New Estimation Results in Regression

TOMÁŠ HOBZA

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

hobza@fjfi.cvut.cz

IGOR VAJDA - Institute of Information Theory and Automation, Czech Academy of Sciences, Prague.
LEANDRO PARDO - Complutense University Madrid

Contents of the talk:

- **Introduction and basic concepts**
- **Definition of median estimator**
- **Asymptotic properties of median estimator**
- **Enhancement of median estimator**
- **Robustness of median estimator**

We are interested in estimation of the parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ in the statistical models with independent real valued observations $Y_1, \ldots, Y_n$

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

where

- ✓ $\boldsymbol{x}_i \in \mathbb{R}^d$ are vectors of explanatory variables (regressors)
- ✓ $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ is a vector of true parameters
- ✓ $\boldsymbol{x}^T \boldsymbol{\beta}$ denotes the scalar product
- ✓ $$\pi(t) = \frac{e^t}{1+e^t} \qquad \forall t \in \mathbb{R}$$
  is the logistic regression function
- ✓ $\mathcal{F} = \{F_\pi : \pi \in (0,1)\}$ is an arbitrary family of distribution functions on $\mathbb{R}$.

- the Bernoulli response functions

$$F_\pi\left(y\right) = \left(1-\pi\right) I\left(0 \leq y < 1\right) + I\left(y \geq 1\right), \qquad \pi \in (0,1)$$

with the jumps

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

define a *discrete Bernoulli model*

- the Bernoulli response functions

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

with the jumps

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

define a *discrete Bernoulli model*

In this case the problem reduces to the *classical logistic regression* with binary observations $Y_1, ..., Y_n$ taking on value

$$\begin{array}{ll} 1 & \text{with probabilities } \pi\left(\boldsymbol{x}_1^T\boldsymbol{\beta}_0\right), \ldots, \pi\left(\boldsymbol{x}_n^T\boldsymbol{\beta}_0\right) \\ 0 & \text{with probabilities } 1-\pi\left(\boldsymbol{x}_1^T\boldsymbol{\beta}_0\right), \ldots, 1-\pi\left(\boldsymbol{x}_n^T\boldsymbol{\beta}_0\right). \end{array}$$

In the classical logistic regression the MLE $\beta_n = \beta_n\left(Y_1, ..., Y_n\right)$ of β_0 minimizes the deviances (negative scores)

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

of the sample $\boldsymbol{Y}_n = \left(Y_1, ..., Y_n\right)$ where

$$d_i\left(\beta\right) = -Y_i \ln \pi_i\left(\beta\right) - \left(1 - Y_i\right) \ln\left(1 - \pi_i\left(\beta\right)\right)$$

are the deviances (negative scores) of individual observations Y_i. Thus

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

Consistency and asymptotic normality with the variances at the *Cramér-Rao lower bound* can be proved.

However, this estimator is too sensitive to outliers among the data $(\boldsymbol{x}_1, Y_1), ..., (\boldsymbol{x}_n, Y_n)$.
Typical outliers are

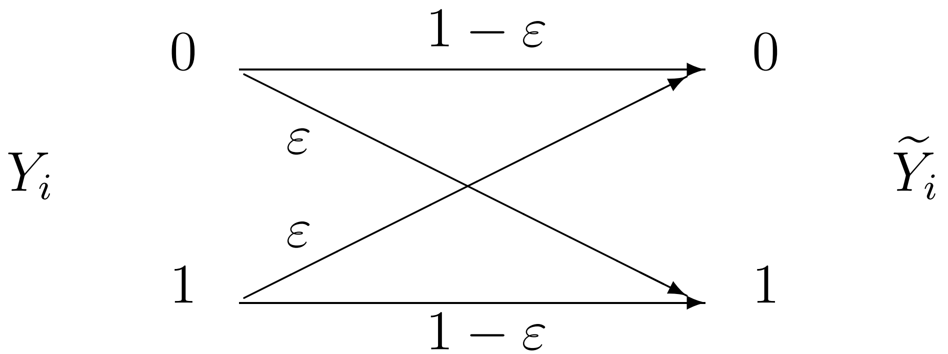
$$Y_i = 0 \qquad \text{when} \qquad \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) \approx 1$$

or

$$Y_i = 1 \qquad \text{when} \qquad \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) \approx 0.$$

