응용통계연구 제 14권 2호, 2001년, pp. 357-367

HUBER의 M-추정함수의 조율상수와 커널추정함수의 평활계수의 관계

박노진[1)]

요 약

Huber의 M-추정함수의 형태는 조율상수가 주어질 때 비로소 그 형태가 결정된다. 조율상수를 커널밀도함수추정량의 평활계수를 이용하여 구하여 보았고, 모의실험을 통해 기존에 사용되는 조율상수들과 그 성능을 비교하여 보았다. 그 결과 새로운 방법에 의해 구해진 조율상수가 기존의 조율상수를 사용하는 경우 보다 모의실험을 통해 얻은 추정치의 분산이 작게되는 경우가 있음을 알았다.

주요용어: 평활계수, Epanechnikov 커널, 커널 밀도함수 추정량.

1. 서론

$x_1, \cdots, x_n$이 밀도함수 $f(x-\theta)$를 갖는 확률표본일 때, Huber (1964)에 의해 제안된 위치모수의 M-추정량은 목적함수

$$\sum_{i=1}^{n} \rho(x_i - T)$$

를 최소화하는 T_n, 또는 음방정식

$$\sum_{i=1}^{n} \psi(x_i - T) = 0$$

의 해인 T_n으로 정의된다. M-추정량을 구함에 있어 M-추정함수는 로우버스트 추정에서 중심적인 역할을 감당하고 있다. 실제로 M-추정함수를 이용하기 위해서는 먼저 함수의 형태를 결정하는 조율상수(tuning constant), 변곡상수(bending constant) 그리고 절사상수(triming constant)같은 상수들의 값을 정해야 한다. 예를 들어, Huber의 M-추정함수는 다음과 같이 주어 지는데

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

$0 < k < \infty$; 함수 $\psi_k(s)$를 자료에 적용하기 위해서 적절한 조율상수 k의 값을 사전에 결정해야 한다. 이론적으로 조율상수의 결정은 추정량의 점근효율(asymptotic efficiency)을 높이고, 영향함수(influence function)의 한계(bound)는 낮추는 방향으로 이루어진다. 예를

1) (300-716) 서울특별시 용산구 한남동, 단국대학교 전산통계학과, 조교수
E-mail: rjpark@dragon.taejon.ac.kr

들어, 표준정규분포함수를 가정하는 경우, 주어진 상수 c에 대하여 점근분산$(= A/B^2)$을 최소로 하고 $\kappa^* = 1 + k^2/A = c$를 만족하는 k로 조 율상수로 삼게된다; 여기서, $A = 2\Phi(k) - 1 - 2k\phi(k) + 2k^2(1 - \Phi(k))$이고 $B = 2\Phi(k) - 1$이다. (자세한 내용은 Hampel, et al. (1986, sec. 2.5d); Staudte and Sheather (1990,p.128)를 참조하시오.) 그런데, 기저모델 확률함수의 형태를 모르는 경우가 더욱 많고, 실제로 그 형태를 안다고 할 지라도 위의 방법에 의거해 조율상수를 구한다는 것은 많은 수고를 요구한다. 더욱이 위의 방법을 사용하기 위해서는 $(1/(2\phi(0)), k/(2\Phi(k) - 1))$에 속하는 적당한 상수 c를 정해야 하는데, 이는 방법의 객관성을 떨어지게 하는 원인이기도 하다. 이러한 계산상의 어려움으로 실제로는 간편하게 1.28과 1.345가 자주 사용되는데, 이는 정규 분포를 가정하는 경우 90% 백분위수가 1.28이고 , k가 1.345일 때, Huber의 M-추정량의 점근효율이 약 0.95가 되기 때문이다.

본 논문에서 소개할 새로운 방법은 기저확률함수에 대한 정보의 존재 유무와 상관없이 조율상수를 구할 수 있는 경험적(empirical)방법이라 하겠는데, 이미 그 효능이 검증된 평활계수를 이용하여 조율상수를 구하는 방법을 소개하겠다.

