TECHNICAL REPORT 0106U005864-MS08

Generalized estimating equations for symmetric distributions observed with admixture

Rostyslav Maiboroda (Kyiv National University, Ukraine)
Olena Sugakova (Kyiv National University, Ukraine)
e-mail: mre@univ.kiev.ua

Abstract

A semiparametric two-component mixture model is considered, in which the distribution of one (primary) component is unknown and assumed symmetric. The distribution of the other component (admixture) is known. Generalized estimating equations are constructed for the estimation of the mixture proportion and the location parameter of the primary component. Asymptotic normality of the estimates is demonstrated and the lower bound for the asymptotic covariance matrix is obtained. An adaptive estimation technique is proposed to obtain the estimates with nearly optimal asymptotic variances.

Key words: adaptive estimating equation, asymptotic normality, finite mixture model, symmetric distribution

Research is supported in part by Swedish Institute grant SI-01424/2007.

Kyiv - 15.02.2009 (Corrected 5.04.2009)

1. Introduction

In this paper we consider semiparametric estimation by a sample from two-component mixture. We are interested in the distribution of one (primary) component of the mixture, while the distribution of the other component (the admixture) is known. Such problems naturally arise in biostatistics, e.g. when the influence of some mutation on a specific phenotypic feature of organisms is investigated. In such studies we usually deal with a sample containing both mutants and wild-type (non-mutant) organisms mistakenly recognized as mutants. Sometimes the distribution of wild-type organisms features is known or can be fitted by some other samples of large size. So the problem is to extract the distribution of the feature for mutants from the mixture of both types when the proportion of mutants in the observed sample is unknown. Another example of such statistical model with an application to a real data is considered in Bordes at al (2006a).

Statistical analysis of mixtures has a long history starting from the papers of Newcomb (1894) and Pearson (1894). An overview of modern statistical mixtures analysis can be found in McLachlan and Peel (2000), Titterington et al. (1985). Most of papers in this field are based on parametric models, since nonparametric statistical mixtures usually are unidentifiable. On the identifiability of parametric finite mixture models see Holzmann et al. (2004,2006). In Hall & Zhou (2003) conditions of identifiability are given for the nonparametric two-component mixture model with multivariate observations in the case when the observed variables are independent in both components of the mixture.

In Lodatko & Maiboroda (2007) observations with admixture are considered in the case when the mixing probabilities are known but different for different observations. Semiparametric generalized method of moments is applied here to the estimation of Euclidian parameters.

In the papers by Hunter et al. (2004,2007), Bordes et al. (2006), Maiboroda (2008) it was observed that the distribution symmetry assumption can lead to identifiable nonparametric models. This idea was applied in Bordes at al (2006a) to the estimation in the model of observations with admixture.

To fix notations, let the feature ξ be observed for n subjects and $\xi_1,\dots,\ \xi_n$ be the observed sample of i.i.d. random variables with pdf

$$\psi(x) = pf(x-a) + (1-p)f_0(x), \tag{1}$$

where $p \in (0,1)$ is the mixing probability, i.e. the probability to observe a subject from the primary component, a is the median of the primary component distribution, f is the pdf of $\xi - a$ for subjects from the primary component, f_0 is the pdf of the admixture. The pdf f_0 is known, f is unknown but assumed to be symmetric around zero, i.e. $f(x) = f(-x)$ for all $x \in \mathbb{R}$. The Euclidian parameters $a \in \mathbb{R}$ and $p \in (0,1)$ are unknown.

In Bordes et al. (2006a) moment estimates and estimates based on the symmetrization technique are considered for the Euclidian parameters a and p. Their $\sqrt{n}$-consistency is demonstrated. A kernel-type estimate for f is also considered and its consistency in L_1 norm is demonstrated. In Sugakova (2009) an estimate for the location parameter a is proposed, based on the generalized estimating equations (GEE) approach, and its asymptotic normality is demonstrated.

Technique of GEE is a useful tool in solving many statistical problems, see e.g. Shao (1998), Small and Wang (2003). Special attention to semiparametric GEE equations is paid in Tsiatis (2006).

In this paper we generalize the results of Sugakova (2009) constructing a vector-GEE estimate $\hat{\boldsymbol{\vartheta}} = (\hat{a}, \hat{p})^T$ for the vector parameter $\boldsymbol{\vartheta} = (a, p)^T$. Asymptotic normality of this estimate is demonstrated and the dispersion matrix (asymptotic covariance matrix) is derived. Then we derive the estimating function which generates the estimate with the least possible dispersion. Unfortunately, this function depends on the unknown parameters of the model. So we apply an adaptive technique of choosing the estimating function by the observed data. As the result an adapted estimate is obtained with nearly optimal dispersion.

The rest of the paper is organized as follows. In section 2 the GEE for the Euclidian parameter estimation is constructed. Asymptotic normality of the estimates is demonstrated in the section 3. Here the lower bound is also obtained for the dispersion matrix. Moment estimate is described in section 4 as an example of GEE estimates. In section 5 we compare our estimates with the best possible estimates in some parametric submodels of our model. Adaptive estimates are described in section 6. Section 7 presents some simulation studies and real data example to evaluate the performance of our method. Section 8 gives some concluding remarks. The proofs are collected in the Appendix.

2. GEE estimate

Here we present a GEE estimate for the parameter $\boldsymbol{\vartheta} = (a, p)^T$ by i.i.d. observations $\xi_1, \ldots, \xi_n$ with the pdf ψ defined by (1). In order to clarify notations we introduce three independent r.v.s: η with pdf f, η_0 with pdf f_0 and

$$\delta = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

Then

$$\xi_j \overset{\mathrm{d}}{=} \delta(\eta + a) + (1-\delta)\eta_0, \tag{2}$$

(here $\overset{\mathrm{d}}{=}$ means equality of distributions).

Let g_1 and g_2 be some odd functions (i.e., $g_i(-x) = -g_i(x)$ for all $x \in \mathbb{R}$) so that for any $\alpha \in \mathbb{R}$, $G_i(\alpha) := \int_{-\infty}^{\infty} g_i(x-\alpha) f_0(x)dx = \mathsf{E}\, g_i(\eta_0 - \alpha) < \infty$ for $i = 1, 2$. Then from (2) we get

$$\mathsf{E}\, g_i(\xi_j - a) = p\, \mathsf{E}\, g_i(\eta) + (1-p)G_i(a) = (1-p)G_i(a), \tag{3}$$

since $\mathsf{E}\, g_i(\eta) = 0$ due to oddness of g_i and symmetry of f.

Equality (3) suggests the following unbiased estimating equation for the Euclidian parameter $\boldsymbol{\vartheta}$:

$$\begin{cases} \hat{g}_1(\alpha) - (1-\pi)G_1(\alpha) = 0, \\ \hat{g}_2(\alpha) - (1-\pi)G_2(\alpha) = 0, \end{cases} \tag{4}$$

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

Any statistics $\hat{\boldsymbol{\vartheta}}_n = (\hat{a}_n, \hat{p}_n)^T$ which is an a.s. solution to (3) w.r.t. $\mathbf{t} = (\alpha, \pi)^T$ we will call a GEE-estimate to $\boldsymbol{\vartheta}$ with the estimating pair (g_1, g_2). (Existence of such statistics for large n can be shown under mild conditions in the usual way, cf. Heyde (1997), Pfanzagl (1969)).

E.g. letting $g_1(x) = x$, $g_2(x) = x^3$ yields the moment estimate considered in Bordes et al. (2006a). Note that (3) usually has many roots, hence the GEE-estimate is not defined by (3) unambiguously. To obtain a consistent estimate in this case one needs some additional procedure of true root selection. An example of such procedure is considered below in section 4.

3. Asymptotics of GEE estimates

Now we assume that a GEE-estimate is consistent and consider its asymptotic distribution.

Denote $B(x) := 2pf'(x)$, $D(x) := f_0(x+a) - f_0(-x+a)$,
$Q(x) := 2pf(x) + (1-p)(f_0(x+a) + f_0(-x+a))$.

Here and below prime means differentiation, $\int h$ means $\int_0^{+\infty} h(x)dx$ for any integrable function h. Denote $\mathbf{S} := \mathbf{S}(g_1, g_2) = (s_{ik})_{i,k=1,2}$, where

$$s_{ik} = \int g_i g_k Q - (1-p)^2 \int g_i D \int g_k D, \tag{5}$$

$$\mathbf{V} = \mathbf{V}(g_1, g_2) := \begin{pmatrix} \int g_1 B & \int g_1 D \\ \int g_2 B & \int g_2 D \end{pmatrix} \tag{6}$$

Theorem 1. *Let $\hat{\boldsymbol{\vartheta}}_n$ be a GEE estimate for $\boldsymbol{\vartheta} = (a,p)^T$ with an estimating pair (g_1, g_2). Assume that*

1.$\hat{\boldsymbol{\vartheta}}_n$ is a consistent estimate of $\boldsymbol{\vartheta}$.

