Теорія Ймовір. та Матем. Статист.
Вип. 91, 2014

Theor. Probability and Math. Statist.
No. 91, 2015, Pages 29–41
http://dx.doi.org/10.1090/tpms/964
Article electronically published on February 3, 2016

ADAPTIVE ESTIMATION FOR A SEMIPARAMETRIC MODEL OF MIXTURE

UDC 519.21

O. V. DORONIN

Abstract. A model of mixture with varying concentrations is considered. It is assumed that the first K of M, $1 \le K \le M$, components of the mixture are parameterized. A technique of the adaptive semiparametric estimation is developed by using the generalized estimating equations. It is proved that the estimators are consistent and asymptotically normal.

1. Introduction

A semiparametric model of mixture can be applied when analyzing the data obtained from an opinion poll. Consider an example of a parliament election with candidates representing two parties. Every citizen may either vote one of the candidates or have another choice, say a voter has a right to claim that he/she does not support either of the candidates. According to the choice made, every voter O is included into one of the subpopulations Ξ_m, $m = 1, \dots, M$, of the sampling population Ξ ($M = 3$ in the case under consideration). The choice of a given voter is not publically known but one can estimate the probability $p^m(O)$ that he/she belongs to a subpopulation Ξ_m, $m = 1, \dots, M$, if one knows the number of votes in favor of every admissible choice.

A researcher may be interested in studying a connection between a particular choice of a voter O and some other characteristic $\xi(O)$ of O (say, income level, social status, marital status, spirituals, etc.).

For definiteness, let $\xi(O)$ be the life satisfaction corresponding to a voter O. As described above, the first two components of the mixture are formed by those citizens who voted candidates representing the first and second parties, respectively. It is natural to assume that the first two components have a normal distribution with different means and equal variances. In the general case, one assumes that the distributions H_m of a characteristic $\xi(O)$ in an m^{th} subpopulation, $m = 1, \dots, K$, belong to a known parametric family of distributions equipped with an Euclidean parameter $t \in \Theta \subset \mathbb{R}^d$. The true value of the parameter is denoted by ϑ and assumed to be unknown. The aim of a researcher is to estimate ϑ. The distributions H_m of the characteristics for other subpopulations, $m = K + 1, \dots, M$, are assumed to be completely unknown.

In the current paper, we follow the technique of adaptive estimators for the semiparametric model of mixture discussed above. This technique can briefly be described as follows: first we estimate the parameter ϑ with the help of a certain pilot estimator $\tilde{\vartheta}_N$, where N denotes the number of observations $\xi_{j;N}$ belonging to the population Ξ_N. Then the estimator $\tilde{\vartheta}_N$ is improved by using an estimator $\hat{\vartheta}_N$ being optimal in a certain sense.

2010 *Mathematics Subject Classification.* Primary 62G05, 62G20, 62F12; Secondary 62P25, 62G30.
Key words and phrases. Adaptive estimation, model of mixture, generalized estimating equations.

This paper is a continuation of [2] where a lower bound is found for the dispersion matrix of the method of generalized estimating equations.

We study the model of mixture with varying concentrations. A discussion of such models can also be found in [3, 4, 5, 6, 9, 12, 14, 15]. The classical model of mixture where the probabilities p^m are the same for all observations $\xi_{j;N}$, $j = 1, \dots, N$, as well as examples of its application is considered in [16, 20]. Some special classes of the model of mixture are investigated in [10, 11, 13].

In Section 2, we provide some auxiliary results. The adaptive estimation technique is discussed in Section 3. Results of simulation are placed in Section 4.

2. A survey of previous results

Let observations $\xi_{j;N} := \xi(O_{j;N})$, $j = 1, \dots, N$, belong to some metric space $\mathfrak{X}$ equipped with a σ-finite measure μ defined on the Borel σ-algebra $\mathfrak{B}(\mathfrak{X})$. By

$$F_m(A) := \mathsf{P}[\xi(O) \in A \mid \operatorname{ind}(O) = m],$$

we denote the distribution of the m^{th} component (in other words, F_m is the distribution of $\xi(O)$ given the object O belongs to the m^{th} subpopulation), $m = 1, \dots, M$, $A \in \mathfrak{B}(\mathfrak{X})$. We assume that the family of probabilities $(p^m_{j;N})_{j=1,\dots,N,m=1,\dots,M}$ is known.

For the sake of convenience, we denote by $\mathbb{O}_d$ the zero vector of the vector space $\mathbb{R}^d$. For a real $m \times n$ matrix A, we write $A \in \mathbb{R}^{m\times n}$.

Let $p_{\cdot;N} := (p^i_{j;N})_{j=1,\dots,N,i=1,\dots,M} \in \mathbb{R}^{N\times M}$ be the matrix constructed from the family of concentrations. Then the Gram matrices (provided they exist) are denoted by

$$\Gamma_N := \frac{1}{N} p^T_{\cdot;N} p_{\cdot;N}, \qquad \Gamma := \lim_{N\to\infty} \Gamma_N. \tag{1}$$

The problem of minimization of the maximal variance of weighted empirical distributions

$$\hat{F}_{i;N}(A) := \frac{1}{N} \sum_{j=1}^{N} a^i_{j;N} \mathbb{I}_{\{\xi_{j;N} \in A\}}$$

is considered in [15]. The minimum in [15] is evaluated among all $A \in \mathfrak{B}(\mathfrak{X})$ and all distributions $F_i(A)$ under the assumption that $\hat{F}_{i;N}(A)$ is unbiased.

The minimax weight coefficients are given by

$$a^i_{\cdot;N} := p_{\cdot;N} \Gamma^+_N e_i, \tag{2}$$

where Γ^+_N denotes the Moore–Penrose pseudo-inverse matrix for Γ_N and where $e_i := (\mathbb{I}_{\{k=i\}})_{k=1,\dots,M} \in \mathbb{R}^M$.

Note that the weight coefficients $a^i_{j;N}$ may be negative and thus the estimator is misleading. To avoid this disadvantage, improved empirical distribution functions $\hat{F}^+_{i;N}$ are considered in [5] to which non-negative weight coefficients $\tilde{a}^i_{j;N}$ correspond:

$$\hat{F}^+_{i;N}(x) := \min\left[1, \sup_{y<x} \hat{F}_{i;N}(y)\right] = \frac{1}{N} \sum_{j=1}^{N} \tilde{a}^i_{j;N} \mathbb{I}_{\{\xi_{j;N} < x\}}. \tag{3}$$

Weighted moments for measurable functions $g^i(x;t)\colon \mathfrak{X} \times \Theta \to \mathbb{R}^d$ are denoted by

$$\hat{g}^i_N(t) := \frac{1}{N} \sum_{j=1}^{N} a^i_{j;N} g^i(\xi_{j;N}; t), \qquad i = 1, \dots, K. \tag{4}$$

Definition 2.1. We say that a random sequence ϕ_N occasionally coincides with a random sequence ψ_N if $\phi_N = \psi_N$ almost surely starting with some random number N.

