Austrian Journal of Statistics
2025, Volume 54, 1–16.
http://www.ajs.or.at/
doi:10.17713/ajs.v54i1.1953

Estimation of Concentrations Parameters in the Model of Mixture with Varying Concentrations

Rostyslav Maiboroda
Taras Shevchenko National University of Kyiv

Vitaliy Miroshnychenko
Taras Shevchenko National University of Kyiv

Olena Sugakova
Taras Shevchenko National University of Kyiv

Abstract

Model of Mixture with Varying Concentrations (MVC) is a generalization of the finite mixture model (FMM) at which the mixing probabilities (concentrations of components in the mixture) vary from observation to observation. In this paper we assume that the components' distributions are completely unknown, while the concentrations are known up to some unknown euclidean parameter. Two approaches are considered to the semiparametric estimation of this parameter in the case of two-component mixture. The Least Squares (LS) estimator is based on fitting the distribution functions of the observations. The Empirical Maximum Likelihood estimator (EML) utilizes some empirical version of the likelihood function. Consistency of the LS estimator is demonstrated. A fast algorithm for the LS estimator calculation is presented. EML and LS estimators are compared via simulations. Both EML and LS estimators show sufficiently good performance in all the experiments. The LS estimator performed better then the EML one for components with different variance. The EML estimator outperformed the LS one for nongaussian components with asymmetric tails.

Keywords: mixture with varying concentrations, semiparametric estimation, empirical maximum likelihood, least squares.

1. Introduction

Finite Mixture Models (FMM) are widely used to describe distribution of heterogeneous statistical data. An introduction to the theory and statistical applications of FMM can be found in (McLachlan and Peel (2000),McNicholas (2017)). Computational aspects of FMM statistical analysis are considerd in Celeux, Fruhwirth-Schnatter, and Robert (2018). Applications to biological and medical data statistics are presented in Schlattmann (2009). In classical FMMs the mixing probabilities (concentrations of components in the mixture) are assumed to be the same for all observations. We consider more flexible model of Mixture with Vary-

ing Concentrations (MVC) where the concentrations of components vary from observation to observation. A brief introduction to MVC with an application to genetics statistical data see in Maiboroda and Sugakova (2012). Some multivariate MVC models with an application to sociological data are considered in Maiboroda, Miroshnichenko, and Sugakova (2022). An interesting example of MVC application to a neuroscience data analysis see in Pidnebesna, Fajnerová, Horáček, and Hlinka (2023). Some statistical tests for MVC models were discussed in Autin and Pouet (2011). In all these papers a nonparametric approach is considered when the distributions of components are completely unknown. A parametric case is discussed in Grün and Leisch (2006).

In theoretical considerations the components' concentrations in MVC are usually assumed to be known. In many applications it is more realistic to assume some parametric model for these concentrations. A Least Squares (LS) approach to estimation of concentrations' parameters when the components' distributions are completely unknown is developed in Maiboroda (2002).

In this paper we consider an alternative approach to the estimation of concentrations' parameters based on a version of smoothed Empirical Maximum Likelihood (EML) technique, Qin and Lawless (1994), Owen (2001). Performance MVC and EML estimators is compared via simulations.

The rest of paper is organized as follows. We consider a motivating example from medical statistics in Section 2. Construction of LS estimator for concentrations' parameters, its consistency and algorithmic implementation are considered in Section 3. The EML estimator is introduced in Section 4. Results of simulations are presented in Section 5. Section 6 contains concluding remarks. Technical proofs are placed in the Appendix.

2. Motivating example

Let there be a medical statistics data on patients with a disease caused by some virus. We observe a result of some biochemical blood test for these patients which is considered as a random variable $\xi \in \mathbb{R}$. The value of ξ for the j-th patient is ξ_j, $j = 1, \dots, n$.

It is known that the disease can be caused by one of the two virus strains (variants): **Strain 1** and **Strain 2**. Let s_j be the strain which caused the disease of the j-th patient. We suppose that the distribution of ξ_j is different for different strains s_j. Let F_i be the CDF of ξ_j for $s_j = i$, i.e.,

$$F_i(x) = \mathsf{P}\{\xi_j < x \mid s_j = i\},\ i = 1, 2.$$

The true values of s_j are not observed. We may only know the probabilities $p_j^i = \mathsf{P}\{s_j = i\}$, $j = 1, \dots, n$. Roughly speaking, the probability p_j^i can be interpreted as the concentration of the i-th strain in the population at the time when the j-th patient was infected. (If we assume that both strains have the same infection rate).

Surely, $p_j^2 = 1 - p_j^1$.

Then the unconditional distribution of ξ_j for unknown s_j is a mixture of the distributions F_i:

$$\mathsf{P}\{\xi_j < x\} = \mathbb{P}_j(x) = p_j^1 F_1(x) + p_j^2 F_2(x). \tag{1}$$

This is the model of two-component mixture with varying concentrations (MVC).

If one knows p_j^1, $j = 1, \dots, n$, then it is possible to estimate F_i by $\xi_1, \dots, \xi_n$ with weighted empirical CDFs Maiboroda and Sugakova (2012). (Observe that when no assumptions are made on F_i the model (1) is identifiable only if p_j^1 are not constants. On the identifiability of (1) see Maiboroda and Sugakova (2012)).

In this presentation we consider a more difficult case when p_j^i are known up to some unknown parameter $\vartheta \in \Theta$:

$$p_j^1 = p_j^1(\vartheta).$$

For example, one may have some theoretical model of the epidemic dynamics which predicts the concentration of Strain1 at the time moment t: $p(t,\vartheta)$, where ϑ are some model parameters. Then, if the j-th observed patient was infected at time t_j,

$$p_j^1 = p(t_j,\vartheta).$$

The Logistic model is, maybe, the simplest example of such models in which one the first strain replaces the other one in the population of deseased persons:

$$p(t,\vartheta) = \frac{1}{1+\exp\left(\frac{t-m}{s}\right)}.$$

Here the unknown parameter is $\vartheta = (m,s)$, where $m \in \mathbb{R}$ is the shift and $s > 0$ is the scale of the logistic transform.

If the changes in concentrations have seasonal pattern one may consider the Harmonic model

$$p(t,\vartheta) = \frac{1}{2}\left(1+\sin\left(2\pi\frac{t-m}{T}\right)\right), \tag{2}$$

where $\vartheta = (m,s)$, $m \in [0, T/2]$ is the shift, $T > 0$ is the period of the harmonic wave. Note that the shift $m = T/2$ leads to the model (1) equivalent to the model $m = 0$ with the interchange of the first and second components.

3. Least squares estimator

3.1. Construction of estimator

So, we consider the data $(\xi_1,\dots,\xi_n)$, such that ξ_j are independent random variables with the CDFs

$$\mathsf{P}\{\xi_j < x\} = \mathbb{P}_j(x) = p_j(\vartheta)F_1(x) + (1-p_j(\vartheta))F_2(x). \tag{3}$$

The CDFs F_i of components are different but completely unknown. The concentrations $p_j(\vartheta)$ of the first component are known up to the unknown parameter $\vartheta \in \Theta$. (In the above examples $\Theta \subseteq \mathbb{R}^d$, but in general, it can be any parametric set). Our aim is to construct an estimator for ϑ by these data. It will be the maximizer of some empirical criterion $\hat{R}$. We start by introduction a theoretical LS criterion R to which $\hat{R}$ is an estimator.

