AN ESTIMATOR OF THE LOCATION PARAMETER
OBTAINED FROM OBSERVATIONS WITH ADMIXTURE

UDC 519.21

O. SUGAKOVA

ABSTRACT. We consider a model of observations from a two-component mixture such that the distribution of the admixture is known, while the distribution of the other component (treated as the primary one) is unknown. We assume that the distribution of the primary component is symmetric about the location parameter. We propose a method for constructing an unbiased estimating equation for the location parameter of the primary component. The asymptotic normality of the corresponding estimators is proved. The exact lower bound for the asymptotic variance is found and estimating functions for which this bound is attained are described. It is shown that the exact lower bound for the estimators under consideration is close to the corresponding bound of effectiveness of parametric estimators for the case where the distributions of both components are Gaussian.

1. INTRODUCTION

This paper is devoted to the estimation of a location parameter of a symmetric distribution for the case where the data contain an admixture that has a known distribution. Problems of this kind arise often when analyzing medical or biological data with the help of statistical methods. A particular example is described in [4, Section 6.1], where the expression of different genes from the so-called microarray data is estimated. The distribution of the data in the problem of [4, Section 6.1] is modeled by a mixture of two components, and one of them (called the primary component) has an unknown distribution, while the other one (called the admixture) has a known distribution.

The statistical analysis of observations from a mixture was known long ago starting with the papers by Newcomb [7] and Pearson [8]. A survey of modern statistical methods for finite mixtures can be found in [6, 10]. Various estimation methods for mixtures proposed till recently are parametric, since a nonparametric approach leads, generally speaking, to nonidentified problems. It was observed in [3, 5] that the assumption that the distributions of both components are symmetric allows one to obtain identifiable models in some cases. This approach was applied in [4] to the problem of estimation of characteristics of a symmetric distribution if the distribution of the admixture is known. In particular, moment estimators are proposed in [4] for the estimation of the location parameter.

In this paper, we consider some generalizations of the estimators proposed in [4] by using the method of generalized estimating equations (see, for example, [9, Section 5.4]). In Section 2, we describe an estimating equation for the generalized estimating equation estimators of the location parameter, obtain conditions for the asymptotic normality of

the estimators, and find their asymptotic variances. Section 3 deals with a particular case where the distribution of the admixture is Gaussian. In Section 4, we obtain an exact lower bound for the dispersion coefficients of generalized estimating equation estimators and the estimating functions are described for which this bound is attained (those functions and the corresponding estimators are called optimal). Some approximate generalized estimating equation estimators are considered in Section 5. We compare the dispersion coefficients of nonparametric generalized estimating equation estimators with those of maximum likelihood estimators in Section 6 for the case where the distributions of both components of the mixture are Gaussian. (Note that the maximum likelihood estimators are asymptotically effective in the problem under consideration.) We show that the asymptotic variance of the optimal generalized estimating equation estimators approaches the asymptotic variance of the maximum likelihood estimator as the difference between the expectations of the components tends to infinity. In other words, the generalized estimating equation estimators are "almost effective" for large samples from mixtures of Gaussian components. The proofs of the main results are placed in Section 8.

Note that the optimal estimating functions depend on unknown characteristics of the distribution of the primary component. Thus the "optimal estimators" cannot be used for the estimation explicitly. One can apply an adaptive approach similar to that proposed in [1] to construct an estimator whose asymptotic variance is equal to that of the optimal estimator. The paper [1] deals with observations with admixture in the model of a mixture with varying concentrations, and this model essentially differs from the model of a finite mixture studied in this paper. Nevertheless, our opinion is that the technique proposed in [1] can be useful for finite mixtures, too.

2. The setting of the problem and construction of the estimator

Assume that each observation may belong to one of two populations (components). The population containing a given observation is unknown. The distribution of the characteristics ξ of an observation depends on a population containing this observation. This distribution is known for one of the components (namely, for the admixture) and unknown for the other one (for the primary component). In what follows we assume that both distributions have densities and, moreover, the density of the primary component is symmetric about the location parameter.

In other words, the observations $\xi_1, \ldots, \xi_N$ are independent identically distributed random variables with the distribution density

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

where $p \in (0, 1]$ means the probability that an observation belongs to the primary component, a is the location parameter of the distribution of ξ for the primary component, $f(x)$ is the distribution density of the deviation between ξ and a for the primary component, and f_0 is the distribution density of ξ for the admixture. We assume that $f(-x) = f(x)$ for all $x \in \mathbb{R}$, that is, that f is an even function.