Definition 2.2. The GEE (method of generalized estimating equations) estimator $\hat{\vartheta}_N$ for the parameter ϑ (see [2]) is defined as a measurable function of the sample $\xi_{1;N}, \dots, \xi_{N;N}$ such that

$$\sum_{k=1}^{K} \hat{g}_N^k(\hat{\vartheta}_N) = \mathbb{O}_d \tag{5}$$

occasionally.

Definition 2.3. Let a sequence of estimators ζ_N, $N \geq 1$, be asymptotically normal, that is, $\sqrt{N} \cdot \zeta_N$ converges in distribution to some random variable with the distribution $\mathcal{N}(a, \Sigma)$. In this case, Σ is called the dispersion coefficient or dispersion matrix if ζ_N is a random number or random matrix, respectively.

Example 2.4. The example considered above can be described in the framework of the model of mixture of three components with varying concentrations. Let all three components be normally distributed (in other words, their distributions are $\mathcal{N}(m_1, \sigma^2)$, $\mathcal{N}(m_2, \sigma^2)$, and $\mathcal{N}(m_3, \sigma_3^2)$, respectively). Let the true parameters of the components be $m_1 = -3$, $m_2 = 2$, $\sigma = 2$, $m_3 = 0$, and $\sigma_3 = 2$ and let the concentrations of the components be $p_{j;N}^{\cdot} := \frac{1}{S_{j;N}}(u_{j;N}^1, u_{j;N}^2, 2u_{j;N}^3)^T$, where $u_{j;N}^i$ are independent uniformly distributed in the interval $[0,1]$ random variables, $i = 1, 2, 3$, $S_{j;N} := u_{j;N}^1 + u_{j;N}^2 + 2u_{j;N}^3$, $j = 1, \dots, N$. Let the distribution of the third component be totally unknown. Then the true value of the parameter can be written as $t = (m_1, m_2, \sigma)^T$. The moment estimator $\tilde{\vartheta}_N$ for ϑ is defined by

$$\tilde{m}_{i;N} := \frac{1}{N} \sum_{j=1}^{N} a_{j;N}^i \xi_{j;N}, \qquad i = 1, 2,$$

$$\tilde{\sigma}_N := \sqrt{\frac{1}{2} \sum_{k=1}^{2} \frac{1}{N} \sum_{j=1}^{N} a_{j;N}^k (\xi_{j;N} - \tilde{m}_{k;N})^2}$$

(provided the expression under the radical is non-negative). The improved moment estimator $\tilde{\vartheta}_N^{impr}$ is defined similarly by using the improved weight coefficients $\tilde{a}_{j;N}^i$ (see (3)). The asymptotic dispersion coefficients of the estimators evaluated in [2] are

$$d^m(\tilde{m}_{1;N}) = 281.204, \qquad d^m(\tilde{m}_{2;N}) = 244.921, \quad \text{and} \quad d^m(\tilde{\sigma}_N) = 217.29.$$

The lower bounds found in [2] are $d^*(\hat{m}_1) = 57.24$, $d^*(\hat{m}_2) = 60.52$, $d^*(\hat{\sigma}) = 19.21$.

It is shown in [2] that one can achieve a substantial decrease of the dispersion coefficients of the estimator by choosing appropriate estimating functions

$$g^k(x; t), \qquad k = 1, \dots, K,$$

as compared to the method of using some simple functions.

3. Adaptive estimation

The optimal functions $g^{*;k}(x;t)$, $k = 1, \dots, K$, where the lower bound for the dispersion matrix of estimators is attained are found in [2]. Note however that these functions depend on the densities of unknown components of the mixture and one cannot use these functions for real estimation of parameters. Thus we propose to use an adaptive technique for the estimation similar to that considered in [15] for the case where the parametric model is specified only for the first component of the mixture.

Let some basis functions $u^k(x;t)\colon \mathfrak{X}\times\Theta\to\mathbb{R}^{L_k}$, $k=1,\dots,K$, be given. Consider the estimating functions $g^k(x;t)$, $k=1,\dots,K$, of the following form

$$g^k(x;t;B^k) := B^k\cdot u^k(x;t), \tag{6}$$

where $B^k\in\mathbb{R}^{d\times L_k}$, $k=1,\dots,K$, are matrices of the coefficients. By $B^{*;k}$, $k=1,\dots,K$, we denote the matrices B^k that minimize the dispersion matrix of the estimator in the class of all GEE-estimators with the estimating functions given by (6) (see a remark to Lemma 3.3). We introduce a partial order for matrices in the sense of Loewner, namely we write for matrices A and B that $A\ge B$ if $A-B$ is a non-negative definite matrix. The matrices $B^{*;k}$ depend on ϑ and on the densities of unknown distributions and thus a certain estimator $\hat{B}_N^k$ has to substitute them in statistical procedures.

The estimator $\hat{\vartheta}_N$ defined by (5) can be constructed as a solution of the equation

$$\sum_{k=1}^{K}\hat{g}_N^k\left(t;\hat{B}_N^k\right)=\mathbb{O}_d \tag{7}$$

with respect to $t\in\Theta$. In the general case, equation (7) is not easy to solve. Similarly, it is not easy to decide whether this equation has a unique solution. Therefore one cannot be convinced in the general case that the estimator $\hat{\vartheta}_N$ exists and is consistent.

We propose to start with a certain $\sqrt{N}$-consistent "pilot" estimator $\tilde{\vartheta}_N$ for ϑ (as $\tilde{\vartheta}_N$, one can take, for example, the method of moments estimator). Put

$$V\left(t;\{B^k\}\right) := \sum_{k=1}^{K}B^k\int\frac{\partial}{\partial t^T}u^k(x;t)\,F_k(dx;t)\in\mathbb{R}^{d\times d}. \tag{8}$$

Then an approximation of equation (7) is of the following form:

$$\mathbb{O}_d=\sum_{k=1}^{K}\hat{g}_N^k(t;\hat{B}_N^k)\approx\sum_{k=1}^{K}\hat{B}_N^k\cdot\hat{u}_N^k(\tilde{\vartheta}_N)+V(\tilde{\vartheta}_N;\{\hat{B}_N^k\})\cdot(t-\tilde{\vartheta}_N). \tag{9}$$

If the matrices $\hat{B}_N^k$ are chosen such that $V\big(\tilde{\vartheta}_N;\{\hat{B}_N^k\}\big)\approx\mathbb{I}_{d\times d}$, then the approximate solution of equation (7) is given by

$$\check{\vartheta}_N := \tilde{\vartheta}_N-\sum_{k=1}^{K}\hat{B}_N^k\cdot\hat{u}_N^k(\tilde{\vartheta}_N). \tag{10}$$

We choose $\check{\vartheta}_N$ as an adaptive estimator for ϑ. We show in what follows that its dispersion matrix is the same as the dispersion matrix $D[\{B^{*;k}\}]$ considered in Lemma 3.3.