2. Huber의 M-추정함수와 평활계수의 관계

f_θ를 θ로 분류되는 확률함수족이라 하자. 최소 L^2 거리 추정량 $\hat{\theta}$은 다음의 식의 해로 정의된다.

$$\nabla_\theta \mu(p, f_\theta) = \nabla_\theta \int (p(x) - f_\theta(x))^2 dx, \tag{2.1}$$

여기서 $p(x), f_\theta(x) \in L^2$를 가정하고 ∇_θ는 θ에 대한 일차미분을 표시한다고 하자. 식(1)은 아래와 같이 다시 표현된다.

$$\int (p(x) - f_\theta(x)) \nabla_\theta f_\theta(x) dx = 0. \tag{2.2}$$

만일 θ이 위치모수이면, $\int f_\theta(x) \nabla_\theta f_\theta(x) dx = (1/2)\nabla_\theta \int g_\theta^2(x) dx = 0$가 되어 식(2)는

$$\int p(x) \nabla_\theta f_\theta(x) dx = \nabla_\theta \int p(x) f_\theta(x) dx = 0. \tag{2.3}$$

가 된다.

확률함수 $f_\theta(x)$로부터 나온 확률표본 $X_1, X_2, \cdots, X_n$에 대하여 확률밀도함수 추정량

$$p(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$$

를 정의하자; 여기서 $K(\cdot)$ 는 커널함수 그리고 h는 평활계수이다 (Silverman, 1986). 식(3)은 이제

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

와 같고, 만일 T_n을 θ의 추정량으로 표시할 때, M-추정함수는

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

와 같이 된다.

$X_1,\cdots,X_n$을 좀더 구체적으로 $f(x-\theta)$를 확률함수로 갖는 독립적이며 동일한 확률표본이라고 가정하면, M-추정함수에 대하여 다음과 같은 결과를 얻게된다.

$$\begin{aligned}
&\int\frac{1}{h}K\left(\frac{x-X_i}{h}\right)f(x-\theta)dx\\
&=\int\frac{1}{h}K\left(\frac{x-\theta}{h}-\frac{X_i-\theta}{h}\right)f(x-\theta)dx\\
&=\int\frac{1}{h}K\left(u-\frac{X_i-\theta}{h}\right)hf(hu)du\\
&=\int\frac{1}{h}K(u-Z_i)hf(hu)du=\int\frac{1}{h}K(Z_i-u)hf(hu)du\\
&=\int\left[K(Z_i)-uK'(Z_i)+\frac{u^2}{2}K''(Z_i)+\cdots\right]f(hu)du\\
&=\frac{1}{h}K(Z_i)\int f(v)dv+\frac{1}{2h^3}K''(Z_i)\int v^2f(v)dv+\ldots,
\end{aligned}$$

여기서 $Z_i=(X_i-\theta)/h$. 따라서, 우리는

$$\begin{aligned}
\psi(X_i;T_n)&=\nabla_\theta\int\frac{1}{h}K\left(\frac{x-X_i}{h}\right)f(x-\theta)dx\Big|_{\theta=T_n}\\
&=\frac{1}{h}\nabla_\theta K(Z_i)\int f(v)dv+\frac{1}{2h^3}\nabla_\theta K''(Z_i)\int v^2f(v)dv+\ldots\Big|_{\theta=T_n}
\end{aligned}$$

라고 할 수 있다.

만일 $K(t)=(3/4)(1-t^2)$, $|t|<1$; 즉 커널함수가 Epanechnikov 커널 이라면, 위 ψ식의 첫 번째 항을 제외하고 모든 항들은 없어진다. 불필요한 상수계수들을 정리하면, 결국

$$\psi(X_i-\theta)=\begin{cases}(X_i-\theta) & ,\ |X_i-\theta|<h\\ 0 & ,\text{otherwise.}\end{cases}$$

가 되고, 다시 표준화를 하면

$$\psi\left(\frac{X_i-\theta}{\sigma}\right)=\begin{cases}(X_i-\theta)/\sigma & ,\ |(X_i-\theta)/\sigma|<h/\sigma\\ 0 & ,\text{otherwise,}\end{cases}$$

가 된다. 이는 *(Huber type) skipped mean* (Hampel et al., 1986 p.154)과 같은 형태를 띤다. 마지막으로 $k=h/\sigma$로 놓고 위 식을 이용하여 모수 θ에 대한 Huber의 대표적인 M-추정함수

$$\psi_k(s)=\begin{cases}s & ,\ |s|<k\\ k & ,\text{otherwise,}\end{cases}\tag{2.4}$$

를 구성 하게된다. 비슷하게, 만일 $K(t) = (3/4)\{1 - (1/5)t^2\}/\sqrt{5}$, $|t| < \sqrt{5}$ 라면, $k = \sqrt{5}(h/\sigma)$로 하는 Huber의 M-추정함수를 얻게 된다. 물론 σ를 모른다면 적당한 추정량으로 대체되어야 하고, 간단하고 로우버스트한 MAD (*median absolute deviation/0.6745*)가 적당하다고 사려된다.

