

Asymptotic normality of modified LS estimator for mixture of nonlinear regressions

Vitalii Miroshnichenko, Rostyslav Maiboroda*

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

vitaliy.miroshnichenko@gmail.com (V. Miroshnichenko), mre@univ.kiev.ua (R. Maiboroda)

Received: 21 August 2020, Revised: 8 November 2020, Accepted: 16 November 2020,
Published online: 8 December 2020

Abstract We consider a mixture with varying concentrations in which each component is described by a nonlinear regression model. A modified least squares estimator is used to estimate the regressions parameters. Asymptotic normality of the derived estimators is demonstrated. This result is applied to confidence sets construction. Performance of the confidence sets is assessed by simulations.

Keywords Finite mixture model, nonlinear regression, mixture with varying concentrations, semiparametric estimation, confidence ellipsoid

2010 MSC 62J05, 62G20

1 Introduction

Nonlinear regression models are widely used in analysis of statistical data [14, 16]. In many applications the observed data are derived from a mixture of components with different dependencies between the variables in different components. In this case a finite mixture model can be used to describe the data [15, 19, 2]. If the concentrations of components in the mixture are different for different observations then the model of mixture with varying concentrations (MVC) can be applied [1, 12, 11]. Parametric models of nonlinear regression mixtures were considered in [5, 4]. Estimation in linear regression MVC models was studied also in [6, 8].

*Corresponding author.

In this paper we adopt a semiparametric approach with the use of modified least squares (mLS) technique. The consistency of mLS estimators in regression MVC models was demonstrated in [9]. Our aim is to derive conditions of mLS estimators asymptotic normality and construct confidence sets for the true values of parameters.

The rest of the paper is organized as follows. In Section 2 we introduce the regression mixture model and the mLS estimator for the regression parameters. Asymptotic behavior of the estimator is discussed in Section 3. Confidence ellipsoids for the parameters are constructed in Section 4. Results of simulations are presented in Section 5. Conclusive remarks are made in Section 6.

2 The model and estimator

In this paper we consider regression technique application to data, which are described by the model of mixture with varying concentrations. It means that each observed subject belongs to one of M different sub-populations (i.e. components of the mixture). We observe n such subjects O_1, \dots, O_n . The true number of component which O_j belongs to will be denoted by κ_j . These numbers are not observed, but one knows the probabilities

$$p_{j;n}^k = \mathbf{P}\{\kappa_j = k\}.$$

These probabilities are called the mixing probabilities or concentrations of the components at j -th observation.

For each subject O_j , one observes a set of numerical variables $\xi_{j;n} = \xi_j = (Y_j, X_j^1, \dots, X_j^m)$, where Y is the response and $\mathbf{X}_j = (X_j^1, \dots, X_j^m)^T$ is the vector of independent variables in the regression model

$$Y = g(\mathbf{X}; \boldsymbol{\vartheta}) + \varepsilon,$$

where g is a known regression function, $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_d)^T$ is a vector of unknown regression coefficients, ε is an unobservable regression error. In fact, the coefficients of the model can be different for different components: $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^{(k)}$ if $\kappa_j = k$. The distribution of ε can also depend on κ_j . These dependencies are described in the following model

$$Y_j = g(\mathbf{X}_j, \boldsymbol{\vartheta}^{(\kappa_j)}) + \varepsilon_j^{(\kappa_j)}. \tag{1}$$

Here $\boldsymbol{\vartheta}^{(k)} \in \Theta^{(k)} \subseteq \mathbb{R}^d$ is the vector of unknown regression coefficients corresponding to the k -th mixture component, $\varepsilon_j^{(k)}, j = 1, \dots, n, k = 1, \dots, M$ are independent random variables with distribution dependent on k but not on j .

We will assume that

$$\mathbf{E} \varepsilon_j^{(k)} = 0, \text{ Var } \varepsilon_j^{(k)} = \sigma^{2(k)} < \infty.$$

(The values of $\sigma^{2(k)}$ are unknown.) The independent variables vectors \mathbf{X}_j are considered as random vectors with distribution possibly dependent on κ_j . It is assumed that the error term $\varepsilon_j^{(\kappa_j)}$ and \mathbf{X}_j are conditionally independent for given κ_j . The vectors (ξ_j, κ_j) are independent for different j .

In what follows we will frequently use expectations and probabilities connected with different mixture components. To present them in a compact form, we introduce formal random vectors $(Y^{(m)}, \mathbf{X}^{(m)}, \varepsilon^{(m)})$ which have conditional distribution of $(Y_j, \mathbf{X}_j, \varepsilon_j^{(\kappa_j)})$ given $\kappa_j = m$, i.e., the distribution of the m -th component.

We will also denote by $\mathbf{p}_{;n}$ the matrix of all concentrations for all observations and all components:

$$\mathbf{p}_{;n} = \begin{pmatrix} p_{1;n}^1 & \cdots & p_{1;n}^M \\ \vdots & \ddots & \vdots \\ p_{n;n}^1 & \cdots & p_{n;n}^M \end{pmatrix},$$

$$\mathbf{p}_{j;n} = (p_{j;n}^1, \dots, p_{j;n}^M)^T, \mathbf{p}_{;n}^m = (p_{1;n}^m, \dots, p_{n;n}^m)^T.$$

Similar notation is used for the weights matrix $\mathbf{a}_{;n}$ introduced below.

We are interested in estimating the parameters $\vartheta^{(k)}$ for different components. The considered estimators are based on the modified least squares (mLS) approach. Namely, we consider the weighted least squares functional

$$J^{(k)}(\mathbf{t}) = \sum_{j=1}^n a_{j;n}^k (Y_j - g(\mathbf{X}_j; \mathbf{t}))^2,$$

where $\mathbf{t} \in \Theta^{(k)}$ is a formal parameter, $a_{j;n}^k$ are some weights aimed to single out the k -th mixture component and suppress influence of all other components on the functional $J^{(k)}$. In this presentation, we restrict ourselves by the minimax weights matrix defined as

$$\mathbf{a}_{;n} = \mathbf{\Gamma}_{;n}^{-1} \mathbf{p}_{;n}, \tag{2}$$

where

$$\mathbf{\Gamma}_{;n} = \mathbf{p}_{;n}^T \mathbf{p}_{;n}.$$

(It is assumed here that $\mathbf{\Gamma}_{;n}$ is nonsingular. See [11, 12] for the minimax properties of these weights.) It is readily seen that