2. $\mathsf{E}(g_i(\xi_1 - a))^2 < \infty$, for $i = 1, 2$.

3. For some $\varepsilon > 0$, $\delta > 0$, $\mathsf{E}\sup_{\alpha:|a-\alpha|<\varepsilon}(g_i'(\xi_1 - \alpha))^{1+\delta} < \infty$, $i = 1, 2$.

4. The functions f', g_i' and G_i' $(i = 1, 2)$ are continuous on $\mathbb{R}$.

5. $\det \mathbf{V}(g_1, g_2) \neq 0$.

Then $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}) \Rightarrow N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(g_1, g_2) := \mathbf{V}^{-1}\mathbf{S}(\mathbf{V}^{-1})^T$.

(Here $\Rightarrow$ denotes weak convergence).

The proof can be obtained by application of standard theorems on asymptotic normality for GEE-estimates (see Appendix).

Our aim now is to obtain the lower bound for the dispersion matrices $\boldsymbol{\Sigma}(g_1, g_2)$ for all possible estimating pairs (g_1, g_2). (The covariances are compared in Lowener order, i.e for any square matrices A and B, $A \le B$ means that $B - A$ is a nonnegatively defined matrix).

Denote

$$\Delta := \int \frac{B^2}{Q} \int \frac{D^2}{Q} - \left(\int \frac{BD}{Q} \right)^2, \tag{7}$$

$$\lambda_{11} := \frac{1}{\Delta} \int \frac{D^2}{Q}, \quad \lambda_{12} = \lambda_{21} := -\frac{1}{\Delta} \int \frac{BD}{Q}, \quad \lambda_{22} := \frac{1}{\Delta} \int \frac{B^2}{Q}, \tag{8}$$

$$g_i^* := \frac{\lambda_{i1} B + \lambda_{i2} D}{Q}. \tag{9}$$

Then, by (5) and (6),

$$\boldsymbol{\Sigma}^* := \boldsymbol{\Sigma}(g_1^*, g_2^*) = \frac{1}{\Delta} \begin{pmatrix} \int \frac{D^2}{Q} & \int \frac{BD}{Q} \\ \int \frac{BD}{Q} & \int \frac{B^2}{Q} - (1-p)^2 \end{pmatrix}. \tag{10}$$

Theorem 2. *Assume that*

1. f *is continuously differentiable on* $\mathbb{R}$.

2. $\int \frac{B^2}{Q} < \infty$.

3. $\Delta \neq 0$.

Then for any estimating pair (g_1, g_2), $\mathbf{\Sigma}^* \le \mathbf{\Sigma}(g_1, g_2)$.

Notes 1. If for some x_0, $Q(x_0) = 0$ then $D(x_0) = 0$ and $B(x_0) = 0$. Hence at such points x where $Q(x)$ vanishes, $g_i(x)$ can be defined arbitrary satisfying the conditions of oddness and continuity only.

2. It is obvious that

$$\int \frac{D^2}{Q} = \int_0^{+\infty} \frac{(f_0(x+a) - f_0(-x+a))^2}{2pf(x) + (1-p)(f_0(x+a) + f_0(-x+a))} dx$$
$$\le \int_0^{+\infty} \frac{2(f_0^2(x+a) + f_0^2(-x+a))}{(1-p)(f_0(x+a) + f_0(-x+a))} dx$$

is always finite since f_0 is a pdf. So $\int \frac{BD}{Q} < \infty$ if $\int \frac{B^2}{D} < \infty$. The assumption $\int \frac{B^2}{D} < \infty$ holds e.g. if $\int (f')^2/f < \infty$, since

$$\int \frac{B^2}{Q} = \int_0^{+\infty} \frac{(2pf'(x))^2}{2pf(x) + (1-p)(f_0(x+a) + f_0(-x+a))} dx \le 4p \int \frac{(f'(x))^2}{f(x)} dx.$$

Corollary 1. *If the assumptions of the theorems 1 and 2 hold then*

1. $\sqrt{n}(\hat{a}_n - a) \Rightarrow N(0, \sigma_a^2(g_1, g_2))$ *for some* $\sigma_a^2(g_1, g_2) \ge \sigma_{a*}^2 = \frac{1}{\Delta} \int \frac{D^2}{Q}$,

2. $\sqrt{n}(\hat{p}_n - p) \Rightarrow N(0, \sigma_p^2(g_1, g_2))$ *for some* $\sigma_p^2(g_1, g_2) \ge \sigma_{a*}^2 = \frac{1}{\Delta}\left(\int \frac{B^2}{Q} - (1-p)^2\right)$.

4. Moment estimate

In this section we discuss the moment estimate for the Euclidian parameter $\boldsymbol{\vartheta}$ under the assumption that the admixture is standard Gaussian, i.e. $f \sim N(0,1)$. (Note that if $f \sim N(a, \sigma^2)$ then a simple rescaling $\xi_j' = (\xi_j - a)/\sigma$ reduces the data to the standard case).

So, let the estimating pair be $g_1(x) = x$, $g_2(x) = x^3$. Then $G_1(\alpha) = -\alpha$, $G_2(\alpha) = -\alpha(3 + \alpha^2)$ and (3) yields the following equations:

$$\hat{m}_1 + (1 - \pi)\alpha = 0 \tag{11}$$

and

$$\alpha H_1(\alpha, \hat{m}_1, \hat{m}_2, \hat{m}_3) = 0, \tag{12}$$

where $\hat{m}_k = \frac{1}{n}\sum_{j=1}^n (\xi_j)^k$ is the k-th empirical moment of the observed sample, $m_k = \mathsf{E}(\xi_1)^k$ is the corresponding theoretical moment and

$$\hat{H}_1(\alpha) = H_1(\alpha, \hat{m}_1, \hat{m}_2, \hat{m}_3) = 3\hat{m}_1 + \hat{m}_3 + (3 - 3\hat{m}_2)\alpha + 2\hat{m}_1\alpha^2. \tag{13}$$

The equation (11) has three roots:

$$\hat{\alpha}_{0,n} = 0,\ \hat{a}_{\pm,n}(\hat{m}_1, \hat{m}_2, \hat{m}_3) = \frac{1}{4\hat{m}_1}\left(3\hat{m}_2 - 3 \pm \sqrt{9 + 24\hat{m}_1^2 - 18\hat{m}_2 + 9\hat{m}_2^2 - 8\hat{m}_1\hat{m}_3}\right).$$

To construct a consistent estimate for a we need to select one of these three roots. Some ideas of such selection procedure based on symmetrization are considered in Bordes et al. (2006a). Here we consider another procedure based on the GEE approach.

Note that $a = 0$ iff $m_1 = 0$. So we select the root 0 as an estimate of a iff $\hat{m}_1$ is near zero, i.e. $|\hat{m}_1| < C_n$, where C_n is some threshold, $C_n \to 0$ as $n \to \infty$.

To choose among $\hat{a}_{+,n}$ and $\hat{a}_{-,n}$ we take a "secondary" estimating pair $g_1(x) = x$, $g_2^*(x) = x^5$. Then $G_2^*(\alpha) = -\alpha(15+10\alpha^2+\alpha^4)$ and from (3) we obtain the following estimating equation for a: $\alpha H_2(\alpha) = 0$, where

$$\hat{H}_2(\alpha) = H_2(\alpha, \hat{m}_1, \hat{m}_2, \hat{m}_3, \hat{m}_4, \hat{m}_5)$$

$$:= -15\hat{m}_1 + \hat{m}_5 + 15\alpha - 5\hat{m}_4\alpha - 10\hat{m}_1\alpha^2 + 10\hat{m}_3\alpha^2 + 10\alpha^3 - 10\hat{m}_2\alpha^3 + 4\hat{m}_1\alpha^4.$$

We will call $\hat{H}_1$ the primary and $\hat{H}_2$ the secondary empirical contrasts. It is obvious that the empirical contrasts $\hat{H}_i(\alpha)$ converge to theoretical contrasts $\bar{H}_i(\alpha) = H_i(\alpha, m_1, m_2, m_3, m_4, m_5)$ as $n \to \infty$ and the zeros of $\hat{H}_i$ converge to corresponding zeros of $\bar{H}_i$. Assume that $a \neq 0$. Since our estimating equations are unbiased, one root of $\bar{H}_i(\alpha) = 0$ is the "true root" $\alpha = a$. The others are the "false roots". We must not use the empirical counterparts of the "false roots" as the estimate. The idea is to take as an estimate the zero of $\hat{H}_1$ which is "nearly zero" of $\hat{H}_2$.

More rigorously, the proposed estimate is

$$\hat{a}_n^{ME} = \begin{cases} 0 & \text{if } |\hat{m}_1| < C_n, \\ \hat{a}_{+,n} & \text{if } |\hat{H}^2(\hat{a}_{+,n})| \leq |\hat{H}^2(\hat{a}_{-,n})| \text{ and } |\hat{m}_1| > C_n, \\ \hat{a}_{-,n} & \text{if } |\hat{H}^2(\hat{a}_{-,n})| \leq |\hat{H}^2(\hat{a}_{+,n})| \text{ and } |\hat{m}_1| > C_n. \end{cases} \tag{14}$$

Form (11) we get the estimate for p: $\hat{p}_n^{ME} := 1 - \hat{a}_n^{ME}/\hat{m}_1$.

