

ADAPTIVE ESTIMATING EQUATIONS FOR A LOCATION PARAMETER CONSTRUCTED BY USING OBSERVATIONS WITH ADMIXTURE

UDC 519.21

R. MAĬBORODA AND O. SUGAKOVA

ABSTRACT. A model of observations with admixture is considered. The distribution of the primary component of the model is symmetric and unknown, while the distribution of the admixture is known. Adaptive estimators are constructed for the median of the distribution of the primary component. The asymptotic variance of this estimator is close to the minimal asymptotic variance among all estimating equations estimators.

1. INTRODUCTION

This paper is a prolongation of [1]. Estimating equations are constructed in [1] for the estimation of the location parameter of a symmetric distribution by using the observations with admixture. A lower bound for the asymptotic variance of the estimating equations estimators is also obtained in [1]. The “optimal” estimating function for which this lower bound is attained depends on unknown parameters and cannot thus be used directly for the estimation.

We study possible adaptive estimating functions constructed from the same observations in order to approximate the optimal estimating function. We consider a class of estimating functions that can be constructed from linear combinations of some fixed basic functions. Optimal (with respect to the minimum of the asymptotic variance of an estimator) coefficients for such linear combinations are found. Estimators for those optimal coefficients are found from the sample. We show that the estimator constructed by using the estimating function with these optimal coefficients has the same asymptotic variance as the estimator with the true optimal coefficients.

Such an adaptive scheme is not new (see, for example, [2]). However it has not been used up to now for observations with admixture. It is worth mentioning that the estimator obtained below is not adaptive in the sense of the paper [3], since its asymptotic variance is larger than that of the optimal estimator that can be constructed if the density f of the centered primary component is known. A true adaptive estimator for the model considered in this paper cannot be constructed at all, since the necessary information conditions are not satisfied in this case (see [3, Section 2.4]). Nevertheless a comparison of these estimators with nonadaptive moment estimators shows that the adaptive procedure improves the effectiveness essentially.

2000 *Mathematics Subject Classification.* Primary 62G07; Secondary 62G20.

Key words and phrases. Method of moments, a finite mixture of probability distributions, consistence, asymptotic normality, coefficient of dispersion.

Supported by the Swedish Institute grant SI-01424/2007.

2. SETTING OF THE PROBLEM

Let the observed data form a sample $\Xi_N = (\xi_1, \dots, \xi_N)$ of independent identically distributed random variables possessing the distribution density

$$(1) \quad \psi(x) = pf(x-a) + (1-p)f_0(x),$$

where $p \in (0, 1)$ is the probability that an observation corresponds to the primary component; a is a median of the distribution of the primary component; f is the distribution density of the deviation of the primary component from a ; f_0 is the distribution density of the admixture.

We assume that the function f_0 is known but the parameters $a \in \mathbb{R}$ and $p \in (0, 1)$ are unknown. The density f is unknown but it is symmetric about the origin, that is, $f(-x) = f(x)$ for all $x \in \mathbb{R}$.

In what follows we assume that f is a continuously differentiable function in \mathbb{R} .

The problem is to estimate the location parameter a of the primary component. (A detailed description of this model and its applications is given in [1, 5].)

An unbiased estimating equation for a (called a generalized estimating equation) is proposed in [1]. This equation is constructed with the help of two odd functions g_1 and g_2 . The equation is written as

$$(2) \quad \hat{h}(\alpha) = 0,$$

where

$$\begin{aligned} \hat{h}(\alpha) &= \frac{1}{N} \sum_{j=1}^N h(\xi_j, \alpha), \\ h(x, \alpha) &= G_2(\alpha)g_1(x-\alpha) - G_1(\alpha)g_2(x-\alpha), \\ G_i(\alpha) &= \int_{-\infty}^{+\infty} g_i(x-\alpha)f_0(x) dx, \quad i = 1, 2. \end{aligned}$$

It is clear that the functions g_1 and g_2 are to be chosen such that $G_1(\alpha) < \infty$ and $G_2(\alpha) < \infty$ for all $\alpha \in \mathbb{R}$.

Any almost sure solution $\hat{a}_N = \hat{a}(\Xi_N)$ of equation (2) is called a GEE-estimator of the parameter a constructed from the estimating pair (g_1, g_2) .

