Discussiones Mathematicae
Probability and Statistics 38 (2018) 21–30
doi:10.7151/dmps.1201

ON DATA-DEPENDENT TUNING CONSTANT OF THE HUBER FAMILY FOR ROBUST ESTIMATOR

Agnieszka Kulawik

Institute of Mathematics
University of Silesia in Katowice
40–007 Katowice, Bankowa 14, Poland

e-mail: Agnieszka.Kulawik@us.edu.pl

and

Stefan Zontek

Faculty of Mathematics, Computer Science and Econometrics
University of Zielona Góra
65–516 Zielona Góra, Szafrana 4A, Poland

e-mail: S.Zontek@wmie.uz.zgora.pl

Abstract

On a base of the location model an idea of a data-dependent choice of tuning constant (truncation level) for a robust estimator is presented. The method uses maximum likelihood estimator in a new model with tuning constant as a nuisance parameter. Some results of computer simulation study are given.

Keywords: Huber's function, shift parameter, tuning constant, robust estimator, asymptotic normality.

2010 Mathematics Subject Classification: 62F35, 62F10.

1. Introduction

Let us consider the family of normal distributions with unknown expected value θ and known variance equal to 1. The classic criterion function defining estimator of θ is the likelihood function. As the solution we obtain the sample mean. It is well known that the sample mean is the best unbiased estimator of θ under model assumption. However, it loses its optimal properties in a neighborhood of the model. Even one observation can have arbitrarily large influence on both the expected value and the variance of the sample mean. To avoid such undesirable property, Huber, 1981, has proposed to replace the quadratic function appearing in the likelihood function by the following function

$$\phi_t(x) = \begin{cases} \frac{1}{2}x^2, & |x| \le t, \\ t|x| - \frac{1}{2}t^2, & |x| > t, \end{cases} \tag{1}$$

where t is a positive number. The choice of the tuning constant t has an influence on properties of Huber's estimator. Under model assumption, if the constant t is sufficiently large, then Huber's estimator is comparable with the sample mean, which has the smallest variance among unbiased estimators. For t sufficiently small, Huber's estimator is comparable with the sample median, which is an unbiased estimator but has larger variance. The situation is changing, when data are contaminated. If the level of contamination increases, then the bias and the variance increase much faster for large value of t than for small. The choice of the constant t can be interpreted as our belief in the deviation from the assumptions. So the value t reflects the level of possibility of having observations, which are outside the model. Our proposition is to include t as a nuisance parameter in new proposed model, which takes into account derogation from the normal model. So if data come from the normal model, then the estimator of t should take values sufficiently large and should choose t sufficiently small depending on the level of contamination. In other words 'optimal' choice of value t is directly connected with unknown level of contamination of the normal model.

The tunning constant for the Huber's function has appeared in several papers. In most cases $t = 1.345$ has been taken (for example Alamgir *et al.* [1]). The value of the tunning constant has also been mentioned by Venables and Ripley [4] as the constant for which about 95% efficiency at the normal is obtained. Another values of tunning constant have also been used. Cantoni and Ronchetti [2] present an example of computing robust estimates of parameters using the tuning constant $t = 1.2$. The mentioned values of the tuning constant were chosen be researchers. An idea of the data-dependent tuning constant has already appeared in literature. In the linear regression model You-Gan Wang *et al.* [5] have proposed a method of obtaining the tunning constant so that the asymptotic efficiency is maximized. We propose another method of obtaining data-dependent tuning constant t. We are concentrated on a simpler model, namely the location normal model with known variance, to verify effectiveness of the proposed method.

This paper consists of five sections. In Section 2 basic information concerning Huber's estimator have been presented. In Section 3 our method of data-dependent choice of tuning constant is described. The section presents also the theorem concerning asymptotic normality of the estimator that has been obtained in this section. In Section 4 there can be found computer simulation results for data from new model described in Section 3. Section 5 presents computer simulation results for estimation of location parameter in normal model for both model and contaminated data. The results concern seven estimators of location parameter (maximum likelihood estimator, four Huber's estimators, median and the estimator from Section 3) and the estimator for tuning constant (described in Section 3).