$$(\mathbf{a}_{;n}^k)^T \mathbf{p}_{;n}^m = \mathbb{1}\{m = k\}. \tag{3}$$

(Here $\mathbb{1}\{A\}$ is the indicator function of an event A .) So

$$\begin{aligned} \mathbb{E} J^{(k)}(\mathbf{t}) &= \sum_{j=1}^n a_{j;n}^k \sum_{m=1}^M p_{j;n}^m \mathbb{E} (Y^{(m)} - g(\mathbf{X}^{(m)}, \mathbf{t}))^2 = \mathbb{E} (Y^{(k)} - g(\mathbf{X}^{(k)}, \mathbf{t}))^2 \\ &\doteq \mathbb{E} (g(\mathbf{X}^{(k)}, \vartheta^{(k)}) - g(\mathbf{X}^{(k)}, \mathbf{t}))^2 + \sigma^{2(k)} \stackrel{\text{def}}{=} \bar{J}(\mathbf{t}). \end{aligned}$$

The minimum of $\bar{J}(\mathbf{t})$ is attained at $\mathbf{t} = \vartheta^{(k)}$. Thus, if $\vartheta^{(k)}$ is the unique minimum point, one expects that under suitable conditions $J^{(k)}(\mathbf{t}) \rightarrow \bar{J}(\mathbf{t})$ by the law of large numbers, and $\operatorname{argmin}_{\mathbf{t} \in \Theta^{(k)}} J(\mathbf{t}) \rightarrow \vartheta^{(k)}$ as $n \rightarrow \infty$. If g is smooth enough, the argmin can be found as a solution to

$$\mathbf{j}^{(k)}(\mathbf{t}) = 0,$$

where $\mathbf{j}^{(k)}(\mathbf{t})$ denotes the vector of partial derivatives of $J^{(k)}(\mathbf{t})$ by each entry of \mathbf{t} .

In what follows we **define** the mLS estimator $\hat{\vartheta}_{;n}^{(k)}$ for $\vartheta^{(k)}$ as a statistic which is a solution to

$$\mathbf{j}^{(k)}(\mathbf{t}) = \sum_{j=1}^n a_{j;n}^k (Y_j - g(\mathbf{X}_j, \mathbf{t})) \dot{\mathbf{g}}(\mathbf{X}_j, \mathbf{t}) = 0, \tag{4}$$

where

$$\dot{\mathbf{g}}(\mathbf{X}_j, \mathbf{t}) = \left(\frac{\partial g(\mathbf{X}_j, \mathbf{t})}{\partial t^1}, \dots, \frac{\partial g(\mathbf{X}_j, \mathbf{t})}{\partial t^d} \right)^T.$$

If there are many solutions to (4) then $\hat{\vartheta}_{;n}^{(k)}$ can be taken any of them, but it must be a measurable function from the observed data (Y_j, \mathbf{X}_j) , $j = 1, \dots, n$.

Note that so defined mLS estimator can be a point of local minimum of $J^{(k)}(\mathbf{t})$. But we still call it mLS since in was obtained by a modification of the LS technique.

3 Asymptotic behavior of mLS estimators

In this section, we consider asymptotic behavior of $\hat{\vartheta}_{;n}^{(k)}$ as the sample size n tends to infinity. Let us start with some general assumptions on the model.

In this paper we make no assumptions on connections between $\mathbf{p}_{;n}$ and $\mathbf{p}_{;m}$, when $n \neq m$ and don't assume that they tend to some limit as $n \rightarrow \infty$. Some assumptions are made only on asymptotic behavior of some averaged characteristics of concentrations.

Note that if a significant fraction of $p_{j;n}^k$ is bounded away from zero, then entries of the matrix $\mathbf{\Gamma}_{;n} = \mathbf{p}_{;n}^T \mathbf{p}_{;n}$ are of order n as $n \rightarrow \infty$. In what follows we will assume that the limit matrix

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{\Gamma}_{;n} = \mathbf{\Gamma} \tag{5}$$

exists and is nonsingular.

Then the weights $a_{j;n}^k$ are of order $1/n$ as $n \rightarrow \infty$ and the sums of the form $\sum_{j=1}^n a_{j;n}^k a_{j;n}^m p_{j;n}^l p_{j;n}^i$ are of order $1/n$ as well.

We will assume that the limits

$$\lim_{n \rightarrow \infty} n \sum_{j=1}^n a_{j;n}^k a_{j;n}^m p_{j;n}^l p_{j;n}^i \stackrel{\text{def}}{=} \langle \mathbf{a}^k \mathbf{a}^m \mathbf{p}^l \mathbf{p}^i \rangle \tag{6}$$

exist, for all $k, m, l, i = 1, \dots, M$. Then

$$\lim_{n \rightarrow \infty} n \sum_{j=1}^n a_{j;n}^k a_{j;n}^m p_{j;n}^l p_{j;n}^i \stackrel{\text{def}}{=} \langle \mathbf{a}^k \mathbf{a}^m \mathbf{p}^l \rangle = \sum_{i=1}^M \langle \mathbf{a}^k \mathbf{a}^m \mathbf{p}^l \mathbf{p}^i \rangle,$$

since $\sum_{i=1}^M p_{j;n}^i = 1$.

We will also denote

$$\mathbf{h}(\xi_j, \mathbf{t}) = (Y_j - g(\mathbf{X}_j, \mathbf{t})) \dot{\mathbf{g}}(\mathbf{X}_j, \mathbf{t}), \tag{7}$$

so $\hat{\boldsymbol{\vartheta}}_{;n}^{(k)}$ is a solution to

$$\sum_{j=1}^n a_{j;n}^k \mathbf{h}(\boldsymbol{\xi}_j, \mathbf{t}) = 0. \tag{8}$$

Note that (8) is an unbiased generalized estimating equation (GEE, see [17], Section 5.4).

Conditions of the $\hat{\boldsymbol{\vartheta}}_{;n}^{(k)}$ consistency are presented in the next statement.

Theorem 1. *Assume the following.*

1. $\boldsymbol{\Gamma}$ is nonsingular.
2. $\Theta^{(k)}$ is a compact set in \mathbb{R}^d .
3. There exists $\delta > 0$ such that $\mathbf{E} |\varepsilon^{(m)}|^\delta < \infty$, $\mathbf{E} \|\mathbf{X}^{(m)}\|^\delta < \infty$ and

$$\mathbf{E} \sup_{\mathbf{t} \in \Theta^{(m)}} \|\mathbf{h}(\boldsymbol{\xi}^{(m)}, \mathbf{t})\|^{1+\delta} < \infty,$$

for all $m = 1, \dots, M$.

