TECHNICAL REPORT 0106U005864-MS09

Nonparametric density estimation 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. We consider three estimates for the pdf of primary component: a naive one, a symmetrized naive estimate and a symmetrized estimate with adaptive weights. Asymptotic behavior and small sample performance of the estimates are investigated. Some rules of thumb for bandwidth selection are discussed.

Key words: asymptotic normality, finite mixture model, symmetric distribution, kernel density estimate, rule of thumb, MISE

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

Kyiv - 14.05.2009

1. Introduction

In biological and medical statistics observations are frequently contaminated by an admixture of subjects which don't belong to the investigated population. Sometimes the distribution of the admixture is known or can be estimated with high accuracy by some additional data. E.g., in genetical studies one is interested in the effect of a prespecified mutation on some phenotypic feature of living organisms. But the observed sample contains both mutants (i.e. the organisms with the mutation in their genotypes) and wild-type organisms (non-mutants) mistakenly considered as mutants.

In Bordes et al. (2006) such data are described by a two-component model in which the component of interest (the primary component) is assumed to be symmetrically distributed and the admixture distribution has a known pdf. This model can be formulated as follows. Let $\xi_1,\dots,\xi_n$ be the values of the observed feature for n subjects. We assume that $\xi_j \in \mathbb{R}$ are i.i.d. with the pdf

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

where $p \in (0,1)$ is the probability to observe a subject from the primary component (mixing probability), a is the median (location parameter) of the primary component, f is the pdf of the observed feature centered by a for primary component subjects, f_0 is the pdf of the admixture. The unknown pdf f is assumed to be symmetric: $f(x) = f(-x)$, $\forall x \in \mathbb{R}$. The pdf f_0 is known. The Euclidian parameters $a \in \mathbb{R}$ and $p \in (0,1)$ are unknown.

In Bordes et al. (2006), Sugakova (2009), Maiboroda & Sugakova (2009) different $\sqrt{n}$-consistent estimates for the Euclidian parameters were proposed. Let $\hat{a}_n$ and $\hat{p}_n$ be such estimates for a and p respectively. Then the pdf f can be estimated by

$$\hat{f}_n^0(x) = \frac{1}{\hat{p}_n}(\hat{\psi}_n(x+\hat{a}_n) - (1-\hat{p}_n)f_0(x+\hat{a}_n)), \tag{2}$$

where $\hat{\psi}_n$ is any nonparametric estimate for pdf ψ by observations $\xi_1,\dots,\xi_n$. This estimate was proposed in Bordes et al. (2006). In this paper the estimate (2) is considered with $\hat{\psi}_n$ being a kernel density estimate. It is shown that under mild conditions the obtained estimate $\hat{f}_n^0$ is L_1 consistent.

Generally speaking, $\hat{f}_n^0(x)$ is not a symmetric function. Symmetrizing $\hat{f}_n^0(x)$ we get a new estimate

$$\hat{f}_n^s(x) = \frac{1}{2}(\hat{f}_n^0(x) + \hat{f}_n^0(-x)). \tag{3}$$

Since the values of $\hat{\psi}(x)$ and $\hat{\psi}(y)$ are asymptotically uncorrelated for large n as $x \neq y$, we can suppose that the symmetrization not only corrects the shape of the estimate but also improves its accuracy.

To check this conjecture we perform asymptotic analysis of these estimates. It demonstrates that in most cases $\hat{f}_n^s$ has less asymptotic variance than $\hat{f}_n^0$, but sometimes the variance of $\hat{f}_n^0$ is better. The asymptotic results enable us to propose the third estimate which is a weighted mean of $\hat{f}_n^0(x)$ and $\hat{f}_n^0(-x)$:

$$\hat{f}_n^a(x) = \frac{\hat{f}_n^0(x)\psi_n(-x+\hat{a}_n) + \hat{f}_n^0(-x)\hat{\psi}_n(x+\hat{a}_n)}{\hat{\psi}_n(x+\hat{a}_n) + \hat{\psi}_n(-x+\hat{a}_n)}. \tag{4}$$

Asymptotic variance of $\hat{f}_n^a(x)$ is better than the asymptotic variances of both $\hat{f}_n^0(x)$ and $\hat{f}_n^s(x)$ for all $x \in \mathbb{R}$. It worth nothing that $\hat{f}_n^a$ is also symmetric.

The asymptotic results on the estimates are contained in section 2. In fact the quality of a density estimate is characterized by its variance and bias. Special bandwidth selection rules are needed to get a balance between these characteristics. We assess the performance the estimates (2-4) via simulations using the mean integrated squared error as the quality criterion and simplest Silverman's type (cf. Silverman, 1986) rules of thumb for the bandwidth selection. These rules are discussed in section 3 and the results of simulations are presented in section 4. The proofs are collected in the Appendix.

2. Asymptotic normality of estimates

To estimate the pdf ψ by the observations $\xi_1,\dots,\xi_n$ we will use the kernel density estimate

$$\hat{\psi}_n(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x-\xi_j}{h}\right), \tag{5}$$

where K is a kernel, i.e. some fixed pdf, $h > 0$ is a bandwidth. The kernel and the bandwidth are parameters of the estimation algorithm to be selected by some optimality criterion, see Shao (1998), section 5.1.3., Silverman (1986).