2. Huber's estimator

Let $\varphi(\cdot)$ denote the density function of the standard normal distribution that is

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right).$$

We consider the normal model with shift parameter θ described by

$$\mathcal{N} = \{\varphi(x-\theta) : \theta \in \mathbb{R}\}.$$

Using the concept of maximum likelihood estimation (m.l.e.), the estimator of θ is defined by

$$\hat{\theta}_n = \operatorname{argmax}_{\theta\in\mathbb{R}} \sum_{i=1}^{n} -\frac{1}{2}(X_i-\theta)^2$$

for independent random variables $X_1, \ldots, X_n$ having distribution $N(\theta, 1)$. Huber [3] has proposed a method of estimation, which in the spirit of m.l.e. can be defined as

$$\hat{\theta}_{n,t} = \operatorname{argmax}_{\theta\in\mathbb{R}} \sum_{i=1}^{n} -\phi_t(X_i-\theta), \tag{2}$$

where ϕ_t is the function given by (1).

Huber's estimator (2) is Fisher consistent and Fréchet differentiable. Cases $t = 0$ and $t = +\infty$ correspond to the median and the maximum likelihood estimator, respectively. The asymptotic distribution of $\sqrt{n}(\hat{\theta}_{n,t} - \theta)$ is normal (see Huber [3]) with expectation zero (at the model) and the variance equal to

$$\int \mathrm{IF}_t^2(x|\theta)\,\mathrm{d}F_\theta(x),$$

where $\mathrm{IF}_t(\cdot|\theta) \colon \mathbb{R} \to \mathbb{R}$ stands for the influence function of the form

$$\mathrm{IF}_t(x|\theta) = \frac{\phi_t'(x-\theta)}{\int \phi_t''(x-\theta)\phi(x-\theta)\,\mathrm{d}x}.$$

3. A method of data-dependent choice of tuning constant

In this section some theory of estimating location parameter θ in normal model using a data-dependent tuning constant t for Huber's estimator will be presented. An idea of this method of estimation is to treat the vector $(\theta, t)^T$ as a vector of parameter in a new model defined by

$$\bar{\mathcal{N}} = \{f_t(x-\theta) : \theta \in \mathbb{R}, t \in (0, +\infty)\}, \tag{3}$$

where f_t is density function of form

$$f_t(x) = \frac{\exp(-\phi_t(x))}{\int \exp(-\phi_t(u))\,\mathrm{d}u} \tag{4}$$

and ϕ_t stands for Huber's function (1).

Let $M\colon (0, +\infty) \to (0, +\infty)$ be a function given for $t > 0$ by

$$M(t) = \int \exp(-\phi_t(u))\,\mathrm{d}u.$$

Lemma 1. *For $t > 0$ we have*

$$M(t) = \sqrt{2\pi}\left(2F_{N(0,1)}(t) - 1\right) + \frac{2}{t}\exp\left(-\frac{1}{2}t^2\right), \tag{5}$$

where $F_{N(0,1)}\colon \mathbb{R} \to (0,1)$ stands for the cumulative distribution function of standard normal distribution.

Proof. For $t > 0$ we have

$$\int_{-t}^{t} \exp(-\phi_t(x))\,\mathrm{d}x = \int_{-t}^{t} \exp\left(-\frac{1}{2}x^2\right)\mathrm{d}x = \sqrt{2\pi}\left(2F_{N(0,1)}(t) - 1\right)$$

and

$$\int_{t}^{+\infty} \exp(-\phi_t(x))\,\mathrm{d}x = \int_{t}^{+\infty} \exp\left(-tx + \frac{1}{2}t^2\right)\mathrm{d}x = \frac{1}{t}\exp\left(-\frac{1}{2}t^2\right),$$

what ends the proof. ∎

Remark 2. Function M is decreasing. Moreover, for $t \to +\infty$, we get $M(t) \to \sqrt{2\pi}$ and

$$f_{+\infty}(x-\theta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\theta)^2\right),$$

what is the density function of the distribution $N(\theta, 1)$. For $t \geq 3$ the difference $M(t) - \sqrt{2\pi}$ is smaller than 0.0007 and the probability that we observe a value greater than 3 (in case of $\theta = 0$) is smaller than 0.0015. So in practice, data from distribution with density function $\phi_t(x-\theta)$, when $t > 3$, are very close to the normal one.

