

ADAPTIVE ESTIMATORS FOR PARAMETERS OF A MIXTURE OF TWO SYMMETRIC DISTRIBUTIONS

UDC 519.21

O. SUGAKOVA

ABSTRACT. A sample is observed from a mixture of two symmetric distributions that differ only by the location parameters. We use the method of estimating equations to estimate unknown parameters of the components. The method works as follows: first, we construct an estimator for the optimal estimating functions; then we use it to construct adaptive estimators. We study the asymptotic behavior of resulting estimators.

1. INTRODUCTION

In this paper, we consider the problem of estimating the parameters of a model of the mixture of two symmetric distributions that differ by a location parameter. There is a number of papers (see, for example, [1]–[4]) devoted to this problem. The identifiability of this model is proved in [2]. Weakly consistent estimators of parameters of the model are constructed in [2] with the help of empirical distribution functions and in [3] with the help of empirical characteristic functions. Moment estimators are proposed in [4] for Euclidean parameters; the asymptotic normality of these parameters is also proved in [4].

In the paper [1], a generalization of moment estimators is considered by using the estimating equations method. The asymptotic normality of estimators obtained by the method of generalized estimating equations (GEE-estimators) is proved and the lower limit for the dispersion coefficients (asymptotic variances) is found in [1]. Unfortunately the lower limit is attained at an estimating function that depends on unknown parameters of the model. Thus an adaptive technique is useful to construct estimators with small dispersion coefficients.

In this paper, we consider GEE-estimators that can be constructed by using linear combinations $\sum \beta_i u_i$ of some fixed basis functions u_i as estimating functions. The exact lower limit $\Sigma(\beta^*)$ for the dispersion matrix of such estimators and corresponding “optimal” coefficients β_i^* of the linear combination are obtained in Section 2. The consistent estimators $\hat{\beta}_i$ for optimal coefficients are described in Section 3. In Section 4, we construct the adaptive estimators of the Euclidean parameters by using $\hat{\beta}_i$ instead of the true optimal coefficients β_i^* . It is proved in Section 4 that the dispersion matrix of these estimators is equal to $\Sigma(\beta^*)$.

2. SETTING OF THE PROBLEM

Consider a sample of independent identically distributed random variables ξ_1, \dots, ξ_n whose distribution is a mixture of two distributions. We assume that these distributions

2010 *Mathematics Subject Classification.* Primary 62G05; Secondary 62G20.

Key words and phrases. Mixture, estimating equations, adaptive estimators for parameters.

differ by a shift of the argument. We further assume that every observation has the density

$$(1) \quad \psi(x) = pf(x - a_1) + (1 - p)f(x - a_2),$$

where p denotes the concentration of the first component in the mixture, $0 < p < 1/2$, where $a_i \in \mathbf{R}$, $i = 1, 2$, is the median of the component i of the mixture, and where f is the symmetric (that is, $f(-x) = f(x)$) density of the distribution of the deviation of an observation from the median (this density is the same for both components of the mixture). We also assume that $f(x)$ is a differentiable function.

Model (1) can be rewritten in other terms, namely

$$(2) \quad \xi_j = a_{\delta_j} + \eta_j,$$

where δ_j is the indicator, that is, $\mathbf{P}\{\delta_j = 1\} = p$, $\mathbf{P}\{\delta_j = 2\} = 1 - p$, and where η_j are independent identically distributed random variables whose density is f .

To estimate the parameters a_1 , a_2 , and p we use, as in the paper [1], the method of generalized estimating equations. Consider a triple of arbitrary odd functions $g_i(x)$, $i = 1, 2, 3$, such that $\mathbf{E}g(\eta_1 + \alpha) < \infty$ for all $\alpha \in \mathbf{R}$.

Then

$$\mathbf{E}g_i(\xi_j - \alpha) = p \int_{-\infty}^{\infty} g_i(x)f(x + \alpha - a_1) dx + (1 - p) \int_{-\infty}^{\infty} g_i(x)f(x + \alpha - a_2) dx$$

and $\mathbf{E}[pg(\xi_j - a_1) + (1 - p)g(\xi_j - a_2)] = 0$. Put

$$\hat{g}_i(\alpha) = \frac{1}{n} \sum_{j=1}^n g_i(\xi_j - \alpha).$$

The statistics $\hat{\theta}_n = (\hat{p}_n, \hat{a}_{1,n}, \hat{a}_{2,n})^T$ is a GEE-estimator of the Euclidean parameter $\theta = (p, a_1, a_2)$ with the estimating triple (g_1, g_2, g_3) if the system of equations

$$(3) \quad \pi \hat{g}_i(\alpha_1) + (1 - \pi) \hat{g}_i(\alpha_2) = 0, \quad i = 1, 2, 3,$$

is satisfied almost surely after the substitutions $\pi = \hat{p}_n$, $\alpha_1 = \hat{a}_{1,n}$, and $\alpha_2 = \hat{a}_{2,n}$.

Consider and fix a certain set M of odd ‘‘basis’’ functions $\{u_m(x)\}_{m=1}^M$. In what follows we restrict our consideration to the estimating triples g_i of the following form:

$$(4) \quad g_i(x) = g_{\beta_i}(x) = \sum_{m=1}^M \beta_{im} u_m(x).$$

Our aim is to find the coefficients β_i^* for which the GEE-estimators with the estimating triple (g_1, g_2, g_3) have the minimal dispersion matrix.

