126) | (0.2145) | (0.2161) | (0.2138) | (0.2278) |
| Huber's | 0.0138 | 0.3307 | 0.7273 | 1.2850 | 1.7566 |
| with $\alpha$=0.5 | (0.2059) | (0.2431) | (0.7727) | (0.4587) | (0.4284) |
| Huber's | 0.0124 | 0.2432 | 0.5649 | 1.0714 | 1.5258 |
| with $\alpha$=0.625 | (0.2133) | (0.2400) | (0.3285) | (0.4886) | (0.5313) |
| Huber's | 0.0093 | 0.1801 | 0.4044 | 0.7890 | 1.1967 |
| with $\alpha$=0.75 | (0.2220) | (0.2390) | (0.2905) | (0.3837) | (0.4956) |
| Huber's | 0.0041 | 0.1591 | 0.4461 | 0.9521 | 1.7471 |
| with $k = 1.28$ | (0.2278) | (0.2494) | (0.2746) | (0.3511) | (0.3513) |
| Huber's | 0.0040 | 0.1656 | 0.4634 | 0.9992 | 1.7858 |
| with $k = 1.36$ | (0.2267) | (0.2486) | (0.2760) | (0.3608) | (0.3373) |
| median | -0.1394 | 0.2571 | 0.2571 | 0.5118 | 0.3821 |
| | (0.2587) | (0.3126) | (0.3126) | (0.3742) | (0.4497) |

Huber's M-estimates with 1.28 and 1.36. That is, the plots are suggesting that there is a trade-off between robustness and efficiency as the level of contaminations increases. However, the ranges of Huber's M-estimates with $k$=1.28 and 1.36 are wider than Huber's M-estimates with $\alpha = 0.75$, that is, Huber's estimates with a fixed bending constant are less stable than Huber's M-estimates with $\alpha = 0.75$.

The fact is that when we consider robustness, 20% contamination at is a lot. We can notice that Huber's M-estimates with $\alpha = 0.75$ perform quite well over Huber's M-estimates with 1.28 and 1.36 at lower levels of contamination. Hence, we propose that Huber's M-estimator based on the algorithm in this article with $\alpha = 0.75$ should be a good alternative to those with $k = 1.28$ and 1.36.

## 2. Examples

In the following three examples, an M-estimate of location is a result of the first iteration of an iterative estimation procedure known as the Newton-Raphson method, that is,

$$T_n = T_n^0 + S_n^0 \sum \psi\left(\frac{X_i - T_n^0}{S_n^0}\right) \Big/ \sum \psi'\left(\frac{X_i - T_n^0}{S_n^0}\right),$$

where $T_n^0$ = median and $S_n^0$ = (Median Absolute Deviation) / 0.6745.

**Example 1: The Cushney and Peebles Data.** In 1904, Cushney and Peebles published their experimental results on "The Action of Optimal Isomers" in the *Journal of Physiology.* The following data are copied from Table 4.1 of Staudte and Sheather (1990, p97), and written in order:

0.0, 0.8, 1.0, 1.2, 1.3, 1.3, 1.4, 1.8, 2.4, 4.6.

- The mean is 1.58 and the median is 1.3.
- The 20% two-sided trimmed mean is 1.40.
- The mean without {0.0, 2.4, 4.6}, which are considered outliers, is 1.26.
- F-distribution: when $k$ = 0.51 ($\alpha$=0.75), 1.18 ($\alpha$=0.625), and 2.36 ($\alpha$=0.5), Huber's estimates are 1.3, 1.37, and 1.40.
- Huber's estimates with $k$ = 1.28 and 1.36 are 1.338 and 1.343.

**Example 2: Lifetimes of EMT6 Cells** (Staudte, R. G. and Sheather, S. J., 1983). Mammalian cells in culture have lifetimes (cell cycle times) typically varying from 8 to 24 hours; the following is a typical data set of lifetimes in hours:

10.4, 10.9, 10.5, 8.8, 8.5, 8.7, 10.4, 7.8, 8.4, 9.1,
9.8,10.3, 9.5, 10.4, 9.0, 8.9, 7.7, 8.2, 9.1, 22.2.

