# Robustnost mediánového odhadu v zobecněném modelu logistické regrese


TOMÁŠ HOBZA

FJFI ČVUT, ÚTIA AV ČR

hobza@km1.fjfi.cvut.cz

IGOR VAJDA - ÚTIA AV ČR
LEANDRO PARDO - Complutense University Madrid

# Contents of the talk:

- **Introduction and basic concepts**
- **Definition of median estimator**
- **Asymptotic properties 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 $Y_1, \ldots, Y_n$

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

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 $\mathbb{R}$.

*Discrete models:*

$F_\pi(y)$ – right-continuous distribution functions with jumps

$$p_\pi(k) = F_\pi(k) - F_\pi(k-0), \qquad k = 0, 1, \ldots$$

summing up to 1

*Continuous models:*

$F_\pi(y)$ – continuous piecewise differentiable distribution functions with densities

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

- 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*

- 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

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

- The discrete geometric response functions

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

with the jumps

$$p_\pi(k) = (1-\pi)\,\pi^k, \quad k = 0, 1, \ldots$$

define a *discrete geometric model* .

- The discrete geometric response functions

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

with the jumps

$$p_\pi(k) = (1-\pi)\,\pi^k, \quad k = 0, 1, \ldots$$

define a *discrete geometric model* .

- the exponential response functions

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

with the densities

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

define a *continuous exponential model* .

**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

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

Member of the class of so-called *least absolute deviation estimators* (or briefly $L_1-$estimators) defined by

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

where $\mu : \Theta \rightarrow \mathbb{R}$ and $u : \mathbb{R} \rightarrow \Theta$ are given functions.

See e.g. Richardson and Bhattacharyya (1987), Yohai (1987), Pollard (1991), Morgenthaler (1992), Arcones (2001), Liese and Vajda (2003, 2004).

*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

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

where $W_i$ 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

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

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

The transformation introduced in the previous definition is statistically sufficient since

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

The median functions $\widetilde{m}\,(\pi) = \widetilde{F}_\pi^{-1}\,(1/2)$ of the transformed observations $\widetilde{Y}_i$ are already one-one on the interval $(0,1)$.

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

$$F_\pi\left(y\right) = \left(1-\pi\right)\ y\, I\left(0 < y \leq 1\right) + \left[1-\pi+\pi\left(y-1\right)\right] I\left(1 < y \leq 2\right).$$

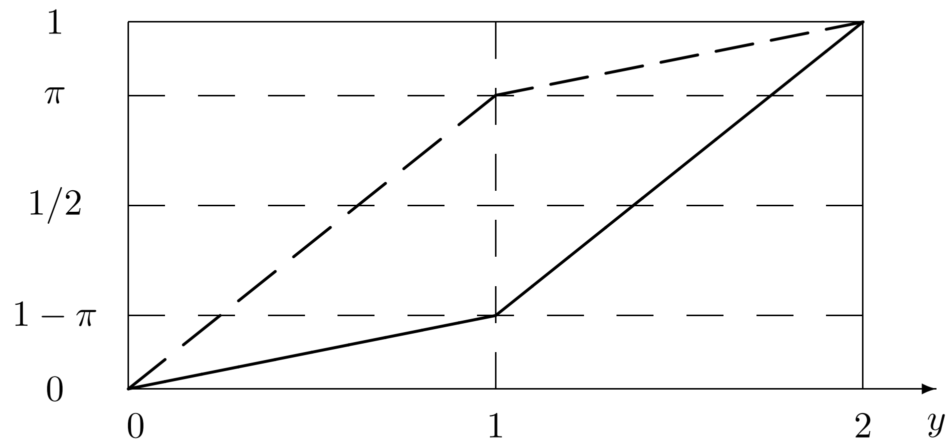



**Figure 1**. $F_\pi\left(y\right)$ full line, $F_{1-\pi}\left(y\right)$ dashed line.

The median function has the form

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

# Assumptions:

**(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

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

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

**(c2)** The smallest eigenvalue of the matrix

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

is positive. Hence for every $\beta \in \mathbb{R}^d$ different from $\beta_0$

$$Q\left(\boldsymbol{x} \in \mathcal{X} : \boldsymbol{x}^T\left(\beta - \beta_0\right) \neq 0\right) > 0.$$

**(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

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

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

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

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

**(c5)** Distributions functions $F_\pi$, $\pi \in (0,1)$ 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

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

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

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

**Theorem 1.** If a continuous logistic regression model or a standardly modified discrete regression model satisfies **(c1)-(c6)** then the median estimator $\widehat{\beta}_n$ of true parameter $\beta_0 \in \mathbb{R}^d$ is consistent if for every $0 < \pi_1 < \pi_2 < 1$ there exists $a > 0$ such that the condition

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

holds.