이제, 평활계수 h의 최적성(optimality)에 대해 생각해보자. 우리가 관심을 갖고 있는 Huber함수의 영향함수가 본질적으로 유계함수이므로, 평활계수의 최적성을 점근분산을 최소화하는 관점에서 규명하자. Fan and Marron (1992) 은 h의 선택에 대한 상대오차의 하한이

$$\sigma^2(f) = \frac{4}{25}\left[\frac{\int (f^{(4)})^2 f}{\{\int (f'')^2\}^2} - 1\right]$$

임을 보였다. 이어서, Kim, Park and Marrron (1994)은

$$n^{1/2}(h/h_0 - 1) \Rightarrow N(0, \sigma^2(f))$$

를 만족하는 h의 최적 선택에 대하여 논하였다

이제, 점근분산을 주어진 h_0에 대하여 급수전개 하면

$$V(\psi_h, F) - V(\psi_{h_0}, F) = \\ (h - h_0)V'(\psi_h, F)|_{h=h_0} + \frac{1}{2}(h - h_0)^2 V''(\psi_h, F)|_{h=h_0} + \cdots$$

가 되고 양변에 $n^{1/2}h_0^{-1}$을 곱한 후, 고차항들을 무시한다면

$$n^{1/2}h_0^{-1}\{V(\psi_h, F) - V(\psi_{h_0}, F)\} \quad \cong \\ n^{1/2}(h/h_0 - 1)V'(\psi_h, F)|_{h=h_0} \quad \Rightarrow \quad N(0, \sigma^2(f)V'^2(\psi_h, F)|_{h=h_0})$$

를 얻게 된다. 점근분산은 (상수 $\times$ $\sigma^2(f)$)의 형태를 띠게 되고, 따라서 Huber의 M-추정함수를 이용함에 있어서, h에 대한 최적의 선택을 구하는 것은 확률함수추정 이론에서 논한 것과 같은 맥락에서 규명될 수 있다. 본 논문에서는 이미 그 최적성이 입증된 평활계수의 추정량들을 이용하여 Huber의 M-추정에 있어서 다양한 평활계수 선택들의 성능을 비교하고자 한다.

3. 모의 실험

모의실험을 위해 20, 40개의 자료로 이루어진 300개의 표본을 $\epsilon = 0.05, 0.10, 0.20$ 에 따라 혼합모형 $(1-\epsilon)N(0,1) + \epsilon N(5,1)$로부터 얻었다. 또한 20, 40개의 자료를 갖는 300개 표본을 Log-Normal(0, 1)로부터 얻었다. 이제 h_s, h_{SJ}, h_{ucv} 그리고 h_{bcv}를 Silverman (1986), Sheather and Jones (1991), *unbiased cross-validation* 그리고 *biased cross-validation*에 의한 h의 선택이라고 각각 표 시하자. 각 표본에 대하여, 모두 9종류의 위치모수 추정치들 - 표본평균; k가 h_s/MAD, $\sqrt{5}h_s$/MAD, h_{SJ}/MAD, h_{ucv}/MAD, h_{bcv}/MAD, 1.28, 1.36인 Huber의 추

정치; 중간값 - 을 구했다. 모의실험 상의 편리함을 의해 소수 두째 자리수로 정규 분포 가정하에서 Huber의 추정량의 점근효율이 0.95에 근접하는 1.36을 1.345대신에 이용하였다.

주어진 모의실험 상황에 따라 점선으로 나누어진 각 부분에 구한 추정치들의 상자 그림들을 위에 나열한 추정치들의 순서로 왼쪽에서 오른쪽으로 그림 1-a와 b에 그려 놓았다. 평활계수 추정량 h_{SJ}, h_{ucv} 그리고 h_{bcv}은 Venables and Ripley(1994)에 제시된 S-Plus함수들을 이용하여 계산했다.

