Теорія Ймовір. та Матем. Статист.
Вип. 70, 2004

Theor. Probability and Math. Statist.
No. 70, 2005, Pages 53–60
S 0094-9000(05)00630-7
Article electronically published on August 26, 2005

ASYMPTOTICS OF EMPIRICAL BAYES RISK IN THE CLASSIFICATION OF A MIXTURE OF TWO COMPONENTS WITH VARYING CONCENTRATIONS

UDC 519.21

YU. O. IVAN'KO

Abstract. We consider the problem of classification for a sample from a mixture of several components. For the problem of classification of a two-component mixture with the space of characteristics $\Re = [a, b] \subset \mathbf{R}$ and smooth distribution densities, we find the precise rate of convergence for the error L_N of the empirical Bayes classifier g_N to the error L^* of the Bayes classifier, namely we prove that

$$N^{4/5}(L_N - L^*) \Rightarrow [A + B\varsigma]^2$$

where ς is a standard normal random variable, and the empirical Bayes classifier g_N is constructed from the kernel estimator of the density of a mixture with varying concentrations. We prove that the kernel estimator with the Epanechnikov kernel is optimal for the empirical Bayes classifier.

1. Introduction

Classification is a quite common procedure when analyzing various data. An example of classification appears in psychology where the behavior of people or their abilities are studied. Other examples are encountered in biological and medical investigations when studying specific features of an illness, or testing new medicines, or determining the influence of environmental factors (such as irradiation or electromagnetic waves). We also mention an example from sociology where a popular approach is to classify people according to their electoral preferences.

Extensive literature is devoted to various problems of classification. The Bayes approach is considered in [3] where the empirical Bayes classifier is studied for a sample whose members are classified. The problem of classification for a mixture with varying concentrations is considered in [4].

In this paper we consider a classification method for which the learning sample is obtained from a sample with varying concentrations. It turns out that the classification problem can be solved in this case under minimal a priori assumptions. We construct the empirical Bayes classifier g_N from the learning sample, and consider the asymptotic properties of its Bayes error L_N. We find the precise rate of convergence of the error L_N to the error L^* of the Bayes classifier in the case of the problem of classification for a two-component mixture with smooth distribution densities, namely we prove that $N^{4/5}(L_N - L^*) \Rightarrow \eta^2$ where $\eta \sim N(A, B^2)$ and A and B are defined by (8).

Our proof is based on results on the asymptotic behavior of estimators of the density of a mixture with varying concentrations obtained in [1], and on results on the behavior of nonhomogeneous empirical functions and measures obtained in [6] and [2].

2000 *Mathematics Subject Classification.* Primary 62H30; Secondary 62C10, 62C12.

2. The setting of the problem

In classification problems, each member of a sample may belong to one of M different classes. Based on observations of some specific characteristics of the members of the sample, the problem is to decide for each member to which class it belongs.

If the distributions of the characteristics are known for all classes of the sample, then the Bayes classifier

$$g^*(X) = \arg\max_k w_k f_k(X) \tag{1}$$

is the optimal solution of the above problem where the random vector X assumes values in the space $\Re = \mathbf{R}^d$.

The random variable Y assumes values in the set $\{1, \dots, M\}$ and is treated as the number of a class containing a member whose characteristics are X;

$$H_k(x) = \mathsf{P}\{x/Y = k\}$$

is the conditional distribution of characteristics in a class k and, by assumption, it has the density $f_k(x)$; $w_k = \mathsf{P}\{Y = k\}$ is the concentration of the class k in the sample.

The classifier $g^*(X)$ minimizes the probability of the error L^*,

$$L^* = \mathsf{P}\{g^*(X) \neq Y\} = \min_g \mathsf{P}\{g(X) \neq Y\} \tag{2}$$

(see [3]).

If the distributions (densities) of characteristics and concentrations of components are unknown, then they are estimated from the learning sample, and the estimators are substituted to the Bayes classifier.