Choose any measure π on $\mathbb{R}$. Let us define

$$Q(\alpha; G_1, G_2) = \frac{1}{n}\sum_{j=1}^{n}\int (\mathbb{P}_j(x) - p_j(\alpha)G_1(x) - (1-p_j(\alpha))G_2(x))^2\pi(dx).$$

Here $\alpha \in \Theta$ is a possible value of ϑ, G_1 and G_2 are any functions considered as candidates on the role of F_1 and F_2 (we do not assume here that G_i are proper distribution functions).

In what follows we use angle brackets $\langle\cdot\rangle_n$ to denote the averaging by $j = 1,\dots,n$, say,

$$\langle p_\cdot(\alpha)p_\cdot(\vartheta)\rangle_n = \frac{1}{n}\sum_{j=1}^{n} p_j(\alpha)p_j(\vartheta)$$

and so on.

So

$$Q(\alpha, G_1, G_2) = \left\langle \int (\mathbb{P}_\cdot(x) - p_\cdot(\alpha)G_1(x) - (1-p_\cdot(\alpha))G_2(x))^2\pi(dx)\right\rangle_n.$$

Since $Q(\alpha, G_1, G_2) \geq 0$ and $Q(\vartheta, F_1, F_2) = 0$, it is obvious that

$$\vartheta \in \operatorname{argmin}_\alpha \min_{G_1,G_2} Q(\alpha, G_1, G_2).$$

There can be some minimizers of this function other then ϑ. Below we consider conditions which exclude such cases.

Let

$$Q(\alpha) = \min_{G_1,G_2} Q(\alpha; G_1, G_2),$$

where min is taken over all possible functions, and

$$u_j(\alpha) = \frac{p_j(\alpha) - \langle p_\cdot(\alpha)\rangle_n}{\sqrt{\Delta_n(\alpha)}},$$

where

$$\Delta_n(\alpha) = \langle (p_\cdot(\alpha))^2\rangle_n - (\langle p_\cdot(\alpha)\rangle_n)^2.$$

I.e. $u_\cdot(\alpha)$ is a centered, normalized version of $p_\cdot(\alpha)$.

Then a simple algebra yields

$$Q(\alpha) = \int \left\langle (\mathbb{P}_\cdot(x) - \langle \mathbb{P}_\cdot(x)\rangle_n)^2 \right\rangle_n \pi(dx) - \int (\langle \mathbb{P}_\cdot(x) u_\cdot(\alpha)\rangle_n)^2 \pi(dx).$$

Since $\int \langle (\mathbb{P}_\cdot(x) - \langle \mathbb{P}_\cdot(x)\rangle_n)^2\rangle_n \, \pi(dx)$ does not depend on α, we obtain

$$\vartheta = \operatorname{argmin}_\alpha Q(\alpha) = \operatorname{argmax}_\alpha R_n(\alpha),$$

where

$$R_n(\alpha) = \int (\langle \mathbb{P}_\cdot(x) u_\cdot(\alpha)\rangle_n)^2 \pi(dx).$$

This is our theoretical criterion.

Observe that

$$\langle \mathbb{P}_\cdot(x) u_\cdot(\alpha)\rangle_n = \mathsf{E}\, \frac{1}{n} \sum_{j=1}^n u_j(\alpha) \mathbf{1}\{\xi_j < x\}.$$

So, we take

$$\hat{R}_n(\alpha) = \int \left(\frac{1}{n} \sum_{j=1}^n u_j(\alpha) \mathbf{1}\{\xi_j < x\} \right)^2 \pi(dx) \tag{4}$$

as our empirical criterion and define the least squares (LS) estimator $\hat{\vartheta}_n^{LS}$ for ϑ as a statistics such that

$$\hat{R}_n(\hat{\vartheta}_n^{LS}) = \sup_{\alpha\in\Theta} \hat{R}_n(\alpha).$$

3.2. Consistency

Here we derive conditions under which the LS estimator is consistent. For two-dimensional MVC models they are less restrictive then the conditions presented in Maiboroda (2002).

In the asymptotic analysis ($n \to \infty$) we consider concentrations which can be different for different n

$$p_j^1 = p_{j;n}^1(\vartheta).$$

So we consider a sequence of samples $(\xi_{1;n}, \dots, \xi_{n;n})$, $n = 1, 2, \dots$ The observations $\xi_{j;n}$ are independent for fixed n. We do not assume any specific relation of observations in samples of different sizes. I.e., there can be $\xi_{j;n_1} = \xi_{j;n_2}$ for some j, or they can be independent. In the model of mixing probabilities of the form $p_j^1 = p(t_j; \vartheta)$ the time moments $t_j = t_{j;n}$ are considered as nonrandom real numbers. There are no specific constrains on the behavior of $t_{j;n}$ for different n.

In what follows we drop the subscript $;n$ to avoid notation complexities.

Lemma. *Assume that*

(1) $\pi(\mathbb{R}) < \infty$,

(2) there exists $\Delta^- > 0$ *such that for all* $n \in \mathbb{N}$ *and all* $\alpha \in \Theta$

$$\Delta_n(\alpha) = \langle (p_\cdot(\alpha))^2 \rangle_n - (\langle p_\cdot(\alpha) \rangle_n)^2 \geq \Delta^-,$$

(3) there exists $V < \infty$ *such that for all* $n \in \mathbb{N}$

$$\sup_{\alpha \in \Theta} \sum_{j=1}^{n-1} |p_{j+1}^1(\alpha) - p_j^1(\alpha)| < V.$$

Then there exists a random variable $\Lambda < \infty$, *such that*

$$\sup_{\alpha \in \Theta} |\hat{R}_n(\alpha) - R_n(\alpha)| \leq \Lambda \sqrt{\frac{\log n}{n}}.$$

(See Appendix for proof).

Let us see, when $R_n(\alpha)$ possess a unique maximum at $\alpha = \vartheta$.

Observe that

$$R_n(\alpha) = \int (\langle \mathbb{P}_\cdot(x) u_\cdot(\alpha) \rangle_n)^2 \pi(dx)$$

$$= \langle p_\cdot(\vartheta) u_\cdot(\alpha) \rangle_n^2 \int (F_1(x) - F_2(x))^2 \pi(dx).$$

We will suppose that $\int (F_1(x) - F_2(x))^2 \pi(dx) > 0$.