Remark 3.1. An arbitrary measurable function $g^k(x;t)$ can be approximated with a given accuracy by functions $B^k u^k(x;t)$ if the basis $u^k(x;t)\in\mathbb{R}^{L_k}$, $k=1,\dots,K$, is sufficiently rich. Thus it is reasonable to conclude that the dispersion matrix of the estimator $\check{\vartheta}_N$ defined as a solution of (7) is close to the lower bound found in [2] if $u^k(x;t)$ are chosen appropriately.

Remark 3.2. The matrix $V(\tilde{\vartheta}_N;\{\hat{B}_N^k\})$ should necessarily be non-degenerate in a neighborhood $\mathfrak{U}\subset\Theta$ of ϑ. Indeed, let $V(\tilde{\vartheta}_N;\{\hat{B}_N^k\})$ be degenerate (this matrix is degenerate in a neighborhood $\tilde{\vartheta}_N$ if $u^k(x;t)$ are continuous). Then there exists a vector $v\in\mathbb{R}^d$, $v\ne 0$, such that $V(\tilde{\vartheta}_N;\{\hat{B}_N^k\})\cdot v=\mathbb{O}_d$. If there exists a solution $\hat{\vartheta}_N$ of approximation (9), then $\hat{\vartheta}_N+cv$ also is a solution of this approximation for an arbitrary $c\in\mathbb{R}$. This means that the bias of the estimator $\hat{\vartheta}_N$ can be made as large as one wants.

3.1. **The asymptotic behavior of the adaptive estimator.** Assume that the distribution functions $F_1, \dots, F_M$ are absolutely continuous with respect to the measure μ. We denote the probability density of every component of the mixture by

$$\begin{aligned} f_k(\cdot;t) &:= \frac{dF_k(\cdot;t)}{d\mu(\cdot)}, \qquad f_k(\cdot) := f_k(\cdot,\vartheta), \qquad k=1,\dots,K, \\ f_k(\cdot) &:= \frac{dF_k(\cdot)}{d\mu(\cdot)}, \qquad k=K+1,\dots,M. \end{aligned} \tag{11}$$

For the sake of convenience, we introduce the matrix formed from all basis functions $u^k(x;t)$, $k=1,\dots,K$:

$$U(x;t) := \operatorname{Diag}(u^{\cdot}(x;t)) := \begin{pmatrix} u^1(x;t) & \mathbb{O}_{L_1} & \cdots & \mathbb{O}_{L_1} \\ \mathbb{O}_{L_2} & u^2(x;t) & \cdots & \mathbb{O}_{L_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O}_{L_K} & \mathbb{O}_{L_K} & \cdots & u^K(x;t) \end{pmatrix} \in \mathbb{R}^{L\times K}.$$

Its expectation is denoted by

$$\bar{U}^m(t) := \begin{cases} \int U(x;t) f_m(x;t)\,\mu(dx), & m=1,\dots,K, \\ \int U(x;t) f_m(x)\,\mu(dx), & m=K+1,\dots,M, \end{cases} \in \mathbb{R}^{L\times K}.$$

Let $\{b_{j;N}\}_{j=1,\dots,N}$ and $\{c_{j;N}\}_{j=1,\dots,N}$ be some families of coefficients. Define the averaging operator $\langle\cdot\rangle_N$ and corresponding arithmetic operations by

$$\langle b_{\cdot;N}\rangle_N := \frac{1}{N}\sum_{j=1}^N b_{j;N}, \qquad \langle b_{\cdot;N}\cdot c_{\cdot;N}\rangle_N := \frac{1}{N}\sum_{j=1}^N b_{j;N}\cdot c_{j;N}. \tag{12}$$

Let random variables η_m have the distribution functions F_m, $m=1,\dots,M$. The following notation is convenient when evaluating the expressions like $\operatorname{Cov}[\hat{u}^k_N, \hat{u}^l_N]$:

$$\begin{aligned} \alpha_{r,s;N} &:= \left(\alpha^{k,l}_{r,s;N}\right)_{k,l=1,\dots,K} \\ &:= \left(\langle a^k_{\cdot;N} a^l_{\cdot;N} p^r_{\cdot;N} p^s_{\cdot;N}\rangle_N\right)_{k,l=1,\dots,K} \in \mathbb{R}^{K\times K}, \qquad r,s=1,\dots,M; \\ \beta_{m;N} &:= \left(\beta^{k,l}_{m;N}\right)_{k,l=1,\dots,K} \\ &:= \left(\langle a^k_{\cdot;N} a^l_{\cdot;N} p^m_{\cdot;N}\rangle_N\right)_{k,l=1,\dots,K} \in \mathbb{R}^{K\times K}, \qquad m=1,\dots,M; \end{aligned}$$

and

$$\begin{aligned} \alpha_{r,s} &:= \lim_{N\to\infty} \alpha_{r,s;N} \in \mathbb{R}^{K\times K}, \qquad r,s=1,\dots,M; \\ \beta_m &:= \lim_{N\to\infty} \beta_{m;N} \in \mathbb{R}^{K\times K}, \qquad m=1,\dots,M. \end{aligned} \tag{13}$$

The matrices $\alpha_{r,s;N}$ are expressed in terms of $\mathsf{E}[u^k(\eta_r)]\,\mathsf{E}[u^l(\eta_s)]^T$, while the matrices $\beta_{m;N}$ are expressed in terms of $\mathsf{E}[u^k(\eta_m)u^l(\eta_m)^T]$.

Using the notation

$$R(x;t) := \sum_{m=1}^K \beta_m f_m(x;t) + \sum_{m=K+1}^M \beta_m f_m(x), \tag{14}$$

the expression $\sum_{m=1}^M \mathsf{E}[U(\eta_m)\beta_m U(\eta_m)^T]$ reduces to $\int U(x)R(x)U(x)^T\,\mu(dx)$.

Let

$$L := \sum_{k=1}^K L_k,$$

where L_k is the dimension of $u^k(x;t)$, $k=1,\dots,K$. Let

$$B := (B^1 \cdots B^K) \in \mathbb{R}^{d\times L}$$

be the matrix constructed from all unknown coefficients.