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 $\mathrm{BSC}(\varepsilon)$ with

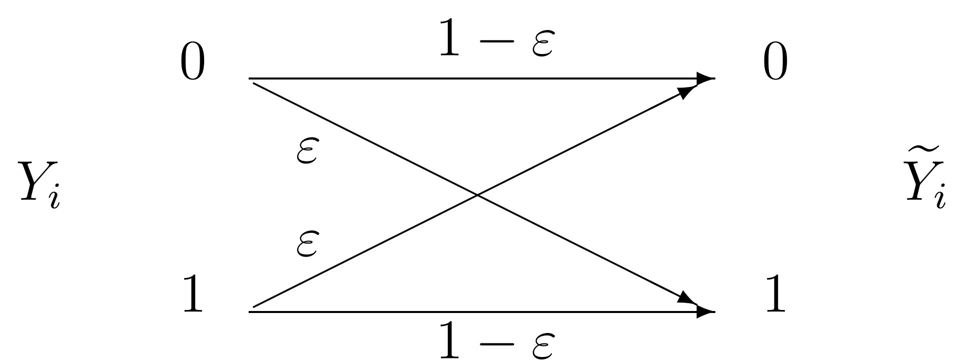
- ✓ independent inputs Y_i
- ✓ additive (mod 2) independent noise $W_i \sim Be\left(\varepsilon\right)$
- ✓ independent outputs $\widetilde{Y}_i = Y_i + W_i$ (mod 2).



Binary Symmetric channel $\mathrm{BSC}(\varepsilon)$.

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 $\mathrm{BSC}(\varepsilon)$ with

- ✓ independent inputs Y_i
- ✓ additive (mod 2) independent noise $W_i \sim Be\left(\varepsilon\right)$
- ✓ independent outputs $\widetilde{Y}_i = Y_i + W_i$ (mod 2).



Binary Symmetric channel $\mathrm{BSC}(\varepsilon)$.

Then $\left(\boldsymbol{x}_1, \widetilde{Y}_1\right), \ldots, \left(\boldsymbol{x}_n, \widetilde{Y}_n\right)$ contain responses $\widetilde{Y}_i$ generated by the stochastic mixture

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

Previous authors replaced the deviances $d_i\left(\beta\right)$ by appropriate functions

$$\varrho\left(d_i\left(\beta\right)\right)$$

of deviances, or even by more general expressions

$$\phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T\beta\right)\right).$$

Previous authors replaced the deviances $d_i(\beta)$ by appropriate functions

$$\varrho\left(d_i\left(\beta\right)\right)$$

of deviances, or even by more general expressions

$$\phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T\beta\right)\right).$$

This lead to M-estimators β_n of the type

$$\beta_n = \arg\min \sum_{i=1}^{n} \varrho\left(d_i\left(\beta\right)\right)$$

and

$$\beta_n = \arg\min \sum_{i=1}^{n} \phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T\beta\right)\right)$$

for $\varrho : (0,\infty) \to \mathbb{R}$ and $\phi : (0,\infty) \times (0,1) \to \mathbb{R}$.

We propose a new robust M-estimator of the logistic regression parameter $\beta_0 \in \mathbb{R}^d$ obtained by application of the classical robust L_1-method.

Definition 1. *The median estimator* $\widehat{\beta}_n$ of the true parameter $\boldsymbol{\beta}_0$ in the general logistic regression model is defined by the formula

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

where $m(\pi)$ is for every $\pi \in (0,1)$ the median

$$m(\pi) = F_\pi^{-1}(1/2) = \inf\left\{y \in \mathbb{R} : F_\pi(y) \geq 1/2\right\}.$$

Condition of applicability: sensitivity of the median function $m(\pi)$ to the change of parameter $\pi \in (0,1)$ (strict monotonicity of $m(\pi)$ on $(0,1)$).

