Robust Median Estimators in General Logistic Regression*

T. Hobza[1]; L. Pardo[2] and I. Vajda[3]

[1]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague.

[2]Department of Statistics and O. R., Complutense University of Madrid, Madrid.

[3]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague.

Abstract

The paper considers generalized logistic regression models which include the classical model with binary responses governed by the Bernoulli law as well as the models with responses governed by general discrete or continuous laws depending on the logistic regression function. It introduces a new median estimator of the logistic regression parameters employing smoothed data in the discrete case. Consistency and asymptotic normality theorems are presented for this estimator. its sensitivity to contaminations of the logistic regression data is extensively studied by simulations and compared with the sensitivity of some robust estimators previously introduced to logistic regression. The median estimator is demonstrated to be more robust than these estimators for higher levels of contamination.

Key words: Logistic regression, Median estimator, Morgenthaler estimator, Bianco and Yohai estimator, Consistency, Asymptotic normality, Robustness.
AMS 2000 Subject Classification: 62F10, 62F12, 62F35.

1 Introduction

In this paper we study estimation of the parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ in the generalized logistic regression where independent observations $Y_1, ..., Y_n$ depend on $\boldsymbol{\beta}_0$ and regressors $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$, all from $\mathbb{R}^d$, and are distributed by $F_{\pi(\boldsymbol{x}_1^T\boldsymbol{\beta}_0)}, ..., F_{\pi(\boldsymbol{x}_n^T\boldsymbol{\beta}_0)}$ for the logistic function $\pi(t) = e^t/(1 + e^t)$ and a given family $(F_\pi : \pi \in (0,1))$ of distribution functions on $\mathbb{R}$. For the Bernoulli distribution functions $F_\pi(y)$ with jumps $1 - \pi$ and π at $y = 0$ and $y = 1$

*This work was supported by the grants DGES PB2003-892, AV CR 107 5403 and MSMT 1M0572

this reduces to the classical logistic regression with binary observations $Y_1, ..., Y_n$ taking on value 1 with probabilities $\pi\left(\boldsymbol{x}_1^T\boldsymbol{\beta}_0\right), ..., \pi\left(\boldsymbol{x}_n^T\boldsymbol{\beta}_0\right)$ and value 0 with the complementary probabilities $1-\pi\left(\boldsymbol{x}_1^T\boldsymbol{\beta}_0\right), ..., 1-\pi\left(\boldsymbol{x}_n^T\boldsymbol{\beta}_0\right)$.

We propose the median estimator

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left|Y_i - m\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right| \tag{1.1}$$

where $m(\pi)$ is the median of the distribution F_π. This estimator was not previously considered in the logistic regression because for the most important Bernoulli model (as well as for all other discrete logistic regression models) the median function $m(\pi) = F_\pi^{-1}(1/2)$ is not sensitive to small variations of the parameter $\pi \in (0,1)$. Therefore the classical median estimator (1.1) cannot be consistent in these models. The originality of our approach consists in the random transformation of discrete integer-valued observations Y_i by the formula

$$\widetilde{Y}_i = Y_i + W_i, \qquad 1 \le i \le n \tag{1.2}$$

where $W_1, ..., W_n$ are mutually (and also on $Y_1, ..., Y_n$) independent random variables uniformly distributed on the interval $(0,1)$. All discrete logistic regression data are supposed to be standardly transformed in this manner. This transformation is statistically sufficient since the original observations Y_i can be recovered from $\widetilde{Y}_i$ as the integer parts $\left[\widetilde{Y}_i\right]$. At the same time the median functions $\widetilde{m}(\pi) = \widetilde{F}_\pi^{-1}(1/2)$ of the transformed observations (1.2) are already one-one on the interval $(0,1)$. We prove a consistency theorem and an asymptotic normality theorem for the median estimators (1.1) in the general logistic regression models under consideration.

It is known (cf. e. g. Hampel et al (1986), Yohai (1987), Jurečková and Sen (1996), Zwanzig (1997)) that the median estimator of parameters of linear and non-linear regression is robust with respect to contamination of observations from the assumed statistical models. This naturally leads to the conjecture that the median estimator proposed in this paper for the general logistic regression is robust too.

In this paper, and also in Hobza et al (2005), we explicitly evaluated in some simple examples the estimates (1.1) and compared them on simulated contaminated as well as noncontaminated data with the MLE and/or some robust estimators tailor-made for the logistic regression, demostrating in this manner the acceptability of the robustness conjecture stated above.

In particular, we compared our median estimator with the L_1-estimator of Morgenthaler (1992) and with the robust estimator of Bianco and Yohai (1996). Our simulations demonstrated that for heavier contaminations and larger sample sizes the robustness of our estimator dominates these two.

2 Models and Estimators

A large class of statistical models assumes independent real valued observations $Y_1, ..., Y_n$ of the form

$$Y_i \sim F_{u\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)}(y), \qquad 1 \leq i \leq n. \tag{2.1}$$

Here $\boldsymbol{x}_i \in \mathbb{R}^d$ are vectors of explanatory variables (regressors), $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ is a vector of true parameters and $\boldsymbol{x}^T\boldsymbol{\beta} = \sum_{j=1}^d x_j\beta_j$ denotes the scalar product of $\boldsymbol{x} = (x_1, ..., x_d)^T \in \mathbb{R}^d$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_d)^T \in \mathbb{R}$. Further, $u : \mathbb{R} \longmapsto \Theta$ 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 to find mappings $\widehat{\boldsymbol{\beta}}_n = \widehat{\boldsymbol{\beta}}_n(Y_1, ..., Y_n)$ from $\mathbb{R}^n$ into $\mathbb{R}^d$ which can be used to estimate the unknown parameters $\boldsymbol{\beta}_0$ on the basis of observations (2.1). Various asymptotic or non-asymptotic properties are usually required from estimators $\widehat{\boldsymbol{\beta}}_n$.

The desirable properties are often found in the class of so-called *least absolute deviation estimators* (*LAD*-estimators, or briefly L_1-estimators) defined by

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \left|Y_i - \mu\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right|. \tag{2.2}$$

From the extensive literature dealing with these estimators one can mention Koenker and Bassett (1978), Richardson and Bhattacharyya (1987), Pollard (19991), Morgenthaler (1992), Chen, Zhao and Wu (1993), Jurečková and Procházka (1994), Knight (1998), Arcones (2001), Liese and Vajda (1999, 2003, 2004) and others cited in these papers.

The L_1-estimators are usually used in the linear regression where the observations

$$Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta}_0 + W_i, \qquad 1 \leq i \leq n \tag{2.3}$$

depend on independent additive zero-mean random errors W_i or in the nonlinear regression where (2.3) is replaced by more general formula

$$Y_i = \mu\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) + W_i, \qquad 1 \leq i \leq n \tag{2.4}$$

for a given smooth function $\mu : \mathbb{R} \longmapsto \mathbb{R}$. If μ is the identity mapping $\mu(y) = y$, then the model (2.4) reduces to the linear regression (2.3). In the linear regression the L_1-estimator (2.2) takes on the simple form

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \left|Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right|. \tag{2.5}$$

This estimator was considered e.g. by Pollard (1991). It is often extended to the class of L_p-estimators with $p \geq 1$ where

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \left|Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right|^p, \tag{2.6}$$

see e. g. Arcones (2001).

The L_1-estimator (2.2) is sometimes applied also beyond the linear and nonlinear regression. For example Morgenthaler (1992) used it for more general models of the type (2.1) with

$$\mu(y) = m(u(y)) \tag{2.7}$$

where $m : \Theta \longmapsto \mathbb{R}$ is the *mean function* of the family $\mathcal{F}$ defined for every $\theta \in \Theta$ by

$$m(\theta) = \int y dF_\theta(y). \tag{2.8}$$

He considered also the L_p-generalizations

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left|Y_i - \mu\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right|^p \tag{2.9}$$

similar to (2.6) for $\mu(y)$ given by (2.7) and (2.8).

Koenker and Bassett (1987), Jurečková and Procházka (1994), Liese and Vajda (2003, 2004) and many other authors generalized the above mentioned L_1-estimators in other way. Namely, instead of the L_1-criterion $L_1(y) = |y|$ they considered the $\widetilde{L}_\alpha-$criteria where

$$\widetilde{L}_\alpha(y) = |y| (\alpha I(y > 0) + (1-\alpha) I(y < 0)),$$

for the indicator function $I(.)$ and a fixed $\alpha \in (0,1)$. Obviously, the symmetric $\widetilde{L}_{1/2}(y) = \frac{1}{2}|y|$ defines the same estimator as $L_1(y) = |y|$.

In this paper we are interested in the special L_1-estimators (2.2) called *median estimators.* They are defined by (2.2) for the function $\mu(y)$ given formally by the same composition formula as (2.7) but with $m : \Theta \mapsto \mathbb{R}$ being the median function of the family $\mathcal{F}$. This function is for every $\theta \in \Theta$ defined by

$$m(\theta) = med(F_\theta) = F_\theta^{-1}(1/2) = \inf\{y \in \mathbb{R} : F_\theta(y) \geq 1/2\}. \tag{2.10}$$

Our attention is restricted to the important subclass of the models (2.1) where

$$Y_i \sim F_{\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)}(y), \qquad 1 \leq i \leq n \tag{2.11}$$

for $\boldsymbol{x}_i$ and $\boldsymbol{\beta}_0$ the same as in (2.1) but for the particular logistic regression function

$$\pi(t) = \frac{e^t}{1+e^t} \qquad \text{for every } t \in \mathbb{R} \tag{2.12}$$

and an arbitrary family $\mathcal{F} = \{F_\pi : \pi \in (0,1)\}$ of distribution functions on $\mathbb{R}$. The models given by (2.11), (2.12) are the *general logistic regression models.* In these models $\pi = \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)$ represents a nonlinear logistic regression and the distribution function F_π specifies the random response to this regression, see e.g. Andersen (1990), Agresti (2002), Pardo et al (2006) and others cited there. Note that here and in the sequel we use the same symbol π for the mapping $\pi(t)$ given by (2.12) and for the parameter of the random response family $\mathcal{F} = \{F_\pi : \pi \in (0,1)\}$. We hope that this will not lead to a confusion.

Important special logistic regression models are obtained if for all $\pi \in (0,1)$ the random response functions $F_\pi(y)$ are either a right-continuous distribution functions with jumps

$$p_\pi(k) = F_\pi(k) - F_\pi(k-0), \qquad k = 0, 1, ... \tag{2.13}$$

summing up to 1, or continuous piecewise differentiable distribution functions with densities

$$f_\pi(y) = \frac{dF_\pi(y)}{dy}, \ y \in \mathbb{R}. \tag{2.14}$$

In the first case we speak about *discrete models* and in the second case about *continuous models.*

For example, the Bernoulli reponse functions

$$F_\pi(y) = (1-\pi)\, I(0 \le y < 1) + I(y \ge 1), \qquad \pi \in (0,1) \tag{2.15}$$

with the jumps (2.13) given by

$$p_\pi(0) = 1-\pi, \quad p_\pi(1) = \pi, \quad p_\pi(k) = 0 \quad \text{for } k > 1 \tag{2.16}$$

define the discrete *Bernoulli models* (2.11). The discrete geometric response functions

$$F_\pi(y) = \sum_{k=0}^{\infty} \left(1-\pi^{k+1}\right) I(k \le y < k+1), \quad \pi \in (0,1) \tag{2.17}$$

with the jumps (2.13) given by

$$p_\pi(k) = (1-\pi)\,\pi^k, \qquad k = 0, 1, ... \tag{2.18}$$

define the *discrete geometric models* (2.11). The exponential response functions

$$F_\pi(y) = 1 - \exp\{-\pi y/(1-\pi)\}\, I(y > 0), \quad \pi \in (0,1) \tag{2.19}$$

with the densities

$$f_\pi(y) = \frac{\pi}{1-\pi} \exp\{-\pi y/(1-\pi\}\, I(y > 0) \tag{2.20}$$

define the continuous *exponential models* (2.11). If the response distributions considered in (2.11) are normal with the densities

$$f_\pi(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-\ln\pi)^2}{2}\right\}, \quad \pi \in (0,1) \tag{2.21}$$

then we speak about *normal logistic regression models.*

The main object of interest of this paper can now be formally defined as follows.

Definition 2.1 The median estimator $\widehat{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the general logistic regression model given by (2.11) and (2.12) is defined by the formula

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| Y_i - m\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)\right| \tag{2.22}$$

where $m(\pi)$ is for every $\pi \in (0,1)$ given by

$$m(\pi) = med\,(F_\pi) = F_\pi^{-1}(1/2) = \inf\{y \in \mathbb{R} : F_\pi(y) \geq 1/2\}. \tag{2.23}$$

The asymptotic theory of the median estimators is presented in this paper for the general logistic regression model (2.11) of arbitrary dimension $d \in \{1, 2, ...\}$. However the applications and simulations we study mainly for the Bernoulli discrete logistic regression models of the small dimensions $d \in \{1, 2\}$. The Bernoulli models are typical and the small dimensions are simpler and transparent enough to provide a good insight into the general theory.

An obvious condition of applicability of the median estimators (2.22) is the sensitivity of the median function (2.23) to the change of parameter $\pi \in (0,1)$. This sensitivity means the strict monotonicity of the median function $m(\pi)$ on its domain $(0,1)$. If $m(\pi)$ is constant on an interval (π_1, π_2) then $m\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)$ will not distinguish between $\boldsymbol{\beta}_0$ and $\widetilde{\boldsymbol{\beta}}_0$ with both $\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)$ and $\pi\left(\boldsymbol{x}_i^T \widetilde{\boldsymbol{\beta}}_0\right)$ belonging to this interval.

For example, in the Bernoulli model (2.15), (2.16) we have the piecewise constant

$$m(\pi) = I(\pi > 1/2) = \begin{cases} 0 & \text{if } \pi \leq 1/2 \\ 1 & \text{if } \pi > 1/2. \end{cases} \tag{2.24}$$

Similarly, in the geometric model (2.17), (2.18) we have

$$m(\pi) = k \quad \text{if} \quad \left(\frac{1}{2}\right)^{1/k} < \pi \leq \left(\frac{1}{2}\right)^{1/(k+1)} \tag{2.25}$$

so that $m(\pi) = 0$ for all $\pi \in (0, 1/2]$, $m(\pi) = 1$ for all $\pi \in \left(1/2, 1/\sqrt{2}\right]$, etc.

The above required strict monotonicity of $m(\pi)$ in the regression models with discrete responses is achieved if we replace these models by their standard modifications defined as follows.

Definition 2.2. The *standard modification* of a discrete logistic regression model (2.11) with arbitrary jumps (2.13) is the continuous logistic regression model

$$\widetilde{Y}_i = Y_i + W_i, \qquad 1 \leq i \leq n, \tag{2.26}$$

where W_i are an independent noise random variables uniformly distributed on the interval $(0,1)$ and independently added to the discrete observations Y_i if the original model (2.11).

It is clear that the probability density function of the continuous observations $\widetilde{Y}_i$ are

$$f_i(\widetilde{y}) = \sum_{k=0}^{\infty} p_i(k)\, I(k \le \widetilde{y} < k+1) \tag{2.27}$$

where

$$p_i(k) = p_{\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)}(k), \qquad k = 0, 1, ... \tag{2.28}$$

are the probabilities of observations Y_i in the original model (2.11).

The continuously distributed observations $\widetilde{Y}_i$ of (2.26) can be viewed as obtained by transmission of the original discrete observations Y_i via an additional observation channel with additive noise W_i. Obviously, each transmission of observations through a channel represents a stochastic transformation which cannot increase the statistical information about $\boldsymbol{\beta}_0$ (cf. e.g. Vajda (1973)). This information typically decreases but in special channels it can be preserved. One of such channels is considered in (2.26). Namely, in (2.26) the channel inputs Y_i can be recovered from the outputs $\widetilde{Y}_i$ by the formula

$$Y_i = \left[\widetilde{Y}_i\right] \text{ a.s.}, \qquad 1 \le i \le n, \tag{2.29}$$

where $[\widetilde{y}] \in \{-1, 0, 1...\}$ denotes the whole part of $\widetilde{y} \in \mathbb{R}$, i.e.,

$$[\widetilde{y}] \le \widetilde{y} < [\widetilde{y}] + 1. \tag{2.30}$$

In other words, the standard modification of a discrete logistic regression model preserves the information contained in the observations.