3. ASYMPTOTIC NORMALITY OF GEE-ESTIMATORS

The asymptotic normality of estimators defined by the system (3) is proved in the paper [1]. Let

$$\begin{aligned} b^+(x) &= -2pf'(x) - (1 - p)(f'(x + a_1 - a_2) + f'(-x + a_1 - a_2)), \\ b^-(x) &= -2(1 - p)f'(x) + p(f'(-x + a_1 - a_2) - f'(x + a_1 - a_2)), \\ d(x) &= f(-x + a_1 - a_2) - f(x + a_1 - a_2). \end{aligned}$$

For all functions g_1 and g_2 defined on \mathbf{R} , we use the following notation:

$$\int g_1 g_2 = \int_{-\infty}^{+\infty} g_1(x) g_2(x) dx.$$

We also consider the matrix

$$Z = Z(\beta) = \begin{pmatrix} \int g_1 b^+ & \int g_1 b^- & \int g_1 d \\ \int g_2 b^+ & \int g_2 b^- & \int g_2 d \\ \int g_3 b^+ & \int g_3 b^- & \int g_3 d \end{pmatrix}.$$

Theorem 3.1. *Assume that*

- 1) $\hat{\theta}_n = (\hat{p}_n, \hat{a}_{1,n}, \hat{a}_{2,n})^T$ is a weakly consistent estimator;
- 2) for $m = 1, \dots, M$ and $k = 1, 2$,

$$\mathbf{E} u_m^2(\xi_1 - a_k) < \infty;$$

- 3) there are $\varepsilon > 0$ and $\delta > 0$ such that

$$\mathbf{E} \sup_{\alpha: |\alpha - a_k| < \varepsilon} |u_m(\xi_1 - \alpha)|^{1+\delta} < \infty$$

and

$$\mathbf{E} \sup_{\alpha: |\alpha - a_k| < \varepsilon} |u'_m(\xi_1 - \alpha)|^{1+\delta} < \infty, \quad m = 1, \dots, M, \quad k = 1, 2;$$

- 4) $u'_m(x)$ and $f'(x)$ are continuous on \mathbf{R} ;
- 5) $u_m(x + \Delta)f(x) \rightarrow 0$ as $x \rightarrow \pm\infty$ for all $\Delta \in \mathbf{R}$;
- 6) all elements of Z are finite and $\det Z \neq 0$.

Then

$$(5) \quad \sqrt{n}((\hat{p}_n, \hat{a}_{1,n}, \hat{a}_{2,n}) - (p, a_1, a_2)) \Rightarrow N(0, \Sigma),$$

where

$$\Sigma = \Sigma(\beta) = Z^{-1}Q(Z^T)^{-1}$$

and where Q is the following matrix:

$$Q = Q(\beta)$$

$$= (\mathbf{E}(pg_i(\xi_1 - a_1) + (1-p)g_i(\xi_1 - a_2))(pg_j(\xi_1 - a_1) + (1-p)g_j(\xi_1 - a_2)))_{i,j=1}^3.$$

This result follows easily from Theorem 3.1 in [1].

We rewrite the matrices Z and Q by taking into account the decomposition in the basis (4). Let

$$b_k^+ = \int_0^\infty b^+(x)u_k(x) dx, \quad b_k^- = \int_0^\infty b^-(x)u_k(x) dx,$$

$$d_k = \int_0^\infty d(x)u_k(x) dx.$$

Then

$$Z = \begin{pmatrix} \sum_{k=1}^M \beta_{1k} b_k^+ & \sum_{k=1}^M \beta_{1k} b_k^- & \sum_{k=1}^M \beta_{1k} d_k \\ \sum_{k=1}^M \beta_{2k} b_k^+ & \sum_{k=1}^M \beta_{2k} b_k^- & \sum_{k=1}^M \beta_{2k} d_k \\ \sum_{k=1}^M \beta_{3k} b_k^+ & \sum_{k=1}^M \beta_{3k} b_k^- & \sum_{k=1}^M \beta_{3k} d_k \end{pmatrix},$$

$$(6) \quad Q = \left(\sum_{k,l=1}^M c_{kl} \beta_{ik} \beta_{jl} \right)_{i,j=1}^3,$$

where, for $k, l = 1, \dots, M$,

$$c_{kl} = \int_{-\infty}^\infty E_k(x)E_l(x)(pf(x) + (1-p)f(x + a_1 - a_2)) dx$$

with

$$E_j(x) = pu_j(x) + (1-p)u_j(x + a_1 - a_2).$$

We further put

$$(7) \quad B = \begin{pmatrix} b_1^+ & b_1^- & d_1 \\ \vdots & \vdots & \vdots \\ b_M^+ & b_M^- & d_M \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \vdots & \vdots & \vdots \\ \beta_{1M} & \beta_{2M} & \beta_{3M} \end{pmatrix}, \quad C = (c_{kl})_{k,l=1,\dots,M}.$$

The result below describes the matrix β for which the matrix Σ is minimal in the sense of the following definition.