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

Theorem 3. *Assume that*

1. $\mathsf{E}\,|\eta|^5 < \infty$.
2. $C_n = Kn^{-\beta}$ *for some constants* $K > 0$, $0 < \beta < 1/2$.
3. $Z(a) \neq 0$.
Then $\hat{a}_n^{ME} \to a$ *in probability as* $n \to \infty$.
Assume further $a \neq 0$ *then* $\hat{p}_n \to p$ *in probability.*

The proof is based on the fact that false zeros of theoretical contrasts $\bar{H}_1$ and $\bar{H}_2$ can't coincide if $Z(a) \neq 0$. See Sugakova (2009) for details.

There are also some other root selection algorithms, see Bordes et al. (2006) which need different assumptions to generate consistent estimates. It seems promising to apply general measures of skewness for this purpose. See Critchley & Jones (2008), Boshnakov (2007), Avérous et al. (1996) for the general theory of such measures.

Under the conditions of theorem 3 (including $a \neq 0$) the moment estimate is asymptotically normal by theorem 1 and its dispersion matrix is $\mathbf{\Sigma}(g_1, g_2)$. E.g. if f is $N(0, s^2)$ then

$$\begin{aligned} \sigma_{ME}^2(a) = p^{-2}(3 + a^2 - 3s^2)^{-2}[&a^4(p(s^2 - 4) + 4) - 6a^2(p(3 - s^2 - s^4) - 3) \\ &+ 3(2 + p(-2 + 3s^2 - 6s^4 + 5s^6)] \end{aligned} \tag{15}$$

5. Comparison to parametric estimates

The lower bound of the theorem 2 is valid for GEE estimates given by (3). Of course, there can be estimates of other form with possibly better asymptotic behavior. Let us compare the bounds of corollary 1 with the dispersions of the best possible estimates for a and p in some parametric submodels of (1).

At the first submodel we fix f, p and f_0 as known, so a is the only one unknown parameter. Then if f and f_0 are smooth enough, the maximum likelihood estimate for a is asymptotically efficient, i.e. its dispersion $\sigma^2_{MLE}(a)$ is the least possible dispersion for all regular estimates and $\sigma^2_{MLE}(a) = \frac{1}{I(a)}$, where

$$I(a) = \int_{-\infty}^{+\infty} \frac{\left(\frac{\partial}{\partial a}(pf(x-a) + (1-p)f_0(x))\right)^2}{\psi(x)} dx$$

is the Fisher information on the parameter a at one observation ξ_j.

Consider the case when $a \to \infty$. Note that if $I_\infty := \int_{-\infty}^{+\infty} \frac{(f'(x))^2}{f(x)} dx < \infty$ and $f_0(x) \to 0$ as $x \to \pm\infty$, then

$$I(a) = \int_{-\infty}^{+\infty} \frac{(pf'(x-a))^2}{pf(x-a) + (1-p)f_0(x)} dx \to pI_\infty. \tag{16}$$

To compare this limit with the result of corollary 1, note that

$$\int \frac{D^2}{Q} \underset{a\to+\infty}{\approx} \int_{-\infty}^{a} \frac{(f_0(t))^2}{2pf(a-t) + (1-p)(f_0(2a-t) + f_0(t))} dt \to \frac{1}{1-p},$$

$$\int \frac{B^2}{Q} \underset{a\to+\infty}{\longrightarrow} 2pI_\infty, \quad \int \frac{DB}{Q} \underset{a\to+\infty}{\longrightarrow} 0.$$

Thus $\Delta \to pI_\infty/(1-p)$ and $\sigma^2_{a*} \to (pI_\infty)^{-1}$.

So, for large a, the best possible GEE estimate for a is nearly as efficient as the efficient estimate in the parametric submodel. But it is not so for the moment estimate: e.g., in the case of gaussian primary component, (15) implies

$$\sigma^2_{ME}(a) \underset{a\to+\infty}{\longrightarrow} \frac{s^2}{p} + \frac{4(1-p)}{p},$$

while $\sigma^2_{MLE} \to s^2/p$.

In the second submodel we assume that f, f_0 and a are fixed known and p is to be estimated. Then, under the assumptions stated above, the least possible dispersion for the regular estimates of p is

$$\sigma^2_{MLE}(p) = \left(\int \frac{(f(x-a) - f_0(x))^2}{pf(x-a) + (1-p)f_0(x)}\right)^{-1} \underset{a\to+\infty}{\longrightarrow} p(1-p)$$

and

$$\sigma^2_{p*} \to \frac{1-p}{pI_\infty} pI_\infty - (1-p)^2 = p(1-p).$$

The case of large a ($a \to \infty$) can be considered as "easily separated admixture", when the primary component can be distinguished from the admixture with the unaided eye (see fig. 1(a)).

To assess the efficiency of the GEE estimates in a more difficult case with "inseparable" admixture we consider a submodel $f \sim N(0, s^2)$, $f_0 \sim N(0,1)$. The unknown parameters are a, p and s^2. In the case $a = p = s = 0.5$ (see fig 1(b)) we get the parametric bounds for dispersions $\sigma^2_{MLE}(a) = 1.3764$, $\sigma^2_{MLE}(p) = 2.1404$. The lower bounds for GEE estimates dispersions are $\sigma^2_{a*} = 1.52587$, $\sigma^2_{p*} = 3.5383$ and for the moment estimate we get $\sigma^2_{ME}(a) = 3.97$, $\sigma^2_{ME}(p) = 6.52$.

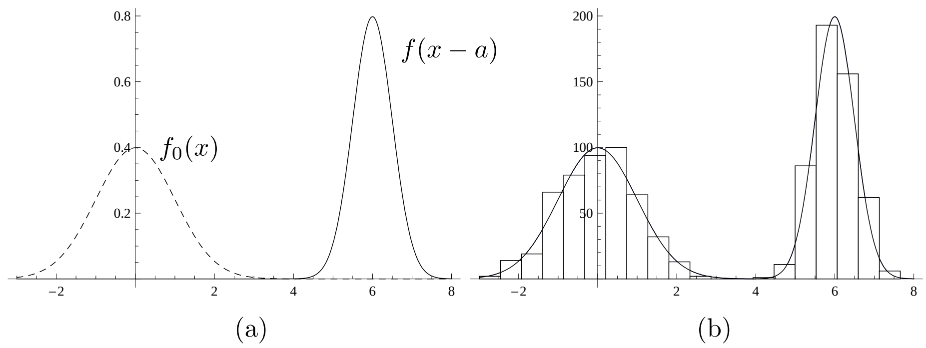


Figure 1. Easily separable admixture: (a)components densities, (b) histogram and density of data

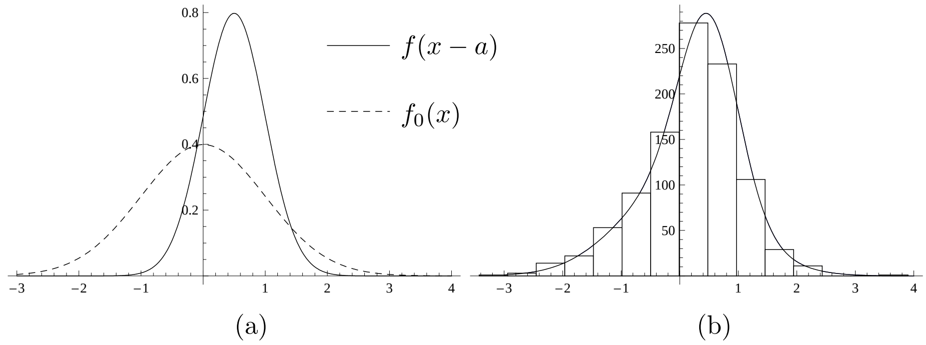


Figure 2. Inseparable admixture: (a)components densities, (b) histogram and density of data

6. Adapted GEE estimates

As we have seen, GEE estimates are potentially quite efficient. But the GEE with the best possible dispersion considered in theorem 2 is infeasible, since its estimating pair (g_1^*, g_2^*) depends on the unknown parameters of the model. To obtain a feasible estimator with dispersion matrix near the best possible for GEE estimates we adopt a two-stage adaptive approach. (Analogical approach to the adaptation of GEE e.g. for the restricted moment model see in Tsiatis (2006), section 4.6).