4. The families of functions $g(\mathbf{x}, \cdot)$, $\mathbf{x} \in \mathbb{R}^M$ and $\dot{\mathbf{g}}(\mathbf{x}, \cdot)$, $\mathbf{x} \in \mathbb{R}^M$ are equicontinuous on $\Theta^{(k)}$.

5. $\mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(k)}, \mathbf{t}) \neq 0$ if $\mathbf{t} \neq \boldsymbol{\vartheta}^{(k)}$.

Then $\hat{\boldsymbol{\vartheta}}_{;n}^{(k)} \rightarrow \boldsymbol{\vartheta}^{(k)}$ in probability as $n \rightarrow \infty$.

Proof. See [9]. □

Now consider the asymptotic normality of $\hat{\boldsymbol{\vartheta}}_{;n}^{(k)}$. We will start with a result formulated in more general terms of GEE estimation.

Assume that $\boldsymbol{\Xi}_{;n} = (\boldsymbol{\xi}_{j;n}, j = 1, \dots, n)$ are random observations in a measurable space \mathfrak{X} described by the model of mixture with varying concentrations (MVC), i.e.,

$$\mathbf{P}\{\boldsymbol{\xi}_{j;n} \in A\} = \sum_{m=1}^M p_{j;n}^m F^{(m)}(A), \tag{9}$$

where $F^{(m)}(A) = \mathbf{P}\{\boldsymbol{\xi}^{(m)} \in A\}$ is the distribution of the observed variable $\boldsymbol{\xi}$ for subjects from the m -th mixture component. Let $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}(F) \in \mathbb{R}^d$ be a functional on a set of possible components' distributions. To estimate $\boldsymbol{\vartheta}^{(k)} = \boldsymbol{\vartheta}(F^{(k)})$, we consider an estimating equation of the form (8), where $\mathbf{h} = (h^1, \dots, h^d)^T$ is some estimating function $\mathbf{h} : \mathfrak{X} \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that

$$\mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(k)}, \boldsymbol{\vartheta}^{(k)}) = 0.$$

(I.e., \mathbf{h} is an unbiased estimating function.) Any statistic $\hat{\boldsymbol{\vartheta}}_{;n}^{(k)}$ is called a GEE-estimator for $\boldsymbol{\vartheta}^{(k)}$ if it is an a.s. solution to (8), i.e.,

$$\mathbf{H}^{(k)}(\hat{\boldsymbol{\vartheta}}_{;n}^{(k)}) \stackrel{\text{def}}{=} \sum_{j=1}^n a_{j;n}^k \mathbf{h}(\boldsymbol{\xi}_j, \hat{\boldsymbol{\vartheta}}_{;n}^{(k)}) = 0 \text{ a.s.} \tag{10}$$

We consider the joint parameters vector $\boldsymbol{\vartheta} = ((\boldsymbol{\vartheta}^{(1)})^T, \dots, (\boldsymbol{\vartheta}^{(M)})^T)^T$ and the corresponding estimator $\hat{\boldsymbol{\vartheta}}_{;n} = ((\hat{\boldsymbol{\vartheta}}_{;n}^{(1)})^T, \dots, (\hat{\boldsymbol{\vartheta}}_{;n}^{(M)})^T)^T$ and derive conditions, under which

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{;n} - \boldsymbol{\vartheta}) \xrightarrow{W} N(0, \mathbf{S}), \quad (11)$$

where \mathbf{S} is some matrix which we now will describe. Denote

$$\mathbf{V}^{(m)} \stackrel{\text{def}}{=} \mathbf{E} \dot{\mathbf{h}}(\boldsymbol{\xi}^{(m)}, \boldsymbol{\vartheta}^{(m)}), \quad (12)$$

where

$$\dot{\mathbf{h}}(\boldsymbol{\xi}^{(m)}, \mathbf{t}) = \begin{pmatrix} \frac{\partial h^1(\boldsymbol{\xi}^{(m)}, \boldsymbol{\vartheta}^{(m)})}{\partial t^1} & \cdots & \frac{\partial h^1(\boldsymbol{\xi}^{(m)}, \boldsymbol{\vartheta}^{(m)})}{\partial t^d} \\ \vdots & \ddots & \vdots \\ \frac{\partial h^d(\boldsymbol{\xi}^{(m)}, \boldsymbol{\vartheta}^{(m)})}{\partial t^1} & \cdots & \frac{\partial h^d(\boldsymbol{\xi}^{(m)}, \boldsymbol{\vartheta}^{(m)})}{\partial t^d} \end{pmatrix}.$$

Then

$$\begin{aligned} \mathbf{Z}^{(m,l)} &= \sum_{i=1}^M \langle \mathbf{a}^m \mathbf{a}^l \mathbf{p}^i \rangle \mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(m)}) \mathbf{h}^T(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(l)}) \\ &- \sum_{i=1}^M \sum_{k=1}^M \langle \mathbf{a}^m \mathbf{a}^l \mathbf{p}^i \mathbf{p}^k \rangle \mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(m)}) \mathbf{E} \mathbf{h}^T(\boldsymbol{\xi}^{(k)}, \boldsymbol{\vartheta}^{(l)}), \\ \mathbf{S} &= \begin{pmatrix} \mathbf{S}^{(1,1)} & \cdots & \mathbf{S}^{(1,M)} \\ \vdots & \ddots & \vdots \\ \mathbf{S}^{(M,1)} & \cdots & \mathbf{S}^{(M,M)} \end{pmatrix}, \end{aligned}$$

where

$$\mathbf{S}^{(m,l)} = (\mathbf{V}^{(m)})^{-1} \mathbf{Z}^{(m,l)} (\mathbf{V}^{(m)})^{-T}. \quad (13)$$

(Here and below $\mathbf{V}^{-T} \stackrel{\text{def}}{=} (\mathbf{V}^{-1})^T$.)