The mixing probability p (the concentration of the primary component), location parameter a, and the function f are unknown, while f_0 is known. The problem is to estimate the parameter a.

For the sake of convenience, we introduce two random variables η and η_0 whose densities are f and f_0, respectively, and a random variable $\delta \in \{0, 1\}$ that is independent of η and η_0 and such that $\mathsf{P}\{\delta = 1\} = p$. Then the distribution of each observation ξ_j coincides with that of $\delta(\eta + a) + (1 - \delta)\eta_0$.

Now we describe the idea on how to construct an estimating equation for a. Let g_i, $i = 1, 2$, be arbitrary odd functions (that is, $g(-x) = -g(x)$ for all $x \in \mathbb{R}$). If for all

$\alpha \in \mathbb{R}$, the expectations $\mathsf{E}\, g_i(\eta - \alpha)$ and $\mathsf{E}\, g_i(\eta_0 - \alpha)$ exist, then

$$\mathsf{E}\, g_i(\xi_j - \alpha) = p\,\mathsf{E}\, g_i(\eta + a - \alpha) + (1-p)\,\mathsf{E}\, g_i(\eta_0 - \alpha).$$

Put

$$(2) \qquad G_i(\alpha) = \mathsf{E}\, g_i(\eta_0 - \alpha) = \int_{-\infty}^{+\infty} g_i(x - \alpha) f_0(x)\, dx = \int_{-\infty}^{+\infty} g_i(x) f_0(x + \alpha)\, dx.$$

Since $f(x)$ is symmetric, $\mathsf{E}\, g_i(\eta) = \int_{-\infty}^{+\infty} g_i(x) f_0(x)\, dx = 0$, whence

$$(3) \qquad\qquad\qquad\qquad \mathsf{E}\, g_i(\xi_j - a) = (1-p) G_i(a).$$

Putting

$$(4) \qquad\qquad\qquad h(x, \alpha) = g_1(x - \alpha) G_2(\alpha) - g_2(x - \alpha) G_1(\alpha),$$

we obtain

$$(5) \qquad\qquad\qquad\qquad\qquad \mathsf{E}\, h(\xi_j, a) = 0.$$

Therefore, one can use h as an estimating function for the construction of an unbiased estimating equation. To show this, we put

$$\hat{h}(\alpha) = \frac{1}{N} \sum_{j=1}^{N} h(\xi_j, \alpha).$$

The statistics $\hat{a}_N$ is called a generalized estimating equation estimator with an estimating function h if

$$(6) \qquad\qquad\qquad\qquad\qquad \hat{h}(\hat{a}_N) = 0.$$

(The pair of the functions (g_1, g_2) used to construct h is called an estimating pair for the estimator $\hat{a}_N$.) Equality (5) implies that $\mathsf{E}\, \hat{h}(a) = 0$, that is, that the equation is unbiased.

If equation (6) has a unique root whatever the data $\xi_1, \ldots, \xi_N$, then the estimator $\hat{a}_N$ is determined uniquely by this equation. If there are several roots, then each of them can be chosen as $\hat{a}_N$ (the only restriction is that $\hat{a}_N$ is measurable as a function of a sample).

The moment estimator for a is proposed in [4]. This estimator can be derived from (4)–(6) by using $g_1(x) = x$ and $g_2(x) = x^3$. The conditions for the consistency of the moment estimator are also found in [4].

In what follows we consider generalized estimating equation estimators with different estimating pairs for which the estimators are consistent; that is, we assume that $\hat{a}_N \to a$ in probability as $N \to \infty$.

Here and in what follows, primed symbols denote derivatives of a corresponding function.

Theorem 2.1. *Let $\hat{a}_N$ be a generalized estimating equation estimator with an estimating pair (g_1, g_2) and let*

1. *$\hat{a}_N$ be a consistent estimator.*
2. *$\mathsf{E}(g_i(\xi_1 - a))^2 < \infty$, $i = 1, 2$.*
3. *For some $\varepsilon > 0$ and $\delta > 0$,*

$$\mathsf{E}\, \sup_{\alpha:\ |a - \alpha| < \varepsilon} (g_i'(\xi_1 - \alpha))^{1+\delta} < \infty, \qquad i = 1, 2.$$

4. *The functions g_i' and G_i' be continuous in $\mathbb{R}$.*
5. *$\mathsf{E}\, \frac{\partial}{\partial \alpha} h(\xi_1, \alpha)\big|_{\alpha = a} \neq 0$.*

Then the distribution of the random variables $\sqrt{N}(\hat{a}_N - a)$ weakly converges as $N \to \infty$ to the Gaussian distribution with zero mean and variance

$$\sigma^2_{(g_1, g_2)}(a) = \sigma^2_h(a) = \frac{\operatorname{Var} h(\xi_1, a)}{\left(\mathsf{E} \frac{\partial}{\partial \alpha} h(\xi_1, \alpha)\big|_{\alpha=a}\right)^2}.$$

The latter result is a trivial corollary of Theorem 5.14 in [9].