Let us fix a class of "basis" even functions $u_1,\dots, u_M$ and consider the estimating functions $g_i(x)$ of the form

$$g_i(x) = g_{\boldsymbol{\beta}_i}(x) := \sum_{m=1}^{M} \beta_{mi} u_m(x), \tag{17}$$

where $\boldsymbol{\beta}_i = (\beta_{1i}, \dots, \beta_{Mi})^T$ are some coefficients vectors. It is easy to find the best coefficients $\boldsymbol{\beta}_i^*$ for which the GEE estimate with the estimating pair $(g^*_{\boldsymbol{\beta}_1}, g^*_{\boldsymbol{\beta}_2})$ has the least dispersion matrix through all estimates with estimating pairs given by (17). These coefficients depend on unknown a, p and f. So at the first stage we estimate $\boldsymbol{\beta}_i^*$ by data and use the obtained estimates $\tilde{\boldsymbol{\beta}}_i$ to construct an adapted estimating pair $(g_{\tilde{\boldsymbol{\beta}}_1}, g_{\tilde{\boldsymbol{\beta}}_2})$. At the second stage this pair is used to obtain a GEE estimator.

Of course, the theory of GEE developed above can't be applied directly to GEE with the estimating pair $(g_{\tilde{\boldsymbol{\beta}}_1}, g_{\tilde{\boldsymbol{\beta}}_2})$, since this pair depends on the sample through $\tilde{\boldsymbol{\beta}}_i$. So we need to analyze asymptotic behavior of such adapted GEE estimates separately. Another difficulty connected with these estimates is that the solution of GEE (3) and true root selection become nontrivial problems when the estimating pair is random.

Therefore we replace (4) by its approximation based on the Taylor expansion of the estimating function in a neighborhood of some "pilot estimate". (Such approximations for maximum likelihood estimators are described e.g. in Shao (1998), section 4.5.3). Assume that there is some $\sqrt{n}$-consistent estimate $\tilde{\boldsymbol{\vartheta}}_n = (\tilde{a}_n, \tilde{p}_n)^T$ for $(a, p)^T$. (It can be e.g. the moment estimate from section 4). We want to improve its convergence rate using the estimating equation

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

where $\mathbf{t} = (\alpha, \pi)^T$, $\mathbf{h}(x, \mathbf{t}) = (h_1(x, \mathbf{t}), h_2(x, \mathbf{t}))^T$, $\hat{\mathbf{h}}(\mathbf{t}) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{h}(\xi_j, \mathbf{t})$.

If $\mathbf{h}$ is smooth enough then (18) is asymptotically equivalent to

$$\hat{\mathbf{h}}(\tilde{\boldsymbol{\vartheta}}_n) + \frac{\partial}{\partial \mathbf{t}} \hat{\mathbf{h}}(\mathbf{t}) \Big|_{\mathbf{t}=\tilde{\boldsymbol{\vartheta}}_n} (\mathbf{t} - \tilde{\boldsymbol{\vartheta}}) = 0.$$

This approximation suggests the following approximated-GEE (aGEE) estimate for $\boldsymbol{\vartheta}$: $\tilde{\boldsymbol{\vartheta}}_n - \left[\frac{\partial}{\partial \mathbf{t}} \hat{\mathbf{h}}(\mathbf{t}) \Big|_{\mathbf{t}=\tilde{\boldsymbol{\vartheta}}_n} \right]^{-1} \hat{\mathbf{h}}(\tilde{\boldsymbol{\vartheta}}_n)$.

In our case the estimating function $\mathbf{h}$ is taken to satisfy (approximately) the normalizing condition

$$\frac{\partial}{\partial \mathbf{t}} \hat{\mathbf{h}}(t) \Big|_{\mathbf{t}=\tilde{\boldsymbol{\vartheta}}_n} \approx \mathbf{E},$$

where $\mathbf{E}$ is the unit matrix. So the resulting aGEE estimate is

$$\check{\boldsymbol{\vartheta}}_n = \tilde{\boldsymbol{\vartheta}}_n - \hat{\mathbf{h}}(\tilde{\boldsymbol{\vartheta}}_n). \tag{19}$$

Now let us see how this estimate works in our case.

6.1. **Best coefficients $\boldsymbol{\beta}_i$.** Here and below $u_1,\dots,u_M$ are some fixed odd continuously differentiable functions on $\mathbb{R}$. Denote

$$U_i(\alpha) := \int_{-\infty}^{+\infty} u_i(x-\alpha) f_0(x)dx,$$

$$q_{ik} := \int u_i u_k Q = \mathsf{E}\, u_i(\xi_1 - a) u_k(\xi_1 - a), \quad \mathbf{Q} := (q_{ik})_{i,k=1}^M, \tag{20}$$

$$d_i := \int u_i D = \mathsf{E}\, u_i(\eta_0 - a) = U_i(a), \quad \mathbf{d} := (d_1,\dots,d_M)^T, \tag{21}$$

(22)

$$b_i := \int u_i B = -p \int_{-\infty}^{+\infty} u_i'(x) f(x) dx = -\mathsf{E}\, u_i'(\xi_1 - a) + (1-p) U_i'(a), \;\; \mathbf{b} := (b_1,\dots,b_M)^T.$$

Then $\int g_{\boldsymbol{\beta}_i} g_{\boldsymbol{\beta}_k} Q = \boldsymbol{\beta}_i^T \mathbf{Q} \boldsymbol{\beta}_k$, $\int g_{\boldsymbol{\beta}_i} B = \mathbf{b}^T \boldsymbol{\beta}_i$, $\int g_{\boldsymbol{\beta}_i} D = \mathbf{d}^T \boldsymbol{\beta}_i$.

Denote

$$\Delta' := \mathbf{d}^T \mathbf{Q}^{-1} \mathbf{d} \cdot \mathbf{b}^T \mathbf{Q}^{-1} \mathbf{b} - (\mathbf{d}^T \mathbf{Q}^{-1} \mathbf{b})^2 \tag{23}$$

$$\lambda_{11}' := \frac{1}{\Delta'} \mathbf{d}^T \mathbf{Q}^{-1} \mathbf{d}, \quad \lambda_{12}' = \lambda_{21}' := -\frac{1}{\Delta'} \mathbf{d}^T \mathbf{Q}^{-1} \mathbf{b}, \quad \lambda_{22}' := \frac{1}{\Delta'} \mathbf{b}^T \mathbf{Q}^{-1} \mathbf{b}, \tag{24}$$

$$\boldsymbol{\beta}_i^* := \mathbf{Q}^{-1} (\lambda_{i1}' \mathbf{b} + \lambda_{i2}' \mathbf{d}). \tag{25}$$

The following theorem is an analogue of theorem 2 for the estimating pairs from functions of the form (17).

Theorem 4. *Assume that*

1. f is continuously differentiable,

2. $|q_{ik}| < \infty$, $|b_i| < \infty$ for all $i,k = 1,\dots,M$,

3. $\det \mathbf{Q} \neq 0$.

Then for any $\boldsymbol{\beta}_i$

$$\boldsymbol{\Sigma}(g_{\boldsymbol{\beta}_1}, g_{\boldsymbol{\beta}_2}) \geq \boldsymbol{\Sigma}^U = \boldsymbol{\Sigma}(g_{\boldsymbol{\beta}_1^*}, g_{\boldsymbol{\beta}_2^*}).$$

If $(\hat{a}_n, \hat{p}_n)^T$ is a consistent GEE with the estimating pair $(g_{\boldsymbol{\beta}_1^*}, g_{\boldsymbol{\beta}_2^*})$ then

$$\sqrt{n}(\hat{a}_n - a) \Rightarrow N(0, \sigma_{a,U}^2), \;\; \textit{where } \sigma_{a,U}^2 = \lambda_{11}',$$

$$\sqrt{n}(\hat{p}_n - p) \Rightarrow N(0, \sigma_{p,U}^2), \;\; \textit{where } \sigma_{p,U}^2 = \lambda_{22}' - (1-p)^2.$$

6.2. **Estimates $\tilde{\boldsymbol{\beta}}_i$ and aGEE estimate.** If $\tilde{a}_n$ and $\tilde{p}_n$ are some pilot estimates to a and p, then the following estimates to $\mathbf{Q}$, $\mathbf{b}$ and $\mathbf{d}$ can be considered, suggested by (20)-(22):

$$\tilde{\mathbf{Q}} = (\tilde{q}_{ik})_{i,k=1}^M, \quad \tilde{q}_{ik} = \frac{1}{n} \sum_{j=1}^n u_i(\xi_j - \tilde{a}_n) u_k(\xi_j - \tilde{a}_n),$$

$$\tilde{\mathbf{d}} = (\tilde{d}_i)_{i=1}^M, \quad \tilde{d}_i = U_i(\tilde{a}_n),$$

$$\tilde{\mathbf{b}} = (\tilde{b}_i)_{i=1}^M, \quad \tilde{b}_i = -\frac{1}{n} \sum_{j=1}^n u_i'(\xi_j - \tilde{a}_n) + (1 - \tilde{p}_n) U_i'(\tilde{a}_n).$$

The estimates $\tilde{\boldsymbol{\beta}}_i$ to $\boldsymbol{\beta}_i^*$ are obtained replacing $\mathbf{Q}$, $\mathbf{d}$ and $\mathbf{b}$ in (23)-(25) by their estimates $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{b}}$.