As a rule, equation (2) has several roots and the choice of a ‘‘correct’’ one for the construction of a consistent estimator is a nontrivial problem. It is shown in [1, 5] that, under some additional assumptions, one can construct GEE-estimators \tilde{a}_N satisfying equality (2). Moreover, these estimators are \sqrt{N} consistent, that is, $\sqrt{N}(\tilde{a}_N - a) = O_p(1)$.

Recall that the equality $\zeta_N = O_p(\varepsilon_N)$ means that

$$\limsup_{N \rightarrow \infty} \mathbf{P}\{|\zeta_N| > C\varepsilon_N\} \rightarrow 0 \quad \text{as } C \rightarrow \infty$$

for a sequence of random variables ζ_N and a sequence of numbers $\varepsilon_N > 0$.

If a mild condition is imposed on the estimating pair and if the estimator \hat{a}_N is consistent, then, according to Theorem 2.1 and equality (11) in [1], the distribution of $\sqrt{N}(\hat{a}_N - a)$ weakly converges to the distribution $N(0, \sigma_{(g_1, g_2)}^2)$, where

$$(3) \quad \sigma_{(g_1, g_2)}^2 = \frac{\int U^2 Q}{(\int UB)^2}$$

(here we use the notation $\int A = \int_0^{+\infty} A(x) dx$),

$$(4) \quad \begin{aligned} U(x) &= G_2(a)g_1(x) - G_1(a)g_2(x), \\ Q(x) &= 2pf(x) + (1-p)(f_0(x+a) + f_0(-x+a)), \\ B(x) &= 2pf'(x). \end{aligned}$$

Let $D(x) = f_0(x+a) - f_0(-x+a)$. Note that

$$(5) \quad \int UD = 0$$

for functions U defined by (4).

The converse statement also holds: if $D \not\equiv 0$, then, given an odd function U satisfying (5), one can find an estimating pair (g_1, g_2) such that (4) holds. Indeed, one can choose $g_1(x) = U(x)/\beta_0$ and find an arbitrary odd function g_2 such that

$$\int g_2(x-\alpha)f_0(x) dx < \infty$$

for all $\alpha \in \mathbb{R}$ and $\beta_0 = \int gD \neq 0$.

Other choices of the estimating pair that lead to the same function U are also possible. These estimating pairs generate different estimators, but they have the same asymptotic variance $\sigma_U^2 = \sigma_{(g_1, g_2)}^2$ defined by (3) if the parameters (a, p, f) of the model are fixed.

It is natural to choose an estimating pair such that the corresponding asymptotic variance is as small as possible. The precise lower bound for σ_U^2 is found in [1] for a wide class of functions U . The corresponding estimating pair cannot be used explicitly for the estimation, since it depends on unknown parameters of the model. Thus we apply the following “linear adaptive” method.

Consider the functions U that can be represented as linear combinations of a finite set of basic odd functions u_1, \dots, u_M , that is,

$$(6) \quad U(x) = U_{\beta} = \sum_{m=1}^M \beta^m u_m(x),$$

where $\beta = (\beta^1, \dots, \beta^M) \in \mathbb{R}^M$ is an arbitrary vector of coefficients. First we find the vector of optimal coefficients β_* for which $\sigma_{U_{\beta}}^2$ attains its minimum for given (unknown) parameters of the model. Then we estimate these optimal coefficients from the sample Ξ_N and obtain the vector of estimators $\tilde{\beta}_N$. The estimated function $U_{\tilde{\beta}_N}$ is used to construct the estimating pair $(\tilde{g}_1, \tilde{g}_2)$ and the corresponding estimating function \tilde{h} . Note that the coefficient β_0 should also be estimated from the sample.

Now one can use the latter estimating function to find an estimator as a solution of the equation $N^{-1} \sum_{j=1}^N \tilde{h}(\xi_j, \alpha) = 0$. However, the question on how to choose a “correct” root of the estimating equation is still open. Therefore the question on whether the estimator is consistent is also open. To avoid this problem one can propose a two-step estimation procedure. The first step is to estimate a with the help of a “pilot” estimator. This estimator should be chosen to be \sqrt{N} -consistent. However, the asymptotic variance of this estimator is far of being optimal. This estimator is improved on the second step with the help of the estimating equation. The second step can also be performed in a variety of ways. For example, as a resulting estimator, one can choose a root of the estimating equation that is the closest one to the pilot estimator. Another approach is to change the true estimating function by its linear approximation in a neighborhood of the pilot estimator and take the solution of the approximating equation as a resulting estimator. This approximating approach for constructing estimators is described in [1]

for nonadaptive estimating equations. We follow the approximating approach when constructing the resulting adaptive estimator. The pilot estimators for a and p are used for the estimation of the optimal coefficients β_* , as well.