The dispersion matrix of the vector $(\hat{u}_N^1(\vartheta),\dots,\hat{u}_N^K(\vartheta))^T$ is denoted by $\mathbb{Z}(\vartheta)$. Then $B\mathbb{Z}(\vartheta)B^T$ is the dispersion matrix of $\sum_{k=1}^K B^k\hat{u}_N^k(\vartheta)$. We also put

$$\mathbb{Z}_1(\vartheta) := \sum_{r,s=1}^{M} \bar{U}(\vartheta)^r \alpha_{r,s} \bar{U}^s(\vartheta)^T \in \mathbb{R}^{L\times L};$$

$$\mathbb{Z}_2(\vartheta) := \int U(x;\vartheta)R(x;\vartheta)U(x;\vartheta)^T\,\mu(dx) \in \mathbb{R}^{L\times L};$$

$$\mathbb{Z}(\vartheta) := \mathbb{Z}_2(\vartheta) - \mathbb{Z}_1(\vartheta) \in \mathbb{R}^{L\times L}.$$

Similarly to [2], let the following assumptions hold for functions $u^k(x;t)$, $k=1,\dots,K$, non-zero matrices B^k, and some open neighborhood $\mathfrak{U}\subset\Theta$ of ϑ.

(a1) The functions $u^1(x;t),\dots,u^K(x;t)$ are differentiable with respect to t for almost all x (mod μ), $t\in\mathfrak{U}$.

(a2) There exists a number $\delta>0$ such that

$$\int \sup_{t\in\mathfrak{U}} \left\| \frac{\partial}{\partial t^T} u^k(x;t) \right\|^{1+\delta} F_i(dx) < \infty, \qquad i=1,\dots,M, \quad k=1,\dots,K.$$

(a3) $\int \left\|u^k(x;\vartheta)\right\|^2 F_i(dx) < \infty$, $i=1,\dots,M$, $k=1,\dots,K$.

(b1) $B^k \int u^k(x;\vartheta)F_k(dx;\vartheta) = \mathbb{O}_d$, $k=1,\dots,K$.

(b2) The matrix $V(t;\{B^k\})$ is finite and non-degenerate for $t\in\mathfrak{U}$ (see (8)).

Lemma 3.3. *Let*

(i) $\tilde{\vartheta}_N$ *be an* $\sqrt{N}$*-consistent estimator of* ϑ;

(ii) *conditions* (a1)–*(a3) hold for the functions* $u^k(x;t)$, $k=1,\dots,K$;

(iii) *conditions* (b1)–*(b2) hold for the matrices* B^k, $k=1,\dots,K$;

(iv) *the limit matrices* $\alpha_{r,s}$ *and* β_m *defined by* (13) *exist for* $m,r,s=1,\dots,M$;

(v) *the adaptive estimator is defined by*

$$\hat{\vartheta}_N := \tilde{\vartheta}_N - V(\tilde{\vartheta}_N;\{B^k\})^{-1}\sum_{k=1}^{K} B^k\hat{u}_N^k(\tilde{\vartheta}_N).$$

Then

$$\sqrt{N}\cdot(\hat{\vartheta}_N-\vartheta) \xrightarrow{W_\vartheta} \mathcal{N}(\mathbb{O}_d, D[\{B^k\}]),$$

where $D[\{B^k\}] := V^{-1}B\mathbb{Z}(\vartheta)B^T(V^{-1})^T$, $V := V(\vartheta;\{B^k\})$.

Proof. The proof is similar to that of Theorem 3.2 in [2]. □

Remark 3.4. The dispersion matrix of the GEE-estimator is found in Theorem 3.2 of [2]. This matrix coincides with the dispersion matrix $D[\{B^k\}]$ obtained in Lemma 3.3 if one chooses $g^k(x;t)=B^k u^k(x;t)$, $k=1,\dots,K$, as estimating functions.

3.2. **Optimal coefficients** B^***.** The derivative of the density of a component with respect to the parameter is denoted by

$$f_k'(x;t) := \frac{\partial f_k(x;t)}{\partial t^T} \in \mathbb{R}^d, \qquad k=1,\dots,K. \tag{15}$$

We introduce matrices $\mathbb{V}_1^k(t)$ to write the condition $V(\vartheta;\{B^k\}) = \mathbb{I}_{d\times d}$ (see (8)) in a simpler form, namely

$$\mathbb{V}_1^k(t) := \int u^k(x;t) f_k'(x;t)^T \,\mu(dx) \in \mathbb{R}^{L_k\times d}, \qquad k=1,\dots,K.$$

It is explained in the proof of Theorem 3.5 why the matrices are written in this form.

Denote the moments of $u^k(x;t)$ by

$$\mathbb{V}_2^k(t) := \int u^k(x;t) f_k(x;t)\,\mu(dx) \in \mathbb{R}^{L_k}, \qquad k=1,\dots,K.$$

Put

$$\mathbb{V}^k(t) := \big(\mathbb{V}_1^k(t)\mathbb{O}_{L_k\times(k-1)}\mathbb{V}_2^k(t)\mathbb{O}_{L_k\times(K-k)}\big) \in \mathbb{R}^{L_k\times(d+K)}, \qquad k=1,\dots,K;$$

$$\mathbb{V}(t) := \begin{pmatrix} \mathbb{V}^1(t) \\ \vdots \\ \mathbb{V}^K(t) \end{pmatrix} \in \mathbb{R}^{L\times(d+K)};$$

$$E_0 := \big(\mathbb{I}_{d\times d}\mathbb{O}_{d\times K}\big) \in \mathbb{R}^{d\times(d+K)}.$$

Note that conditions **(b1)** and $V(\vartheta;\{B^k\}) = \mathbb{I}_{d\times d}$ can be written as $B\mathbb{V}(\vartheta) = -E_0$ (see the proof of Theorem 3.5 for more detail).

Finally, the optimal coefficients for $u^k(x;\vartheta)$, $k=1,\dots,K$, are given by

$$B^* := -E_0\big[\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1}\mathbb{V}(\vartheta)\big]^{-1}\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1} \in \mathbb{R}^{d\times L}.$$

Theorem 3.5. *Let all the assumptions of Lemma* 3.3 *hold. Further let*

(i) *the derivatives of the densities defined by* (15) *exist and be continuous*;

(ii) *the matrices* $\mathbb{Z}(\vartheta)$ *and* $\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1}\mathbb{V}(\vartheta)$ *exist and be non-degenerate.*

Then $D[\{B^k\}] \ge D[\{B^{*;k}\}]$ *in the sense that the matrix* $(D[\{B^k\}] - D[\{B^{*;k}\}])$ *is non-negative definite.*

Proof. Restriction imposed on the test functions.