Let $X_1, \dots, X_n$ be independent, identically distributed random variables with density function $f_t \in \bar{\mathcal{N}}$. The m.l.e. of η is given by

$$\tilde{\eta}_n = \left(\tilde{\theta}_n, \tilde{t}_n\right)^T = \operatorname{argmax}_{(\theta,t)^T \in \mathbb{R}\times(0,+\infty)} \sum_{i=1}^{n} -\left(\ln M(t) + \phi_t(X_i - \theta)\right). \tag{6}$$

Theorem 3. *The asymptotic distribution of* $\sqrt{n}(\tilde{\eta}_n - \eta)$ *is normal with expectation zero* (*at the model*) *and the covariance matrix equal*

$$\begin{bmatrix} D_1(t)^{-1} & 0 \\ 0 & D_2(t)^{-1} \end{bmatrix},$$

where

$$D_1(t) = \frac{\sqrt{2\pi}}{M(t)} \left(2F_{N(0,1)}(t) - 1\right)$$

and

$$D_2(t) = \frac{4}{t^3 M^2(t)} \exp\left(-\frac{1}{2}t^2\right) \left(\frac{1}{t}\exp\left(-\frac{1}{2}t^2\right) + \sqrt{2\pi}\left(2F_{N(0,1)}(t) - 1\right)\right).$$

Proof. The model $\bar{\mathcal{N}}$ given by (3) does not satisfy the regularity conditions (see for example Zacks [6]) because the function

$$\psi_t(x) = \phi_t'(x) = \begin{cases} -t, & x < -t, \\ x, & |x| \le t, \\ t, & x > t. \end{cases} \tag{7}$$

is not differentiable at $x = -t$ and $x = t$. To avoid the problem we can smoothly modify function ϕ (using third degree polynomial, for example) in a neighbourhood of points $-t$ and t. Moreover, if neighbourhoods are sufficiently small, then the modification has an insignificant influence on values of estimator (6).

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be neighbourhoods of points $-t$ and t, respectively and let $\tilde{\phi}_t \colon \mathbb{R} \to [0, +\infty)$ be a function such that $\tilde{\phi}_t(x) = \phi_t(x)$ for $x \in \mathcal{D} = \mathbb{R} \setminus (\mathcal{H}_1 \cup \mathcal{H}_2)$ and $\tilde{\phi}_t$ satisfies regularity conditions. For $x - \theta \in \mathcal{D}$ we have

$$\frac{\partial^2}{\partial\theta\partial t} \ln \frac{\exp(-\tilde{\phi}_t(x-\theta))}{M(t)} = \begin{cases} 1, & x - \theta > t, \\ 0, & |x - \theta| \le t, \\ -1, & x - \theta < -t, \end{cases}$$

$$\frac{\partial^2}{\partial\theta^2} \ln \frac{\exp(-\tilde{\phi}_t(x-\theta))}{M(t)} = \begin{cases} -1, & |x - \theta| \le t, \\ 0, & |x - \theta| > t \end{cases}$$

and

$$\frac{\partial^2}{\partial t^2} \ln \frac{\exp(-\tilde{\phi}_t(x-\theta))}{M(t)} = \frac{\frac{4}{t^4}\exp(-t^2) - \exp\left(-\frac{1}{2}t^2\right)\left(\frac{2}{t} + \frac{4}{t^3}\right) M(t)}{M^2(t)} + \begin{cases} 0, & |x - \theta| \le t, \\ 1, & |x - \theta| > t. \end{cases}$$

Hence

$$\mathbb{E}_\eta\left[-\frac{\partial^2}{\partial\theta\partial t}\ln f_t(X-\theta)\right]=0,$$

$$D_1(t)=\mathbb{E}_\eta\left[-\frac{\partial^2}{\partial\theta^2}\ln f_t(X-\theta)\right]=\frac{\sqrt{2\pi}}{M(t)}\left(2F_{N(0,1)}(t)-1\right)$$

and

$$\begin{aligned}D_2(t)&=\mathbb{E}_\eta\left[-\frac{\partial^2}{\partial t^2}\ln f_t(X-\theta)\right]\\&=\frac{4}{t^3M^2(t)}\exp\Big(-\frac{1}{2}t^2\Big)\left(\frac{1}{t}\exp\Big(-\frac{1}{2}t^2\Big)+\sqrt{2\pi}\left(2F_{N(0,1)}(t)-1\right)\right).\end{aligned}$$