By (2.26), (2.27) and (2.11), the standardly modified discrete logistic regression model is of the form

$$\widetilde{Y}_i \sim \widetilde{F}_{\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)}(y), \qquad 1 \le i \le n \tag{2.31}$$

where $\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)$ is the same as in (2.11) but the response family $\widetilde{\mathcal{F}} = \left\{\widetilde{F}_\pi : \pi \in (0,1)\right\}$ is continuous with the densities

$$\widetilde{f}_\pi(y) = \sum_{k=1}^{\infty} p_\pi(k)\, I(k \le y < k+1) \tag{2.32}$$

for the discrete probabilities $p_\pi(k)$ of the original model introduced in (2.13).

For example, the *standardly modified Bernoulli model* with the discrete probabilities (2.16) is the continuous model

$$Y_i \sim f_i(y) = \left(1 - \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right) I(0 \le y < 1) + \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) I(1 \le y < 2), \tag{2.33}$$

and the *standardly modified geometric model with the discrete probabilities* (2.18) is the continuous model

$$Y_i \sim f_i(y) = \left(1 - \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right) \sum_{k=0}^{\infty} \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)^k I(k \le y < k+1). \tag{2.34}$$

By (2.33), the standardly modified Bernoulli model has the response distribution function

$$F_\pi(y) = (1-\pi)\ \ y\, I(0<y\le 1) + [1-\pi+\pi(y-1)]\, I(1<y\le 2) \tag{2.35}$$

and consequently the median function

$$m(\pi) = 1 + \frac{\pi - 1/2}{\pi \vee (1-\pi)} = 1 + \frac{\pi - 1/2}{1/2 + |\pi - 1/2|}, \qquad \pi \in (0,1). \tag{2.36}$$

This median function is strictly increasing on (0,1) with the bell-shaped continuous positive derivative

$$m'(\pi) = \frac{1}{2\left[\pi^2 \vee (1-\pi)^2\right]}.$$

Note that here and in the sequel,

$$a \vee b = \max\{a,b\} \qquad \text{and} \qquad a \wedge b = \min\{a,b\}. \tag{2.37}$$

For the standardly modified geometric model the response distribution function

$$F_\pi(y) = \sum_{k=0}^{\infty} \left[1-\pi^k + \pi^k(1-\pi)(y-k)\right] I(k<y\le k+1) \tag{2.38}$$

is piecewise linear connecting the planar points

$$\left(k; F_\pi(k-1)\right) = \left(k; 1-\pi^k\right) \ \text{ for } k = 0,1,...$$

Therefore the median function is

$$m(\pi) = k + \frac{\pi^k - 1/2}{\pi^k(1-\pi)} \qquad \text{for} \qquad \left(\frac{1}{2}\right)^{1/k} < \pi \le \left(\frac{1}{2}\right)^{1/(k+1)} \tag{2.39}$$

where the term

$$\frac{\pi^k - 1/2}{\pi^k(1-\pi)}$$

continuously and strictly increases from 0 to 1 for $1/2 \le \pi^k \le (1/2)^{k/(k+1)}$. Therefore the median function $m(\pi)$ is continuous and strictly increasing in the domain $\pi \in (0,1)$.

3 Asymptotic for general median estimator

In this section we study the asymptotic of the median estimator from Definition 2.1. Remind that this estimator defined by the formula

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_\beta \sum_{i=1}^{n} \left|Y_i - m\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)\right| \tag{3.1}$$

estimates the true parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ of the general logistic regression with independent responses $\left(Y_i \sim F_{\pi(\boldsymbol{x}_i^T\boldsymbol{\beta}_0)} : 1 \le i \le n\right)$ to given regressors $(\boldsymbol{x}_i : 1 \le i \le n)$. Here $\pi(t) = e^t/(1+e^t)$ is a mapping $\mathbb{R} \longmapsto (0,1)$ and $m(\pi) = med(F_\pi)$ is a mapping $(0,1) \longmapsto \mathbb{R}$ depending on a given class of distribution functions $(F_\pi : \pi \in (0,1))$ specifying the responses.

Our results will be based on the theorems of Liese and Vajda (1999, 2003, 2004) concerning asymptotics of general $M-$estimators of parameters in structural statistical models. Let us start with conditions of consistency for the estimator (3.1). Liese and Vajda (1999) studied the consistency of more general estimators

$$\boldsymbol{\beta}_n = \arg\min_{\beta} \sum_{i=1}^{n} \rho\left(Y_i - \tau\left(u\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)\right) \tag{3.2}$$

where Y_i are observations from the general model (2.1) and $\rho : \mathbb{R} \longmapsto (0,\infty]$ and $\tau : \Theta \longmapsto \mathbb{R}$ are given functions. We see that our median estimator (3.1) is a special case of (3.2) for $\rho(y) = |y|$, $\Theta = (0,1)$ and $\tau = m$.

Adapted to the present situation, the consistency conditions of Theorem 2 and Lemmas 8 and 9 in Liese and Vajda (1999) are as follows.

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

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\boldsymbol{x}_i} \tag{3.3}$$

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

Remark 3.1. If the regressors $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$ 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 (3.3) 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

$$\boldsymbol{\Sigma} = \int_{\mathcal{X}} \boldsymbol{x}\boldsymbol{x}^T dQ(\boldsymbol{x}) \tag{3.4}$$

is positive. Hence for every $\boldsymbol{\beta} \in \mathbb{R}^d$ different from $\boldsymbol{\beta}_0$

$$Q\left(\boldsymbol{x} \in \mathcal{X} : \boldsymbol{x}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \neq 0\right) > 0. \tag{3.5}$$

(c3) Distributions functions $F_\pi(y)$ are continuous in both arguments $\pi \in (0,1)$ and $y \in (0,\infty)$. Moreover, for each $\pi \in (0,1)$ the function $F_\pi(y)$ has a density $f_\pi(y) = dF_\pi(y)/dy$ and

$$\int_{-\infty}^{+\infty} |y|\, f_\pi(y)\, dy < \infty, \qquad \pi \in (0,1). \tag{3.6}$$

(c4) Distributions functions F_π, $\pi \in (0,1)$ are increasing on certain intervals $I_\pi \subseteq \mathbb{R}$ in the strict sense

$$F_\pi(y_1) < F_\pi(y_2) \text{ for } y_1 < y_2 \text{ from } I_\pi \tag{3.7}$$

and constant on the complements $\mathbb{R} - I_\pi$.

(c5) Distributions functions F_π, $\pi \in (0,1)$ are stochastically ordered in the sense that for every $0 < \pi_1 < \pi_2 < 1$ and $y \in \mathbb{R}$ it holds $F_{\pi_1}(y) \geq F_{\pi_2}(y)$ where

$$F_{\pi_1}(y) > F_{\pi_2}(y) \text{ if } y \in I_{\pi_1} \cup I_{\pi_2}. \tag{3.8}$$

We shall show that **(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).

The function $\rho(y) = |y|$ is continuous, nondecreasing on $(0,\infty)$ and nonincreasing on $(-\infty,0)$ with $\rho(0) = 0$. Therefore (E1+) holds. Moreover, $\overline{\rho}(y) = \rho(y) \wedge \rho(-y)$ satisfies the relation

$$\overline{\rho}(\infty) \equiv \lim_{y\to\infty} \overline{\rho}(y) = \infty. \tag{3.9}$$

The function $\pi(t) = e^t/(1+e^t)$ is strictly increasing and continuous in the variable $t \in \mathbb{R}$. By **(c4),** $m(\pi) = med(F_\pi)$ is for every $\pi \in (0,1)$ unique and belongs to the interior of I_π. By **(c5),**

$$1/2 = F_{\pi_1}(med(\pi_1)) \geq F_{\pi_2}(med(\pi_1))$$

so that $med(\pi_2) \geq med(\pi_1)$. The assumption $med(\pi_2) = med(\pi_1)$ for $\pi_1 < \pi_2$ contradicts **(c3)** and **(c5)** because by **(c3)** $med(\pi_1)$ belongs not only to I_{π_1} but also to I_{π_2} for π_2 sufficiently close to π_1. In this case **(c5)** implies

$$F_{\pi_1}(med(\pi_1)) > F_{\pi_2}(med(\pi_1))$$

so that $med(\pi_2) > med(\pi_1)$. This proves that under **(c3)-(c5)** the function $m(\pi)$ is strictly increasing on $(0,1)$. The assumption of a jump of $m(\pi)$ at some $\pi = \pi_0 \in (0,1)$ contradicts **(c3)** and **(c4)**. Therefore **(c3)-(c5)** imply that $m(\pi)$ is continuous and strictly increasing on $(0,1)$. This means that the composed function

$$\varphi(t) = m(\pi(t)) \tag{3.10}$$

is continuous and strictly increasing in the variable $t \in \mathbb{R}$, which is the above mentioned assumption (E2). Moreover, we see that $\varphi(t)$ is bounded on $\mathbb{R}$ if and only if $m(\pi)$ is bounded on $(0,1)$ and $\overline{\varphi}(t) = |\varphi(t)| \wedge |\varphi(-t)|$ satisfies the relation

$$\overline{\varphi}(\infty) \equiv \lim_{t\to\infty} \overline{\varphi}(t) = \infty \tag{3.11}$$

if and only if $m(\pi)$ is unbounded on $(0,1)$ in the sense

$$\lim_{\pi\uparrow 1} m(\pi) = \infty \qquad \text{and} \qquad \lim_{\pi\downarrow 0} m(\pi) = -\infty. \tag{3.12}$$

The assumption (EM1) requires for every $0 < \pi_1 < \pi_2 < 1$ and every $t \in \mathbb{R}$

$$\lim_{N\to\infty} \sup_{\pi_1<\pi<\pi_2} \int_N^{\infty} |y-t|\, dF_\pi(y) = 0$$

and

$$\lim_{N\to -\infty} \sup_{\pi_1<\pi<\pi_2} \int_{-\infty}^{N} |y-t|\, dF_\pi(y) = 0.$$

We shall prove that the first relation follows from **(c1)** and **(c5).** Proof of the second relation is similar. Since for large $y \in \mathbb{R}$ we have $|y-t| \le y + |t|$ and, by **(c5),**

$$\sup_{\pi_1<\pi<\pi_2} \int_N^{\infty} dF_\pi(y) = \sup_{\pi_1<\pi<\pi_2} (1 - F_\pi(N)) = 1 - F_{\pi_2}(N),$$

the first relation will be proved if we prove that

$$\lim_{N\to\infty} \sup_{\pi_1<\pi<\pi_2} \int_N^{\infty} y\, dF_\pi(y) = 0. \tag{3.13}$$

By (3.6), for large N and every $\pi \in (0,1)$

$$N(1 - F_\pi(N)) \le \int_N^{\infty} y\, dF_\pi(y) \underset{N\to\infty}{\longrightarrow} 0.$$

Hence the integration by parts for the Lebesgue-Stiltjes integrals implies that

$$\int_N^{\infty} d\left[y(1 - F_\pi(y))\right] = \int_N^{\infty} (1 - F_\pi(y))\, dy - \int_N^{\infty} y\, dF_\pi(y)$$

so that

$$\int_N^{\infty} y\, dF_\pi(y) = N(1 - F_\pi(N)) + \int_N^{\infty} (1 - F_\pi(y))\, dy.$$

By **(c5),** this implies that

$$\sup_{\pi_1<\pi<\pi_2} \int_N^\infty y\, dF_\pi \le N\left(1 - F_{\pi_1}(N)\right) + \int_N^\infty \left(1 - F_{\pi_2}(y)\right) dy.$$

By (3.6), the right-hand side tends for $N \to \infty$ to zero which completes the desired proof of (3.13).

In view of (3.9), the condition (EM2) reduces to

$$\int |y - m(\pi)|\, dF_\pi(y) < \infty, \qquad \pi \in (0,1),$$

which follows from (3.6) in **(c3)**.

The continuity $F_\pi(y) \to F_{\pi_0}(y)$ for every $y \in \mathbb{R}$ and $\pi \to \pi_0$ required in (M1) follows from **(c3)** and the strict monotonicity of $\pi(t)$ required in (M2) was already clarified. Assumptions (M3) and (M4) concerning the regressors coincide with those in **(c1).**

Thus we proved that under **(c1)-(c5)** all assumptions of Theorem 2 and Lemma 9 in Liese and Vajda (1999) hold. In Theorem 2 and Lemma 9 we meet also the following condition.

(c6) The median function $m(\pi)$ is either bounded on $(0,1)$ or unbounded in the sense of (3.12).

The condition (3.12) is equivalent to (3.11), i.e. to $\overline{\varphi}(\infty) = \infty$. By Lemma 8 in Liese and Vajda (1999), in this case the sufficient condition of Lemma 9 reduces to (3.5) assumed in **(c2).** Hence, by the cited Theorem 2 and Lemmas 8, 9, under **(c1)-(c6)** the median estimator $\widehat{\boldsymbol{\beta}}_n$ consistently estimates true parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ provided the probability measure Q of **(c1)** and the function φ of (3.10) define a new function

$$\boldsymbol{m}(\boldsymbol{\beta}) = \int_{\mathbb{R}} \int_{\mathcal{X}} \left|y - \varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right| dF_{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)}(y)\, dQ(\boldsymbol{x}) \tag{3.14}$$

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

$$\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| \ge \varepsilon} \boldsymbol{m}(\boldsymbol{\beta}) > \boldsymbol{m}(\boldsymbol{\beta}_0). \tag{3.15}$$

This important fact will be used in the proof of the following theorem.

Theorem 3.1 If a continuous logistic regression model or a standardly modified discrete regression model satisfies **(c1)-(c6)** then the median estimator $\widehat{\boldsymbol{\beta}}_n$ of true parameter

$\boldsymbol{\beta}_0 \in \mathbb{R}^d$ is consistent if for every $0 < \pi_1 < \pi_2 < 1$ there exists $a > 0$ such that the densities f_π assumed in **(c3)** and the median function $m(\pi)$ satisfy the condition

$$\Lambda(a) \equiv \inf_{|y| \le a} \left(\inf_{\pi_1 \le \pi \le \pi_2} f_\pi (m(\pi) + y) \right) > 0. \tag{3.16}$$

Proof. By what was said above, it suffices to prove that if **(c1)-(c6)** holds then (3.16) implies (3.15). Put for φ of (3.10)

$$\Delta = \Delta(\boldsymbol{x}, \boldsymbol{\beta}) = \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right) - \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) \tag{3.17}$$

and

$$Z = Y - \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right).$$

Then the density of Z is

$$g_{\boldsymbol{x}}(z) = f_{\pi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)}\left(z + \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)\right), \qquad z \in \mathbb{R},$$

and

$$\boldsymbol{m}(\boldsymbol{\beta}) - \boldsymbol{m}(\boldsymbol{\beta}_0) = \int_{\mathcal{X}} \left[w\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)\right] dQ(\boldsymbol{x}) \tag{3.18}$$

for

$$\begin{aligned} w\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) &= E\left|Y - \varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}\right)\right| \\ &= E\left|Z + \Delta(\boldsymbol{x}, \boldsymbol{\beta})\right| \qquad \text{(cf. (3.17))}. \end{aligned}$$

The difference

$$w\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right) = E\left(|Z + \Delta(\boldsymbol{x}, \boldsymbol{\beta})| - |Z|\right) \tag{3.19}$$

will be estimated by using the generalized Taylor formula

$$|Z + \Delta| - |Z| = \mathcal{D}|Z|\,\Delta + \mathcal{R}(Z, \Delta) \tag{3.20}$$

valid for all real Δ where

$$\mathcal{D}|z| = I(0 \le z < \infty) - I(-\infty < z < 0) \tag{3.21}$$

is the right-hand derivative of the function $|z|$ for $z \in \mathbb{R}$ and $\mathcal{R}(z, \Delta) = (z + \Delta) \cdot I(-\Delta < z < 0)$ is a remainder in the formula (3.20). This follows from the generalized Taylor expansion arbitrary convex functions established in (2.7) of Liese and Vajda (2003). Since $med(Z) = 0$, it holds $E\mathcal{D}|Z| = 0$. Therefore we get from (3.19), (3.20)

$$\begin{aligned} w\left(\boldsymbol{x}^T \boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right) &= E\,\mathcal{R}(Z, \Delta) \\ &= \int (z + \Delta)\, I(-\Delta < z < 0)\, g_{\boldsymbol{x}}(z)\, dz \\ &= \int_0^\Delta (\Delta - z)\, g_{\boldsymbol{x}}(-z)\, dz \\ &= \int_0^\Delta (\Delta - z)\, f_{\pi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right) - z\right) dz. \end{aligned}$$

