Теорія Ймовір. та Матем. Статист.
Вип. 73, 2005

Theor. Probability and Math. Statist.
No. 73, 2006, Pages 99–108
S 0094-9000(07)00684-9
Article electronically published on January 17, 2007

ESTIMATION OF THE DENSITY OF A DISTRIBUTION FROM DATA WITH AN ADMIXTURE

UDC 519.21

N. LODATKO AND R. MAĬBORODA

Abstract. We consider the problem of estimation of a density from observations of a two-component mixture with varying concentrations. It is assumed that the distribution of the first component is unknown, while a parametric model is (perhaps) available for the second component. Applying the sieve maximum likelihood method we construct histogram-type estimators for the densities of distributions of the components and estimators for unknown parameters of the second component. We prove the consistency of the estimators and obtain estimates for the rate of convergence.

1. Introduction and survey of the literature

The problem of the analysis of data with admixture appears quite often in biostatistics and in medical statistics. Below we discuss one of the possible examples of such data.

Let N patients be under medical evaluation, since a preliminary diagnosis (say, "atypical pneumonia") is made for all of them. A certain medical test is ordered for every patient; let the result of the test be ξ. The complete blood count is an example of such a test. The test results are given by a sample

$$\Xi_N = (\xi_1, \dots, \xi_N),$$

where ξ_j is the value of ξ for the patient j. Assume that the distribution of ξ for atypical pneumonia patients is different from that for healthy persons. Denote the density of the distribution of ξ for atypical pneumonia patients by h_1, and that for healthy persons by h_2. If all N persons are atypical pneumonia patients, then Ξ_N is an ordinary sample from the density h_1. Unfortunately, atypical pneumonia is not always easily distinguished from acute bronchitis and other respiratory infections. Thus it is likely that the sample Ξ_N is mixed with observations ξ_j from the density h_2. However one can estimate the probability w_j that person j is an atypical pneumonia patient. This can be done by a thorough study of the symptoms of person j. In such a case, the density of ξ_j is a mixture of the densities h_1 and h_2, namely

$$w_j h_1(x) + (1 - w_j) h_2(x).$$

How can one estimate h_1 in such a case?

This paper is devoted to the problem of the estimation of densities from data with admixture for the case where the density h_1 of the principal component is unknown and has to be estimated.

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

Key words and phrases. Sieve maximum likelihood method, histogram, a mixture with varying concentrations, consistency.

The assumptions imposed on the density of the admixture h_2 may vary to some extent; for example,

(1) h_2 is unknown like h_1 (the nonparametric case);
(2) h_2 is known; for example, it can be estimated from observations ξ for persons who do not suffer from atypical pneumonia (the deterministic case);
(3) h_2 is known up to certain parameters (a parametric model of data with admixture).

The third case appears if the distribution of ξ in the sample for persons who are ill but are not atypical pneumonia patients is of the same type as that for healthy persons. One can assume, for example, that both distributions are Gaussian but that they have different mean values.

Applying the sieve maximum likelihood method we construct estimators of the histogram type for all three cases mentioned above. For the third case, we also construct estimators of unknown parameters of the distribution of the admixture. We prove that the estimators are consistent and provide estimates for the rate of convergence of the estimators to the true values.

The method of constructing the estimators is described in Section 2. The main results are stated in Section 3, and the proofs are given in Section **??**.

The general approach to analyze mixtures with varying concentration is described in [3]. The density of a distribution is estimated in [4, 2] with the help of kernel estimators for the nonparametric case. Histogram estimators are studied in [5] for homogeneous samples. A description of the sieve maximum likelihood method can be found in [6].

2. The setting of the problem

For the asymptotic analysis of estimators, it is reasonable to view the sample Ξ_N as an element of a scheme of series, namely $\Xi_N = (\xi_{1:N}, \dots, \xi_{N:N})$, where $\xi_{j:N}$ are independent random variables for any fixed N, and

$$\mathsf{P}\{\xi_{j:N} < x\} = w_{j:N} H_1(x) + (1 - w_{j:N}) H_2(x),$$

where H_1 is the distribution function of the principal component, H_2 is the distribution function of the admixture, and $w_{j:N}$ is the concentration of the principal component in the mixture when the observation j is made; that is, $w_{j:N}$ is the probability that the observation j is taken from the principal component.