Condition of applicability: sensitivity of the median function $m(\pi)$ to the change of parameter $\pi \in (0,1)$ (strict monotonicity of $m(\pi)$ on $(0,1)$).

- in the discrete Bernoulli model

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

Condition of applicability: sensitivity of the median function $m(\pi)$ to the change of parameter $\pi \in (0,1)$ (strict monotonicity of $m(\pi)$ on $(0,1)$).

- in the discrete Bernoulli model

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

- in the discrete geometric model

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

Definition 2. The *standard modification* of a discrete logistic regression model is the continuous logistic regression model with the observations

$$Z_i = Y_i + U_i, \qquad 1 \leq i \leq n,$$

where U_i are an independent noise random variables uniformly distributed on the interval $(0, 1)$.

Definition 2. The *standard modification* of a discrete logistic regression model is the continuous logistic regression model with the observations

$$Z_i = Y_i + U_i, \qquad 1 \le i \le n,$$

where U_i are an independent noise random variables uniformly distributed on the interval $(0, 1)$.

The transformation introduced in the previous definition is statistically sufficient since

$$Y_i = [Z_i] \text{ a.s.}, \qquad 1 \le i \le n.$$

The median functions $m(p) = F_p^{-1}(1/2)$ of the transformed observations Z_i are already one-one on the interval $(0, 1)$.

E.g. the standardly modified Bernoulli model is the continuous model with the response function

$$F_p(z) = (1-p)\ z\, I(0 < z \leq 1) + [1 - p + p(z-1)]\, I(1 < z \leq 2)\,.$$

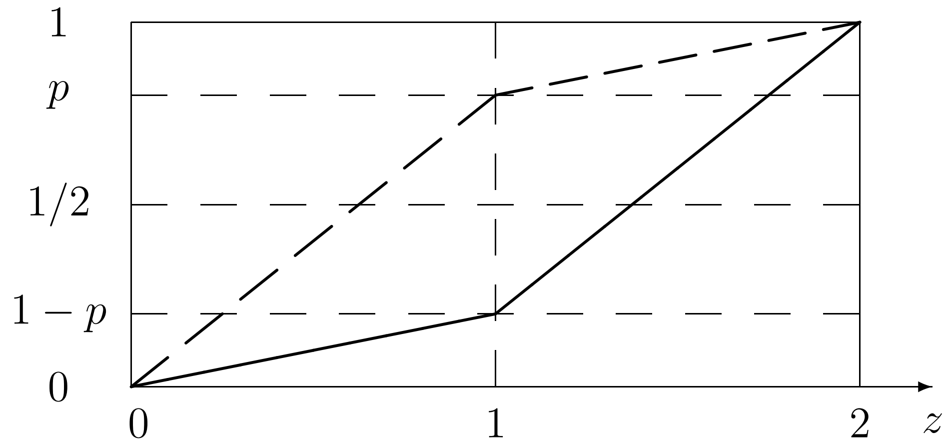


Figure 1. $F_p(z)$ full line, $F_{1-p}(z)$ dashed line. The median function has the form