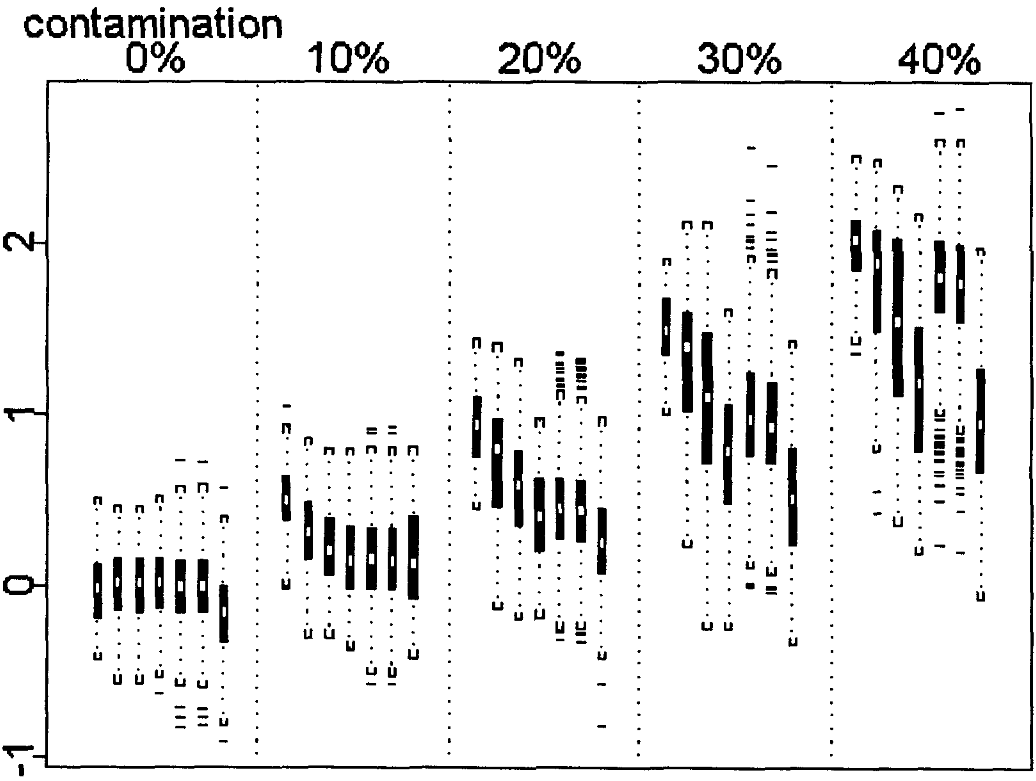
- The mean and median are 9.93 and 9.1, respectively.
- F-distribution: when $k$ = 0.53 ($\alpha$=0.75), 2.43 ($\alpha$=0.4), and 3.07 ($\alpha$=0.5) Huber's estimates are 9.25, 9.42, and 9.45.
- Huber's estimates are 9.375 when both $k$ = 1.28 and 1.36.

**Example 3: Self-Awareness Data** (Wilcox, 1997, p33). Dana (1990) conducted a study dealing with self-awareness and self-evaluation. One segment of his study measured the time subjects could keep a portion of an apparatus in contact with a specified target. The following shows some data for one of the groups:

77, 87, 88, 114, 151, 210, 219, 246, 253, 262,
296, 299, 306, 376, 428, 515, 666, 1310, 2611.

- The sample mean and the sample trimmed means with 10% and 20% trimmings are 448, 343, and 283.
- Huber's M-estimate with $k = 1.28, 1.36$ are 285.1 and 2.8853.
- F-distribution: when $k = 1.01$ ($\alpha$=0.75), 1.68 ($\alpha$=0.4), and 2.24 ($\alpha$=0.5) Huber's estimates are 227.56, 298.68, and 314.72.

**Figure 1**: Box-plots of the estimates under various contaminations; sample means, Huber's M-estimates based on $\alpha = 0.5, 0.625$ and $0.75$, and Huber's M-estimates with $k = 1.28$ and $1.36$, and medians in order from the left to the right in each section.



## References

1. Cook, R. D. (1977a). Detection of outliers in Linear Regression, *Technometrics*, 19, 15-18.

2. Cook, R. D. (1977b). Letter to the editor, *Technometrics*, 19, 348.

3. Cook, R. D., Peña, D., and Weisberg, S. (1988). The Likelihood Displacement: A Unifying Principle for Influence Measures, *Communications in Statistics - Theory and Methods*, 17(3), 623-640.

4. Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman & Hall, New York.

5. Cushny, A. R. and Peebles, A. R. (1904). The Action of Optical Isomers II, Hyoscines, *Journal of Physiology* , 32, 501-510.

6. Dana, E. (1990). *Salience of the self and salience of standards: Attempts to match self to standard.*, Unpublished Ph.D. dissertation, University of Southern California.

7. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics, the Approach Based on Influence Functions*, John Wiley & Sons, New York.

8. Huber, P. J. (1981). *Robust Statistics*, John Wiley & Sons, New York.

9. Staudte, R. G. and Sheather, S. J. (1990).*Robust Estimation and Testing*, John Wiley & Sons, New York.

10. Wilcox, R. R. (1997). *Introduction to Robust Estimation and Hypothesis Testing*, Academic Press, New York.