In what follows we assume that the supports of the distributions H_i belong to a finite interval. Without loss of generality, we assume that all the supports coincide with $[0, 1]$. We also assume that the densities h_i of the distributions H_i exist with respect to the Lebesgue measure and moreover

(A) there are numbers $0 < u < U < \infty$ such that $u \le h_i(x) \le U$ for all $x \in [0, 1]$.

To construct the estimators we follow the sieve maximum likelihood method. We seek an estimator $\hat{h}_i^N(x)$ for $h_i(x)$ among step functions of the form

$$g_k(x) = \sum_{k=1}^{K_N} g_k \, \mathbb{1}\{x \in A_k\}, \tag{1}$$

where K_N is the number of subintervals of the partition A_k, $A_k = [t_{k-1}, t_k)$ for $k = 1, \dots, K_N - 1$, $A_{K_N} = [t_{K_N-1}, t_{K_N}]$, $t_k = k/K_N$, and g_i are arbitrary positive numbers such that

$$\frac{1}{K_N} \sum_{k=1}^{K_N} g_k = 1.$$

Let $0 < u_N < U_N < \infty$ be some real numbers. We denote by $\mathcal{H}_N$ the set of all functions (1) such that $u_N \le g_k \le U_N$, $k = 1, \dots, K_N$. Let

$$\chi_{kj} = \mathbb{1}\{\xi_{j:N} \in A_k\}.$$

If the true densities of the components are $h_i = g_i(x)$ and

$$g_i(x) = \sum_{k=1}^{K_N} g_{ki}\, \mathbb{1}\{x \in A_k\},$$

then the logarithm of the likelihood function constructed from the sample Ξ_N is given by

$$\begin{aligned} l(g_1, g_2) &= \sum_{j=1}^{N} \ln\left(w_{j:N} \sum_{k=1}^{K_N} g_{k1}\chi_{kj} + (1 - w_{j:N}) \sum_{k=1}^{K_N} g_{k2}\chi_{kj} \right) \\ &= \sum_{j=1}^{N} \sum_{k=1}^{K_N} \chi_{kj} \ln(w_{j:N} g_{k1} + (1 - w_{j:N}) g_{k2}). \end{aligned}$$

If both densities h_1 and h_2 are unknown, then one can choose the pair $(\hat{h}_1^N, \hat{h}_2^N) \in \mathcal{H}_N \times \mathcal{H}_N = \mathcal{H}_N^2$ that maximizes l on $\mathcal{H}_N^2$ as an estimator for (h_1, h_2):

$$(2) \qquad \left(\hat{h}_1^N, \hat{h}_2^N\right) = \operatorname*{arg\,max}_{(g_1, g_2) \in \mathcal{H}_N^2} l(g_1, g_2).$$

If arg max is attained at several pairs of functions, then any of them can serve as an estimator for (h_1, h_2).

The sieve maximum likelihood method leads to histogram estimators of densities in the case of homogeneous data, that is, in the case of pure samples. By analogy, any estimator of the form (1) is called a histogram.

If the density of the second component h_2 is known up to a parameter ϑ, then

$$h_2(x) = h_2(x; \vartheta),$$

where $\vartheta \in \Theta$ is the unknown parameter. Put

$$\bar{h}_2(x, \vartheta) = \sum_{k=1}^{K_N} \bar{h}_{k2}(\vartheta)\, \mathbb{1}\{x \in A_k\},$$

where

$$\bar{h}_{k2} = \bar{h}_{k2}(\vartheta) = \frac{1}{s_N} \int_{t_{k-1}}^{t_k} h_2(x, \vartheta)\, dx.$$

Let $\Theta_N = \{\vartheta \in \Theta \colon \bar{h}_2(\cdot, \vartheta) \in \mathcal{H}_N\}$. As an estimator for the pair of unknown parameters of the distribution of (h_1, ϑ) we take

$$(3) \qquad \left(\hat{h}_1^N, \hat{\vartheta}^N\right) = \operatorname*{arg\,max}_{g_1 \in \mathcal{H}_N, \tau \in \Theta_N} l(g_1, \bar{h}_2(\cdot, \tau));$$

that is, we seek the maximum among those “histograms” whose second component is obtained by averaging the density corresponding to the parametric model.