To clarify the presentation of the main ideas, in this paper we consider the estimation problem in the case when f and f_0 are twice continuously differentiable functions. (Possible generalizations on other Holder classes of density functions are straightforward.) It is a well known fact (see e.g. Borovkov (1998) section 1.10, Ibragimov & Khasminsky (1979), section 4.4-4.5) that in this case the optimal bandwidth is of order $n^{-1/5}$ as $n \to \infty$ and the optimal kernel must be of the second order, i.e.

$$\int_{-\infty}^{\infty} zK(z)dz = 0, \quad \int_{-\infty}^{\infty} z^2K(z)dz = D < \infty, \quad \int_{-\infty}^{\infty} K^2(z)dz = d^2 < \infty.$$

In what follows we use more restrictive assumptions on the kernel.

Assumption A. *K is a finitely supported pdf (i.e. for some $-\infty < a < b < \infty$, $K(x) = 0$ as $x \notin [a,b]$). K is a second order kernel and its derivative K' has bounded variation on $[a,b]$.*

The following assumption is made on the data distributions.

Assumption B. *Pdfs f and f_0 are twice continuously differentiable and $\sup_x |f_0'(x)| < \infty$, $\sup_x |f'(x)| < \infty$, $\sup_x |f''(x)| < \infty$, $\int_{-\infty}^{\infty} x^2 f(x)dx < \infty$, $\int_{-\infty}^{\infty} x^2 f_0(x)dx < \infty$.*

To construct the estimates for f we use some estimates $\hat{a}_n$ and $\hat{p}_n$ for the Euclidian parameters a and p.

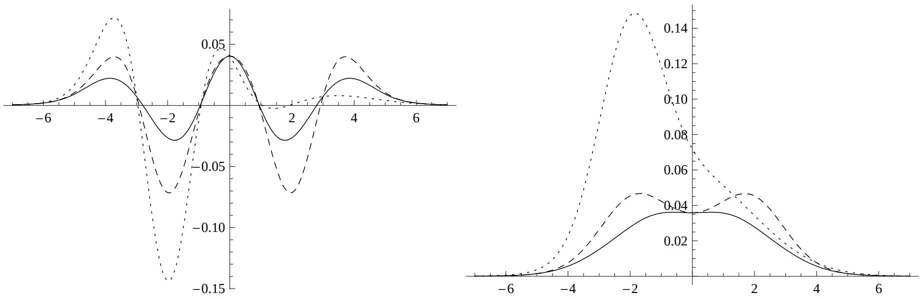
Assumption C. *The estimates $\hat{a}_n$ and $\hat{p}_n$ are $\sqrt{n}$-consistent, i.e.*

$$\sqrt{n}(\hat{a}_n - a) = O_p(1), \quad \sqrt{n}(\hat{p}_n - p) = O_p(1).$$

The last assumption determines the behavior of the bandwidth h as the sample size $n \to \infty$:

Assumption D. *$h = h_n = Hn^{-1/5}$ for some $0 < H < \infty$.*

Note that this assumption can be weakened to $\lim_{n\to\infty} n^{-1/5}h_n = H$.



(a) (b)

FIGURE 1. Quality of estimates for $f \sim N(0,4)$, $f_0 \sim N(0,1)$, $a = 2$, $p = 0.5$.(Dotted line for $\hat{f}_n^0$, dashed line for $\hat{f}^s$, solid line for $\hat{f}^a$). (a)bias (b) variance.

Theorem 1. *Under the assumptions (A)-(D) for any $x_1,\dots x_m \in \mathbb{R}$ the random vector $Z_n = n^{2/5}(\hat{f}_n^0(x_i) - f(x_i))_{i=1}^m$ converges weakly to $Z_\infty = (\zeta_1,\dots,\zeta_m)$, where ζ_i are independent Gaussian r.v.s with*

$$\mathsf{E}\,\zeta_i = b_0(x_i) = \frac{D^2H^2}{2p}\psi''(x_i + a),$$
$$\operatorname{Var}\zeta_i = \sigma_0^2(x_i) = \frac{d^2}{Hp^2}\psi(x_i + a).$$

Corollary 1. *Under the assumptions (A)-(D) for any $x \in \mathbb{R}$*

$$n^{2/5}(\hat{f}_n^s(x) - f(x)) \Rightarrow N(b_s(x), \sigma_s^2(x)),$$

where

$$b_s(x) = \frac{D^2H^2}{4p}(\psi''(x+a) + \psi''(-x+a)), \quad \sigma_s^2(x) = \frac{d^2}{4Hp^2}(\psi(x+a) + \psi(-x+a)).$$

A typical example of asymptotic variances $\sigma_0^2(x)$ and $\sigma_s^2(x)$ as functions of x is represented in fig. 1b. It is readily seen that the maximal value of $\sigma_0^2(x)$ is more than two times higher than of $\sigma_s^2(x)$. But there exists an interval at which $\sigma_s^2(x) > \sigma_0^2(x)$. This effect has a simple explanation. Obviously, if two independent estimates with nearly the same variances are averaged, the variance of the mean is less than the variances of summands. On the other hand, if the first estimate is significantly worse than the second one, then the averaging deteriorates the better estimate. The symmetrized estimate $\hat{f}_n^s(x)$ has the asymptotic variance worse than $\hat{f}_n^0(x)$ at the points x at which $\hat{f}^0(-x)$ is significantly worse than $\hat{f}^0(x)$ due to the admixture influence.