Definition 3.1. The inequality $A \leq B$ for two matrices A and B means that the matrix $B - A$ is nonnegative definite.

Theorem 3.2. *If $\det C \neq 0$ and the rank of the matrix B is equal to 3, then*

$$\Sigma(\beta) \geq \Sigma(\beta^*)$$

for an arbitrary matrix $\beta = (\beta_{ij})$, $i = 1, 2, 3$; $j = 1, \dots, M$, where

$$\beta^* = C^{-1}B(B^T C^{-1}B)^{-1}.$$

Proof of Theorem 3.2. We need the following auxiliary result [1].

Lemma 3.1. *Let a matrix g consist of the row vectors g_i , $i = 1, 2, 3$, while a matrix S consists of the row vectors s_i , $i = 1, 2, 3$. Denote by $\langle a, b \rangle$ the scalar product of two vectors of the space \mathbf{R}^M . Let $\Gamma(g) = (\langle g_i, g_j \rangle)_{i,j=1}^3$ and $V(g) = (\langle g_i, s_j \rangle)_{i,j=1}^3$. Finally, let V_0 be a fixed matrix and $\det \Gamma(S) \neq 0$.*

Then

$$\Gamma(g) \geq \Gamma(g^*)$$

for an arbitrary set of vectors g_i such that $V(g) = V_0$, where $g^* = V_0 \Gamma(S)^{-1} S$. Moreover,

$$\Gamma(g^*) = V_0 \Gamma(S)^{-1} V_0^T.$$

We identify the functions $g_i = \sum_{k=1}^M \beta_{ik} u_k$ with the vectors $\beta_i = (\beta_{i1}, \dots, \beta_{iM})$. Without loss of generality we assume that $Z(g_1, g_2, g_3) = E$, where E is the unit matrix and where $Z(g_1, g_2, g_3)$ is the matrix Z corresponding to the triple of estimating functions (g_1, g_2, g_3) .

Indeed, let (g_1, g_2, g_3) be a triple of odd estimating functions, let

$$L = (l_{ik})_{i,k=1}^3 \in \mathbf{R}^{3 \times 3}$$

be a nondegenerate matrix, and let $\tilde{g}_i(x) = l_{i1}g_1(x) + l_{i2}g_2(x) + l_{i3}g_3(x)$.

It does not matter which estimating triple, either (g_1, g_2, g_3) or $(\tilde{g}_1, \tilde{g}_2, \tilde{g}_3)$, is used in the system of estimating equations (3), since the solution is the same. It is easy to see that

$$Z(\tilde{g}_1, \tilde{g}_2, \tilde{g}_3) = LZ(g_1, g_2, g_3).$$

If $\det Z(g_1, g_2, g_3) \neq 0$, then we put $L = Z^{-1}(g_1, g_2, g_3)$ and obtain $Z(\tilde{g}_1, \tilde{g}_2, \tilde{g}_3) = E$.

Define the scalar product of the vectors as follows:

$$\langle \beta_i, \beta_j \rangle = \sum_{k,l=1}^M c_{kl} \beta_{ik} \beta_{jl},$$

where the c_{kl} are given by (6).

Imposing the normalizing condition, $Z = V_0 = E$, and using the notation of Lemma 3.1 and Theorem 3.2 we obtain

$$\Sigma = Q = \Gamma(\beta), \quad V(\beta) = (\langle \beta_i, s_j \rangle)_{i,j=1}^3.$$

The vectors s_j are determined by the condition that the scalar products are equal:

$$\langle \beta_i, s_j \rangle = \beta_i^T C s_j = \beta_i^T b^j, \quad i, j = 1, 2, 3.$$

Here $b^1 = b^+$, $b^2 = b^-$, and $b^3 = d$ are the corresponding columns of the matrix B . Then

$$s_1 = C^{-1}b^+, \quad s_2 = C^{-1}b^-, \quad s_3 = C^{-1}d.$$

By Lemma 3.1, the minimum of the limit dispersion matrix is attained at

$$\beta^* = C^{-1}B(B^T C^{-1}B)^{-1}$$

and it is equal to

$$\Gamma(g^*) = \Sigma(\beta^*) = \Gamma^{-1}(C^{-1}B) = (B^T C^{-1}B)^{-1}. \quad \square$$

4. ESTIMATES OF THE OPTIMAL COEFFICIENTS β^*

The matrices B and C in Theorem 3.2 are expressed in terms of the unknown parameters p , a_1 , and a_2 . A natural problem is to find consistent estimators for the elements of the matrices B and C .

In the paper [4], strong consistent estimators for the Euclidean parameters p , a_1 , and a_2 are constructed with the help of the method of moments. The construction in [4] requires the condition

$$(8) \quad \mathbf{E} |\eta_j|^5 < \infty.$$

By \hat{p}_n , $\hat{a}_{1,n}$, and $\hat{a}_{2,n}$, we denote arbitrary consistent estimators of the unknown parameters p , a_1 , and a_2 , respectively. An example of such GEE-estimators is given by the method of moments studied in [1] and corresponding to the functions $g_1(x) = x$, $g_2(x) = x^3$, and $g_3(x) = x^5$.