It is clear that the resulting estimator is not a GEE-estimator in the sense of (2). Nevertheless we show that its asymptotic variance is equal to $\sigma_{U_{\beta_*}}^2$, that is, to the least possible value of GEE-estimators constructed with the help of linear combinations of basic functions.

3. OPTIMAL COEFFICIENTS β_*

We start with finding the coefficients β_* for which $\sigma_{U_{\beta}}^2$ attains its minimal value. In what follows we assume that u_m , $m = 1, \dots, M$, are odd, continuously differentiable functions. It is easy to see that

$$\int U_{\beta}^2 Q = \beta^T \mathbf{Q}$$

for all $\beta \in \mathbb{R}^M$, where $\mathbf{Q} = (q_{ik})_{i,k=1}^M$,

$$(7) \quad \begin{aligned} q_{ik} &= \int u_i u_k Q = \int_{-\infty}^{+\infty} u_i(x-a) u_k(x-a) (pf(x-a) + (1-p)f_0(x)) dx \\ &= \mathbb{E} u_i(\xi_1 - a) u_k(\xi_1 - a), \end{aligned}$$

$$(8) \quad \int U_{\beta} B = \mathbf{B}\beta,$$

and where $\mathbf{B} = (b_1, \dots, b_M)^T$,

$$(9) \quad \begin{aligned} b_i &= \int u_i B = \int_{-\infty}^{+\infty} u_i(x) p f'(x) dx = -p \int_{-\infty}^{+\infty} u_i'(x) f(x) dx \\ &= -\mathbb{E} u_i'(\xi_1 - a) + (1-p) \mathbb{E} u_i'(\eta_0 - a). \end{aligned}$$

Here η_0 is a random variable with the distribution density f_0 . For the latter equality we used the property that

$$\mathbb{E} v(\xi_1) = \int_{-\infty}^{+\infty} v(x) \psi(x) dx = p \int_{-\infty}^{+\infty} v(x) f(x-a) dx + (1-p) \int_{-\infty}^{+\infty} v(x) f_0(x) dx$$

for an arbitrary function v .

Finally $\int U_{\beta} D = \mathbf{D}^T \beta$, where $\mathbf{D} = (d_1, \dots, d_M)^T$ and

$$(10) \quad d_m = \int u_m D = \mathbb{E} u_m(\eta_0 - a).$$

Considering equalities (3) and (5) we conclude that one needs to solve the minimization problem $\beta^T \mathbf{Q} \beta / (\mathbf{B}^T \beta)^2 \rightarrow \min$ given $\mathbf{D}^T \beta = 0$ to determine the coefficients β_* . Since both the minimization criterion and restrictions do not change when passing from β to $c\beta$, where $c \in \mathbb{R}$ is an arbitrary number, one can introduce an extra normalizing condition $\mathbf{B}^T \beta = 1$. Then the problem can be written as follows:

$$\begin{cases} \beta^T \mathbf{Q} \beta \rightarrow \min, \\ \mathbf{D}^T \beta = 0, \\ \mathbf{B}^T \beta = 1. \end{cases}$$

Solving this problem by the Lagrange multipliers method we show that if $\det \mathbf{Q} \neq 0$, then the minimum of $\sigma_{U_{\beta}}^2$ is attained at the point $\beta = \beta_* = (\beta_*^1, \dots, \beta_*^M)^T$, where

$$(11) \quad \beta_* = \mathbf{Q}^{-1}(\lambda_1 \mathbf{B} + \lambda_2 \mathbf{D}),$$

$$(12) \quad \lambda_1 = \frac{1}{\Delta} \mathbf{D}^T \mathbf{Q}^{-1} \mathbf{D}, \quad \lambda_2 = \frac{1}{\Delta} \mathbf{D}^T \mathbf{Q}^{-1} \mathbf{B},$$

$$(13) \quad \Delta = \mathbf{D}^T \mathbf{Q}^{-1} \mathbf{D} \cdot \mathbf{B}^T \mathbf{Q}^{-1} \mathbf{B} - (\mathbf{D}^T \mathbf{Q}^{-1} \mathbf{B})^2.$$

Moreover $\sigma_{U_{\beta_*}}^2 = \lambda_1$.

