

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspiRobust median estimator in logistic regression[☆]T. Hobza^a, L. Pardo^{b,*}, I. Vajda^c^aFaculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, Czech Republic^bDepartment of Statistics and O.R., Complutense University of Madrid, Madrid, Spain^cInstitute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic

ARTICLE INFO

Article history:

Received 26 January 2006

Received in revised form

11 September 2007

Accepted 1 February 2008

Available online 28 March 2008

MSC:

62F10

62F12

62F35

Keywords:

Logistic regression

MLE

Morgenthaler estimator

Bianco and Yohai estimator

Croux and Haselbroeck estimator

Median estimator

Consistency

Asymptotic normality

Robustness

ABSTRACT

This paper introduces a median estimator of the logistic regression parameters. It is defined as the classical L_1 -estimator applied to continuous data Z_1, \dots, Z_n obtained by a statistical smoothing of the original binary logistic regression observations Y_1, \dots, Y_n . Consistency and asymptotic normality of this estimator are proved. A method called enhancement is introduced which in some cases increases the efficiency of this estimator. Sensitivity to contaminations and leverage points is studied by simulations and compared in this manner with the sensitivity of some robust estimators previously introduced to the logistic regression. The new estimator appears to be more robust for larger sample sizes and higher levels of contamination.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction and basic concepts

In this paper we study estimation of Euclidean parameters $\beta_0 \in \mathbb{R}^d$ in logistic regression based on independent observations $Y_1 \sim \text{Be}(\pi_1), \dots, Y_n \sim \text{Be}(\pi_n)$ where the Bernoulli parameters $\pi_i = \Pr(Y_i = 1) = EY_i$ depend on β_0 and regressors $\mathbf{x}_1, \dots, \mathbf{x}_n$ from \mathbb{R}^d ,

$$\pi_i = \pi_i(\beta_0) = \pi(\mathbf{x}_i^T \beta_0). \quad (1.1)$$

Here and elsewhere in the paper, $\mathbf{x}^T \beta = \sum_{j=1}^d x_j \beta_j$ denotes the scalar product of arbitrary vectors $\mathbf{x} = (x_1, \dots, x_d)^T$, $\beta = (\beta_1, \dots, \beta_d)^T$ from \mathbb{R}^d and

$$\pi(t) = \frac{e^t}{1 + e^t} \quad (1.2)$$

[☆] This work was supported by the grants MTM 2006-06872, MŠMT 1M 0572 and MPO FI-IM3/136.

* Corresponding author.

E-mail address: Leandro_Pardo@mat.ucm.es (L. Pardo).

the standard logistic function. The MLE $\beta_n = \beta_n(Y_1, \dots, Y_n)$ of β_0 minimizes the deviances (negative scores)

$$\mathcal{D}_n(\beta) = \sum_{i=1}^n d_i(\beta)$$

of the sample $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ where

$$d_i(\beta) = -\ln[\pi_i(\beta)^{Y_i}(1 - \pi_i(\beta))^{1-Y_i}] = -Y_i \ln \pi_i(\beta) - (1 - Y_i) \ln(1 - \pi_i(\beta)) \tag{1.3}$$

are the deviances (negative scores) of individual observations Y_i for $\pi_i(\beta) = \pi(\mathbf{x}_i^T \beta)$. Thus

$$\beta_n = \arg \min \mathcal{D}_n(\beta) = \arg \min \sum_{i=1}^n d_i(\beta). \tag{1.4}$$

Notice that the expected deviances are of the form

$$E\mathcal{D}_n(\beta) = \sum_{i=1}^n E d_i(\beta) = \sum_{i=1}^n [I(\pi_i(\beta_0) \parallel \pi_i(\beta)) + H(\pi_i(\beta_0))], \tag{1.5}$$

where

$$I(p_0 \parallel p) = p_0 \ln \frac{p_0}{p} + (1 - p_0) \ln \frac{1 - p_0}{1 - p}$$

is the non-negative information divergence of the Bernoulli models $\text{Be}(p_0)$ and $\text{Be}(p)$ which reduces to zero iff $p_0 = p$ and

$$H(p_0) = -p_0 \ln p_0 - (1 - p_0) \ln(1 - p_0)$$

is the entropy of the model $\text{Be}(p_0)$ which does not depend on p . Therefore β_0 is the unique $\arg \min E\mathcal{D}_n(\beta)$ unless there exists $\beta \neq \beta_0$ with the property

$$\pi_i(\beta) = \pi_i(\beta_0) \quad (\text{i.e. } \mathbf{x}_i^T(\beta - \beta_0) = 0) \quad \text{for all } 1 \leq i \leq n.$$

This leads to the consistency of MLE unless all regressors $\mathbf{x}_1, \mathbf{x}_2, \dots$ are on a hyperplane in \mathbb{R}^d (cf. Andersen, 1990; Agresti, 2002; Pardo et al., 2006 and references therein).

However, this optimistic picture dramatically changes as soon as the true models $\text{Be}(\pi_i)$ for $\pi_i = \pi(\mathbf{x}_i^T \beta_0)$ are positively ε -contaminated by some alternative Bernoulli models $\text{Be}(p_i)$, e.g. by $\text{Be}(1 - \pi_i)$. Then the true models are

$$\text{Be}(\pi_i + \varepsilon(1 - 2\pi_i)) = (1 - \varepsilon) \text{Be}(\pi_i) + \varepsilon \text{Be}(1 - \pi_i) \tag{1.6}$$

for some $0 < \varepsilon < 1$ which differ from $\text{Be}(\pi_i)$ unless $\pi_i = \frac{1}{2}$. Hence β_0 remains to be the unique $\arg \min E\mathcal{D}_n(\beta)$ for all sufficiently large n only in the trivial case where

$$\pi(\mathbf{x}_i^T \beta_0) = 1/2 \quad (\text{i.e. } \mathbf{x}_i^T \beta_0 = 0) \quad \text{for all } i \geq 1. \tag{1.7}$$

In the remaining cases even small contaminations $\varepsilon > 0$ may lead to large expected deviances $d_i(\beta_0)$ for some $i \geq 1$, thus pushing the MLE's $\beta_n = \arg \min \mathcal{D}_n(\beta)$ far away from the true value of β_0 . Indeed, the expected deviances $d_i(\beta_0)$ are in the contaminated models (1.6) given by the formula

$$E d_i(\beta_0) = I(\pi_i + \varepsilon(1 - 2\pi_i) \parallel \pi_i) + H(\pi_i + \varepsilon(1 - 2\pi_i)) \quad (\text{cf. (1.5)}).$$

Hence for small π_i (in symbols $\pi_i \approx 0$) we get $1 - \pi_i \approx 1$ and $\pi_i + \varepsilon(1 - 2\pi_i) \approx \varepsilon$. This means that for all $n \geq i$

$$E\mathcal{D}_n(\beta_0) \geq E d_i(\beta_0) \approx \varepsilon \ln \frac{\varepsilon}{\pi_i} + (1 - \varepsilon) \ln(1 - \varepsilon) + H(\varepsilon),$$

where the right-hand term tends to infinity as $\pi_i \rightarrow 0$. At the same time $\beta_1 = \mathbf{0}$ satisfies $\mathbf{x}_i^T \beta_1 = 0$ and therefore $\pi(\mathbf{x}_i^T \beta_1) = \frac{1}{2}$ (cf. (1.7)) so that for $\beta_0 \neq \mathbf{0}$

$$E\mathcal{D}_n(\beta_1) = 0 < E\mathcal{D}_n(\beta_0).$$

This means that, as stated above, the MLE β_n will be far from the true parameter β_0 .

In order to restrict the undesired influence of large deviances resulting from contaminations of logistic regression models $\text{Be}(\pi_i) = \text{Be}(\pi(\mathbf{x}_i^T \boldsymbol{\beta}_0))$, previous authors replaced the deviances $d_i(\boldsymbol{\beta})$ in definition (1.4) by appropriate functions $\varrho(d_i(\boldsymbol{\beta}))$ of deviances, or even by more general expressions $\phi(Y_i, \pi(\mathbf{x}_i^T \boldsymbol{\beta}))$. This led to M -estimators $\boldsymbol{\beta}_n$ of the type

$$\boldsymbol{\beta}_n = \arg \min \sum_{i=1}^n \varrho(d_i(\boldsymbol{\beta})) \tag{1.8}$$

and

$$\boldsymbol{\beta}_n = \arg \min \sum_{i=1}^n \phi(Y_i, \pi(\mathbf{x}_i^T \boldsymbol{\beta})) \tag{1.9}$$

for $\varrho : (0, \infty) \rightarrow \mathbb{R}$ and $\phi : (0, \infty) \times (0, 1) \rightarrow \mathbb{R}$, or to the related M -estimators $\boldsymbol{\beta}_n$ which solve for a suitable $\boldsymbol{\psi} : \mathbb{R}^{2d+1} \rightarrow \mathbb{R}^d$ the equations

$$\sum_{i=1}^n \boldsymbol{\psi}(Y_i, \mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{0}. \tag{1.10}$$

The robust estimator $\boldsymbol{\beta}_n^{(0)}$ defined by (1.8) for a special function $\varrho(t)$ increasing with rate \sqrt{t} as $t \rightarrow \infty$ was proposed by Pregibon (1982). Morgenthaler (1992) proposed the robust estimator $\boldsymbol{\beta}_n^{(1)}$ defined by (1.9) for the function

$$\phi(Y, \pi(\mathbf{x}^T \boldsymbol{\beta})) = \frac{|Y - \pi(\mathbf{x}^T \boldsymbol{\beta})|}{\sqrt{\pi(\mathbf{x}^T \boldsymbol{\beta})(1 - \pi(\mathbf{x}^T \boldsymbol{\beta}))}}.$$

In order to improve asymptotic properties he redefined $\boldsymbol{\beta}_n^{(1)}$ as the solution of (1.10) for the function

$$\boldsymbol{\psi}(Y, \mathbf{x}, \boldsymbol{\beta}) = \sqrt{\pi(\mathbf{x}^T \boldsymbol{\beta})(1 - \pi(\mathbf{x}^T \boldsymbol{\beta}))}(Y - \pi(\mathbf{x}^T \boldsymbol{\beta}))\mathbf{x}. \tag{1.11}$$

This estimator will be called *Morg-estimator* in the sequel.

Bianco and Yohai (1996) improved the asymptotic properties and also the robustness of the Pregibon's $\boldsymbol{\beta}_n^{(0)}$ by introducing the class of M -estimators defined as minimizers (1.9) for

$$\phi(Y_i, \pi(\mathbf{x}_i^T \boldsymbol{\beta})) = \varrho(d_i(\boldsymbol{\beta})) + \varrho_0(\pi(\mathbf{x}_i^T \boldsymbol{\beta})), \tag{1.12}$$

where $\varrho(t)$ is bounded and differentiable on $(0, \infty)$ with a derivative $\varrho'(t)$ and the compensator function ϱ_0 is of the form

$$\varrho_0(p) = \varrho_1(p) + \varrho_1(1 - p) \tag{1.13}$$

for ϱ_1 depending on ϱ by the formula

$$\varrho_1(p) = \int_0^p \varrho'(-\ln t) dt, \quad p \in (0, 1). \tag{1.14}$$

These authors found pleasing statistical properties in the particular case where

$$\varrho(0) = 0 \quad \text{and} \quad \varrho'(t) = (1 - t)\mathbf{I}(0 < t < 1) \tag{1.15}$$

and $\mathbf{I}(\cdot)$ denotes the indicator function. This M -estimator is denoted by $\boldsymbol{\beta}_n^{(2)}$ and called *BY-estimator* in the sequel. Croux and Haesbroeck (2003) found that even more pleasing statistical properties are obtained when the BY-estimator from (1.15) is replaced by the alternative estimator from the general Bianco–Yohai class defined by

$$\varrho(0) = 0 \quad \text{and} \quad \varrho'(t) = e^{-\sqrt{1/2}}\mathbf{I}(0 < t < 1/2) + e^{-\sqrt{t}}\mathbf{I}(t \geq 1/2). \tag{1.16}$$

This particular M -estimator is denoted by $\boldsymbol{\beta}_n^{(3)}$ and called *CH-estimator* in the sequel.