The greater the pdf $f_0(x)$ is, the worse $\hat{f}_n^0(x)$ is as an estimate for $f(x)$. Since $f_0(x)$ is known it seems natural to use the symmetrized estimate $\hat{f}_n^s(x)$ for points

x at which $f_0(-x+\hat{a}_n)$ is not too large and the raw $\hat{f}^0(x)$ at all other points. But what does it mean "not too large"? To answer this question correctly we need to compare $f_0(-x+a)$ with the unknown value of $f(-x)$ which is to be estimated.

Another idea is to average the estimates $\hat{f}^0(x)$ and $\hat{f}^0(-x)$ with some weights w_+ and w_- taking into account different reliability of these estimates. To derive a consistent estimate we need $w_+ + w_- = 1$. The variance of the weighted mean

$$\hat{f}^w(x) = w_+ \hat{f}^0_n(x) + w_- \hat{f}^0_n(-x) \tag{6}$$

is minimal if the weights are taken reciprocally in proportion to the variances of the estimates. So, this way leads to the (infeasible) estimate

$$\hat{f}^\sigma_n(x) = \frac{\sigma_0^2(-x)\hat{f}^0_n(x) + \sigma_0^2(x)\hat{f}^0_n(-x)}{\sigma_0^2(-x) + \sigma_0^2(x)}. \tag{7}$$

Replacing here $\sigma_0^2(\pm x)$ by their estimates $\hat{\sigma}_0^2(x) = d^2\hat{\psi}(x+\hat{a})/(H\hat{p}^2)$, we derive the adaptively weighted estimate $\hat{f}^a_n$ from (4).

Theorem 2. *Under the assumptions (A)-(D) for any $x \in \mathbb{R}$ such that $f(x) > 0$,*

$$n^{2/5}(\hat{f}^a_n(x) - f(x)) \Rightarrow N(b_a(x), \sigma_a^2(x))$$

where

$$b_a(x) = \frac{D^2H^2[\psi(-x+a)\psi''(x+a) + \psi(x+a)\psi''(-x+a)]}{2p[\psi(x+a) + \psi(-x+a)]},$$

$$\sigma_a^2(x) = \frac{d^2\psi(x+a)\psi(-x+a)}{Hp^2[\psi(x+a) + \psi(-x+a)]}.$$

Note that the asymptotic variance $\sigma_a^2(x)$ of $\hat{f}^a_n(x)$ is the same as of "the best" infeasible estimate $\hat{f}^\sigma_n(x)$ from (7). So it is not more than $\sigma_0^2(x)$ and $\sigma_s^2(x)$ for all $x \in \mathbb{R}$.

Up to now we have been dealing with the asymptotic variances of our estimates only. Comparison of biases provides less clear results. It is seen in fig. 1a that the asymptotic bias $b_a(x)$ of $\hat{f}^a_n(x)$ can be worse than that of $\hat{f}^s_n(x)$ at some x. One can try to estimate the bias of $\hat{f}^w(x)$ by some plug-in estimate analogical to the one used above for the variance estimation. Such estimate can be used then to derive the weights w_+ and w_- optimizing some balanced measure of the estimate quality, e.g. the mean squared error: $\mathrm{MSE}[\hat{f}^w(x)] = \mathrm{Var}\,\hat{f}^w(x) + \Big(\mathrm{Bias}[\hat{f}^w(x)]\Big)^2$.

There are significant difficulties on this way. The plug-in estimators of bias based on the formula $n^{2/5}\,\mathrm{Bias}[\hat{f}^0_n(x)] \approx D^2H^2\psi''(x)/(2p)$ are less accurate than their counterparts for the variance, since estimates for ψ'' have very slow convergence. Straightforward bootstrap techniques don't capture the bias of kernel density estimates at all, see Hall P. (1990,1992).

Therefore in many applications (e.g. when a confidence interval for f is constructed) the undersmoothing technique is applied, see Ho & Lee (2008), Hall (1992). In our framework the undersmoothed estimate can be described as an estimate $\hat{f}^w_n$ with the parameter H (and respectively, the bandwidth h) taken so small that its bias can be neglected in comparison to the scattering caused by the variance. Then a confidence interval can be constructed for f based on the estimates $\hat{f}_n$ and $\hat{\sigma}^2$ only. When the undersmoothing paradigm is applied, $\hat{f}^a_n$ seems the most attractive estimate through all the estimates of the form (6).