The adapted estimate $\check{\boldsymbol{\vartheta}}_n$ is given by (19), where

$$\hat{h}_i(\tilde{a}_n, \tilde{p}_n) = \hat{g}_{\tilde{\boldsymbol{\beta}}_i}(\tilde{a}_n) - (1 - \tilde{p}_n) G_{\tilde{\boldsymbol{\beta}}_i}(\tilde{a}_n)$$

$$= \sum_{m=1}^{M} \tilde{\beta}_{mi} \hat{u}_i(\tilde{a}_n) - (1 - \tilde{p}_n) \sum_{m=1}^{M} \tilde{\beta}_{mi} U_i(\tilde{a}_n).$$

It is obvious from (23)-(25) that

$$\sum_{m=1}^{M} \tilde{\beta}_{m1} U_i(\tilde{a}_n) = \tilde{\mathbf{d}}^T \tilde{\boldsymbol{\beta}}_1 = 0 \text{ and } \sum_{m=1}^{M} \tilde{\beta}_{m2} U_i(\tilde{a}_n) = \tilde{\mathbf{d}}^T \tilde{\boldsymbol{\beta}}_2 = 1.$$

Hence (19) yields

$$\check{a}_n = \tilde{a}_n - \sum_{m=1}^{M} \tilde{\beta}_{m1} \hat{u}_i(\tilde{a}_n), \tag{26}$$

$$\check{p}_n = 1 - \sum_{m=1}^{M} \tilde{\beta}_{m2} \hat{u}_i(\tilde{a}_n), \tag{27}$$

Theorem 5. *Assume that*

1. The pilot estimates $\tilde{a}_n$, $\tilde{p}_n$ are $\sqrt{n}$-consistent.

For all $m = 1, \ldots, M$

2. $\mathsf{E}((u_m(\xi_1 - a))^2 < \infty$.

3. u_m are twice continuously differentiable on $\mathbb{R}$.

4. For some $\varepsilon > 0$,

$$e_{1,m} = \mathsf{E} \sup_{\alpha: |a-\alpha| < \varepsilon} (u'_m(\xi_1 - \alpha))^2 < \infty,$$

$$e_{2,m} = \mathsf{E} \sup_{\alpha: |a-\alpha| < \varepsilon} |u''_m(\xi_1 - \alpha)| < \infty.$$

5. $\det \mathbf{Q} \neq 0$.

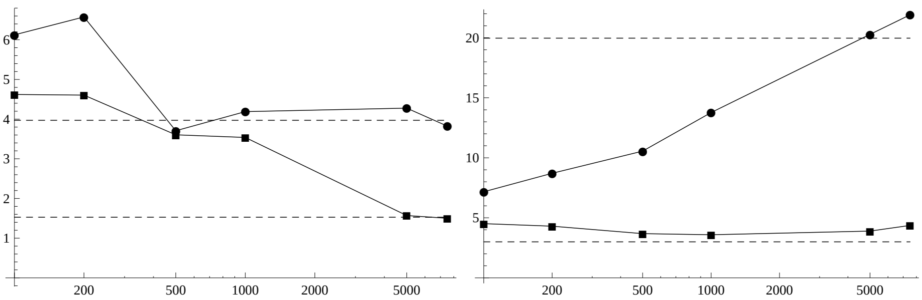
Then $\sqrt{n}(\check{a}_n - a) \Rightarrow N(0, \sigma^2_{a,U})$, $\sqrt{n}(\check{p}_n - p) \Rightarrow N(0, \sigma^2_{p,U})$, where $\sigma^2_{a,U}$, $\sigma^2_{p,U}$ are defined in theorem 4.

7. Simulations and real data example

We carried out a small simulation study to assess the performance of the proposed algorithms on samples of fixed size. The algorithms were implemented in Mathematica 6. B-splines with knots placed equidistantly on an interval $[0, T]$ were used as the basis functions $u_1, \ldots, u_M$. For $x \in [-T, 0]$ the functions were defined as $u_i(x) = -u_i(-x)$. We didn't apply any true root selection algorithm in this study, taking the empirical counterpart of the "true root" as the moment estimate in all cases. The Moore-Penrose inverse matrix was used instead of the true inverse $\tilde{\mathbf{Q}}^{-1}$ to obtain more stable estimates for small sample sizes. Estimates for p were truncated on $[0, 1]$. Empirical means and variances were calculated over estimates by 1000 simulated samples.

In the first simulated example we took $f_0 \sim N(0, 1)$, $f \sim N(0, 0.25)$, $a = 0.5$, $p = 0.5$ (see fig. 2(b) for the distribution of data). The number of basis functions was $M = 5$, the interval half-length $T = 4$.

The results of simulations are presented in table 1 and in fig. 3(a).



(a) (b)

FIGURE 3. Dispersion of estimates for a: $\bullet$ — $\hat{a}^{ME}$, $\blacksquare$ — $\check{a}_n$. (a)First example: variances , (b) Second example: IQ. (Log scale on horizontal n-axis).

n	Bias $\hat{a}^{ME}$	Bias $\check{a}$	Var $\hat{a}^{ME}$	Var $\check{a}$	Bias $\hat{p}^{ME}$	Bias $\check{p}$	Var $\hat{p}^{ME}$	Var $\check{p}$
100	0.135	-0.751	6.12	4.62	0.507	2.36	6.06	6.14
200	0.239	-1.10	6.57	4.60	0.359	2.76	7.02	9.64
500	-0.111	-1.33	3.70	3.60	0.484	2.85	6.41	13.5
1000	0.0782	-0.839	4.18	3.53	0.134	1.70	6.24	12.1
5000	-0.0611	-0.151	4.27	1.56	0.158	0.292	7.01	4.41
7500	0.013	-0.0554	3.82	1.50	-0.073	0.201	5.99	3.25
∞	0	0	3.97	1.527	0	0	6.52	3.54

TABLE 1. Simulations results for the inseparable admixture

Here Bias means the bias of a corresponding estimate multiplied by $\sqrt{n}$ and variances are multiplied by n. The asymptotic values of the characteristics are placed at the bottom line.

Recall that the dispersions of the infeasible best possible GEEs are $\sigma^2_{a*} = 1.5258$ and $\sigma^2_{p*} = 3.538$.

In the second simulated example we used standard normal distribution for the admixture and Student-T_ν distribution with $\nu = 4$ degrees of freedom for the primary component to study the influence of heavy tails on the behavior of the estimates. (Note that for $\nu \le 3$ the third moment of Student-T_ν distribution is infinite so we should not expect the moment estimate to perform adequately). $M = 5$ basis functions were used in the aGEE estimates.

The simulations show that in this case the estimates distributions are also heavy-tailed (is is readily seen by typical outliers at histograms of estimates, see fig. 4). The variances of the normalized estimates don't converge to dispersions. E.g. for $n = 7500$, $n\,\mathrm{Var}(\hat{a}^{ME}_n) = 503.22$, $n\,\mathrm{Var}(\check{a}_n) = 41.076$, while $\sigma^2_{ME}(a) = 219.152$, $\sigma^2_{a*} = 4.94391$. Biases (normalized by $\sqrt{n}$) are also far from 0. On the other hand,

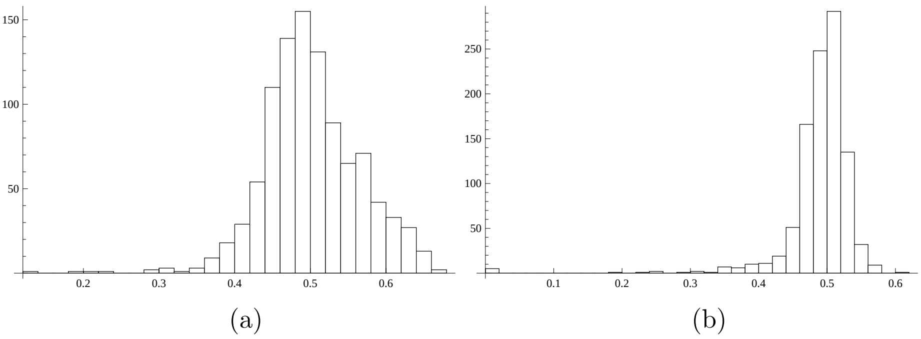


FIGURE 4. Histograms of estimates for p for heavy-tailed primary component. Sample size ($n = 1000$) (a)$\hat{p}_n^{ME}$, (b) $\check{p}_n$.