Since $\mathcal{X} \subset \mathbb{R}^d$ is bounded, the values

$$\pi_1 = \inf_{\boldsymbol{x}\in\mathcal{X}} \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) \qquad \text{and} \qquad \pi_2 = \sup_{\boldsymbol{x}\in\mathcal{X}} \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)$$

are bounded away from 0 and 1. Thus, taking into account that $\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) = m\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)$, we see from (3.16) that we can find $a > 0$ such that

$$\inf_{|z|\le a} \inf_{\boldsymbol{x}\in\mathcal{X}} f_{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) - z\right) \ge \Lambda(a) > 0.$$

This implies that if $0 < b < a$ then for every $|\Delta(\boldsymbol{x},\boldsymbol{\beta})| > b$ it holds

$$w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) \ge \frac{b^2}{2}\Lambda(a).$$

Hence, by (3.18), for every $0 < b < a$ we get

$$\boldsymbol{m}(\boldsymbol{\beta}) - \boldsymbol{m}(\boldsymbol{\beta}_0) \ge \frac{b^2}{2}\Lambda(a)\, Q\left(\mathcal{X}_{b,\boldsymbol{\beta}}\right) \tag{3.22}$$

for the subset of regressors

$$\mathcal{X}_{b,\boldsymbol{\beta}} = \{\, \boldsymbol{x} \in \mathcal{X} : |\Delta(\, \boldsymbol{x},\boldsymbol{\beta})| \ge b\}.$$

By **(c2),** the smallest eigenvalue $\lambda(\boldsymbol{\Sigma})$ of the matrix (3.4) is positive. Further, for every $\tau > 0$

$$\begin{aligned}
\lambda(\boldsymbol{\Sigma})\left\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\right\|^2 \le\; & (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \boldsymbol{\Sigma}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\
=\; & \int_{\mathcal{X}} \left(\boldsymbol{x}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right)^2 dQ(\boldsymbol{x}) \\
\le\; & \|\mathcal{X}\| \,.\, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right) + \tau^2
\end{aligned}$$

where $\|\mathcal{X}\|$ stands for $\max\|\boldsymbol{x}\|$ on $\mathcal{X}$ and $\mathcal{X}^0_{\tau,\boldsymbol{\beta}} = \left\{\, \boldsymbol{x} \in \mathcal{X} : \left|\boldsymbol{x}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right| > \tau\right\}$. From here we see that for all $\varepsilon > 0$ and all sufficiently small $\tau > 0$

$$\psi(\tau,\varepsilon) \equiv \inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \ge \varepsilon} Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right) > 0. \tag{3.23}$$

Since we proved earlier that $\varphi(t)$ of (3.10) is strictly increasing on $\mathbb{R}$, the function

$$\phi(\tau) \equiv \inf_{\substack{|t| \le \|\mathcal{X}\| . \|\boldsymbol{\beta}_0\| \\ |s-t| \ge \tau}} |\varphi(s) - \varphi(t)|$$

is positive in the domain $\tau > 0$ and, obviously,

$$\mathcal{X}_{\phi(\tau),\boldsymbol{\beta}} \supseteq \mathcal{X}^0_{\tau,\boldsymbol{\beta}}.$$

Further, $\varphi(t)$ of (3.10) was proved to be continuous so that $\phi(\tau) < a$ for all sufficiently small $\tau > 0$. Consequently (3.22) implies for any $\varepsilon > 0$

$$\begin{aligned}\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon}[\boldsymbol{m}(\boldsymbol{\beta})-\boldsymbol{m}(\boldsymbol{\beta}_0)] &\geq \tfrac{\phi(\tau)^2}{2}\Lambda(a)\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon} Q\left(\mathcal{X}_{\phi(\tau),\boldsymbol{\beta}}\right)\\ &\geq \tfrac{\phi(\tau)^2}{2}\Lambda(a)\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon} Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right)\\ &= \tfrac{\phi(\tau)^2}{2}\Lambda(a)\,\psi(\tau,\varepsilon).\end{aligned}$$

By (3.23), the last product is positive which proves the desired relation (3.15).
■

Now we formulate conditions which are in a combination with **(c1)-(c6)** sufficient for the asymptotic normality of the median estimators $\widehat{\boldsymbol{\beta}}_n$.

(c7) The quantile function $m(\pi)$ is differentiable on $(0,1)$ and the derivative $m'(\pi)$ is locally Lipschitz in the sense that for every $\pi_0 \in (0,1)$ there exists a constant $L(\pi_0)$ such that
$$|m'(\pi) - m'(\pi_0)| \leq L(\pi_0)\,|\pi - \pi_0|. \tag{3.24}$$

(c8) The densities f_π assumed in (**c3**) satisfy for every $0 < \pi_1 < \pi_2 < 1$ the condition
$$\lim_{y\to 0}\sup_{\pi_1\leq\pi\leq\pi_2} |f_\pi(m(\pi)+y) - f_\pi(m(\pi))| = 0. \tag{3.25}$$

Under (**c6**)**,** the function φ of (3.10) is continuously differentiable with the derivative $\varphi'(t) = m'(\pi(t))\pi'(t)$ where $\pi'(t) = \pi(t)(1-\pi(t))$. Let us introduce similar notation as in the proof of Theorem 3.1, namely let for $i = 1, 2, ...$

$$\begin{aligned}\Delta_i(\boldsymbol{\beta}) &= \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) - \varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right), \qquad \boldsymbol{\beta} \in \mathbb{R}^d,\\ Z_i &= Y_i - \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right),\end{aligned}$$

where
$$\widetilde{f}_i(z) = f_{\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)}\left(z + \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right), \qquad z \in \mathbb{R},$$
is the probability density function of Z_i. The functions $\Delta_i(\boldsymbol{\beta})$ are continuously differentiable on $\mathbb{R}^d$ with gradients
$$grad\,(\Delta_i(\boldsymbol{\beta})) = -\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\boldsymbol{x}_i.$$

Therefore the linear term $\mathcal{L}_n(h)$ considered in (2.3) of Liese and Vajda (2004) is given here by
$$\mathcal{L}_n(\boldsymbol{h}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{D}\,|Z_i|\,\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\boldsymbol{x}_i^T\boldsymbol{h}, \qquad \boldsymbol{h} \in \mathbb{R}^d,$$

where $\mathcal{D}|z|$ denotes the right-hand derivative (3.21). Since $E\,\mathcal{D}|Z_i| = 0$, the variance of $\mathcal{L}_n(\boldsymbol{h})$ is $\boldsymbol{h}^T\Sigma_n\boldsymbol{h}$ for the matrix given in accordance with (2.5) of Liese and Vajda (2004) by

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

But $E\left(\mathcal{D}|Z_i|\right)^2 = 1$, so that we can write this matrix in the form

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

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

$$\lim_{n\to\infty}\Sigma_n = \Sigma \equiv \int_{\mathcal{X}} \left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)^2 \boldsymbol{x}^T\boldsymbol{x}\,dQ(\boldsymbol{x}) \tag{3.26}$$

where Q is the limit measure from **(c1)**.

The next step is evaluation of the matrices

$$\mathcal{Q}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(0)\,\nabla\varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\left(\nabla\varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)^T$$

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 on p. 467 in Liese and Vajda (2003). By the definition of $\mathcal{D}|z|$ in (3.21), for $\pi_i = \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$ and $\varphi_i = \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$

$$\begin{aligned} G_i(t) = {} & EI(Z_i + t > 0) - EI(Z_i + t \le 0) \\ = {} & EI(Y_i > \varphi_i - t) - EI(Y_i \le \varphi_i - t) \\ = {} & 1 - 2F_{\pi_i}(\varphi_i - t). \end{aligned}$$

Thus $g_i(t) = 2f_{\pi_i}(\varphi_i - t)$ and

$$g_i(0) = 2f_{\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right).$$

Therefore the matrices $\mathcal{Q}_n$ may be given by

$$\mathcal{Q}_n = 2\int_{\mathcal{X}} f_{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)^2 \boldsymbol{x}^T\boldsymbol{x}\,dQ_n(\boldsymbol{x}).$$

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

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

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

$$\lim_{n\to\infty} \mathcal{Q}_n = \mathcal{Q} \;\equiv\; 2\int_{\mathcal{X}} f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)} \left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)^2 \boldsymbol{x}^T\boldsymbol{x}\, dQ\left(\boldsymbol{x}\right). \tag{3.27}$$

Finally, $\mathcal{D}\rho\left(Z_i\right) = \mathcal{D}\left|Z_i\right|$ is in the present situation bounded and $grad\left(\Delta_i\left(\boldsymbol{\beta}_0\right)\right) = -\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\boldsymbol{x}_i$ is bounded uniformly for all possible $\boldsymbol{x}_i \in \mathcal{X}$. Since Z_i and $\Delta_i(\beta)$ were denoted in Liese and Vajda (2004) by X_i and $f_i(\beta)$, this means that the Liapunov condition (2.6) of Liese and Vajda (2004) holds. Similarly, one can verify that the conditions **(C3), (C4)** of Liese and Vajda (2003) as well as (2.39), (2.40) ibid. hold. Thus, by Lemma 3 in Liese and Vajda (2003), **(C5)** and **(C6)** ibid. hold too.

Thus we can conclude that if (**c1**)-(**c8**) hold then all assumptions of Theorem 1 in Liese and Vajda (2004) are satisfied and the following assertion is proved.

Theorem 3.2 Let a logistic regression model considered in Theorem 3.1 satisfy **(c1)-(c8)** and let the condition (3.16) of Theorem 3.1 hold. If the limit matrix $\mathcal{Q}$ in (3.27) is positive definite then the median estimator $\widehat{\boldsymbol{\beta}}_n$ of true parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ is asymptotically normal in the sense that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N\left(\mathbf{0},\ \mathcal{Q}^{-1}\Sigma\mathcal{Q}^{-1}\right) \tag{3.28}$$

for Σ given by (3.26).

Proof. See above.

■

4 Applications to the Bernoulli model

In this section we apply the theory of Section 3 to the general Bernoulli model introduced in Section 2. As argued in Section 2, for this application it is convenient to consider the standard modification of this model with the distribution functions $F_\pi\left(y\right),\ \pi \in (0,1)$ of the form given in Fig. 4.1 and the corresponding densities

$$f_\pi\left(y\right) = \left(1-\pi\right) I\left(0 < y \le 1\right) + \pi I\left(1 < y < 2\right),\ y \in \mathbb{R}. \tag{4.1}$$

It follows from here that (**c3)-(c5)** hold. The median function $m\left(\pi\right)$ is given on $(0,1)$ by (2.36), from where it is seen that it is bounded on $(0,1)$ and continuously differentiable with bounded derivative. Therefore the model under consideration satisfies also (**c6**) and (**c7**).

The conditions (**c8**) and (3.16) will be proved for arbitrary $0 < \pi_1 < \pi_2 < 1$ if we prove them separately for $\pi_1 = 1/2 < \pi_2 < 1$ and $0 < \pi_1 < \pi_2 = 1/2$. We shall prove them for $\pi_1 = 1/2 < \pi_2 < 1$ only since for the alternative the proof is similar. Let $\pi > 1/2$ be arbitrary. It is seen from (2.36) that

$$1 + \frac{1}{2} > m\left(\pi\right) = 1 + \frac{2\pi - 1}{2\pi} > 1$$

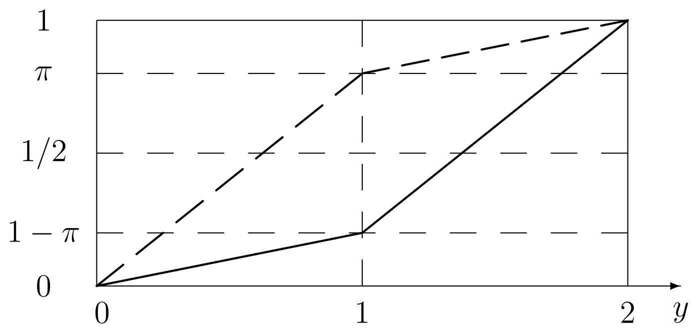

Figure 4.1: $F_{\pi}(y)$ full line, $F_{1-\pi}(y)$ dashed line.

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

$$f_{\pi}(m(\pi)+y) = f_{\pi}(m(\pi)) = \pi$$

unless $m(\pi)+y \leq 1$ in which case $f_{\pi}(m(\pi)+y) = 1-\pi$. Thus

$$\inf_{1/2 \leq \pi \leq \pi_2} f_{\pi}(m(\pi)+y) \geq 1-\pi$$

which implies (3.16). Further, the absolute difference $|f_{\pi}(m(\pi)+y) - f_{\pi}(m(\pi))|$ is either zero or $\pi - (1-\pi) = 2\pi - 1$. This difference will be maximized if we take maximal π satisfying the inequality $m(\pi)+y \leq 1$ for the fixed y. Since $m(\pi)$ is increasing in π, this means that

$$\sup_{1/2 \leq \pi < \pi_2} |f_{\pi}(m(\pi)+y) - f_{\pi}(m(\pi))| = f_{\pi_*}(m(\pi_*)+y) - f_{\pi_*}(m(\pi_*))$$

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

$$\sup_{1/2 \leq \pi < \pi_2} |f_{\pi}(m(\pi)+y) - f_{\pi}(m(\pi))| \leq 2\pi_* - 1 = \frac{|y|}{1-|y|}$$

which implies (**c8).**

It follows from (3.10) that

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

Therefore

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

Further

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

Consequently,

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

Therefore we get from (3.26) that

$$\Sigma = \frac{1}{4}\int_{\mathcal{X}} e^{-2|\boldsymbol{x}^T\boldsymbol{\beta}_0|}\boldsymbol{x}\boldsymbol{x}^T dQ(\boldsymbol{x}) \tag{4.2}$$

and from (3.27) that

$$\mathcal{Q} = \frac{1}{2}\int_{\mathcal{X}} \frac{e^{-|\boldsymbol{x}^T\boldsymbol{\beta}_0|}}{1+e^{|\boldsymbol{x}^T\boldsymbol{\beta}_0|}}\ \boldsymbol{x}\boldsymbol{x}^T dQ(\boldsymbol{x}). \tag{4.3}$$

Since we proved that the model under consideration satisfies **(c3)-(c8)** and (3.16), Theorem 8.2 implies the following assertion.

Corollary 4.1. If the regressors in the above considered logistic model with binary responses satisfy **(c1)** and **(c2)** and the matrix $\mathcal{Q}$ given by (4.3) is positive definite then the median estimator $\widehat{\boldsymbol{\beta}}_n$ of the true parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ is asymptotically normal in the sense

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(\mathbf{0}, \mathcal{Q}^{-1}\Sigma\mathcal{Q}^{-1}\right) \tag{4.4}$$

for Σ given in (4.2).

■

Example 4.1. The univariate Bernoulli logistic regression model with identical regressors $x_1 = x_2 = ... = 1$ is characterized by observations $Y_i \sim F_{\pi(\beta_0)}, 1 \le i \le n$ where

$$\pi(\beta) = \frac{e^{\beta}}{1+e^{\beta}},\ \beta \in \mathbb{R}$$

and F_π is the Bernoulli response distribution function considered in Figure 4.1. This model satisfies (**c1**) and (**c2**) with the Dirac measure $Q = \delta_1$ concentrated at $1 \in \mathcal{X} = \{1\} \subset \mathbb{R}$.Therefore we get from (4.2) and (4.3) that

$$\Sigma = \frac{1}{4}e^{-2|\beta_0|}, \qquad \mathcal{Q} = \frac{e^{-|\beta_0|}}{2\left(1+e^{|\beta_0|}\right)}.$$

But

$$\begin{aligned}\pi(t) \wedge (1-\pi(t)) &= \frac{1}{1+e^{-t}} \wedge \frac{1}{1+e^{t}} \\ &= \frac{1}{1+e^{|t|}}\end{aligned}$$

so that

$$\mathcal{Q}^{-1}\Sigma\mathcal{Q}^{-1} = \left(1+e^{|\beta_0|}\right)^2 = \frac{1}{\pi\left(\beta_0\right)^2 \wedge \left(1-\pi\left(\beta_0\right)\right)^2}.$$

This result coincides with (6.22) and (6.23) obtained in Section 6 below by a direct calculation.

5 Contaminated Bernoulli models

The median estimators $\widehat{\boldsymbol{\beta}}_n$ of the true logistic regression parameters $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ introduced in Section 2 and studied in Section 3 were defined by means of the least absolute deviations principle. The original justification of this principle in the papers cited after formula (2.2) was the robustness of the resulting estimators, namely their resistance to gross errors in the observations $Y_1, ..., Y_n$. This means that one can expect more robustness from the median estimator proposed by us for the general logistic regression than from the classical MLE's or from the estimators obtained by the least square method (L_2-estimators) and applied to the logistic regression. Similar robust alternatives to the classical MLE's seem so far been considered only for the logistic regression with Bernoulli responses. We shall describe the two most recent of those known to us.