Consider the entries of the matrices B and C . By construction,

$$\mathbf{E} g'_i(\xi_1 - a_1) = \sum_{k=1}^M \beta_{ik} \mathbf{E} u'_k(\xi_1 - a_1) = \sum_{k=1}^M \beta_{ik} b_k^+,$$

whence

$$b_k^+ = \mathbf{E} u'_k(\xi_1 - a_1).$$

Similarly

$$(9) \quad \begin{aligned} b_k^- &= \mathbf{E} u'_k(\xi_1 - a_2), & d_k &= \frac{1}{p} \mathbf{E} u_k(\xi_1 - a_2), \\ c_{kl} &= \mathbf{E} (p u_k(\xi_1) + (1-p) u_k(\xi_1 + a_1 - a_2)) (p u_l(\xi_1) + (1-p) u_l(\xi_1 + a_1 - a_2)), \\ & & k, l &= 1, \dots, M. \end{aligned}$$

Replacing the theoretical moments in these equalities by their empirical counterparts as well as replacing the parameters by their method of moments estimators we obtain the following estimators for the entries of the matrix B :

$$(10) \quad \begin{aligned} \hat{b}_{k,n}^+ &= \frac{1}{n} \sum_{j=1}^n u'_k(\xi_j - \hat{a}_{1,n}), & \hat{b}_{k,n}^- &= \frac{1}{n} \sum_{j=1}^n u'_k(\xi_j - \hat{a}_{2,n}), \\ \hat{d}_{k,n} &= \frac{1}{n \hat{p}_n} \sum_{j=1}^n u_k(\xi_j - \hat{a}_{2,n}), \end{aligned}$$

for $k = 1, \dots, M$, and those of the matrix C :

$$(11) \quad \hat{c}_{kl,n} = \frac{1}{n} \sum_{j=1}^n (\hat{p}_n u_k(\xi_j - \hat{a}_{1,n}) + (1 - \hat{p}_n) u_k(\xi_j - \hat{a}_{2,n})) \\ \times (\hat{p}_n u_l(\xi_j - \hat{a}_{1,n}) + (1 - \hat{p}_n) u_l(\xi_j - \hat{a}_{2,n}))$$

for $k, l = 1, \dots, M$.

Denote by \hat{B} and \hat{C} the matrices constituted by the elements defined by equalities (10) and (11), respectively.

An open interval centered at zero is denoted by Θ .

Theorem 4.1. *Assume that there exist open intervals Θ_i , $i = 1, 2, 3$, such that*

- (i) $\mathbf{E} \left(\sup_{\gamma \in \Theta_1} |u_k''(\xi_1 - a_i + \gamma)| \right) < \infty, \quad i = 1, 2, k = 1, \dots, M;$
- (ii) $\mathbf{E} \left(\sup_{\gamma \in \Theta_2} (u_k'(\xi_1 - a_i + \gamma))^2 \right) < \infty, \quad i = 1, 2, k = 1, \dots, M;$
- (iii) $\mathbf{E} \left(\sup_{\gamma \in \Theta_3} (u_k(\xi_1 - a_i + \gamma))^2 \right) < \infty, \quad i = 1, 2, k = 1, \dots, M;$
- (iv) *the estimators $\hat{p}_n, \hat{a}_{1,n}, \hat{a}_{2,n}$ are consistent.*

Then the estimators $\hat{b}_{k,n}^+, \hat{b}_{k,n}^-, \hat{d}_{k,n}$, and $\hat{c}_{kl,n}$, $k, l = 1, \dots, M$, defined by equalities (10) and (11) are consistent.

Proof of Theorem 4.1. We rewrite the estimator $\hat{b}_{k,n}^+$ as follows:

$$\hat{b}_{k,n}^+ - b_k^+ = \varepsilon_1^n + \varepsilon_2^n,$$

where

$$\varepsilon_1^n = \frac{1}{n} \sum_{j=1}^n [u_k'(\xi_j - \hat{a}_{1,n}) - u_k'(\xi_j - a_1)], \quad \varepsilon_2^n = \frac{1}{n} \sum_{j=1}^n u_k'(\xi_j - a_1) - \mathbf{E} u_k'(\xi_j - a_1).$$

According to the strong law of large numbers, $\varepsilon_2^n \rightarrow 0$ almost surely. We rewrite the first term ε_1^n with the help of Taylor's formula:

$$\varepsilon_1^n = \frac{1}{n} \sum_{j=1}^n u_k''(\xi_j - a_1 + \theta(a_1 - \hat{a}_{1,n}))(a_1 - \hat{a}_{1,n}) = S_n(a_1 - \hat{a}_{1,n}).$$

Here θ is a random variable assuming values in the interval $[0, 1]$.