Since

$$\langle p_\cdot(\vartheta) u_\cdot(\alpha) \rangle_n = \langle u_\cdot(\vartheta) u_\cdot(\alpha) \rangle_n / \Delta_n(\vartheta)$$

and $\langle u_\cdot(\vartheta) \rangle_n^2 = 1$, $R_n(\alpha)$ attains the unique maiximum at ϑ iff

$$|r_n(\alpha, \vartheta)| < 1$$

for all $\alpha \neq \vartheta$, where

$$r_n(\alpha, \vartheta) = \langle u_\cdot(\vartheta) u_\cdot(\alpha) \rangle_n = \frac{\langle (p_\cdot(\vartheta) - \langle p_\cdot(\vartheta) \rangle_n)(p_\cdot(\alpha) - \langle p_\cdot(\alpha) \rangle_n) \rangle_n}{\sqrt{\Delta_n(\alpha) \Delta_n(\vartheta)}}$$

can be considered as a "sample correlation" between the hypothetical first component concentrations for the unknown parameter values α and ϑ.

To obtain an asymptotic result (consistency) the assumption $|r_n(\alpha, \vartheta)| < 1$ must be strengthened.

Combining this observation with the Lemma we obtain

Theorem. *Assume that* Θ *is compact and the following conditions hold.*

(1) $\pi(\mathbb{R}) < \infty$,

(2) There exists $\Delta^- > 0$ *such that for all* $n \in \mathbb{N}$ *and all* $\alpha \in \Theta$

$$\Delta_n(\alpha) = \langle (p_\cdot(\alpha))^2 \rangle_n - (\langle p_\cdot(\alpha) \rangle_n)^2 \geq \Delta^-,$$

(3) there exists $V < \infty$ *such that for all* $n \in \mathbb{N}$

$$\sup_{\alpha \in \Theta} \sum_{j=1}^{n-1} |p_{j+1}^1(\alpha) - p_j^1(\alpha)| < V,$$

(4) $\int (F_1(x) - F_2(x))^2 \pi(dx) > 0$,

(5) for all $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{\alpha:\ |\alpha-\vartheta|>\delta} |r_n(\alpha,\vartheta)| < 1.$$

Then $\hat{\vartheta}_n^{LS} \to \vartheta$ in probability as $n \to \infty$.

3.3. Algorithmic issues

Fast calculation algorithm

Recall that the estimator $\hat{\vartheta}_n^{LS}$ is the maximizer of

$$\hat{R}_n(\alpha) = \int \left(\frac{1}{n} \sum_{j=1}^{n} u_j(\alpha) \mathbf{1}\{\xi_j < x\} \right)^2 \pi(dx)$$

$$= \frac{1}{n^2} \sum_{i,j=1}^{n} u_i(\alpha) u_j(\alpha) \int \mathbf{1}\{\xi_i < x\} \int \mathbf{1}\{\xi_j < x\} \pi(dx)$$

$$= \frac{1}{n^2} \sum_{i,j=1}^{n} u_i(\alpha) u_j(\alpha) \bar{\pi}(\max(\xi_i, \xi_j)),$$

where

$$\bar{\pi}(z) = \int \mathbf{1}\{z < x\} \pi(dx) = \pi\{x \ : \ x > z\}.$$

Direct calculation by this formula needs $\sim Cn^2$ operations. A simple reordering of terms allows us to reduce this calculation complexity to $\sim Cn \log(n)$.

Let $\xi_{[1]} \le \xi_{[2]} \le \cdots \le \xi_{[n]}$ be the order statistics of ξ_j, $j = 1, \ldots, n$, $u_{[j]}(\alpha)$, $j = 1, \ldots, n$ are the weights $u_j(\alpha)$ reordered together with $\xi_{[j]}$. Then

$$\hat{R}_n(\alpha) = \frac{1}{n^2} \sum_{i,j=1}^{n} u_{[i]}(\alpha) u_{[j]}(\alpha) \bar{\pi}(\max(\xi_{[i]}, \xi_{[j]})).$$

Observe that

$$\pi(\max(\xi_{[i]}, \xi_{[j]})) = \pi_{[j]} = \pi(\xi_{[j]}) \text{ if } i \le j.$$

So

$$\hat{R}_n(\alpha) = \frac{1}{n^2} \left(2 \sum_{j=2}^{n} u_{[j]}(\alpha) U_j \pi_{[j]} + \sum_{j=1}^{n} (u_{[j]}(\alpha))^2 \pi_{[j]} \right), \tag{5}$$

where $U_i = \sum_{j=1}^{i-1} u_{[j]}(\alpha)$.

To calculate $\hat{R}_n(\alpha)$ by (5) one needs to sort the sample $\xi_j, j = 1, \ldots, n$ in the ascending order ($\sim Cn \log(n)$ operations), calculate U_i recusively for $i = 1, \ldots, n$ ($\sim Cn$ operations), and apply (5) ($\sim Cn$ operations).

When some numeric procedure is used to find maximum of $\hat{R}_n(\alpha)$, one needs to calculate this function many times at different points α. Note that the reordering of ξ_j and calculation of $\pi_{[j]}$ can be done once at the beginning of this procedure and the results can be used in calculation of $\hat{R}_n(\alpha)$ for all needed α. This makes the calculation of $\hat{\vartheta}_n$ a rather fast procedure.

Choice of integration measure

How to choose the integration measure π in our empirical criterion $\hat{R}(\alpha)$? Recall that the theoretical criterion, which we would like to maximize is

$$R_n(\alpha) = \langle p_\cdot(\vartheta) u_\cdot(\alpha) \rangle_n^2 \int (F_1(x) - F_2(x))^2 \pi(dx).$$

So we need $\int (F_1(x) - F_2(x))^2 \pi(dx) > 0$ to obtain a consistent estimate. But F_1 and F_2 are unknown. So it is prudent to take π with support on all $\mathbb{R}$. Then any difference of the components' distributions will be visible by our criterion. On the other hand, if most weight of π is placed in the domain where no ξ_j is observed the estimate should be inefficient.

In our simulation experiments we took Gaussian π with the mean $\bar{\xi} = \frac{1}{n}\sum_{j=1}^{n} \xi_j$ and the variance $S^2(\xi) = \frac{1}{n}\sum_{j=1}^{n}(\xi_j - \bar{\xi})^2$. The results show that with this choice of π the estimator performs satisfactory in simple cases, when the components' distributions differ significantly.

4. Smoothed empirical likelihood estimator

LS estimator uses difference between CDFs of components. It is well known that estimators based on densities usually are more efficient then the ones based on CDFs. This observation is utilized in the Maximum Likelihood (ML) estimation technique which leads to asymptotically efficient estimators in many parametric problems. There exists also a nonparametric Maximum Empirical Likelihood (MEL) estimation technique (see Owen (2001),Qin and Lawless (1994),Vexler and Boca Raton (2018)).

Application of standard ML to parametric models usually doesn't lead to consistent estimators. This is the case, e.g., for Gaussian mixtures when the means and variances of components are unknown. In nonparametric models direct application of MEL also is not possible. So, to derive an estimator for ϑ in the model (3) we apply a combined technique. To derive an empirical version of likelihood we estimate the PDFs f_i of the mixture components by kernel estimators $\hat{f}_{i;n}$ and use the estimators $\hat{f}_{i;n}$ instead of the true f_i in the likelihood.