Theorem 2. *Let the following assumptions hold.*

1. $\boldsymbol{\vartheta}$ is an inner point of $\Theta = \Theta^{(1)} \times \dots \times \Theta^{(M)}$.
2. $\mathbf{h}(\mathbf{x}, \mathbf{t})$ is continuously differentiable by \mathbf{t} , for almost all $\mathbf{x} \pmod{F^{(m)}}$ for all $m = 1, \dots, M$.
3. For some $\delta > 0$ and some open ball B , such that $\boldsymbol{\vartheta} \in B \subseteq \Theta$

$$\mathbf{E} \sup_{\mathbf{t} \in B} \left\| \dot{\mathbf{h}}(\boldsymbol{\xi}^{(m)}, \mathbf{t}) \right\|^{1+\delta} < \infty,$$

for all $m = 1, \dots, M$.

4. $\mathbf{E} \|\mathbf{h}(\boldsymbol{\xi}^{(m)}, \boldsymbol{\vartheta})\|^2 < \infty$, for all $m = 1, \dots, M$.
5. The matrices $\mathbf{V}^{(m)}$ are finite and nonsingular, for all $m = 1, \dots, M$.
6. The limits $\langle \mathbf{a}^k \mathbf{a}^m \mathbf{p}^l \mathbf{p}^i \rangle$ defined in (6) exist, for all $k, m, l, i = 1, \dots, M$.
7. The matrix $\mathbf{\Gamma}$ defined in (5) exists and is nonsingular.
8. $\hat{\boldsymbol{\vartheta}}_{;n}$ is a consistent estimator of $\boldsymbol{\vartheta}$.

Then (11) holds with the matrix \mathbf{S} defined in (13).

Proof. The **proof of the theorem** is quite standard. Applying the Taylor expansion to the LHS of (10) one obtains

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{;n}^{(k)} - \boldsymbol{\vartheta}^{(k)}) = -[\dot{\mathbf{H}}^{(k)}(\zeta)]^{-1}(\sqrt{n}\mathbf{H}^{(k)}(\boldsymbol{\vartheta}^{(k)})),$$

where ζ is an intermediate point between $\hat{\boldsymbol{\vartheta}}_{;n}^{(k)}$ and $\boldsymbol{\vartheta}^{(k)}$. In view of Assumptions 2–5 and 7 of the theorem, by the same way as in [17], Theorem 5.14 and Lemma 5.3., it can be shown that as $n \rightarrow \infty$,

$$\dot{\mathbf{H}}^{(k)}(\zeta) \rightarrow \mathbf{E} \dot{\mathbf{H}}^{(k)}(\boldsymbol{\vartheta}^{(k)}).$$

Then a straightforward calculation with (3) in mind yields

$$\mathbf{E} \dot{\mathbf{H}}^{(k)}(\boldsymbol{\vartheta}^{(k)}) = \mathbf{V}^{(k)}.$$

Note that by (3)

$$\mathbf{E} \mathbf{H}^{(k)}(\boldsymbol{\vartheta}^{(k)}) = 0.$$

Hence

$$\begin{aligned} \text{Cov}(\mathbf{H}^{(m)}(\boldsymbol{\vartheta}^{(m)}), \mathbf{H}^{(l)}(\boldsymbol{\vartheta}^{(l)})) &= \sum_{j=1}^n a_{j;n}^m a_{j;n}^l \text{Cov}(\mathbf{h}(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(m)}), \mathbf{h}(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(l)})), \\ \text{Cov}(\mathbf{h}(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(m)}), \mathbf{h}(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(l)})) &= \mathbf{E} \mathbf{h}(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(m)}) \mathbf{h}^T(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(l)}) \\ &\quad - \mathbf{E} \mathbf{h}(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(m)}) (\mathbf{E} \mathbf{h}(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(l)}))^T \\ &= \sum_{i=1}^M p_j^i \mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(m)}) \mathbf{h}^T(\boldsymbol{\xi}_j, \boldsymbol{\vartheta}^{(l)}) - \sum_{i,k=1}^M \mathbf{E} p_j^i p_j^k \mathbf{h}(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(m)}) (\mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(k)}, \boldsymbol{\vartheta}^{(l)}))^T. \end{aligned}$$

So,

$$\lim_{n \rightarrow \infty} n \text{Cov}(\mathbf{H}^{(m)}(\boldsymbol{\vartheta}^{(m)}), \mathbf{H}^{(l)}(\boldsymbol{\vartheta}^{(l)})) = \mathbf{Z}^{(m,l)}. \tag{14}$$

Then, applying the central limit theorem with the Lindeberg’s condition as in the proof of Theorem 3.1.1 in [11] one shows that the system of vectors $(\sqrt{n}\mathbf{H}^{(k)}(\boldsymbol{\vartheta}^{(k)}), k = 1, \dots, M)$ converge weakly to the Gaussian system of vectors $(\mathbf{u}^{(k)}, k = 1, \dots, M)$, such that

$$\mathbf{E} \mathbf{u}^{(k)} = 0, \mathbf{E} \mathbf{u}^{(k)} (\mathbf{u}^{(m)})^T = \mathbf{Z}^{(k,m)}.$$

This implies that the system of vectors $(\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{;n}^{(k)} - \boldsymbol{\vartheta}^{(k)}), k = 1, \dots, M)$ converges weakly to the system $((\mathbf{V}^{(k)})^{-1} \mathbf{u}^{(k)}, k = 1, \dots, M)$.

This result is just the statement of the theorem. □

Return to the regression mixture model (1). Obviously it is a partial case of the MVC model (9). How the matrices $\mathbf{V}^{(m)}$ and $\mathbf{Z}^{(m,l)}$ can be represented for the regression mixture model?