Balancing of variance and bias terms in the total estimate error needs more deliberate bandwidth selection techniques, such as the cross-validation, see Hall et al. (1992), Bowman (1984), Rudemo (1982), bootstrap h selection with pilot bandwidth depending on h (see Chacón et al., 2008) and others, cf. Devroye (1997), Wand and Jones (1995). In the next section we consider a very simple example of such techniques based on the asymptotic mean integrated squared error (MISE) minimization and compare the performance of our estimates via simulations.

3. Rules of thumb for MISE optimization

MISE is maybe the most popular integral criterion of pdf estimates quality. It is defined by

$$\mathrm{MISE}[\hat{f}] = \mathsf{E}\int_{-\infty}^{\infty}(\hat{f}(x) - f(x))^2dx,$$

where f is the estimated pdf, $\hat{f}$ is its estimate.

It is a well known fact (cf. Borovkov, 1998) that under the assumptions (A), (B) and (D) if $\varphi := \int_{-\infty}^{\infty}(\psi''(x))^2dx < \infty$ then the asymptotic MISE of the kernel estimate $\hat{\psi}_n$ defined by (5) is

$$\mathrm{aMISE}[\hat{\psi}] := \lim_{n\to\infty} n^{4/5}\,\mathrm{MISE}[\psi_n]$$
$$= \lim_{n\to\infty}\left[\frac{D}{n^{1/5}h} + \frac{n^{4/5}h^4}{4}d^2\varphi\right] = \frac{D}{H} + \frac{H^4d^2\varphi}{4}.$$

The value of $\mathrm{aMISE}(\hat{\psi})$ is minimized by $H = \left(\frac{D}{d^2\varphi}\right)^{1/5}$. Many rules for the selection of h are based on the replacement of φ in this formula by some its approximation. Say, one can observe that for a Gaussian pdf ψ with the variance S^2, $\varphi = 3S^{-5}/(8\sqrt{\pi}) \approx 0.212S^{-5}$. Replacing S^2 by its empirical counterpart

$$\hat{S}^2 = \frac{1}{n}\sum_{j=1}^{n}(\xi_j - \bar{\xi})^2,\ \bar{\xi} = \frac{1}{n}\sum_{j=1}^{n}\xi_j, \tag{8}$$

we obtain the popular “Silverman's Rule of Thumb” (RoT) for the bandwidth selection:

$$\hat{h}_n^{RoT} = \left(\frac{8\sqrt{\pi}D}{3d^2n}\right)^{1/5}\hat{S}.$$

One expects that the bandwidth choice $\hat{h}_n^{RoT}$ is appropriate if the estimated density is nearly Gaussian, say it is unimodal, fairly symmetric and do not have too fat tails. Clearly, the mixed density ψ defined by (1) should not possess these fine qualities. It is not usually unimodal even if f and f_0 are. It must not be symmetric if we want to estimate p and a consistently by the data. And there is no a priori evidence on the lightness of its tails.

In order to obtain more adequate bandwidth selection rule we propose to use (1) with zero mean Gaussian f as a model for the true ψ. To simplify the presentation, let f_0 be the standard Gaussian pdf. Then

$$\varphi = \varphi(s,a,p) = \frac{1}{8\sqrt{\pi}s^4(1+s)^{9/2}}\left(e^{-\frac{a^2}{2+2s}}A(s,p,a) + B(s,p)\right),$$

where s^2 is the variance of the primary component f, p is the mixing probability,

$$A(s,p,a) = -8\sqrt{2}(-1+p)ps^4\left(a^4 - 6a^2(1+s) + 3(1+s)^2\right),$$

$$B(s,p) = s^{3/2}(1+s)^{9/2}\left(s^{5/2} - 2ps^{5/2} + p^2\left(1+s^{5/2}\right)\right).$$

Then the modified RoT (mRoT) for our model is

$$\hat{h}_n^{mRoT} = \left(\frac{D}{d^2\varphi(\hat{s}_n, \hat{a}_n, \hat{p}_n)n}\right)^{1/5}, \tag{9}$$

where $\hat{a}_n$ and $\hat{p}_n$ are some estimates for the Euclidian parameters a and p, $\hat{s}^2$ is an estimate for the variance s^2 of f. To estimate s^2 one can use the equation

$$\mathsf{E}(\xi_j)^2 = p(s^2 + a^2) + (1-p)$$

(we keep in mind that $f_0 \sim N(0,1)$ here). So the estimate $\hat{s}_n^2$ for s^2 is

$$\hat{s}_n^2 = \frac{1}{\hat{p}_n}(\hat{m}_n^{(2)} - (1-\hat{p}_n)) - (\hat{a}_n)^2, \tag{10}$$

where $\hat{m}_n^{(2)}$ is some estimate for the second order moment of observations $\mathsf{E}\,\xi_j^2$. The usual second empirical moment $\frac{1}{n}\sum_{j=1}^n \xi_j^2$ can be impractical here if one deals with heavy tailed observations. To avoid the problems with outliers we propose to use the truncated second moment

$$\hat{m}_n^2 = \frac{1}{n(1-\alpha)} \sum_{j \geq n\alpha/2}^{j \leq n(1-\alpha/2)} (\xi_{j:n})^2, \tag{11}$$

where $\xi_{1:n} \leq \xi_{2:n} \leq \cdots \leq \xi_{n:n}$ are the order statistics of the sample (ξ_j, $j = 1,\ldots,n$), α is some truncation level. In the simulation experiments below we took α to be very low and fixed ($\alpha = 0.05$). Theoretically it could be better to use $\alpha = \alpha_n \to 0$ as $n \to \infty$.