So, assume that there exist PDFs $f_i(x)$ of distributions F_i, $i = 1, 2$. Then the PDF of ξ_j is

$$p_j(\vartheta) f_1(x) + (1 - p_j(\vartheta)) f_2(x)$$

Then, if f_i are known, the log-likelihood for ϑ estimation is

$$l(\alpha) = \sum_{j=1}^{n} \log(p_j(\alpha) f_1(\xi_j) + (1 - p_j(\alpha)) f_2(\xi_j)). \tag{6}$$

In this (parametric) case the maximum likelihood estimator for ϑ is defined as

$$\hat{\vartheta}_n^{PML} = \operatorname{argmax}_{\alpha \in \Theta} l(\alpha). \tag{7}$$

Conditions of asymptotic optimality of this estimator can be obtained by usual parametric statistics methods (e.g. Section III.4 in Ibragimov and Has'minskii (1981)).

In our case the PDFs f_i are unknown. So we will estimate them under the assumption that the unknown ϑ is equal to some fixed value α and then replace f_i in (6) by these estimators. Weighted kernel density estimators are used for this purpose of the form:

$$\hat{f}_{i;n}(x; \alpha) = \frac{1}{nh} \sum_{j=1}^{n} a_j^i(\alpha) K\left(\frac{x - \xi_j}{h}\right), \tag{8}$$

where K is a kernel, i.e. some PDF, h is a bandwidth, and $a_j^i(\alpha)$ are weights separating the ith component from other mixture components.

We consider the weights of the form

$$a_j^i(\alpha) = \sum_{k=1}^{2} \gamma_k^i(\alpha) p_j^k(\alpha), \tag{9}$$

where $p_j^1(\alpha) = p_j(\alpha)$, $p_j^2(\alpha) = 1 - p_j(\alpha)$ are the concentrations of the first and second components which correspond to the unknown parameter value α.

The coefficients $\gamma_k^i(\alpha)$ are taken to satisfy the condition:

$$\langle a_\cdot^i(\alpha) p_\cdot^k(\alpha)\rangle_n = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases}, \text{ for } i,k = 1,2.$$

A straightforward computation provides

$$\begin{aligned} \gamma_1^1(\alpha) &= \langle p_\cdot^2(\alpha) p_\cdot^2(\alpha)\rangle_n / \Delta_n(\alpha),\ \gamma_1^2(\alpha) = \langle p_\cdot^1(\alpha) p_\cdot^2(\alpha)\rangle_n / \Delta_n(\alpha), \\ \gamma_2^1(\alpha) &= \langle p_\cdot^1(\alpha) p_\cdot^2(\alpha)\rangle_n / \Delta_n(\alpha),\ \gamma_2^2(\alpha) = -\langle p_\cdot^1(\alpha) p_\cdot^1(\alpha)\rangle_n / \Delta_n(\alpha). \end{aligned} \tag{10}$$

The weights a_j^i given by (9) with (10) are called the minimax weights. Their properties are described in Maiboroda and Sugakova (2012).

It is shown in Sugakova (1999) that if $\alpha = \vartheta$ is the true value of the unknown parameter, then, under suitable assumptions, the estimators $\hat{f}_{i;n}(x,\vartheta)$ converge to the true values of $f_i(x)$, $i = 1,2$. The optimal convergence rate of the estimator is attained if

$$h = h_{i;n} = \left(\frac{A_n^i d^2}{n D^4 \varphi_i}\right)^{1/5}, \tag{11}$$

where $A_n^i = \langle (a^i)^2\rangle_n$, $d = \int K^2(z)dz$, $D^2 = \int z^2 K(z)dz$,

$$\varphi_i = \int \left(\frac{\partial^2}{\partial x^2} f_i(x;\vartheta)\right)^2 dx$$

(see Theorem 3 in Sugakova (1999) for details).

Since φ_i is unknown it should be estimated/approximated to obtain a feasible bandwidth selection rule. We adopt a simple approximation similar to the Silverman's rule of thumb. In this approximation φ_i for the unknown f_i is replaced by φ_i of normal density with the same variance as of f_i. This variance is estimated by the weighted sample variance

$$S_i^2(\alpha) = \frac{1}{n}\sum_{j=1}^n a_j^{i\pm}(\alpha)(\xi_j - \bar{\xi}_i(\alpha)),$$

where

$$\bar{\xi}_i(\alpha) = \frac{1}{n}\sum_{j=1}^n a_j^{i\pm}(\alpha)\xi_j.$$

Here $(a_j^{i\pm}(\alpha), j = 1,\dots,n)$ is the vector of weights obtained by correction of $(a_j^i(\alpha), j = 1,\dots,n)$. The correction allows to make all the weights non-negative (see Maiboroda and Kubaichuk (2005)).

So we obtain the following modified Silverman's rule of thumb

$$\hat{h}_{i;n}(\alpha) = \left(\frac{8\sqrt{\pi} d^2}{3D^2 n}\langle (a^i(\alpha))^2\rangle_n\right)^{1/5} S_{i;n}(\alpha). \tag{12}$$

In what follows we use $h = \hat{h}_{i;n}(\alpha)$ in $\hat{f}_{i;n}(x;\alpha)$ defined by (8).

If $\alpha \neq \vartheta$, $\hat{f}_{i;n}(x;\alpha)$ converges to some linear combination $z f_1(x) + (1-z) f_2(x)$ of the components' PDFs. Note that z here not necessarily belongs to $[0,1]$, so this limit can attain negative values for some x. So, to construct an empirical version of the loglikelihood (6) we truncate the obtained density estimator away of zero by some threshold $t > 0$. The resulting empirical quasi-loglikelihood is:

$$\hat{l}_n(\alpha) = \sum_{j=1}^n \log(\max(t, p_j(\alpha)\hat{f}_{1;n}(\xi_j;\alpha) + (1-p_j(\alpha))\hat{f}_{2;n}(\xi_j;\alpha))). \tag{13}$$

Then the empirical quasi-likelihood estimator for ϑ is defined by

$$\hat{\vartheta}_n^{EML} = \operatorname{argmax}_{\alpha\in\Theta} \hat{l}_n(\alpha). \tag{14}$$

5. Simulation results

We performed a small simulation study to compare the performance of the least squares estimator $\hat{\vartheta}_n^{LS}$ and the empirical quasi-likelihood estimator $\hat{\vartheta}_n^{EML}$. In these comparisons we consider also the parametric estimator $\hat{\vartheta}_n^{PML}$ defined in (7). This estimator utilizes much more prior information on the data distribution then $\hat{\vartheta}_n^{LS}$ and $\hat{\vartheta}_n^{EML}$. So one expects that the bias and variance of $\hat{\vartheta}_n^{PML}$ will be smaller then of the nonparametric estimators for any version of tuning parameters, such as the measure π for $\hat{\vartheta}_n^{LS}$ or the bandwidth h and the threshold t for $\hat{\vartheta}_n^{EML}$. In this sense $\hat{\vartheta}_n^{PML}$ represents a benchmark for possible improvements of the considered estimation techniques.

In the experiments we simulated data from the harmonic model (1-2) with known period $T = 1$ and unknown m. For the sample size n the covariate values t_j were taken as $t_j = j/n$. The true value of m to be estimated was $m_0 = 0.25$. The components' distributions F_1 and F_2 were different in different experiments (see below).