Assume that \mathbf{h} is defined by (7) and the function $g(\mathbf{x}, \mathbf{t})$ has second derivatives by \mathbf{t} :

$$\ddot{\mathbf{g}}(\mathbf{X}_j, \mathbf{t}) = \begin{pmatrix} \frac{\partial^2 g(\mathbf{X}_j, \mathbf{t})}{\partial t^1 \partial t^1} & \cdots & \frac{\partial^2 g(\mathbf{X}_j, \mathbf{t})}{\partial t^1 \partial t^d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g(\mathbf{X}_j, \mathbf{t})}{\partial t^d \partial t^1} & \cdots & \frac{\partial^2 g(\mathbf{X}_j, \mathbf{t})}{\partial t^d \partial t^d} \end{pmatrix}.$$

Then

$$\begin{aligned} \mathbf{V}^{(m)} &= \mathbf{E} \frac{\partial}{\partial \mathbf{t}} [(g(\mathbf{X}^{(m)}, \boldsymbol{\vartheta}^{(m)}) + \varepsilon^{(m)} - g(\mathbf{X}^{(m)}, \mathbf{t})) \dot{\mathbf{g}}(\mathbf{X}^{(m)}, \mathbf{t})] \Big|_{\mathbf{t}=\boldsymbol{\vartheta}^{(m)}} \\ &= -\mathbf{E} \dot{\mathbf{g}}(\mathbf{X}^{(m)}, \boldsymbol{\vartheta}^{(m)}) (\dot{\mathbf{g}}(\mathbf{X}^{(m)}, \boldsymbol{\vartheta}^{(m)}))^T + \mathbf{E} \varepsilon^{(m)} \ddot{\mathbf{g}}(\mathbf{X}^{(m)}, \boldsymbol{\vartheta}^{(m)}). \end{aligned}$$

The second term is zero since $\varepsilon^{(m)}$ is independent from $\mathbf{X}^{(m)}$ and $\mathbf{E} \varepsilon^{(m)} = 0$. So

$$\mathbf{V}^{(m)} = -\mathbf{E} \dot{\mathbf{g}}(\mathbf{X}^{(m)}, \boldsymbol{\vartheta}^{(m)}) (\dot{\mathbf{g}}(\mathbf{X}^{(m)}, \boldsymbol{\vartheta}^{(m)}))^T. \quad (15)$$

A similar algebra yields

$$\begin{aligned} \mathbf{Z}^{(m,k)} &= \sum_{l=1}^M \langle \mathbf{a}^m \mathbf{a}^k \mathbf{p}^l \rangle [\sigma^{2(l)} A(l, k, m) + B(l, k, m)] \\ &- \sum_{i,l=1}^M \langle \mathbf{a}^m \mathbf{a}^k \mathbf{p}^l \mathbf{p}^i \rangle (G(i, i, k) - G(i, k, k))(G(l, l, m) - G(l, m, m)), \end{aligned}$$

where

$$\begin{aligned} A(l, k, m) &= \mathbf{E} \dot{\mathbf{g}}(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(k)}) \dot{\mathbf{g}}^T(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(m)}), \\ B(l, k, m) &= \mathbf{E} (g(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(l)}) - g(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(k)})) \\ &\times (g(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(l)}) - g(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(m)})) \dot{\mathbf{g}}(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(k)}) \dot{\mathbf{g}}^T(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(m)}), \\ G(l, m, k) &= \mathbf{E} g(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(m)}) \dot{\mathbf{g}}(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(k)}). \end{aligned}$$

4 Confidence ellipsoids for regression parameters

Apply the results of Section 3 to the construction of asymptotic confidence sets for $\boldsymbol{\vartheta}^{(k)}$. For any $\mathbf{t} \in \mathbb{R}^d$ and any nonsingular $\mathbf{S} \in \mathbb{R}^{d \times d}$, define

$$T^{(k)}(\mathbf{t}, \mathbf{S}) = n(\hat{\boldsymbol{\vartheta}}_{;n}^{(k)} - \mathbf{t})^T \mathbf{S}^{-1} (\hat{\boldsymbol{\vartheta}}_{;n}^{(k)} - \mathbf{t}).$$

It is obvious that if Theorem 2 holds and $\mathbf{S}^{(k,k)}$ is nonsingular, then

$$T^{(k)}(\boldsymbol{\vartheta}^{(k)}, \mathbf{S}^{(k,k)}) \xrightarrow{W} \chi_d^2, \quad (16)$$

where χ_d^2 is the χ^2 -distribution with d degrees of freedom. Note that (16) holds also if $\mathbf{S}^{(k,k)}$ is replaced by a consistent estimator $\hat{\mathbf{S}}_{;n}^{(k,k)}$. Let χ_α be the α -upper quantile of χ_d^2 . Then the set

$$B_{;n}^k(\alpha) = \{\mathbf{t} \in \mathbb{R}^d : T^{(k)}(\boldsymbol{\vartheta}^{(k)}, \hat{\mathbf{S}}^{(k,k)}) < \chi_\alpha\}$$

is an asymptotic α -level confidence set for $\boldsymbol{\vartheta}^{(k)}$ in the sense that

$$\mathbf{P}\{\boldsymbol{\vartheta}^{(k)} \in B_{;n}^k(\alpha)\} \rightarrow 1 - \alpha$$

as $n \rightarrow \infty$.

To accomplish the confidence set construction, we need convenient conditions for the $\mathbf{S}^{(k,k)}$ nonsingularity and a consistent estimator of this matrix.

4.1 Nonsingularity of $\mathbf{S}^{(k,k)}$

Since

$$\mathbf{S}^{(k,k)} = (\mathbf{V}^{(k)})^{-1} \mathbf{Z}^{(k,k)} (\mathbf{V}^{(k)})^{-T},$$

we need conditions for the nonsingularity of $\mathbf{V}^{(k)}$ and $\mathbf{Z}^{(k,k)}$.

Assumption \mathbf{I}_{lk} . For all $\mathbf{c} \in \mathbb{R}^d$ such that $\mathbf{c} \neq 0$,

$$\mathbf{P}\{\mathbf{c}^T \dot{\mathbf{g}}(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(k)}) \neq 0\} > 0.$$

This assumption means that the functions

$$\dot{g}_i(\cdot) = \left. \frac{\partial}{\partial t^i} g(\cdot, \mathbf{t}) \right|_{\mathbf{t}=\boldsymbol{\vartheta}^{(k)}}, \quad i = 1, \dots, d$$

are linearly independent a.s. with respect to the distribution of \mathbf{X} for the l -th component.

Lemma 1. Assume that the matrix

$$\mathbf{A}^{l,k} = \mathbf{E} \dot{\mathbf{g}}(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(k)}) (\dot{\mathbf{g}}(\mathbf{X}^{(l)}, \boldsymbol{\vartheta}^{(k)}))^T$$

exists, is finite and assumption \mathbf{I}_{lk} holds. Then \mathbf{A} is nonsingular.