Note, that the idea of popular Silverman's "better rule of thumb" (Hardle et al., 2004), in which the interquartile range is utilized to estimate the variance, can't be used directly in our case since it is based on the relation between the interquartile range and variance valid for Gaussian distributions only.

Surely, MISE($\hat{\psi}_n$) and MISE($\hat{f}_n^0$) are different. But, as it is demonstrated in Appendix, $n^{2/5}(\hat{f}_n^0(x) - f(x)) \sim \frac{n^{2/5}}{p}(\hat{\psi}_n(x) - \psi(x))$, so one expects that the choice of h optimal for MISE($\hat{\psi}_n$) minimization should also be good for MISE($\hat{f}_n^0$) minimization.

4. Simulation results

In the simulation experiments we used the moment estimates $\hat{a}_n$ and $\hat{p}_n$ for a and p described in Bordes et al. (2006), Maiboroda & Sugakova (2009). The moment estimate $\hat{a}_n$ is derived as a solution to corresponding generalized estimating equation. In fact, this equation has three roots:

$$\hat{\alpha}_{0,n} = 0,$$

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

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

where $\hat{m}_k = \frac{1}{n}\sum_{j=1}^n (\xi_j)^k$. There are different rules for selection of the "true" root yielding a consistent estimate, discussed in Bordes et al. (2006), Maiboroda & Sugakova (2009). In this presentation we didn't use them, taking directly the

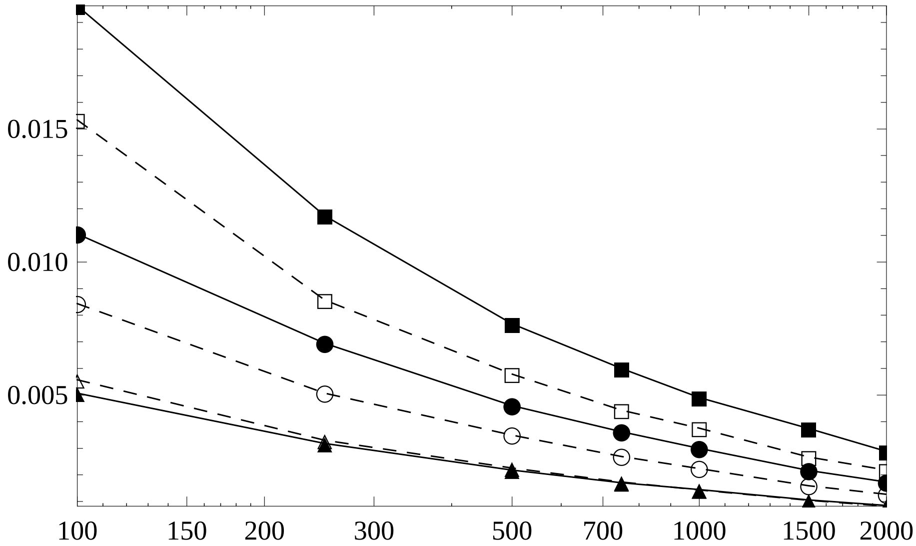


FIGURE 2. MISE for the example 1. Solid lines for $\hat{h}^{RoT}$, dashed ones for $\hat{h}^{mRoT}$. $\square$ for $\hat{f}_n^0$, $\circ$ for $\hat{f}_n^s$, $\triangle$ for $\hat{f}_n^a$.

$\hat{h}_n$	RoT			mRoT		
n	$\hat{f}_n^0$	$\hat{f}_n^s$	$\hat{f}_n^a$	$\hat{f}_n^0$	$\hat{f}_n^s$	$\hat{f}_n^a$
100	0.0196268	0.0110584	0.0050736	0.0153381	0.00843476	0.00557842
250	0.0117411	0.00694704	0.00318095	0.00856864	0.00506838	0.00330121
500	0.00767588	0.00460257	0.0021817	0.00578782	0.00349479	0.00225732
750	0.00600432	0.00362277	0.00170438	0.0044321	0.00269315	0.00173252
1000	0.00490611	0.00299481	0.00144874	0.00375372	0.00223889	0.00143337
1500	0.00374382	0.00216507	0.0010568	0.00266725	0.00159258	0.00104137
2000	0.00288292	0.00173214	0.000855187	0.00217863	0.00127051	0.000827089

TABLE 1. MISE of estimates for the example 1.

root $\hat{a}_{+,n} = \hat{a}_n$ which is the "true" one in all the cases considered below. Then $\hat{p}_n = \hat{m}_1/\hat{a}_n$.

In all the experiments the integral $\int_{-\infty}^{\infty}(\hat{f}_n(x) - f(x))^2 dx$ was calculated for estimates $\hat{f}_n = \hat{f}_n^0, \hat{f}_n^s$ and $\hat{f}_n^a$ by 1000 samples of size n and then averaged by all the samples. Both $\hat{h}_n^{RoT}$ and $\hat{h}_n^{mRoT}$ were used. The sample size was changed from 100 to 2000. The admixture was $f_0 \sim N(0,1)$ in all the experiments.