We considered samples of size $n = 100, 250, 500, 1000, 2000$ and 5000. For each sample size n, $B = 1000$ samples $X^{(b)}$, $b = 1, \dots B$ were generated. Three estimators were calculated for each $X^{(b)}$: the least squares estimator $\hat{m}_{(b)}^{LS}$, the empirical quasi-likelihood estimator $\hat{m}_{(b)}^{EML}$ and the parametric maximum likelihood estimator $\hat{m}_{(b)}^{PML}$. The biases (b) and standard deviations (sd) of these estimators were approximated by

$$\mathrm{b}^* = \bar{m}^* - m_0,\ \mathrm{sd}^* = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{m}_{(b)}^* - \bar{m}^*)^2},\ \bar{m}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{m}_{(b)}^*,$$

where $*$ means any of symbols LS, EML or PML.

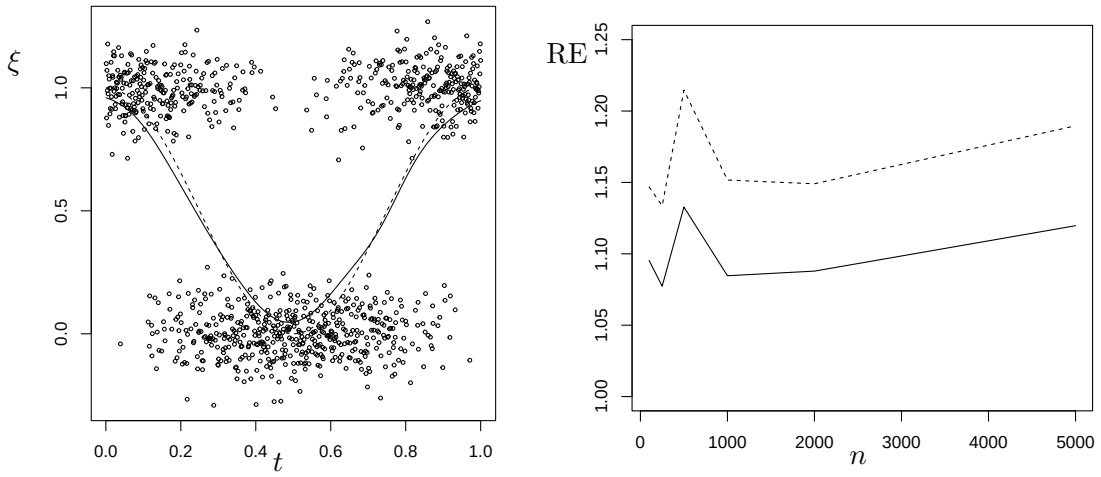
Experiment 1. In this experiment we consider the case when the components' distributions differ only in mean. Namely, $F_1 \sim N(0, 0.01)$, $F_2 \sim N(1, 0.01)$. The variance of the observations is so small that the components can be separated with the naked eye. A typical scatterplot of (t_j, ξ_j) for such data of size $n = 1000$ is represented on Fig. 1(a). By (3)

$$\mathsf{E}\,\xi_j = \mu_2 + (\mu_1 - \mu_2)p_j(\vartheta),$$

where μ_i are the means of distributions F_i, $i = 1, 2$. So, in this case the shape of $p(t_j, \vartheta) = p_j(\vartheta)$ can be deduced from a nonparametric estimate of $\mathsf{E}\,\xi_j$ as a function of t_j. On Fig 1(a) the solid line represents the Nadaraya-Watson estimator for $\mathsf{E}\,\xi_j$ and the dashed line is the graph of $1 - p(t, \vartheta_0)$. In the case when the component means differ significantly, such figures allow to perform a graphical diagnostic of the concentrations model.

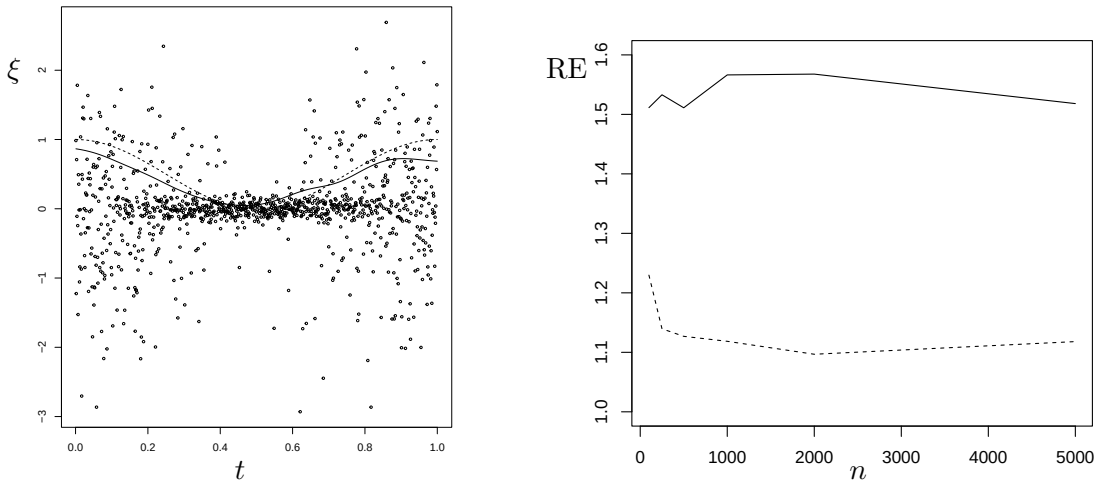
Biases and standard deviations of the three estimators considered in the experiment are presented at the Table 1. It is readily seen that for all the estimators the bias is negligible in comparison to the standard deviation. As it could be expected, the PML possesses the smallest variance for all sample sizes. But the nonparametric LS and EML estimators also show relatively good performance. On Fig 1 (b) we present graphs of relative efficiency (RE) of LS and EML w.r.t. PML, i.e. $\mathrm{RE}^{\mathrm{LS}} = \mathrm{Sd}^{\mathrm{LS}}/\mathrm{Sd}^{\mathrm{PML}}$ and $\mathrm{RE}^{\mathrm{EML}} = \mathrm{Sd}^{\mathrm{LS}}/\mathrm{Sd}^{\mathrm{PML}}$. They show that the Sd of EML is nearly 10% worse than of PNL and this grows to 20% for LS. So, only a narrow margin is left for the improvement of PML or LS in this case.