If h_2 is known, then one can use (3) for the set Θ of parameters containing a single element. Thus

$$\hat{h}_1 = \operatorname*{arg\,max}_{g_1 \in \mathcal{H}_N} l(g_1, \bar{h}_2(\cdot))$$

if h_2 is known.

3. Main results

Let $s_N = 1/K_N$ be the length of the intervals of the partition A_k and let

$$\bar{w}_N = \frac{1}{N}\sum_{j=1}^{N} w_{j:N}$$

be the "mean concentration" of the principal component in the mixture.

The following result contains conditions for the $L_2[0,1]$ consistency of histogram estimators for the nonparametric case, that is, in the case of functions defined on [0,1] and whose norm is given by

$$\|a\|_2 = \left(\int_0^1 a^2(x)\,dx\right)^{1/2}.$$

Theorem 3.1. *Let condition* (A) *hold. Assume that*

(1) *there exists $c > 0$ such that $N^{-1}\sum_{j=1}^{N}(w_{j:N} - \bar{w}_N)^2 > c$ for all N,*
(2) *$s_N \to 0$ and $s_N N \to \infty$ as $N \to \infty$,*
(3) *$U_N \to \infty$, $u_n \to 0$,*

$$\frac{U_N^5}{N s_N u_N^4} \to 0, \qquad \frac{\ln^2 u_N}{s_N N} \to 0 \qquad \text{as } N \to \infty.$$

Then

$$\left\|\hat{h}_i^N - h_i\right\|_2 \to 0$$

in probability as $N \to \infty$ for the estimators $\hat{h}_i^N$, $i = 1, 2$, defined by (2).

In the case of a parametric model for the admixture, let

$$\omega(s) = \sup_{|x-y|\le s}\sup_{\tau\in\Theta}|h_2(x,\tau) - h_2(y,\tau)|$$

be the uniform modulus of continuity of the density h_2 for all possible values of the unknown parameter.

The conditions for consistency are contained in the following result.

Theorem 3.2. *Let condition* (A) *and assumptions (1)–(3) of Theorem* 3.1 *hold. Moreover, let*

(1) *$\omega(s) \to 0$ as $s \to 0$,*
(2) *Θ is a compact set in the semimetrics $\rho(\alpha,\tau) = \|h_2(\cdot,\alpha) - h_2(\cdot,\tau)\|_2$,*
(3) *$\rho(\tau,\vartheta) > 0$ for $\tau \ne \vartheta$.*

Then $\|\hat{h}_1^N - h^1\|_2 \to 0$ and $\rho(\hat{\vartheta}^N,\vartheta) \to 0$ in probability as $N \to \infty$ for the estimators $\hat{h}_1^N$ and $\hat{\vartheta}^N$ defined by (3).

Introducing extra conditions imposed on the densities h_1 and h_2 one can obtain the rate of convergence of the estimators. For example, let

$$\omega_i(s) = \sup_{|x-y|\le s}|h_i(x) - h_i(y)|.$$

Consider the nonparametric case. By the symbol C we denote all positive finite constants; C may even denote different constants in the same relation if this causes no confusion.

Theorem 3.3. *Let condition* (A) *and assumption (1) of Theorem* 3.1 *hold. Moreover assume that*

(1) *$\omega_i(s) \le Cs^\beta$ for some $\beta > 0$,*
(2) *$s_N = CN^{-\alpha}$ for some $\alpha > 0$,*
(3) *$U_N = C\ln N$ and $u_N = C/\ln N$.*

Then for the estimators defined by (2) *and all* γ *such that*

$$0 < \gamma < \alpha\beta, \qquad \gamma < (1-\alpha)/4,$$

there are constants $C_1, C_2 > 0$ *such that*

$$\mathsf{P}\big\{\|\hat{h}_i^N - h_i\|_2 > \lambda N^{-\gamma}\big\} \le \frac{C_1 \ln N}{N^{1-\alpha}(\lambda N^{-\gamma} - C_2 N^{-\alpha\beta})^4}$$

for all $\lambda > 0$ *and all sufficiently large* N.