Extensions of the Morgenthaler-type M -estimators defined by Eq. (1.10) were studied later by several authors, e.g. Kordzakhia et al. (2001), Adimari and Ventura (2001), Rousseeuw and Christmann (2003), Gervini (2005) and others cited there.

In this paper we propose a new robust M -estimator of the logistic regression parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ obtained by application of the classical robust L_1 -method (cf. Hampel et al., 1986; Jurečková and Sen, 1996; Zwanzig, 1997) to the continuous data

$$Z_i = Y_i + U_i, \quad 1 \leq i \leq n \tag{1.17}$$

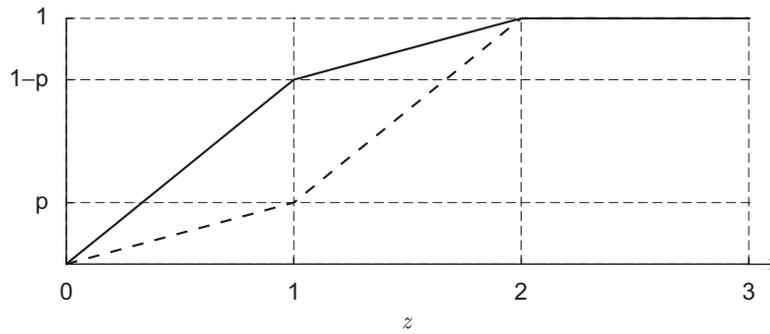


Fig. 1. $F_p(z)$ full line, $F_{1-p}(z)$ dashed line.

obtained by adding mutually and on Y_i independently $U(0, 1)$ -distributed (i.e. uniformly on $(0, 1)$ distributed) random variables U_i to the mutually independent above introduced observations

$$Y_i \sim \text{Be}(\pi(\mathbf{x}_i^T \boldsymbol{\beta}_0)). \tag{1.18}$$

In other words, we define the estimator

$$\hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n |Z_i - m(\pi(\mathbf{x}_i^T \boldsymbol{\beta}))| \tag{1.19}$$

for Z_i given by (1.17), (1.18) and for the median function

$$m(p) = F_p^{-1}(1/2) = \inf\{z \in \mathbb{R} : F_p(z) \geq 1/2\}$$

corresponding to the class of distribution functions F_p of the random variables

$$Z = \text{Be}(p) + U(0, 1)$$

when the parameter p varies in the closed interval $[0, 1]$. Obviously, for each $p \in [0, 1]$ and $z \in \mathbb{R}$

$$F_p(z) = (1 - p)z\mathbf{I}(0 < z \leq 1) + (1 - 2p + pz)\mathbf{I}(1 < z \leq 2) + \mathbf{I}(z > 2) \tag{1.20}$$

and the median function has the explicit form

$$m(p) = 1 + \frac{p - 1/2}{p \vee (1 - p)}, \quad 0 \leq p \leq 1. \tag{1.21}$$

Here and in the sequel we use the notation

$$a \vee b = \max\{a, b\} \quad \text{and} \quad a \wedge b = \min\{a, b\}.$$

The graphs of functions $F_p(z)$ and $m(p)$ are presented in Figs. 1 and 2. Fig. 2 also displays the inverse median function

$$m^{-1}(z) = \frac{z - 1/2}{2 - z} \mathbf{I}(1/2 \leq z \leq 1) + \frac{1}{2(2 - z)} \mathbf{I}(1 < z \leq 3/2) \tag{1.22}$$

used in the sequel.

Definition 1.1. Operation (1.17) is called a *statistical smoothing* of the discrete observations Y_i .

Definition 1.2. The estimator $\hat{\boldsymbol{\beta}}_n$ defined by (1.17)–(1.19) is a *median estimator* of logistic regression parameters $\boldsymbol{\beta}_0$ called *Med-estimator* in the sequel.

Remark 1.1. Median function $m(p)$ corresponding to the family of continuous random variables $Z = \text{Be}(p) + U(0, 1)$ is strictly increasing in $p \in [0, 1]$. Since the logistic function is strictly increasing too, the argument $m(\pi(\mathbf{x}^T \boldsymbol{\beta}))$ in (1.19) detects every change of the product $\mathbf{x}^T \boldsymbol{\beta}$. Contrary to this, the median function $\tilde{m}(p)$ corresponding to the family of discrete random variables $Y = \text{Be}(p)$ is piecewise constant in $p \in (0, 1)$ so that $\tilde{m}(\pi(\mathbf{x}^T \boldsymbol{\beta}))$ is insensitive to small variations of the product $\mathbf{x}^T \boldsymbol{\beta}$. Therefore the robust L_1 -estimation cannot be applied directly to the logistic regression data Y_i , i.e., the estimator

$$\tilde{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n |Y_i - \tilde{m}(\pi(\mathbf{x}_i^T \boldsymbol{\beta}))|$$

is not of a too much practical interest.

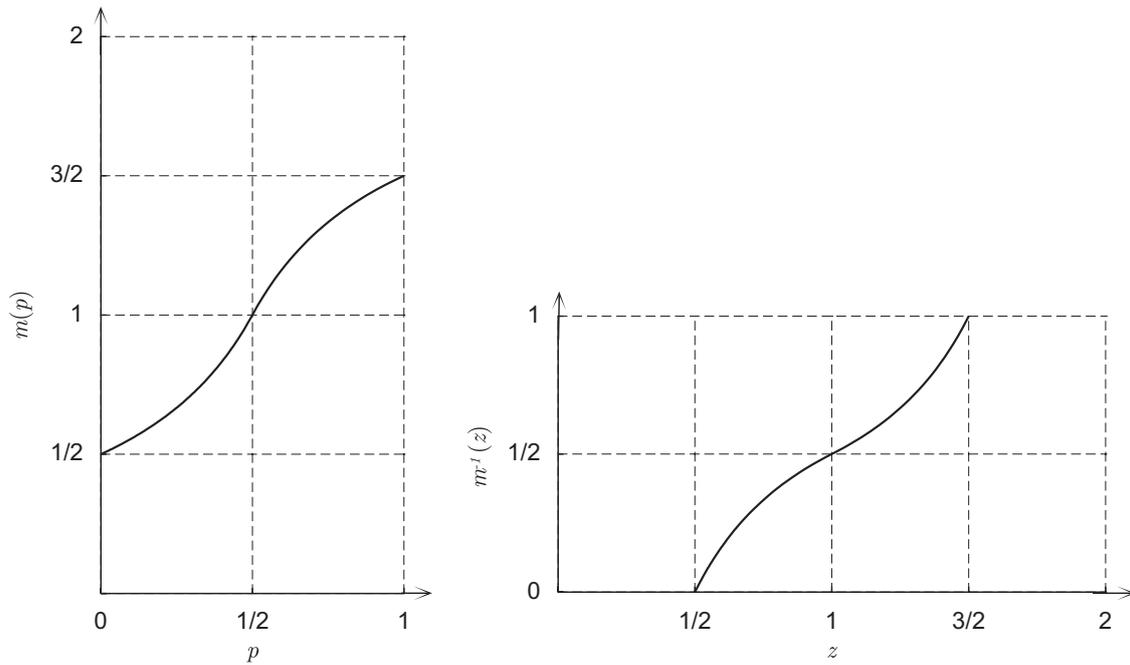


Fig. 2. Median function $m(p)$ and its inverse $m^{-1}(z)$.

Remark 1.2. The operation of statistical smoothing $Z_i = T(Y_i)$ given by (1.17) is statistically sufficient because the image Z_i allows complete recovery of the original Y_i by applying the integer-part operation $[\cdot]$ to Z_i . It is equivalent to the observation of discrete data through a semicontinuous channel considered in the information theory. Such an observation procedure goes in the opposite direction to the statistical quantization frequently applied to continuous data. The quantization is usually accompanied by the loss of information so that it is not statistically sufficient.

Remark 1.3. The statistical smoothing (1.17) can also be applied to integer valued observations Y_i in other discrete statistical models. This makes the statistical methods developed for continuous models more widely applicable in discrete statistics than just in the particular situation studied in this paper.

In the sections that follow we establish some desirable statistical properties of the Med-estimators such as the consistency and asymptotic normality. In simple situations we compare asymptotic variances of these estimators with the asymptotic variances of the above-mentioned selected classical estimators. We verify the robustness of our Med-estimators by demonstrating their low sensitivity to high leverage outliers and also by demonstrating that they outperform the above-mentioned classical robust estimators in certain special situations (e.g. heavy contaminations and large sample sizes). Our conclusions are based partly on simulations of the models used in the previous literature for mutual comparison of various estimators in logistic regression.

2. Asymptotic theory

A large class of statistical models assume independent real valued observations Z_1, \dots, Z_n of the form

$$Z_i \sim F_{u(\mathbf{x}_i^T \boldsymbol{\beta}_0)}(z), \quad 1 \leq i \leq n. \tag{2.1}$$

Similarly as above, here $\mathbf{x}_i \in \mathbb{R}^d$ are vectors of explanatory variables (regressors), $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ is a vector of true parameters and $\mathbf{x}_i^T \boldsymbol{\beta}$ the scalar product. Further, $u : \mathbb{R} \rightarrow \Theta$ is a smooth mapping and $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ a family of distribution functions on \mathbb{R} with an interval parameter space $\Theta \subseteq \mathbb{R}$. The basic statistical problem related to these models is finding mappings $\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_n(Z_1, \dots, Z_n)$ from \mathbb{R}^n to \mathbb{R}^d which can be used to estimate the unknown parameters $\boldsymbol{\beta}_0$ on the basis of observations (2.1).

Various desirable asymptotic or non-asymptotic properties are usually required from estimators $\hat{\boldsymbol{\beta}}_n$. Such properties are often found in the class of so-called *least absolute deviation estimators* (briefly L_1 -estimators) defined by

$$\hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n |Z_i - m(u(\mathbf{x}_i^T \boldsymbol{\beta}))| \tag{2.2}$$

for a given function $m : \Theta \rightarrow \mathbb{R}$. From the extensive literature related to these estimators one can mention e.g. Koenker and Basset (1978), Richardson and Bhattacharyya (1987), Yohai (1987), Pollard (1991), Morgenthaler (1992), Chen et al. (1993), Jurečková and Procházka (1994), Knight (1998), Arcones (2001), Liese and Vajda (1999, 2003, 2004) and Maronna et al. (2006).

In this section we study the asymptotics of the median estimator $\widehat{\beta}_n$ from (1.19) which estimates the true parameter $\beta_0 \in \mathbb{R}^d$ of the general logistic regression using the statistically smoothed responses

$$Z_i = Y_i + U_i \sim F_{\pi(\mathbf{x}_i^T \beta_0)}(z) \quad (\text{cf. (1.17)})$$

to the regressors \mathbf{x}_i where $\pi(t)$ and $F_p(z)$ are given by (1.2) and (1.20). We see that our data Z_i as well as the estimator $\widehat{\beta}_n$ are special cases of (2.1) and (2.2) for $\pi(t)$ and $F_p(z)$ given by (1.2) and (1.20) and for $m(p)$ given by (1.21).

Our results are based on what Liese and Vajda (1999, 2003, 2004) proved concerning the general median estimators (2.2) of parameters β_0 in the general statistical models (2.1). Next follow conditions (c1)–(c8) for consistency and asymptotic normality established by these authors and applied to the present median estimators (1.19).

(c1) The regressors $\mathbf{x}_1, \mathbf{x}_2, \dots$ are from a compact set $\mathcal{X} \subset \mathbb{R}^d$ and the probability measures

$$Q_n = \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{x}_i} \tag{2.3}$$

tend weakly for $n \rightarrow \infty$ to a probability measure Q on Borel subsets of \mathcal{X} .