Morgenthaler (1992) introduced an L_1-estimator $\boldsymbol{\beta}_n^{(1)}$ for the logistic regression with linear responses Y_i given by (2.1) for the Bernoulli distribution function $F_\pi(y)$ given in (2.15). He started with the weighted L_1-estimator

$$\boldsymbol{\beta}_n^{(0)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \frac{\left|\Delta_i(\boldsymbol{\beta})\right|}{\sigma_i(\boldsymbol{\beta})},$$

where

$$\sigma_i^2(\boldsymbol{\beta}) = \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\left(1-\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right), \quad \Delta_i(\boldsymbol{\beta}) = Y_i - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right),$$

more precisely with the solutions $\boldsymbol{\beta}_n^{(0)}$ of the system of equations $U_n^{(0)}(\boldsymbol{\beta}) = 0$, for

$$U_n^{(0)}(\boldsymbol{\beta}) = D^T\left\{diag\left(\sigma_1(\boldsymbol{\beta}), ..., \sigma_n(\boldsymbol{\beta})\right)\right\}^{-1/2}\left(sgn\,\Delta_1(\boldsymbol{\beta}), ..., sgn\,\Delta_n(\boldsymbol{\beta})\right)^T$$

where $D = \left(D_{ij} = \partial\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)/\partial\beta_j\right)_{i,j=1}^T$ and *sgn* denotes the sign. Since the resulting estimator $\boldsymbol{\beta}_n^{(0)}$ was inconsistent, he proposed a slight modification $\boldsymbol{\beta}_n^{(1)}$ which solves the equation $U_n^{(1)}(\boldsymbol{\beta}) = 0$ for the centered version $U_n^{(1)}(\boldsymbol{\beta}) = U_n^{(0)}(\boldsymbol{\beta}) - E_{\boldsymbol{\beta}}U_n^{(0)}(\boldsymbol{\beta})$. We shall find an explicit formula for $U_n^{(1)}$. It is easy to see that

$$E_{\boldsymbol{\beta}}U_n^{(0)}(\boldsymbol{\beta}) = D^T\left\{diag\left(\sigma_1(\boldsymbol{\beta}), ..., \sigma_n(\boldsymbol{\beta})\right)\right\}^{-1/2}\left(2\mu_1(\boldsymbol{\beta}) - 1, ..., 2\mu_n(\boldsymbol{\beta}) - 1\right)^T$$

for the expectation functions

$$\mu_i(\boldsymbol{\beta}) = E_{\boldsymbol{\beta}}Y_i = \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right) \tag{5.5}$$

and, moreover,

$$sgn\,\Delta_i(\boldsymbol{\beta}) - (2\mu_i(\boldsymbol{\beta}) - 1) = 2\Delta_i(\boldsymbol{\beta}).$$

Therefore

$$U_n^{(1)}(\boldsymbol{\beta}) = D^T\left\{diag\left(\sigma_1(\boldsymbol{\beta}), ..., \sigma_n(\boldsymbol{\beta})\right)\right\}^{-1/2}\left(\Delta_1(\boldsymbol{\beta}), ..., \Delta_n(\boldsymbol{\beta})\right)^T.$$

Since

$$D = \left(diag\left(\sigma_1^2(\boldsymbol{\beta}), ..., \sigma_1^2(\boldsymbol{\beta})\right)\right)\left(\boldsymbol{x}_1, ..., \boldsymbol{x}_n\right)^T,$$

we finally obtain the d-variate functions $U_n^{(1)}(\boldsymbol{\beta})$ in the form

$$\begin{aligned} U_n^{(1)}(\boldsymbol{\beta}) = &\ \left(\boldsymbol{x}_1, ..., \boldsymbol{x}_n\right)\left\{diag\left(\sigma_1(\boldsymbol{\beta}), ..., \sigma_n(\boldsymbol{\beta})\right)\right\}\left(\Delta_1(\boldsymbol{\beta}), ..., \Delta_n(\boldsymbol{\beta})\right) \\ = &\ \sum_{i=1}^n \sigma_i(\boldsymbol{\beta})\Delta_i(\boldsymbol{\beta})\boldsymbol{x}_i \\ = &\ \sum_{i=1}^n \sqrt{\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\left(1 - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)}\left(Y_i - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)\boldsymbol{x}_i. \end{aligned} \tag{5.6}$$

An alternative robust estimator $\boldsymbol{\beta}_n^{(2)}$ for the logistic regression was proposed by Bianco and Yohai (1996) who also assumed the Bernoulli responses Y_i given in (2.1) for the $F_\pi(y)$ of (2.15). They started with the MLE

$$\boldsymbol{\beta}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n D_i(\boldsymbol{\beta}) \tag{5.7}$$

where

$$D_i(\boldsymbol{\beta}) = -Y_i \ln \mu_i(\boldsymbol{\beta}) - (1 - Y_i)\ln\left(1 - \mu_i(\boldsymbol{\beta})\right) \tag{5.8}$$

and $\mu_i(\boldsymbol{\beta}) = \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)$ are the expectation functions (5.5). If the data $(\boldsymbol{x}_1, Y_1), ..., (\boldsymbol{x}_n, Y_n)$ are from the assumed model, i.e. if $Y_i \sim Be\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)$ then the expected value of the sum minimized in (5.7) is

$$E\left(\sum_{i=1}^n D_i(\boldsymbol{\beta})\right) = \sum_{i=1}^n \left[-\mu_i(\boldsymbol{\beta}_0)\ln\mu_i(\boldsymbol{\beta}) - \left(1 - \mu_i(\boldsymbol{\beta}_0)\right)\ln\left(1 - \mu_i(\boldsymbol{\beta})\right)\right]$$

which is minimized by $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ since for every $1 \le i \le n$

$$\mu_i(\boldsymbol{\beta}_0) \ln \frac{\mu_i(\boldsymbol{\beta}_0)}{\mu_i(\boldsymbol{\beta})} + (1 - \mu_i(\boldsymbol{\beta}_0)) \ln \frac{1 - \mu_i(\boldsymbol{\beta}_0)}{1 - \mu_i(\boldsymbol{\beta})} \ge 0.$$

It follows from here that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ is the only minimum of $E\left(\sum_{i=1}^n D_i(\boldsymbol{\beta})\right)$ unless all $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are from a hyperplane in $\mathbb{R}^d$, i.e., unless $\boldsymbol{x}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = 0$ for some $\boldsymbol{\beta} \ne \boldsymbol{\beta}_0$ and all $1 \le i \le n$. Under mild regularity this property of $E\left(\sum_{i=1}^n D_i(\boldsymbol{\beta})\right)$ already guarantees nice asymptotic properties of $\boldsymbol{\beta}_n$ like consistency and asymptotic normality with the variances at the Cramér-Rao lower bound. However, this estimator is too sensitive to the gross errors (outliers) among the data $(\boldsymbol{x}_1, Y_1), ..., (\boldsymbol{x}_n, Y_n)$ which are the pairs $(\boldsymbol{x}_i, Y_i)$ where Y_i are not generated by the Bernoulli model $Be\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)$. Typical outliers are $Y_i = 0$ and $\boldsymbol{x}_i$ leading to $\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) \approx 1$ or $Y_i = 1$ and $\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) \approx 0$. A simple source of outliers taking place with a probability $0 < \varepsilon < 1/2$ is the transmission of the true observations $Y_i \sim Be\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)$ through a binary symmetric channel presented in Fig. 5.1.

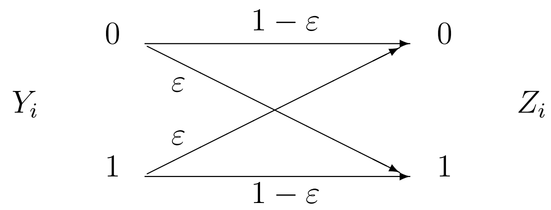

Figure 5.1: Binary Symmetric channel BSC(ε) with independent inputs Y_i, additive (mod 2) independent noise $W_i \sim Be(\varepsilon)$ and independent outputs $Z_i = Y_i + W_i$ (mod 2).

Then the actual data $(\boldsymbol{x}_1, Z_1), ..., (\boldsymbol{x}_n, Z_n)$ contain responses Z_i generated by the stochastic mixture

$$(1 - \varepsilon) Be\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right) + \varepsilon Be\left(1 - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)$$

of the Bernoulli models with parameters $\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$ and $1 - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$. The outliers $(\boldsymbol{x}_i, Y_i)$ with $Y_i \sim Be\left(1 - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)$ and $\mu_i(\boldsymbol{\beta}_0) = \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$ very close to 0 or 1 lead with a high probability to very large values of $D_i(\boldsymbol{\beta}_0)$, thus pushing the MLE estimate (5.7) away from the true value $\boldsymbol{\beta}_0$. The resulting effect is a sharp loss of consistency.

To restrict the influence of the outliers $(\boldsymbol{x}_i, 0)$ with probabilities $\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) = 1 - \delta_i$ close to 1 and the outliers $(\boldsymbol{x}_i, 1)$ with probabilities $\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) = \delta_i$ close to 0, both of them leading to large $D_i(\boldsymbol{\beta}_0) = -\ln \delta_i$, Bianco and Yohai (1996) replaced the MLE $\boldsymbol{\beta}_n$ by the M-estimator

$$\boldsymbol{\beta}_n^{(0)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho(D_i(\boldsymbol{\beta})) \tag{5.9}$$

where

$$\rho(y) = \left(y - \frac{y^2}{2c}\right) I(0 \le y \le c) + \frac{c}{2} I(y > c) \tag{5.10}$$

is a hard-limiter defined on the real line and specified by a limiting constant $c > 0$. However, the hard-limiting violates the consistency since $E\left[\sum_{i=1}^{n} \rho(D_i(\boldsymbol{\beta}))\right]$ is no more minimized at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Similarly as in the case of $\boldsymbol{\beta}_n^{(0)}$ of Morgenthaler (1992), Bianco and Yohai introduced a bias-correcting term into (5.9). They proved that if $c > \ln 2 = 0.7$ and

$$G(\pi) = \int_0^{\pi} \rho'(-\ln t)\, dt \quad \text{for } \pi \in (0,1) \tag{5.11}$$

then $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ minimizes the expectation

$$\sum_{i=1}^{n} \left[E\rho(D_i(\boldsymbol{\beta})) + G(\mu_i(\boldsymbol{\beta})) + G(1 - \mu_i(\boldsymbol{\beta}))\right]$$

and that this minimum is unique unless all $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are from a hyperplane in $\mathbb{R}^d$. Therefore the consistent robust estimator of Bianco and Yohai for the logistic regression with binary responses is

$$\boldsymbol{\beta}_n^{(2)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left[\rho(D_i(\boldsymbol{\beta})) + G(\mu_i(\boldsymbol{\beta})) + G(1 - \mu_i(\boldsymbol{\beta}))\right] \tag{5.12}$$

for ρ defined by (5.10) with $c > \ln 2$, $D_i(\boldsymbol{\beta})$ defined by (5.8), $G(y)$ defined by (5.11) and for regressors $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ satisfying some regularity assumptions. (E.g., they are assumed to be independent realizations of a $d-$variate random vector $\boldsymbol{X}$ such that $\Pr\left(\boldsymbol{X}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \neq 0\right) = 1$ for every $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$).

The robust alternatives $\boldsymbol{\beta}_n^{(1)}$ and $\boldsymbol{\beta}_n^{(2)}$ to our median estimator $\widehat{\boldsymbol{\beta}}_n$ will be compared with $\widehat{\boldsymbol{\beta}}_n$ on simulated contaminated data from the logistic regression models to which these alternatives were designed, i.e. from the Bernoulli models. These simulations are presented in Section 9 and Section 11. The robustness of $\widehat{\boldsymbol{\beta}}_n$ will also be demonstrated on simulated contaminated data from some logistic regression models with non-binary responses, namely from the models with geometric responses. These results will be given in Section 10.

6 Identical regressors in univariate Bernoulli models

In this section we study the simple special case, namely the Bernoulli models, of dimension $d = 1$ with all univariate regressors identical, $x_1 = x_2 = ... = x \in \mathbb{R}$. For simplicity we put $x = 1$. Thus we estimate a parameter $\beta_0 \in \mathbb{R}$ using the independent logistic regression observations

$$Y_i \sim F_{\pi(\beta_0)}(y), \ 1 \le i \le n, \tag{6.1}$$

where

$$\pi(\beta) = \frac{e^{\beta}}{1+e^{\beta}},\ \beta \in \mathbb{R} \tag{6.2}$$

and F_π is the Bernoulli response distribution function of (2.15).

The Fisher information $I(\pi)$ in the Bernoulli model $Be(\pi)$ is

$$I(\pi) = \frac{1}{\pi(1-\pi)} \tag{6.3}$$

and the MLE of a true parameter, $\pi_0 \in (0,1)$ is the relative frequency

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{6.4}$$

satisfying the relation

$$\sqrt{n}\left(\overline{Y}_n - \pi_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(0, \pi_0\left(1-\pi_0\right)\right). \tag{6.5}$$

Fisher information in the model Bernoulli logistic regression model (6.1) is

$$\begin{aligned} \mathcal{J}(\beta_0) &= \frac{\left[\pi'(\beta_0)\right]^2}{\pi(\beta_0)} + \frac{\left[(1-\pi(\beta_0))'\right]^2}{1-\pi(\beta_0)} \\ &= \frac{\left[\pi'(\beta_0)\right]^2}{\pi(\beta_0)(1-\pi(\beta_0))} = \frac{\left[\pi(\beta_0)(1-\pi(\beta_0))\right]^2}{\pi(\beta_0)(1-\pi(\beta_0))} \\ &= \pi(\beta_0)(1-\pi(\beta_0)) = \frac{e^{\beta_0}}{(1+e^{\beta_0})^2}. \end{aligned} \tag{6.6}$$

The MLE β_n in this model is

$$\beta_n = \phi\left(\overline{Y}_n\right) \quad \text{where} \quad \phi(y) = \pi^{-1}(y) = \ln\frac{y}{1-y}.$$

For $\pi_0 = \pi(\beta_0)$ we have

$$\phi\left(\overline{Y}_n\right) - \phi(\pi_0) = \phi'(\pi_0)\left(\overline{Y}_n - \pi_0\right) + o\left(\overline{Y}_n - \pi_0\right). \tag{6.7}$$

Since

$$\phi\left(\overline{Y}_n\right) = \beta_n,\ \phi(\pi_0) = \pi^{-1}(\pi(\beta_0)) = \beta_0,\ \phi'(y) = \frac{1}{y(1-y)},$$

we get from (6.7)

$$\sqrt{n}\left(\beta_n - \beta_0\right) = \frac{1}{\pi_0(1-\pi_0)}\sqrt{n}\left(\overline{Y}_n - \pi_0\right) + o\left(\sqrt{n}\left(\overline{Y}_n - \pi_0\right)\right).$$

Hence by (6.5),

$$\sqrt{n}\left(\beta_n-\beta_0\right)=\frac{1}{\pi_0\left(1-\pi_0\right)}\sqrt{n}\left(\overline{Y}_n-\pi_0\right)+o_p(1)$$

$$\xrightarrow[n\to\infty]{\mathcal{D}} N\left(0,\frac{\pi_0\left(1-\pi_0\right)}{\left[\pi_0\left(1-\pi_0\right)\right]^2}\right) \tag{6.8}$$

$$= N\left(0,\frac{1}{\pi_0\left(1-\pi_0\right)}\right)=N\left(0,\frac{\left(1+e^{\beta_0}\right)^2}{e^{\beta_0}}\right).$$

In this derivation we used only the asymptotic normality (6.5). Hence the estimator

$$\beta_n=\ln\frac{\overline{Y}_n}{1-\overline{Y}_n} \tag{6.9}$$

consistently estimates β_0 in every model $\widetilde{\pi}(\beta)$ such that $\widetilde{\pi}(\beta_0)=\pi(\beta_0)$ and, using the Taylor expansion (6.7), we find that

$$\sqrt{n}\left(\beta_n-\beta_0\right)\xrightarrow[n\to\infty]{\mathcal{D}} N\left(0,\frac{1}{\pi\left(\beta_0\right)\left(1-\pi\left(\beta_0\right)\right)}\right)\quad\text{(cf. (6.8).}$$