Proof. Observe that $\mathbf{A}^{l,k}$ is the Gram matrix of the set of functions $G = (\dot{g}_1, \dots, \dot{g}_d)$ in the L_2 space of functions on \mathbb{R}^d with inner product

$$(f, g) = \mathbf{E} f(\mathbf{X}^{(l)}) g(\mathbf{X}^{(l)}).$$

Assumption \mathbf{I}_{lk} implies that the functions in G are linearly independent in this space. So, their Gram matrix is nonsingular. \square

Theorem 3. Assume that the matrix $\mathbf{Z}^{(k,k)}$ exists, is finite, assumption \mathbf{I}_{kk} holds and $\sigma^{2(k)} > 0$. Then $\mathbf{S}^{(k,k)}$ exists and is nonsingular.

Proof. From $\mathbf{V}^{(k)} = -\mathbf{A}^{k,k}$ one readily obtains the nonsingularity of $\mathbf{V}^{(k,k)}$. Show nonsingularity of $\mathbf{Z}^{(k,k)}$.

In what follows \geq means the Loewner order for matrices, i.e., $\mathbf{A} \geq \mathbf{Z}$ means that $\mathbf{A} - \mathbf{Z}$ is a positive semidefinite matrix.

Observe that

$$\begin{aligned} \text{Cov}[\mathbf{h}(\xi_j, \boldsymbol{\vartheta}^{(k)})] &= \mathbf{E} \left[\text{Cov}[\mathbf{h}(\xi_j, \boldsymbol{\vartheta}^{(k)}) | \kappa_j] \right] + \text{Cov} \left[\mathbf{E}[\mathbf{h}(\xi_j, \boldsymbol{\vartheta}^{(k)}) | \kappa_j] \right] \\ &\geq \mathbf{E} \left[\text{Cov}[\mathbf{h}(\xi_j, \boldsymbol{\vartheta}^{(k)}) | \kappa_j] \right] = \sum_{l=1}^M p_{j;n}^l \text{Cov}[\mathbf{h}(\xi^{(l)}, \boldsymbol{\vartheta}^{(k)})] \\ &\geq p_{j;n}^k \text{Cov}[\mathbf{h}(\xi^{(k)}, \boldsymbol{\vartheta}^{(k)})] = p_{j;n}^k \sigma^{2(k)} \mathbf{A}^{k,k}. \end{aligned}$$

So, by (14)

$$\mathbf{Z}^{(k,k)} \geq \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^k \rangle \sigma^{2(k)} \mathbf{A}^{k,k}.$$

Since $\mathbf{A}^{k,k} \geq 0$ and $\det \mathbf{A}^{k,k} \neq 0$, to prove the theorem, it is enough to show that $\langle (\mathbf{a}^{(k)})^2 \mathbf{p}^k \rangle > 0$.

To do this, observe that by (3)

$$\sum_{j=1}^n a_{j;n}^k p_{j;n}^k = 1.$$

Then

$$\sum_{j=1}^n a_{j;n}^k p_{j;n}^k \mathbb{1}\{a_{j;n}^k > 1/(2n)\} = 1 - \sum_{j=1}^n a_{j;n}^k p_{j;n}^k \mathbb{1}\{a_{j;n}^k \leq 1/(2n)\} \geq 1/2,$$

since $0 \leq p_{j;n}^k \leq 1$, and

$$n \sum_{j=1}^n (a_{j;n}^k)^2 p_{j;n}^k \geq n \sum_{j=1}^n (a_{j;n}^k p_{j;n}^k) (a_{j;n}^k \mathbb{1}\{a_{j;n}^k > 1/(2n)\}) \geq 1/4.$$

Therefore,

$$\langle (\mathbf{a}^{(k)})^2 \mathbf{p}^k \rangle = \lim_{n \rightarrow \infty} n \sum_{j=1}^n (a_{j;n}^k)^2 p_{j;n}^k \geq 1/4.$$

□

4.2 Estimation of $\mathbf{S}^{(k,k)}$

There are at least two ways to estimate $\mathbf{S}^{(k,k)}$. The first is based on the plug-in technique. Namely, we construct empirical counterparts to $\mathbf{V}^{(k)}$ and $\mathbf{Z}^{(k,k)}$ and substitute them into (13) to obtain an estimator for $\mathbf{S}^{(k,k)}$. Formula (15) suggests the following estimator of $\mathbf{V}^{(k)}$:

$$\hat{\mathbf{V}}_{;n}^k = - \sum_{j=1}^n a_{j;n}^k \hat{\mathbf{g}}(\mathbf{X}_j, \hat{\boldsymbol{\vartheta}}_{;n}^{(m)}) (\hat{\mathbf{g}}(\mathbf{X}_j, \hat{\boldsymbol{\vartheta}}_{;n}^{(m)}))^T.$$

Estimation of $\mathbf{Z}^{(k,k)}$ is more complicated. We can estimate $\mathbf{M}^{(i,k)} = \mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(k)})$ by

$$\hat{\mathbf{M}}_{;n}^{(i,k)} = \sum_{j=1}^n a_{j;n}^i \mathbf{h}(\boldsymbol{\xi}_{j;n}, \hat{\boldsymbol{\vartheta}}_{;n}^{(k)}),$$

and

$$\mathbf{D}^{(i,k)} = \mathbf{E} \mathbf{h}(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(k)}) \mathbf{h}^T(\boldsymbol{\xi}^{(i)}, \boldsymbol{\vartheta}^{(k)})$$

by

$$\hat{\mathbf{D}}_{;n}^{(i,k)} = \sum_{j=1}^n a_{j;n}^i \mathbf{h}(\boldsymbol{\xi}_{j;n}, \hat{\boldsymbol{\vartheta}}_{;n}^{(k)}) \mathbf{h}^T(\boldsymbol{\xi}_{j;n}, \hat{\boldsymbol{\vartheta}}_{;n}^{(k)}).$$

We also replace the limits $\langle (\mathbf{a}^{(i)})^2 \mathbf{p}^l \rangle$ and $\langle (\mathbf{a}^{(i)})^2 \mathbf{p}^l \mathbf{p}^m \rangle$ with their approximations

$$\alpha(i, l) = \sum_{j=1}^n (a_{j;n}^i)^2 p_{j;n}^l, \quad \alpha(k, i, l) = \sum_{j=1}^n (a_{j;n}^k)^2 p_{j;n}^i p_{j;n}^l.$$

Then the estimator of $\mathbf{Z}^{(k,k)}$ is

$$\hat{\mathbf{Z}}_{;n}^{(k,k)} = \sum_{i=1}^M \alpha(k, i) \hat{\mathbf{D}}_{;n}^{(i,k)} - \sum_{i,l=1}^M \alpha(k, i, l) \hat{\mathbf{M}}_{;n}^{(i,k)} (\hat{\mathbf{M}}_{;n}^{(l,k)})^T.$$

Now, the resulting plug-in estimator for $\mathbf{S}^{(k,k)}$ is

$$plug \hat{\mathbf{S}}_{;n}^{(k,k)} = (\hat{\mathbf{V}}_{;n}^k)^{-1} \hat{\mathbf{Z}}_{;n}^{(k,k)} (\hat{\mathbf{V}}_{;n}^k)^{-T}.$$

By the same methods as in Theorem 5.15 in [17] it can be shown that under assumptions of Theorem 2, this estimator is consistent.