Substituting the estimators into (1) we obtain the classifier

$$g_N(X) = \arg\max_k \hat{w}_N^k \hat{f}_N^k(X), \tag{3}$$

called the empirical Bayes estimator. We denote by

$$L_N = \mathsf{P}\{g_N(X) \neq Y/D_N\} \tag{4}$$

the conditional probability of the wrong classification for such a classifier if the learning sample

$$\left\{\xi_j^N, j = 1, \dots, N\right\}$$

is fixed and $D_N = \sigma\{\xi_j^N, j = \overline{1, N}\}$.

Note that L_N is a Borel function of the sample, and hence it is a random variable.

Example 2.1. If the classification of members of a learning sample $\{\xi_j^N, j = \overline{1, N}\}$ is known, that is, if the number Y_j of a class containing the member j of the learning sample is known, and if the concentrations of the classes in the sample are constant, then the relative frequence

$$\hat{w}_N^k = \frac{1}{N} \sum_{j=1}^N I\{Y_j = k\}$$

can serve as an estimator of the concentration of the class k. The densities of characteristics of components can be estimated with the help of the kernel estimator

$$\hat{f}_N^k(x) = \frac{1}{N h_N^d} \sum_{j=1}^N K\left(\frac{x - \xi_j^N}{h_N}\right) I\{Y_j = k\}.$$

A natural question arises on how large is the difference between the probability of the wrong classification of the empirical Bayes classifier and that of the Bayes classifier.

It is proved in [3], under certain assumptions, that given $\varepsilon > 0$ there is n_0 such that

$$\mathsf{P}\{L_N - L^* > \varepsilon\} \le \exp(-A_\varepsilon N)$$

for $N > n_0$, where A_ε is a constant depending on ε and independent of N.

Below we consider the case where the learning sample $\{\xi_j^N, j = 1, \dots, N\}$ is taken from a mixture with varying concentrations. The classification is unknown for such samples. On the other hand, the probability that a given member belongs to a certain class is known. The concentration of components is different for different samples. Denote by $\operatorname{ind}(j)$ the number of a class containing the member j. The true value of $\operatorname{ind}(j)$ is unknown; instead the concentration $w_{j,N}^k = \mathsf{P}\{\operatorname{ind}(j) = k\}$ of the component k in a mixture for the member j is known.

The distribution of characteristics of a member is then given by

$$\mathsf{P}\left\{\xi_j^N \in A\right\} = \mu(A) = \sum_{k=1}^{M} w_{j,N}^k H_k(A), \qquad A \in \Re = \mathbf{R}^d, \tag{5}$$

where $H_k(A) = \mathsf{P}\{\xi_j^N \in A / \operatorname{ind}(j) = k\}$ is the conditional distribution of characteristics in the class k. Based on this information, our goal is to classify new observations, that is, to decide which of M classes contains a given member if the characteristics of the member are known.

In order to construct the empirical Bayes classifier (3) we estimate the densities of components from a learning sample taken from a mixture with varying concentrations. As an estimator in this case we consider the kernel estimator of the density constructed in [1], namely

$$\hat{f}_N^k(x) = \frac{1}{N h_N^d} \sum_{j=1}^{N} a_{j,N}^k K\left(\frac{x - \xi_j^N}{h_N}\right) \tag{6}$$

where $K(x)$ is a kernel, that is, a density of some probability distribution on $\Re$;

$$a_{j,N}^k = \frac{1}{\det \Gamma_N} \sum_{i=1}^{M} (-1)^{k+i} \gamma_{ki} w_{j,N}^k$$

are weight coefficients defined for $\det \Gamma_N \ne 0$; $\Gamma_N = \left(\langle w_\cdot^k, w_\cdot^l \rangle\right)_{k,l=1}^M$ is the Gram matrix; $\langle w_\cdot^k, w_\cdot^l \rangle = N^{-1} \sum_{i=1}^M w_{j,N}^k w_{j,N}^l$; γ_{ki} is the principal minor of Γ_N.

Assume that

- a1) $H_k(x)$ possesses the density $f_k(x)$, $1 \le k \le M$;
- a2) $\hat{f}_N^k(x)$ are estimators of the form (6) such that $a = \sup_{k,j,N} a_{j,N}^k < \infty$;
- a3) $h_N \to 0$ and $N h_N^d \to \infty$ as $N \to \infty$.