The variance $\sigma^2_h(a)$ of the limit Gaussian distribution of normalized estimators is called the asymptotic variance of the estimator $\hat{a}_N$. To express $\sigma^2_h(a)$ via the characteristics of our model we introduce the following notation.

For an arbitrary function $b \colon \mathbb{R} \to \mathbb{R}$, let $b^{s+}(x) = b(x) + b(-x)$ and

$$b^{s-}(x) = b(x) - b(-x).$$

The operations $s+$ and $s-$ correspond to the argument x; that is, for example,

$$b^{s+}(x + a) = b(x + a) + b(-x + a).$$

For functions $A \colon [0, \infty) \to \mathbb{R}$, put $\int A = \int_0^\infty A(x)\, dx$.

Since $\mathsf{E}\, h(\xi_1, a) = 0$ by (5) and since the function

$$(7) \qquad\qquad\qquad U(x) = G_2(a)g_1(x) - G_1(a)g_2(x)$$

is odd (which means that $(U(x))^2$ is an even function), we have

$$\operatorname{Var} h(\xi_1, a) = \int_{-\infty}^{\infty} \big(G_2(a)g_1(x-a) - G_1(a)g_2(x-a)\big)^2 \big(pf(x-a) + (1-p)f_0(x)\big)\, dx$$

$$= \int_0^{\infty} \big(G_2(a)g_1(x) - G_1(a)g_2(x)\big)^2 \big(2pf(x) + (1-p)f_0(x+a)^{s+}\big)\, dx.$$

Putting

$$(8) \qquad\qquad\qquad Q(x) = 2pf(x) + (1-p)f_0(x+a)^{s+},$$

we obtain

$$(9) \qquad\qquad\qquad \operatorname{Var} h(\xi, a) = \int U^2 Q.$$

Now we consider

$$\frac{\partial}{\partial \alpha} \mathsf{E}\, h(\xi_1, \alpha)\Big|_{\alpha=a}$$

$$= \frac{\partial}{\partial \alpha} \int_{-\infty}^{+\infty} \big(G_2(\alpha)g_1(x) - G_1(\alpha)g_2(x)\big)\big(pf(x-a+\alpha) + (1-p)f_0(x+\alpha)\big)\, dx\Big|_{\alpha=a}$$

$$= (1-p)\big(G_2'(a)G_1(a) - G_2(a)G_1'(a)\big) + \int_{-\infty}^{+\infty} U(x)(1-p)f_0'(x+a)\, dx$$

$$+ \int_0^{+\infty} U(x)2pf'(x)\, dx.$$

Here and in what follows we suppose that the change of the order of integration and differentiation is justified.

Taking into account that

$$G_i'(a) = \frac{\partial}{\partial \alpha} \int_{-\infty}^{+\infty} g_i(x)(1-p)f_0(x+\alpha)\, dx\Big|_{\alpha=a} = \int_{-\infty}^{+\infty} g_i(x)(1-p)f_0'(x+a)\, dx,$$

we get

$$\int_{-\infty}^{+\infty} U(x)(1-p)f_0'(x+a)\, dx = (1-p)\big(G_2(a)G_1'(a) - G_2'(a)G_1(a)\big),$$

whence

$$\frac{\partial}{\partial\alpha}\,\mathsf{E}\,h(\xi_1,\alpha)\bigg|_{\alpha=a} = \int UB,$$

where

(10)
$$B(x) = 2pf'(x).$$

Finally we obtain

(11)
$$\sigma_h^2(a) = \frac{\int U^2 Q}{(\int UB)^2}.$$

3. The moment estimator for the case of a Gaussian admixture

As an application of the estimators described in Section 2, we consider the model (1) where the distribution of the admixture is standard Gaussian (that is, $\eta_0 \sim N(0,1)$). As an estimating pair, we take $g_1(x) = x$ and $g_2(x) = x^3$. Then

$$G_1(\alpha) = -\alpha, \qquad G_2(\alpha) = -\alpha\left(3+\alpha^2\right),$$

and

$$h(x,\alpha) = \alpha\left(x^3 - 6x^2\alpha + x\left(6\alpha^2 - 3\right) + 6\alpha\right).$$