The second approach to estimation of $\mathbf{S}^{(k,k)}$ is based on the jackknife technique. Consider the dataset $\Xi_{;-i,n} = (\xi_{1;n}, \dots, \xi_{i-1;n}, \xi_{i+1;n}, \dots, \xi_{n;n})$, which consists of all observations from $\Xi_{;n}$ without the i -th one. Similarly, the matrix $\mathbf{p}_{;-i,n}$ contains all rows of $\mathbf{p}_{;n}$ except the i th one, $\mathbf{\Gamma}_{;-i,n} = \mathbf{p}_{;-i,n} \mathbf{p}_{;-i,n}^T$ and $\mathbf{a}_{;-i,n} = \mathbf{\Gamma}_{;-i,n}^{-1} \mathbf{p}_{;-i,n}$. Then $\hat{\boldsymbol{\theta}}_{;-i,n}^{(k)}$ is the GEE estimator $\hat{\boldsymbol{\theta}}^{(k)}$ constructed by the data $\Xi_{;-i,n}$ with the weights $\mathbf{a}_{;-i,n}$, i.e., it is a solution to the estimating equation

$$\sum_{j \neq i} a_{j;-i,n}^{(k)} (Y_j - g(\mathbf{X}_j; \mathbf{t}))^2 = 0.$$

The jackknife estimator of $\mathbf{S}^{(k,k)}$ is defined as

$$jn \hat{\mathbf{S}}_{;n}^{(k,k)} = n \sum_{i=1}^n (\hat{\boldsymbol{\theta}}_{;-i,n}^{(k)} - \hat{\boldsymbol{\theta}}_{;n}^{(k)}) (\hat{\boldsymbol{\theta}}_{;-i,n}^{(k)} - \hat{\boldsymbol{\theta}}_{;n}^{(k)})^T.$$

Jackknife estimators for i.i.d. sample are considered in [18]. Consistency of jackknife is demonstrated in [8] for samples from mixtures with varying concentrations in which the components are described by linear errors-in-variables regression models. We do not state the consistency conditions for $jn \hat{\mathbf{S}}_{;n}^{(k,k)}$, but analyze its applicability to the confidence ellipsoids construction in a small simulation study.

5 Simulations results

In the simulation study, the performance of confidence ellipsoids constructed in Section 4 is tested on $N = 1000$ simulated samples in each experiment. In all the experiments, we constructed confidence ellipsoids for the regression parameters with nominal covering probability 95% and calculated the obtained covering frequencies, i.e., the percent of ellipsoids which cover the true parameter vector.

The data were generated from a mixture of two components (i.e., $M = 2$) with mixing probabilities which also were obtained by random generation:

$$P_{j;n}^k = \frac{u_j^k}{u_j^1 + u_j^2}, \quad k = 1, 2,$$

where $u_j^i, i = 1, 2, j = 1, \dots, n$ are independent uniformly distributed on $[0, 1]$.

Table 1. True parameters values for the regression model

m	1	2
$\mu^{(m)}$	0.0	1.0
$\Sigma^{(m)}$	2.0	2.0
$\vartheta_0^{(m)}$	0.5	0.5
$\vartheta_1^{(m)}$	2	-1/3

Table 2. Covering frequencies for normal regression errors

n	first component			second component		
	oracle	plug-in	jk	oracle	plug-in	jk
100	0.668	0.942	0.955	0.729	0.939	0.953
500	0.929	0.931	0.956	0.948	0.934	0.939
1 000	0.954	0.951	0.95	0.944	0.937	0.939
5 000	0.959	0.951	0.943	0.952	0.940	0.931
7 500	0.961	0.942	0.933	0.951	0.938	0.957
1 000	0.954	0.949	0.944	0.944	0.947	0.954

Each observation contains two variables (Y, X) , and their distribution for the m th component follows the logistic regression model with continuous response:

$$Y^{(m)} = g(X^{(m)}, \vartheta^{(m)}) + \varepsilon^{(m)}, \quad X^{(m)} \sim N(\mu^{(m)}, \Sigma^{(m)}),$$

$$g(X, \vartheta) = \frac{1}{1 + \exp(\vartheta_0 + \vartheta_1 X)}.$$

Here $\vartheta^{(m)} = (\vartheta_0^{(m)}, \vartheta_1^{(m)})^T$ is the vector of unknown regression parameters for the m th component to be estimated. The true values of parameters with which the data were generated are presented in Table 1. $\varepsilon^{(m)}$ are zero mean regression errors independent of $X^{(m)}$, and their distributions were different in different experiments.

In each experiment, we calculated

- (i) oracle 95% covering sets at which the true value of \mathbf{S} is used;
- (ii) plug-in 95% confidence ellipsoids based on $^{plug}\hat{\mathbf{S}}_{;n}^{(k,k)}$;
- (iii) jackknife ellipsoids based on $^{jk}\hat{\mathbf{S}}_{;n}^{(k,k)}$.

The ellipsoids were constructed by 1000 simulated samples and covering frequencies were calculated. These frequencies are presented in the tables, for each experiment.

Experiment 1. Here the error terms were zero mean normal with the variance $\sigma^{2(k)} = 0.25$ for $k = 1, 2$. The resulting covering frequencies are presented in Table 2. It seems that the accuracy of the plug-in confidence ellipsoids in this experiment is not high, but enough for the practical purposes for sample sizes larger than 1000. The plug-in ellipsoids accuracy is nearly the same as for the oracle covering sets, so the observed deviations of the covering frequencies from the nominal confidence probability can not be explained by errors in \mathbf{S} estimation. The jackknife ellipsoids are almost as accurate as the plug-in ones.