First we prove that S_n is a stochastically bounded sequence:

$$(12) \quad \sup_n \mathbf{P} \{|S_n| > C\} \rightarrow 0, \quad C \rightarrow \infty.$$

For an arbitrary $C > 0$,

$$\mathbf{P} \{|S_n| > C\} \leq \mathbf{P} \{\theta(a_1 - \hat{a}_{1,n}) \notin \Theta_1\} + \mathbf{P} \{|S_n| > C; \theta(a_1 - \hat{a}_{1,n}) \in \Theta_1\}.$$

The first probability on the right-hand side approaches zero as $n \rightarrow \infty$, since $\hat{a}_{1,n} \xrightarrow{P} a_1$. We estimate the second term by condition (i) of Theorem 4.1:

$$\begin{aligned} & \mathbb{P} \left\{ \frac{1}{n} \left| \sum_{j=1}^n u_k''(\xi_j - a_1 + \theta(a_1 - \hat{a}_{1,n})) \right| > C; \theta(a_1 - \hat{a}_{1,n}) \in \Theta_1 \right\} \\ & \leq \mathbb{P} \left\{ \frac{1}{n} \sum_{j=1}^n \sup_{\gamma \in \Theta_1} |u_k''(\xi_j - a_1 + \gamma)| > C \right\} \\ & \leq \frac{\mathbb{E} \sup_{\gamma \in \Theta_1} |u_k''(\xi_j - a_1 + \gamma)|}{C} \rightarrow 0, \quad C \rightarrow \infty. \end{aligned}$$

Therefore condition (12) is proved.

Since S_n is a stochastically bounded sequence and $\hat{a}_{1,n} - a_1 \xrightarrow{P} 0$, we have $\varepsilon_1^n \xrightarrow{P} 0$ as $n \rightarrow \infty$.

Thus we proved that the estimator $\hat{b}_{k,n}^+$ is consistent. The consistency of the estimator $\hat{b}_{k,n}^-$ is proved in the same manner.

Now we prove that the estimator $\hat{d}_{k,n}$ is consistent. We have

$$\hat{d}_{k,n} - d_k = \varepsilon_3^n + \varepsilon_4^n + \varepsilon_5^n,$$

where

$$\begin{aligned} \varepsilon_3^n &= \frac{p - \hat{p}_n}{p\hat{p}_n} \frac{1}{n} \sum_{j=1}^n u_k(\xi_j - \hat{a}_{2,n}), & \varepsilon_4^n &= \frac{1}{pn} \sum_{j=1}^n (u_k(\xi_j - \hat{a}_{2,n}) - u_k(\xi_j - a_2)), \\ \varepsilon_5^n &= \frac{1}{pn} \sum_{j=1}^n u_k(\xi_j - a_2) - \frac{1}{p} \mathbb{E} u_k(\xi_j - a_2). \end{aligned}$$

We consider every term separately. First, $\varepsilon_5^n \rightarrow 0$ almost surely by the strong law of large numbers. Further,

$$\varepsilon_4^n = \frac{1}{pn} \sum_{j=1}^n u_k'(\xi_j - a_2 + \theta(a_2 - \hat{a}_{2,n}))(a_2 - \hat{a}_{2,n}).$$

Here θ denotes a random variable assuming values in the interval $[0, 1]$. Applying the inequality

$$\frac{1}{n} \sum_{j=1}^n y_j \leq \sqrt{\frac{1}{n} \sum_{j=1}^n y_j^2}$$

and condition (ii) of the theorem, we show that the sequence

$$\frac{1}{n} \sum_{j=1}^n u_k'(\xi_j - a_2 + \theta(a_2 - \hat{a}_{2,n}))$$

is stochastically bounded. Hence $\varepsilon_4^n \xrightarrow{P} 0$ as $n \rightarrow \infty$.

We prove similarly that

$$\varepsilon_3^n = \frac{1}{n} \sum_{j=1}^n u_k(\xi_j - a_2 + a_2 - \hat{a}_{2,n})(\hat{p}_n - p) \xrightarrow{P} 0, \quad n \rightarrow \infty,$$

by condition (iii). Thus $\hat{d}_{k,n} - d_k \xrightarrow{P} 0$ as $n \rightarrow \infty$.

Finally we prove that $\hat{c}_{kl,n}$ is consistent. First we write

$$\hat{c}_{kl,n} - c_{kl} = \delta_1^n + \delta_2^n + \delta_3^n + \delta_4^n,$$

where

$$\begin{aligned}\delta_1^n &= \frac{\hat{p}_n^2}{n} \sum_{j=1}^n u_k(\xi_j - \hat{a}_{1,n})u_l(\xi_j - \hat{a}_{1,n}) - p^2 \mathbf{E} u_k(\xi_1 - a_1)u_l(\xi_1 - a_1), \\ \delta_2^n &= \frac{\hat{p}_n(1 - \hat{p}_n)}{n} \sum_{j=1}^n u_k(\xi_j - \hat{a}_{1,n})u_l(\xi_j - \hat{a}_{2,n}) - p(1 - p) \mathbf{E} u_k(\xi_1 - a_1)u_l(\xi_1 - a_2), \\ \delta_3^n &= \frac{\hat{p}_n(1 - \hat{p}_n)}{n} \sum_{j=1}^n u_k(\xi_j - \hat{a}_{2,n})u_l(\xi_j - \hat{a}_{1,n}) - p(1 - p) \mathbf{E} u_k(\xi_1 - a_2)u_l(\xi_1 - a_1), \\ \delta_4^n &= \frac{(1 - \hat{p}_n)^2}{n} \sum_{j=1}^n u_k(\xi_j - \hat{a}_{2,n})u_l(\xi_j - \hat{a}_{2,n}) - (1 - p)^2 \mathbf{E} u_k(\xi_1 - a_2)u_l(\xi_1 - a_2).\end{aligned}$$