Canceling the root $\alpha = 0$, we arrive at the estimating equation

$$\hat{m}_3 - 6\hat{m}_2\alpha + \hat{m}_1\left(6\alpha^2 - 3\right) + 6\alpha = 0,$$

where $\hat{m}_i = N^{-1}\sum_{j=1}^{N}(\xi_j)^i$ are sampling moments of order i and where $m_i = \mathsf{E}(\xi_1)^i$. The latter equation has two roots

$$\hat{a}_{\pm,N} = \hat{a}_{\pm,N}(\hat{m}_1,\hat{m}_2,\hat{m}_3) = \frac{3\hat{m}_2 - 3 \pm \sqrt{9 + 24\hat{m}_1^2 - 18\hat{m}_2^2 - 8\hat{m}_1\hat{m}_3}}{4\hat{m}_1}.$$

Thus, to construct the estimator $\hat{a}_N$, one has to choose one of the three roots of the estimating equation, namely either 0, or $\hat{a}_-$, or $\hat{a}_+$. If $p > 0$, then the case of $a = 0$ holds if and only if $m_1 = 0$. This means that the zero root is worthwhile to take as an estimator of a only if $|\hat{m}_1| < C_N$, where C_N is a certain level such that $C_N \to 0$ as $N \to \infty$.

To make a choice between $\hat{a}_+$ and $\hat{a}_-$ one can use the following approach. Consider $g_1(x) = x$ and $g_2^*(x) = x^5$ as an estimating pair. Then

$$G_2^*(\alpha) = -\alpha\left(15 + 10\alpha^2 + \alpha^4\right)$$

and thus the estimating equation becomes of the following form:

$$\hat{h}^*(\alpha) = -15\hat{m}_1 + \hat{m}_5 + 15\alpha - 5\hat{m}_4\alpha - 10\hat{m}_1\alpha^2 + 10\hat{m}_3\alpha^2 + 10\alpha^3 - 10\hat{m}_2\alpha^3 + 4\hat{m}_1\alpha^4 = 0.$$

When α is changed for the "correct" root $\hat{a}_{\pm}$ on the right hand side of the latter equation, the value of the function $\hat{h}^*(\alpha)$ should be close to zero. This reasoning leads us to the following moment estimator $\hat{a}_N^{\mathrm{moment}}$ for a:

(12)
$$\hat{a}_N^{\mathrm{moment}} = \begin{cases} 0 & \text{if } |\hat{m}_1| < C_N, \\ \hat{a}_+ & \text{if } |\hat{h}^*(\hat{a}_+)| \le |\hat{h}^*(\hat{a}_-)| \text{ and } |\hat{m}_1| > C_N, \\ \hat{a}_- & \text{if } |\hat{h}^*(\hat{a}_-)| \le |\hat{h}^*(\hat{a}_+)| \text{ and } |\hat{m}_1| > C_N. \end{cases}$$

Put $e_i = \mathsf{E}(\eta)^i$ and $Z(a) = -9 + 9e_2^3 + 5e_4a^2 - a^4 - 3e_2^2(9 + 5a^2) + e_2(27 + a^4)$.

Theorem 3.1. *Let*

1. $\mathsf{E}\,|\eta|^5 < \infty$.
2. $C_N = KN^{-\beta}$, *where* $K > 0$ *and* $0 < \beta < 1/2$ *are some constants.*
3. $Z(a) \neq 0$.

Then $\hat{a}_N^{\mathrm{moment}} \to a$ *in probability as* $N \to \infty$.

Now let $\eta \sim N(0, s^2)$ (that is, the primary component has the Gaussian distribution with mean a and variance s^2). Considering equality (11), we obtain the asymptotic variance of the moment estimator:

$$
\sigma_{\mathrm{moment}}^2(a) = \frac{1}{p^2\,(3 + a^2 - 3s^2)^2}\Big(a^4\left(p\left(s^2 - 4\right) + 4\right) - 6a^2\left(p\left(3 - s^2 - s^4\right) - 3\right)
$$
$$
(13) \qquad\qquad\qquad\qquad + 3\left(2 + p\left(-2 + 3s^2 - 6s^4 + 5s^6\right)\right)\Big).
$$

For example, if $p = 0.5$, $a = 0.5$, and $s = 0.5$, we obtain $\sigma_{\mathrm{moment}}^2 = 3.97$.