Experiment 2. Let us consider now the components different in variance only: $F_1 \sim N(0, 0.01)$, $F_2 \sim N(0, 1)$. A typical sample of size $n = 1000$ is presented on Fig. 2. The mean of ξ_j is the same for all j, but the shape of $p(t)$ can be seen from the variance (second moment)



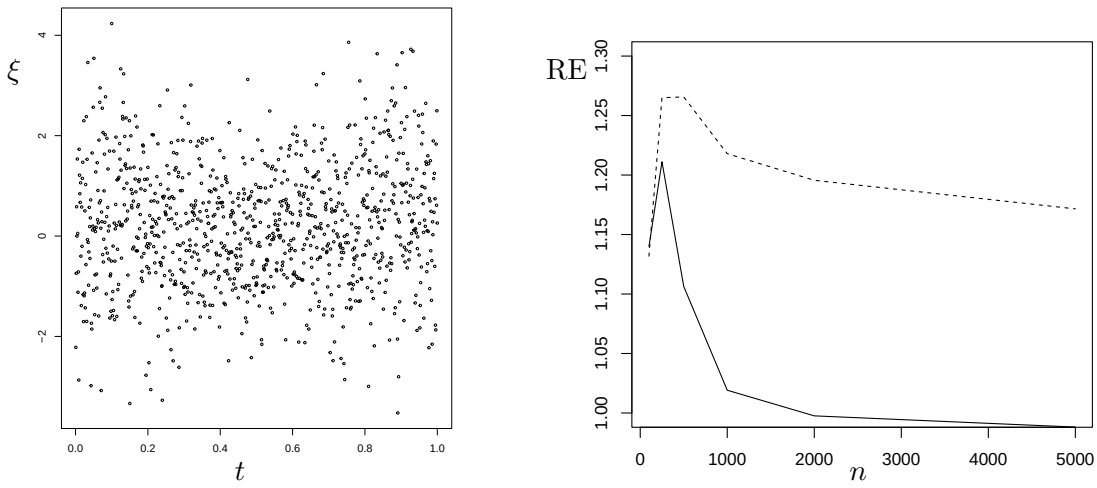
(a) (b)

Figure 1: Results of Example 1. (a) Typical data scatterplot with the second component concentration (dashed line) and the Nadaraya-Watson estimator of mean of ξ_j (solid line), (b) $\mathrm{Sd}^{LS}/\mathrm{Sd}^{PML}$ (dashed line), $\mathrm{Sd}^{EML}/\mathrm{Sd}^{PML}$ (solid line).



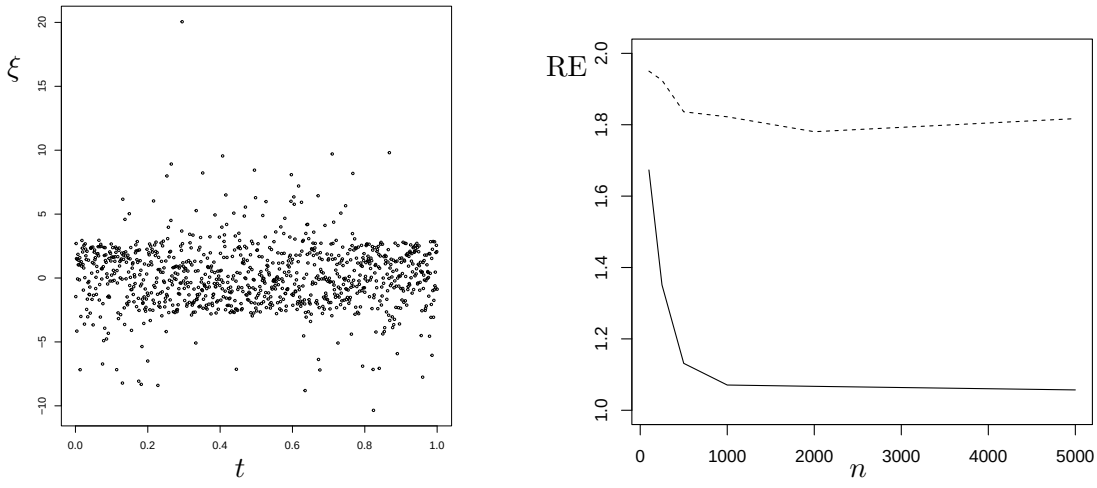
(a) (b)

Figure 2: Results of Example 2. (a) Typical data scatterplot with the second component concentration (dashed line) and the Nadaraya-Watson estimator of mean of ξ_j^2 (solid line), (b) $\mathrm{Sd}^{LS}/\mathrm{Sd}^{PML}$ (dashed line), $\mathrm{Sd}^{EML}/\mathrm{Sd}^{PML}$ (solid line).



(a) (b)

Figure 3: Results of Example 3. (a) Typical data scatterplot, (b) $\mathrm{Sd}^{LS}/\mathrm{Sd}^{PML}$ (dashed line), $\mathrm{Sd}^{EML}/\mathrm{Sd}^{PML}$ (solid line).



(a) (b)

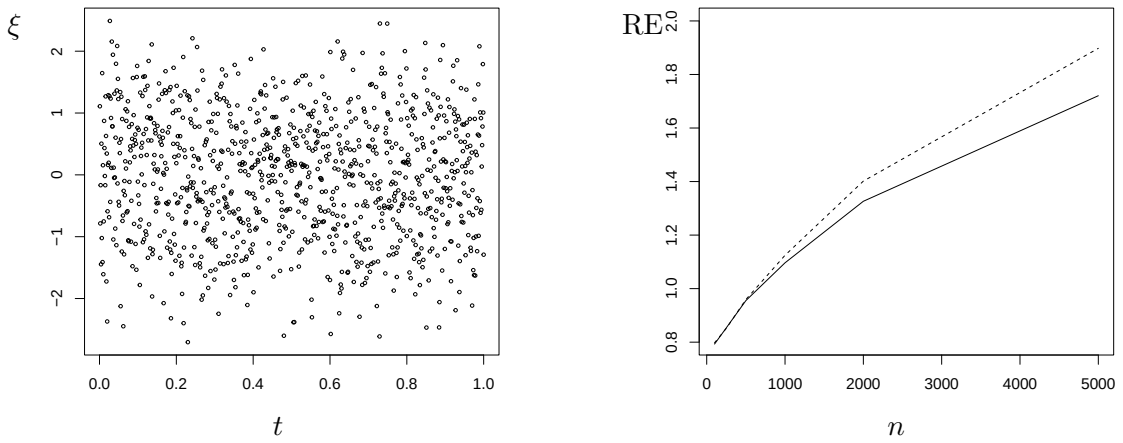
Figure 4: Results of Example 4. (a) Typical data scatterplot, (b) $\mathrm{Sd}^{LSE}/\mathrm{Sd}^{PML}$ (dashed line), $\mathrm{Sd}^{EML}/\mathrm{Sd}^{PML}$ (solid line).

of ξ_j. Results of the experiment are represented in Table 2. Here LS performs significantly better then the EML estimator and its variance is quite near to the variance of PML.

Experiment 3. Now consider the case when both the mean and the variance change. We let $F_1 \sim N(0,1)$, $F_2 \sim N(0.5,2)$. Despite the differences, the distributions of the components are much closer to each other than in the previous experiments and the changes in the series of the observations are less visible (See Fig. 3). The results of the experiment are placed in Table 3

In this experiment EML performs nearly as good as PML for sample sizes larger than 1000. The LS is slightly worse.