Remark 2.1. If the regressors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are independently generated by a probability measure Q on the Borel subsets of a compact set $\mathcal{X} \subset \mathbb{R}^d$ then (c1) holds almost surely for these \mathcal{X} and Q . For example, if the dimension $d = 1$, then by the Glivenko theorem, the empirical probability measure (2.3) tends almost surely to Q in the Kolmogorov distance. But the convergence in this distance implies the weak convergence required by (c1).

(c2) The smallest eigenvalue of the matrix

$$\Sigma = \int_{\mathcal{X}} \mathbf{x}\mathbf{x}^T dQ(\mathbf{x}) \tag{2.4}$$

is positive. Hence for every $\beta \in \mathbb{R}^d$ different to β_0

$$Q(\mathbf{x} \in \mathcal{X} : \mathbf{x}^T(\beta - \beta_0) \neq 0) > 0. \tag{2.5}$$

The following conditions (c3)–(c5) obviously hold for the distribution functions $F_p(z)$ under consideration and their densities

$$f_p(z) = (1 - p)\mathbf{I}(0 < z \leq 1) + p\mathbf{I}(1 < z < 2), \quad z \in \mathbb{R}. \tag{2.6}$$

(c3) Distributions functions $F_p(z)$ are continuous in both arguments $p \in (0, 1), z \in 0, \infty$ and

$$\int_{-\infty}^{+\infty} |z| f_p(z) dz = \frac{1}{2} + p \quad \text{for each } p \in (0, 1). \tag{2.7}$$

(c4) Distributions functions $F_p(z), p \in (0, 1)$ are increasing in $z \in [0, 2]$ in the strict sense

$$F_p(z_1) < F_p(z_2) \quad \text{for } z_1 < z_2 \text{ from } [0, 2] \tag{2.8}$$

and constant for $z \in \mathbb{R} - [0, 2]$.

(c5) Distribution functions $F_p(z), p \in (0, 1)$ are stochastically ordered in the sense that for every $0 < p_1 < p_2 < 1$ and $z \in \mathbb{R}$ it holds $F_{p_1}(z) \geq F_{p_2}(z)$ where

$$F_{p_1}(z) > F_{p_2}(z) \quad \text{if } z \in [0, 2]. \tag{2.9}$$

The present conditions (c1)–(c5) imply the assumptions (E1+), (E2), (EM1), (EM2) and (M1)–(M4) of Theorem 2 and Lemmas 8 and 9 in Liese and Vajda (1999). For a detailed proof of this assertion we refer to Section 3 of Hobza et al. (2005). We shall prove that the present model satisfies also the following less evident conditions of consistency and asymptotic normality considered in Liese and Vajda (1999).

(c6) For every $0 < p_1 < p_2 < 1$ there exists $a > 0$ such that densities (2.6) and the median function $m(p)$ satisfy the condition

$$\Lambda(a) \equiv \inf_{|y| \leq a} \left(\inf_{p_1 \leq p \leq p_2} f_p(m(p) + y) \right) > 0. \tag{2.10}$$

(c7) The quantile function $m(p)$ is differentiable on $(0, 1)$ and the derivative $m'(p)$ is locally Lipschitz in the sense that for every $p_0 \in (0, 1)$

$$|m'(p) - m'(p_0)| \leq 2|p - p_0|.$$

(c8) Densities (2.6) satisfy for every $0 < p_1 < p_2 < 1$ the condition

$$\lim_{y \rightarrow 0} \sup_{p_1 \leq p \leq p_2} |f_p(m(p) + y) - f_p(m(p))| = 0. \tag{2.11}$$

Lemma 2.1. *In the present model conditions (c6)–(c8) hold.*

Proof. (I) Condition (2.10) can be verified separately for $p_1 = \frac{1}{2} < p_2 < 1$ and $0 < p_1 < p_2 = \frac{1}{2}$. We shall do this for $p_1 = \frac{1}{2} < p_2 < 1$ as the alternative can be verified similarly. Let $p > \frac{1}{2}$ be arbitrary. By (1.21),

$$\frac{3}{2} > m(p) = 1 + \frac{2p - 1}{2p} > 1$$

so that if $y \neq 0$ with $|y| \leq \frac{1}{2}$ is fixed then

$$f_p(m(p) + y) = f_p(m(p)) = p$$

unless $m(p) + y \leq 1$ in which case $f_p(m(p) + y) = 1 - p$. Thus

$$\inf_{1/2 \leq p \leq p_2} f_p(m(p) + y) \geq 1 - p$$

so that (2.10) holds.

(II) The median function (1.21) is continuously differentiable on the interval $(0, 1)$ with the positive derivative

$$m'(p) = \frac{1}{2[p \vee (1 - p)]^2} \tag{2.12}$$

bounded above by 2. Therefore (c7) holds too.

(III) Similarly as (2.10), condition (2.11) can be verified separately for $p_1 = \frac{1}{2} < p_2 < 1$ and $0 < p_1 < p_2 = \frac{1}{2}$. We shall do this for $p_1 = \frac{1}{2} < p_2 < 1$ as the alternative can be verified similarly. Let $y \neq 0$ with $|y| \leq \frac{1}{2}$ be arbitrarily fixed. Then

$$\inf_{1/2 \leq p \leq p_2} f_p(m(p) + y) \geq 1 - p.$$

The absolute difference $|f_p(m(p) + y) - f_p(m(p))|$ is either zero or $p - (1 - p) = 2p - 1$. This difference will be maximized if we take maximal p satisfying the inequality $m(p) + y \leq 1$ for the fixed y under consideration. Since $m(p)$ is increasing in p , this means that

$$\sup_{1/2 \leq p < p_2} |f_p(m(p) + y) - f_p(m(p))| = f_{p_*}(m(p_*) + y) - f_{p_*}(m(p_*)),$$

where p_* solves the equation $m(p) + y = 1$. Solutions p_* exist only for $y < 0$ (i.e. for $-\frac{1}{2} < y < 0$) and then $p_* = 1/[2(1 - |y|)]$. Thus we proved that

$$\sup_{1/2 \leq p < p_2} |f_p(m(p) + y) - f_p(m(p))| \leq 2p_* - 1 = \frac{|y|}{1 - |y|}$$

which implies (2.11) and completes the whole proof. \square

The median function $m(p)$ of (1.21) is bounded on $[0, 1]$. By Lemma 8 in Liese and Vajda (1999), this means that the sufficient condition of Lemma 9 in Liese and Vajda (1999) reduces to (2.5) assumed in (c2). Hence, by Theorem 2 and Lemmas 8 and 9 in Liese and Vajda (1999), under (c1)–(c5) our Med-estimator $\hat{\beta}_n$ consistently estimates the true $\beta_0 \in \mathbb{R}^d$ provided the measure Q of (c1) defines the function

$$\mathbf{m}(\beta) = \int_{\mathbb{R}} \int_{\mathcal{X}} |y - \varphi(\mathbf{x}^T \beta)| dF_{\pi(\mathbf{x}^T \beta)}(y) dQ(\mathbf{x}) \quad \text{for } \varphi(t) = m(\pi(t)) \tag{2.13}$$

of variable $\beta \in \mathbb{R}^d$ satisfying for every $\varepsilon > 0$ the identifiability condition

$$\inf_{\|\beta - \beta_0\| \geq \varepsilon} \mathbf{m}(\beta) > \mathbf{m}(\beta_0). \tag{2.14}$$

This is used in the proof of the following consistency theorem.

Theorem 2.1. *If the regressors of the model under consideration satisfy (c1) and (c2) then the Med-estimator $\widehat{\beta}_n$ consistently estimates the model parameters β_0 .*

Proof. By assumptions and by what was said above, (c1)–(c8) hold. Hence it suffices to prove (2.14). For this proof of a technical nature we refer to in Hobza et al. (2005, pp. 14–16) or Hobza et al. (2006, pp. 10–12). \square

The next problem is the asymptotic normality of the Med-estimator $\widehat{\beta}_n$. Let us start with the observation that the function φ defined in (2.13) is continuously differentiable with the derivative $\varphi'(t) = m'(\pi(t))\pi'(t)$ where $\pi'(t) = \pi(t)(1 - \pi(t))$. Further, define for $i = 1, 2, \dots$, the function

$$\Delta_i(\beta) = \varphi(\mathbf{x}_i^T \beta_0) - \varphi(\mathbf{x}_i^T \beta), \quad \beta \in \mathbb{R}^d,$$

and the random variables

$$Z_i = Y_i - \varphi(\mathbf{x}_i^T \beta_0),$$

with probability densities

$$\tilde{f}_i(z) = f_{\pi(\mathbf{x}_i^T \beta_0)}(z + \varphi(\mathbf{x}_i^T \beta_0)), \quad z \in \mathbb{R}.$$

The functions $\Delta_i(\beta)$ are continuously differentiable on \mathbb{R}^d with gradients

$$\text{grad}(\Delta_i(\beta)) = -\varphi'(\mathbf{x}_i^T \beta)\mathbf{x}_i.$$

Therefore the linear term $\mathcal{L}_n(h)$ considered in (2.3) of Liese and Vajda (2004) is given here by

$$\mathcal{L}_n(\mathbf{h}) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n D^+|Z_i| \varphi'(\mathbf{x}_i^T \beta) \mathbf{x}_i^T \mathbf{h}, \quad \mathbf{h} \in \mathbb{R}^d,$$

where $D^+|z|$ denotes the right-hand derivative of the function $|z|$,

$$D^+|z| = \mathbf{I}(0 \leq z < \infty) - \mathbf{I}(-\infty < z < 0). \tag{2.15}$$

Since $ED^+|Z_i| = 0$, the variance of $\mathcal{L}_n(\mathbf{h})$ is $\mathbf{h}^T \Sigma_n \mathbf{h}$ for the matrix Σ_n given in accordance with (2.5) of Liese and Vajda (1999) by

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n E(D^+|Z_i|)^2 (\varphi'(\mathbf{x}_i^T \beta))^2 \mathbf{x}_i \mathbf{x}_i^T.$$

But $E(D^+|Z_i|)^2 = 1$ so that we can write the matrix Σ_n in the integral form

$$\Sigma_n = \int_{\mathcal{X}} (\varphi'(\mathbf{x}^T \beta))^2 \mathbf{x}^T \mathbf{x} dQ_n(\mathbf{x}),$$

where Q_n is the empirical measure from (c1). Since $\varphi'(\mathbf{x}^T \beta)$ is continuous and bounded on \mathcal{X} , it holds

$$\lim_{n \rightarrow \infty} \Sigma_n = \Sigma \equiv \int_{\mathcal{X}} (\varphi'(\mathbf{x}^T \beta))^2 \mathbf{x}^T \mathbf{x} dQ(\mathbf{x}), \tag{2.16}$$

where Q is the limit measure from (c1).