Put $g^k(x;t) := B^k u^k(x;t)$. Condition **(b1)** implies that

$$\begin{aligned}\mathbb{O}_d &= \frac{\partial}{\partial t_i}\int g^k(x;t) f_k(x;t)\,\mu(dx) \\ &= \int\left(\frac{\partial}{\partial t_i} g^k(x;t)\right) f_k(x;t)\,\mu(dx) + \int g^k(x;t)\left(\frac{\partial}{\partial t_i} f_k(x;t)\right)\mu(dx),\end{aligned}$$

whence

$$V(\vartheta;\{B^k\}) = -\sum_{k=1}^K \int g^k(x;t) f_k'(x;t)^T\,\mu(dx).$$

Without loss of generality one can assume that $V(\vartheta;\{B^k\}) = \mathbb{I}_{d\times d}$. Indeed, the GEE-estimator $\hat{\vartheta}_N$ does not change if the functions $g^k(x;t)$ are changed by

$$V(\vartheta;\{B^k\})^{-1} g^k(x;t), \qquad k=1,\dots,K.$$

Therefore the restrictions imposed on the functions $g^k(x;t)$, $k=1,\dots,K$, can be written as follows:

$$\begin{cases} -\sum_{k=1}^K \int g^k(x;\vartheta)\cdot f_k'(x;\vartheta)^T\,\mu(dx) = \mathbb{I}_{d\times d}, \\ \int g^k(x;\vartheta) f_k(x;\vartheta)\,\mu(dx) = \mathbb{O}_d, \end{cases} \qquad k=1,\dots,K. \tag{16}$$

Now conditions (16) are rewritten in the following form:

$$\begin{cases} -\sum_{k=1}^K B^k\mathbb{V}_1^k(\vartheta) = \mathbb{I}_{d\times d}, \\ B^k\mathbb{V}_2^k(\vartheta) = \mathbb{O}_d, \end{cases} \qquad k=1,\dots,K,$$

or, equivalently,

$$B\mathbb{V}(\vartheta) = -E_0. \tag{17}$$

Problem of minimization. Theorem 3.5 follows if

$$c^T(D[\{B^k\}] - D[\{B^{*;k}\}])c \geq 0$$

for an arbitrary $c \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d\times R}$ for which conditions (17) hold. In other words, one has to minimize $c^T B\mathbb{Z}(\vartheta)B^T c$ for all $c \in \mathbb{R}^d$ under restrictions (17).

Put $b := B^T c$. Then the problem of minimization of $c^T B\mathbb{Z}(\vartheta)B^T c$ under conditions (17) takes the form

$$\begin{cases} b^T\mathbb{Z}(\vartheta)b \to \min; \\ \mathbb{V}(\vartheta)^T b = -E_0^T c. \end{cases} \tag{18}$$

Optimal coefficients. We introduce the variable $\lambda \in \mathbb{R}^R$. The Lagrange functional corresponding to the problem of minimization (18) is given by

$$\mathfrak{L}(b) := b^T\mathbb{Z}(\vartheta)b - 2\left(\mathbb{V}(\vartheta)^T b + E_0^T c\right)^T \lambda.$$

We find a stationary point of this functional. We derive from the equation $\frac{\partial \mathfrak{L}(\cdot)}{\partial b} = 0$ that $b^* = \mathbb{Z}(\vartheta)^{-1}\mathbb{V}(\vartheta)\lambda^*$. Conditions (18) imply $\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1}\mathbb{V}(\vartheta)\lambda^* = -E_0^T c$, whence

$$\lambda^* = -\left[\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1}\mathbb{V}(\vartheta)\right]^{-1}E_0^T c.$$

Thus, $b^* = -\mathbb{Z}(\vartheta)^{-1}\mathbb{V}(\vartheta)\left[\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1}\mathbb{V}(\vartheta)\right]^{-1}E_0^T c$.

Since $c \in \mathbb{R}^d$ is arbitrary and $b^* = {B^*}^T c$, we get

$$B^* = -E_0\left[\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1}\mathbb{Z}(\vartheta)\right]^{-1}\mathbb{V}(\vartheta)^T\mathbb{Z}(\vartheta)^{-1}.$$

Checking the correctness. The functional $b^T\mathbb{Z}(\vartheta)b$ that has been minimized in problem (18) is a quadratic form in an affine space. Since the matrix $\mathbb{Z}(\vartheta)$ is non-negative definite (as the covariance matrix of a certain random vector), the stationary point of this functional is the point of its global minimum. □

Remark 3.6. Conditions **(b1)** and **(b2)** hold for the matrices $B^{*;k}$, $k = 1, \dots, K$, by construction (see the proof of Theorem 3.5). Moreover, $V(\vartheta; \{B^{*;k}\}) = \mathbb{I}_{d\times d}$.

Example 3.7. Consider the model of mixture discussed in Example 2.4. To construct the adaptive estimator, we take a family of seven uniform B-splines $u(x; m, \sigma) \in \mathbb{R}^7$ with nodes at the points

$$\{m + k \cdot \sigma\}_{k=-5,\dots,5}.$$

Let $u^1(x;t) := u(x; t_1, t_3)$ and $u^2(x;t) := u(x; t_2, t_3)$. The dispersion coefficients for such an adaptive estimator and with optimal B^* equal $d^a(\hat{m}^1) = 62.77$, $d^a(\hat{m}^2) = 66.40$, and $d^a(\hat{\sigma}) = 22.16$, respectively. Therefore, the dispersion coefficients of the adaptive estimator are quite close to the lower bound for dispersion coefficients of the GEE-estimators.

3.3. **Estimation of B^*.** The optimal coefficients B^* can be estimated as follows:

$$\hat{U}_N^m := \frac{1}{N}\sum_{j=1}^N a_{j;N}^m U(\xi_{j;N};\tilde{\vartheta}_N) \in \mathbb{R}^{L\times K}, \qquad m=1,\dots,M;$$

$$\hat{\mathbb{Z}}_{1;N} := \sum_{r,s=1}^M \hat{U}_N^r \alpha_{r,s;N} (\hat{U}_N^s)^T \in \mathbb{R}^{L\times L};$$

$$\hat{\mathbb{Z}}_{2;N} := \frac{1}{N}\sum_{j=1}^N U(\xi_{j;N};\tilde{\vartheta}_N)\left[\sum_{m=1}^M a_{j;N}^m \beta_{m;N}\right] U(\xi_{j;N};\tilde{\vartheta}_N)^T \in \mathbb{R}^{L\times L};$$

$$\hat{\mathbb{Z}}_N := \hat{\mathbb{Z}}_{2;N} - \hat{\mathbb{Z}}_{1;N};$$

$$\hat{B}_N^* := -E_0\left[\mathbb{V}(\tilde{\vartheta}_N)^T \hat{\mathbb{Z}}_N^{-1}\mathbb{V}(\tilde{\vartheta}_N)\right]^{-1}\mathbb{V}(\tilde{\vartheta}_N)^T\hat{\mathbb{Z}}_N^{-1} \in \mathbb{R}^{d\times L}.$$