5%, 10%의 혼합 모형 하에서는 모든 추정치들이 대체로 비슷한 분포를 보인다. 그러나, 혼합 비율이 큰 경우 h_s/MAD를 이용한 Huber의 M-추정치, 1.28과 1.36을 이용한 Huber의 M-추정 치, 그리고 중간값들이 다른 추정치들에 비해 로우버스트함을 알 수 있다 (그림 3.1). 분산의 경우에는 h_{SJ}/MAD, h_{ucv}/MAD 그리고 h_{bcv}/MAD 같은 최적 평활계수 추정량을 이용하는 Huber의 M-추정치들이 예상대로 다른 추정치들에 비해 작은 표본 분산을 갖는다 (표 3.1). 추정치들의 로우버스트성 정도와 분산의 감소가 서로 반비례함을 본 모의실험에서도 관찰할 수 있다. Log-normal 분포의 경우 h_s/MAD, 1.28 그리고 1.36을 이용하는 Huber의 M-추 정치들이 다른 추정치들 보다 로우버스트한 면이나 분산의 면에서 바람직한 분포를 보인다 (그림 3.2, 표 3.1). 특별히, h_s/MAD를 이용한 Huber의 M-추정치들이 1.28과 1.36을 이용한 추정치 들 보다 혼합율이 가장 큰 20%인 경우와 Log-normal의 경우에 분산이 작게 나타난다 (표 3.1).

모의 실험결과 h_s/MAD를 이용한 Huber의 M-추정치들이 1.28과 1.36을 이용하는 경우와 비교하여, 비슷하거나 때로는 우월한 성능을 보임으로, 기존에 널리 사용되는 조율상수들과 같이 사용한다면 좋은 결과를 얻을 수 있으리라 생각된다.

4. 사례 분석

모의 실험과 동일한 추정치들을 교과서들에 나오는 대표적인 예들에 적용하여 보았다. 본 논문에서 제시한 평활계수를 이용한 M-추정치들이 기존의 로우버스트 추정치들과 같이 유용하게 사용될 수 있음을 볼 수 있다.

자료 1: The Cushney and Peebles Data. 1904년 Cushney and Peebles는 "The Action of Optimal Isomers" 라는 제목 하에 관찰자료를 *Journal of Physiology*에 소개하고 있다. 아래 자료는 Staudte and Sheather(1990)의 책 97페이지의 Table 4.1에서 가져와 크기 순서로 정리하였다.

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

- 평균은 1.58 그리고 중간값은 1.3.
- 20% 양쪽 절사 평균은 1.40.
- 이상치로 의심되는 {0.0, 2.4, 4.6}을 제거한 평균은 1.26.

- Huber : 1.34 (h_s/MAD); 1.40 ($\sqrt{5}h_s$/MAD); 1.338 (1.28); 1.342 (1.345); 1.343(1.36); 1.37 (h_{SJ}/MAD); 1.54 (h_{bcv}/MAD); 1.42 (h_{ucv}/MAD).

자료 2: Lifetimes of EMT6 Cells (Staudte, R. G. and Sheather, S. J., 1983). 말라리아 세포의 배양 시간 자료이다.

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.

- 평균, 중간값이 각각 9.93, 9.1 이다.
- Huber : 9.37 (h_s/MAD); 9.47 ($\sqrt{5}h_s$/MAD); 9.38 (1.28);9.38 (1.345); 9.38 (1.36); 9.39 (h_{SJ}/MAD); 9.47 (h_{BCV}/MAD); 9.47 (h_{UCV}/MAD).

자료 3: Self-Awareness Data (Wilcox, 1997, p33). Dana (1990)는 자아인식과 자아평가에 대한 연구를 실시했다. 그의 연구의 한 부분은 대상들이 특정한 목표에 집중하는 시간을 재는 것이었다. 아래의 자료는 한 집단에 대한 집중 시간에 자료이다.

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

- 평균, 중간값, 10% 그리고 20% 절사평균은 각각 448, 262, 343, 그리고 283 이다.
- Huber : 311.02 (h_s/MAD); 362.19 ($\sqrt{5}h_s$/MAD); 285.16 (1.28); 288.06 (1.345); 288.54 (1.36); 313.26 (h_{SJ}/MAD); 411.5 (h_{BCV}/MAD); 327.99 (h_{UCV}/MAD).