Let us now evaluate the matrices

$$\mathcal{Q}_n = \frac{1}{n} \sum_{i=1}^n g_i(0) \nabla \varphi(\mathbf{x}_i^T \beta_0) (\nabla \varphi(\mathbf{x}_i^T \beta_0))^T,$$

where $g_i(t)$ denote derivatives of the functions $G_i(t) = E\mathcal{D}|Z_i + t|$ of variable $t \in \mathbb{R}$ introduced in Liese and Vajda (2003, p. 467). By formula (2.15) for $D^+|z|$, if we put $\pi_i = \pi(\mathbf{x}_i^T \beta_0)$ and $\varphi_i = \varphi(\mathbf{x}_i^T \beta_0)$ then

$$G_i(t) = E\mathbf{I}(Z_i + t > 0) - E\mathbf{I}(Z_i + t \leq 0) = E\mathbf{I}(Y_i > \varphi_i - t) - E\mathbf{I}(Y_i \leq \varphi_i - t) = 1 - 2F\pi_i(\varphi_i - t).$$

Thus $g_i(t) = 2f\pi_i(\varphi_i - t)$ and

$$g_i(0) = 2f_{\pi(\mathbf{x}_i^T \beta_0)}(\varphi(\mathbf{x}_i^T \beta_0)).$$

Therefore the matrices \mathcal{Q}_n may be represented as the integrals

$$\mathcal{Q}_n = 2 \int_{\mathcal{X}} f_{\pi(\mathbf{x}^T \beta_0)}(\varphi(\mathbf{x}^T \beta_0)) (\varphi'(\mathbf{x}^T \beta_0))^2 \mathbf{x}^T \mathbf{x} dQ_n(\mathbf{x}).$$

Since $\varphi'(\mathbf{x}_i^T \boldsymbol{\beta}_0)$ is continuous and bounded on \mathcal{X} and, by (c8), the function

$$f_{\pi(\mathbf{x}^T \boldsymbol{\beta}_0)}(\varphi(\mathbf{x}^T \boldsymbol{\beta}_0)) = f_{\pi(\mathbf{x}^T \boldsymbol{\beta}_0)}(m(\pi(\mathbf{x}^T \boldsymbol{\beta}_0)))$$

is continuous and bounded on \mathcal{X} too, it holds

$$\lim_{n \rightarrow \infty} \mathcal{Q}_n = \mathcal{Q} \equiv 2 \int_{\mathcal{X}} f_{\pi(\mathbf{x}^T \boldsymbol{\beta}_0)}(\varphi(\mathbf{x}^T \boldsymbol{\beta}_0)) (\varphi'(\mathbf{x}^T \boldsymbol{\beta}_0))^2 \mathbf{x}^T \mathbf{x} dQ(\mathbf{x}). \tag{2.17}$$

Finally, $D^+ \rho(Z_i) = D^+ |Z_i|$ is in the present situation bounded and $\nabla f_i(\boldsymbol{\beta}_0) = \text{grad}(A_i(\boldsymbol{\beta}_0)) = -\varphi'(\mathbf{x}_i^T \boldsymbol{\beta}_0) \mathbf{x}_i$ is bounded uniformly for all possible $\mathbf{x}_i \in \mathcal{X}$. Consequently the Liapunov condition (2.6) of Liese and Vajda (2004) holds. Similarly, one can verify that conditions (2.39) and (2.40) as well as (C3) and (C4) in Liese and Vajda (2003) hold. Thus, by Lemma 3 in Liese and Vajda (2003), the remaining conditions (C5) and (C6) stated there hold too.

To finalize the evaluation of the matrices Σ and \mathcal{Q} given in (2.16), (2.17), take into account that the definition of $\varphi(t)$ in (2.13) implies in the present situation

$$\varphi(t) = \begin{cases} \frac{3}{2} - \frac{e^{-t}}{2} & \text{if } t \geq 0, \\ \frac{1}{2} + \frac{e^t}{2} & \text{if } t < 0. \end{cases}$$

Therefore

$$\varphi'(t) = \frac{e^{-|t|}}{2} \quad \text{if } t \in \mathbb{R}.$$

Further

$$f_{\pi(t)}(\varphi(t)) = \pi(t) \vee (1 - \pi(t)) = \frac{1}{1 + e^{-t}} \vee \frac{1}{1 + e^t} = \frac{1}{1 + e^{-|t|}} = \frac{e^{|t|}}{1 + e^{|t|}}.$$

Consequently,

$$f_{\pi(t)}(\varphi(t)) (\varphi'(t))^2 = \frac{e^{-|t|}}{4(1 + e^{|t|})}$$

and by (2.16), (2.17)

$$\Sigma = \frac{1}{4} \int_{\mathcal{X}} e^{-2|\mathbf{x}^T \boldsymbol{\beta}_0|} \mathbf{x} \mathbf{x}^T dQ(\mathbf{x}) \tag{2.18}$$

and

$$\mathcal{Q} = \frac{1}{2} \int_{\mathcal{X}} \frac{e^{-|\mathbf{x}^T \boldsymbol{\beta}_0|}}{1 + e^{|\mathbf{x}^T \boldsymbol{\beta}_0|}} \mathbf{x} \mathbf{x}^T dQ(\mathbf{x}). \tag{2.19}$$

Thus we can conclude that if (c1) and (c2) hold then all assumptions of Theorem 1 in Liese and Vajda (2004) are satisfied and the following assertion follows from there.

Theorem 2.2. *Let the regressors of the model under consideration satisfy (c1) and (c2). If the limit matrix \mathcal{Q} in (2.19) is positive definite then the Med-estimator $\hat{\boldsymbol{\beta}}_n$ of the model parameters $\boldsymbol{\beta}_0$ is asymptotically normal in the sense that*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(\mathbf{0}, \mathcal{Q}^{-1} \Sigma \mathcal{Q}^{-1}) \tag{2.20}$$

for Σ given in (2.18).

Proof. See above. \square

Example 2.1. Most simple is the application of Theorem 2.2 to the univariate logistic regression

$$Y_i \sim \text{Be}(\pi(x_i \beta_0)), \quad 1 \leq i \leq n \tag{2.21}$$

with identical regressors $x_1 = x_2 = \dots = 1$ and $\beta_0 \in \mathbb{R}$. In this model the Med-estimator $\hat{\beta}_n \in \mathbb{R}$ is defined by

$$\hat{\beta}_n = \arg \min_{\beta} \sum_{i=1}^n |Z_i - m(\pi(\beta))| \quad \text{for } Z_i = Y_i + U_i \quad (\text{cf. (1.17)–(1.19)}). \tag{2.22}$$

Then (c1) and (c2) hold for the Dirac measure $Q = \delta_1$ concentrated at the point $x_1 = x_2 = \dots = 1$ from the singleton $\mathcal{X} = \{1\} \subset \mathbb{R}$. Therefore from (2.18) and (2.19) we get

$$\Sigma = \frac{1}{4} e^{-2|\beta_0|}, \quad \varrho = \frac{e^{-|\beta_0|}}{2(1 + e^{|\beta_0|})}$$

so that $\varrho^{-1} \Sigma \varrho^{-1} = [1 + e^{|\beta_0|}]^2$ and according to (2.20)

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, [1 + e^{|\beta_0|}]^2). \tag{2.23}$$

Note that this asymptotic normality can also be verified directly from the central limit theorem applied to the explicit formula

$$\widehat{\beta}_n = \pi^{-1}(m^{-1}(Z_{(n/2)})) \tag{2.24}$$

which can easily be obtained from definition (2.22), where $Z_{(n/2)}$ denotes the median of Z_1, \dots, Z_n , $m^{-1}(z)$ is the inverse function from (1.22) also given in Fig. 2 and

$$\pi^{-1}(p) = \ln \frac{p}{1-p}, \quad 0 < p < 1 \tag{2.25}$$

is the inverse to the logistic function $\pi(t)$ of (1.2). This derivation can be found in Hobza et al. (2006, p. 16).

Notice that

$$1 + e^{|\beta_0|} = [\pi(\beta_0) \wedge (1 - \pi(\beta_0))]^{-1}. \tag{2.26}$$

Hence the asymptotic variance $[1 + e^{|\beta_0|}]^2 = [\pi(\beta_0) \wedge (1 - \pi(\beta_0))]^{-2}$ in (2.23) is strictly larger than the reversed Fisher information

$$\mathcal{J}^{-1}(\beta_0) = [\pi(\beta_0)(1 - \pi(\beta_0))]^{-1}$$

in the statistical model $\text{Be}(\pi(\beta_0))$. Therefore the Med-estimator $\widehat{\beta}_n$ of (2.22) or (2.24) is not asymptotically efficient. Its subefficiency can to some extent be suppressed by the method of enhancing introduced in Section 3. In Section 4 we shall see that this subefficiency is compensated by the very desirable property of robustness with respect to contaminations of the model $\text{Be}(\pi(\beta_0))$.

3. Enhancement

As illustrated on the lines above, application of the L_1 -estimators (2.2) in discrete statistical models with observations Y_i , $1 \leq i \leq n$ statistically smoothed into the continuous form (2.1) is usually accompanied by a loss of efficiency achievable in the original discrete models. By choosing a simple concrete L_1 -estimator we can try to find a method for suppression of this inefficiency.

An example of an L_1 -estimator which is even simpler than the Med-estimator (2.22), (2.24) is the median estimator

$$\widehat{p}_n = \arg \min_p \sum_{i=1}^n |Z_i - m(p)| = m^{-1}(Z_{(n/2)}) \tag{3.1}$$

of the Bernoulli parameter $p_0 \in (0, 1)$ based on the smoothed versions $Z_i = Y_i + U_i$ (cf. (1.17)) of the original discrete observations $Y_i \sim \text{Be}(p_0)$. In (3.1), $m(p)$ is the median function (1.21), $m^{-1}(z)$ is its inverse (1.22) and $Z_{(n/2)}$ is the median of Z_1, \dots, Z_n . It is well known that the Fisher information in the model $\text{Be}(p_0)$ is $\mathcal{J}(p_0) = 1/[p_0(1 - p_0)]$ and that the MLE

$$p_n = \frac{1}{n} \sum_{i=1}^n Y_i \tag{3.2}$$

of p_0 is asymptotically efficient in the sense

$$\sqrt{n}(p_n - p_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1/\mathcal{J}(p_0)) = N(0, p_0(1 - p_0)). \tag{3.3}$$

On the other hand, it is easy to prove (see Hobza et al., 2006, p. 15) that the Med-estimator \widehat{p}_n is asymptotically normal in the sense

$$\sqrt{n}(\widehat{p}_n - p_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, [p_0 \vee (1 - p_0)]^2), \tag{3.4}$$

where

$$[p_0 \vee (1 - p_0)]^2 \geq p_0(1 - p_0).$$

Since this inequality is strict unless $p_0 = \frac{1}{2}$, the Med-estimator \widehat{p}_n is asymptotically less efficient than the MLE p_n except in the special case when $p_0 = \frac{1}{2}$.

The set of statistically smoothed data $Z_i = Y_i + U_i$, $1 \leq i \leq n$ can be expanded by considering for $k > 1$ the matrix of data

$$Z_{ij} = Y_i + U_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq k, \tag{3.5}$$

where U_{ij} are $U(0, 1)$ -distributed and mutually as well as on Y_1, \dots, Y_n independent random variables. If a method of processing the data Z_1, \dots, Z_n is statistically optimal in an appropriate sense, like e.g. the MLE

$$\tilde{p}_n(Z_1, \dots, Z_n) = \arg \max_p \prod_{i=1}^n f_p(Z_i) = \arg \max_p \prod_{i=1}^n p^{Y_i} (1-p)^{1-Y_i}$$

coinciding with the classical Bernoulli MLE $p_n = p_n(Y_1, \dots, Y_n)$ introduced in (3.2), then its performance cannot be improved by expanding the sufficient statistic (Z_1, \dots, Z_n) . For example, it is easy to see that the MLE

$$\tilde{p}_n(Z_{11}, \dots, Z_{nn}) = \arg \max_p \prod_{j=1}^k \prod_{i=1}^n f_p(Z_{ij})$$

"enhanced" by utilizing the expanded data set (3.5) coincides with the previous MLE $\tilde{p}_n(Z_1, \dots, Z_n) \equiv p_n$. On the other hand, if the method is suboptimal, like for example the median estimator $\hat{p}_n = \hat{p}_n(Z_1, \dots, Z_n)$ introduced in (3.1), then its performance can be improved by using the expanded data set (3.5).