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

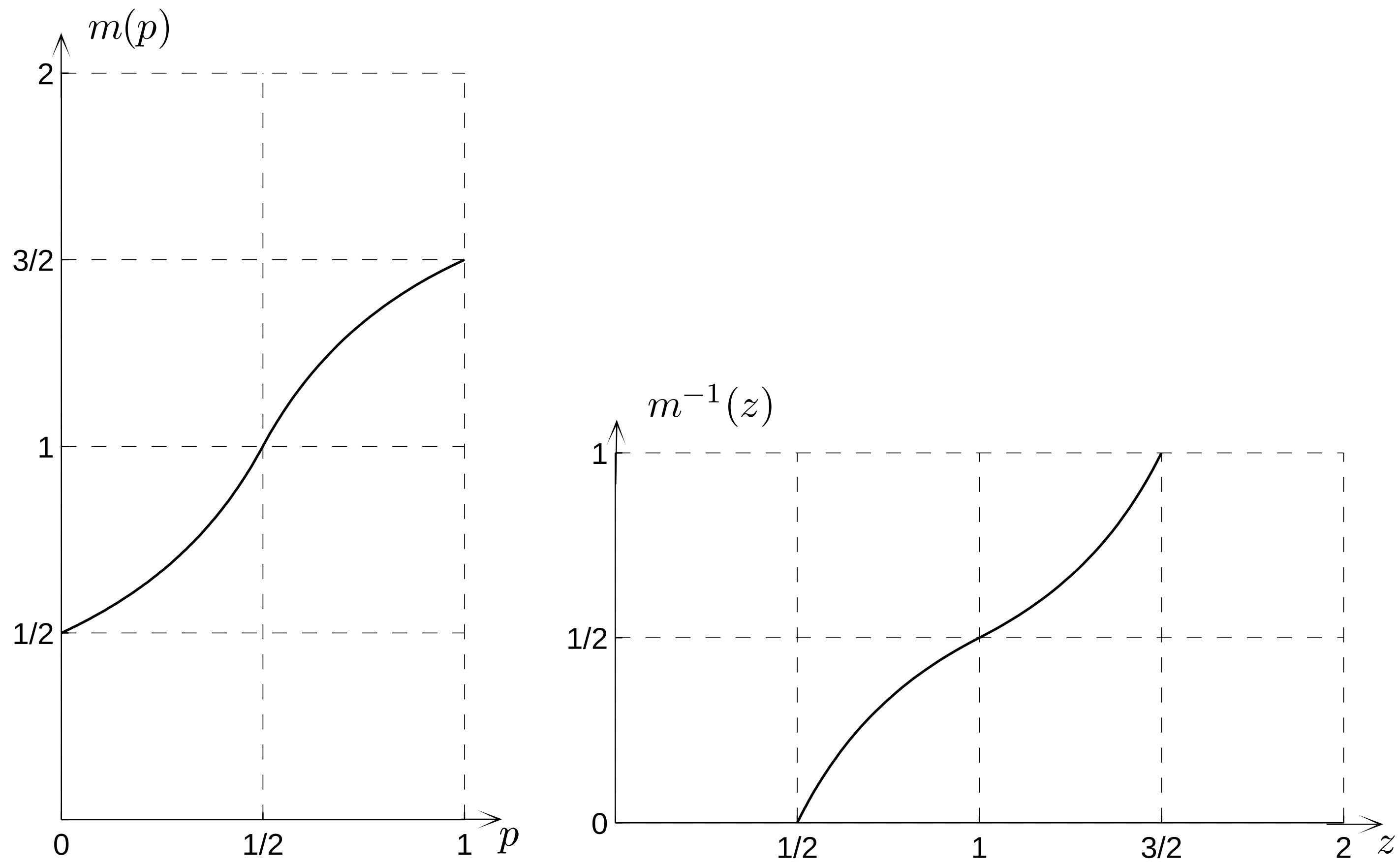


Figure 2. Median function $m(p)$ and its inverse $m^{-1}(z)$.

Theorem 1. If the regressors of the model under consideration satisfy **(c1)**, **(c2)** then the median estimator $\widehat{\beta}_n$ applied to the standard modification of the model consistently estimates the model parameters β_0.

Theorem 1. If the regressors of the model under consideration satisfy **(c1)**, **(c2)** then the median estimator $\widehat{\beta}_n$ applied to the standard modification of the model consistently estimates the model parameters β_0.

Theorem 2. Let the regressors of the model under consideration satisfy (**c1**), (**c2**). If the matrix $\mathcal{Q}$ is positive definite then the median estimator $\widehat{\beta}_n$ of the model parameters β_0 is asymptotically normal in the sense that

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} N\left(\mathbf{0},\ \mathcal{Q}^{-1}\Sigma\mathcal{Q}^{-1}\right).$$

Let us study the simple special case of dimension $d = 1$ with all univariate regressors identical,

$$x_1 = x_2 = ... = 1 \in \mathbb{R}.$$

Thus we estimate a parameter $\beta_0 \in \mathbb{R}$ using the independent observations

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

where

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

and F_π is the geometric response distribution function.