Remark. 1. The case where the admixture has a Gaussian but not standard distribution can be reduced to the case considered above by subtracting the expectation from all members of the sample and then dividing the result by the square root of the variance of the admixture. It is clear that the estimator for the location parameter obtained from such a normalized sample should be transformed to the initial scale by inverting the normalization.

2. If $a = 0$, then the moment estimator defined by equality (12) has a "supereffective" point $\sigma_{\mathrm{moment}}^2(0) = 0$. As known, the property of supereffectiveness is a disadvantage rather an advantage of an estimator, since supereffective estimators have an unacceptably high bias in a neighborhood of a supereffective point. The question on how to construct a better version of the moment estimator with a correct behavior at zero requires a special investigation.

3. Condition 3 in the statement of the theorem seems artificial. As shown in [4], the model (1) is unidentifiable even for symmetric distributions without extra assumptions imposed on the distribution of the primary component of the model. Some other examples (also nontrivial) are considered in [4] where models are nonidentifiable but consistent estimators exist under some extra assumptions.

4. The lower bound for the asymptotic variance of generalized estimating equation estimators

Choosing different estimating pairs (g_1, g_2), one obtains different generalized estimating equation estimators $\hat{a}_N$ with different asymptotic variances. A natural question arises on the minimal value of the asymptotic variance of the generalized estimating equation estimator for a given probabilistic characteristics of the model. This is a question we are going to discuss further.

Equality (11) implies that the asymptotic variance $\sigma_{(g_1, g_2)}^2$ depends only on the function $U\colon [0, +\infty) \to \mathbb{R}$ defined by (7) if a is fixed, but it does not explicitly depend on the functions g_1 and g_2. For negative x, the function U is defined by $U(-x) = -U(x)$. Now we describe the functions that can be obtained by substituting different functions g_1 and g_2 in equality (7).

Put $D(x) = f_0^{s-}(x + a)$. Then $G_i(a) = \int g_i D$.

Lemma 4.1. *The function* $U\colon [0, +\infty) \to \mathbb{R}$ *is represented in the form* (7) *if and only if* $\int U D = 0$.

Proof. Let equality (7) hold. Then

$$\int UD = \int g_2 D \cdot \int g_1 D - \int g_1 D \cdot \int g_2 D = 0.$$

Now let $\int UD = 0$. Consider an arbitrary function Z such that $\beta = \int ZD \neq 0$. Put $g_1(x) = U(x) - (2\beta)^{-1}Z(x)$ and $g_2(x) = U(x) + (2\beta)^{-1}Z(x)$. Then

$G_2(a)g_1(x) - G_1(a)g_2(x)$

$$= \int \left(U + \frac{1}{2\beta}Z(x)\right)D \cdot \left(U - \frac{1}{2\beta}Z(x)\right) - \int \left(U - \frac{1}{2\beta}Z(x)\right)D \cdot \left(U + \frac{1}{2\beta}Z(x)\right)$$

$$= U(x).$$

The lemma is proved. □

Therefore the minimal asymptotic variance is a solution of the minimization problem for the ratio $\int U^2 Q / (\int UB)^2$ over all functions U such that $\int UD = 0$. Note that this ratio does not change if U is changed by βU for an arbitrary $\beta > 0$. This allows us to choose the normalizing term β such that $\int UB = 1$. Thus the minimization problem becomes of the form

$$\begin{cases} \int U^2 Q \to \min & \text{(with respect to } U), \\ \int UD = 0, \\ \int UB = 1. \end{cases}$$

Using the Lagrange multipliers method we prove that the minimum of $\int U^2 Q$ is attained at the function

$$U^* = \frac{\lambda_1 B + \lambda_2 D}{Q},$$

where

$$\lambda_1 = \frac{1}{\Delta}\int \frac{D^2}{Q}, \qquad \lambda_2 = -\frac{1}{\Delta}\int \frac{BD}{Q},$$

$$\Delta = \int \frac{B^2}{Q}\int \frac{D^2}{Q} - \left(\int \frac{BD}{Q}\right)^2.$$

Moreover the minimal value of the asymptotic variance is

$$\sigma_*^2(a) = \inf_{g_1,g_2} \sigma_{g_1,g_2}^2(a) = \lambda_1.$$

5. Approximate generalized estimating equation estimator

Finding the roots of the estimating equation (6) can be a complicated problem for a given estimating pair. Another problem is to choose the roots in order to construct a consistent estimator, and that problem is quite complicated even in the simplest cases (as we have seen in Section 3). Consider a method for constructing estimators allowing one to approximate the generalized estimating equation estimators described in Section 2 (these estimators have the same asymptotic behavior as the generalized estimating equation estimators). The approximations of the generalized estimating equation estimators are called the approximate generalized estimating equation estimators.