5. 맺는 말

Huber의 M-추정함수의 조율상수를 구함에 있어 그 방법의 복잡성으로, 두 상수 1.28과 1.345이 널리 사용되고 있다. 새로이 제안된 조율상수는 이미 많은 연구가 되어 있는 평활계수를 사용하여 계산의 용의성과 객관성을 보장함을 보였다.

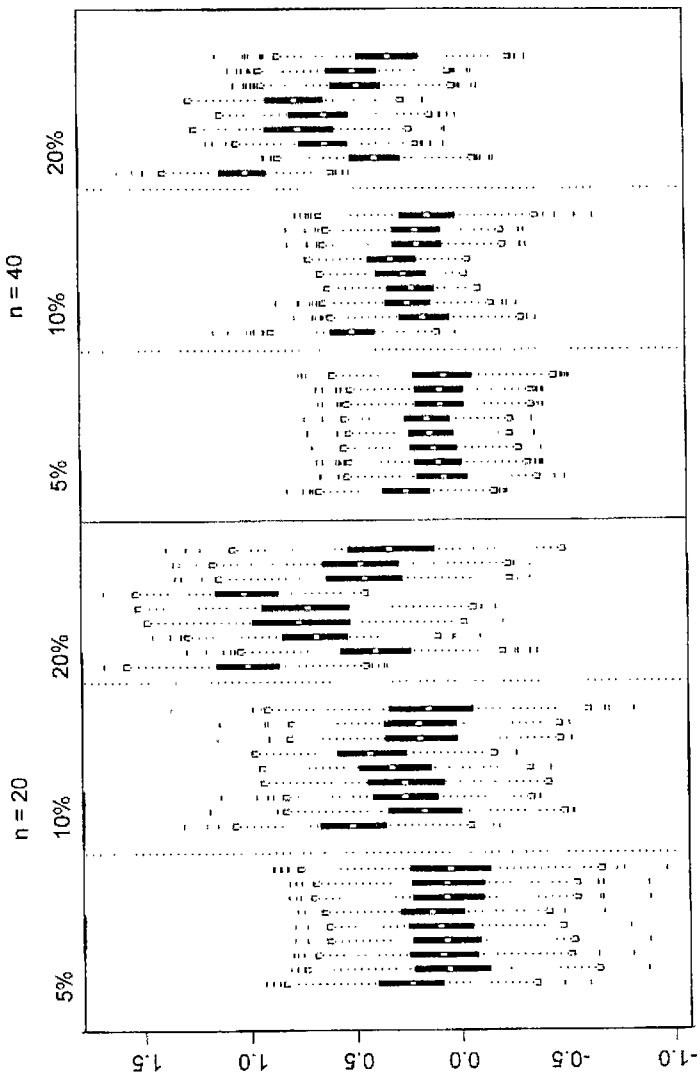

그림 2.1(a): 혼합정규분포에 근거한 모의 실험에서 구한 추정치들의 상자도표

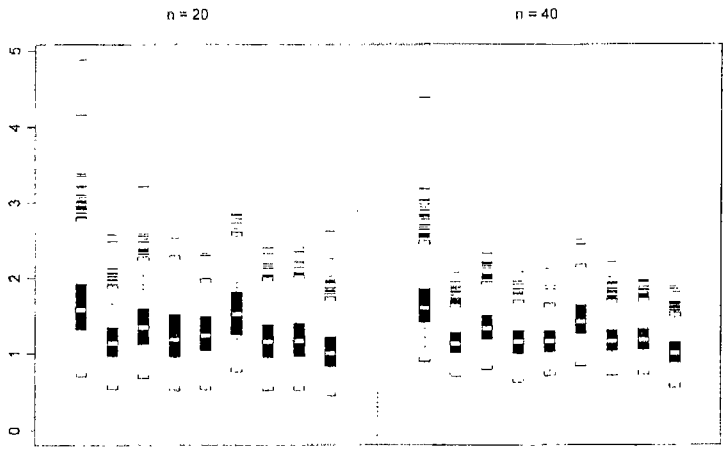

그림 2.1(b): Log-normal분포에 근거한 모의 실험에서 구한 추정치들의 상자도표

표 2.1: 추정치들의 표본평균과 표본분산

	$n=20$			$n=40$			Log-normal	
	5%	10%	20%	5%	10%	20%	$n=20$	$n=40$
Sample	0.2395	0.5087	1.0030	0.2507	0.4962	1.0023	1.6533	1.6546
Mean	(0.0518)	(0.0496)	(0.0464)	(0.0262)	(0.0234)	(0.0241)	(0.1289)	(0.1241)
Huber's	0.0569	0.1668	0.3942	0.0763	0.1561	0.3839	1.1725	1.1541
w/h_s	(0.0635)	(0.0612)	(0.0624)	(0.0318)	(0.0296)	(0.0315)	(0.0805)	(0.0412)
Huber's	0.0891	0.2612	0.6826	0.0967	0.2300	0.6231	1.3858	1.3579
$w/\sqrt{5}h_s$	(0.0551)	(0.0528)	(0.0540)	(0.0277)	(0.0249)	(0.0266)	(0.1200)	(0.0584)
Huber's	0.0729	0.2611	0.7435	0.1218	0.2193	0.7276	1.2359	1.1659
w/h_{SJ}	(0.0614)	(0.0697)	(0.1089)	(0.0232)	(0.0205)	(0.0548)	(0.1417)	(0.0675)
Huber's	0.1002	0.3086	0.7251	0.1334	0.2601	0.6265	1.2706	1.1721
w/h_{UCV}	(0.0591)	(0.0625)	(0.0919)	(0.0272)	(0.0247)	(0.0425)	(0.1012)	(0.0478)
Huber's	0.1391	0.4186	1.0078	0.1520	0.3046	0.7614	1.5588	1.4532
w/h_{BCV}	(0.0565)	(0.0543)	(0.0486)	(0.0267)	(0.0233)	(0.0424)	(0.1663)	(0.0742)
Huber's	0.0670	0.1853	0.4473	0.0864	0.1836	0.4677	1.1811	1.1845
$w/1.28$	(0.0595)	(0.0594)	(0.0706)	(0.0332)	(0.0274)	(0.0332)	(0.0889)	(0.0460)
Huber's	0.0694	0.1914	0.4654	0.0887	0.1897	0.4868	1.1963	1.2005
$w/1.36$	(0.0589)	(0.0589)	(0.0715)	(0.0336)	(0.0272)	(0.0336)	(0.0912)	(0.0472)
Median	0.0509	0.1427	0.3250	0.0697	0.1371	0.3225	1.0370	1.0237
	(0.0797)	(0.0818)	(0.0933)	(0.0420)	(0.0390)	(0.0465)	(0.0774)	(0.0413)

참고문헌

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

[2] 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.