Experiment 2. Here we consider bounded regression errors, namely $\varepsilon^{(k)}$ are uniform on $[-0.25, 0.25]$. The resulting covering frequencies are presented in Table 3.

Table 3. Covering frequencies for uniform regression errors

n	first component			second component		
	oracle	plug-in	jk	oracle	plug-in	jk
100	0.593	0.952	0.909	0.684	0.964	0.939
500	0.907	0.929	0.938	0.924	0.929	0.934
1 000	0.917	0.959	0.946	0.951	0.944	0.939
5 000	0.938	0.941	0.934	0.947	0.959	0.933
7 500	0.934	0.948	0.948	0.958	0.956	0.944
1 000	0.937	0.947	0.943	0.955	0.945	0.950

Table 4. Covering frequencies for uniform regression errors

n	first component			second component		
	oracle	plug-in	jk	oracle	plug-in	jk
100	0.568	0.942	0.903	0.701	0.932	0.931
500	0.900	0.938	0.944	0.917	0.929	0.956
1 000	0.936	0.932	0.926	0.944	0.931	0.919
5 000	0.930	0.949	0.945	0.942	0.936	0.948
7 500	0.946	0.939	0.953	0.941	0.953	0.926
1 000	0.955	0.947	0.935	0.925	0.944	0.936

It seems that the accuracy of plug-in and jackknife ellipsoids is nearly the same as in Experiment 1. Paradoxically, the oracle ellipsoids perform somewhat worse than the ones in Experiment 1.

Experiment 3. Here we compare the ellipsoids accuracy on the regression with heavy-tailed errors. The errors are taken with distribution of $\eta/10$, where η has Student's t distribution with four degrees of freedom. The results are presented in the Table 4. In this case, the accuracy of jackknife ellipsoids seems significantly worse than in the Experiments 1 and 2. The plug-in ellipsoids show nearly the same performance as in the previous experiments.

6 Conclusion

We presented theoretical results on the asymptotic normality of the modified least squares estimators for mixtures of nonlinear regressions. These results were applied to construction of confidence ellipsoids for the regression coefficients. Simulation results show that the proposed ellipsoids can be used for large enough samples.

References

- [1] Autin, F., Pouet, Ch.: Test on the components of mixture densities. *Stat. Risk. Model.* **28**(4), 389–410 (2011). [MR2877572](https://doi.org/10.1524/strm.2011.1065). <https://doi.org/10.1524/strm.2011.1065>
- [2] Benaglia, T., Chauveau, D., Hunter, D.R., Young, D.S.: *mixtools*: An R Package for Analyzing Finite Mixture Models. *J. Stat. Softw.* **32**(6), 1–29 (2009)
- [3] Borovkov, A.A.: *Mathematical statistics*. Gordon and Breach Science Publishers, Amsterdam (1998). [MR1712750](https://doi.org/10.1080/00949650802590261)
- [4] Faria, S., Soromenho, G.: Fitting mixtures of linear regressions. *J. Stat. Comput. Simul.* **80**(2), 201–225 (2010). <https://doi.org/10.1080/00949650802590261>

- [5] Grün, B., Friedrich, L.: Fitting finite mixtures of linear regression models with varying & fixed effects in R. In: Rizzi, A., Vichi, M. (eds.) *Compstat 2006 – Proceedings in Computational Statistics*, pp. 853–860. Physica Verlag, Heidelberg, Germany (2006). [MR2173118](#)
- [6] Liubashenko, D., Maiboroda, R.: Linear regression by observations from mixture with varying concentrations. *Mod. Stoch. Theory Appl.* **2**(4), 343–353 (2015). [MR3456142](#). <https://doi.org/10.15559/15-VMSTA41>
- [7] Maiboroda, R.: *Statistical analysis of mixtures*. Kyiv University Publishers, Kyiv (2003). (in Ukrainian)
- [8] Maiboroda, R., Sugakova, O.: Jackknife covariance matrix estimation for observations from mixture. *Mod. Stoch. Theory Appl.* **6**(4), 495–513 (2019). [MR4047396](#). <https://doi.org/10.15559/19-vmsta145>
- [9] Miroshnichenko, V.O.: Generalized least squares estimates for mixture of nonlinear regressions. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics & Mathematics* **219**(5).
- [10] Maiboroda, R., Kubaichuk, O.: Asymptotic normality of improved weighted empirical distribution functions. *Theory Probab. Math. Stat.* **69**, 95–102 (2004). [MR2110908](#). <https://doi.org/10.1090/S0094-9000-05-00617-4>
- [11] Maiboroda, R.E., Sugakova, O.V.: *Estimation and classification by observations from mixture*. Kuiv University Publishers, Kyiv (2008). (in Ukrainian)
- [12] Maiboroda, R., Sugakova, O.: Statistics of mixtures with varying concentrations with application to DNA microarray data analysis. *J. Nonparametr. Stat.* **24**(1), 201–205 (2012). [MR2885834](#). <https://doi.org/10.1080/10485252.2011.630076>
- [13] Maiboroda, R.E., Sugakova, O.V., Doronin, A.V.: Generalized estimating equations for mixtures with varying concentrations. *Can. J. Stat.* **41**(2), 217–236 (2013). [MR3061876](#). <https://doi.org/10.1002/cjs.11170>
- [14] Masiuk, S., Kukush, A., Shklyar, S., Chepurny, M., Likhtarov, I. (eds.): *Radiation Risk Estimation: Based on Measurement Error Models*, 2nd edn. de Gruyter series in Mathematics and Life (2017). [MR3726857](#)
- [15] McLachlan, G., Krishnan, T.: *The EM Algorithm and Extensions*, 2nd edn. Wiley (2008). [MR2392878](#). <https://doi.org/10.1002/9780470191613>
- [16] Seber, G.A.F., Wild, C.J.: *Nonlinear Regression*. Wiley (2003). [MR0986070](#). <https://doi.org/10.1002/0471725315>
- [17] Shao, J.: *Mathematical statistics*. Springer, New York (1998). [MR1670883](#)
- [18] Shao, J., Tu, D.: *The Jackknife and Bootstrap*. Springer, NY (1995). [MR1351010](#). <https://doi.org/10.1007/978-1-4612-0795-5>
- [19] Titterton, D.M., Smith, A.F., Makov, U.E.: *Analysis of Finite Mixture Distributions*. Wiley, New York (1985). [MR0838090](#)