n	Med $\hat{a}^{ME}$	Med $\check{a}$	IQ $\hat{a}^{ME}$	IQ $\check{a}$	Med $\hat{p}^{ME}$	Med $\check{p}$	IQ $\hat{p}^{ME}$	IQ $\check{p}$
100	0.556	0.489	7.17	4.51	-0.0331	0.489	1.69	1.07
200	-0.0106	0.402	8.71	4.30	0.0335	0.118	1.82	1.08
500	0.105	0.584	10.6	3.68	-0.025	0.0161	2.20	1.01
1000	-0.479	0.787	13.8	3.59	0.0459	-0.108	2.88	1.17
5000	-0.148	0.876	20.3	3.89	0.0689	-0.611	3.82	1.22
7500	-1.20	0.807	21.9	4.37	0.147	-0.180	4.137	1.057
∞	0	0	19.97	3.00	0	0	3.60	0.946

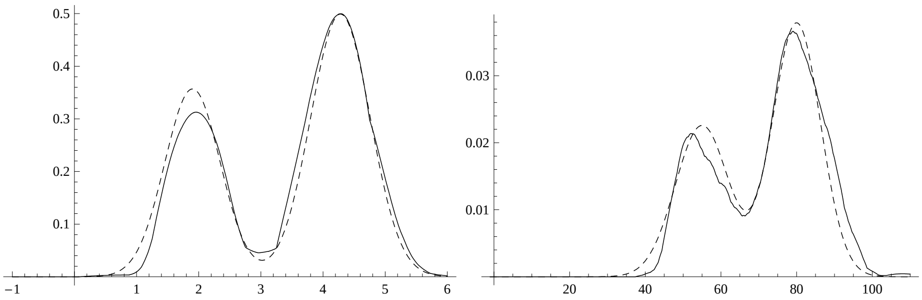
TABLE 2. Simulations results for heavy tailed primary component

the variances of the adapted estimates become less than of the moment ones at the sample sizes $n \geq 500$. The biases of the adaptive estimates are nearly the same as of the moment ones at these sample sizes.

Therefore we used such robust measures of quality as the median and interquartile range to compare the performance of the estimates in this case. The results are presented in e table 2 and in fig. 3(b).

Here Med $\check{a}$ means $\sqrt{n}(\text{Median}(\check{a}_n) - a)$, IQ $\check{a}$ means the interquartile range of $\check{a}$ multiplied by $\sqrt{n}$ and analogous symbols are used for the other estimates. Convergence to asymptotic values is here not so clear as in the previous example, but aGEE estimates overperform the moment ones in dispersion for almost all sample sizes.

In both cases the moment estimates of the location parameter demonstrate significantly better bias then the aGEE ones. In the first example the gain in variance overwhelms the loss in bias only for large sample sizes $n > 1000$. In the second example aGEE estimates are preferable even at $n = 500$.



(a) (b)

FIGURE 5. Kernel density estimates for Geyser data (solid line) and normal mixture approximations (dashed line) (a) *duration*, (b) *waiting*.

Data	a			p	
	Normal	Moment	aGEE	Moment	aGEE
duration	4.23	4.90847	4.09514	0.276	0.290
	1.9	2.222	2.188	0.398	0.294
waiting	80	79.58	80.892	0.296	0.3499
	55	48.95	54.29	0.247	0.434

TABLE 3. Analysis of the Geyser data

It is traditional to use the Old Faithful Geyser data from Azzalini & Bowman (1990) to compare performance of statistical algorithms for two-component mixture analysis. These data contain two variables *duration* and *waiting* for 299 cases. The *duration* data distribution can be nicely fitted by the two-component normal mixture $p_1 f_1^d(x) + p_2 f_2^d$ with $f_1^d \sim N(1.9, 0.2)$, $f_2^d \sim N(4.23, 0.23)$, $p_1 = 0.4$, $p_2 = 0.6$ (see fig. 5(a)). For illustrative purposes we took at first f_1^d as the admixture and the other component as the primary one. In the second experiment f_2^d was the admixture.

The *waiting* distribution can't be approximated by a mixture of two normals, moreover, it seems that the left component in it is not symmetric (see fig 5(b)). Despite this observation we used $p_1 f_1^w(x) + p_2 f_2^w$ with $f_1^w \sim N(55, 50)$, $f_2^w \sim N(80, 40)$ to approximate this distribution. The component f_1^w was the admixture in the third experiment, f_2^w in the fourth one. The results are presented in table 3.

Estimates of location parameter a seem consistent with the estimates of the peaks locations in the density of the observations made by eye.

8. Conclusion

We propose a new way to estimate the Euclidian parameters in the model of observations with symmetric distribution with admixture. The main idea behind our approach is to construct a GEE with the data-adapted estimating function yielding approximately the least possible dispersion of the estimates.

Asymptotic analysis demonstrates potential efficiency of such estimates. Results of simulations show that the adapted GEE estimates of location parameter demonstrate less scattering then simple moment estimates for almost all sample sizes. For mixing probabilities aGEE are also preferable when the sample size is large enough. But aGEE estimates possess inappropriately large bias for small sample sizes. Standard bias reduction techniques (such as bootstrap ones) can be used as medicines to cure this problem.

The number of basis functions M for the aGEE can be considered as a regularization parameter in the nonparametric problem of best estimating pair approximation. Large M allows better approximation of (g_1^*, g_2^*) by $(g_{\tilde{\boldsymbol{\beta}}_1}, g_{\tilde{\boldsymbol{\beta}}_2})$. But the larger M is the more unknown coefficients $\tilde{\beta}_{ki}$ are to be estimated.

Special algorithms for optimal selection of M and the range T as well as more deliberate nonlinear techniques of adaptation should be developed to make the estimates more efficient in real data analysis.

Appendix

Proof of the theorem 1. Denote $\mathbf{t} = (\alpha, \pi)^T$, $h_i(x, \mathbf{t}) = g_i(x - \alpha) - (1 - \pi)G_i(\alpha)$, $\mathbf{h}(x, \mathbf{t}) = (h_1(x, t), h_2(x, t))^T$, $\hat{\mathbf{h}}(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^n \mathbf{h}(\xi_j, \mathbf{t})$. Then the estimating equation (4) can be rewritten as

$$\hat{\mathbf{h}}(\mathbf{t}) = 0.$$

Applying to this equation theorem 5.14 and lemma 5.3. from Shao (1998) we obtain that

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \to N(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \mathbf{M}^{-1}\mathbf{C}(\mathbf{M}^T)^{-1}$, $\mathbf{M} = \frac{\partial}{\partial \mathbf{t}} \mathsf{E}\, \mathbf{h}(\xi_1, \mathbf{t})\big|_{\mathbf{t}=\boldsymbol{\vartheta}}$, $\mathbf{C} = \mathrm{Cov}(\mathbf{h}(\xi_1, \boldsymbol{\vartheta}))$ is the covariance matrix of $\mathbf{h}(\xi_1, \boldsymbol{\vartheta})$. So, to complete the proof, we need to show that $\mathbf{M} = \mathbf{V}$, $\mathbf{C} = \mathbf{S}$.

Note that, for any odd function g, $G(a) = \mathsf{E}\, g(\eta_0 - a) = \int_{-\infty}^{+\infty} g(x - a) f_0(x) dx = \int_{-\infty}^{+\infty} g(x) f_0(x + a) dx = \int gD$ and, for any even function u,

$$\mathsf{E}\, u(\xi_1 - a) = \int_{-\infty}^{+\infty} u(x)(pf(x) + (1 - p) f_0(x + a)) dx = \int uQ.$$

Hence for $i, k = 1, 2$,

$$\begin{aligned}\mathrm{Cov}(h_i(\xi_1, \boldsymbol{\vartheta}), h_k(\xi_1, \boldsymbol{\vartheta})) &= \mathsf{E}\, g_i(\xi_1 - a) g_k(\xi_1 - a) - (1 - p)^2\, \mathsf{E}\, g_i(\eta_0 - a)\, \mathsf{E}\, g_k(\eta_0 - a) \\ &= \int g_i g_k Q - (1 - p)^2 \int g_i D \int g_k D.\end{aligned}$$

So $\mathrm{Cov}(\mathbf{h}(\xi_1, \boldsymbol{\vartheta})) = \mathbf{S}$.

Since g_i is odd

$$\frac{\partial}{\partial \alpha} \mathsf{E}\, h_i(\xi_1, \mathbf{t})\bigg|_{\mathbf{t}=\boldsymbol{\vartheta}} = \frac{\partial}{\partial \alpha} \int_{-\infty}^{+\infty} \pi g_i(x) f(x - a - \alpha) dx\bigg|_{\mathbf{t}=\boldsymbol{\vartheta}} = \int gB$$

and

$$\frac{\partial}{\partial \pi} \mathsf{E}\, h_i(\xi_1, \boldsymbol{\vartheta})\bigg|_{\mathbf{t}=\boldsymbol{\vartheta}} = G_i(a) = \int g_i D.$$

Hence $\mathbf{M} = \mathbf{V}$.