■

4. Simulation results for the new model

In simulation study m.l.e. of vector of parameters $\eta=(\theta,t)^T$ of distribution from model $\bar{\mathcal{N}}$ defined by (3) was considered. The computer simulations concerned the following cases:

$$\mathcal{C}=\big\{(\theta,t)^T:\theta=0,\ t\in\{0.25,0.50,0.75,1.00,1.25,1.50,1.75,2.00,2.50,3.00,4.00,6.00\}\big\}.$$

For each $\eta\in\mathcal{C}$, $n=100$ observations were generated randomly. The procedure was repeated $N=5000$ times. For each $i\in\{1,\dots,N\}$, estimates $\tilde{\theta}_{100,i}$ and $\tilde{t}_{100,i}$ of parameters θ and t were obtained using m.l.e. Table 1 presents results of the computer simulations. The table consists of six columns. The first and the fourth column present vectors of parameters from $\mathcal{C}$, respectively. The second and the fifth column stand for results concerning parameter θ. The third and the sixth column — for results concerning parameter t. For each density function the results have been grouped into three rows. The first one is for the averages of the 5000 estimates, namely $\tilde{\theta}=\frac{1}{N}\sum_{i=1}^{N}\tilde{\theta}_{100,i}$ and $\tilde{t}=\frac{1}{N}\sum_{i=1}^{N}\tilde{t}_{100,i}$. The second row stands for sample standard deviations for the estimates (numbers in round brackets). The third row gives the "true" asymptotic values for the standard deviations (numbers in square brackets).

The discussion falls naturally into two parts. The first part concerns estimation of θ. All averages for parameter θ are close to the model one ($\theta=0$). Moreover, the greater t is, the smaller sample standard deviation for θ is. The cases with $t\geq 3$ are nearly the same as results getting from m.l.e. of parameter θ for samples from $N(\theta,1)$. What is more, sample standard deviations are very close to the asymptotic ones.

The second part of the discussion concerns estimation of parameter t. These results are more complicated. For $t \in \{0.25, 0.50, 0.75, 1.00\}$ the averages for parameter t are close to the model ones. For the cases also sample standard deviations are close to the asymptotic ones. Clear differences between averages and model values for parameter t can be noticed for the cases $t \in \{1.25, 1.50, 1.75, 2.00, 2.50\}$. The differences increase from 0.128 ($t = 1.25$) to 1.27 ($t = 2.50$). It can be said that the bias of considered estimator of parameter t is greater then zero. Clear differences are also visible between sample standard deviations and asymptotic standard deviations. The biggest difference can be noticed for $t = 1.5$. For the case of t, sample standard deviation is about 300% higher than corresponding asymptotic one. For $t \in \{1.50, 1.75, 2.00, 2.50\}$ the differences between sample standard deviations and asymptotic ones decrease (for $t = 2.50$ the difference is close to zero).

Results concerning $t \in \{3.00, 4.00, 6.00\}$ have been also included. Namely, samples then seem to be normally distributed $N(\theta, 1)$ (see also Remark 2). That is confirmed by all obtained values for estimator of parameter θ and by averages for parameter t (all the values are close to each other). For the cases of t, Remark 2 can also be an explanation for increasing differences between sample standard deviations and asymptotic ones. The differences would be smaller for samples of greater size.

Table 1. Results for the new model data.