Lemma 3.8. *Assume that*

(i) *conditions* (a1) *and* (a3) *hold*;

(ii) *for some open neighborhood* $\mathfrak{U}$, $\vartheta \in \mathfrak{U} \subset \Theta$,

$$\int \sup_{t\in\mathfrak{U}} \left\|\frac{\partial}{\partial t} u^k(x;t)\right\|^2 F_i(dx) < \infty, \qquad i=1,\dots,M,\ k=1,\dots,K,$$

(*this is a stronger version of condition* (a2));

(iii) $\det\Gamma \neq 0$;

(iv) *the limit matrices* $\alpha_{r,s} \in \mathbb{R}^{K\times K}$ *and* $\beta_m \in \mathbb{R}^{K\times K}$ *exist for* $r,s,m=1,\dots,M$;

(v) $\tilde{\vartheta}_N \to \vartheta$ *in probability.*

Then $\hat{\mathbb{Z}}_N \to \mathbb{Z}(\vartheta)$ *in probability.*

Proof. The proof follows the lines of that of Lemma 2 in [15].

Let $\kappa(\cdot)$ be the index of a component in the mixture, and let $\lambda(\cdot)$ be a coordinate of the vector $u^k(x;t)$:

$$\begin{aligned} \kappa(1) := \dots := \kappa(L_1) := 1; &\qquad \lambda(1) := 1,\dots,\lambda(L_1) := L_1; \\ \vdots &\qquad \vdots \\ \kappa(L-L_K+1) := \dots := \kappa(L) := K; &\qquad \lambda(L-L_K+1) := 1,\dots,\lambda(L) := L_K. \end{aligned}$$

Let $k,l \in \{1,\dots,K\}$. Below we obtain some useful expressions for $(\mathbb{Z}_1)_{k,l}$ and $(\mathbb{Z}_2)_{k,l}$. First,

$$(\mathbb{Z}_1)_{k,l} = \sum_{r,s=1}^M \left\langle a^{\kappa(k)}a^{\kappa(l)}p^r p^s\right\rangle \int u_{\lambda(k)}^{\kappa(k)}(x;\vartheta) f_r(x)\,\mu(dx) \int u_{\lambda(l)}^{\kappa(l)}(x;\vartheta) f_s(x)\,\mu(dx).$$

By Lemmas 4 and 1 of [15] with

$$b_{j;N} := a_{j;N}^{\kappa(k)} a_{j;N}^{\kappa(l)}, \qquad u(x) := u_{\lambda(k)}^{\kappa(k)}(x;\vartheta), \qquad v(x) := u_{\lambda(l)}^{\kappa(l)}(x;\vartheta)$$

we obtain the convergence $(\hat{\mathbb{Z}}_{1;N})_{k,l} \to (\mathbb{Z}_1)_{k,l}$ in probability.

Second,

$$(\mathbb{Z}_2)_{k,l} = \sum_{m=1}^M \left\langle a^{\kappa(k)}a^{\kappa(l)}p^m\right\rangle \int u_{\lambda(k)}^{\kappa(k)}(x;\vartheta) u_{\lambda(l)}^{\kappa(l)}(x;\vartheta) f_m(x)\,\mu(dx).$$

By Lemma 3 of [15] with $b_{j;N} := a_{j;N}^{\kappa(k)} a_{j;N}^{\kappa(l)}$ and $v(x;t) := u_{\lambda(k)}^{\kappa(k)}(x;t) u_{\lambda(l)}^{\kappa(l)}(x;t)$ we prove the convergence $(\hat{\mathbb{Z}}_{2;N})_{k,l} \to (\mathbb{Z}_2)_{k,l}$ in probability. □

3.4. **Adaptive estimator.** The adaptive estimator is defined by

$$\check{\vartheta}_N := \tilde{\vartheta}_N - \sum_{k=1}^{K} \hat{B}_N^{*;k} \cdot \hat{u}_N^k\big(\tilde{\vartheta}_N\big).$$

Theorem 3.9. *Assume that conditions* (i)–(iv) *of Lemma* 3.8 *hold. Further let*

(i) *the matrix* $\mathbb{V}(\vartheta)$ *be of full rank*;
(ii) *the estimator* $\tilde{\vartheta}_N$ *of* ϑ *be* $\sqrt{N}$*-consistent.*

Then

$$\sqrt{N} \cdot (\check{\vartheta}_N - \vartheta) \xrightarrow{W} \mathcal{N}\big(\mathbb{O}_d, D\big[\big\{B^{*;k}\big\}\big]\big).$$

Proof. The proof follows the lines of that of Theorem 4 of [15].

We represent $\sqrt{N}(\check{\vartheta}_N - \vartheta)$ as follows:

$$\sqrt{N}(\check{\vartheta}_N - \vartheta) = \sqrt{N}\left(\tilde{\vartheta}_N - \vartheta - \sum_{k=1}^{K} \hat{B}_N^{*;k}\hat{u}_N^k\big(\tilde{\vartheta}_N\big)\right) = -\sqrt{N}\sum_{k=1}^{K} \hat{B}_N^{*;k}\hat{u}_N^k(\vartheta) + \varepsilon_1,$$

where

$$\varepsilon_1 := \sqrt{N}\left(\mathbb{I}_{d\times d} - \sum_{k=1}^{K} \hat{B}_N^{*;k}\frac{\partial}{\partial t}\hat{u}_N^k(t)\,\bigg|_{t=\zeta}\right)(\tilde{\vartheta}_N - \vartheta)$$

and ζ is an intermediate point between ϑ and $\tilde{\vartheta}_N$.