**(c7)** The quantile function $m(\pi)$ is differentiable on $(0,1)$ and the derivative $m'(\pi)$ is locally Lipschitz.

**(c8)** The densities $f_\pi$ satisfy for every $0 < \pi_1 < \pi_2 < 1$ the condition

$$\lim_{y \to 0} \sup_{\pi_1 \le \pi \le \pi_2} \left| f_\pi\left(m(\pi) + y\right) - f_\pi\left(m(\pi)\right)\right| = 0.$$

**(c7)** The quantile function $m(\pi)$ is differentiable on $(0,1)$ and the derivative $m'(\pi)$ is locally Lipschitz.

**(c8)** The densities $f_\pi$ satisfy for every $0 < \pi_1 < \pi_2 < 1$ the condition

$$\lim_{y \to 0} \sup_{\pi_1 \le \pi \le \pi_2} \left| f_\pi\left(m(\pi) + y\right) - f_\pi\left(m(\pi)\right)\right| = 0.$$

**Theorem 2.** Let a logistic regression model satisfy the conditions of Theorem 1 and **(c7)**, **(c8)**. Then the median estimator $\widehat{\beta}_n$ of true parameter $\beta_0 \in \mathbb{R}^d$ is asymptotically normal in the sense that

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

## Robustness of median estimator

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.

**There is a hope:** the median estimator for the general logistic regression is robust too.

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.

**There is a hope:** the median estimator for the general logistic regression is robust too.

We will demonstrate, by means of simulation experiment, the robustness of the median estimator and to compare it with the robustness of some well known estimators tailor-made for robust estimation in logistic regression. For this comparison we selected:

- the $L_1$-estimator of Morgenthaler (1992)
- the $M$-estimator of Bianco and Yohai (1996).

Morgenthaler (1992) started with the weighted $L_1$-estimator

$$\boldsymbol{\beta}_n^{(0)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \frac{\left| Y_i - \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right|}{\sqrt{\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\left(1 - \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)}},$$

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

Morgenthaler (1992) started with the weighted $L_1$-estimator

$$\boldsymbol{\beta}_n^{(0)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \frac{\left|Y_i - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right|}{\sqrt{\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\left(1-\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)}},$$

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

A slight modification $\boldsymbol{\beta}_n^{(1)}$ which solves the equation

$$U_n^{(1)}(\boldsymbol{\beta}) = U_n^{(0)}(\boldsymbol{\beta}) - E_\beta U_n^{(0)}(\boldsymbol{\beta}) = 0$$

were proposed.

Explicit formula for $U_n^{(1)}$ can be found

$$U_n^{(1)}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sqrt{\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\left(1-\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)}\left(Y_i - \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)\boldsymbol{x}_i.$$

Bianco and Yohai (1996) started with the MLE

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

where

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

and $\mu_i\left(\boldsymbol{\beta}\right) = E_{\boldsymbol{\beta}} Y_i = \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)$. They proved consistency and asymptotic normality with the variances at the *Cramér-Rao lower bound*.

Bianco and Yohai (1996) started with the MLE

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

where

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

and $\mu_i\left(\boldsymbol{\beta}\right) = E_{\boldsymbol{\beta}} Y_i = \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)$. They proved consistency and asymptotic normality with the variances at the *Cramér-Rao lower bound*.

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

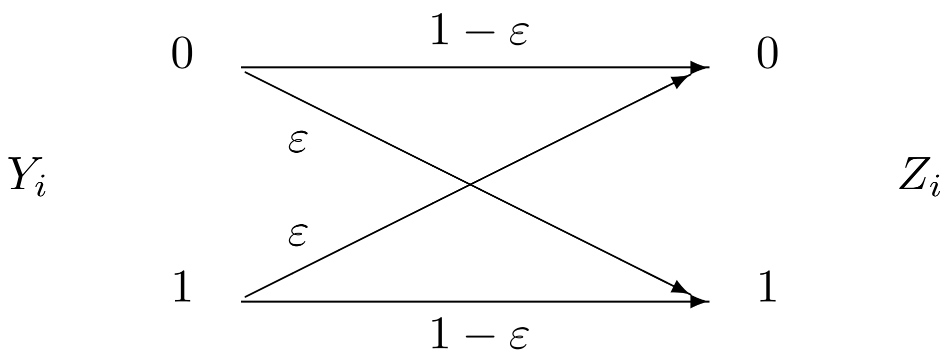
$$Y_i = 0 \qquad \text{when} \qquad \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) \approx 1$$