Therefore we have proved the following result.

Theorem 3.1. *Let an estimating pair (g_1, g_2) be such that the function U defined by (4) is of the form $U(x) = U_{\beta}(x) = \sum_{i=1}^M \beta^i u_i(x)$, where u_i are continuously differentiable odd functions. Assume that $\det \mathbf{Q} \neq 0$ and $\mathbf{D} \neq 0$.*

Then the asymptotic variance of the GEE-estimator \hat{a}_N obtained from the estimating pair (g_1, g_2) cannot be less than the number λ_1 defined by the first equality in (12). The value of λ_1 is attained for the point $\beta = \beta_$ defined by equality (11).*

4. ESTIMATORS OF OPTIMAL COEFFICIENTS

Let \tilde{a}_N and \tilde{p}_N be consistent pilot estimators for the parameters a and p , respectively. Note that one can take the moment estimators [1, 5] as pilot estimators. By definitions (7)–(10), one can consider the following estimators for \mathbf{Q} , \mathbf{D} , and \mathbf{B} :

$$\begin{aligned} \tilde{\mathbf{Q}} &= (\tilde{q}_{ik})_{i,k=1}^M, & \tilde{q}_{ik} &= \frac{1}{N} \sum_{j=1}^N u_i(\xi_j - \tilde{a}_N) u_k(\xi_j - \tilde{a}_N), \\ \tilde{\mathbf{D}} &= (\tilde{d}_i)_{i=1}^M, & \tilde{d}_i &= \int_{-\infty}^{+\infty} u_i(x - \tilde{a}_N) f_0(x) dx, \\ \tilde{\mathbf{B}} &= (\tilde{b}_i)_{i=1}^M, & \tilde{b}_i &= -\frac{1}{N} \sum_{j=1}^N u'_i(\xi_j - \tilde{a}_N) + (1 - \tilde{p}_N) \int_{-\infty}^{+\infty} u'_i(x - \tilde{a}_N) f_0(x) dx. \end{aligned}$$

The estimator $\tilde{\beta}_N = (\tilde{\beta}_N^1, \dots, \tilde{\beta}_N^M)^T$ for β_* is determined by equalities (11)–(13) where the true values \mathbf{Q} , \mathbf{B} , and \mathbf{D} are changed by their approximate values $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{D}}$, respectively.

Theorem 4.1. *Assume that*

- 1) *the pilot estimators \tilde{a}_N and \tilde{p}_N are \sqrt{N} -consistent.*

Further we assume that, for all $m = 1, \dots, M$,

- 2) $\mathbf{E}(u_m(\xi_1 - a))^2 < \infty$;
- 3) u_m is a twice continuously differentiable function;
- 4) for some $\varepsilon > 0$,

$$\begin{aligned} e_{1,m} &= \mathbf{E} \sup_{\alpha: |a-\alpha| < \varepsilon} (u'_m(\xi_1 - \alpha))^2 < \infty, \\ e_{2,m} &= \mathbf{E} \sup_{\alpha: |a-\alpha| < \varepsilon} |u''_m(\xi_1 - \alpha)| < \infty. \end{aligned}$$

Then \tilde{q}_{ik} , \tilde{b}_m , \tilde{d}_m , and $\tilde{\beta}_^m$ are \sqrt{N} -consistent estimators of q_{ik} , b_m , d_m , and β_*^m , respectively.*

5. ADAPTIVE ESTIMATOR

Now we describe an estimator of the parameter a with the help of the adaptive function $U_{\tilde{\beta}}$. We add an odd function u_0 to the collection of the basic functions u_1, \dots, u_M . The function u_0 may coincide with u_i for some $i \in \{1, \dots, M\}$. Put

$$G_i(\alpha) = \int_{-\infty}^{+\infty} u_i(x - \alpha) f_0(x) dx.$$

Let \tilde{a}_N and \tilde{p}_N be the pilot estimators for a and p , respectively. Consider the estimating function

$$h_{\beta^0, \beta}(x, \alpha) = \frac{1}{\beta^0} \sum_{m=1}^M \beta^m (u_m(x - \alpha) G_0(\alpha) - G_i(\alpha) u_0(x - \alpha)).$$