Theorem 2.1 ([4]). *If assumptions a1)–a3) hold for all $\varepsilon > 0$, then there is n_0 such that*

$$\mathsf{P}\{L_N - L^* > \varepsilon\} \le \exp(-A_\varepsilon N), \qquad N > n_0, \tag{7}$$

where the constant A_ε depends on ε and does not depend on N.

3. Main result

Below we find the rate of convergence of the error L_N of the empirical Bayes classifier to the error L^* of the Bayes classifier in the case of two-component mixtures ($M = 2$) and if $\xi_j \in [a, b]$ where a and b are finite numbers.

Let a sample $\{\xi_j^N, j = 1, N\}$ be taken from a two-component mixture. Assume that the concentrations $\{w_{j,N}^k\}_{j=1}^{N}{}_{k=1}^{2}$ of the components are different for different members of the sample. Let a random variable X be the characteristic of the members of the sample and let X assume values in $\Re = [a, b] \subset \mathbf{R}$ where $a < b$ are real numbers. Denote by λ_1 and $\lambda_2 = 1 - \lambda_1$ the concentrations of components.

Assume that

b1) the Bayes classifier splits the space of characteristics $\Re = [a, b] \subset \mathbf{R}$ into two connected sets, that is, there is only one point x_0 where the graphs of $\lambda_1 f_1(x)$ and $\lambda_2 f_2(x)$ intersect;

b2) $f_k(x)$ are continuous and for some $c < \infty$

$$f_k(x) < c, \qquad 1 \le k \le M;$$

b3) there exist the derivatives $f_k'(x)$ and $f_k''(x)$ and they are bounded in a neighborhood of x_0;

b4) the tangents of functions $\lambda_1 f_1(x)$ and $\lambda_2 f_2(x)$ at the point x_0 have different slopes, that is, $\lambda_1 f_1'(x_0) \ne \lambda_2 f_2'(x_0)$;

b5) the limits

$$\sigma_k^2(x) = \lim_{N\to\infty} \sum_{r=1}^{M} \left\langle (a_\cdot^k)^2, (w_\cdot^r) \right\rangle_N \cdot f_r(x) < \infty, \qquad 1 \le k \le M,$$

exist;

b6) a kernel K is such that

$$\sup_{x\in\Re} |K(x)| \le \overline{K} < \infty, \qquad \sup_{x\in\Re} |K'(x)| \le \overline{K} < \infty,$$
$$\operatorname*{Var}_{x\in\Re} K(x) \le \overline{K} < \infty, \qquad \operatorname*{Var}_{x\in\Re} K'(x) \le \overline{K} < \infty.$$

b7) $\int_{-\infty}^{\infty} zK(z)\,dz = 0$, $D^2 = \int_{-\infty}^{\infty} z^2 K(z)\,dz < \infty$, $d^2 = \int_{-\infty}^{\infty} K^2(z)\,dz < \infty$;

b8) $h_N \to 0$ and $Nh_N \to \infty$ as $N \to \infty$;

b9) $\sup_{k,j,N} a_{j,N}^k < \infty$ and $\sup_{k,N} \operatorname{Var}_j a_{j,N}^k < \infty$.

Put

$$\sigma^2(x) = \lim_{N\to\infty} \sum_{r=1}^{M} \left\langle \left(\sum_{l=1}^{M} \lambda_l a_\cdot^l \right)^2, (w_\cdot^r) \right\rangle_N \cdot f_r(x),$$
$$\phi_s(x_0) = \lambda_2 f_2^{(s)}(x_0) - \lambda_1 f_1^{(s)}(x_0).$$

Theorem 3.1. *It follows from assumptions b1)–b9) for $h_N = c/N^{1/5}$ that*

$$N^{4/5}(L_N - L^*) \Rightarrow [A + B\varsigma]^2 \tag{8}$$

where ς is a standard normal random variable,

$$A = D^2 c^{2/5} \frac{\phi_2(x_0)}{2\sqrt{2\phi_1(x_0)}}, \qquad B = \frac{d}{c^{1/10}} \frac{\sigma(x_0)}{\sqrt{2\phi_1(x_0)}}.$$

Remark 3.1. Assumptions b5) and b3) imply that $\sigma^2(x)$ and $\phi_s(x_0)$ are finite for $1 \le s \le 2$ and $x \in \Re$.