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

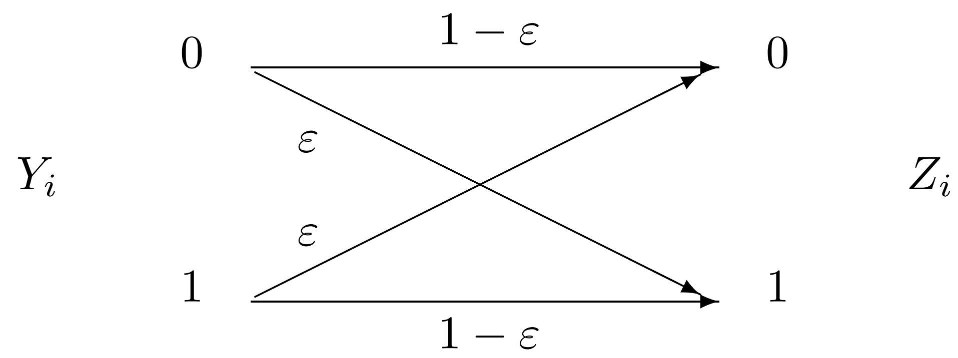
- ✓ independent inputs $Y_i$
- ✓ additive (mod 2) independent noise $W_i \sim Be\left(\varepsilon\right)$
- ✓ independent outputs $Z_i = Y_i + W_i$ (mod 2).




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 $Z_i = Y_i + W_i$ (mod 2).




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

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

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

To restrict the influence of the outliers Bianco and Yohai proposed the modified estimator

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

where

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

is a hard-limiter and terms involving the function

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

represent a bias correction.

## Simulation experiment 1

We 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 $F_\pi$ is the geometric response distribution function.

In this case:

- the MLE $\beta_n = \ln\left(\overline{Y}_n\right)$ of $\beta_0$ satisfies the relation

$$\sqrt{n}\left(\beta_n - \beta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N\left(0, 1/\pi\left(\beta_0\right)\right)$$

In this case:

- the MLE $\beta_n = \ln\left(\overline{Y}_n\right)$ of $\beta_0$ satisfies the relation

$$\sqrt{n}\left(\beta_n - \beta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} N\left(0, 1/\pi\left(\beta_0\right)\right)$$

- the median estimator $\widehat{\beta}_n$ of $\beta_0$ is given by

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

where

$$\widehat{\pi}_n = m^{-1}\left(Y_{(n/2)}\right)$$

and

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

# Simulation experiment 1

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 $k = 1, 2, \ldots,$.

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

| $\varepsilon$ | Estimator | *MAE*(50) | *MAE*(100) | *MAE*(200) |
|---|---|---|---|---|
| 0.00 | $\beta_n$ | 0.156 | 0.109 | 0.076 |
| | $\widehat{\beta}_n$ | 0.343 | 0.271 | 0.214 |
| 0.05 | $\beta_n$ | 0.227 | 0.202 | 0.170 |
| | $\widehat{\beta}_n$ | 0.342 | 0.268 | 0.208 |
| 0.10 | $\beta_n$ | 0.298 | 0.283 | 0.306 |
| | $\widehat{\beta}_n$ | 0.344 | 0.266 | 0.205 |
| 0.20 | $\beta_n$ | 0.439 | 0.466 | 0.510 |
| | $\widehat{\beta}_n$ | 0.346 | 0.258 | 0.197 |
| 0.30 | $\beta_n$ | 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 $\beta_n$ and $\widehat{\beta}_n$ for the sample sizes $n \in \{50, 100, 200\}$ and true parameter $\beta_0 = 1/4$.

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

**Table 1**: Mean absolute errors *MAE*$(n)$ of the estimators $\beta_n$ and $\widehat{\beta}_n$ for the sample sizes $n \in \{50, 100, 200\}$ and true parameter $\beta_0 = 1/2$.

In the simulation experiment we study particularly the *discrete Bernoulli models* of the small dimension $d = 2$.

- ✓ they are typical and the small dimensions are simpler and sufficient to provide an insight into the general properties of estimators.
- ✓ the considered logistic regression models are the same as used in Bianco and Yohai (1996) for demonstration of robustness of their estimator.