$(\theta, t)^T$	$\tilde{\theta}$	$\tilde{t}$	$(\theta, t)^T$	$\tilde{\theta}$	$\tilde{t}$
$(0, 0.25)^T$	0.002 (0.429) [0.408]	0.254 (0.026) [0.025]	$(0, 1.75)^T$	0 (0.106) [0.105]	2.487 (1.406) [0.408]
$(0, 0.50)^T$	0.001 (0.221) [0.216]	0.508 (0.053) [0.051]	$(0, 2.00)^T$	0 (0.103) [0.103]	3.090 (1.564) [0.620]
$(0, 0.75)^T$	-0.001 (0.159) [0.157]	0.765 (0.087) [0.082]	$(0, 2.50)^T$	0 (0.101) [0.101]	3.770 (1.490) [1.499]
$(0, 1.00)^T$	-0.004 (0.130) [0.131]	1.023 (0.155) [0.123]	$(0, 3.00)^T$	0 (0.101) [0.100]	3.978 (1.445) [3.906]
$(0, 1.25)^T$	-0.002 (0.117) [0.117]	1.378 (0.581) [0.183]	$(0, 4.00)^T$	0 (0.102) [0.100]	4.013 (1.417) [34.577]
$(0, 1.50)^T$	0 (0.111) [0.110]	1.912 (1.094) [0.272]	$(0, 6.00)^T$	0 (0.100) [0.100]	4.013 (1.412) [9427.405]

5. Simulation results for estimation of location parameter in normal model

In previous section we discussed results concerning m.l.e. of vector of parameters $\eta = (\theta, t)^T$ of distribution from model defined by (3), where f_t is density function of form (4). In the following section the estimator is used for samples normally distributed with variance equal 1. Both uncontaminated and contaminated data are involved. The estimator allows us to estimate true expected value of normal distribution. The estimating of θ depends on tuning constant t which is also estimated. Simulation results will show that the greater level of contamination is, the smaller tuning constant we get.

In simulation study the following levels of contamination have been considered:

$$l \in \{0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10\}.$$

For each level of contamination a sample that consists of $n = 100$ elements was generated randomly. More precisely, $(1 - l) * 100$ elements have been taken from $N(0, 1)$ and $l * 100$ elements — from $N(10, 25)$, where 25 denotes variance of the normal distribution. The procedure was repeated 5000 times. We treat $\theta = 0$ as true expected value and $N(10, 25)$ is a contaminating distribution. The following estimators were considered:

- m.l.e.(θ) for the arithmetic mean,
- Huber's estimators (2) for $t \in \{2, 1.5, 1, 0.75\}$ ($\hat{\theta}_{t=2}$, $\hat{\theta}_{t=1.5}$, $\hat{\theta}_{t=1}$, $\hat{\theta}_{t=0.75}$),
- median,
- $(\tilde{\theta}, \tilde{t})^T$ as the estimator (6).

The results are presented in Table 2. The table consists of nine columns. The first one gives the level of contamination. The other columns are for results concerning mentioned estimators. For each level of contamination, the results are presented in two rows. The first row is for the averages of the 5000 estimates. The second row (numbers in round brackets) stands for sample standard deviations for the estimates.

Computer simulation results concern most of all seven estimators of expected value of distribution $N(0, 1)$. One of them (m.l.e.) is not robust - it is not worth taking it as an estimator of expected value in case of contaminated data ($l > 0$). The other estimators ($\hat{\theta}_{t=2}$, $\hat{\theta}_{t=1.5}$, $\hat{\theta}_{t=1}$, $\hat{\theta}_{t=0.75}$, $\tilde{\theta}$, median) are robust. In case of model data ($l = 0$) averages are close to $\theta = 0$ but median has the largest sample standard deviation and estimators $\hat{\theta}_{t=2}$, $\hat{\theta}_{t=1.5}$, $\hat{\theta}_{t=1}$, $\hat{\theta}_{t=0.75}$, $\tilde{\theta}$ have similar sample standard deviations. For contaminated data bias of estimator $\tilde{\theta}$ seems to be very close to median's one but sample standard deviation for $\tilde{\theta}$ is smaller than median's one. However, the bigger level of contamination is, the closer to median estimator $\tilde{\theta}$ is. When level of contamination increases, the bias of estimator $\tilde{\theta}$ gets greater slowly and biases of Huber's estimators increase faster. Sample standard deviations for Huber's estimators seem no to change significantly. What is more, for the level of contamination where $\tilde{t} \approx 1.5$ ($l = 0.01$), estimators $\hat{\theta}_{t=1.5}$ and $\tilde{\theta}$ are almost the same. Similar results can be noticed in case of $\tilde{t} \approx 1$ ($l = 0.03$, $\hat{\theta}_{t=1}$ and $\tilde{\theta}$ are comparable) and in case of $\tilde{t} \approx 0.75$ ($l = 0.06$, $\hat{\theta}_{t=0.75}$ and $\tilde{\theta}$ are comparable).