Remark 3.2. We seek the unknown parameter c such that

$$c = B\frac{d^2}{D}$$

where B is some constant. Relation (8) is of the form

$$N^{4/5}(L_N - L^*) \Rightarrow (d^2D)^{4/5}\left[B^{2/5}\frac{\phi_2(x_0)}{2\sqrt{2\phi_1(x_0)}} + \frac{1}{B^{1/10}}\frac{\sigma(x_0)}{\sqrt{2\phi_1(x_0)}}\varsigma\right]^2 \tag{9}$$

in this case.

Relation (9) allows one to choose the optimal kernel K as a function minimizing d^2D. An optimal solution of this problem is found by Epanechnikov:

$$K(z) = \frac{3}{4}\left(1 - |z|^2\right), \qquad |z| \le 1,$$

(see [3, Lemma 18]). Note that the optimal kernel does not depend on the distribution of characteristics of members of the sample.

4. Nonhomogeneous empirical measures and functions

Let $\hat{\mu}_N^k(A, a_{\cdot,\cdot}^k) = N^{-1}\sum_{j=1}^N a_{j,N}^k I\{\xi_j^N \in A\}$, $A \in \Re = \mathbf{R}^d$. The estimator $\hat{\mu}_N^k(A, a_{\cdot,\cdot}^k)$ is proposed in [2] as the estimator of the conditional distribution of characteristics $H_k(A) = \mathsf{P}\{\xi_j^N \in A/\operatorname{ind}(j) = k\}$ in the class k. It can be found as a function that minimizes the functional $J(a_{\cdot,\cdot}^k) = \sup \mathsf{E}_{A,H_k}\left|\hat{\mu}_N^k(A, a_{\cdot,\cdot}^k) - H_k(A)\right|^2$ in the class of all weight vectors $a_{\cdot,\cdot}^k$ for which $\hat{\mu}_N^k(A, a_{\cdot,\cdot}^k)$ is an unbiased estimator of $H_k(A)$.

Theorem 4.1. *Let $K(x)$ be a measurable function on $\Re = \mathbf{R}^d$ such that*

$$\sup_{x\in\Re}|K(x)| \le \overline{K} < \infty.$$

Consider the collection $\Im_K$ of sets A of the form $A = \{x\colon K(x) < c\}$ for all possible c. Then

$$\sup_{x\in\Re}\left|\hat{f}_N^k(x) - f_k(x)\right| \le 2\overline{K}\sup_{A\in F_K}\left|\hat{\mu}_N^k(A, a_{\cdot,\cdot}^k) - H_k(A)\right|.$$

Proof. It follows from the definition of Lebesgue integral that

$$f_k(x) = \lim_{n\to\infty}\sum_{j=-n}^{n}\frac{j\overline{K}}{n}H_k\left(A_j^n/A_{j-1}^n\right)$$

where $A_j^n = \left\{x\colon K(x) < (j/n)\overline{K}\right\}$. Similarly

$$\hat{f}_N^k(x) = \lim_{n\to\infty}\sum_{j=-n}^{n}\frac{j\overline{K}}{n}\hat{\mu}_N^k\left(A_j^n/A_{j-1}^n, a_{\cdot,\cdot}^k\right).$$

Then

$$\begin{aligned}&\left|\hat{f}_N^k(x) - f_k(x)\right| \\ &\quad\le \lim_{n\to\infty}\sum_{j=-n}^{n}\frac{j\overline{K}}{n}\left(\left|\hat{\mu}_N^k\left(A_j^n, a_{\cdot,\cdot}^k\right) - H_k\left(A_j^n\right)\right| + \left|\hat{\mu}_N^k\left(A_{j-1}^n, a_{\cdot,\cdot}^k\right) - H_k\left(A_{j-1}^n\right)\right|\right) \\ &\quad\le 2\overline{K}\sup_{A\in F_K}\left|\hat{\mu}_N^k\left(A, a_{\cdot,\cdot}^k\right) - H_k(A)\right|. \qquad \square\end{aligned}$$