But β_n need not be the MLE in the model $\widetilde{\pi}(\beta)$. Such estimator is

$$\widetilde{\beta}_n=\widetilde{\pi}^{-1}\left(\overline{Y}_n\right) \tag{6.10}$$

for which

$$\sqrt{n}\left(\widetilde{\beta}_n-\beta_0\right)\xrightarrow[n\to\infty]{\mathcal{D}} N\left(0,\frac{\widetilde{\pi}\left(\beta_0\right)\left(1-\widetilde{\pi}\left(\beta_0\right)\right)}{\left[\widetilde{\pi}'\left(\beta_0\right)\right]^2}\right) \tag{6.11}$$

because the Fisher information in this model is

$$\widetilde{\mathcal{J}}\left(\beta_0\right)=\frac{\left[\widetilde{\pi}'\left(\beta_0\right)\right]^2}{\widetilde{\pi}\left(\beta_0\right)\left(1-\widetilde{\pi}\left(\beta_0\right)\right)}\qquad\text{(cf. (6.6).} \tag{6.12}$$

For example, the model

$$\widetilde{\pi}(\beta)=\frac{e^{\beta}}{3}\wedge 1,\qquad \beta\in\mathbb{R} \tag{6.13}$$

for the particular value $\beta_0=\ln 2$ satisfies the relation

$$\widetilde{\pi}\left(\beta_0\right)=\frac{e^{\ln 2}}{3}=\frac{2}{3}=\frac{e^{\ln 2}}{1+e^{\ln 2}}=\pi\left(\beta_0\right)$$

where the model $\pi(\beta)$ is given by (6.2). The Fisher information in this model is in the domain $\beta<\ln 3$ given by

$$\widetilde{\mathcal{J}}(\beta)=\frac{\left[\widetilde{\pi}'(\beta)\right]^2}{\widetilde{\pi}(\beta)\left(1-\widetilde{\pi}(\beta)\right)}=\frac{\left[\widetilde{\pi}(\beta)\right]^2}{\widetilde{\pi}(\beta)\left(1-\widetilde{\pi}(\beta)\right)}=\frac{\widetilde{\pi}(\beta)}{1-\widetilde{\pi}(\beta)}.$$

By (6.10) and (6.11), the MLE

$$\widetilde{\beta}_n = \widetilde{\pi}^{-1}\left(\overline{Y}_n\right) = \ln(3\overline{Y}_n)$$

satisfies the relation

$$\sqrt{n}\left(\widetilde{\beta}_n - \beta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(0, \frac{1-\widetilde{\pi}\left(\beta_0\right)}{\widetilde{\pi}\left(\beta_0\right)}\right) \tag{6.14}$$

and

$$\frac{1}{\widetilde{\mathcal{J}}\left(\beta_0\right)} = \frac{1-\widetilde{\pi}\left(\beta_0\right)}{\widetilde{\pi}\left(\beta_0\right)} < \frac{1}{\widetilde{\pi}\left(\beta_0\right)\left(1-\widetilde{\pi}\left(\beta_0\right)\right)} = \frac{1}{\mathcal{J}\left(\beta_0\right)}. \tag{6.15}$$

We see from here if $\pi\left(\beta\right)$ is not the correct model then the MLE β_n may lead to the asymptotic variance $\sigma^2 = 1/\mathcal{J}\left(\beta_0\right)$ much larger than that given by the achievable Cramér-Rao bound $\widetilde{\sigma}^2 = 1/\widetilde{\mathcal{J}}\left(\beta_0\right)$.

Let us now look what is obtained from the median estimator $\widehat{\beta}_n$ defined by (2.22) using the continuously modified data

$$Y_i \sim f_{\pi_0}\left(y\right) = \left(1-\pi_0\right) I\left(0 < y \le 1\right) + \pi_0 I\left(1 < y \le 2\right) \tag{6.16}$$

(cf. (2.33)) where $\pi_0 = \pi\left(\beta_0\right)$. This estimator is defined by

$$\widehat{\beta}_n = \arg\min_{\beta} \sum_{i=1}^{n} \left|Y_i - m\left(\pi\left(\beta\right)\right)\right|$$

for $m\left(\pi\right)$ given by (2.36) and $\pi\left(\beta\right)$ given by (6.2). Let $Y_{(1)} \le ..... \le Y_{(n)}$ be the ordered sample $Y_1, ..., Y_n$ and denote by $Y_{(n/2)}$ the sample median. As well known,

$$Y_{(n/2)} = \arg\min_{m} \sum_{i=1}^{n} \left|Y_i - m\right|. \tag{6.17}$$

Therefore

$$\widehat{\beta}_n = \phi\left(m^{-1}\left(Y_{(n/2)}\right)\right) = \ln \frac{m^{-1}\left(Y_{(n/2)}\right)}{1-m^{-1}\left(Y_{(n/2)}\right)},$$

where $\phi\left(y\right) = \ln\left(y/\left(1-y\right)\right)$ is inverse to $\pi\left(\beta\right)$. In other words,

$$\widehat{\beta}_n = \phi\left(\widehat{\pi}_n\right) \tag{6.18}$$

for

$$\widehat{\pi}_n = \arg\min_{\pi} \sum_{i=1}^{n} \left|Y_i - m\left(\pi\right)\right| = m^{-1}\left(Y_{(n/2)}\right). \tag{6.19}$$

By p. 490 in Rényi (1970), the random observations (6.16) satisfy the limit law

$$\sqrt{n}\frac{Y_{(n/2)} - m\left(\pi_0\right)}{1/\left[2 f_{\pi_0}\left(m\left(\pi_0\right)\right)\right]} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N(0,1). \tag{6.20}$$

Suppose that $\pi_0 \geq 1/2$ in which case $f_{\pi_0}(m(\pi_0)) = \pi_0$. Then

$$\sqrt{n}\left(Y_{(n/2)} - m(\pi_0)\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(0, \frac{1}{4\pi_0^2}\right). \tag{6.21}$$

For $\phi(y) = \ln(y/(1-y))$ and $\psi(y) = \phi(m^{-1}(y))$ we obtain

$$\begin{aligned}
\psi'(y) = {} & \phi'(m^{-1}(y))(m^{-1}(y))' \\
= {} & \frac{1}{m^{-1}(y)(1-m^{-1}(y))}(m^{-1}(y))' \\
= {} & \frac{1}{m^{-1}(y)(1-m^{-1}(y))}\begin{cases} \dfrac{1}{2(2-y)^2} \equiv 2(m^{-1}(y))^2 & \text{if } 1 \leq y \leq 3/2 \\ \dfrac{1}{2y^2} \equiv 2(1-m^{-1}(y))^2 & \text{if } 1/2 \leq y \leq 1 \end{cases} \\
= {} & \begin{cases} \dfrac{2(m^{-1}(y))}{1-m^{-1}(y)} & \text{if } 1 \leq y \leq 3/2 \\ \dfrac{2(1-m^{-1}(y))}{m^{-1}(y)} & \text{if } 1/2 \leq y \leq 1\ . \end{cases}
\end{aligned}$$

Therefore

$$\psi'(m(\pi_0)) = \begin{cases} \dfrac{2\pi_0}{1-\pi_0} & \text{if } 1/2 \leq \pi_0 \leq 1 \\ \dfrac{2(1-\pi_0)}{\pi_0} & \text{if } 0 < \pi_0 \leq 1/2\ . \end{cases}$$

Using similar Taylor expansion as (6.7), we obtain for $\pi_0 \geq 1/2$ from (6.21)

$$\sqrt{n}\left(\psi\left(Y_{(n/2)}\right) - \psi(m(\pi_0))\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(0, \left(\psi'(m(\pi_0))\right)^2 \frac{1}{4\pi_0^2}\right).$$

Hence for $\pi_0 \in [1/2, 1)$

$$\sqrt{n}\left(\phi\left(m^{-1}\left(Y_{(n/2)}\right)\right) - \phi\left(m^{-1}(m(\pi_0))\right)\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(0, \left(\frac{2\pi_0}{1-\pi_0}\right)^2 \frac{1}{4\pi_0^2}\right).$$

By (6.19), $Y_{(n/2)} = m\left(\pi\left(\widehat{\beta}_n\right)\right)$ where $\pi\left(\widehat{\beta}_n\right) = \phi^{-1}\left(\widehat{\beta}_n\right)$. Further, $\phi(\pi_0) = \phi(\pi(\beta_0))$ $= \pi^{-1}(\pi(\beta_0)) = \beta_0$. Therefore we obtain from the last limit relation

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(0, \frac{1}{(1-\pi(\beta_0))^2}\right). \tag{6.22}$$

Similarly for $\pi_0 \in (0, 1/2]$ we obtain

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N\left(0, \frac{1}{\left(\pi\left(\beta_0\right)\right)^2}\right), \tag{6.23}$$

i. e. the asymptotic variance of $\widehat{\beta}_n$ is the maximum of $1/\left(\pi\left(\beta_0\right)\right)^2$ and $1/\left(1-\pi\left(\beta_0\right)\right)^2$. By (6.15), this maximum satisfies the inequality

$$\frac{1}{\left(\pi\left(\beta_0\right)\right)^2} \vee \frac{1}{\left(1-\pi\left(\beta_0\right)\right)^2} \geq \frac{1}{\mathcal{J}\left(\beta_0\right)} = \frac{1}{\pi\left(\beta_0\right)\left(1-\pi\left(\beta_0\right)\right)},$$

i.e. the asymptotic variance of the median estimator $\widehat{\beta}_n$ in the present simple model $\pi\left(\beta\right) = e^{\beta}/\left(1+e^{\beta}\right)$ exceeds that of the MLE β_n given by (6.9).

However, it follows from the derivation of (6.22) that the asymptotic variance$1/(\pi(\beta_0))^2$ $\vee[1/\left(1-\pi\left(\beta_0\right)\right)^2]$ exceeds the asymptotic variance $1/\mathcal{J}\left(\beta_0\right) = 1/\left[\pi\left(\beta_0\right)\left(1-\pi\left(\beta_0\right)\right)\right]$ of β_n in every model $\widetilde{\pi}\left(\beta\right)$ with true β_0 satisfying the equality $\widetilde{\pi}\left(\beta_0\right) = \pi\left(\beta_0\right)$. For example, for the model $\pi(\beta_0)$ given by (6.13) and $\beta_0 = \ln 2$ we obtain the asymptotic standard deviation

$$\frac{1}{1-\pi\left(\beta_0\right)} = \frac{1}{1-2/3} = 3$$

for $\widehat{\beta}_n$ and

$$\frac{1}{\sqrt{\pi\left(\beta_0\right)\left(1-\pi\left(\beta_0\right)\right)}} = \frac{3}{\sqrt{2}} = 2.12$$

for β_n. At the same time, the smallest standard deviation given by the Cramér-Rao bound and achieved by the true MLE $\widetilde{\beta}_n = \ln(3\overline{Y}_n)$ is

$$\sqrt{\frac{1-\pi\left(\beta_0\right)}{\pi\left(\beta_0\right)}} = \frac{1}{\sqrt{2}} = 0.71.$$

We see that for wrongly specified models the quality of the median estimator $\widehat{\beta}_n$ may be comparable to that of the false MLE β_n but it can never be better. To demonstrate advantages of $\widehat{\beta}_n$ over the false MLE β_n we need less trivial univariate Bernoulli models where the regressors $x_1, ..., x_n$ are still simple in the sense that they are univariate but they are not more identical (see Section 8 below).

7 Identical regressors in univariate geometric models

Here we study similar special case $x_1 = x_2 = ... = 1 \in \mathbb{R}$ as in the previous section, with the model given by (6.1) and (6.2), but with the Bernouli response function $F_\pi\left(y\right)$

of (2.15) replaced by the geometric response function of (2.17). The Fisher information in the geometric model is

$$I(\pi) = \frac{1}{\pi(1-\pi)^2}. \tag{7.1}$$

Moreover

$$EY_i = \frac{\pi_0}{1-\pi_0} \quad \text{and} \quad E(Y_i - EY_i)^2 = \frac{\pi_0}{(1-\pi_0)^2}$$

for

$$\pi_0 = \pi(\beta_0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}. \tag{7.2}$$

Further, the sample mean $\overline{Y}_n$ given in (6.4) is the MLE of the function $\pi_0/(1-\pi_0)$ of the true parameter $\beta_0 \in \mathbb{R}$ with the property

$$\sqrt{n}\left(\overline{Y}_n - \frac{\pi_0}{1-\pi_0}\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N\left(0, \frac{\pi_0}{(1-\pi_0)^2}\right). \tag{7.3}$$

From (7.3) and from the Taylor formula

$$\phi\left(\overline{Y}_n\right) - \phi\left(\frac{\pi_0}{1-\pi_0}\right) = \phi'\left(\frac{\pi_0}{1-\pi_0}\right)\left(\overline{Y}_n - \frac{\pi_0}{1-\pi_0}\right) + o_p\left(\overline{Y}_n - \frac{\pi_0}{1-\pi_0}\right)$$

for $\phi(y) = y/(1+y)$ we obtain

$$\sqrt{n}\left(\phi\left(\overline{Y}_n\right) - \phi\left(\frac{\pi_0}{1-\pi_0}\right)\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N\left(0, \phi'\left(\frac{\pi_0}{1-\pi_0}\right)^2 \frac{\pi_0}{(1-\pi_0)^2}\right).$$

Since $\phi(\pi/(1-\pi)) = \pi$, this means that the estimator $\pi_n = \overline{Y}_n/\left(1+\overline{Y}_n\right)$ of π_0 satisfies the relation

$$\begin{aligned}\sqrt{n}(\pi_n - \pi_0) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} & \; N\left(0, (1-\pi_0)^4 \frac{\pi_0}{(1-\pi_0)^2}\right) \\ = & \; N\left(0, \pi_0(1-\pi_0)^2\right) \\ = & \; N(0, 1/I(\pi_0)) \qquad \text{(cf. (7.1))}.\end{aligned} \tag{7.4}$$

Further, since

$$\frac{\pi_0}{1-\pi_0} = e^{\beta_0},$$

we get from (7.3) and from the Taylor formula

$$\ln\left(\overline{Y}_n\right) - \ln e^{\beta_0} = \frac{1}{e^{\beta_0}}\left(\overline{Y}_n - e^{\beta_0}\right) + o_p\left(\overline{Y}_n - e^{\beta_0}\right)$$

that the MLE $\beta_n = \ln\left(\overline{Y}_n\right)$ of β_0 satisfies the relation

$$\begin{aligned}\sqrt{n}\left(\beta_n - \beta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} & \; N\left(0, \frac{e^{\beta_0}\left(1+e^{\beta_0}\right)}{e^{2\beta_0}}\right) \\ = & \; N\left(0, 1/\pi\left(\beta_0\right)\right) \\ = & \; N\left(0, 1/\mathcal{J}\left(\beta_0\right)\right)\end{aligned} \tag{7.5}$$

where

$$\mathcal{J}(\beta) = \sum_{k=0}^{\infty} \frac{\left(dp_{\pi(\beta)}(k)/d\beta\right)^2}{p_{\pi(\beta)}(k)} = \pi(\beta)$$

with $p_\pi(k)$ denoting the jumps (2.18) of the geometric response function is the Fisher information in the logistic regression model (6.1) and (6.2) with geometric responses.