4. Proofs of theorems

For $g_{k1}, g_{k2} > 0$, put

$$\hat{J}_k(g_{k1}, g_{k2}) = \frac{1}{N} \sum_{j=1}^{N} \chi_{kj} \ln\big(w_{j:N} g_{k1} + (1 - w_{j:N}) g_{k2}\big).$$

Then

$$\frac{1}{N} l(g_1, g_2) = \hat{J}(g_1, g_2) := \sum_{k=1}^{K_N} \hat{J}_k(g_{k1}, g_{k2}).$$

In what follows we denote by the same symbol g_i both the step function

$$\sum_{k=1}^{K_N} g_{ki}\, \mathbb{1}\{x \in A_k\}$$

and the set of numbers $(g_{ki}, k = 1, \dots, K_N)$.

Let η_i be random variables with the distribution H_i. Let

$$\bar{h}_{ki} = \frac{1}{s_N}\, \mathsf{P}\{\eta_i \in A_k\} = \frac{1}{s_N} \int_{t_{k-1}}^{t_k} h_i(x)\, dx,$$

$$\bar{h}_i(x) = \sum_{k=1}^{K_N} \bar{h}_{ki}\, \mathbb{1}\{x \in A_k\},$$

$$\begin{aligned} J_k(g_{k1}, g_{k2}) &= \mathsf{E}\, \hat{J}_k(g_{k1}, g_{k2}) \\ &= \frac{1}{N} \sum_{j=1}^{N} s_N \big(w_{j:N} \bar{h}_{k1} + (1 - w_{j:N}) \bar{h}_{k2}\big) \cdot \ln\big(w_{j:N} g_{k1} + (1 - w_{j:N}) g_{k2}\big), \end{aligned}$$

$$J(g_1, g_2) = \sum_{k=1}^{K_N} J_k(g_{k1}, g_{k2}).$$

Lemma 4.1. *Let condition* (A) *hold. Then*

$$\begin{aligned} &\mathsf{P}\left\{ \sup_{g_1, g_2 \in \mathcal{H}_N} \big|\hat{J}(g_1, g_2) - J(g_1, g_2)\big| > \varepsilon \right\} \\ &\qquad \le \frac{4 U_N}{s_N N \varepsilon^2} \left(\frac{U_N^4}{u_N^4} + 2 \frac{U_N^2}{u_N^2} + \max\big(\ln^2(u_N), \ln^2(U_N)\big) \right) \end{aligned}$$

for all $\varepsilon > 0$.

Proof. Put $f(g_{k1},g_{k2})=\hat{J}_k(g_{k1},g_{k2})-J_k(g_{k1},g_{k2})$. Applying twice the Newton–Leibniz formula with respect to the variables t_1 and t_2 we obtain

$$\begin{aligned} f(x_1,x_2) &= \int_{u_N}^{x_2}\int_{u_N}^{x_1} \frac{\partial^2}{\partial t_1 \partial t_2} f(t_1,t_2)\,dt_1\,dt_2 \\ &\quad + \int_{u_N}^{x_2} \frac{\partial}{\partial t_2} f(u_N,t_2)\,dt_2 + \int_{u_N}^{x_1} \frac{\partial}{\partial t_1} f(t_1,u_N)\,dt_1 + f(u_N,u_N). \end{aligned} \tag{4}$$

Now we use the Cauchy–Schwarz inequality for the first three terms in (4) and get

$$\begin{aligned} \sup_{x_1,x_2\in[u_N,U_N]} |f(x_1,x_2)| \le & \left(\int_{u_N}^{U_N}\int_{u_N}^{U_N} \left(\frac{\partial^2}{\partial t_1\partial t_2} f(t_1,t_2)\right)^2 dt_1\,dt_2 (U_N-u_n)^2\right)^{1/2} \\ & + \left(\int_{u_N}^{U_N} \left(\frac{\partial}{\partial t_2} f(u_N,t_2)\right)^2 dt_2 (U_N-u_n)\right)^{1/2} \\ & + \left(\int_{u_N}^{U_N} \left(\frac{\partial}{\partial t_1} f(t_1,u_N)\right)^2 dt_1 (U_N-u_n)\right)^{1/2} + f(u_N,u_N). \end{aligned}$$