[3] Fan, J. and Marron, J.S. (1992). On optimal data-based bandwidth selection in kernel density estimation, *Biometrika* **78**, 263-269.

[4] Hampel, F.R., Rousseeuw, P.J. and Ronchetti, E. (1981). The Change-of-Variance Curve And Optimal Redescending M-estimators, *Journal of the American Statistical Association* **76**, 643-648.

[5] 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.

[6] Huber, P.J. (1964). Robust estimation of location parameters. *Annals of Mathematical Statistics*, **35**, 73-101.

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

[8] Kim, W.C., Park, B.U. and Marron, J.S. (1994). Asymptotically best bandwidth selectors in kernel density estimation, *Statistics and Probability Letters* **19**, 119-127.

[9] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

[10] Sheather, S.J. and Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society, Ser. B* **53**, 683-690.

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

[12] Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.

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

[2000년 11월 접수, 2001년 6월 채택]

The Bending Constant in Huber's Function in Terms of a Bandwidth in Density Estimator

Ro Jin Pak[1)]

ABSTRACT

Among many M-estimating functions, we have focused on Huber's M-estimation function. The shape is determined by a bending constant, and we have shown that the bending constant can be expressed in terms of the bandwidth for a kernel based density estimator. We have compared by simulations how M-estimators for the location parameter are distributed according to various choices of bandwidths.

Keywords: Bandwidth; Epanechnikov kernel; Kernel; Density estimator.

1) Assistant Professor, Department of Computer Sciences and statistics, Dankook University.
E-mail: rjpark@dragon.taejon.ac.kr