To construct the approximate generalized estimating equation estimators, we start with a "pilot" consistent estimator $\tilde{a}_N$ for a. We represent a solution of (6) in the form of $\hat{a}_N = \tilde{a}_N + \delta_N$ and expand $\hat{h}$ in Taylor's series in a neighborhood of $\tilde{a}_N$:

$$0 = \hat{h}(\hat{a}_N) \simeq \hat{h}(\tilde{a}_N) + \hat{h}'(\tilde{a}_N)\delta_N,$$

where $\hat{h}'(\tilde{a}_N) = \frac{\partial}{\partial \alpha}\hat{h}(\alpha)\big|_{\alpha=\tilde{a}_N}$. Thus $\delta_N \simeq -\hat{h}(\tilde{a}_N)/\hat{h}'(\tilde{a}_N)$.

The approximate generalized estimating equation estimator is given by
$$\check{a}_N = \tilde{a}_N - \hat{h}(\tilde{a}_N)/\hat{h}'(\tilde{a}_N).$$

Theorem 5.1. *Let assumptions 2–5 of Theorem 2.1 hold. Suppose that*
$$\sqrt{N}(\tilde{a}_N - a) = O_p(1)$$
as $N \to \infty$. Then the distribution of $\sqrt{N}(\check{a}_N - a)$ weakly converges to the Gaussian distribution with zero mean and variance $\sigma_h^2(a)$.

Remark. If the admixture is Gaussian, then one can take the moment estimator $\hat{a}_N^{\mathrm{moment}}$ defined by (12) as a pilot estimator of $\tilde{a}_N$. By Theorem 3.1, $\sqrt{N}\,(\tilde{a}_N^{\mathrm{moment}} - a) = O_p(1)$.

6. A COMPARISON WITH THE PARAMETRIC ESTIMATOR

In order to reach a conclusion concerning the effectiveness of the generalized estimating equation estimators and approximate generalized estimating equation estimators introduced above, we compare the exact lower bound for the asymptotic variance of maximum likelihood estimators obtained in Section 4 in the case where the distributions of both components are Gaussian and the parameter a tends to infinity. In fact, the larger the parameter a, the simpler is the problem of distinguishing observations belonging to the primary component from those belonging to the admixture whose expectation is equal to zero. In other words, we consider a "simple" estimation problem.

Let

(14) $$\eta_0 \sim N(0, 1), \qquad \eta \sim N\left(0, s^2\right).$$

Denote by φ_s the density of η_0 and by φ the density of η. Using the notation introduced in Section 4,
$$B(x) = 2px\varphi_s(x)/s^2, \qquad D(x) = \varphi(x + a) - \varphi(x - a),$$
and $Q(x) = 2p\varphi_s(x) + (1 - p)(\varphi(x + a) + \varphi(x - a))$. It is easy to see that
$$\int \frac{B^2}{Q} \xrightarrow[a \to +\infty]{} 2p \int_0^\infty \frac{\left(s^{-2}x\varphi_s(x)\right)^2}{\varphi_s(x)} \, dx = \frac{p}{s^2},$$
$$\int \frac{D^2}{Q} \underset{a \to +\infty}{\simeq} \int_0^\infty \frac{(\varphi(x + a))^2}{(1 - p)\varphi(x + a)} \, dx \xrightarrow[a \to +\infty]{} \frac{1}{1 - p},$$
$$\int \frac{BD}{Q} \xrightarrow[a \to +\infty]{} 0.$$

Thus $\Delta \to p/(s^2(1 - p))$ and the lower bound for the asymptotic variance of generalized estimating equation estimators is such that $\sigma_*^2(a) \to s^2/p$ as $a \to +\infty$.