We have

$$\mathsf{E} \sup_{x_1,x_2\in[u_N,U_N]} (f(x_1,x_2))^2 \le 4\left(R_1(U_N-u_n)^2 + R_2(U_N-u_N) + R_3(U_N-u_N) + R_4\right),$$

where

$$R_1 = \int_{u_N}^{U_N}\int_{u_N}^{U_N} \left(\frac{\partial^2}{\partial t_1\partial t_2} f(t_1,t_2)\right)^2 dt_1\,dt_2, \qquad R_2 = \int_{u_N}^{U_N} \left(\frac{\partial}{\partial t_2} f(u_N,t_2)\right)^2 dt_2,$$

$$R_3 = \int_{u_N}^{U_N} \left(\frac{\partial}{\partial t_1} f(t_1,u_N)\right)^2 dt_1, \qquad R_4 = f^2(u_N,u_N).$$

Now we consider every term R_i, $1\le i\le 4$, separately. First,

$$\begin{aligned} R_1 = \frac{1}{N^2}\int_{u_N}^{U_N}\int_{u_N}^{U_N} \sum_{j,l=1}^{N} & \frac{w_{j:N}(1-w_{j:N})w_{l:N}(1-w_{l:N})}{(w_{j:N}t_1+(1-w_{j:N})t_2)^2(w_{l:N}t_1+(1-w_{l:N})t_2)^2} \\ & \times \mathsf{E}(\chi_{jk}-\mathsf{E}\,\chi_{jk})(\chi_{lk}-\mathsf{E}\,\chi_{lk})\,dt_1\,dt_2. \end{aligned}$$

If $l\neq j$, then

$$\mathsf{E}(\chi_{jk}-\mathsf{E}\,\chi_{jk})(\chi_{lk}-\mathsf{E}\,\chi_{lk}) = 0,$$

while if $l=j$, then

$$\mathsf{E}(\chi_{jk}-\mathsf{E}\,\chi_{jk})(\chi_{lk}-\mathsf{E}\,\chi_{lk}) \le \mathsf{E}\,\chi_{j,k}.$$

Thus

$$R_1 \le \frac{1}{N^2}\sum_{j=1}^{N}\int_{u_N}^{U_N}\int_{u_N}^{U_N} \frac{1}{u_N^4} s_N\left(w_{j:N}\bar{h}_{k1}+(1-w_{j:N})\bar{h}_{k2}\right)dt_1\,dt_2 \le \frac{s_N(U_N-u_n)^2U}{Nu_N^4}.$$

Then we obtain for R_2 and R_3 that

$$R_2 \le \int_{u_N}^{U_N}\sum_{j=1}^{N} \frac{w_{j:N}^2}{(w_{j:N}t_1+(1-w_{j:N})u_N)^2}\,\mathsf{E}(\chi_{jk}-\mathsf{E}\,\chi_{jk})^2\,dt_1 \le \frac{s_N(U_N-u_N)U}{Nu_N^2},$$

and similarly

$$R_3 \le \frac{s_N(U_N-u_N)U}{Nu_N^2}.$$

For R_4, we have

$$R_4 \le \frac{1}{N^2}\sum_{j=1}^{N} \ln^2(u_N)\,\mathsf{E}(\chi_{j:N} - \mathsf{E}\,\chi_{j:N})^2 \le \frac{1}{N}\ln^2(u_N)s_N U.$$