In the simulation experiment we suppose that the geometric distribution function $F_{\pi(\beta_0)}(y)$ is replaced by the mixture

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

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

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

at $k = 1, 2, ...,$.

In the simulation experiment we suppose that the geometric distribution function $F_{\pi(\beta_0)}(y)$ is replaced by the mixture

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

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

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

at $k = 1, 2, ...,$.

In tables we present for $\beta_0 \in \{1/4, 1/2\}$ 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 of the median estimates $\widehat{\beta}_n$.

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

Table 1: 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 $\beta_0 = 1/4$.

By choosing a simple concrete L_1-estimator we will illustrate a method for suppression of the inefficiency.

Let us suppose the median estimator

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

of the Bernoulli parameter $p_0 \in (0,1)$ based on the smoothed versions $Z_i = Y_i + U_i$ of the original discrete observations $Y_i \sim Be(p_0)$.

The MLE

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

of p_0 is asymptotically efficient in the sense

$$\sqrt{n}\,(p_n - p_0) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, 1/\mathcal{J}(p_0)) = N\left(0, p_0\left(1 - p_0\right)\right).$$

On the other hand the median estimator $\widehat{p}_n$ is asymptotically normal in the sense

$$\sqrt{n}\,(\widehat{p}_n - p_0) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N(0, [p_0 \vee (1-p_0)]^2)$$

where

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

On the other hand the median estimator $\widehat{p}_n$ is asymptotically normal in the sense

$$\sqrt{n}\left(\widehat{p}_n - p_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N(0, \left[p_0 \vee (1-p_0)\right]^2)$$

where

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

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

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

where U_{ij} are mutually and also on $Y_1, ..., Y_n$ independent $U\left(0,1\right)$-distributed random variables.

On the other hand the median estimator $\widehat{p}_n$ is asymptotically normal in the sense

$$\sqrt{n}\,(\widehat{p}_n - p_0) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} N(0, [p_0 \vee (1-p_0)]^2)$$

where

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

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

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

where U_{ij} are mutually and also on $Y_1, ..., Y_n$ independent $U\,(0,1)$-distributed random variables.

We define the k-*enhanced median estimator*

$$\widehat{p}_{n*k} = \arg\min_p \sum_{j=1}^{k}\sum_{i=1}^{n} |Z_{ij} - m\,(p)| = m^{-1}\left(Z_{(nk/2)}\right)$$

of the Bernoulli parameter $p_0 \in (0,1)\,.$

Theorem 3. The k-enhanced median estimator $\widehat{p}_{n*k}$ is asymptotically optimal in the sense that for each $n \geq 1$

$$\sigma^2\left(\widehat{p}_{n*k}\right) \underset{k\to\infty}{\longrightarrow} \sigma^2\left(p_n\right) \quad \text{and} \quad \mathsf{e}\left(\widehat{p}_{n*k}\right) \underset{k\to\infty}{\longrightarrow} \mathsf{e}\left(p_n\right)$$

where

$$\sigma^2\left(\widehat{p}_{n*k}\right) = E\left(\widehat{p}_{n*k} - p_0\right)^2 \qquad \mathsf{e}\left(\widehat{p}_{n*k}\right) = E\left|\widehat{p}_{n*k} - p_0\right|$$

are the expected squared and absolute errors of the estimators $\widehat{p}_{n*k}$ and

$$\sigma^2\left(p_n\right) = E\left(p_n - p_0\right)^2 = \frac{p_0\left(1-p_0\right)}{n}, \quad \mathsf{e}\left(p_n\right) = E\left|p_n - p_0\right|$$

are the expected squared and absolute errors of the classical MLE p_n.

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

Table 2: Analysis of the proposed smoothing method in the Bernoulli model $Be(p_0), p_0 = 0.2$.

Definition 3 . For every $k \geq 1$ we define the k-*enhanced median estimator* $\widehat{\beta}_{n*k}$ of the parameters of logistic regression β_0 by the condition

$$\widehat{\beta}_{n*k} = \arg\min_{\beta} \sum_{j=1}^{k} \sum_{i=1}^{n} \left| Z_{ij} - m\left(\pi\left(\boldsymbol{x}_i^T \beta\right)\right) \right|$$

where $Z_{ij} = Y_i + U_{ij}$.