The following theorem motivates this section. It deals with the *k-enhanced median estimator*

$$\hat{p}_{n*k} = \arg \min_p \sum_{j=1}^k \sum_{i=1}^n |Z_{ij} - m(p)| = m^{-1}(Z_{(nk/2)}) \tag{3.6}$$

of the Bernoulli parameter $p_0 \in (0, 1)$ where $m(p)$ is the same as in (3.1), Z_{11}, \dots, Z_{nk} are the smoothed observations (3.5) and $Z_{(nk/2)}$ is the median of all these observations. More precisely, it deals with the expected squared and absolute errors

$$\sigma^2(\hat{p}_{n*k}) = E(\hat{p}_{n*k} - p_0)^2 \quad \text{and} \quad e(\hat{p}_{n*k}) = E|\hat{p}_{n*k} - p_0| \tag{3.7}$$

of the estimators \hat{p}_{n*k} and compares them with the expected squared error

$$\sigma^2(p_n) = E(p_n - p_0)^2 = \frac{p_0(1-p_0)}{n} \tag{3.8}$$

and the expected absolute error

$$e(p_n) = E|p_n - p_0| = \frac{1}{n} \sum_{r=0}^n |r - np_0| \binom{n}{r} p_0^r (1-p_0)^{n-r} \tag{3.9}$$

of the classical MLE p_n minimizing the expected squared error (variance) in the class of all unbiased estimators (cf. Mood et al., 1974, Example 31, p. 322).

Theorem 3.1. *The k-enhanced median estimator \hat{p}_{n*k} is asymptotically optimal in the sense that for each $n \geq 1$*

$$\sigma^2(\hat{p}_{n*k}) \xrightarrow[k \rightarrow \infty]{} \sigma^2(p_n) \quad \text{and} \quad e(\hat{p}_{n*k}) \xrightarrow[k \rightarrow \infty]{} e(p_n). \tag{3.10}$$

Proof. See Hobza et al. (2006, pp. 18–21). \square

The method of smoothing and the subsequent median estimation originally introduced for continuous statistical models is illustrated experimentally in Table 1.

This table also illustrates the loss of efficiency resulting from the median estimation and its compensation by means of the enhancement procedure introduced above. Among other it verifies the convergences established in Theorem 3.1. A more complete table also illustrating $p_0 = 0.1$ and 0.5 can be found as Table 1 in Hobza et al. (2006).

Motivated by Theorem 3.1 and its experimental verifications, we extend our Definition 1.2 as follows.

Definition 3.1. For every $k \geq 1$ we define the *k-enhanced median estimator* (briefly, *k-Med-estimator*) $\hat{\beta}_{n*k}$ of the parameters of logistic regression β_0 by the condition

$$\hat{\beta}_{n*k} = \arg \min_{\beta} \sum_{j=1}^k \sum_{i=1}^n |Z_{ij} - m(\pi(\mathbf{x}_i^T \beta))|, \tag{3.11}$$

Table 1
Analysis of the proposed smoothing method in the Bernoulli model $Be(p_0)$

p_0	\tilde{p}_n	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
		MAE	STD	MAE	STD	MAE	STD	MAE	STD
0.2	p_n	0.097	0.127	0.071	0.090	0.045	0.057	0.032	0.040
	\hat{p}_n	0.214	0.326	0.153	0.214	0.091	0.119	0.065	0.082
	\hat{p}_{n*5}	0.133	0.171	0.093	0.118	0.058	0.073	0.039	0.049
	\hat{p}_{n*10}	0.118	0.149	0.082	0.104	0.052	0.066	0.035	0.044
	\hat{p}_{n*50}	0.105	0.132	0.074	0.092	0.047	0.059	0.033	0.041
	\hat{p}_{n*100}	0.104	0.131	0.073	0.090	0.046	0.058	0.032	0.040

Compared are two estimators \tilde{p}_n , namely the MLE p_n and the Med-estimator \hat{p}_n together with its enhanced versions \hat{p}_{n*k} for selected $k > 1$. The table presents the mean absolute errors (MAE) and standard deviations (STD) (cf. (4.14), (4.15)) for given sample sizes n . The column minima are printed bold.

where the random variables Z_{ij} are defined by (3.5) for the same logistic $Y_i \sim Be(\pi(\mathbf{x}_i^T \boldsymbol{\beta}_0))$ as considered in the special case $k = 1$ in Definition 1.2.

Obviously, if $k = 1$ then the estimator introduced by this definition coincides with the median estimator of Definition 1.2.

Example 3.1. Let us consider the simple logistic regression model of Example 2.1 with an unknown parameter $\beta_0 \in \mathbb{R}$ and the Med-estimator $\hat{\beta}_n$ satisfying (2.23). The MLE in this example is

$$\beta_n = \pi^{-1}(p_n) \quad \text{for } p_n = \frac{1}{n} \sum_{i=1}^n Y_i \quad (\text{cf. (3.2)}), \tag{3.12}$$

where π^{-1} is the inverse function of $\pi(t)$ in (1.2) and

$$Y_i \sim Be(p_0) \quad \text{for } p_0 = \pi(\beta_0). \tag{3.13}$$

Combining the Taylor expansion of $\pi^{-1}(p)$ around $p_0 = \pi(\beta_0)$ with (3.3) and using $d/dp(\pi^{-1}(p)) = [p(1 - p)]^{-1}$, we get

$$\sqrt{n}(\beta_n - \beta_0) \xrightarrow{D} N(0, [\pi(\beta_0)(1 - \pi(\beta_0))]^{-1}) \quad (\text{cf. (2.23), (2.26)}), \tag{3.14}$$

where the asymptotic variance

$$[\pi(\beta_0)(1 - \pi(\beta_0))]^{-1} = (e^{-\beta_0} + 1)(e^{\beta_0} + 1) \tag{3.15}$$

of the MLE β_n is minimal in the class of all unbiased estimators of β_0 . In particular, for all $\beta_0 \neq 0$ it is smaller than the asymptotic variance $(e^{|\beta_0|} + 1)^2$ of the Med-estimator $\hat{\beta}_n$ found in (2.23).

In Hobza et al. (2006, p. 22) we proved that the k -Med-estimators $\hat{\beta}_{n*k}$ satisfy for every $k \geq 1$ the limit law

$$\sqrt{n}(\hat{\beta}_{n*k} - \beta_0) \xrightarrow{D} N(0, s_k^2(\beta_0)). \tag{3.16}$$

Here $s_k^2(\beta_0)$ is an asymptotic variance of $\hat{\beta}_{n*k}$ for fixed k and $n \rightarrow \infty$, tending to the minimal asymptotic variance (3.15) for fixed n and $k \rightarrow \infty$, i.e.,

$$s_k^2(\beta_0) \xrightarrow{k \rightarrow \infty} [\pi(\beta_0)(1 - \pi(\beta_0))]^{-1}. \tag{3.17}$$

Thus in the present example the enhancing enables to bring the asymptotic performance of the *Med-estimator* arbitrarily close to the optimal MLE. This result is not surprising—it follows directly from Theorem 3.1 since in the present example $Y_i \sim Be(\pi(\beta_0))$ and instead of $p_0 = \pi(\beta_0)$ we estimate the one–one related $\beta_0 = \pi^{-1}(p_0)$ by $\hat{\beta}_{n*k} = \pi^{-1}(\hat{p}_{n*k})$.

Remark 3.1. A systematic general theory of k -enhanced median estimators in the logistic regression exceeds the scope of the present paper. In our research we carried out a simulation study which among others demonstrated that for more general models $Y_i \sim Be(\pi(\mathbf{x}_i^T \boldsymbol{\beta}_0))$ the k -Med-estimators remain systematically for all k outperformed by the MLE's. Some results of our simulations are in Tables 2–5 of the next section. The previously mentioned conclusion is to some extent seen from the rows of Tables 4 and 5 corresponding to $\varepsilon = 0$. The tables of Section 4 at the same time demonstrate that the enhancement improves the efficiency of median estimators but in some situations deteriorates their robustness.

Table 2

Mean absolute errors (4.14), standard deviations (4.15) and rejection rates RR for selected estimators $\tilde{\beta}_n$ of the true parameter $\beta_0 = -\ln 4$ of the simple logistic regression model $\text{Be}(\pi(x_i\beta_0))$ for random independent x_i with $\Pr(x_i = 1) = \Pr(x_i = -1) = \frac{1}{2}$

ε	$\tilde{\beta}_n$	$n = 50$			$n = 100$			$n = 500$			$n = 1000$		
		MAE	STD	RR (%)									
0	$\hat{\beta}_n$	0.273	0.350	0	0.195	0.248	0	0.088	0.110	0	0.065	0.081	0
	$\hat{\beta}_n$	0.579	0.854	7	0.436	0.617	0	0.180	0.246	0	0.121	0.153	0
	$\hat{\beta}_{n*5}$	0.371	0.504	1	0.256	0.338	0	0.114	0.142	0	0.082	0.104	0
	$\hat{\beta}_{n*10}$	0.324	0.424	0	0.233	0.297	0	0.102	0.128	0	0.073	0.091	0
	$\hat{\beta}_{n*50}$	0.291	0.376	0	0.201	0.256	0	0.091	0.114	0	0.067	0.083	0
0.1	$\hat{\beta}_n$	0.393	0.463	0	0.358	0.411	0	0.343	0.357	0	0.339	0.346	0
	$\hat{\beta}_n$	0.563	0.708	3	0.448	0.617	0	0.346	0.380	0	0.335	0.354	0
	$\hat{\beta}_{n*5}$	0.432	0.509	0	0.373	0.432	0	0.338	0.357	0	0.338	0.348	0
	$\hat{\beta}_{n*10}$	0.417	0.498	0	0.365	0.418	0	0.341	0.357	0	0.337	0.346	0
	$\hat{\beta}_{n*50}$	0.399	0.467	0	0.360	0.413	0	0.342	0.357	0	0.339	0.346	0
0.2	$\hat{\beta}_n$	0.632	0.695	0	0.617	0.654	0	0.631	0.638	0	0.630	0.634	0
	$\hat{\beta}_n$	0.685	0.783	1	0.624	0.685	0	0.623	0.640	0	0.626	0.633	0
	$\hat{\beta}_{n*5}$	0.631	0.702	0	0.615	0.659	0	0.630	0.639	0	0.629	0.634	0
	$\hat{\beta}_{n*10}$	0.629	0.700	0	0.616	0.657	0	0.632	0.639	0	0.629	0.633	0
	$\hat{\beta}_{n*50}$	0.633	0.696	0	0.619	0.656	0	0.631	0.638	0	0.630	0.634	0
0.3	$\hat{\beta}_n$	0.890	0.941	0	0.895	0.920	0	0.894	0.899	0	0.896	0.899	0
	$\hat{\beta}_n$	0.874	0.950	0	0.879	0.920	0	0.885	0.894	0	0.894	0.897	0
	$\hat{\beta}_{n*5}$	0.883	0.940	0	0.892	0.920	0	0.893	0.899	0	0.896	0.898	0
	$\hat{\beta}_{n*10}$	0.887	0.940	0	0.893	0.919	0	0.892	0.897	0	0.896	0.899	0
	$\hat{\beta}_{n*50}$	0.889	0.940	0	0.895	0.920	0	0.893	0.898	0	0.896	0.899	0

Compared is the common value β_n of the Morg-, CH- and BY-estimators with the median estimators $\hat{\beta}_n$ and their k -enhancements $\hat{\beta}_{n*k}$. The errors are evaluated for the contaminated data $Y_i \sim (1 - \varepsilon) \text{Be}(\pi(x_i\beta_0)) + \varepsilon \text{Be}(1 - \pi(x_i\beta_0))$. The achieved minima are printed bold.