Thus

$$\mathsf{E}\sup_{t_1,t_2\in[u_N,U_N]}\left|\hat{J}_k(t_1,t_2) - J_k(t_1,t_2)\right|^2 \le \frac{s_N D}{N}$$

for

$$D = 4U\left(\frac{(U_N-u_N)^4}{u_N^4} + 2\frac{(U_N-u_N)^2}{u_N^2} + \ln^2(u_N)\right),$$

whence we obtain that

$$\begin{aligned}
&\left(\mathsf{E}\sup_{g_1,g_2\in\mathcal{H}_N}\left|\hat{J}(t_1,t_2) - J(t_1,t_2)\right|^2\right)^{1/2} \\
&\quad\le \left(\mathsf{E}\left(\sum_{k=1}^{K_N}\sup_{g_{1k},g_{2k}\in[u_N,U_N]}\left|\hat{J}_k(g_{1k},g_{2k}) - J_k(g_{1k},g_{2k})\right|\right)^2\right)^{1/2} \\
&\quad\le \left(K_N\sum_{k=1}^{K_N}\mathsf{E}\sup_{g_{1k},g_{2k}\in[u_N,U_N]}\left(\hat{J}_k(g_{1k},g_{2k}) - J_k(g_{1k},g_{2k})\right)^2\right)^{1/2} \\
&\quad\le \sqrt{K_N^2\frac{s_N D}{N}} = \sqrt{\frac{D}{s_N N}}.
\end{aligned}$$

Now we use the Chebyshev inequality:

$$\mathsf{P}\left\{\sup_{g_1,g_2\in\mathcal{H}_N}\left|\hat{J}(t_1,t_2) - J(t_1,t_2)\right| \ge \varepsilon\right\} \le \frac{D}{s_N N\varepsilon^2}.$$

Thus the lemma is proved. □

Proof of Theorem 3.1. Put

$$B_N(\varepsilon) = \left\{\sup_{g_1,g_2\in\mathcal{H}_N}\left|\hat{J}(g_1,g_2) - J(g_1,g_2)\right| > \varepsilon\right\}.$$

According to Lemma 4.1,

$$\mathsf{P}\{B_N(\varepsilon)\} \to 0 \quad \text{as } N\to\infty$$

for all $\varepsilon > 0$. Denote by $\bar{B}_N(\varepsilon)$ the complement of the event $B_N(\varepsilon)$. If the event $\bar{B}_N(\varepsilon)$ occurs, then

$$\left|\hat{J}(\hat{h}_1^N,\hat{h}_2^N) - J(\hat{h}_1^N,\hat{h}_2^N)\right| < \varepsilon$$

and $|\hat{J}(\bar{h}_1,\bar{h}_2) - J(\bar{h}_1,\bar{h}_2)| < \varepsilon$. Thus

$$J\big(\hat{h}_1^N,\hat{h}_2^N\big) \ge \hat{J}\big(\hat{h}_1^N,\hat{h}_2^N\big) - \varepsilon \ge \hat{J}\big(\bar{h}_1,\bar{h}_2\big) - \varepsilon \ge J\big(\bar{h}_1,\bar{h}_2\big) - 2\varepsilon,$$

whence it follows that

$$0 \ge J\big(\hat{h}_1^N,\hat{h}_2^N\big) - J\big(\bar{h}_1,\bar{h}_2\big) \ge -2\varepsilon.$$

Let

$$\rho_{1N}^j := \int_0^1 \ln\left(\frac{w_{j:N}\hat{h}_1^N(x) + (1-w_{j:N})\hat{h}_2^N(x)}{w_{j:N}\bar{h}_1(x) + (1-w_{j:N})\bar{h}_2(x)}\right)\big(w_{j:N}\bar{h}_1(x) + (1-w_{j:N})\bar{h}_2(x)\big)\,dx.$$

By the definition of $J(h_1,h_2)$ we have

$$\rho_{1N} := \left|\frac{1}{N}\sum_{j=1}^{N}\rho_{1N}^j\right| \le 2\varepsilon.$$