**Compared estimates:**

- $\widehat{\beta}_n$ (Median)
- $\beta_n^{(1)}$ (Morg)
- $\beta_n^{(2)}$ (B&Y)
- $\beta_n$ (MLE)

**Compared estimates:**

- $\widehat{\boldsymbol{\beta}}_n$ (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 $\xi_i$ i.i.d $N(0,1)-$ distributed regressors.

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

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

# Simulation experiment 2

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

# Simulation experiment 2

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

We selected

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

and

$$n \in \{400, 800, 1600\}$$

and computed the mean absolute errors

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

for 1000 simulated realizations of $(Y_1, \ldots, Y_n)$.

| $\varepsilon$ | $(\widehat{\beta}_1, \widehat{\beta}_2)$ | *MAE*(400) | NEF | *MAE*(800) | NEF | *MAE*(1600) | NEF |
|---|---|---|---|---|---|---|---|
| 0 | MLE | 0.246 | 0 | 0.176 | 0 | 0.126 | 0 |
| | Morg | 0.270 | 0 | 0.193 | 0 | 0.135 | 0 |
| | B&Y | 0.340 | 0 | 0.232 | 0 | 0.161 | 0 |
| | Median | 1.128 | 7 | 0.678 | 0 | 0.378 | 0 |
| 0.05 | MLE | 1.011 | 0 | 1.037 | 0 | 1.048 | 0 |
| | Morg | 0.731 | 16 | 0.760 | 3 | 0.768 | 0 |
| | B&Y | 0.525 | 0 | 0.520 | 0 | 0.528 | 0 |
| | Median | 1.118 | 0 | 0.634 | 0 | 0.514 | 0 |
| 0.1 | MLE | 1.368 | 0 | 1.414 | 0 | 1.441 | 0 |
| | Morg | 3.518 | 1826 | 4.836 | 2568 | 5.114 | 3391 |
| | B&Y | 0.796 | 0 | 0.888 | 0 | 0.942 | 0 |
| | Median | 1.070 | 1 | 0.789 | 0 | 0.792 | 0 |
| 0.15 | MLE | 1.814 | 0 | 1.819 | 0 | 1.819 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 1.590 | 0 | 1.608 | 0 | 1.609 | 0 |
| | Median | 1.317 | 2 | 1.336 | 0 | 1.369 | 0 |
| 0.2 | MLE | 2.029 | 0 | 2.036 | 0 | 2.036 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 1.940 | 0 | 1.951 | 0 | 1.953 | 0 |
| | Median | 1.643 | 0 | 1.716 | 0 | 1.737 | 0 |

*MAE(n)* in the model with $Pr(Y = 1) = 0.2$

| $\varepsilon$ | $(\widehat{\beta}_1, \widehat{\beta}_2)$ | $MAE(400)$ | NEF | $MAE(800)$ | NEF | $MAE(1600)$ | NEF |
|---|---|---|---|---|---|---|---|
| 0 | MLE | 0.263 | 0 | 0.194 | 0 | 0.136 | 0 |
| | Morg | 0.295 | 0 | 0.211 | 0 | 0.149 | 0 |
| | B&Y | 0.365 | 0 | 0.249 | 0 | 0.173 | 0 |
| | Median | 1.105 | 7 | 0.671 | 0 | 0.404 | 0 |
| 0.05 | MLE | 1.102 | 0 | 1.137 | 0 | 1.145 | 0 |
| | Morg | 0.753 | 17 | 0.791 | 2 | 0.792 | 0 |
| | B&Y | 0.553 | 0 | 0.543 | 0 | 0.537 | 0 |
| | Median | 1.056 | 1 | 0.629 | 0 | 0.526 | 0 |
| 0.1 | MLE | 1.488 | 0 | 1.535 | 0 | 1.562 | 0 |
| | Morg | 2.470 | 1622 | 2.526 | 2086 | 3.153 | 3024 |
| | B&Y | 0.804 | 0 | 0.898 | 0 | 0.948 | 0 |
| | Median | 0.959 | 5 | 0.835 | 0 | 0.809 | 0 |
| 0.15 | MLE | 1.943 | 0 | 1.948 | 0 | 1.948 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 1.637 | 0 | 1.654 | 0 | 1.657 | 0 |
| | Median | 1.395 | 1 | 1.364 | 0 | 1.402 | 0 |
| 0.2 | MLE | 2.161 | 0 | 2.168 | 0 | 2.168 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 2.031 | 0 | 2.043 | 0 | 2.046 | 0 |
| | Median | 1.810 | 0 | 1.769 | 0 | 1.799 | 0 |