Table 3

The same simulations as in Table 2 evaluated for $n(1 - \varepsilon)$ standard logistic regression data $Y_i \sim \text{Be}(\pi(x_i\beta_0))$ for $1 \leq i \leq n(1 - \varepsilon)$ and $n\varepsilon$ leverage points $Y_i \sim \text{Be}(\pi(-10x_i\beta_0))$ for $n(1 - \varepsilon) + 1 \leq i \leq n$

ε	$\tilde{\beta}_n$	$n = 50$			$n = 100$			$n = 500$			$n = 1000$		
		MAE	STD	RR (%)									
0	$\hat{\beta}_n$	0.273	0.350	0	0.195	0.248	0	0.088	0.110	0	0.065	0.081	0
	$\hat{\beta}_n$	0.579	0.854	7	0.436	0.617	2	0.180	0.246	0	0.121	0.153	0
	$\hat{\beta}_{n*5}$	0.371	0.504	1	0.256	0.338	0	0.114	0.142	0	0.082	0.104	0
	$\hat{\beta}_{n*10}$	0.324	0.424	0	0.233	0.297	0	0.102	0.128	0	0.073	0.091	0
	$\hat{\beta}_{n*50}$	0.291	0.376	0	0.201	0.256	0	0.091	0.114	0	0.067	0.083	0
0.1	$\hat{\beta}_n$	0.454	0.520	0	0.441	0.478	0	0.441	0.449	0	0.443	0.447	0
	$\hat{\beta}_n$	0.575	0.718	1	0.474	0.541	0	0.425	0.451	0	0.436	0.449	0
	$\hat{\beta}_{n*5}$	0.470	0.539	0	0.441	0.489	0	0.439	0.450	0	0.443	0.449	0
	$\hat{\beta}_{n*10}$	0.461	0.529	0	0.440	0.484	0	0.442	0.452	0	0.442	0.447	0
	$\hat{\beta}_{n*50}$	0.457	0.522	0	0.443	0.481	0	0.441	0.449	0	0.443	0.447	0
0.2	$\hat{\beta}_n$	0.791	0.820	0	0.809	0.823	0	0.811	0.815	0	0.811	0.813	0
	$\hat{\beta}_n$	0.774	0.835	0	0.783	0.821	0	0.806	0.813	0	0.808	0.811	0
	$\hat{\beta}_{n*5}$	0.779	0.818	0	0.803	0.822	0	0.810	0.814	0	0.811	0.813	0
	$\hat{\beta}_{n*10}$	0.791	0.824	0	0.806	0.823	0	0.811	0.814	0	0.811	0.813	0
	$\hat{\beta}_{n*50}$	0.791	0.821	0	0.809	0.824	0	0.812	0.816	0	0.811	0.813	0
0.3	$\hat{\beta}_n$	1.135	1.151	0	1.141	1.149	0	1.143	1.145	0	1.143	1.144	0
	$\hat{\beta}_n$	1.092	1.126	0	1.119	1.134	0	1.138	1.141	0	1.140	1.142	0
	$\hat{\beta}_{n*5}$	1.127	1.147	0	1.134	1.144	0	1.142	1.144	0	1.143	1.143	0
	$\hat{\beta}_{n*10}$	1.129	1.147	0	1.140	1.148	0	1.142	1.144	0	1.143	1.144	0
	$\hat{\beta}_{n*50}$	1.135	1.151	0	1.141	1.149	0	1.143	1.145	0	1.143	1.144	0

The achieved minima are printed bold.

Table 4

Mean absolute errors (4.16) and rejection rates RR for selected estimators $\tilde{\beta}_n = (\tilde{\beta}_{n0}, \tilde{\beta}_{n1})^T$ of the true parameter $\beta_0 = (\beta_{00}, \beta_{01})^T = (-2.82, 2.82)^T$ in the ε -contaminated logistic regression model (4.17) with $\Pr(Y = 1) = 0.2$

ε	$\tilde{\beta}_n$	$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
		MAE	RR (%)						
0	MLE	0.86	1	0.58	0	0.24	0	0.16	0
	Morg	0.95	3	0.64	0	0.26	0	0.18	0
	CH	0.84	1	0.58	0	0.25	0	0.17	0
	BY	1.53	8	0.94	1	0.31	0	0.21	0
	Med	3.27	18	2.55	9	0.92	0	0.50	0
	5-Med	2.79	22	1.64	10	0.52	0	0.33	0
	10-Med	2.50	24	1.73	10	0.46	0	0.31	0
0.05	MLE	1.07	0	0.98	0	1.05	0	1.05	0
	Morg	1.06	1	0.83	0	0.77	0	0.77	0
	CH	0.98	0	0.85	0	0.84	0	0.84	0
	BY	1.48	5	0.91	1	0.55	0	0.53	0
	Med	2.72	13	2.52	8	0.85	0	0.59	0
	5-Med	2.09	18	1.54	8	0.58	0	0.49	0
	10-Med	2.10	19	1.44	9	0.55	0	0.48	0
0.1	MLE	1.42	0	1.47	0	1.51	0	1.52	0
	Morg	1.35	0	1.29	0	1.32	0	1.34	0
	CH	1.33	0	1.33	0	1.36	0	1.37	0
	BY	1.45	3	1.22	0	1.07	0	1.10	0
	Med	2.70	11	2.29	4	0.98	0	0.92	0
	5-Med	1.91	14	1.59	5	0.92	0	0.93	0
	10-Med	1.82	16	1.57	6	0.91	0	0.93	0
0.2	MLE	1.99	0	2.01	0	2.03	0	2.03	0
	Morg	1.94	0	1.96	0	1.97	0	1.98	0
	CH	1.94	0	1.96	0	1.97	0	1.98	0
	BY	1.90	0	1.89	0	1.94	0	1.95	0
	Med	2.38	5	1.99	2	1.68	0	1.71	0
	5-Med	2.03	5	1.88	2	1.72	0	1.73	0
	10-Med	2.04	8	1.73	3	1.73	0	1.73	0
0.3	MLE	2.34	0	2.34	0	2.35	0	2.36	0
	Morg	2.32	0	2.32	0	2.34	0	2.34	0
	CH	2.32	0	2.32	0	2.34	0	2.34	0
	BY	2.31	0	2.32	0	2.34	0	2.34	0
	Med	2.46	3	2.27	1	2.23	0	2.24	0
	5-Med	2.27	2	2.24	0	2.25	0	2.25	0
	10-Med	2.29	3	2.22	0	2.25	0	2.25	0

The achieved minima are printed bold.

4. Robustness

The median estimator $\hat{\beta}_n$ of logistic regression parameters $\beta_0 \in \mathbb{R}^d$ was defined in Section 1.1 by means of the least absolute deviations principle (the L_1 -method principle) which is motivated in the classical statistical literature by the robustness of the corresponding statistical procedures (cf. Hampel et al., 1986 or Jurečková and Sen, 1996). Therefore we expected that this estimator would be robust and this robustness was in fact our main motivation for this research. A secondary motivation was to demonstrate that the general theory of consistency and asymptotic normality of M -estimators developed by Liese and Vajda (1999, 2003, 2004) is applicable in particular concrete situations.

The primary purpose of this section is verification of the desired robustness of $\hat{\beta}_n$. The Med-estimator $\hat{\beta}_n$ is in this respect compared with several robust estimators known from the previous literature and discussed in Section 1.1, namely with the Morgenthaler's Morg-estimator $\beta_n^{(1)}$, the Bianco and Yohai's BY-estimator $\beta_n^{(2)}$ and the Croux and Haesbroeck's the CH-estimator $\beta_n^{(3)}$. For the sake of completeness, we have also included in our comparisons the MLE β_n . The robustness is compared by means of simulated performances of all these estimators in the logistic models $\text{Be}(\pi(\mathbf{x}^T \beta_0))$ ε -contaminated at the levels $0 \leq \varepsilon \leq 0.3$ by the alternative data source $\text{Be}(1 - \pi(\mathbf{x}^T \beta_0))$, or contaminated at the same levels ε by the leverage points from logistic models $\text{Be}(\pi(\tilde{\mathbf{x}}^T \beta_0))$ with strongly distorted regressors $\tilde{\mathbf{x}}$ in place of \mathbf{x} . A secondary purpose of this section is to verify the effect of enhancing on the performance of the Med-estimator $\hat{\beta}_n$.

Our simulations are at first generated by a simple model where the correctness of simulations can be verified by theoretical means such as the central limit theorem, and where the eventual inaccuracies of computational algorithms can be detected and excluded. As the next step, they are generated by a more realistic model used formerly by Bianco and Yohai (1996) for verification of the robustness of their BY-estimator $\beta_n^{(2)}$.

Table 5

Mean absolute errors (4.16) and rejection rates RR for selected estimators $\tilde{\beta}_n = (\tilde{\beta}_{n0}, \tilde{\beta}_{n1})^T$ of the true parameter $\beta_0 = (\beta_{00}, \beta_{01})^T = (-2.82, 2.82)^T$ in the logistic regression model (4.20) ε -contaminated by leverage points and preserving $\Pr(Y = 1) = 0.2$

ε	$\tilde{\beta}_n$	$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
		MAE	RR (%)						
0	MLE	0.86	1	0.58	0	0.24	0	0.16	0
	Morg	0.95	3	0.64	0	0.26	0	0.18	0
	CH	0.84	1	0.58	0	0.25	0	0.17	0
	BY	1.53	8	0.94	1	0.31	0	0.21	0
	Med	3.27	18	2.55	9	0.92	0	0.50	0
	5-Med	2.79	22	1.64	10	0.52	0	0.33	0
	10-Med	2.50	24	1.73	10	0.46	0	0.31	0
0.05	MLE	1.07	0	1.00	0	1.06	0	1.07	0
	Morg	1.05	1	0.85	0	0.82	0	0.82	0
	CH	0.99	0	0.87	0	0.88	0	0.89	0
	BY	1.36	4	0.91	1	0.62	0	0.61	0
	Med	2.66	12	2.36	7	0.83	0	0.62	0
	5-Med	2.00	17	1.78	6	0.64	0	0.59	0
	10-Med	1.95	17	1.55	8	0.62	0	0.58	0
0.1	MLE	1.48	0	1.50	0	1.55	0	1.55	0
	Morg	1.42	1	1.34	0	1.40	0	1.40	0
	CH	1.41	0	1.37	0	1.42	0	1.42	0
	BY	1.44	3	1.17	0	1.21	0	1.22	0
	Med	2.46	10	1.72	3	1.11	0	1.11	0
	5-Med	1.95	12	1.46	4	1.10	0	1.13	0
	10-Med	1.96	15	1.36	5	1.10	0	1.13	0
0.2	MLE	2.04	0	2.05	0	2.06	0	2.07	0
	Morg	2.01	0	2.01	0	2.02	0	2.03	0
	CH	2.01	0	2.01	0	2.02	0	2.03	0
	BY	1.97	0	1.97	0	2.00	0	2.01	0
	Med	2.22	5	2.00	2	1.83	0	1.88	0
	5-Med	1.96	5	1.90	2	1.88	0	1.89	0
	10-Med	2.04	5	1.84	1	1.88	0	1.89	0
0.3	MLE	2.37	0	2.37	0	2.39	0	2.39	0
	Morg	2.37	0	2.36	0	2.38	0	2.38	0
	CH	2.37	0	2.36	0	2.38	0	2.38	0
	BY	2.36	0	2.36	0	2.38	0	2.38	0
	Med	2.36	2	2.28	0	2.33	0	2.34	0
	5-Med	2.31	1	2.29	0	2.34	0	2.34	0
	10-Med	2.32	2	2.29	0	2.34	0	2.34	0

The achieved minima are printed bold.