At the same time,

$$\sqrt{N}\sum_{k=1}^{K} \hat{B}_N^{*;k}\hat{u}_N^k(\vartheta) = \sqrt{N}\sum_{k=1}^{K} B^{*;k}\hat{u}_N^k(\vartheta) + \varepsilon_2,$$

where $\varepsilon_2 := \sqrt{N}\sum_{k=1}^{K}(\hat{B}_N^{*;k} - B^{*;k})\hat{u}_N^k(\vartheta)$. Thus

$$\sqrt{N}(\check{\vartheta}_N - \vartheta) = -\sqrt{N}\sum_{k=1}^{K} B^{*;k}\hat{u}_N^k(\vartheta) + \varepsilon_1 + \varepsilon_2.$$

Now we determine ε_2. By construction, $B^*\mathbb{V}(\vartheta) = E_0$ and $\hat{B}_N^*\mathbb{V}(\vartheta) = E_0$, whence $B^{*;k}\bar{u}^k(\vartheta) = \mathbb{O}_d$ and $\hat{B}^{*;k}\bar{u}^k(\vartheta) = \mathbb{O}_d$, where

$$\bar{u}(\vartheta) := \int u^k(x;\vartheta) f_k(x;\vartheta)\,\mu(dx), \qquad k = 1,\dots,K.$$

Hence

$$\varepsilon_2 = \sum_{k=1}^{K} \sqrt{N}\big(\hat{B}_N^{*;k} - B^{*;k}\big)\big(\hat{u}_N^k(\vartheta) - \bar{u}^k(\vartheta)\big).$$

Since $\hat{B}_N^{*;k} \to B^{*;k}$ in probability and $\mathsf{E}\left[\big(\sqrt{N}(\hat{u}_N^k(\vartheta) - \bar{u}^k(\vartheta))\big)^2\right]$ possesses the limit, we get $\varepsilon_2 = o_p(1)$ as $N \to \infty$.

Then we determine ε_1. By construction of B^*,

$$\mathsf{E}\left[\sum_{k=1}^{K} B^{*;k}\frac{\partial}{\partial t}\hat{u}^k(t)\bigg|_{t=\vartheta}\right] = \sum_{k=1}^{K}\int B^{*;k}\frac{\partial}{\partial t}u^k(t)\bigg|_{t=\vartheta} F_k(dx;\vartheta) = V\big(\vartheta;\big\{B^{*;k}\big\}\big) = \mathbb{I}_{d\times d}.$$

Lemma 3 of [15] implies that

$$\hat{B}_N^{*;k}\frac{\partial}{\partial t}\hat{u}^k(t)\bigg|_{t=\zeta} - B^{*;k}\frac{\partial}{\partial t}\hat{u}^k(t)\bigg|_{t=\vartheta} \to \mathbb{O}_{d\times d}$$

in probability, $k = 1, \dots, K$. Thus $\varepsilon_1 = o_p(1)$ as $N \to \infty$. Therefore

$$\sqrt{N}(\check{\vartheta}_N - \vartheta) = -\sqrt{N} \sum_{k=1}^{K} B^{*;k} \hat{u}_N^k(\vartheta) + o_p(1).$$

Similarly to Lemma 3.3,

$$\sqrt{N} \sum_{k=1}^{K} B^{*;k} \hat{u}_N^k(\vartheta) \xrightarrow{W} \mathcal{N}\big(\mathbb{O}_d, D\big[\{B^{*;k}\}\big]\big).$$

Therefore, $\sqrt{N} \cdot (\check{\vartheta}_N - \vartheta) \xrightarrow{W} \mathcal{N}(\mathbb{O}_d, D[\{B^{*;k}\}])$. □

4. Simulation

The model of mixture considered in Example 2.4 was simulated with the help of the software *Mathematica*. Simulated were samples of different sizes. The number of observations in these samples were 50, 100, 250, 500, 750, 1 000, 2 000, and 5 000. Evaluated from the simulated data were the moment estimators, improved moment estimators (see Example 2.4), and adaptive estimators (see Example 3.7) constructed from the improved moment estimators. The behavior of estimators is checked for multiple samples generated for each size (2,000 times for each size). The results of simulation are seen in Figure 1.

The quality of estimators is measured in the sense of their mean square error (MSE) multiplied by the number of observations in a sample. It is worth mentioning that one observes large deviations from the average for adaptive estimators in small samples (whose size does not exceed 500). Thus, in addition to MSE, the deviation of estimators is estimated by RobVar $:= N \cdot (\hat{Q}_N(3/4) - \hat{Q}_N(1/4))^2/\gamma^2$, where N is the size of a sample, $\hat{Q}_N(\alpha)$ an estimate of the quantile of level α, and γ the interquartile range of the standard normal distribution (being approximately equal to 1.34898).

The simulation shows that $N \cdot$MSE tends to the dispersion coefficient for all estimators as N grows. This means that the bias is negligible for each of the estimators. The adaptive estimator seems to be better than the moment estimator if the size of a sample is larger than 500 which makes it reasonable to advise the use of adaptive estimators for moderate and large samples. For smaller sizes (about 500 observations), the adaptive estimator is worth using if it differs from the pilot estimator inessential. One can propose the following procedure to make the latter observation precise: construct a confidence interval for ϑ by using the pilot estimator and check whether or not the adaptive estimator falls into the confidence interval (the variance of moment estimators is found in [2]).

5. Concluding remarks

We considered a model of a finite mixture with varying concentrations, where a part of components is described as a submodel with parameters. We propose a technique for the adaptive estimation of parameters which, in fact, is a realization of the method of GEE-estimators. For adaptive estimators, conditions for the consistency and asymptotic normality are proposed and dispersion matrices are found. The dispersion coefficients of the adaptive estimators are quite close to the lower bound of the GEE-estimators found in [2]. The behavior of adaptive estimators is demonstrated for simulated data.