Consider δ_1^n . We represent δ_1^n as follows: $\delta_1^n = \varepsilon_6^n + \varepsilon_7^n + \varepsilon_8^n$, where

$$\begin{aligned}\varepsilon_6^n &= \frac{\hat{p}_n^2}{n} \sum_{j=1}^n [u_k(\xi_j - \hat{a}_{1,n})u_l(\xi_j - \hat{a}_{1,n}) - u_k(\xi_j - a_1)u_l(\xi_j - a_1)], \\ \varepsilon_7^n &= (\hat{p}_n^2 - p^2) \frac{1}{n} \sum_{j=1}^n u_k(\xi_j - a_1)u_l(\xi_j - a_1), \\ \varepsilon_8^n &= \frac{p^2}{n} \sum_{j=1}^n u_k(\xi_j - a_1)u_l(\xi_j - a_1) - p^2 \mathbf{E} u_k(\xi_1 - a_1)u_l(\xi_1 - a_1).\end{aligned}\tag{13}$$

By the law of large numbers, $\varepsilon_8^n \xrightarrow{P} 0$ as $n \rightarrow \infty$. Since $\hat{p}_n^2 - p^2 \xrightarrow{P} 0$ as $n \rightarrow \infty$ and since the sequence $\frac{1}{n} \sum_{j=1}^n u_k(\xi_j - a_1)u_l(\xi_j - a_1)$ is stochastically bounded,

$$\varepsilon_7^n \xrightarrow{P} 0, \quad n \rightarrow \infty.\tag{14}$$

Now we consider

$$\varepsilon_6^n = \frac{\hat{p}_n^2}{n} \sum_{j=1}^n (u'_k u_l + u_k u'_l) (\xi_j - a_1 + \theta(a_1 - \hat{a}_{1,n}))(a_1 - \hat{a}_{1,n}).$$

Here θ denotes a random variable assuming values in the interval $[0, 1]$. Applying the inequality

$$\frac{1}{n} \sum_{j=1}^n (u'_k u_l + u_k u'_l) \leq \sqrt{\frac{1}{n} \sum_{j=1}^n (u'_k)^2} \cdot \sqrt{\frac{1}{n} \sum_{j=1}^n (u_l)^2} + \sqrt{\frac{1}{n} \sum_{j=1}^n (u'_l)^2} \cdot \sqrt{\frac{1}{n} \sum_{j=1}^n (u_k)^2},$$

we derive from conditions (ii) and (iii) of the theorem that the sequence

$$\frac{1}{n} \sum_{j=1}^n (u'_k u_l + u_k u'_l)$$

(as well as \hat{p}_n^2) is stochastically bounded in probability and therefore $\hat{a}_{1,n} - a_1 \xrightarrow{P} 0$ as $n \rightarrow \infty$. Thus

$$\varepsilon_6^n \xrightarrow{P} 0, \quad n \rightarrow \infty.\tag{15}$$

Now (13)–(15) imply that $\delta_1^n \xrightarrow{P} 0$ as $n \rightarrow \infty$.

The term δ_4^n is similar to δ_1^n and the proof of the convergence $\delta_4^n \xrightarrow{P} 0$ as $n \rightarrow \infty$ is the same as that of the convergence $\delta_1^n \xrightarrow{P} 0$ as $n \rightarrow \infty$.

Consider δ_2^n . Put $\delta_2^n = \varepsilon_9^n + \varepsilon_{10}^n + \varepsilon_{11}^n$, where

$$\begin{aligned}\varepsilon_9^n &= \frac{\hat{p}_n(1 - \hat{p}_n)}{n} \sum_{j=1}^n [u_k(\xi_j - \hat{a}_{1,n})u_l(\xi_j - \hat{a}_{2,n}) - u_k(\xi_j - a_1)u_l(\xi_j - a_2)], \\ \varepsilon_{10}^n &= \frac{(\hat{p}_n - p)(1 - \hat{p}_n - p)}{n} \sum_{j=1}^n u_k(\xi_j - a_1)u_l(\xi_j - a_2), \\ \varepsilon_{11}^n &= p(1 - p) \left[\frac{1}{n} \sum_{j=1}^n u_k(\xi_j - a_1)u_l(\xi_j - a_2) - \mathbb{E} u_k(\xi_1 - a_1)u_l(\xi_1 - a_2) \right].\end{aligned}$$

By the law of large numbers, $\varepsilon_{11}^n \xrightarrow{p} 0$ as $n \rightarrow \infty$. It is clear that ε_{10}^n is a product of a stochastically bounded sequence and a sequence approaching zero in probability. Hence $\varepsilon_{10}^n \xrightarrow{p} 0$ as $n \rightarrow \infty$.