(I) The first model is $\text{Be}(\pi(x\beta_0))$ for a real valued parameter β_0 and regressors x attaining at time $1 \leq i \leq n$ mutually independent random values $x_i \in X = \{-1, 1\}$ with equal probabilities $\Pr(x_i = -1) = \Pr(x_i = 1) = \frac{1}{2}$. We use the concrete parameter value $\beta_0 = -\ln 4$ for which $p_0 = \pi(\beta_0) = 0.2$. For the levels $\varepsilon \in \{0, 0.1, 0.2, 0.3\}$ and $1 \leq i \leq n$ we consider the contaminated logistic regression data

$$Y_i \sim (1 - \varepsilon) \text{Be}(\pi(x_i\beta_0)) + \varepsilon \text{Be}(1 - \pi(x_i\beta_0)), \tag{4.1}$$

i.e. instead of the assumed Bernoulli data $Y_i \sim \text{Be}(\pi(x_i\beta_0))$ we generate the distorted Bernoulli data

$$Y_i \sim \text{Be}(x_i\pi(\beta_\varepsilon)) \quad \text{for } \beta_\varepsilon = \pi^{-1}(\pi(\beta_0) + \varepsilon(1 - 2\pi(\beta_0))) \tag{4.2}$$

(cf. (1.6) and (1.2)). In this simulation model the above-mentioned estimators $\beta_n^{(1)} - \beta_n^{(3)}$ and $\hat{\beta}_n$ take on the univariate forms $\beta_n^{(1)} - \beta_n^{(3)}$ and $\hat{\beta}_n$. Moreover, their formulas considerably simplify. For derivation of these formulas it is useful to consider the sets

$$A_n^+ = \{1 \leq i \leq n : x_i = 1\}, \quad A_n^- = \{1, \dots, n\} - A_n^+ \tag{4.3}$$

and the statistics

$$\tilde{Y}_i = \begin{cases} Y_i & \text{if } i \in A_n^+, \\ 1 - Y_i & \text{if } i \in A_n^-. \end{cases} \tag{4.4}$$

Clearly,

$$\tilde{Y}_i \sim \text{Be}(\pi(\beta_\varepsilon)) \quad \text{for all } 1 \leq i \leq n \text{ and } \beta_\varepsilon \text{ given by (4.2)}. \tag{4.5}$$

By (1.10) and (1.11), the Morg-estimator $\beta_n^{(1)}$ solves the equation

$$\sum_{i \in A_n^+} (Y_i - \pi(\beta)) - \sum_{i \in A_n^-} (Y_i - 1 + \pi(\beta)) = 0,$$

i.e. coincides with the MLE β_n specified in the model of (4.5) by

$$\beta_n = \pi^{-1}(\tilde{p}_n) \quad \text{for } \tilde{p}_n = \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i \tag{4.6}$$

for \tilde{Y}_i given in (4.5).

As to the BY-estimator $\beta_n^{(2)}$, one can deduce from (1.3), (1.9) and (1.12)–(1.14) that $\beta_n^{(2)} = \pi^{-1}(\alpha_n)$ for

$$\alpha_n = \arg \min_{\alpha \in (0,1)} L(p_n, \alpha), \tag{4.7}$$

where

$$L(p_n, \alpha) = \tilde{p}_n \rho(-\ln \alpha) + (1 - \tilde{p}_n) \rho(-\ln(1 - \alpha)) + \rho_0(\alpha)$$

for \tilde{p}_n given by (4.6) and the functions ρ, ρ_0 that appear in (1.12). It is easy to see that

$$\frac{d}{d\alpha} L(\tilde{p}_n, \alpha) = (\alpha - \tilde{p}_n) \left[\frac{\rho'(-\ln \alpha)}{\alpha} + \frac{\rho'(-\ln(1 - \alpha))}{1 - \alpha} \right],$$

where by (1.15) the derivative $\rho'(t)$ is non-negative for $t > 0$ and positive for $0 < t < 1$. Therefore $\alpha_n = \tilde{p}_n$ is the unique arg min in (4.7) which means that $\beta_n^{(2)}$ coincides with the estimator β_n given in (4.6).

The CH-estimator $\beta_n^{(3)}$ differs from $\beta_n^{(2)}$ by a somewhat modified function $\rho'(t)$ and an argument similar to that given above leads to the conclusion that in the model under consideration $\beta_n^{(3)}$ coincides with the estimator β_n of (4.6) also.

Let us now turn our attention to our Med-estimator $\hat{\beta}_n$ and its k -enhanced versions $\hat{\beta}_{n*k}$. By Definition 3.1,

$$\hat{\beta}_n = \pi^{-1}(\hat{p}_n) = \ln \frac{\hat{p}_n}{1 - \hat{p}_n}, \tag{4.8}$$

where

$$\hat{p}_n = \arg \min_p \left(\sum_{i \in A_n^+} |Z_i - m(p)| + \sum_{i \in A_n^-} |Z_i - m(1 - p)| \right) = \arg \min_p \left(\sum_{i \in A_n^+} |Y_i + U_i - m(p)| + \sum_{i \in A_n^-} |Y_i + U_i - 2 + m(p)| \right).$$

The second equality holds because from (1.21) we deduce that $m(1 - p) = 2 - m(p)$. But

$$|Y_i + U_i - 2 + m(p)| = |\tilde{Y}_i + \tilde{U}_i - m(p)|,$$

where \tilde{Y}_i is defined by (4.4) and $\tilde{U}_i = 1 - U_i$ is the same uniform $U(0, 1)$ -variable as U_i . Therefore

$$\hat{p}_n = \arg \min_p \left(\sum_{i=1}^n |\tilde{Z}_i - m(p)| \right) = m^{-1}(\tilde{Z}_{(n/2)}) \tag{4.9}$$

for

$$\tilde{Z}_i = \tilde{Y}_i + U_i \quad (\text{cf. (4.5), (1.17)}) \quad \text{and} \quad \tilde{Z}_{(n/2)} = \text{median}(\tilde{Z}_1, \dots, \tilde{Z}_n). \tag{4.10}$$

Similarly,

$$\hat{\beta}_{n*k} = \pi^{-1}(\hat{p}_{n*k}) = \ln \frac{\hat{p}_{n*k}}{1 - \hat{p}_{n*k}}, \tag{4.11}$$

where

$$\hat{p}_{n*k} = \arg \min_p \left(\sum_{j=1}^k \sum_{i=1}^n |\tilde{Z}_{ij} - m(p)| \right) = m^{-1}(\tilde{Z}_{nk/2}) \tag{4.12}$$

for
$$\tilde{Z}_{ij} = \tilde{Y}_i + U_{ij} \quad (\text{cf. (4.5)}) \quad \text{and} \quad \tilde{Z}_{(nk/2)} = \text{median}(\tilde{Z}_{11}, \dots, \tilde{Z}_{kn}), \quad (4.13)$$

where U_{ij} are independent and uniformly distributed on $(0, 1)$.

In Table 2 1000 realizations $\tilde{\beta}_n^{(1)}, \dots, \tilde{\beta}_n^{(1000)}$ have been used to evaluate the *mean absolute errors*

$$\text{MAE} = \frac{1}{1000} \sum_{l=1}^{1000} |\tilde{\beta}_n^{(l)} - \beta_0|, \quad (4.14)$$

standard deviations (roots of the *mean squared errors*)

$$\text{STD} = \left(\frac{1}{1000} \sum_{l=1}^{1000} (\tilde{\beta}_n^{(l)} - \beta_0)^2 \right)^{1/2} \quad (4.15)$$

and also the *rejection rates* RR, i.e. the percentages of those data vectors $\tilde{\mathbf{Z}}_n = (\tilde{Z}_1, \dots, \tilde{Z}_n)$ or data matrices $\tilde{\mathbf{Z}}_{k,n} = (\tilde{Z}_{11}, \dots, \tilde{Z}_{kn})$ for which the corresponding estimator $\tilde{\beta}_n$ or $\tilde{\beta}_{n*k}$ was undefined because the median $\tilde{Z}_{(n/2)}$ or $\tilde{Z}_{(nk/2)}$ was outside of the definition domain $(\frac{1}{2}, \frac{3}{2})$ of the inverse median function $m^{-1}(z)$ (cf. Fig. 2).

From the first (uncontaminated) sector of Table 2 one can verify that $\sqrt{n} \text{STD}(\beta_n)$ agrees very well already for not too large n with the asymptotic standard deviation

$$\sigma(\text{MLE}) = \frac{1}{\sqrt{\pi(\beta_0)(1 - \pi(\beta_0))}} = \frac{1}{\sqrt{0.2 \times 0.8}} = 2.5$$

obtained from the limit theorem of Example 2.1. E.g., $\sqrt{50} \text{STD}(\beta_{50}) = 2.47$, or $\sqrt{100} \text{STD}(\beta_{100}) = 2.48$. Similarly one can verify that for the Med-estimator the scaled standard deviation $\sqrt{n} \text{STD}(\hat{\beta}_n)$ tends with increasing n to the corresponding theoretic value

$$\sigma(\text{Med}) = \frac{1}{\pi(\beta_0)} = \frac{1}{0.2} = 5,$$

e.g. $\sqrt{100} \text{STD}(\hat{\beta}_{100}) = 6.17$ and $\sqrt{500} \text{STD}(\hat{\beta}_{500}) = 5.5$.

Similarly, from the ε -contaminated sector of Table 2 one can verify that $\sqrt{n} \text{STD}(\beta_n)$ and $\sqrt{n} \text{STD}(\hat{\beta}_n)$ approximate the theoretical standard deviations

$$\sigma_n(\text{MLE}) = \sqrt{n(\beta_\varepsilon - \beta_0)^2 + 1/[\pi(\beta_\varepsilon)(1 - \pi(\beta_\varepsilon))]}$$

and

$$\sigma_n(\text{Med}) = \sqrt{n(\beta_\varepsilon - \beta_0)^2 + (e^{|\beta_\varepsilon|} + 1)^2}.$$

E.g., for $\varepsilon=0.2$ we get $\beta_\varepsilon = -0.7538$ and $(\beta_\varepsilon - \beta_0)^2 = 0.4000$ so that we compute $\sigma_n(\text{MLE}) = 6.68$ while in the 0.2-contaminated sector of Table 2 we find $\sqrt{100} \text{STD}(\beta_{100}) = 6.54$. Similarly we compute $\sigma_n(\text{Med}) = 7.05$ while in the table we find $\sqrt{100} \text{STD}(\hat{\beta}_{100}) = 6.85$ which is a satisfactory agreement.

All sectors of Table 2 also clearly indicate that for fixed n the error measures $\text{STD}(\hat{\beta}_{n*k})$ and $\text{MAE}(\hat{\beta}_{n*k})$ tend, with increasing k , to the respective values $\text{STD}(\beta_n)$ and $\text{MAE}(\beta_n)$, in accordance with what has been asserted by Theorem 3.1. However, the main message of Table 2 is that for larger sample sizes n our Med-estimators $\hat{\beta}_n$ better resist to higher levels of contamination than the remaining four estimators known from the previous literature.

Table 3 in some sense more strongly supports the observations and conclusions drawn from Table 2. In this table the ε -fractions of the data Y_i were generated by the false regressors $\tilde{x}_i = -10x_i$ so that they represent leverage points in the common sense of the regression analysis.

(II) Let us now turn to the second model and the corresponding results in Tables 4 and 5. These tables present similar characteristics as Tables 2 and 3 but based on simulations in more realistic logistic regression models with bivariate parameters $\beta = (\beta_0, \beta_1)^T \in \mathbb{R}^2$ and bivariate regressors $\mathbf{x} = (x_0, x_1)^T \in \mathbb{R}^2$. This means among others that in these tables the MAE formula (4.14) is replaced by

$$\text{MAE} = \frac{1}{2000} \sum_{l=1}^{1000} (|\tilde{\beta}_{n0}^{(l)} - \beta_{00}| + |\tilde{\beta}_{n1}^{(l)} - \beta_{01}|) \quad (4.16)$$

obtained from 1000 simulated realizations $\tilde{\beta}_n^{(l)} = (\tilde{\beta}_{n0}^{(l)}, \tilde{\beta}_{n1}^{(l)})^T$ of an estimator $\tilde{\beta}_n = (\tilde{\beta}_{n0}, \tilde{\beta}_{n1})^T$ of true parameters $\beta_0 = (\beta_{00}, \beta_{01})^T$.

In order to achieve an ideal comparability, all selected estimators $\tilde{\beta}_n$ are evaluated for the same simulated data vectors $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ and the related smoothed data vectors or matrices

$$\mathbf{Z}_n = (Y_1 + U_1, \dots, Y_n + U_n) \quad \text{or} \quad \mathbf{Z}_{k,n} = (Y_1 + U_{11}, \dots, Y_n + U_{kn}) \quad \text{cf. ((1.17), (3.5)).}$$

If numerical evaluation of one of these estimators fails then the corresponding \mathbf{Y}_n is rejected and replaced by a new independent realization. This procedure is repeated until the computer successfully numerically evaluates 1000 realizations $\tilde{\beta}_n^{(1)}, \dots, \tilde{\beta}_n^{(1000)}$ for each of the selected estimators $\tilde{\beta}_n$. The *rejection rate* RR then specifies for each selected estimator the percentage of the data vectors \mathbf{Y}_n rejected during evaluation of the desired 1000 realizations.