Corollary 4.1. *If the assumptions of Theorem 4.1 hold, $\Im_K$ is a Vapnik–Chervonenkis class, and condition b9) is satisfied, then*

$$\sup_{x\in\Re}\left|\hat{f}_N^k(x) - f_k(x)\right| \le \Lambda\sqrt{\frac{\ln N}{N}}$$

where Λ is a random variable such that $\Lambda < \infty$ almost surely.

Proof. This is a straightforward corollary of Theorem 4.1 and Theorem 2.4.1 in [6]. ☐

Corollary 4.2. *If the assumptions of Theorem 4.1 hold, $\Re = [a, b]$, the kernel $K(x)$ is a function of bounded variation, and condition b9) is satisfied, then Corollary 4.1 holds.*

Proof. Since $K(x)$ is a function of bounded variation, we get $K(x) = K_1(x) - K_2(x)$ where $K_1(x)$ and $K_2(x)$ are increasing functions such that $|K_j(x)| \le 2\overline{K}$. Thus it is necessary to consider the case of an increasing kernel $K(x)$. The assumptions of Corollary 4.1 hold in this case, since the collection of intervals $[a, c)$ is a Vapnik–Chervonenkis class and the sets of the collection $\Im_K$ are intervals $[a, c)$ if K_1 and K_2 are monotone. ☐

Theorem 4.2. *Assume that conditions b1)–b4), b6), and b9) hold. Then there is a number N_0 such that for all $N > N_0$ the equation $\lambda_1 \hat{f}_N^1(x) = \lambda_2 \hat{f}_N^2(x)$ almost surely has a unique solution x_N, and moreover $x_N \to x_0$ as $N \to \infty$.*

Proof. First we prove that any sequence of elements x_N of the sets of solutions of the equations $\lambda_1 \hat{f}_N^1(x) = \lambda_2 \hat{f}_N^2(x)$ almost surely converges to x_0 as $N \to \infty$.

Let $v(x) := |\lambda_1 f_1(x) - \lambda_2 f_2(x)|$. Condition b1) implies that there exists a unique x_0 such that $v(x_0) = 0$, and moreover $x_0 = \min_x v(x)$.

Let $v_N(x) := \left|\lambda_1 \hat{f}_N^1(x) - \lambda_2 \hat{f}_N^2(x)\right|$. Note that $\{v_N(x)\}$ is a sequence of random functions.

We apply Theorem 4.3 to $v(x)$ and $\{v_N(x)\}$. Conditions b6) and b9) imply Corollary 4.2, whence condition c1) follows. Condition c2) follows explicitly from b1). ☐

Theorem 4.3. *Assume that*

c1) *there exists a nonrandom function $v(x)$, $x \in \Re = [a, b] \subset \mathbf{R}$, such that*

$$\sup_{x \in \Re} |v_N(x) - v(x)| \to 0$$

as $N \to \infty$;

c2) *there is a unique point $x_0 \in \Re$ such that $v(x_0) < v(x)$ for $x \ne x_0$.*

Put $A_N = \{x\colon v_N(x) = \min_z v_N(z)\}$. Then any sequence $\{x_N\}$ such that $x_N \in A_N$ for all $N \ge 1$ converges almost surely to x_0 as $N \to \infty$.

Proof. Assume that there is a sequence $x'_N \in \{x_N\}$ that does not converge to x_0. Then there is a subsequence $\{N_k\}$ such that the limit $x' = \lim_{N\to\infty} x'_{N_k}$ exists and $x' \ne x_0$. By construction

$$v_N(x'_{N_k}) \le v_N(x), \qquad x \in \Re. \tag{10}$$

Since $v(x)$ is continuous, we obtain from c1) that

$$\sup_{x \in \Re} \left|v_N(x'_{N_k}) - v(x')\right| \to 0$$

as $N \to \infty$. Passing to the limit in (10) as $N \to \infty$ we obtain

$$v(x') \le v(x_0)$$

contradicting condition c2).