Experiment 4. In this experiment we consider two non-Gaussian distributions with the same mean and variance. Namely, let η be a chi-square distributed random variable with three degrees of freedom. Then F_1 is the distribution of $\eta - \mathsf{E}\,\eta$ and F_2 is the distribution of $\mathsf{E}\,\eta - \eta$. A typical example of such data is presented on Fig. 4 (a).



(a) (b)

Figure 5: Results of Example 5. (a) Typical data scatterplot, (b) $\mathrm{Sd}^{LSE}/\mathrm{Sd}^{PML}$ (dashed line), $\mathrm{Sd}^{EML}/\mathrm{Sd}^{PML}$ (solid line).

The change of components concentrations in the mixture can be observed by the observations in tail domains of their distributions. But the mean and variance of ξ_j are constant for all j.

Results of simulations are presented in Table 4 and on Fig. 4. The EML estimator shows very good performance, say, for the sample size $n = 5000$ it is only 6% worse than PML. The LS is not so efficient in this case. It shows significantly worse performance than EML and PML.

Experiment 5. In the previous experiments the changes in the observations' distributions were visible at the (t, ξ)-scatterplots. Now we consider an example of data for which this is not the case. In these data the first component is standard normal, while the second one is itself a mixture of two sub-components with mixing probabilities $p_1 = p_2 = 1/2$. The first sub-component is $N(a, s^2)$ and the second one is $N(-a, s^2)$, where $a = 1/\sqrt{2}$, $s^2 = 1/2$. It is readily seen that with these parameters the mean and variance of the second component are the same as of the first one.

A typical sample is presented on Fig. 5 (a), the results of the experiment are on Fig. 5 (b) and in the Table 5. The EML preforms slightly better then the LS. The standard deviations of the estimators are significantly worse then in the previous experiments. But these deviations decrease toward zero as n increases, so the estimators observe this "invisible" change in distribution. It is interesting to note that the PML here performs even slightly worse then LS and EML for small sample sizes.

Table 1: Results of Experiment 1

	Bias			Sd		
n	PML	LS	EML	PML	LS	EML
100	0.000354	0.000855	0.000602	0.0165	0.01891	0.0181
250	-0.000181	-0.000357	-0.000437	0.0107	0.0121	0.0115
500	0.000422	0.000479	0.000515	0.00741	0.00900	0.00839
1000	-6.92E-005	-9.85E-005	-1.33E-004	0.00526	0.00606	0.00571
2000	0.000150	0.000222	0.000160	0.00379	0.00435	0.00412
5000	7.56E-005	1.32E-004	1.21E-004	0.00238	0.00283	0.00266

Table 2: Results of Experiment 2

	Bias			Sd		
n	PML	LS	EML	PML	LS	EML
100	9.15E-4	-0.000431	-0.000952	0.0234	0.0289	0.0355
250	-2.75E-4	-2.99E-5	1.67E-4	0.0146	0.0167	0.0224
500	0.000761	0.000719	0.000921	0.0109	0.0123	0.0165
1000	8.47E-5	1.24E-004	2.45E-4	0.00761	0.00851	0.0119
2000	0.000280	0.000463	0.000666	0.00531	0.00582	0.00832
5000	0.000196	0.000237	0.000427	0.00345	0.00386	0.00524

Table 3: Results of Experiment 3

	Bias			Sd		
n	PML	LS	EML	PML	LS	EML
100	0.00388	0.00133	0.00146	0.085	0.0962	0.0968
250	-0.000514	-0.00240	-0.000868	0.0512	0.0648	0.0621
500	0.000315	-0.000117	0.000781	0.0366	0.0463	0.0405
1000	0.000125	-0.000781	-0.000305	0.0249	0.0303	0.0254
2000	8.24E-5	1.54E-4	2.69E-4	0.0182	0.0218	0.0182
5000	0.000501	0.000358	0.000520	0.0116	0.0136	0.01146

Table 4: Results of Experiment 4

	Bias			Sd		
n	PML	LS	EML	PML	LS	EML
100	-0.00179	0.00829	0.00976	0.0496	0.0939	0.0828
250	-0.00257	0.00159	0.000374	0.03177	0.0615	0.0454
500	0.000861	0.00125	0.000532	0.0224	0.0402	0.0247
1000	-0.000217	0.000225	-5.80E-5	0.0157	0.0277	0.0168
2000	-1.21E-4	-3.47E-4	-1.47E-005	0.0109	0.0200	0.0117
5000	8.81E-5	-2.41E-4	1.20E-004	0.00698	0.0129	0.00744

Table 5: Results of Experiment 5

	Bias			Sd		
n	PML	LS	EML	PML	LS	EML
100	-0.00709	0.00046	0.00203	0.17836	0.14227	0.14142
250	0.00417	0.00498	0.00468	0.16349	0.13921	0.13953
500	0.00442	-0.00329	-0.00331	0.14759	0.14157	0.14098
1000	-0.00298	-0.00428	-0.00511	0.12515	0.14074	0.13722
2000	0.00214	0.00557	0.00510	0.09007	0.12624	0.11946
5000	0.00061	-0.00658	-0.00548	0.05608	0.10642	0.09649

6. Concluding remarks

We considered two nonparametric approaches to the estimation of concentrations parameter in two-component mixture: least squares and empirical maximum likelihood. Parametric

maximum likelihood estimation, which needs much more information of the components' distributions, was considered as a benchmark of possible estimator efficiency. Both EML and LS estimators show sufficiently good performance in all the experiments. The LS estimator performed better then the EML one for components with different variance. The EML estimator outperformed the LS one for nongaussian components with asymmetric tails. Throughout the study the biases of the estimators were negligible in comparison to their variances. In most experiments the variances of EML and LS was quite near to PML, therefore, one should not expect a significant improvement in their performance by more deliberate choice of tuning parameters. Of course, these observations are made by a restricted number of experiments with only one simple model of concentrations. More experiments should be made to derive more reliable conclusions.

Let us note some ways of the estimation technique improvement. LS estimation can be improved by more deliberate choice of the integration measure π for the criterion $\hat{R}_n(\alpha)$ in (4). In the EML estimator one may try to choose more accurately the bandwidth h and the threshold t in the criterion $\hat{l}_n(\alpha)$ in (13). And a combination of LS and EML can be considered in which the LS is used to derive a pilot estimator which is then improved by a shift in direction of empirical likelihood maximization. Further work is needed to analyze efficiency of these proposals.

Appendix

Proof of Lemma. Let

$$\hat{U}_n(x,\alpha) = \frac{1}{n}\sum_{j=1}^{n} u_j(\alpha)\mathbf{1}\{\xi_j < x\} \text{ and } \bar{U}_n(x,\alpha) = \mathsf{E}\,\hat{U}_n(x,\alpha).$$

By the second part of theorem 2.4.3 in Maiboroda (2003), there exists a random variable $\tilde{\Lambda} < \infty$ such that

$$\sup_x |\hat{U}_n(x,\alpha) - \bar{U}_n(x,\alpha)| \le \tilde{\Lambda}\sqrt{\frac{\log n}{n}}\Big[\sup_{1\le j\le n}|u_{j;n}(\alpha)| + \sum_{j=1}^{n-1}|u_{j+1;n}(\alpha) - u_{j;n}(\alpha)|\Big]$$

for all $n = 1, 2, \dots$ (This is a corollary from the Vapnik — Chervonenkis inequality for observations from MVC).