Eample 1. $f \sim N(0,4)$, $a = 2$, $p = 0.5$. The results of simulation are presented in Table 1 and Fig. 2. They demonstrate that $\hat{f}_n^a$ overperforms $\hat{f}_n^s$ and $\hat{f}_n^0$ in all cases. The use of modified RoT improves $\hat{f}_n^s$ and $\hat{f}_n^0$ significantly and causes insignificant deterioration of $\hat{f}_n^a$ for small sample sizes ($n \le 750$).

Example 2. In this example we are interested in the performance of our estimates on heavy tailed distributions. So the distribution of primary component f

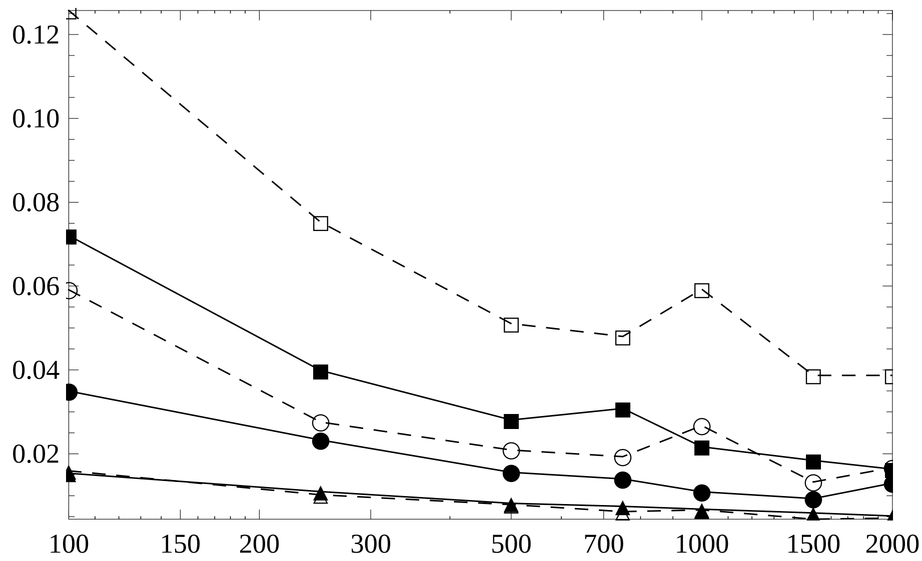


FIGURE 3. MISE for the example 2. Solid lines for $\hat{h}^{RoT}$, dashed ones for $\hat{h}^{mRoT}$. $\square$ for $\hat{f}_n^0$, $\circ$ for $\hat{f}_n^s$, $\triangle$ for $\hat{f}_n^a$.

$\hat{h}_n$	RoT			mRoT		
n	$\hat{f}_n^0$	$\hat{f}_n^s$	$\hat{f}_n^a$	$\hat{f}_n^0$	$\hat{f}_n^s$	$\hat{f}_n^a$
100	0.0720441	0.0349648	0.0153465	0.125742	0.0591245	0.0159273
250	0.0398813	0.0232679	0.0109806	0.0752593	0.0275706	0.010199
500	0.0280436	0.0155846	0.00817104	0.051032	0.020907	0.00785878
750	0.0307909	0.0139871	0.00746996	0.0479655	0.0192972	0.00619378
1000	0.0216628	0.0109195	0.00679174	0.0592487	0.0266879	0.00658356
1500	0.018407	0.00929889	0.00587021	0.0387086	0.0132671	0.00438548
2000	0.0163254	0.0130204	0.0051492	0.0387058	0.0166901	0.00467035

TABLE 2. MISE of estimates for the example 2

is taken to be Student-T with d=4 degrees of freedom (if d were less than 4 there would be no third moment of f and the moment estimate $\hat{a}_n$ would be inconsistent). The Euclidian parameters are $a = 2$, $p = 0.5$. The results of the experiment are presented in Table 2 and Fig. 3. Here we see that $\hat{f}_n^a$ also overperforms $\hat{f}_n^s$ and $\hat{f}_n^0$ in all cases. But the modification of RoT deteriorates $\hat{f}_n^s$ and $\hat{f}_n^0$. The estimate $\hat{f}_n^a$ now is improved slightly by the modification of RoT for sample sizes larger than 100. This effect can be explained by the bad behavior of the estimate $\hat{a}_n$ in this case. It seems tempting to use some modifications of $\hat{a}_n$ appropriate for heavy tailed distributions such as those considered in Maiboroda & Sugakova (2009).