It follows from Theorem 4.2 that all the solutions of the equation $\lambda_1 \hat{f}_N^1(x) = \lambda_2 \hat{f}_N^2(x)$ approach x_0 as $N \to \infty$. Now we prove that the solution is unique, indeed. Condition b4) implies that there exists $R > 0$ such that

$$v'(x) \ne 0$$

for all $x \in B(x_0, R)$. Consider an arbitrary number ε such that $0 < \varepsilon < R$. Assume that, in the ε-neighborhood of x_0 and for all N, there exist solutions x'_N and x''_N of the equation $v_N(x) = 0$. Then, by the Rolle theorem there is $c \in B(x_0, \varepsilon)$ such that $v'_N(c) = 0$. It follows from conditions b6) and b9) and Corollary 4.2 that $\sup_{x\in\Re}|v'_N(x)-v'(x)|\to 0$ as $N\to\infty$. This contradiction proves the theorem. □

5. Proof of the main result

It follows from Theorem 4.2 that, starting from some N, the equation

$$\lambda_1\hat f^1_N(x)=\lambda_2\hat f^2_N(x)$$

almost surely has a unique solution x_N. Condition b1) implies that the errors of classifiers (1) and (3) can be represented as follows:

$$L^*=\lambda_2\int_{-\infty}^{x_0}f_2(x)\,dx+\lambda_1\int_{x_0}^{\infty}f_1(x)\,dx,\qquad L_N=\lambda_2\int_{-\infty}^{x_N}f_2(x)\,dx+\lambda_1\int_{x_N}^{\infty}f_1(x)\,dx.$$

This implies

$$L_N-L^*=\lambda_1\int_{x_N}^{x_0}f_1(x)\,dx-\lambda_2\int_{x_N}^{x_0}f_2(x)\,dx=\int_{x_N}^{x_0}[\lambda_1f_1(x)-\lambda_2f_2(x)]\,dx.$$

Put $\Delta_N=x_0-x_N$. Then

$$\begin{aligned}L_N-L^*&=\int_{x_0-\Delta_N}^{x_0}[\lambda_1f_1(x)-\lambda_2f_2(x)]\,dx\qquad |t=x-x_0|\\&=\int_{-\Delta_N}^{0}[\lambda_1f_1(x_0+t)-\lambda_2f_2(x_0+t)]\,dt.\end{aligned}$$

Now we expand the function $g(x):=\lambda_1f_1(x)-\lambda_2f_2(x)$ into the Taylor series and let $N\to\infty$. According to Theorem 4.2, $\Delta_N\to 0$ as $N\to\infty$. Thus

$$\begin{aligned}&\int_{-\Delta_N}^{0}\big(\lambda_1f_1(x_0)-\lambda_2f_2(x_0)+t(\lambda_1f_1'(x_0)-\lambda_2f_2'(x_0))+o(t)\big)\,dt\\&=\int_{-\Delta_N}^{0}t(\lambda_1f_1'(x_0)-\lambda_2f_2'(x_0))\,dt+o(\Delta_N^2)\\&=\frac{\Delta_N^2}{2}\,|\lambda_2f_2'(x_0)-\lambda_1f_1'(x_0)|+o(\Delta_N^2).\end{aligned}\tag{11}$$

Therefore the problem on the rate of convergence of L_N-L^* is reduced to the same problem for Δ_N as $N\to\infty$.

Let $g_N(x):=\lambda_1\hat f^1_N(x)-\lambda_2\hat f^2_N(x)$ and $g_N(x_N)=0$. Since $x_N=x_0-\Delta_N$, we get

$$\begin{aligned}0=g_N(x_0-\Delta_N)&=g_N(x_0)-\Delta_Ng_N'(x_0),\qquad \Delta_N\to 0,\\ \Delta_N=\frac{g_N(x_0)}{g_N'(x_0)}&=\frac{\lambda_2\hat f^2_N(x_0)-\lambda_1\hat f^1_N(x_0)}{[\lambda_2\hat f^2_N(x_0)-\lambda_1\hat f^1_N(x_0)]'_x}\\&=\frac{\lambda_2(\hat f^2_N(x_0)-f_2(x_0))-\lambda_1(\hat f^1_N(x_0)-f_1(x_0))}{\phi_1(x_0)}.\end{aligned}$$

We used condition b1) in the latter equality.