*MAE(n)* in the model with $Pr(Y = 1) = 0.3$

| $\varepsilon$ | $(\widehat{\beta}_1, \widehat{\beta}_2)$ | $MAE(400)$ | NEF | $MAE(800)$ | NEF | $MAE(1600)$ | NEF |
|---|---|---|---|---|---|---|---|
| 0 | MLE | 0.264 | 0 | 0.187 | 0 | 0.137 | 0 |
| | Morg | 0.288 | 0 | 0.205 | 0 | 0.149 | 0 |
| | B&Y | 0.346 | 0 | 0.244 | 0 | 0.171 | 0 |
| | Median | 1.169 | 5 | 0.607 | 0 | 0.396 | 0 |
| 0.05 | MLE | 1.018 | 0 | 1.049 | 0 | 1.056 | 0 |
| | Morg | 0.686 | 19 | 0.718 | 1 | 0.718 | 0 |
| | B&Y | 0.510 | 0 | 0.492 | 0 | 0.484 | 0 |
| | Median | 1.028 | 4 | 0.614 | 0 | 0.489 | 0 |
| 0.1 | MLE | 1.378 | 0 | 1.413 | 0 | 1.439 | 0 |
| | Morg | 2.359 | 1643 | 2.347 | 2009 | 2.378 | 2845 |
| | B&Y | 0.742 | 0 | 0.814 | 0 | 0.871 | 0 |
| | Median | 0.888 | 2 | 0.736 | 0 | 0.729 | 0 |
| 0.15 | MLE | 1.786 | 0 | 1.793 | 0 | 1.793 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 1.481 | 0 | 1.507 | 0 | 1.511 | 0 |
| | Median | 1.275 | 0 | 1.237 | 0 | 1.281 | 0 |
| 0.2 | MLE | 1.985 | 0 | 1.993 | 0 | 1.993 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 1.856 | 0 | 1.869 | 0 | 1.870 | 0 |
| | Median | 1.563 | 0 | 1.622 | 0 | 1.638 | 0 |

*MAE(n)* in the model with $Pr(Y = 1) = 0.4$

| $\varepsilon$ | $(\widehat{\beta}_1, \widehat{\beta}_2)$ | *MAE*(400) | NEF | *MAE*(800) | NEF | *MAE*(1600) | NEF |
|---|---|---|---|---|---|---|---|
| 0 | MLE | 0.262 | 0 | 0.182 | 0 | 0.128 | 0 |
| | Morg | 0.283 | 0 | 0.196 | 0 | 0.139 | 0 |
| | B&Y | 0.342 | 0 | 0.229 | 0 | 0.159 | 0 |
| | Median | 0.963 | 4 | 0.559 | 0 | 0.341 | 0 |
| 0.05 | MLE | 0.886 | 0 | 0.893 | 0 | 0.887 | 0 |
| | Morg | 0.626 | 14 | 0.626 | 0 | 0.614 | 0 |
| | B&Y | 0.477 | 0 | 0.445 | 0 | 0.427 | 0 |
| | Median | 0.882 | 6 | 0.537 | 1 | 0.439 | 0 |
| 0.1 | MLE | 1.173 | 0 | 1.188 | 0 | 1.199 | 0 |
| | Morg | 1.684 | 1762 | 1.695 | 2004 | 1.579 | 2686 |
| | B&Y | 0.664 | 0 | 0.709 | 0 | 0.736 | 0 |
| | Median | 0.887 | 3 | 0.673 | 0 | 0.644 | 0 |
| 0.15 | MLE | 1.502 | 0 | 1.492 | 0 | 1.484 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 1.263 | 0 | 1.262 | 0 | 1.258 | 0 |
| | Median | 1.129 | 0 | 1.070 | 0 | 1.084 | 0 |
| 0.2 | MLE | 1.662 | 0 | 1.654 | 0 | 1.646 | 0 |
| | Morg | - | - | - | - | - | - |
| | B&Y | 1.558 | 0 | 1.552 | 0 | 1.545 | 0 |
| | Median | 1.337 | 4 | 1.357 | 0 | 1.368 | 0 |

*MAE(n)* in the model with $Pr(Y = 1) = 0.5$

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.