Let us now consider the median estimator

$$\widehat{\pi}_n = \arg\min_{\pi} \sum_{i=1}^{n} \left|Y_i - m(\pi)\right| \tag{7.6}$$

of π_0 for the median function $m(\pi)$ given in (2.39). By (6.17), $\widehat{\pi}_n$ is the solution of the equation

$$m\left(\widehat{\pi}_n\right) = Y_{(n/2)}, \qquad \text{i.e.} \qquad \widehat{\pi}_n = m^{-1}\left(Y_{(n/2)}\right), \tag{7.7}$$

and the median estimator $\widehat{\beta}_n$ of β_0 defined by (2.22) is given, by

$$\widehat{\beta}_n = \ln \frac{\widehat{\pi}_n}{1-\widehat{\pi}_n}. \tag{7.8}$$

Let us analyze the case $\beta_0 < 0$, i.e. $\pi_0 < 1/2$. In the domain $\pi < 1/2$ we get from (2.39)

$$m(\pi) = \frac{1}{2(1-\pi)} \in (1/2, 1). \tag{7.9}$$

Thus if $Y_{(n/2)} \geq 1/2$ then (7.7) and (7.8) imply

$$\widehat{\pi}_n = \frac{2Y_{(n/2)} - 1}{2Y_{(n/2)}} \qquad \text{and} \qquad \widehat{\beta}_n = \ln\left(2Y_{(n/2)} - 1\right). \tag{7.10}$$

Since

$$\psi'(y) = 2\left(1-\psi(y)\right)^2 \qquad \text{for} \qquad \psi(y) = \frac{2y-1}{2y}$$

and $\psi\left(m\left(\pi_0\right)\right) = \pi_0$, it holds the Taylor formula

$$\psi\left(Y_{(n/2)}\right) - \pi_0 = 2\left(1-\pi_0\right)^2\left(Y_{(n/2)} - m\left(\pi_0\right)\right) + o_p\left(Y_{(n/2)} - m\left(\pi_0\right)\right). \tag{7.11}$$

By p. 490 in Rényi, the relation (6.20) remains valid for the present median formula (7.9) and the present density

$$f_{\pi}(y) = (1-\pi)\sum_{k=0}^{\infty}\pi^{k} I(k < y \leq k+1). \tag{7.12}$$

Since for $\pi_0 < 1/2$ we have $m(\pi_0) < 1$, it holds $f_{\pi_0}(m(\pi_0)) = 1 - \pi_0$ and we get from (6.20)

$$\sqrt{n}\left(Y_{(n/2)} - m(\pi_0)\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N\left(0, \frac{1}{4(1-\pi_0)^2}\right). \tag{7.13}$$

This together with (7.11) implies

$$\sqrt{n}\left(\widehat{\pi}_n - \pi_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N\left(0, (1-\pi_0)^2\right). \tag{7.14}$$

Similarly the definition of $\widehat{\beta}_n$ in (7.10) together with (7.13) implies

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \quad &\underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \quad N\left(0, \left(\frac{2}{2m(\pi_0)-1}\right)^2 \cdot \frac{1}{4(1-\pi_0)^2}\right) \\ &= \quad N\left(0, \frac{4(1-\pi_0)^2}{\pi_0^2} \cdot \frac{1}{4(1-\pi_0)^2}\right) \\ &= \quad N\left(0, 1/\pi^2(\beta_0)\right).\end{aligned} \tag{7.15}$$

If we take into account that $\pi_0 = \pi(\beta_0) < 1/2$, we see that the asymptotic variance $(1-\pi_0)^2$ of the estimator $\widehat{\pi}_n$ in (7.14) is larger than the asymptotic variance $\pi_0(1-\pi_0)^2$ of π_n in (7.4). Similarly the asymptotic variance $1/\pi^2(\beta_0)$ of the median estimator $\widehat{\beta}_n$ in (7.15) is larger than the asymptotic variance $1/\pi(\beta_0)$ of the MLE β_n in (7.5).

In spite of that the MLE β_n is asymptotically preferable to the median estimator $\widehat{\beta}_n$ under the true geometric model (6.1), (6.2), the opposite asymptotic preference can be obtained if an alternative model producing outliers takes places with an arbitrary small probability $\varepsilon > 0$. To see this, suppose that the geometric distribution function $F_{\pi(\beta_0)}(y)$ is replaced by the mixture

$$(1-\varepsilon)F_{\pi(\beta_0)}(y) + \varepsilon G(y) \tag{7.16}$$

where $G(y)$ is a step function on $\mathbb{R}$ with the jumps

$$G(k) - G(k-0) = \frac{1}{k(k+1)} \tag{7.17}$$

at $k = 1, 2, ...$, i.e. where

$$G(k) = 1 - \frac{1}{k+1} \qquad \text{for } k = 0, 1, \dots\,. \tag{7.18}$$

If the responses $Y_1, ..., Y_n$ are distributed by (7.16) then the sample mean $\overline{Y}_n$ tends to the expectation

$$(1-\varepsilon)\,\frac{\pi(\beta_0)}{1-\pi(\beta_0)} + \varepsilon \sum_{k=1}^{\infty} \frac{k}{k(k+1)} \tag{7.19}$$

which is infinite as soon as $\varepsilon > 0$. Here for all $\varepsilon > 0$ the MLE $\widehat{\beta}_n = \ln\left(\overline{Y}_n\right)$ fails to meet the asymptotic normality (7.5) because it surely diverges to infinity. On the other hand, the median function $m_\varepsilon(\pi)$ of the equivalent continuous modification of the model (7.16) is finite and tends to $m(\pi)$ for $\varepsilon \downarrow 0$. Hence the asymptotic bias of the median estimator $\widehat{\beta}_n$ will vanish for $\varepsilon \downarrow 0$. If $\pi < 1/2$ then $m_\varepsilon(\pi)$ is given by the formula

$$m_\varepsilon(\pi) = \frac{1}{2(1-\varepsilon)(1-\pi)+\varepsilon} \qquad \text{(cf. (7.9))} \tag{7.20}$$

so that instead of (7.10) we obtain from (7.7) and (7.8)

$$\widehat{\pi}_n = \frac{2Y_{(n/2)} - 1 - \varepsilon Y_{(n/2)}}{2Y_{(n/2)}(1-\varepsilon)} \qquad \text{and} \qquad \widehat{\beta}_n = \ln\left(\frac{2Y_{(n/2)} - 2}{1-\varepsilon Y_{(n/2)}} + 1\right). \tag{7.21}$$

8 Binary regressors in univariate Bernoulli models

In this section we study the same Bernoulli model as in Section 7, of the dimension $d = 1$, but with the univariate regressors $x_1, ..., x_n$ taking on two different values. For simplicity we suppose that n is even and

$$x_1 = x_2 = ... = x_{n/2} = 1, \qquad x_{n/2+1} = x_{n/2+2} = ... = x_n = -1.$$

We consider the binary responses

$$Y_i \sim F_{\pi(\beta_0)}(y), \qquad 1 \le i \le n/2 \tag{8.1}$$

and

$$Y_i \sim F_{1-\pi(\beta_0)}(y), \qquad n/2 < i \le n \tag{8.2}$$

for

$$\pi(\beta) = \frac{e^\beta}{1+e^\beta} \qquad \text{and} \qquad 1-\pi(\beta) = \frac{e^{-\beta}}{1+e^{-\beta}} = \frac{1}{1+e^\beta} \tag{8.3}$$

where $F_\pi(y)$ is the distribution of (2.35) corresponding to the standardly modified Bernoulli regression with the median function $m(\pi)$ given by (2.36). We shall evaluate the three robust estimators $\widehat{\beta}_n$, $\beta_n^{(1)}$ and $\beta_n^{(2)}$ of β_0 introduced in Section 2.

By (2.22), our median estimator $\widehat{\beta}_n$ is defined by

$$\widehat{\beta}_n = \arg\min_\beta \left(\sum_{i=1}^{n/2} |Y_i - m(\pi(\beta))| + \sum_{i=n/2+1}^{n} |Y_i - m(1-\pi(\beta))| \right) \tag{8.4}$$

for the standardly modified Bernoulli logistic regression data Y_i. Obviously,

$$\widehat{\beta}_n = \ln \frac{\widehat{\pi}_n}{1-\widehat{\pi}_n} \tag{8.5}$$

where $\widehat{\pi}_n$ is the median estimator of $\pi_0 = \pi\left(\beta_0\right)$ defined by

$$\widehat{\pi}_n = \arg\min_{\pi} \left(\sum_{i=1}^{n/2} \left|Y_i - m\left(\pi\right)\right| + \sum_{i=n/2+1}^{n} \left|Y_i - m\left(1-\pi\right)\right| \right). \tag{8.6}$$

Since $m\left(\pi\right)$ of (2.36) is strictly monotone in $\pi \in (0,1)$ and

$$\phi\left(m\right) = \sum_{i=1}^{k} \left|Y_i - m\right|$$

is decreasing in the domain $m \in \left(-\infty, Y_{(k/2)}\right)$, the function

$$\psi\left(\pi\right) = \sum_{i=1}^{n/2} \left|Y_i - m\left(\pi\right)\right| + \sum_{i=n/2+1}^{n} \left|Y_i - m\left(1-\pi\right)\right|$$

is decreasing in the domain

$$\pi \in \left(0, \widehat{\pi}_n^+ \wedge \widehat{\pi}_n^-\right]$$

and increasing in the domain

$$\pi \in \left(\widehat{\pi}_n^+ \wedge \widehat{\pi}_n^-, \infty\right]$$

where

$$\widehat{\pi}_n^+ = m^{-1}\left(Y_{(n/4)}\right), \ \widehat{\pi}_n^- = 1 - m^{-1}\left(Y_{(3n/4)}\right). \tag{8.7}$$

Thus the minimization in (8.6) can be restricted to the interval

$$\pi \in \left[\widehat{\pi}_n^+ \wedge \widehat{\pi}_n^-,\ \widehat{\pi}_n^+ \vee \widehat{\pi}_n^-\right]$$

for $\widehat{\pi}_n^+$ and $\widehat{\pi}_n^-$ given by (8.7). But we can prove more.

Theorem 8.1. The median estimates $\widehat{\beta}_n$ and $\widehat{\pi}_n$ are a.s. uniquely defined by (8.4), (8.6) and it holds

$$\widehat{\pi}_n = med\left(Y_1, ..., Y_{n/2}, 2 - Y_{n/2+1}, ..., 2 - Y_n\right) \tag{8.8}$$

while $\widehat{\beta}_n$ is obtained by applying (8.5) to (8.8).

Proof. It follows from (2.36) that $m(\pi) = 2 - m(1-\pi)$. Therefore

$$\sum_{i=n/2+1}^{n} \left|Y_i - m\left(1-\pi\right)\right| = \sum_{i=n/2+1}^{n} \left|2 - Y_i - m\left(\pi\right)\right|$$

and

$$\widehat{\pi}_n = \arg\min_{\pi} \sum_{i=1}^{n} |Z_i - m(\pi)| = m^{-1}\left(Z_{(n/2)}\right)$$

where

$$Z_i = \begin{cases} Y_i & \text{when} \quad 1 \le i \le n/2 \\ 2 - Y_i & \text{when} \quad n/2 \le i \le n \end{cases}$$

and the median $Z_{(n/2)}$ is a.s. unique $\arg\min_m \sum_{i=1}^{n} |Z_i - m|$. The statement of Theorem 5.1 is clear from here.

■

The robust estimator $\beta_n^{(1)}$ of Morgenthaler (1992) solves the equation $U_n^{(1)}(\beta) = 0$ where, by (5.6),

$$\begin{aligned} U_n^{(1)}(\beta) &= \sum_{i=1}^{n/2} \sqrt{\pi(\beta)(1-\pi(\beta))}\,(Y_i - \pi(\beta)) - \sum_{i=n/2+1}^{n} \sqrt{\pi(\beta)(1-\pi(\beta))}\,(Y_i - 1 + \pi(\beta)) \\ &= \sqrt{\pi(\beta)(1-\pi(\beta))}\left[\left(\sum_{i=1}^{n/2} Y_i - \sum_{i=n/2+1}^{n} Y_i + \tfrac{n}{2} - n\pi(\beta)\right)\right] \end{aligned}$$

for given binary responses Y_i. Hence

$$\beta_n^{(1)} = \ln \frac{\pi_n^{(1)}}{1 - \pi_n^{(1)}} \tag{8.9}$$

for

$$\pi_n^{(1)} = \frac{\overline{Y}_n^{+} + 1 - \overline{Y}_n^{-}}{2} \tag{8.10}$$

where

$$\overline{Y}_n^{+} = \frac{1}{n/2} \sum_{i=1}^{n/2} Y_i \qquad \text{and} \qquad \overline{Y}_n^{-} = \frac{1}{n/2} \sum_{i=n/2+1}^{n} Y_i \tag{8.11}$$

are the average regressors corresponding to the explanatory variables $x_i = +1$ and $x_i = -1$, respectively.

The robust estimator $\beta_n^{(2)}$ of Bianco and Yohai (1996) is defined by (5.12) and depends on a constant $c > 0$ figuring in (5.10) . The consistency of $\beta_n^{(2)}$ was proved for $c > \ln 2$ so that we can consider $c = 1$. This means that

$$\beta_n^{(2)} = \arg\min_{\beta} \left(M^{+}(\beta) + M^{-}(\beta)\right) \tag{8.12}$$

for

$$M^{+}(\beta) = \sum_{i=1}^{n/2} \left[\rho\left(-Y_i \ln \pi(\beta) - (1 - Y_i)\ln(1 - \pi(\beta))\right) + H(\pi(\beta))\right]$$

and

$$M^{-}(\beta)=\sum_{i=n/2+1}^{n}\left[\rho\left(-Y_i\ln\left(1-\pi\left(\beta\right)\right)-\left(1-Y_i\right)\ln\pi\left(\beta\right)\right)+H\left(\pi\left(\beta\right)\right)\right].$$

In these formulas

$$\rho(y)=\left(y-\frac{y^2}{2}\right)I\left(0\leq y\leq 1\right)+\frac{1}{2}I\left(y>1\right) \tag{8.13}$$

(cf. (5.10)) and

$$H(\pi)=G(\pi)+G(1-\pi) \tag{8.14}$$

where

$$G(\pi)=I\left(\pi>1/e\right)\left(\pi\ln\pi+\frac{1}{e}\right),\ \pi\in(0,1)\qquad \text{(cf.(5.11))}. \tag{8.15}$$

We see from here that

$$\beta_n^{(2)}=\ln\frac{\pi_n^{(2)}}{1-\pi_n^{(2)}} \tag{8.16}$$

where

$$\pi_n^{(2)}=\arg\min_{\pi}\left[N^{+}(\pi)+N^{-}(\pi)+nH(\pi)\right]$$

estimates the value $\pi_0=\pi\left(\beta_0\right)$. In the last formula

$$\begin{aligned}N^{+}(\pi) &= \sum_{i=1}^{n/2}\left[\rho\left(-Y_i\ln\pi-\left(1-Y_i\right)\ln\left(1-\pi\right)\right)\right]\\ &= \tfrac{n}{2}\left[\rho\left(-\ln\pi\right)\overline{Y}_n^{+}+\rho\left(-\ln\left(1-\pi\right)\right)\left(1-\overline{Y}_n^{+}\right)\right]\end{aligned}$$

for $\overline{Y}_n^{+}$ defined by (8.11) and, similarly,

$$N^{-}(\pi)=\frac{n}{2}\left[\rho\left(-\ln\left(1-\pi\right)\right)\overline{Y}_n^{-}+\rho\left(-\ln\pi\right)\left(1-\overline{Y}_n^{-}\right)\right]$$

for $\overline{Y}_n^{-}$ defined by (8.11). Therefore

$$\frac{1}{n}\left(N^{+}(\pi)+N^{-}(\pi)+nH(\pi)\right)=p_n\rho\left(-\ln\pi\right)+\left(1-p_n\right)\rho\left(-\ln\left(1-\pi\right)\right)+H(\pi)$$

where

$$p_n=\frac{\overline{Y}_n^{+}+1-\overline{Y}_n^{-}}{2}\quad\text{and}\quad 1-p_n=\frac{\overline{Y}_n^{-}+1-\overline{Y}_n^{+}}{2}. \tag{8.17}$$

Consequently,

$$\pi_n^{(2)}=\arg\min_{\pi}L\left(p_n,\pi\right) \tag{8.18}$$

where

$$L\left(p_n,\pi\right)=p_n\rho\left(-\ln\pi\right)+\left(1-p_n\right)\rho\left(-\ln\left(1-\pi\right)\right)+G(\pi)+G(1-\pi) \tag{8.19}$$

for p_n given by (8.17), ρ given by (8.13) and G given by (8.15).

Theorem 8.2. The estimators $\pi_n^{(2)}$ and $\beta_n^{(2)}$ are uniquely defined by (8.18) and (8.12) and satisfy the relations

$$\pi_n^{(2)} = p_n = \frac{\overline{Y}_n^+ + 1 - \overline{Y}_n^-}{2} \tag{8.20}$$

and

$$\beta_n^{(2)} = \ln \frac{p_n^{(2)}}{1 - p_n^{(2)}} = \ln \frac{\overline{Y}_n^+ + 1 - \overline{Y}_n^-}{\overline{Y}_n^- + 1 - \overline{Y}_n^+} \tag{8.21}$$

Proof. In view of (8.16), it suffices to prove that (8.20) is the unique minimizer of the function (8.19) in the domain $\pi \in (0,1)$. But

$$\begin{aligned} \frac{d}{d\pi} L(p_n, \pi) = & -\frac{p_n}{\pi}\rho'(-\ln \pi) + \frac{1-p_n}{1-\pi}\rho'(-\ln(1-\pi)) \\ + & \rho'(-\ln \pi) - \rho'(-\ln(1-\pi)) \\ = & (\pi - p_n)\left[\frac{\rho'(-\ln \pi)}{\pi} + \frac{\rho'(-\ln(1-\pi))}{1-\pi}\right] \end{aligned}$$

where $\rho'(y) = (1-y)\, I(0 < y \le 1)$. Hence the expression in the brackets is

$$\frac{1+\ln \pi}{\pi} I(\pi > 1/e) + \frac{1 + \ln(1-\pi)}{1-\pi} I(\pi < 1 - 1/e)$$

which is positive for every $\pi \in (0,1)$. Therefore $\pi = p_n$ is the unique minimum of $L(p_n, \pi)$ in the domain $\pi \in (0,1)$.