□

To prove theorem 2 we need one auxiliary statement from Hilbert space geometry. Let $\mathbb{H}$ be a Hilbert space with an inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$, $\mathbf{b} = (b_1, \dots, b_d)^T \in \mathbb{H}^d$ be some fixed column vector with entries from $\mathbb{H}$. For any $\mathbf{g} = (g_1, \dots, g_d)^T \in \mathbb{H}^d$, $\boldsymbol{\Gamma}(\mathbf{g}) = (\langle g_i, g_k\rangle)_{i,k=1}^d$ denotes the Gram matrix for the entries of $\mathbf{g}$, $\mathbf{V}(\mathbf{g}) = (\langle g_i, b_k\rangle)_{i,k=1}^d$.

Lemma 1. *Assume that* $\mathbf{V}_0 \in \mathbb{R}^{d\times d}$ *is some fixed matrix,* $\det \boldsymbol{\Gamma}(\mathbf{b}) \neq 0$. *Then for any* $\mathbf{g} \in \mathbb{H}^d$, *so that* $\mathbf{V}(\mathbf{g}) = \mathbf{V}_0$,

$$\boldsymbol{\Gamma}(\mathbf{g}) \geq \boldsymbol{\Gamma}(\mathbf{g}^*),$$

where $\mathbf{g}^* = \mathbf{V}_0 \boldsymbol{\Gamma}(\mathbf{b})^{-1}\mathbf{b}$.

Furthermore $\mathbf{V}\mathbf{g}^* = \mathbf{V}_0$, $\boldsymbol{\Gamma}(\mathbf{g}^*) = \mathbf{V}_0 \boldsymbol{\Gamma}(\mathbf{b})\mathbf{V}_0^T$.

Proof. Denote by $\mathcal{V}$ the linear sub-space in $\mathbb{H}$ spanned by $b_1,\dots,b_d$. For any $g \in \mathbb{H}$, $g^= := \mathrm{Pr}_{\mathcal{V}}\, g$ is the orthogonal projection of g onto $\mathcal{V}$ and $g^\perp := \mathrm{Pr}_{\mathcal{V}^\perp}\, g$ is the orthogonal projection of g onto the orthogonal complement to $\mathcal{V}$, The operations $\mathbf{g}^=$ and $\mathbf{g}^\perp$ are applied to $\mathbf{g} \in \mathbb{H}^d$ coordinatewise.

Note that $\mathbf{V}(\mathbf{g}) = \mathbf{V}(\mathbf{g}^=) + \mathbf{V}(\mathbf{g}^\perp) = \mathbf{V}(\mathbf{g}^=)$ and, for any $\mathbf{c} \in \mathbb{R}^d$,

$$\mathbf{c}^T\boldsymbol{\Gamma}(\mathbf{g})\mathbf{c} = \|\mathbf{c}^T\mathbf{g}\|^2 = \|\mathbf{c}^T\mathbf{g}^=\|^2 + \|\mathbf{c}^T\mathbf{g}^\perp\|^2 \geq \mathbf{c}^T\boldsymbol{\Gamma}(\mathbf{g}^=)\mathbf{c}.$$

Hence if $\mathbf{V}(\mathbf{g}) = \mathbf{V}_0$ then $\mathbf{V}(\mathbf{g}^=) = \mathbf{V}_0$ and $\boldsymbol{\Gamma}(\mathbf{g}) \geq \boldsymbol{\Gamma}(\mathbf{g}^=)$. Since $\mathbf{g}^= \in \mathcal{V}$, it can be represented in the form $\mathbf{g}^= = \mathbf{C}\mathbf{b}$, with some matrix $\mathbf{C} \in \mathbb{R}^{d\times d}$. Then $\mathbf{V}_0 = \mathbf{V}(\mathbf{g}^=) = \mathbf{C}\mathbf{V}(\mathbf{b}) = \mathbf{C}\boldsymbol{\Gamma}(\mathbf{b})$ yields $\mathbf{C} = \mathbf{V}_0\boldsymbol{\Gamma}(\mathbf{b})^{-1}$ and $\mathbf{g}^= = \mathbf{g}^*$.

The matrix $\boldsymbol{\Gamma}(\mathbf{g}^*)$ can be calculated straightforwardly.

□

Proof of theorem 2. Let (g_1, g_2) be some estimating pair, $\mathbf{L} = (l_{ik})_{i,k=1}^2 \in \mathbb{R}^{2\times 2}$ be some nonsingular matrix, $\tilde{g}_i(x) = l_{i1}g_1(x) + l_{i2}g_2(x)$. Note that the GEE (3) with the estimating pair $(\tilde{g}_1, \tilde{g}_2)$ is equivalent to GEE with the initial pair (g_1, g_2): any GEE estimate generated by $(\tilde{g}_1, \tilde{g}_2)$ satisfy (3) with the estimating pair (g_1, g_2) and vice versa.

It is readily seen that $\mathbf{V}(\tilde{g}_1, \tilde{g}_2) = \mathbf{L}\mathbf{V}(g_1, g_2)$, so, if $\det \mathbf{V}(g_1, g_2) \neq 0$ letting $\mathbf{L} = \mathbf{V}(g_1, g_2)^{-1}$, we obtain that any estimating pair is equivalent to some pair (g_1, g_2) with $\mathbf{V}(g_1, g_2) = \mathbf{E}$, where $\mathbf{E}$ is the 2×2-unit matrix. For such pairs

$$\boldsymbol{\Sigma}(g_1, g_2) = \begin{pmatrix} \int (g_1)^2 Q & \int g_1 g_2 Q \\ \int g_1 g_2 Q & \int (g_2)^2 Q - (1-p)^2 \end{pmatrix}.$$

Application of lemma 1 with $\langle g_1, g_2\rangle = \int g_1 g_2 Q$, $b_1 = B/Q$, $b_2 = D/Q$, $\mathbf{V}_0 = \mathbf{E}$ completes the proof.

□

Proof of theorem 4 is analogous to the proofs of theorems 1 and 2.

Proof of theorem 5 will be divided into four lemmas.

Let $v : \mathbb{R} \to \mathbb{R}$ be some nonrandom smooth function, $a \in \mathbb{R}$ and a_n be some sequence of random variables. Denote $\hat{v}_n(\alpha) = \frac{1}{n}\sum_{j=1}^n v(\xi_j - \alpha)$, $V(\alpha) = \mathsf{E}\, v(\xi_j - \alpha)$, $\bar{v}_\varepsilon(x) = \sup_{\alpha:|a-\alpha|<\varepsilon} |v'(x-\alpha)|$.

Lemma 2. *Assume that*
1. $\mathsf{E}\, v(\xi_1 - a)^2 < \infty$.
2. $\sqrt{n}(a_n - a) = O_p(1)$.
3. For some $\varepsilon > 0$, $e = \mathsf{E}\, \bar{v}_\varepsilon(\xi_1) < \infty$.
Then $\sqrt{n}(\hat{v}_n(a_n) - V(a)) = O_p(1)$.

Proof. Note that

$$\sqrt{n}(\hat{v}_n(a_n) - V(a)) = \sqrt{n}(\hat{v}_n(a_n) - \hat{v}_n(a)) + \sqrt{n}(\hat{v}_n(a) - V(a)).$$

The second term is $O_p(1)$ by assumption 1. For the first term we write

$$\sqrt{n}(\hat{v}_n(a) - V(a)) = \frac{1}{n}\sum_{j=1}^{n} v'(\xi_j - \zeta_n)\sqrt{n}(a_n - a), \tag{28}$$

where ζ is some intermediate point between a_n and a.

By assumption 2, $\sqrt{n}(a_n - a) = O_p(1)$. Consider

$$\frac{1}{n}\sum_{j=1}^{n} v'(\xi_j - \zeta_n) =: S_n = S_n \mathbb{1}\{|\zeta_n - a| > \varepsilon\} + S_n \mathbb{1}\{|\zeta_n - a| < \varepsilon\}.$$

Since $a_n \to a$ in probability, $S_n \mathbb{1}\{|\zeta_n - a| > \varepsilon\} = o_p(1)$. Then, in view of assumption 3,

$$|S_n \mathbb{1}\{|\zeta_n - a| < \varepsilon\}| \le \frac{1}{n}\sum_{j=1}^{n} \bar{v}(\xi_j) = O_p(1)$$

by the law of large numbers. So $S_n = O_p(1)$. Substituting this into (28), by assumption 2 we obtain the statement of the lemma.□

Let us now introduce (infeasible) estimators

$$a_n^* = \tilde{a}_n - \sum_{m=1}^{M} \beta_{m1}^*(\hat{u}_m(\tilde{a}_n) - (1 - \tilde{p}_n)U_m(\tilde{a}_n)),$$

$$p_n^* = \tilde{p}_n - \sum_{m=1}^{M} \beta_{m2}^*(\hat{u}_m(\tilde{a}_n) - (1 - \tilde{p}_n)U_m(\tilde{a}_n)),$$

$\boldsymbol{\vartheta}_n^* = (a_n^*, p_n^*)^T$.