The adaptive estimator is given by

$$\check{a}_N = \tilde{a}_N - \frac{1}{N} \sum_{j=1}^N h_{G_0(\tilde{a}_N), \tilde{\beta}_N}(\xi_j, \tilde{a}_N) = \tilde{a}_N - \sum_{m=1}^M \tilde{\beta}_N^m \left(\hat{u}_m(\tilde{a}_N) - \frac{G_m(\tilde{a}_N)}{G_0(\tilde{a}_N)} \hat{u}_0(\tilde{a}_N) \right),$$

where

$$\hat{u}_m(\alpha) = \frac{1}{N} \sum_{j=1}^N u_m(\xi_j - \alpha).$$

Remark. If the same pilot estimator is used for the estimation of optimal coefficients and for the construction of \check{a}_N , then $\tilde{\mathbf{D}}^T \tilde{\beta} = 0$ implies that $\sum_{m=1}^M G_m(\tilde{a}_N) \beta_N^m = 0$, whence we obtain

$$\check{a}_N = \tilde{a}_N - \sum_{m=1}^M \tilde{\beta}_N^m \hat{u}_m(\tilde{a}_N).$$

Theorem 5.1. *Assume that*

- 1) $\det \mathbf{Q} \neq 0$;
- 2) $\beta_*^0 = G_0(a) \neq 0$;
- 3) *the assumptions of Theorem 4.1 hold and moreover conditions 2)–4) hold for $m = 0$ as well.*

Then $\sqrt{N}(\check{a}_N - a) \Rightarrow N(0, \sigma_{U_{\beta_}}^2)$.*

Here and in what follows, the symbol \Rightarrow stands for the weak convergence.

6. RESULTS OF SIMULATION

In order to obtain information on about the quality of the adaptive estimator for samples of a fixed size, we simulated samples from the distribution density ψ defined in (1) with

$$f(x) = s^{-1} \varphi(x/s) \quad \text{and} \quad f_0(x) = \varphi(x),$$

where φ is the standard normal density and $p = a = s = 0.5$. Therefore both the primary component and the admixture have a normal distribution for the simulated data.

As shown in [1] for the Gaussian case, the asymptotic variance of the nonadaptive moment estimator \tilde{a}_N (that is, of the GEE-estimator with the estimating pair $g_1(x) = x$, $g_2(x) = x^3$) is equal to $\sigma_{\text{moment}}^2 = 3.97$, while the minimal variance of a GEE-estimator is $\sigma_*^2 = 1.34$.

For the adaptive procedure, we have chosen the following three basic functions:

$$u_1(x) = x, \quad u_2(x) = x^3, \quad u_3(x) = x^5.$$

For the moment estimator proposed in [1], the function u_3 is used in the additional estimating pair to pick up the correct root of the estimating equation and, in fact, the

adaptive estimator is constructed with the same estimating functions as those used to construct the moment estimator. The asymptotic variance of the adaptive estimator described in Section 5 equals $\sigma_{\text{adapt}}^2 = 2.27$ for our case; that is, the asymptotic variance is between σ_{moment}^2 and σ_*^2 .

The moment estimator \tilde{a}_N and the estimator $\tilde{p}_N = 1 - (\bar{\xi} - \tilde{a}_N)/(\mathbf{E}\eta_0 - \tilde{a}_N)$, where $\bar{\xi} = N^{-1} \sum_{j=1}^N \xi_j$, have been used as pilot estimators for the adaptive procedure.

The results for the comparison of moment and adaptive estimators are given in the table below.