■

We see from (8.9), (8.10) and (8.20), (8.21) that in the model considered in the present section the estimators $\beta_n^{(1)}$ of Morgenthaler (1992) and $\beta_n^{(2)}$ of Bianco and Yohai (1996) coincide. In the following section we shall denote by $\widetilde{\beta}_n$ the common value of $\beta_n^{(1)}$ and $\beta_n^{(2)}$ from (8.9) and (8.21) and by $\widetilde{\pi}_n$ the common value of $\pi_n^{(1)}$ and $\pi_n^{(2)}$ from (8.10) and (8.20). It is clear that $\widetilde{\pi}_n$ differs from $\widehat{\pi}_n$ given by (8.8) and, consequently, our median estimator $\widehat{\beta}_n$ differs from the estimator $\widetilde{\beta}_n$ of Morgenthaler, Bianco and Yohai.

9 Simulations: Binary regressors in univariate Bernoulli models

In this section we study the model of logistic regression of Section 8 with the binary regressors x_i, $1 \le i \le n$, uniformly distributed on $\{1, -1\}$ and with the independent Bernoulli responses $Y_i = 1$ and $Y_i = 0$ taken on with the corresponding binomial probabilities

$$\pi_0 = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \quad \text{and} \quad 1 - \pi_0 = \frac{e^{-\beta_0}}{1 + e^{-\beta_0}} = \frac{1}{1 + e^{\beta_0}}.$$

We restrict ourselves to the parameters $\beta_0 \in \{1/4,\ 1/2, 1\}$ and the sample sizes $n \in \{50, 100\}$. The ideal data set

$$\left\{x_i,\ Y_i \sim Be\left(e^{x_i\beta_0}/\left(1+e^{x_i\beta_0}\right)\right)\right\}, \qquad 1 \le i \le n,$$

was contamined by applying the binary symmetric channel of Fig. 5.1 to the responses Y_i. Therefore the true data set was defined by

$$\left\{x_i,\ Y_i \sim (1-\varepsilon)\, Be\left(e^{x_i\beta_0}/\left(1+e^{x_i\beta_0}\right)\right) + \varepsilon Be\left(e^{-x_i\beta_0}/\left(1+e^{-x_i\beta_0}\right)\right)\right\} \tag{9.1}$$

for $1 \le i \le n$. This means that we considered the contamination introduced in Section 5. The level ε of this contamination was taken from the set $\{0.1, 0.2\}$. We compared the common values $\widetilde{\beta}_n$ of the estimator of Morgenthaler and that of Bianco and Yohai introduced in Section 5 and explicitly evaluated for the present model in (8.9), (8.10) and (8.20), (8.21) with the median estimates $\widehat{\beta}_n$ explicitly evaluated in in (8.5) and (8.8). Namely, we simulated 1000 times the random data (9.1) for $1 \le i \le n$ and evaluated the resulting estimates

$$\widetilde{\beta}_n(l),\ \widehat{\beta}_n(l), \qquad 1 \le l \le 1000.$$

In Tables 9.1 and 9.2 we present the possibility to compare the estimates $\widetilde{\beta}_n$ and $\widehat{\beta}_n$ for $n = 50$ and $n = 100$ on the basis of the mean values

$$\widetilde{\beta}_{0,n} = \frac{1}{1000}\sum_{l=1}^{1000}\widetilde{\beta}_n(l), \qquad \widehat{\beta}_{0,n} = \frac{1}{1000}\sum_{l=1}^{1000}\widehat{\beta}_n(l) \tag{9.2}$$

and the standard deviations

$$\widetilde{\sigma}_n = \left(\frac{1}{1000}\sum_{l=1}^{1000}\left[\widetilde{\beta}_n(l) - \widetilde{\beta}_{0,n}\right]^2\right)^{1/2}, \qquad \widehat{\sigma}_n = \left(\frac{1}{1000}\sum_{l=1}^{1000}\left[\widehat{\beta}_n(l) - \widehat{\beta}_{0,n}\right]^2\right)^{1/2}. \tag{9.3}$$

We see that in the examined model our estimates $\widehat{\beta}_n$ resist better to the increasing contamination levels ε than those of Morgenthaler or Bianco and Yohai in the sense that, in average, $\widehat{\beta}_n$ are closer to the true β_0 than $\widetilde{\beta}_n$. The dispersions of estimates $\widehat{\beta}_n$ around the mean values $\widehat{\beta}_{0,n}$ exceed the dispersion of $\widetilde{\beta}_n$ around $\widetilde{\beta}_{0,n}$ but for $n = 100$ these dispersions are mutually comparable.

From Tables 9.1 and 9.2 one can draw the conclusion that the median estimator $\widehat{\beta}_n$ deserves to be studied alongside with the estimators of Morgenthaler (1992) and Bianco and Yohai (1996), because it seems to be more robust with respect to heavy contaminations than these two.

	$\beta_0 = \mathbf{1/4}$		$\beta_0 = \mathbf{1/2}$		$\beta_0 = \mathbf{1}$	
	$\varepsilon = 0.1$	$\varepsilon = 0.2$	$\varepsilon = 0.1$	$\varepsilon = 0.2$	$\varepsilon = 0.1$	$\varepsilon = 0.2$
$\widetilde{\beta}_{0,50}$	**0.21**	**0.14**	**0.41**	**0.30**	**0.80**	**0.58**
$\widetilde{\sigma}_{50}$	0.08	0.08	0.09	0.09	0.10	0.09
$\widehat{\beta}_{0,50}$	**0.36**	**0.27**	**0.62**	**0.47**	**1.04**	**0.77**
$\widehat{\sigma}_{50}$	0.17	0.16	0.30	0.21	0.40	0.29

Table 9.1: Means $\widetilde{\beta}_{0,n}$ and $\widehat{\beta}_{0,n}$ and standard deviations $\widetilde{\sigma}_n$, $\widehat{\sigma}_n$ of 1000 realizations of the estimates $\widetilde{\beta}_n$, $\widehat{\beta}_n$ for the sample size $n = 50$.

	$\beta_0 = \mathbf{1/4}$		$\beta_0 = \mathbf{1/2}$		$\beta_0 = \mathbf{1}$	
	$\varepsilon = 0.1$	$\varepsilon = 0.2$	$\varepsilon = 0.1$	$\varepsilon = 0.2$	$\varepsilon = 0.1$	$\varepsilon = 0.2$
$\widetilde{\beta}_{0,100}$	**0.19**	**0.15**	**0.40**	**0.31**	**0.78**	**0.58**
$\widetilde{\sigma}_{100}$	0.04	0.04	0.04	0.04	0.05	0.04
$\widehat{\beta}_{0,100}$	**0.26**	**0.22**	**0.48**	**0.39**	**0.90**	**0.67**
$\widehat{\sigma}_{100}$	0.07	0.06	0.09	0.08	0.15	0.11

Table 9.2: Means $\widetilde{\beta}_{0,n}$ and $\widehat{\beta}_{0,n}$ and standard deviations $\widetilde{\sigma}_n$, $\widehat{\sigma}_n$ of 1000 realizations of the estimates $\widetilde{\beta}_n$, $\widehat{\beta}_n$ for the sample size $n = 100$.

We verified more precisely the above mentioned hint of better resistance of the median estimator to contamination of data than the resistance of the robust estimators of Morgenthaler and Biancco-Yohai. To this end we evaluated for $\beta_0 \in \{1/4, 1/2, 1\}$ the mean absolute errors

$$MAE(n) = \frac{1}{1000} \sum_{l=1}^{1000} \left| \widehat{\beta}_n(l) - \beta_0 \right|$$

of the median estimates $\widehat{\beta}_n$ and similar mean absolute errors

$$MAE(n) = \frac{1}{1000} \sum_{l=1}^{1000} \left| \widetilde{\beta}_n(l) - \beta_0 \right|$$

of the coinciding estimates $\widetilde{\beta}_n$ of Morgenthaler and Bianco-Yohai. The results are in Tables 9.3-9.5.

Let us see which of the estimators $\widetilde{\beta}_n$, $\widehat{\beta}_n$ is better in the sense of the *MAE(n)*. Looking first at the middle column in Table 9.4 we see that for no contamination (i.e. $\varepsilon = 0$) $\widetilde{\beta}_{100}$ is considerably better than $\widehat{\beta}_{100}$. But for $\varepsilon = 0.2$ these two estimators are equivalent and if $\varepsilon \geq 0.3$ then $\widehat{\beta}_{100}$ is better than $\widetilde{\beta}_{100}$. For the larger sample size $n = 200$ the domination of $\widetilde{\beta}_{200}$ over $\widehat{\beta}_{200}$ at $\varepsilon = 0$ is less dramatic than for $n = 100$ but the reversed domination at $\varepsilon = 0.3$ is better visible. For smaller sample size $n = 50$ the change is opposite and $\widetilde{\beta}_{50}$ in fact dominates $\widehat{\beta}_{50}$ for all contamination levels $0 \leq \varepsilon \leq 0.3$.

The conclusions deduced from Table 9.4 are even more evident from Table 9.5 because the difference between the Bernoulli parameters

$$e^{x_i\beta_0} / \left(1 + e^{x_i\beta_0}\right) \text{ and } e^{-x_i\beta_0} / \left(1 + e^{-x_i\beta_0}\right)$$

in (9.1) is larger for $\beta_0 = 1$ than for $\beta_0 = 1/2$, i.e. the influence of the contamination in (9.1) on the observations is larger for $\beta_0 = 1$ than for $\beta_0 = 1/2$. Since the same difference is smaller for $\beta_0 = 1/4$ than for $\beta_0 = 1/2$, the conclusions deduced from Table 9.4 are less evidently supported by Table 9.3.

ε	Estimator	$MAE(50)$	NEF	$MAE(100)$	NEF	$MAE(200)$	NEF
0.00	$\widetilde{\beta}_n$	0.229	0	0.170	0	0.116	0
	$\widehat{\beta}_n$	0.330	4	0.208	0	0.140	0
0.05	$\widetilde{\beta}_n$	0.233	0	0.171	0	0.117	0
	$\widehat{\beta}_n$	0.317	1	0.206	0	0.135	0
0.10	$\widetilde{\beta}_n$	0.227	0	0.167	0	0.119	0
	$\widehat{\beta}_n$	0.299	0	0.200	0	0.134	0
0.20	$\widetilde{\beta}_n$	0.243	0	0.183	0	0.144	0
	$\widehat{\beta}_n$	0.288	2	0.201	0	0.142	0
0.30	$\widetilde{\beta}_n$	0.267	0	0.208	0	0.172	0
	$\widehat{\beta}_n$	0.312	2	0.212	0	0.166	0

Table 9.3: Mean absolute errors MAE(n) of the two estimators $\widetilde{\beta}_n$ and $\widehat{\beta}_n$ for the sample size $n \in \{50, 100, 200\}$ and true parameter $\beta_0 = 1/4$. The column NEF presents the number of simulation vectors $(Y_1, ..., Y_n)$ for which the evaluation of the corresponding estimates failed.

ε	Estimator	$MAE(50)$	NEF	$MAE(100)$	NEF	$MAE(200)$	NEF
0.00	$\widetilde{\beta}_n$	0.237	0	0.165	0	0.120	0
	$\widehat{\beta}_n$	0.397	7	0.233	0	0.166	0
0.05	$\widetilde{\beta}_n$	0.239	0	0.171	0	0.125	0
	$\widehat{\beta}_n$	0.359	3	0.227	0	0.154	0
0.10	$\widetilde{\beta}_n$	0.240	0	0.181	0	0.139	0
	$\widehat{\beta}_n$	0.327	4	0.219	0	0.160	0
0.20	$\widetilde{\beta}_n$	0.281	0	0.239	0	0.212	0
	$\widehat{\beta}_n$	0.324	3	0.248	0	0.203	0
0.30	$\widetilde{\beta}_n$	0.355	0	0.321	0	0.303	0
	$\widehat{\beta}_n$	0.368	3	0.301	0	0.279	0

Table 9.4: The same as in Table 9.3 for $\beta_0 = 1/2$.

ε	Estimator	$MAE(50)$	NEF	$MAE(100)$	NEF	$MAE(200)$	NEF
0.00	$\widetilde{\beta}_n$	0.259	0	0.188	1	0.125	0
	$\widehat{\beta}_n$	0.535	60	0.377	1	0.232	0
0.05	$\widetilde{\beta}_n$	0.272	0	0.204	2	0.157	0
	$\widehat{\beta}_n$	0.484	36	0.329	2	0.221	0
0.10	$\widetilde{\beta}_n$	0.303	0	0.249	2	0.228	0
	$\widehat{\beta}_n$	0.437	20	0.314	2	0.246	0
0.20	$\widetilde{\beta}_n$	0.438	0	0.430	0	0.427	0
	$\widehat{\beta}_n$	0.464	7	0.409	0	0.397	0
0.30	$\widetilde{\beta}_n$	0.622	0	0.633	0	0.619	0
	$\widehat{\beta}_n$	0.583	3	0.569	0	0.582	0

Table 9.5: The same as in Table 9.3 for $\beta_0 = 1$.

10 Simulations: Identical regressors in univariate geometric models

In this section we study the model of logistic regression of Section 7 with the identical regressors $x_i = 1$, $1 \le i \le n$, and contaminated geometric responses $Y_i \sim (1-\varepsilon)\, F_{\pi_0}(y) + \varepsilon G(y)$, for $1 \le i \le n$, where $F_{\pi_0}(y)$ is the geometric distribution function with jumps $(1-\pi_0)\,\pi_0^k$ at $k = 0, 1, ...$ and $G(y)$ is the step distribution function with jumps (7.17) at $k = 1, 2, ...$ We consider $\pi_0 = e^{\beta_0} / \left(1 + e^{\beta_0}\right)$ for the same true parameters $\beta_0 \in \{1/4, 1/2, 1\}$ and the same sample sizes $n \in \{50, 100\}$ as in Section 9. We compare the MLE $\beta_n = \ln\left(\overline{Y}_n\right)$ and the median estimator $\widehat{\beta}_n$ defined by (7.7), (7.8). Similarly as in the previous section, for $\varepsilon = 0.1$ and $\varepsilon = 0.3$, we simulated 1000 times the contaminated data sets $\{Y_1, ..., Y_n\}$ leading to realizations $\beta_n(l), \widehat{\beta}_n(l), \ 1 \le l \le 1000$ of the estimates $\beta_n, \widehat{\beta}_n$. Using these realizations we present in Tables 10.1 and 10.2 the corresponding means $\beta_{0,n}$ and $\widehat{\beta}_{0,n}$ and standard deviations σ_n, $\widehat{\sigma}_n$ obtained from obvious modifications of formulas (9.2) and (9.3). We see from these tables that in heavily contaminated geometric models the median estimator $\widehat{\beta}_n$ deviates from the true values β_0 in average much less than the MLE β_n, and also that the values of $\widehat{\beta}_n$ are less dispersed than those of β_n. Therefore these tables justify the deeper interest in the median estimator $\widehat{\beta}_n$ introduced in this paper.

	$\beta_0 = \mathbf{1/4}$		$\beta_0 = \mathbf{1/2}$		$\beta_0 = \mathbf{1}$	
	$\varepsilon = 0.1$	$\varepsilon = 0.3$	$\varepsilon = 0.1$	$\varepsilon = 0.3$	$\varepsilon = 0.1$	$\varepsilon = 0.3$
$\beta_{0,50}$	**0.41**	**0.80**	**0.61**	**0.90**	**1.06**	**1.18**
σ_{50}	0.20	0.56	0.17	0.10	0.12	0.07
$\widehat{\beta}_{0,50}$	**0.46**	**0.40**	**0.73**	**0.64**	**1.13**	**1.06**
$\widehat{\sigma}_{50}$	0.12	0.14	0.09	0.12	0.07	0.09

Table 10.1: Means $\beta_{0,n}$ and $\widehat{\beta}_{0,n}$ and standard deviations σ_n, $\widehat{\sigma}_n$ of 1000 realizations of the estimates β_n, $\widehat{\beta}_n$ for the sample size $n = 50$.