By the Lagrange formula,

$$\begin{aligned}\varepsilon_9^n &= \frac{\hat{p}_n(1 - \hat{p}_n)}{n} \sum_{j=1}^n [u'_k(\xi_j - a_1 + \theta_1(a_1 - \hat{a}_{1,n}))u_l(\xi_j - a_2 + \theta_2(a_2 - \hat{a}_{2,n}))(a_1 - \hat{a}_{1,n}) \\ &\quad + u'_l(\xi_j - a_2 + \theta_2(a_2 - \hat{a}_{2,n}))u_k(\xi_j - a_1 + \theta_1(a_1 - \hat{a}_{1,n}))(a_2 - \hat{a}_{2,n})].\end{aligned}$$

Here θ_1 and θ_2 are two random variables assuming values in the interval $[0, 1]$. Using conditions (ii) and (iii) of the theorem one can prove that $\varepsilon_9^n \xrightarrow{p} 0$ as $n \rightarrow \infty$.

Therefore $\delta_2^n \xrightarrow{p} 0$ as $n \rightarrow \infty$. The proof of the convergence $\delta_3^n \xrightarrow{p} 0$ as $n \rightarrow \infty$ is completely analogous to the above proof. The theorem is proved. \square

Corollary 4.1. *Assume that all the conditions of Theorem 4.1 hold. If $\det C \neq 0$ and $\text{rank}(B) = 3$, then*

$$\hat{\beta} \xrightarrow{p} \beta^*,$$

where $\hat{\beta} = \hat{C}^{-1}\hat{B}(\hat{B}^T\hat{C}^{-1}\hat{B})^{-1}$.

Proof of Corollary 4.1. Theorem 4.1 implies that $\hat{B} \xrightarrow{p} B$ and $\hat{C} \xrightarrow{p} C$ as $n \rightarrow \infty$ in the sense that every entry of the matrices converges. Thus

$$\hat{C}^{-1}\hat{B}(\hat{B}^T\hat{C}^{-1}\hat{B})^{-1} \rightarrow \beta^*. \quad \square$$

5. ADAPTIVE ESTIMATORS OF EUCLIDEAN PARAMETERS

According to Theorem 4.1, one can construct consistent estimators $\{\hat{\beta}_{im}\}$ for the optimal coefficients $\{\beta_{im}^*\}$ for $i = 1, 2, 3$ and $m = 1, \dots, M$. Substituting these estimators into (4) we obtain the estimators for the optimal estimating functions

$$(16) \quad g_{\hat{\beta}_i}(x) = \sum_{m=1}^M \hat{\beta}_{im} u_m(x).$$

In turn, substituting equality (16) into (3) we obtain the following adaptive system of equations:

$$(17) \quad \pi \hat{g}_{\hat{\beta}_i}(\alpha_1) + (1 - \pi) \hat{g}_{\hat{\beta}_i}(\alpha_2) = 0, \quad i = 1, 2, 3.$$

Introducing the notation

$$\hat{\mathbf{h}}_{\hat{\beta}}(\mathbf{t}) = \begin{pmatrix} \pi \hat{g}_{\hat{\beta}_1}(\alpha_1) + (1 - \pi) \hat{g}_{\hat{\beta}_1}(\alpha_2) \\ \pi \hat{g}_{\hat{\beta}_2}(\alpha_1) + (1 - \pi) \hat{g}_{\hat{\beta}_2}(\alpha_2) \\ \pi \hat{g}_{\hat{\beta}_3}(\alpha_1) + (1 - \pi) \hat{g}_{\hat{\beta}_3}(\alpha_2) \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \pi \\ \alpha_1 \\ \alpha_2 \end{pmatrix},$$

we get the following adaptive estimating equation:

$$(18) \quad \hat{h}(t) = 0,$$

which is equivalent to the preceding one. It is a rather complicated problem to solve this equation. Thus we are going to construct an approximation to the solution of equation (18).

Denote by

$$\tilde{\theta}_n = \begin{pmatrix} \tilde{p}_n \\ \tilde{a}_{1,n} \\ \tilde{a}_{2,n} \end{pmatrix}$$

a pilot estimator of the unknown parameter $\theta = (p, a_1, a_2)^T$. Note that this estimator is \sqrt{n} -consistent. For example, one can take the moment estimator as $\tilde{\theta}_n$. Using the Taylor decomposition for the left-hand side of equation (18) at the point $\tilde{\theta}_n$ we get the following approximating equation:

$$\hat{h}_{\hat{\beta}}(t) \approx \hat{h}_{\hat{\beta}}(\tilde{\theta}_n) + \hat{h}'_{\hat{\beta}}(\tilde{\theta}_n)(t - \tilde{\theta}_n) = 0,$$

where $\hat{h}'_{\hat{\beta}}(t)$ is the Jacobi matrix

$$\hat{h}'_{\hat{\beta}}(t) = \frac{D(h_1, h_2, h_3)}{D(\pi, \alpha_1, \alpha_2)}.$$

Solving the latter equation with respect to t and taking into account the approximate equality $\hat{h}'_{\hat{\beta}}(\theta) \approx E$ (it holds by the normalization condition of Theorem 3.2) we obtain the so-called approximate adaptive estimator of the unknown parameter

$$(19) \quad \check{\theta}_n = \tilde{\theta}_n - \hat{h}_{\hat{\beta}}(\tilde{\theta}_n).$$

Theorem 5.1. *Assume that*

- 1) a pilot estimator $\tilde{\theta}_n$ is \sqrt{n} -consistent;
- 2) conditions (i)–(iii) of Theorem 4.1 hold;
- 3) conditions 4) and 5) of Theorem 3.1 hold;
- 4) $\det C \neq 0$, the rank of the matrix B is equal to 3, where the matrices C and B are defined by equalities (6) and (7), respectively.