N	50	75	100	250	500	1000	2000	4000
$N \cdot s_{\text{moment}}^2$	15.68	9.59	8.26	4.37	4.85	3.99	3.93	3.88
$N \cdot s_{\text{adapt}}^2$	10.88	4.43	4.49	3.38	3.30	2.67	2.42	2.33
b_{moment}	0.07	0.018	0.03	0.008	0.008	0.002	0.0008	0.00008
b_{adapt}	-0.097	-0.063	-0.037	-0.029	-0.0076	-0.0012	0.0024	0.0017
$N \cdot MSE_{\text{moment}}$	15.93	9.62	8.36	4.39	4.89	3.995	3.93	3.876
$N \cdot MSE_{\text{adapt}}$	11.35	4.72	4.63	3.59	3.33	2.67	2.43	2.34
P	0.796	0.886	0.897	0.985	0.999	1	1	1

The number N in the above table denotes the number of observations, the word “moment” in the subscripts indicates the moment estimators, while the word “adapt” stands for the adaptive estimators, s^2 denotes the sampling variance, MSE is the mean square deviation of the estimators, and b is the bias of the estimators. The values of these characteristics are evaluated from 1000 simulated samples. As the moment estimator, we always choose the “correct” root of the estimating equation. Hence these characteristics do not involve an error due to the incorrect choice of a root of the estimating equation for the moment estimator. The frequency P of a correct choice of the root for the moment estimator is given in the bottom row of the table.

As seen from the above table the asymptotic results give an adequate picture of the variances of estimators only if the size of the samples is large enough (starting with 1000 observations). Nevertheless, the variance of adaptive estimators is essentially better than the variance of moment estimators for all sample sizes. The bias of the adaptive estimators is larger than that of the moment estimators. On the other hand, the contribution of the bias to the mean square deviation of the estimator is negligible as compared to the contribution of the variance.

7. CONCLUDING REMARKS

The adaptive procedure for the estimating equations described above allows one to construct estimators for the median of a symmetric distribution from observations with admixture. The adaptive estimators turn out to be essentially better than nonadaptive moment estimators. The question on how to choose optimally the basic functions u_m is a matter for further considerations.

8. PROOF OF THE RESULTS

Proof of Theorem 4.1. 1. Fix two arbitrary integer numbers $i, k \in \{1, \dots, M\}$. We show that $\sqrt{N}(\tilde{q}_{ik} - q_{ik}) = O_p(1)$. First we estimate

$$J = \left| \sqrt{N}(\tilde{q}_{ik} - q_{ik}) \right| = \left| \frac{1}{\sqrt{N}} \sum_{j=1}^N (u_i(\xi_j - \tilde{a}_N)u_k(\xi_j - \tilde{a}_N) - \mathbf{E} u_i(\xi_j - a)u_k(\xi_j - a)) \right|$$

$$\leq J_1 + J_2,$$

where

$$J_1 = \frac{1}{\sqrt{N}} \sum_{j=1}^N (u_i(\xi_j - a)u_k(\xi_j - a) - \mathbf{E} u_i(\xi_j - a)u_k(\xi_j - a)) = O_p(1)$$

in view of condition 2) of Theorem 4.1 and Chebyshev's inequality. Then

$$\begin{aligned} J_2 &= \left| \frac{1}{\sqrt{N}} \sum_{j=1}^N (u_i(\xi_j - \tilde{a}_N)u_k(\xi_j - \tilde{a}_N) - u_i(\xi_j - a)u_k(\xi_j - a)) \right| \\ &\leq J_3(\tilde{a}_N - a)\sqrt{N}, \end{aligned}$$

where

$$J_3 = \frac{1}{N} \sum_{j=1}^N |u'_i(\xi_j - \zeta_N)u_k(\xi_j - \zeta_N) + u_i(\xi_j - \zeta_N)u'_k(\xi_j - \zeta_N)|,$$

and ζ_N is an intermediate point between \tilde{a}_N and a . Since \tilde{a}_N is a \sqrt{N} -consistent estimator of a , it remains to prove that $J_3 = O_p(1)$.

Note that

$$J_3 \leq \frac{1}{2}(S'_i + S_i + S'_k + S_k),$$

where

$$\begin{aligned} S_m &= \frac{1}{N} \sum_{j=1}^N (u_m(\xi_j - \zeta_N))^2, \\ S'_m &= \frac{1}{N} \sum_{j=1}^N (u'_m(\xi_j - \zeta_N))^2. \end{aligned}$$

Using the method of [6] we will show that $S'_m = O_p(1)$. According to condition 4), there exists a number $\varepsilon > 0$ such that $e_{1,m} < \infty$. Consider

$$(14) \quad S'_m = S'_m \mathbb{1}\{|\zeta_N - a| > \varepsilon\} + S'_m \mathbb{1}\{|\zeta_N - a| < \varepsilon\}.$$

Here and in what follows, $\mathbb{1}\{A\}$ denotes the indicator of a set A . Since ζ_N is an intermediate point between a and its consistent estimator \tilde{a}_N , we have

$$\mathbf{P}\{S'_m \mathbb{1}\{|\zeta_N - a| > \varepsilon\} > C\} \leq \mathbf{P}\{\mathbb{1}\{|\zeta_N - a| > \varepsilon\} > 0\} \rightarrow 0 \quad \text{as } N \rightarrow \infty$$

for all $C > 0$. Thus $S'_m \mathbb{1}\{|\zeta_N - a| > \varepsilon\} = O_p(1)$.