Every term ρ_{1N}^j in the latter sum is the Kullback–Leibler divergence between densities $w_{j:N}\hat{h}_1^N(x) + (1 - w_{j:N})\hat{h}_2^N(x)$ and $w_{j:N}\bar{h}_1(x) + (1 - w_{j:N})\bar{h}_2(x)$. Using the relation between the Kullback–Leibler divergence and the Hellinger distance [1] for

$$\rho_{2N}^j := \int_0^1 \left(\sqrt{w_{j:N}\bar{h}_1^N(x) + (1 - w_{j:N})\bar{h}_2(x)} - \sqrt{w_{j:N}\bar{h}_1^N(x) + (1 - w_{j:N})\bar{h}_2(x)}\right)^2 dx$$

we get $\rho_{2N}^j \le \rho_{1N}^j$ and

$$\rho_{2N} := \frac{1}{N}\sum_{j=1}^N \rho_{2N}^j \le \rho_{1N} \le 2\varepsilon.$$

Let

$$\rho_{3N}^j := \int_0^1 \Big(\big(w_{j:N}\hat{h}_1^N(x) + (1 - w_{j:N})\hat{h}_2^N(x)\big) - \big(w_{j:N}\bar{h}_1(x) + (1 - w_{j:N})\bar{h}_2(x)\big)\Big)^2 dx.$$

Since

$$|\sqrt{a} - \sqrt{b}| = |a - b|/(\sqrt{a} + \sqrt{b}) \ge |a - b|/\big(2\max(\sqrt{a}, \sqrt{b})\big)$$

for all positive a and b, we get

$$\rho_{3N} := \frac{1}{N}\sum_{j=1}^N \rho_{3N}^j \le 2\max\big(\sqrt{U_N}\sqrt{U}\big)\rho_{2N} \le 4\sqrt{U_N}\varepsilon$$

if N is sufficiently large.

Let $b_{11} = N^{-1}\sum_{j=1}^N w_{j:N}^2$,

$$b_{12} = b_{21} = N^{-1}\sum_{j=1}^N w_{j:N}(1 - w_{j:N}), \qquad b_{22} = N^{-1}\sum_{j=1}^N (1 - w_{j:N})^2,$$

and $B = (b_{ik})_{i,k=1}^2$. We have

$$\rho_{3N} = b_{11}\big\|\hat{h}_1^N - \bar{h}_1\big\|_2^2 + 2b_{12}\big\|\hat{h}_1^N - \bar{h}_1\big\|_2 \cdot \big\|\hat{h}_2^N - \bar{h}_2\big\|_2 + b_{22}\big\|\hat{h}_2^N - \bar{h}_2\big\|_2^2.$$

Thus

$$\rho_{3N} \ge \lambda_{\min}\big(\|\hat{h}_1 - \bar{h}_1\|_2^2 + \|\hat{h}_2 - \bar{h}_2\|_2^2\big),$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimal and maximal eigenvalues of the matrix B, respectively. Since the entries of the matrix B do not exceed 1, we conclude that $\lambda_{\max} \le 2$. According to assumption (1),

$$\lambda_{\min}\lambda_{\max} = \det B \ge c,$$

whence $\lambda_{\min} > c/2$. Therefore

$$\big\|\bar{h}_i - \hat{h}_i^N\big\|_2^2 \le \frac{8\sqrt{U_N}\varepsilon}{c}.$$

Let z be an arbitrary positive number and $\varepsilon = cz/(8\sqrt{U_N})$. Applying Lemma 4.1 we prove that

$$\mathsf{P}\big\{\|\bar{h}_i - \hat{h}_i^N\|_2^2 \ge z\big\} \le \frac{36UU_N}{cs_N N z^2}\left(\frac{U_N^4}{u_N^4} + 2\frac{U_N^2}{u_N^2} + \ln^2(u_N)\right). \tag{5}$$

The right hand side of the latter inequality approaches zero as $N \to \infty$ in view of assumptions (2) and (3). Now we use Theorem 6 of [5] to show that

$$\int_0^1 |\bar{h}_i(x) - h_i(x)|\, dx \to 0 \quad \text{as } s_N \to 0.$$

Condition (A) yields

$$|h_i(x)| \le U;$$

hence $|\bar{h}_i(x)| \le U$. Thus

$$\|\bar{h}_i - h_i\|_2^2 \le 2U \int_0^1 \left|\bar{h}_i(x) - h_i(x)\right| dx \to 0$$

as $N \to \infty$. Taking into account (5) we complete the proof of the theorem. □