Compared estimates:

- $\widehat{\boldsymbol{\beta}}_n$ (Median)
- $\widehat{\boldsymbol{\beta}}_{n*k}$ (k-enhanced Median)
- $\boldsymbol{\beta}_n^{(1)}$ (Morg)
- $\boldsymbol{\beta}_n^{(2)}$ (B&Y)
- $\boldsymbol{\beta}_n$ (MLE)

Compared estimates:

- $\widehat{\boldsymbol{\beta}}_n$ (Median)
- $\widehat{\boldsymbol{\beta}}_{n*k}$ (k-enhanced Median)
- $\boldsymbol{\beta}_n^{(1)}$ (Morg)
- $\boldsymbol{\beta}_n^{(2)}$ (B&Y)
- $\boldsymbol{\beta}_n$ (MLE)

Estimates are evaluated from the same simulated data

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

for a fixed $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02})$ and $\boldsymbol{x}_i = (1, \xi_i)$, where ξ_i i.i.d $N(0,1)-$, and the related smoothed data vectors or matrices

$$\boldsymbol{Z}_n = (Y_1 + U_1, \ldots, Y_n + U_n) \quad \text{or} \quad \boldsymbol{Z}_{k,n} = (Y_1 + U_{11}, \ldots, Y_n + U_{kn}).$$

The same specifications are used as in Bianco and Yohai (1996)

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

	$\Pr(Y_i = 1) = 0.2$	$\Pr(Y_i = 1) = 0.3$	$\Pr(Y_i = 1) = 0.4$	$\Pr(Y_i = 1) = 0.5$
β_{01}	-2.82	-2.16	-1.16	0
β_{02}	2.82	3.71	4.20	4.36

Simulation experiment 2

The same specifications are used as in Bianco and Yohai (1996)

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

	$\Pr(Y_i = 1) = 0.2$	$\Pr(Y_i = 1) = 0.3$	$\Pr(Y_i = 1) = 0.4$	$\Pr(Y_i = 1) = 0.5$
β_{01}	-2.82	-2.16	-1.16	0
β_{02}	2.82	3.71	4.20	4.36

We selected

$$\varepsilon \in \{0, 0.05, 0.1, 0.15, 0.2\}$$

and

$$n \in \{50, 100, 500, 1000\}$$

and computed the mean absolute errors

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

Simulation experiment 2

		$n = 50$	$n = 100$	$n = 500$	$n = 1000$
ε	$\widetilde{\beta}_n$	MAE	MAE	MAE	MAE
0	MLE	**0.862**	**0.579**	**0.235**	**0.159**
	Morg	0.951	0.639	0.256	0.176
	BY	1.527	0.942	0.312	0.212
	Med	3.267	2.554	0.925	0.497
	5-Med	2.786	1.637	0.521	0.327
	10-Med	2.496	1.730	0.464	0.306
0.05	MLE	1.067	0.979	1.047	1.048
	Morg	**1.062**	**0.834**	0.769	0.771
	BY	1.482	0.915	0.553	0.527
	Med	2.715	2.524	0.848	0.587
	5-Med	2.093	1.544	0.579	0.487
	10-Med	2.098	1.444	**0.547**	**0.475**
0.1	MLE	1.425	1.473	1.510	1.525
	Morg	**1.347**	1.290	1.320	1.338
	BY	1.446	**1.220**	1.072	1.100
	Med	2.701	2.293	0.977	**0.921**
	5-Med	1.907	1.589	0.920	0.931
	10-Med	1.819	1.569	**0.909**	0.934
0.2	MLE	1.989	2.011	2.029	2.034
	Morg	1.939	1.956	1.975	1.979
	BY	**1.897**	1.891	1.941	1.948
	Med	2.381	1.988	**1.683**	**1.710**
	5-Med	2.029	1.879	1.721	1.732
	10-Med	2.042	**1.735**	1.725	1.734

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.

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

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

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

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

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

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.