Now we consider the estimation problem for the parameter a for the case where the distributions of the components are given by (14) and where s^2 and p are known. In this case, the maximum likelihood estimator is asymptotically effective and its asymptotic variance is equal to $\sigma_{MLE}^2(a) = 1/I$, where I denotes the Fisher information for the unknown parameter contained in a single observation:
$$I = \int_{-\infty}^{+\infty} \frac{\left(\frac{\partial}{\partial a}\left(p\varphi_s(x - a) + (1 - p)\varphi(x)\right)^2\right)^2}{p\varphi_s(x - a) + (1 - p)\varphi(x)} \, dx \xrightarrow[a \to +\infty]{} p/s^2.$$

Thus $\sigma_{MLE}^2(a) \to s^2/p$ as $a \to +\infty$; that is, the asymptotic variance of an asymptotically effective parametric estimator has the same limit as that of nonparametric generalized estimating equation estimators and of approximate generalized estimating equation estimators.

Example. Let $p = 0.5$, $a = 0.5$, and $s = 0.5$. Then $\sigma^2_{\text{moment}} = 3.97$ and $\sigma^2_* = 1.34141$. The asymptotic variance of the maximum likelihood estimator is $\sigma^2_{MLE} = 1.31579$ if p and s^2 are unknown. This means that the relative asymptotic effectiveness of generalized estimating equation estimators with an optimal estimating pair is $\sigma^2_{MLE}/\sigma^2_* = 0.980898$.

This is a comparatively "complicated" case, since a does not approach infinity. Nevertheless, the optimal generalized estimating equation estimator is not worse than the asymptotically effective maximum likelihood estimator and almost three times better than the moment estimator: $\sigma^2_*/\sigma^2_{\text{moment}} = 0.337887$.

7. Remarks

The above results confirm a potentially high quality of the generalized estimating equation estimators and approximate generalized estimating equation estimators. However these estimators cannot be used together with the optimal estimating pairs, since the optimal pair depends on unknown parameters of the model. To avoid this problem, one can apply an adaptive approach similar to that used when constructing approximate generalized estimating equation estimators. Namely, the approach is

- first, to estimate the parameters of the model by using nonoptimal estimators;
- and then, to use these nonoptimal estimators to construct a "quasioptimal" estimating pair.

An approximate generalized estimating equation estimator with such a quasioptimal estimating pair may also have nice asymptotic properties. Which adaptive estimator attains the lower bound of optimality for generalized estimating equation estimators and what are the conditions for that should be investigated in forthcoming publications.

8. Proofs of theorems

Proof of Theorem 3.1. The proof is trivial if $a = 0$ and thus $m_1 = 0$. Let $m_1 \neq 0$. Applying the law of large numbers we obtain

$$\hat{a}_\pm(\hat{m}_1, \hat{m}_2, \hat{m}_3) \to \hat{a}_\pm(m_1, m_2, m_3) \quad \text{a.s.} \quad \text{as } N \to \infty.$$

Expressing m_i in terms of a, p, and e_k we get

$$\hat{a}_\pm(m_1, m_2, m_3) = \frac{p\left(3a^2 + 3e_2 - 3\right) \pm \sqrt{\delta}}{4ap},$$

where $\delta = p(a^2 + 3e_2 - 3)$. In other words, if $\delta > 0$, then $\hat{a}_+$ is a "correct" root, while $\hat{a}_-$ is a "correct" root if $\delta < 0$.

Applying the law of large numbers again we prove that

$$\hat{h}^*(\hat{a}_\pm) \to Z(a)\left(\delta \pm \sqrt{\delta^2}\right) / \left(8a^4\right) = h^*_\pm.$$

This means that $h^*_\pm = 0$ for a correct root $\hat{a}_\pm$. If a root is incorrect, then $h^*_\pm = 0$ only if $Z(a) = 0$.

Let $\delta > 0$ and $Z(a) \neq 0$. Then $\hat{a}_+ \to a$, $\hat{h}^*(\hat{a}_+) \to 0$, and $\hat{h}^*(\hat{a}_-) \to h^*_- \neq 0$. Thus $\mathsf{P}\{|\hat{h}^*(\hat{a}_+)| < |\hat{h}^*(\hat{a}_-)|\} \to 1$ as $N \to \infty$. Therefore if $\delta > 0$, then

$$\mathsf{P}\{\hat{a}_N = \hat{a}_+\} \to 1 \quad \text{and} \quad \hat{a}_N \to a \quad \text{in probability.}$$