Proof of Theorem 3.2. Similarly to the proof of Theorem 3.1, we apply Lemma 4.1 to prove that $\|\hat{h}_1^N - \bar{h}_1\|_2 \to 0$ and $\|\bar{h}_2(\cdot, \hat{\vartheta}^N) - \hat{h}_2(\cdot, \vartheta)\|_2 \to 0$ as $N \to \infty$. Given arbitrary $x \in [0, 1]$ choose A_i such that $x \in A_i$. Then

$$\left|\bar{h}_2(x, \tau) - h_2(x, \tau)\right| \le \frac{1}{s_N} \int_{A_i} |h_2(y, \tau) - h_2(x, \tau)|\, dy \le \omega(s_N)$$

for all $\tau \in \Theta$. This implies that $\|\bar{h}_2(\cdot, \tau) - h_2(\cdot, \tau)\|_2 \le \omega(s_N) \to 0$ as $N \to \infty$ and $\|h_2(\cdot, \hat{\vartheta}^N) - h_2(\cdot, \vartheta)\|_2 \to 0$ in probability. Taking into account assumptions (2) and (3) we get $\hat{\vartheta}^N \to \vartheta$ in probability as $N \to \infty$.

The theorem is proved. □

Proof of Theorem 3.3. Similarly to the proof of Theorem 3.2 we obtain

$$\|\hat{h}_i - h_i\|_2 \le \omega(s_N).$$

Thus

$$\begin{aligned} \mathsf{P}\{\|\hat{h}_i^N - h_i\|_2 \ge \lambda N^{-\gamma}\} &\le \mathsf{P}\{\|\hat{h}_i^N - h_i\|_2 + \|\bar{h}_i - h_i\|_2 \ge \lambda N^{-\gamma}\} \\ &\le \mathsf{P}\{\|\hat{h}_i^N - h_i\|_2 + \omega(s_N) \ge \lambda N^{-\gamma}\} \\ &\le \mathsf{P}\{\|\hat{h}_i^N - h_i\|_2 \ge \lambda N^{-\gamma} - CN^{-\alpha\beta}\}. \end{aligned}$$

Applying (5) we complete the proof of the theorem. □

5. Concluding remarks

We constructed sieve maximal likelihood estimators for the densities of distributions from data with admixture and proved their consistency. The inequality obtained in Theorem 3.3 provides a rough upper bound for the rate of convergence of these estimators. A further investigation is necessary to establish a better asymptotics of these estimators and to study their behavior for small samples.

Bibliography

1. A. A. Borovkov, *Mathematical Statistics*, "Nauka", Moscow, 1984; English transl., Gordon and Breach Science Publishers, Amsterdam, 1998. MR782295 (86i:62001)
2. Yu. O. Ivan'ko, *Asymptotic behavior of kernel estimators of densities and their derivatives constructed from a mixture with varying concentrations*, Visnyk Kyiv Nat. Univ., Ser. Mathematics and Mechanics **9–10** (2003), 29–35.
3. R. E. Maĭboroda, *Statistical Analysis of Mixtures. A Course of Lectures*, Kyiv University, Kyiv, 2003. (Ukrainian)
4. O. V. Sugakova, *Asymptotics of a kernel estimate for the density of a distribution constructed from observations of a mixture with varying concentration*, Teor. Ĭmovīr. Mat. Stat. **59** (1998), 156–166; English transl. in Theor. Probability and Math. Statist. **59** (1999), 161–171. MR1793776

5. L. Devroye and L. Gyorfi, *Nonparametric Density Estimation. The L_1 View*, Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics, John Wiley & Sons, Inc., New York, 1985. MR780746 (86i:62065)
6. X. Shen, *On methods of sieves and penalization*, Ann. Statist. **25** (1997), no. 6, 2555–2591. MR1604416 (2000a:62074)

Department of Probability Theory and Mathematical Statistics, Faculty for Mathematics and Mechanics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kyiv 03127, Ukraine

E-mail address: lodatko@yandex.ru

Department of Probability Theory and Mathematical Statistics, Faculty for Mathematics and Mechanics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kyiv 03127, Ukraine

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

Received 4/DEC/2004

Translated by S. KVASKO