Assumption 2 of the Lemma with $0 \le p_{j;n}(\alpha) \le 1$ imply

$\sup_{1\le j\le n}|u_{j;n}(\alpha)| \le 2/\Delta^-$. Similarly $\sum_{j=1}^{n-1}|u_{j+1;n}(\alpha) - u_{j;n}(\alpha)| \le 2V/\Delta^-$.

Observe that $\sup_{x,\alpha}|\hat{U}_n(x,\alpha)| \le 2/\Delta^-$ and $\sup_{x,\alpha}|\bar{U}_n(x,\alpha)| \le 2/\Delta^-$.

So

$$|\hat{R}_n(\alpha) - R_n(\alpha)| = \left|\int_{\mathbb{R}} \Big(\hat{U}_n(x,\alpha)\Big)^2 - \Big(\bar{U}_n(x,\alpha)\Big)^2 \pi(dx)\right|$$

$$\le \int_{\mathbb{R}} |\hat{U}_n(x,\alpha) + \bar{U}_n(x,\alpha)| \times |\hat{U}_n(x,\alpha) - \bar{U}_n(x,\alpha)|\pi(dx)$$

$$\le \frac{4}{\Delta^-}\tilde{\Lambda}\sqrt{\frac{\log n}{n}}\left[\frac{2}{\Delta^-} + \frac{2V}{\Delta^-}\right]\pi(\mathbb{R}).$$

This is just the statement of the Lemma. □

Proof of Theorem. Let $\lambda_n = \Lambda\sqrt{\log n/n}$, where Λ is defined in the Lemma.

By the definition of $\hat{\vartheta}_n$, $\hat{R}_n(\hat{\vartheta}_n) \ge \hat{R}_n(\vartheta)$. So, by the Lemma,

$$R_n(\hat{\vartheta}_n) + \lambda_n \ge \hat{R}_n(\hat{\vartheta}_n) \ge \hat{R}_n(\vartheta) \ge R_n(\vartheta) - \lambda_n. \tag{15}$$

Assumptions 4-5 of the Theorem imply that for any $\delta > 0$, there exist $\varepsilon > 0$ and n_0, such that

$$R_n(\vartheta) \geq R_n(\alpha) + \varepsilon \tag{16}$$

for all $n > n_0$ and all α, such that $|\vartheta - \alpha| > \delta$.

Combining (15) with (16) we obtain that $|\hat{\vartheta}_n - \vartheta| < \delta$ for all n, such that $n > n_0$ and $2\lambda_n < \varepsilon$. Since $\lambda_n \to 0$ a.s. this implies $\hat{\vartheta}_n \to \vartheta$ a.s. as $n \to \infty$. □

References

Autin F, Pouet C (2011). "Test on the Components of Mixture Densities." *Statistics & Risk Modelling*, **28**(4), 389–410. doi:10.1524/strm.2011.1065. URL https://www.degruyter.com/document/doi/10.1524/strm.2011.1065/html.

Celeux G, Fruhwirth-Schnatter S, Robert CP (eds.) (2018). *Handbook of Mixture Analysis*. Chapman and Hall/CRC. ISBN 978-1-4987-6381-3.

Grün B, Leisch F (2006). "Fitting Finite Mixtures of Linear Regression Models with Varying & Fixed Effects in R." In A Rizzi, M Vichi (eds.), *Compstat 2006 - Proceedings in Computational Statistics*, pp. 853–860. ISBN 978-3790817089.

Ibragimov IA, Has'minskii RZ (1981). *Statistical Estimation: Asymptotic Theory*. Springer-Verlag. ISBN 9780387905235.

Maiboroda R, Kubaichuk O (2005). "Improved Estimators for Moments Constructed form Observations of a Mixture." *Theory of Probability and Mathematical Statistics*, **70**, 83–92. doi:10.1090/S0094-9000-05-00642-3.

Maiboroda R, Miroshnichenko V, Sugakova O (2022). "Jackknife for Nonlinear Estimating Equations." *Modern Stochastics: Theory and Applications*, **9**(4), 377 – 399. doi:10.15559/22-VMSTA208.

Maiboroda R, Sugakova O (2012). "Statistics of Mixtures with Varying Concentrations with Application to DNA Microarray Data Analysis." *Nonparametric statistics*, **24**(1), 201–215. doi:10.1080/10485252.2011.630076.

Maiboroda RE (2002). "The Least Square Estimate for Parameters of Concentrations of Varying Mixtures I. The Consistency." *Theory of Probability and Mathematical Statistics*, **64**, 53–69. URL https://probability.knu.ua/tims/issues-new/64/PDF/12.pdf.

Maiboroda RE (2003). *Statistical Analysis of Mixtures*. Kyiv University Publishers. (in Ukrainian) ISBN 966-594-521-1.

McLachlan GJ, Peel D (2000). *Finite Mixture Models*. Wiley. doi:10.1002/0471721182. ISBN:9780471721185.

McNicholas PD (2017). *Mixture Model-based Classification*. Chapman and Hall CRC. doi:10.1201/9781315373577. ISBN:9781315373577.

Owen AB (2001). *Empirical Likelihood*. Chapman and Hall CRC. doi:10.1201/9781420036152. ISBN:9780429128998.

Pidnebesna A, Fajnerová I, Horáček J, Hlinka J (2023). "Mixture Components Inference for Sparse Regression: Introduction and Application for Estimation of Neuronal Signal from fMRI BOLD." *Applied Mathematical Modelling*, **116**, 735–748. doi:10.1016/j.apm.2022.11.034.

Qin J, Lawless JF (1994). "Empirical Likelihood and General Estimating Equations." *The Annals of Statistics*, **22**(1), 300–325. doi:10.1214/aos/1176325370.

Schlattmann P (2009). *Medical Applications of Finite Mixture Models*. Springer-Verlag. doi:10.1007/978-3-540-68651-4. ISBN: 978-3-540-68650-7.

Sugakova OV (1999). "Asymptotics of a Kernel Estimate for the Density of a Distribution Constructed from Observations of a Mixture with Varying Concentration." *Theory of Probability and Mathematic Statistics*, **59**, 161–171. URL https://probability.knu.ua/tims/issues-new/59/PDF/21.pdf.

Vexler A, Boca Raton JY (2018). *Empirical Likelihood Methods in Biomedicine and Health*. Chapman and Hall CRC. ISBN: 9781032401812.

Affiliation:

Rostyslav Maiboroda
Taras Sevchenko National
University of Kyiv
64/13, Volodymyrska Street, City of Kyiv, Ukraine, 01601
E-mail: rostmaiboroda@knu.ua
URL: https://probability.knu.ua/index.php?page=userinfo&person=mre&lan=en

Austrian Journal of Statistics
published by the Austrian Society of Statistics
Volume 54
2025
http://www.ajs.or.at/
http://www.osg.or.at/
Submitted: 2024-05-16
Accepted: 2024-08-04