The case $\delta < 0$ is considered analogously. $\qquad\square$

Proof of Theorem 5.1. Note that $\hat{h}(\tilde{a}_N) = \hat{h}(a) + \hat{h}'(\zeta)(\tilde{a}_N - a)$, where ζ is an intermediate point between $\tilde{a}_N$ and a. Thus

$$\sqrt{N}(\check{a}_N - a) = \sqrt{N}(\check{a}_N - \tilde{a}_N + \tilde{a}_N - a) = -\sqrt{N}\frac{\hat{h}(\tilde{a}_N)}{\hat{h}'(\tilde{a}_N)} + \sqrt{N}\frac{\hat{h}(\tilde{a}_N) - \hat{h}(a)}{\hat{h}'(\zeta)} = I_1 + I_2,$$

where
$$I_1 = \sqrt{N}\left(\frac{\hat{h}(\tilde{a}_N)}{\hat{h}'(\zeta)} - \frac{\hat{h}(\tilde{a}_N)}{\hat{h}'(\tilde{a}_N)}\right), \qquad I_2 = \sqrt{N}\frac{\hat{h}(a)}{\hat{h}'(\zeta)}.$$

Assumptions 2–4 of Theorem 2.1 and Lemma 5.3 of [9] imply that

$$(15) \qquad \sup_{\alpha\,:\,|a-\alpha|<\varepsilon}\left|\hat{h}'(\alpha) - \mathsf{E}\,\frac{\partial}{\partial\alpha}h(\xi_1,\alpha)\right| \to 0 \quad \text{in probability} \quad \text{as } N\to\infty.$$

Since $\tilde{a}_N$ is consistent, $\zeta \to a$, whence $\hat{h}'(\zeta) \to \mathsf{E}\,\frac{\partial}{\partial\alpha}h(\xi_1,\alpha)\big|_{\alpha=a}$ by (15), since $\mathsf{E}\,\frac{\partial}{\partial\alpha}h(\xi_1,\alpha)$ is continuous with respect to α.

Now assumption 2 and $\mathsf{E}\,\hat{h}(a) = 0$ imply that $\sqrt{N}(\hat{a}-a) \Rightarrow N(0, \operatorname{Var} h(\xi_1,a))$. Hence $I_2 \Rightarrow N(0, \sigma_{\hat{h}}^2(a))$.

To complete the proof of the theorem, one needs to check that $I_1 \to 0$ in probability. We have $I_1 = J_1 \cdot J_2$, where $J_1 = \sqrt{N}\hat{h}(\tilde{a}_N)$, $J_2 = (\hat{h}'(\tilde{a}_N) - \hat{h}'(\zeta))/(\hat{h}'(\tilde{a}_N)\hat{h}'(\zeta))$. Considering (15) we obtain $J_2 \to 0$ by assumption 5.

Now we estimate $J_1 = \sqrt{N}\hat{h}(a) + \hat{h}'(\zeta)\sqrt{N}(\tilde{a}_N - a)$. The first term is bounded in probability, since its variance is bounded. The second term is bounded, since $\sqrt{N}(\tilde{a}_N - a)$ is bounded by assumptions of the theorem. Therefore $I_1 \to 0$ in probability.

The theorem is proved. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Bibliography

1. N. Lodatko and R. Maĭboroda, *An adaptive moment estimator of a parameter of a distribution constructed from observations with admixture*, Teor. Imovirnost. Matem. Statyst. **75** (2006), 61–70; English transl. in Theor. Probability and Math. Statist. **75** (2007), 71–82. MR2321182 (2008g:62101)
2. A. A. Borovkov, *Mathematical Statistics*, Nauka, Moscow, 1984; English transl., Gordon and Breach Science Publishers, Amsterdam, 1998. MR782295 (86i:62001); MR1712750 (2000f:62003)
3. L. Bordes, S. Mottelet, and P. Vandekerkhove, *Semiparametric estimation of a two-component mixture model*, Ann. Statist. **34** (2006), no. 3, 1204–1232. MR2278356 (2008e:62064)
4. 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)
5. D. R. Hunter, S. Wang, and T. R. Hettmansperger, *Inference for mixtures of symmetric distributions*, Ann. Statist. **35** (2007), 224–251. MR2332275 (2008g:62079)
6. G. J. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, 2000. MR1789474 (2002b:62025)
7. S. Newcomb, *A generalized theory of the combination of observations so as to obtain the best result*, Amer. J. Math. **8** (1886), 343–366. MR1505430
8. K. Pearson, *Contribution to the mathematical theory of evolution*, Phil. Trans. Roy. Soc. A **185** (1894), 71–110.
9. J. Shao, *Mathematical Statistics*, 2nd ed., Springer-Verlag, New York, 2003. MR2002723 (2004g:62002)
10. D. M. Titterington, A. F. Smith, and U. E. Makov, *Analysis of Finite Mixture Distributions*, Wiley, New York, 1985. MR0838090 (87j:62033)

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`