In case of Huber's estimator we have to choose the tuning constant by ourselves. The estimator $\tilde{\theta}$ is based on Huber's estimator but the choice of tuning constant is chosen without any inference of the researcher. The choice of the tunning constant depends on the data related to the level of contamination. The greater level of contamination is, the smaller tuning constant t has been chosen.

Table 2. Data from $N(0,1)$ contaminated by $N(10,25)$

l	m.l.e.(θ)	$\hat{\theta}_{t=2}$	$\hat{\theta}_{t=1.5}$	$\hat{\theta}_{t=1}$	$\hat{\theta}_{t=0.75}$	median	$\tilde{\theta}$	$\tilde{t}$
0	0	0	0	0	0	0	0	4.021
	(0.100)	(0.101)	(0.102)	(0.105)	(0.108)	(0.123)	(0.101)	(1.418)
0.01	0.100	0.020	0.017	0.014	0.014	0.014	0.016	1.551
	(0.110)	(0.102)	(0.103)	(0.107)	(0.109)	(0.124)	(0.105)	(0.786)
0.02	0.203	0.041	0.034	0.029	0.028	0.025	0.031	1.153
	(0.121)	(0.101)	(0.103)	(0.106)	(0.109)	(0.125)	(0.105)	(0.252)
0.03	0.299	0.062	0.051	0.043	0.041	0.038	0.044	1.005
	(0.132)	(0.105)	(0.107)	(0.110)	(0.113)	(0.128)	(0.110)	(0.154)
0.04	0.402	0.083	0.069	0.058	0.055	0.049	0.057	0.894
	(0.142)	(0.105)	(0.106)	(0.109)	(0.113)	(0.130)	(0.111)	(0.108)
0.05	0.502	0.105	0.087	0.074	0.069	0.062	0.070	0.815
	(0.148)	(0.103)	(0.105)	(0.108)	(0.111)	(0.128)	(0.110)	(0.092)
0.06	0.602	0.128	0.106	0.090	0.084	0.077	0.084	0.754
	(0.157)	(0.106)	(0.107)	(0.111)	(0.113)	(0.130)	(0.114)	(0.083)
0.07	0.696	0.149	0.123	0.104	0.097	0.089	0.096	0.705
	(0.164)	(0.107)	(0.108)	(0.111)	(0.114)	(0.130)	(0.115)	(0.074)
0.08	0.798	0.171	0.142	0.120	0.112	0.102	0.110	0.658
	(0.175)	(0.108)	(0.109)	(0.112)	(0.115)	(0.132)	(0.116)	(0.068)
0.09	0.895	0.193	0.159	0.134	0.126	0.114	0.122	0.620
	(0.176)	(0.105)	(0.106)	(0.109)	(0.113)	(0.130)	(0.115)	(0.062)
0.10	0.998	0.219	0.181	0.154	0.144	0.132	0.139	0.585
	(0.184)	(0.108)	(0.109)	(0.112)	(0.115)	(0.133)	(0.118)	(0.057)

References

[1] Alamgir, Ali, A., S.A. Khan, D.M. Khan and U. Khalil, *A new efficient redescending M-estimator: Alamgir redescending M-estimator*, Res. J. Recent Sci. **2** (8) (2013) 79–91.

[2] E. Cantoni and E. Ronchetti, *Robust inference for generalized linear models*, J. Amer. Stat. Assoc. **96** (455) (2001) 1022–1030.

[3] P.J. Huber, Robust Statistics (Wiley, New York, 1981).

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

[5] Y. Wang, X. Lin, M. Zhu and Z. Bai, *Robust estimation using the Huber function with a data-dependent tuning constant*, J. Comput. Graph. Stat. **16** (2) (2007) 468–481.

[6] S. Zacks, The Theory of Statistical Inference (John Wiley and Sons, New York, 1971).

Received 15 May 2018
Revised 8 August 2019