For the second term in (14) we have

$$S'_m \mathbb{1}\{|\zeta_N - a| < \varepsilon\} \leq \frac{1}{N} \sum_{j=1}^N \sup_{\alpha: |a-\alpha| < \varepsilon} (u'_m(\xi_j - \alpha))^2 = O_p(1)$$

by condition 4) and the law of large numbers.

Hence $S'_m = O_p(1)$. Similarly we prove that $S_m = O_p(1)$. Therefore $J_3 = O_p(1)$, whence $J \leq J_1 + J_3\sqrt{N}(\tilde{a}_N - a) = O_p(1)$.

2. We prove that \tilde{d}_m is a \sqrt{N} -consistent estimator of d_m . First,

$$\mathbf{P}\{\sqrt{N}(\tilde{d}_m - d_m) > C\} \leq \mathbf{P}\{\sqrt{N}(\tilde{d}_m - d_m) > C, |\tilde{a}_N - a| \leq \varepsilon\} + \mathbf{P}\{|\tilde{a}_N - a| > \varepsilon\}.$$

Note that the second probability tends to 0. The first probability is estimated as follows:

$$\begin{aligned} & \left| \sqrt{N}(\tilde{d}_m - d_m) \mathbb{1}\{|\tilde{a}_N - a| < \varepsilon\} \right| \\ & \leq \sqrt{N} \int_{-\infty}^{+\infty} |u_i(x - \tilde{a}_N) - u_i(x - a)| \mathbb{1}\{|\tilde{a}_N - a| < \varepsilon\} f_0(x) dx \\ & \leq \int_{-\infty}^{+\infty} f_0(x) \sup_{\zeta: |\zeta - a| < \varepsilon} |u'_m(x - \zeta)| \cdot |\sqrt{N}(\tilde{a}_N - a)| dx \leq C' e_{1,m} |\sqrt{N}(\tilde{a}_N - a)|. \end{aligned}$$

Since \tilde{a}_N is a \sqrt{N} -consistent estimator, the latter result implies that \tilde{d}_m also is a \sqrt{N} -consistent estimator.

3. The proof of the \sqrt{N} -consistence of the estimator \tilde{b}_m follows the same idea of the estimation as that used in Steps 1 and 2 above. Since the construction of an estimator requires u'_m , the proof for \tilde{b}_m requires the extra conditions on u''_m listed in the statement of the theorem.

4. The \sqrt{N} -consistence of \tilde{b}_N^m follows from the \sqrt{N} -consistence of the estimators \tilde{d}_{ik} , \tilde{b}_m , and \tilde{d}_m , since the limit matrix \mathbf{D} is nondegenerate.

The theorem is proved. □

Proof of Theorem 5.1. Consider

$$(15) \quad a_N^* = \tilde{a}_N - \frac{1}{N} \sum_{j=1}^N h_{\beta_*^0, \beta_*}(\xi_j, \tilde{a}_N).$$

We show that

$$\sqrt{N}(\tilde{a}_N - a_N^*) \rightarrow 0$$

in probability and that

$$(16) \quad \sqrt{N}(a_N^* - a) \Rightarrow N(0, \sigma_{\tilde{U}_{\beta_*}}^2).$$

This implies the statement of the theorem, indeed.