Example 3. Now f is taken to be the pdf of Beta distribution with parameters $\alpha = \beta = 3$. So, we can see how the departure of normality changes the estimates performance in the light tailed case. The Euclidian parameters of the model are $a = 2$, $p = 0.5$. The results are displayed in fig. 4 and in table 3. It is readily

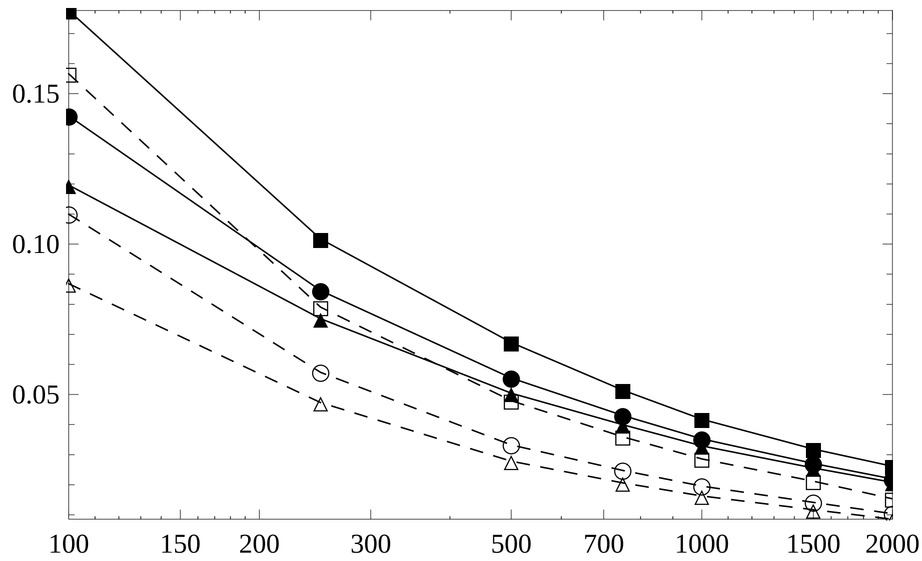


FIGURE 4. MISE for the example 3. Solid lines for $\hat{h}^{RoT}$, dashed ones for $\hat{h}^{mRoT}$. $\square$ for $\hat{f}_n^0$, $\circ$ for $\hat{f}_n^s$, $\triangle$ for $\hat{f}_n^a$.

$\hat{h}_n$	RoT			mRoT		
n	$\hat{f}_n^0$	$\hat{f}_n^s$	$\hat{f}_n^a$	$\hat{f}_n^0$	$\hat{f}_n^s$	$\hat{f}_n^a$
100	0.177646	0.142616	0.119593	0.156615	0.109951	0.0868058
250	0.101728	0.084551	0.0751758	0.0790001	0.0573581	0.0473337
500	0.0672797	0.0554897	0.0504437	0.0479307	0.0332667	0.0278323
750	0.0514419	0.0430169	0.0399633	0.0359726	0.0247899	0.020638
1000	0.0417739	0.0352523	0.0329516	0.0286076	0.0195597	0.0162556
1500	0.031889	0.0270768	0.0255979	0.0211883	0.0141618	0.0116775
2000	0.0261339	0.0219275	0.0207947	0.015355	0.010396	0.00854644

TABLE 3. MISE of estimates for the example 3.

seen that $\hat{f}_n^a$ is still the best estimate for all sample sizes. The use of modified RoT improves all the estimates for all n as well.

5. CONCLUSION

We considered three estimates for the symmetric primary component pdf in the model of observations with admixture. Asymptotic analysis and results of simulations demonstrate that symmetrization of the estimate proposed by Bordes et al. significantly improves the variance and integrated mean squared error of the estimate. The adaptively weighted estimate $\hat{f}_n^a$ shows the best performance in almost all cases for all sample sizes. The modification of rule of thumb for the bandwidth selection can improve the estimates of pdf if the estimates of Euclidian parameters used in mRoT are good enough. More complicated algorithms of bandwidth selection (e.g. cross-validation) should be developed to make these estimates more accurate.

Appendix

Sketch proof of theorem 1. Denote

$$\hat{f}_n^*(x) = \frac{1}{p}(\hat{\psi}_n(x+a) - (1-p)f_0(x+a)).$$

It is an infeasible estimate for $f(x)$ if a or p is unknown. Since

$$n^{2/5}(\hat{f}_n^*(x) - f(x)) = \frac{1}{p}n^{2/5}(\hat{\psi}_n(x+a) - \psi(x+a)),$$

by the usual theorem on the asymptotic normality of $\hat{\psi}_n$ (Borovkov (1998) section 1.10) we obtain

$$Z_n^* = (n^{2/5}(\hat{f}_n^*(x_i) - f(x_i)), i = 1, \ldots, m) \Rightarrow Z_\infty. \tag{12}$$

The key observation is that

$$\hat{f}_n^0(x) - \hat{f}^*(x) = o_p(n^{-2/5}). \tag{13}$$

To show this note that by the assumption C, $\hat{a}_n - a = O_p(n^{-1/2})$, $\hat{p}_n - p = O_p(n^{-1/2})$, so

$$\frac{1-\hat{p}_n}{\hat{p}_n} f_0(x+\hat{a}_n) - \frac{1-p}{p} f_0(x+a) = O_p(n^{-1/2}),$$

since f_0' is bounded.