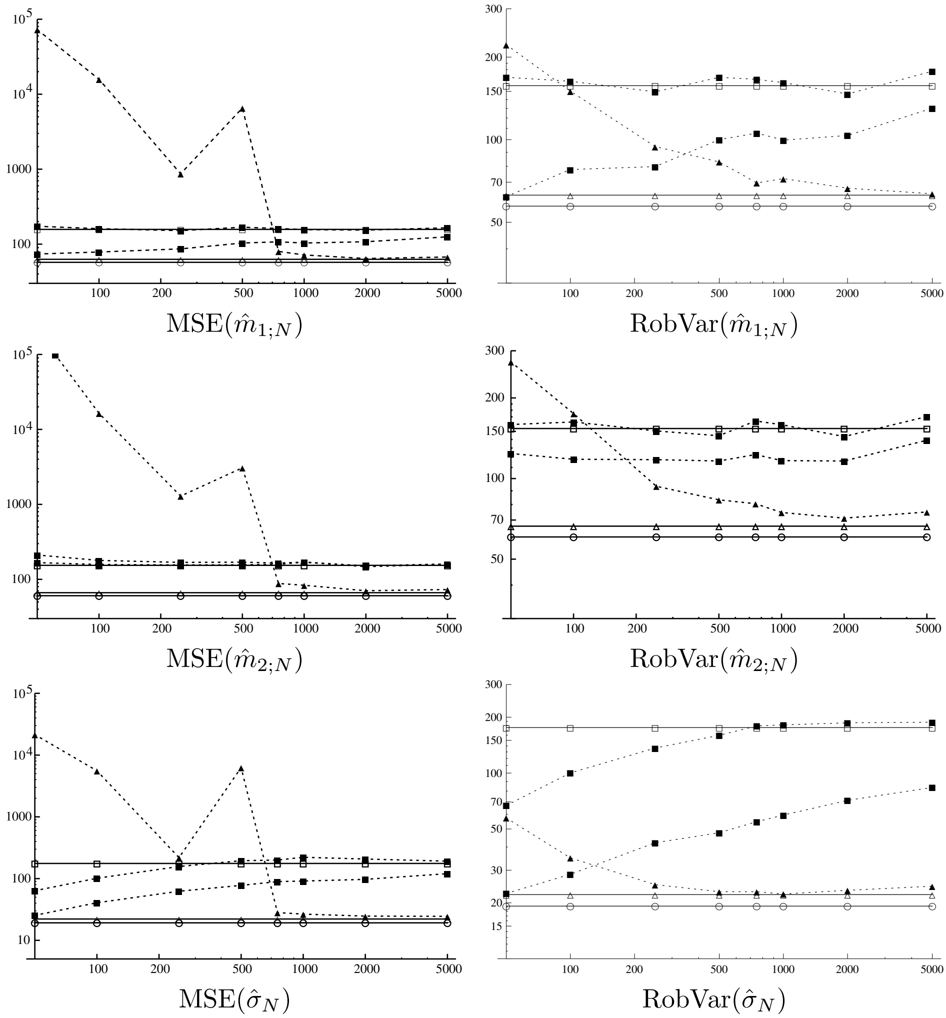

FIGURE 1. Dispersion of estimators. MSE is the mean square deviation of the estimator multiplied by the number of observations. RobVar is the interquartile range (an estimator of the variance) multiplied by the number of observations. The symbols □ and △ correspond to the theoretical variance of the moment estimator and adaptive estimator, respectively; the symbol ∘ corresponds to the lower bound. The symbols ■ and ▲ denote MSE and RobVar, respectively. The upper line with ■ corresponds to the moment estimator, while the lower one corresponds to the improved estimator.

BIBLIOGRAPHY

1. A. A. Borovkov, *Mathematical Statistics*, “Nauka”, Moscow, 1984; English transl., Translated from the Russian by A. Moullagaliev and revised by the author, Gordon and Breach Science Publishers, Amsterdam, 1998. MR1712750 (2000f:62003)
2. O. V. Doronin, *Lower bound of the dispersion matrix for the semiparametric estimation in a model of mixture*, Teor. Imovirnost. Matem. Statist. (accepted)

3. O. V. Doronin, *Robust estimates for mixtures with a Gaussian component*, Visnyk Kyiv National Taras Shevchenko University. Mathematics and Mechanics (2012), no. 1, 18–23. (Ukrainian)
4. N. Lodatko and R. Maĭboroda, *An adaptive moment estimator of a parameter of a distribution constructed from observations with admixture*, Teor. Imovirnost. Matem. Statist. **75** (2006), 61–70; English transl in Theor. Probability and Math. Statist. **75** (2007), 71–82. MR2321182 (2008g:62101)
5. R. E. Maiboroda and O. V. Sugakova, *An Estimation and Classification by Observations from a Mixture*, Kyiv University Press, Kyiv, 2008. (Ukrainian)
6. D. I. Pokhyl'ko, *Wavelet estimators of a density constructed from observations of a mixture*, Teor. Imovirnost. Matem. Statist. **70** (2004), 121–130; English transl in Theor. Probability and Math. Statist. **70** (2005), 135–145.
7. A. M. Shcherbina, *Estimation of the mean value in a model of mixtures with varying concentrations*, Teor. Imovirnost. Matem. Statist. **84** (2011), 142–154; English transl in Theor. Probability and Math. Statist. **84** (2012), 151–164. MR2857425 (2012h:62130)
8. A. M. Shcherbina, *Estimation of the parameters of the binomial distribution in a model of mixture*, Teor. Imovirnost. Matem. Statist. **86** (2012), 182–192; English transl in Theor. Probability and Math. Statist. **86** (2013) 205–217. MR2986460
9. F. Autin and Ch. Pouet, *Test on the components of mixture densities*, Statistics & Risk Modelling **28** (2011), no. 4, 389–410. MR2877572
10. L. Bordes, C. Delmas, and P. Vandekerkhove, *Semiparametric Estimation of a two-component mixture model where one component is known*, Scand. J. Statist. **33** (2006), 733–752. MR2300913 (2008f:62049)
11. P. Hall and X.-H. Zhou, *Nonparametric estimation of component distributions in a multivariable mixture*, Ann. Statist. **31** (2003), no. 1, 201–224. MR1962504 (2003m:62105)
12. R. E. Maiboroda and O. O. Kubaichuk, *Improved estimators for moments constructed from observations of a mixture*, Teor. Imovir. Mat. Stat. **70** (2004), 74–81; English transl. in Theor. Probability and Math. Statist. **70** (2005), 83–92.
13. R. Maiboroda and O. Sugakova, *Nonparametric density estimation for symmetric distributions by contaminated data*, Metrica **75** (2012), no. 1, 109–126. MR2878111
14. R. Maiboroda and O. Sugakova, *Statistics of mixtures with varying concentrations with application to DNA microarray data analysis*, J. Nonparam. Statist. **24** (2012), no. 1, 201–205. MR2885834
15. R. E. Maiboroda, O. V. Sugakova, and A. V. Doronin, *Generalized estimating equations for mixtures with varying concentrations*, Canadian J. Statist. **41** (2013), no. 2, 217–236. MR3061876
16. G. J. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, 2000. MR1789474 (2002b:62025)
17. J. Shao, *Mathematical Statistics*, Springer-Verlag, New York, 1998. MR2002723 (2004g:62002)
18. O. Sugakova, *Adaptive estimates for the parameter of a mixture of two symmetric distributions*, Teor. Imovir. Mat. Stat. **82** (2010), 146–155 (Ukrainian, with English, Russian and Ukrainian summaries); English transl., Theor. Probability and Math. Statist. **82** (2011), 149–159. MR2790490 (2011m:62130)
19. O. Sugakova, *Empirical Bayesian classification for observations with an admixture*, Teor. Imovir. Mat. Stat. 84 (2011), 155–162 (Ukrainian, with English, Russian and Ukrainian summaries); English transl. Theor. Probability and Math. Statist. **84** (2012), 165–172. MR2857426 (2012f:62128)
20. D. M. Titterington, A. F. Smith, and U. E. Makov, *Analysis of Finite Mixture Distributions*, Wiley, New York, 1985.

Department of Probability Theory, Statistics, and Actuarial Mathematics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Volodymyrs'ka Street, 64, Kyiv 01601, Ukraine

E-mail address: al_doronin@ukr.net

Received 04/MAR/2014
Translated by N. SEMENOV