Then

$$\sqrt{n}(\check{\theta}_n - \theta) \Rightarrow N(0, \Sigma^*), \quad n \rightarrow \infty.$$

Proof of Theorem 5.1. We rewrite $\sqrt{n}(\check{\theta}_n - \theta)$ as follows:

$$(20) \quad \begin{aligned} \sqrt{n}(\check{\theta}_n - \theta) &= \sqrt{n}(\tilde{\theta}_n - \hat{h}_{\hat{\beta}}(\tilde{\theta}_n) - \theta) \\ &= -\sqrt{n}\hat{h}_{\hat{\beta}}(\theta) + (\tilde{\theta}_n - \theta + \hat{h}_{\hat{\beta}}(\theta) - \hat{h}_{\hat{\beta}}(\tilde{\theta}_n)) \\ &= -\sqrt{n}\hat{h}_{\hat{\beta}}(\theta) + (\tilde{\theta}_n - \theta)\sqrt{n} \left(E - \frac{\partial}{\partial t} \hat{h}_{\hat{\beta}}(t) \Big|_{t=\zeta} \right), \end{aligned}$$

where ζ is a point of the line segment connecting the points θ and $\tilde{\theta}_n$.

We consider each term on the right-hand side of (20) separately. Put

$$h_{u_i}(x, t) = \pi u_i(x - \alpha_1) + (1 - \pi) u_i(x - \alpha_2), \quad \bar{h}(x, t) = \begin{pmatrix} h_{u_1}(x, t) \\ \vdots \\ h_{u_M}(x, t) \end{pmatrix},$$

$$\hat{h}(\theta) = \frac{1}{n} \sum_{j=1}^n \bar{h}(\xi_j, \theta).$$

Then

$$h_\beta(x, t) = \beta^T \bar{h}(x, t).$$

Therefore the first term in (20) is of the form

$$-\sqrt{n}\hat{h}_{\hat{\beta}}(\theta) = -\sqrt{n}\hat{h}_{\beta^*}(\theta) + (\hat{\beta} - \beta^*)^T \sqrt{n}\hat{h}(\theta).$$

Using relation (5) and Theorem 3.2,

$$\sqrt{n}\hat{h}_{\beta^*}(\theta) \Rightarrow N(0, \Sigma^*).$$

Condition 2) of the theorem implies that $\sqrt{n}\hat{h}(\theta) = O_p(1)$. Since $(\hat{\beta} - \beta^*) \xrightarrow{p} 0$ by Corollary 4.1,

$$-\sqrt{n}\hat{h}_{\hat{\beta}}(\theta) \Rightarrow N(0, \Sigma^*), \quad n \rightarrow \infty.$$

Consider the second term in (20). We have

$$\left. \frac{\partial}{\partial t} \hat{h}_{\hat{\beta}}(t) \right|_{t=\zeta} = \hat{\beta}^T \left. \frac{\partial}{\partial t} \hat{h}_{\hat{\beta}}(t) \right|_{t=\zeta} = \hat{\beta}^T \hat{B}(\zeta),$$

where the entries of the matrix \hat{B} are given by (10) with ζ being substituted for θ . By Corollary 4.1 we get $\hat{B} \rightarrow B$ in the sense that every entry of the matrices converges. Hence

$$\hat{\beta}^T \hat{B}(\zeta) \rightarrow E = \beta^T B$$

and

$$E - \left. \frac{\partial}{\partial t} \hat{h}_{\hat{\beta}}(t) \right|_{t=\zeta} \rightarrow 0$$

in probability. Since $\tilde{\theta}_n$ is a \sqrt{n} -consistent estimator of θ , the second term in (20) approaches 0. The theorem is proved. \square

BIBLIOGRAPHY

1. R. Maïboroda and O. Sugakova, *Estimation of Euclidean parameters of a mixture of two symmetric distributions*, Ukrain. Mat. Zh. **62** (2010), 945–953. (Ukrainian)
2. L. Bordes, S. Mottelet, and P. Vandekerkhove, *Semiparametric estimation of a two-component mixture model*, Ann. Statist. **34** (2006), 1204–1232. MR2278356 (2008e:62064)
3. D. R. Hunter, S. Wang, and T. R. Hettmansperger, *Inference for mixtures of symmetric distributions*, Ann. Statist. **35** (2007), 224–251. MR2332275 (2008g:62079)
4. R. Maïboroda, *Estimation of mean positions and concentrations from observations of a two-component mixture of symmetric distributions*, Teor. Imovir. Mat. Stat. **78** (2008), 132–140; English transl. in Theory Probab. Math. Statist. **78** (2009), 147–156. MR2446855 (2010b:62134)
5. J. Shao, *Mathematical Statistics*, Springer-Verlag, New York, 1998. MR2002723 (2004g:62002)

DEPARTMENT OF MATHEMATICS AND THEORETICAL RADIOPHYSICS, FACULTY FOR RADIOPHYSICS, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE 2, KIEV 03127, UKRAINE
E-mail address: sugak@univ.kiev.ua

Received 17/FEB/2010

Translated by S. KVASKO