	$\beta_0 = \mathbf{1/4}$		$\beta_0 = \mathbf{1/2}$		$\beta_0 = \mathbf{1}$	
	$\varepsilon = 0.1$	$\varepsilon = 0.3$	$\varepsilon = 0.1$	$\varepsilon = 0.3$	$\varepsilon = 0.1$	$\varepsilon = 0.3$
$\beta_{0,100}$	**0.48**	**0.84**	**0.67**	**0.96**	**1.07**	**1.25**
σ_{100}	0.24	0.40	0.20	0.35	0.07	0.30
$\widehat{\beta}_{0,100}$	**0.42**	**0.36**	**0.71**	**0.60**	**1.21**	**1.04**
$\widehat{\sigma}_{100}$	0.07	0.08	0.05	0.07	0.04	0.05

Table 10.2: Means $\beta_{0,n}$ and $\widehat{\beta}_{0,n}$ and standard deviations σ_n, $\widehat{\sigma}_n$ of 1000 realizations of the estimates β_n, $\widehat{\beta}_n$ for the sample size $n = 100$.

Next follow Tables 10.3-10.5 presenting for $\beta_0 \in \{1/4, 1/2, 1\}$ the mean absolute errors

$$MAE(n) = \frac{1}{1000} \sum_{l=1}^{1000} |\beta_n(l) - \beta_0|$$

of the MLE's β_n and similar mean absolute errors

$$MAE(n) = \frac{1}{1000} \sum_{l=1}^{1000} \left|\widehat{\beta}_n(l) - \beta_0\right|$$

of the median estimates $\widehat{\beta}_n$. The situation is similar as observed in Tables 9.3-9.5 except that the mean errors of the estimates β_{100} and $\widehat{\beta}_{100}$ in Table 10.4 become equal already for

$\varepsilon \approx 0.1$ and the domination of $\widehat{\beta}_n$ in the sense of robustness is more evidently demostrated by these tables than by Tables 9.3-9.5 of Section 9.

ε	Estimator	$MAE(50)$	$MAE(100)$	$MAE(200)$
0.00	$\widetilde{\beta}_n$	0.156	0.109	0.076
	$\widehat{\beta}_n$	0.343	0.271	0.214
0.05	$\widetilde{\beta}_n$	0.227	0.202	0.170
	$\widehat{\beta}_n$	0.342	0.268	0.208
0.10	$\widetilde{\beta}_n$	0.298	0.283	0.306
	$\widehat{\beta}_n$	0.344	0.266	0.205
0.20	$\widetilde{\beta}_n$	0.439	0.466	0.510
	$\widehat{\beta}_n$	0.346	0.258	0.197
0.30	$\widetilde{\beta}_n$	0.559	0.696	0.725
	$\widehat{\beta}_n$	0.340	0.246	0.189

Table 10.3: Mean absolute errors $MAE(n)$ of the estimators β_n and $\widehat{\beta}_n$ for the sample sizes $n \in \{50, 100, 200\}$ and true parameter $\beta_0 = 1/4$..

ε	Estimator	$MAE(50)$	$MAE(100)$	$MAE(200)$
0.00	$\widetilde{\beta}_n$	0.147	0.104	0.072
	$\widehat{\beta}_n$	0.333	0.274	0.235
0.05	$\widetilde{\beta}_n$	0.206	0.178	0.144
	$\widehat{\beta}_n$	0.327	0.266	0.223
0.10	$\widetilde{\beta}_n$	0.260	0.239	0.252
	$\widehat{\beta}_n$	0.321	0.254	0.212
0.20	$\widetilde{\beta}_n$	0.372	0.380	0.409
	$\widehat{\beta}_n$	0.317	0.233	0.185
0.30	$\widetilde{\beta}_n$	0.466	0.571	0.584
	$\widehat{\beta}_n$	0.307	0.217	0.166

Table 10.4: The same as in Table 10.3 for $\beta_0 = 1/2$.

ε	Estimator	$MAE(50)$	$MAE(100)$	$MAE(200)$
0.00	$\widetilde{\beta}_n$	0.136	0.096	0.067
	$\widehat{\beta}_n$	0.338	0.290	0.261
0.05	$\widetilde{\beta}_n$	0.172	0.144	0.108
	$\widehat{\beta}_n$	0.317	0.267	0.231
0.10	$\widetilde{\beta}_n$	0.209	0.180	0.178
	$\widehat{\beta}_n$	0.298	0.239	0.198
0.20	$\widetilde{\beta}_n$	0.284	0.263	0.263
	$\widehat{\beta}_n$	0.265	0.188	0.142
0.30	$\widetilde{\beta}_n$	0.338	0.386	0.366
	$\widehat{\beta}_n$	0.248	0.167	0.117

Table 10.5: The same as in Table 10.3 for $\beta_0 = 1$.

11 Simulations: Random regressors in bivariate Bernoulli models

In this section we compare performance of our median estimator $\widehat{\beta}_n$ with performances of two robust estimators discussed in Section 5, namely the Morgenthaler estimator $\boldsymbol{\beta}_n^{(1)}$ and the Bianco-Yohai estimator $\boldsymbol{\beta}_n^{(2)}$. The first of them is defined as solution of the equation $U_n^{(1)}(\boldsymbol{\beta}) = 0$ for $U_n^{(1)}(\boldsymbol{\beta})$ defined by (5.6) and the other is defined by (5.10) and (5.12). The constant c used in (5.10) will be equal to $-\ln 0.03 \approx \ln 33.3$ which is the value used in the simulations of Bianco and Yohai (1996). We use the same simulated data as used by Bianco and Yohai, namely the independent realizations

$$Y_i \sim (1-\varepsilon)\, Be\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right) + \varepsilon Be\left(1 - \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right), \qquad 1 \le i \le n$$

for a fixed $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02})$ and $\boldsymbol{x}_i = (1, \xi_i)$ where ξ_i are random mutually independent $N(0,1)-$distributed regressors. Four different data sources will be used, defined by the conditions

$$E\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) \in \{0.2, 0.3, 0.4, 0.5\}\,. \tag{11.1}$$

These expectations coincide with probabilities $\Pr(Y_i = 1)$ and are one-one related to the parameters $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02})$. Values of these parameters corresponding to the conditions (11.1) are given under the Tables 11.1-11.4 below.

In these tables one can find for $\varepsilon \in \{0, 0.05, 0.1, 0.2\}$ and $n \in \{200, 500, 1000\}$ the mean absolute deviations

$$MAE(n) = \frac{1}{2000} \sum_{l=1}^{1000} \left(\left|\beta_{n1}(l) - \beta_{01}\right| + \left|\beta_{n2}(l) - \beta_{02}\right|\right)$$

for 1000 simulated realizations of $(Y_1, ..., Y_n)$ and the corresponding 1000 values $\boldsymbol{\beta}_n(l) = (\beta_{n1}(l), \beta_{n2}(l))$ of four estimates $\boldsymbol{\beta}_n = (\beta_{n1}, \beta_{n2})$, namely the MLE defined by (5.7) and the above mentioned $\boldsymbol{\beta}_n^{(1)}$, $\boldsymbol{\beta}_n^{(2)}$ and $\widehat{\boldsymbol{\beta}}_n$ denoted briefly as Morg, B&Y and Median.

Using Tables 11.1-11.4 one compare performances of these four estimators measured by the corresponding mean absolute errors *MAE(n)*. We see from the first rows that if there is not contamination ($\varepsilon = 0$) then the best estimator is MLE. For light and medium contaminations ($0 < \varepsilon < 0.1$) the best is the estimator $\beta_n^{(2)}$ of Bianco and Yohai (1996). For heavier contamination ($\varepsilon \geq 0.1$) the best is our median estimator $\widehat{\beta}_n$.

The Morgenthaler's $\widehat{\beta}_n$ is outperformed by B&Y and Median in each of the present contamination model. Moreover, it faces evaluation problems when for the minimization and solving equations are used subroutines from the IMSL numerical package. This is indicated by the NEF numbers increasing with the contamination level to unacceptable levels for $\varepsilon > 0.05$. Note that NEF is the count of the simulated realizations of $(Y_1, ..., Y_n)$ for which either the estimate cannot be evaluated or it is evaluated but its absolute error exceeds 50.

ε	$(\widehat{\beta}_1, \widehat{\beta}_2)$	$MAE(200)$	NEF	$MAE(500)$	NEF	$MAE(1000)$	NEF
0	MLE	0.368	0	0.227	0	0.159	0
	Morg	0.414	0	0.246	0	0.173	0
	B&Y	0.591	1	0.297	0	0.203	0
	Median	2.003	36	0.869	2	0.515	0
0.05	MLE	0.980	0	1.027	0	1.037	0
	Morg	0.776	65	0.778	5	0.761	0
	B&Y	0.608	1	0.530	0	0.526	0
	Median	1.581	29	0.780	1	0.576	0
0.1	MLE	1.297	0	1.371	0	1.415	0
	Morg	2.106	1615	1.775	1900	4.759	2406
	B&Y	0.754	0	0.822	0	0.914	0
	Median	1.613	17	0.836	2	0.786	0
0.2	MLE	2.006	0	2.036	0	2.037	0
	Morg	-	-	-	-	-	-
	B&Y	1.892	0	1.954	0	1.954	0
	Median	1.814	14	1.704	0	1.731	0

Table 11.1: Mean absolute errors *MAE(n)* of the four estimators in the model of Bianco and Yohai with $\Pr(Y=1) = 0.2$ and the true parameters $(\beta_{01}, \beta_{02}) = (-2.82, 2.82)$. Column NEF presents the numbers of simulated observation vectors $(Y_1, ..., Y_n)$ for which the evaluation of the corresponding estimates failed. If NEF exceeds 10000, neither *MAE(n)* nor NEF is presented.

ε	$(\widehat{\beta}_1, \widehat{\beta}_2)$	$MAE(200)$	NEF	$MAE(500)$	NEF	$MAE(1000)$	NEF
0	MLE	0.389	0	0.244	0	0.167	0
	Morg	0.431	0	0.269	0	0.184	0
	B&Y	0.597	1	0.327	0	0.218	0
	Median	1.864	43	0.980	6	0.531	0
0.05	MLE	1.075	0	1.124	0	1.136	0
	Morg	0.811	73	0.771	8	0.779	0
	B&Y	0.642	0	0.549	0	0.522	0
	Median	1.562	26	0.801	0	0.613	0
0.1	MLE	1.428	0	1.494	0	1.542	0
	Morg	1.525	1507	1.428	1858	2.463	2218
	B&Y	0.789	0	0.819	0	0.924	0
	Median	1.786	37	1.022	1	0.811	0
0.2	MLE	2.109	0	2.167	0	2.172	0
	Morg	-	-	-	-	-	-
	B&Y	1.901	0	2.045	0	2.049	0
	Median	1.905	10	1.795	0	1.789	0

Table 11.2: The same as in Table 11.1 for $\Pr(Y = 1) = 0.3$ corresponding to the parameters $(\beta_{01}, \beta_{02}) = (-2.16, 3.71)$.

ε	$(\widehat{\beta}_1, \widehat{\beta}_2)$	$MAE(200)$	NEF	$MAE(500)$	NEF	$MAE(1000)$	NEF
0	MLE	0.376	0	0.241	0	0.168	0
	Morg	0.417	0	0.267	0	0.181	0
	B&Y	0.571	0	0.321	0	0.210	0
	Median	1.926	51	0.942	3	0.518	0
0.05	MLE	0.998	0	1.038	0	1.054	0
	Morg	0.681	71	0.706	10	0.720	0
	B&Y	0.596	0	0.510	0	0.489	0
	Median	1.529	32	0.835	2	0.547	0
0.1	MLE	1.320	0	1.386	0	1.426	0
	Morg	1.268	1417	1.242	1803	2.359	2160
	B&Y	0.701	0	0.762	0	0.834	0
	Median	1.583	38	0.896	0	0.755	0
0.2	MLE	1.927	0	1.990	0	1.994	0
	Morg	-	-	-	-	-	-
	B&Y	1.716	1	1.858	0	1.870	0
	Median	1.671	18	1.600	0	1.628	0

Table 11.3: The same as in Table 11.1 for $\Pr(Y = 1) = 0.4$ corresponding to the parameters $(\beta_{01}, \beta_{02}) = (-1.16, 4.20)$.

ε	$(\widehat{\beta}_1, \widehat{\beta}_2)$	$MAE(200)$	NEF(50)	$MAE(500)$	NEF	$MAE(1000)$	NEF
	MLE	0.383	0	0.234	0	0.168	0
0	Morg	0.424	0	0.260	0	0.181	0
	B&Y	0.567	1	0.314	0	0.209	0
	Median	1.593	39	0.774	0	0.480	0
	MLE	0.885	0	0.894	0	0.892	0
0.05	Morg	0.642	75	0.628	11	0.625	0
	B&Y	0.571	0	0.457	0	0.442	0
	Median	1.335	31	0.733	5	0.519	0
	MLE	1.139	0	1.174	0	1.192	0
0.1	Morg	0.948	1499	0.956	1788	1.616	2122
	B&Y	0.647	2	0.679	0	0.717	0
	Median	1.446	29	0.768	0	0.647	0
	MLE	1.613	0	1.662	0	1.654	0
0.2	Morg	-	-	-	-	-	-
	B&Y	1.420	0	1.557	0	1.555	0
	Median	1.413	9	1.359	0	1.375	0

Table 11.4: The same as in Table 11.1 for $\Pr(Y = 1) = 0.5$ corresponding to the parameters $(\beta_{01}, \beta_{02}) = (0, 4.36)$.

Reference

[1] A. Agresti (2002). *Categorical Data Analysis* (Second Edition). Wiley, New York.

[2] E. B. Andersen (1990). *The Statistical Analysis of Categorical data.* Springer-Verlag, Berlin.

[3] M. A. Arcones (2001), Asymptotic distribution of regression M-estimators. *Journal of Statistical Plannning and Inference,* **97,** 235-261.

[4] A. M. Bianco and V. J. Yohai (1996), Robust estimation in the logistic regression model. *Robust Statistics, Data Analysis, and Computer Intensive Methods* (Schloss Thurnau, 1994), 17-34. Lecture Notes in Statistics, 109, Springer, New York.

[5] X. R. Chen, L. Zhao and Y. Wu (1993), On conditions of consistency of ML_1N estimates. *Statistica Sinica,* **3,** 9-18.

[6] J. Jurečková and B. Procházka (1994), Regression quantiles and trimmed least squares estimator in nonlinear regression model. *Nonparametric Statistics,* **3,** 201-222.

[7] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti, W. A. Stahel (1986), *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[8] J. Jurečková and P. K. Sen (1996), *Robust Statistical Procedures.* Wiley, New York.

[9] K. Knight (1998), Limiting distributions for L_1 regression estimators under general conditions, *Annals Statistics,* **26**, 755-770.

[10] R. Koenker and G. Basset (1987), Regression quantiles. *Econometrica,* **46,** 33-50.

[11] F. Liese and I. Vajda (1999), M-estimators of structural parameters in pseudolinear models. *Applications of Mathematics,* **44,** 245-270.

[12] F. Liese and I. Vajda (2003), A general asymptotic theory of M-estimators, I. *Mathematical Methods of Statistics,* **12,** 454-477.

[13] F. Liese and I. Vajda (2004), A general asymptotic theory of M-estimators, II. *Mathematical Methods of Statistics,* **13,** 82-95.

[14] S. Morgenthaler (1992), Least-absolute-derivations fits for generalized linear models. *Biometrika,* **79**, 747-754.

[15] J. A. Pardo, L. Pardo and M. C. Pardo (2006), Testing in logistic regression models based on ϕ-divergences measures. *Journal of Statistical Planning and Inference*, **136,** 982-1006.

[16] D. Pollard (1991), Asymptotic for least absolute derivation regression estimators. *Econometric Theory* **7,** 186-199.

[17] G. D. Richardson and B. B. Bhattacharyya (1987), Consistent L_1-estimates in nonlinear regression for a noncompact parameter space. *Sankhya,* **49,** ser. A, 377-387.

[18] V. J. Yohai (1987), High breakdown point high efficiency robust estimates for regression. *Annals Statistics,* **15,** 692–656.

[19] S. Zwanzig (1997), *On L_1-norm Estimators in Nonlinear Regression and in Nonlinear Error-in-Variables Models.* IMS Lecture Notes 31, 101-118, Hayward.