Lemma 3. *Under the assumptions of theorem 5,*

$$\sqrt{n}(a_n^* - a) \Rightarrow N(0, \sigma_{a,U}^2), \quad \sqrt{n}(p_n^* - p) \Rightarrow N(0, \sigma_{p,U}^2).$$

Proof. Denote

$$h_i(x, \mathbf{t}) = \sum_{m=1}^{M} \beta_{mi}^*(u_m(x - \alpha) - (1 - \pi)U_{mi}(\alpha)), \quad \hat{h}_i(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n} h_i(\xi_j, \mathbf{t}),$$

$\hat{\mathbf{h}}(\mathbf{t}) = (\hat{h}_1(\mathbf{t}), \hat{h}_2(\mathbf{t}))^T$.

Note that $\boldsymbol{\vartheta}_n^* = \tilde{\boldsymbol{\vartheta}}_n - \hat{\mathbf{h}}(\tilde{\boldsymbol{\vartheta}}_n)$ and

$$\hat{\mathbf{h}}(\tilde{\boldsymbol{\vartheta}}_n) = \hat{\mathbf{h}}(\boldsymbol{\vartheta}) + \mathbf{M}_n(\tilde{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}),$$

where

$$\mathbf{M}_n = \frac{\partial}{\partial \mathbf{t}}\hat{\mathbf{h}}(\mathbf{t})\Big|_{\mathbf{t}=\zeta_n},$$

ζ_n is an intermediate point between $\tilde{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}$.

Hence

$$\sqrt{n}(\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}) = \sqrt{n}(\tilde{\boldsymbol{\vartheta}}_n - \hat{\mathbf{h}}(\tilde{\boldsymbol{\vartheta}}) - \boldsymbol{\vartheta})$$

$$= \sqrt{n}(\tilde{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta} - \hat{\mathbf{h}}(\boldsymbol{\vartheta}) - \mathbf{M}_n(\tilde{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}))$$

$$= \sqrt{n}\hat{\mathbf{h}}(\boldsymbol{\vartheta}) + \sqrt{n}(\mathbf{E} - \mathbf{M}_n)(\tilde{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}).$$

Applying lemma 2 we obtain $\mathbf{M}_n \to \mathsf{E}\,\mathbf{M}$ in probability, where $\mathbf{M} = \frac{\partial}{\partial \boldsymbol{\vartheta}}\hat{\mathbf{h}}(\boldsymbol{\vartheta})$. By the proof of theorem 1, $\mathsf{E}\,\mathbf{M} = \mathbf{V}(g_{\boldsymbol{\beta}_1}, g_{\boldsymbol{\beta}_2})$, where $\mathbf{V}$ is defined by (6), $g_{\boldsymbol{\beta}_i} = \sum_{m=1}^M \beta^*_{mi} u_i$. From (23)-(25) we obtain $\mathbf{V}(g_{\boldsymbol{\beta}_1}, g_{\boldsymbol{\beta}_2}) = \mathbf{E}$. So

$$\sqrt{n}(\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}) \sim_p \sqrt{n}\hat{\mathbf{h}}(\boldsymbol{\vartheta}) \Rightarrow N(0, \boldsymbol{\Sigma}(g_{\boldsymbol{\beta}_1}, g_{\boldsymbol{\beta}_2})).$$

□

Lemma 4. *Under the assumptions of theorem 5,* $\sqrt{n}|\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^*| = O_p(1)$.

Proof. Applying lemma 3 to functions u_i, $u_i u_k$, u_i', $i,k = 1,m$ we obtain $\sqrt{n}$-consistency of $\tilde{d}_i$, $\tilde{q}_{ik}$ and $\tilde{b}_i$ as estimators for d_i, q_{ik} and b_i. Then the definitions (23)-(25) yield the statement of the lemma.□

Lemma 5. *Under the assumptions of the theorem 5,*

$$\sqrt{n}|\check{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_n^*| = o_p(1) \text{ as } n \to \infty.$$

Proof. By the definition of $\tilde{\boldsymbol{\beta}}_i$, $\sum_{m=1}^M \tilde{\beta}_{mi} U_m(\tilde{a}_n) = \tilde{\mathbf{d}}^T \tilde{\boldsymbol{\beta}}_i$ is 0 as $i = 1$ and 1 as $i = 2$. Hence

$$\check{a}_n = \tilde{a}_n - \sum_{m=1}^M \tilde{\beta}_{m1}(\hat{u}_m(\tilde{a}_n) - (1 - \tilde{p}_n)U_m(\tilde{a}_n)),$$

$$\check{p}_n = \tilde{p}_n - \sum_{m=1}^M \tilde{\beta}_{m2}(\hat{u}_m(\tilde{a}_n) - (1 - \tilde{p}_n)U_m(\tilde{a}_n))$$

and for some $C < \infty$,

$$|\check{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_n^*| \le C \sum_{m=1}^M \sum_{i=1}^2 |\tilde{\beta}_{mi} - \beta^*_{mi}||\hat{u}_m(\tilde{a}_n) - (1 - \tilde{p}_n)U_m(\tilde{a}_n)|.$$

In view of lemma 4, we need only to show that

$$\hat{u}_m(\tilde{a}_n) - (1 - \tilde{p}_n)U_m(\tilde{a}_n) = o_p(1) \tag{29}$$

to complete the proof.

Since $\mathsf{E}\,\hat{u}_m(a) = (1 - p)U_m(a)$, applying lemma 2 we obtain

$$\hat{u}_m(\tilde{a}_n) \to (1 - p)U_m(a).$$

Then consistency of the pilot estimates $\tilde{a}_n$, $\tilde{p}_n$ yields (29).□

Combining lemmas 3 and 5 we obtain the statement of theorem 5.

□

References

[1] Avérous, J., Fougéres A.L. & Meste, M. (1996). Tailweight with respect to mode for unimodal distributions. *Statist. Probab. Lett.* **28**, 367-373.

[2] Azzalini, A. and Bowman, A. W. (1990). A Look at Some Data on the Old Faithful Geyser. *Applied Statistics* **39**, 357-365.

[3] Boshnakov, G. N. (2007). Some measures for asymetry of distributions. *Statist. Probab. Lett.* **77**, 1111-1116.

[4] Bordes L., Mottelet S., Vandekerkhove P. (2006). Semiparametric estimation of a two-component mixture model.*Ann. Statist.* **34**, 1204-1232.

[5] Bordes L., Delmas C., Vandekerkhove P. (2006a). Semiparametric Estimation of a two-component Mixture model where one component is known.*Scand. J. Statist.* **33**, 733-752.

[6] Critchley, F., Jones, M.C. (2008). Asymetry and gradient asymetry functions: density-based skewness and kurtosis.*Scand. J. Statist.* **35**, 415-437.

[7] Hall P., Zhou X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann.Statist.*, **31**, 201-224.

[8] Heyde, C.C. (1997).*Quasi-likelihood and its application* Springer, New York.

[9] Holzmann, H., Munk, A. & Stratmann, B. (2004). Identifiability of finite mixtures — with applications to circular distributions. *Sankhya* **66**, 440-449.

[10] Holzmann, H., Munk, A. & Gneiting T. (2006). Identifiability of finite mixtures of elliptical distributions. *Scand. J. Statist.* **33**, 753-763.

[11] Hunter D.R., Wang S., Hettmansperger T.R. (2004). Inference for mixtures of symmetric distributions. *Technical Report 04-01, Penn State University, Philadelphia.*

[12] Hunter D.R., Wang S., Hettmansperger T.R. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35**, 224-251.

[13] Lodatko, N., Maiboroda R. (2007). An adaptive moment estimator of a parameter of a distribution constructed from observations with admixture. *Theory of Probability and Mathematical Statistics* **75**, 71-82.

[14] Maiboroda, R. (2008). Estimation of locations and mixing probabilities by observations from two-component mixture of symmetric distributions.*Teorija Imovirnosti ta Matematychna Statystyka* **78**, 133-141 (in Ukrainian).

[15] McLachlan, G.J., Peel, D. (2000). *Finite Mixture Models.* Wiley, New York.

[16] Newcomb, S. (1894). A generalized theory of the combination of observations so as to obtain the best result.*Amer. J. Math.* **8**, 343-366.

[17] Pearson, K. (1894). Contribution to the mathematical theory of evolution.*Phil. Trans. Roy. Soc. A.* **185**, 71-110.

[18] Pfanzagl, G. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14**, 249-273.

[19] Shao, J. (1998). *Mathematical statistics.* Springer-Verlag, New York.

[20] Small C.G., Wang J. (2003). Numerical methods for nonlinear estimating equations.- Clarendon Press, Oxford, 309p.

[21] Sugakova O. (2009). Estimation of location parameter by observations with admixture.*Teorija Imovirnosti ta Matematychna Statystyka* **80**, to appear (in Ukrainian).

[22] Titterington, D.M., Smith, A.F., Makov, O.E. (1985). *Analysis of Finite Mixture Distributions.* Wiley, New York.

[23] Tsiatis, A. A. (2006). *Semiparametric theory and missing data.* Springer, New York.

Dept. of Probability and Statistics, Faculty of Mechanics and Mathematics, Kyiv National University

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