Therefore we only need to show that

$$\hat{\psi}_n(x+\hat{a}_n) - \hat{\psi}_n(x+a) = O_p(n^{-2/5}).$$

Note that

$$\hat{\psi}_n(x+\hat{a}_n) - \hat{\psi}_n(x+a) = \hat{\psi}_n'(\zeta)(\hat{a}_n - a), \tag{14}$$

where ζ is some intermediate point between $x+\hat{a}_n$ and $x+a$. Denote

$$\bar{\psi}_n'(x) = \mathsf{E}\,\hat{\psi}_n'(x) = \psi'(x) + O(h_n^2).$$

It is obvious that $\bar{\psi}_n'(x)$ is uniformly (by x and n) bounded, so

$$\bar{\psi}_n'(\zeta)(\hat{a}_n - a) = O(n^{-1/2}). \tag{15}$$

Then

$$|\hat{\psi}_n'(x) - \bar{\psi}_n'(x)| = \frac{1}{h^2}\left|\int_{-\infty}^{\infty} K'\left(\frac{x-y}{h}\right) d(F_n(y) - F(y))\right|,$$

where F is the cdf of ξ_1, F_n is the empirical cdf of the sample $(\xi_j,\ j = 1, \ldots, n)$. By the Vapnik-Chervonenkis inequality (see e.g. Devroye (1996), chapter 7.9) for all $\varepsilon > 0$

$$\mathsf{P}\left\{\sup_y |F_n(y) - F(y)| > \varepsilon\right\} \le 6(2n+1)\exp\left(-\frac{\varepsilon^2(n-1)}{2}\right),$$

so

$$J_n = \sup_y |F_n(y) - F(y)| = O_p\left(\frac{\log n}{\sqrt{n}}\right)$$

and

$$\sup_x |\hat{\psi}'_n(x) - \bar{\psi}'(x)| \le \frac{1}{h^2} V_{K'} J_n = O_p\left(\frac{\log n}{n^{1/10}}\right), \tag{16}$$

where $V_{K'}$ is the variation of K' on $\mathbb{R}$. Combining (14-16) we get

$$|\hat{\psi}_n(x + \hat{a}_n) - \hat{\psi}_n(x + a)| = O_p(n^{1/2}),$$

so (12) holds. This with (13) completes the proof. □

Proof of theorem 2. Denote

$$w_+^n = \frac{\hat{\psi}_n(-x + \hat{a}_n)}{\hat{\psi}_n(-x + \hat{a}_n) + \hat{\psi}_n(x + \hat{a}_n)}, \quad w_-^n = 1 - w_+^n.$$

Then

$$n^{2/5}(\hat{f}_n^a(x) - f(x)) = w_+^n(\hat{f}_n^0(x) - f(x)) + w_-^n(\hat{f}_n^0(-x) - f(-x)).$$

Since

$$w_+^n \to \frac{\psi_n(-x + a)}{\psi_n(-x + a) + \psi(x + a)} \text{ as } n \to \infty \text{ in probability,}$$

the statement of theorem 2 follows from theorem 1.

□

References

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

[2] Borovkov A.A. (1998) *Mathematical Statistics.* Gordon and Breach Science Publishers, Amsterdam.

[3] Bowman A.W., (1984) An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71**, 353-360.

[4] Chacón J.E., Montanero J., Nogales A.G. (2008) Bootstrap bandwidth selection using an h-dependent pilot bandwidth. *Scand. J. Statist.* **35**, 139-157.

[5] Devroye, L., Györfi, L., and Lugosi, G. (1996). A Probabilistic Theory of Pattern Recognition, Springer, New York.

[6] Devroye, L. (1997) Universal smoothing factor selection in density estimation: theory and practice (with discussion). *Test* **6**, 223-320.

[7] Hall P. (1990) Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32**, 177-203.

[8] Hall P. (1992) Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20**, 675-694.

[9] Hall P., Marron J.S., Park B.U. (1992) Smoothed cross-validation. *Probab. Theory Related Fields* **92**, 1-20.

[10] Hardle W., Muller M., Sperlich S., Werwatz A. (2004) *Nonparametric and Semiparametric Models.* Berlin: Springer-Verlag.

[11] Ho Y.H.S., Lee S.M.S. (2008) Iterated bootstrap-t confidence intervals for density functions.*Scand. J. Statist.* **35**, 295-308.

[12] Ibragimov, I.A.,Khasminsky R.Z. (1979). *Asymptotic Estimation Theory.* Nauka, Moscow (in Russian).

[13] Maiboroda R, Sugakova O. Generalized estimating equations for symmetric distributions observed with admixture. *Technical report TR 0106U005864-MS08* http://probability.univ.kiev.ua/texrep/MSSymAdmixBi.pdf.

[14] Rudemo M. (1982) Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* **9**, 65-78.

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

[16] Silverman B.W. (1986)*Density estimation for statistics and data analysis.* Chapman & Hall, London.

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

[18] Sugakova O. Density estimation by observations with admixture. *Technical report TR 0106U005864-S08* http://probability.univ.kiev.ua/texrep/SSymAdmixPDF.pdf.

[19] Wand M.P. Jones M.C. (1995) *Kernel smoothing.* Chapman & Hall, London.

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

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

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

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