Let us explain what is meant by the numerical evaluation of estimates $\tilde{\beta}_n^{(l)} = (\tilde{\beta}_{n0}^{(l)}, \tilde{\beta}_{n1}^{(l)})^T$ used in the formulas (4.16) and leading to the MAE's and RR's that appear in the columns and rows of Tables 4 and 5. It is the evaluation in accordance with the definition of each corresponding estimator given above, using the iteration procedures presented in the *IMSL C Numerical Libraries*, version 3.0. The minimization of a function of two variables uses a quasi-Newton method (for details see Dennis and Schnabel, 1983, Appendix A), and systems of equations are solved using a modified Powell hybrid algorithm (for further description see Moré et al., 1980). The initial iteration seeds for the MLE $\beta_n = (\beta_{n0}, \beta_{n1})^T$ were the true parameters $\beta_0 = (\beta_{00}, \beta_{01})^T$ and the initial iteration seeds for all the remaining estimates $\tilde{\beta}_n = (\tilde{\beta}_{n0}, \tilde{\beta}_{n1})^T$ were the MLE's $\beta_n = (\beta_{n0}, \beta_{n1})^T$. Since the values of the estimates of Tables 4 and 5 are numerically less reliable than those of Tables 2 and 3, the MAE's presented in these tables are rounded to two decimal places.

Table 4 was obtained from the simulated data $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ generated by the contaminated logistic source

$$Y_i \sim (1 - \varepsilon) \text{Be}(\pi(\mathbf{x}_i^T \beta_0)) + \varepsilon \text{Be}(1 - \pi(\mathbf{x}_i^T \beta_0)) \quad \text{cf. (4.1)} \tag{4.17}$$

with the random regressors

$$\mathbf{x}_i = (x_{i0} \equiv 1, x_{i1} \sim \text{N}(0, 1))^T \tag{4.18}$$

and true parameters $\beta_0 = (\beta_{00}, \beta_{01})^T = (-2.82, 2.82)^T$ leading to the probability

$$\text{Pr}(Y_i = 1) \equiv E\pi(\mathbf{x}_i^T \beta_0) = 0.2. \tag{4.19}$$

This data source was used previously by Bianco and Yohai (1996) to demonstrate experimentally the robustness of their BY-estimator denoted above as $\beta_n^{(2)}$.

Table 5 was obtained for the simulated data $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ with the contaminated data source (4.17) replaced by

$$Y_i \sim (1 - \varepsilon) \text{Be}(\pi(\mathbf{x}_i^T \beta_0)) + \varepsilon \text{Be}(\pi(\tilde{\mathbf{x}}_i^T \beta_0)) \tag{4.20}$$

with the random regressors \mathbf{x}_i and true parameters β_0 as used in Table 4, but with different random regressors $\tilde{\mathbf{x}}_i$, given by the formula

$$\tilde{\mathbf{x}}_i = \left(1, \tilde{x}_{i1} = \beta_{00} + 4 \text{sign} \left[-\frac{\beta_{00}}{\beta_{01}} - x_{i1} \right] \beta_{01} \right), \quad x_{i1} \sim \text{N}(0, 1) \quad \text{cf. (4.18)}. \tag{4.21}$$

We see that in (4.20) the source $\text{Be}(1 - \pi(\mathbf{x}_i^T \beta_0))$ of contaminating data is replaced by the source $\text{Be}(\pi(\tilde{\mathbf{x}}_i^T \beta_0))$ of leverage points where the regressors $\tilde{\mathbf{x}}_i = \tilde{\mathbf{x}}_i(\beta_0)$ given by (4.21) are characterized by the property

$$\pi(\mathbf{x}_i^T \beta_0) > \frac{1}{2} \quad \text{implies} \quad \pi(\tilde{\mathbf{x}}_i^T \beta_0) \approx 0$$

and

$$\pi(\mathbf{x}_i^T \beta_0) \leq \frac{1}{2} \quad \text{implies} \quad \pi(\tilde{\mathbf{x}}_i^T \beta_0) \approx 1.$$

We see that the results in Tables 4 and 5 basically agree with those in Tables 2 and 3. An obvious difference between these pairs of tables is that in the data sources (4.17)–(4.19) used in the latter tables the MLE β_n , the Morg-estimator $\beta_n^{(1)}$, the BY-estimator $\beta_n^{(2)}$ and the CH-estimator $\beta_n^{(3)}$ mutually differ. Therefore Tables 4 and 5 compare our Med-estimator $\hat{\beta}_n$ and its k -enhanced versions $\hat{\beta}_{n*k}$ with four different estimators. This means that the reading and interpretation of these two tables is more complicated than the reading and interpretation of Tables 2 and 3.

It is easy to notice another important difference between Tables 4, 5 and Tables 2 and 3. Namely, in the present tables the MAE of $\hat{\beta}_{n*k}$ does not seem to converge for increasing k to the MAE of the MLE β_n . But the MAE($\hat{\beta}_{n*k}$) still preserves the remaining properties observed in Tables 2 and 3: For $\varepsilon = 0$ and increasing k it monotonically decreases when $50 \leq n \leq 1000$ remains fixed. For $\varepsilon > 0$ and increasing k it decreases first and then increases, and the larger the sample size n the more visible this phenomenon is. Another feature shared with Tables 2 and 3 is that the conclusions drawn from the classical contaminated models in Table 4 are even more convincingly supported by Table 5 that was obtained from the models contaminated by leverage points.

But the main message in Tables 4 and 5 remains the same as in Tables 2 and 3: Our Med-estimator $\hat{\beta}_n$ is more resistant to the distortions of logistic regression models by contaminations and leverage points than the remaining four estimators when the levels of distortions are higher and the sample sizes are medium (around $n \approx 100$) or large (around $n \approx 1000$). For the medium sample sizes it is convenient to use the k -enhancing for $k \approx 10$.

The preferences between the remaining estimators β_n and $\beta_n^{(1)} - \beta_n^{(3)}$ which can be drawn from Tables 4 and 5 agree with what has been previously published in the literature, in particular with what was stated in Bianco and Yohai (1996) and Croux and Haesbroeck (2003).

Let us remark that in their simulation study Bianco and Yohai (1996) considered not only the true parameters $\beta_0 = (-2.82, 2.82)^T$ leading to the probability $\Pr(Y_i = 1)$ given in (4.19) but also true $\beta_0 = (\beta_{00}, \beta_{01})^T$ leading to the probabilities 0.3, 0.4 and 0.5. Alternatives of the Tables 4 and 5 obtained by simulating data corresponding to these β_0 can be found in Hobza et al. (2006). These tables provide additional evidence for the conclusions drawn in this paper from Tables 4 and 5. Note that the tables in Hobza et al. (2006) are larger, and also contain the results for the k -Med-estimator with $k = 50$.

Acknowledgment

The authors thank the referees and the associate editor for careful reading of the paper and many interesting improvements. Special thanks are due to the referee who proposed the procedure called statistical enhancing in this paper. This procedure has proved to be able to increase the efficiency of our estimators, in some cases up to the level achieved by the MLE. It seems that it will also remain useful in applications of statistical smoothing in another discrete statistical models.

References

- Adimari, G., Ventura, L., 2001. Robust inference for generalized linear models with application to logistic regression. *Statist. Probab. Lett.* 55 (4), 413–419.
- Agresti, A., 2002. *Categorical Data Analysis*. second ed. Wiley, New York.
- Andersen, E.B., 1990. *The Statistical Analysis of Categorical data*. Springer, New York.
- Arcones, M.A., 2001. Asymptotic distribution of regression M -estimators. *J. Statist. Plann. Inference* 97, 235–261.
- Bianco, A.M., Yohai, V.J., 1996. Robust estimation in the logistic regression model. In: *Robust Statistics, Data Analysis, and Computer Intensive Methods* (Schloss Thurnau, 1994). *Lecture Notes in Statistics*, vol. 109, Springer, New York, pp. 17–34.
- Chen, X.R., Zhao, L., Wu, Y., 1993. On conditions of consistency of ML_1N estimates. *Statist. Sinica* 3, 9–18.
- Croux, C., Haesbroeck, G., 2003. Implementing the Bianco and Yohai estimator for logistic regression. *Comput. Statist. Data Anal.* 44, 273–295.
- Dennis Jr., J.E., Schnabel, R.B., 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, NJ.
- Gervini, D., 2005. Robust adaptive estimators for binary regression models. *J. Statist. Plann. Inference* 131, 297–311.
- Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.M., Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- Hobza, T., Pardo, L., Vajda, I., 2005. Median estimators in generalized logistic regression. Research Report DAR-UTIA 2005/40, Institute of Information Theory, Prague (available at <http://dar.site.cas.cz/?publication=1007>).
- Hobza, T., Pardo, L., Vajda, I., 2006. Robust median estimators in logistic regression. Research Report DAR-UTIA 2006/31, Institute of Information Theory, Prague (available at <http://dar.site.cas.cz/?publication=1089>).
- Jurečková, J., Procházka, B., 1994. Regression quantiles and trimmed least squares estimator in nonlinear regression model. *Nonparametric Statist.* 3, 201–222.
- Jurečková, J., Sen, P.K., 1996. *Robust Statistical Procedures*. Wiley, New York.
- Knight, K., 1998. Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* 26, 755–770.
- Koenker, R., Basset, G., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Kordzakhia, N., Mishra, G.D., Reiersølmoen, L., 2001. Robust estimation in the logistic regression model. *J. Statist. Plann. Inference* 98, 211–223.
- Liese, F., Vajda, I., 1999. M -estimators of structural parameters in pseudolinear models. *Appl. Math.* 44, 245–270.
- Liese, F., Vajda, I., 2003. A general asymptotic theory of M -estimators I. *Math. Methods Statist.* 12, 454–477.
- Liese, F., Vajda, I., 2004. A general asymptotic theory of M -estimators II. *Math. Methods Statist.* 13, 82–95.
- Maronna, R.A., Martin, R.D., Yohai, V.J., 2006. *Robust Statistics. Theory and Methods*. Wiley, New York.
- Mood, A.M., Graybill, F.A., Boes, D.C., 1974. *Introduction to the Theory of Statistics*. McGraw-Hill, New York.
- Moré, J., Burton, G., Kenneth, H., 1980. User Guide for MINPACK-1. Argonne National Laboratory Report ANL-80-74, Argonne, Illinois.
- Morgenthaler, S., 1992. Least-absolute-deviations fits for generalized linear models. *Biometrika* 79, 747–754.
- Pardo, J.A., Pardo, L., Pardo, M.C., 2006. Testing in logistic regression models based on ϕ -divergences measures. *J. Statist. Plann. Inference* 136, 982–1006.
- Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
- Pregibon, D., 1982. Resistant fits for some commonly used logistic models with medical applications. *Biometrics* 38, 485–498.
- Richardson, G.D., Bhattacharyya, B.B., 1987. Consistent L_1 -estimates in nonlinear regression for a noncompact parameter space. *Sankhya Ser. A* 49, 377–387.
- Rousseeuw, P.J., Christmann, A., 2003. Robustness against separation and outliers in logistic regression. *Comput. Statist. Data Anal.* 43, 315–332.
- Yohai, V.J., 1987. High breakdown point high efficiency robust estimates for regression. *Ann. Statist.* 15, 642–656.
- Zwanzig, S., 1997. On L_1 -norm Estimators in Nonlinear Regression and in Nonlinear Error-in-Variables Models. *IMS Lecture Notes*, vol. 31, Hayward, pp. 101–118.