Lemma 2 of [5] implies for $h_N=c/N^{1/5}$ that

$$\begin{aligned}&N^{2/5}\left[\lambda_2\left(\hat f^2_N(x_0)-f_2(x_0)\right)-\lambda_1\left(\hat f^1_N(x_0)-f_1(x_0)\right)\right]\\&\Rightarrow\left[\frac{D^2c^{2/5}}{2}\phi_2(x_0)+\frac{d}{c^{1/10}}\sigma(x_0)\varsigma\right]\end{aligned}$$

where ς is a standard random variable,

$$\sigma^2(x) = \lim_{N\to\infty} \sum_{r=1}^{M} \left\langle \left(\sum_{l=1}^{M} \lambda_l a_{\cdot}^l \right)^2, (w_{\cdot}^r) \right\rangle_N \cdot f_r(x).$$

This implies that $N^{2/5}\Delta_N \Rightarrow [D^2 c^{2/5} \frac{\phi_2(x_0)}{2\phi_1(x_0)} + \frac{d}{c^{1/10}} \frac{\sigma(x_0)}{\phi_1(x_0)} \varsigma]$ and

$$N^{4/5}(L_N - L^*) \Rightarrow \left[D^2 c^{2/5} \frac{\phi_2(x_0)}{2\sqrt{2\phi_1(x_0)}} + \frac{d}{c^{1/10}} \frac{\sigma(x_0)}{\sqrt{2\phi_1(x_0)}} \varsigma \right]^2 .$$

Bibliography

1. O. V. Sugakova, *Asymptotics of a kernel estimate for distribution density constructed from observations of a mixture with varying concentrations*, Teor. Imovirnost. ta Matem. Statist. **59** (1998), 156–166; English transl. in Theor. Probability Math. Statist. **59** (1999), 161–171. MR1793776
2. R. E. Maĭboroda, *Estimators of components of a mixture with varying concentrations*, Ukrain. Mat. Zh. **48** (1996), no. 4, 562–566; English transl. in Ukrainian Math. J. **48** (1996) no. 4, 618–622. MR1417019 (97j:62055)
3. L. Devroye and L. Györfi, *Nonparametric Density Estimation. The L_1-View*, Wiley, New York, 1985. MR0780746 (86i:62065)
4. Yu. O. Ivan'ko and R. E. Maĭboroda, *Exponential estimates for the empirical Bayes risk in the classification of a mixture with varying concentrations*, Ukrain. Mat. Zh. **54** (2002), no. 10, 1421–1428; English transl. in Ukrainian Math. J. **54** (2002), no. 10, 1722–1731. MR2015493 (2004g:62022)
5. Yu. O. Ivan'ko, *Asymptotic behavior of kernel estimators of densities and their derivatives constructed from samples with varying concentrations*, Visnyk Kyiv. Univer. **9** (2003), 29–35. (Ukrainian)
6. R. E. Maĭboroda, *Statistical Analysis of Mixtures*, Kyiv University, Kyiv, 2003. (Ukrainian)
7. A. A. Borovkov, *Probability Theory*, "Nauka", Moscow, 1986; English transl., Gordon and Breach, Amsterdam, 1998. MR0891184 (88c:60001); MR1711261 (2000f:60001)
8. V. V. Anisimov and A. V. Seilhamer, *Asymptotic properties of extreme sets of random fields*, Teor. Imovirnost. ta Matem. Statist. **51** (1994), pp. 30–38; English transl. in Theor. Probability Math. Statist. **51** (1995), 29–38. MR1445049 (97m:60070)

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

E-mail address: ivanko@lemma-insur.com.ua

Received 4/APR/2003

Translated by OLEG KLESOV