First we prove (15). Let

$$\hat{h}_N^m(\alpha) = \frac{1}{N} \sum_{j=1}^N (u_m(\xi_j - \alpha)G_0(\alpha) - u_0(\xi_j - \alpha)G_m(\alpha)).$$

As in the proof of Theorem 4.1 we prove that $\tilde{\beta}_N^0 = G_0(\tilde{a}_N)$ is a \sqrt{N} -consistent estimator of β_*^0 . Thus

$$\begin{aligned} \left| \sqrt{N}(\tilde{a}_N - a_N^*) \right| &= \left| \frac{1}{\sqrt{N}} \sum_{j=1}^N \left(h_{\tilde{\beta}_N^0, \tilde{\beta}}(\xi_j, \tilde{a}_N) - h_{\beta_*^0, \beta_*}(\xi_j, \tilde{a}_N) \right) \right| \\ &\leq \sum_{j=1}^M \sqrt{N} \left| \frac{\tilde{\beta}_N^m}{\tilde{\beta}_N^0} - \frac{\beta_*^m}{\beta_*^0} \right| \cdot |\hat{h}_N^m(\tilde{a}_N)| \leq \sum_{m=1}^M O_p(1) |\hat{h}^m(\tilde{a}_N)|. \end{aligned}$$

Using again the method of the proof of Theorem 4.1 we obtain that

$$\hat{h}_N^m(\tilde{a}_N) \rightarrow \mathbf{E} \hat{h}_N^m(a) = 0$$

in probability. Hence $|\sqrt{N}(\tilde{a}_N - a_N^*)| \rightarrow 0$ in probability as $N \rightarrow \infty$.

Now we prove (16). Note that

$$\hat{h}_{\beta^*}(\tilde{a}_N) = \hat{h}_{\beta^*}(a) + (\tilde{a}_N - a)\hat{h}'_{\beta^*}(\zeta)$$

for some intermediate point ζ between a and \tilde{a}_N . Hence

$$\begin{aligned}\sqrt{N}(a_N^* - a) &= \sqrt{N}(\tilde{a}_N - \hat{h}_{\beta^*}(\tilde{a}_N)) = \sqrt{N}(\tilde{a}_N - (\tilde{a}_N - a)\hat{h}'_{\beta^*}(\zeta) + \hat{h}_{\beta^*}(a)) \\ &= \sqrt{N}(\tilde{a}_N - a)(1 - \hat{h}'_{\beta^*}(\zeta)) + \sqrt{N}\hat{h}_{\beta^*}(a).\end{aligned}$$

As shown in [1], $\mathbf{E} \hat{h}'_{\beta} = \int U_{\beta} B$. By construction of β^* , we have $\int U_{\beta^*} B = \mathbf{B}^T \beta^* = 1$. Since $\zeta \rightarrow a$ as $N \rightarrow \infty$ and since $\mathbf{E} \hat{h}'_{\beta^*} = \int U_{\beta^*} B = 1$, we get $\hat{h}'_{\beta^*}(\zeta) \rightarrow 1$.

Therefore

$$\sqrt{N}(a_N^* - a) \sim \sqrt{N}\hat{h}_{\beta^*}(a) \Rightarrow N(0, \sigma_{U_{\beta^*}}^2).$$

The theorem is proved. □

BIBLIOGRAPHY

1. O. V. Sugakova, *An estimator of the location parameter from observations with admixture*, Teor. Imovirnost. Matem. Statyst. **80** (2009), 128–137; English transl. in this issue. MR2541959 (2010e:62077)
2. A. A. Tsiatis, *Semiparametric Theory and Missing Data*, Springer, Berlin, 2006. MR2233926 (2007g:62009)
3. P. Bickel, C. A. J. Klaassen, Ya. Ritov, and J. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, J. Hopkins University Press, Baltimore–London, 1993. MR1245941 (94m:62007)
4. A. A. Borovkov, *Mathematical Statistics*, Nauka, Moscow, 1984; English transl., Gordon and Breach Science Publishers, Amsterdam, 1998. MR782295 (86i:62001); MR1712750 (2000f:62003)
5. L. Bordes, C. Delmas, and P. Vandekerckhove, *Semiparametric estimation of a two-component mixture model where one component is known*, Scand. J. Statist. **33** (2006), 733–752. MR2300913 (2008f:62049)
6. J. Shao, *Mathematical Statistics*, Springer-Verlag, New York, 1998. MR2002723 (2004g:62002)

DEPARTMENT OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS, FACULTY FOR MECHANICS AND MATHEMATICS, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE, 2, KIEV 03127, UKRAINE

E-mail address: sugak@univ.kiev.ua

DEPARTMENT OF HIGHER 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 16/SEP/2008

Translated by S. V. KVASKO