ISSN 2347-1921
Volume 12 Number 05
Journal of Advances in Mathematics

The asymptotic behaviour of threshold-based classification rules in case of three prescribed classes

Oksana Kubaychuk
Department of Computer Science, National University of Food Technologies, Kyiv, Ukraine
kubaychuk@gmail.com

ABSTRACT

We consider the problem of the classification of an object from the observation after its numerical characteristic in case of three prescribed classes. We also study a problem on finding and asymptotic behaviour of threshold-based classification rules constructed from a sample from a mixture with varying concentrations.

Keywords

classification rule; mixture with varying concentrations; estimator; threshold.

Academic Discipline And Sub-Disciplines

Mathematics, Probability and Statistics

SUBJECT CLASSIFICATION

62G05, 62G20, 60G15

1. INTRODUCTION

Object classification by its numerical characteristic is an important theoretical problem and has practical significance, for example, the definition of a person as “not healthy”, if the temperature of its body exceeds 37°C. To solve this problem we consider the threshold-based rule. According to this rule, an object is classified to belong to the first class if its characteristic does not exceed a threshold 37°C; otherwise, an object is classified to belong to the second class. The empirical Bayes classification (EBC) (Devroye and Giorfi, 1985; Ivan’ko and Maiboroda, 2002) and minimization of the empirical risk (ERM) (Vapnik, 1989; Vapnik, 1996) are widely used methods to estimate the best threshold. The case when the learning sample is obtained from a mixture with varying concentrations is considered in (Ivan’ko and Maiboroda, 2006).

However, it is often necessary to classify an object in case of more than one threshold, for example, the definition of a person as «not healthy», if the temperature of its body exceeds 37°C or lower then 36°C. Another example: the person is sick, if the level of its haemoglobin exceeds 84 units or lower than 72 units. In particular, this problem is discussed in (Kubaychuk, 2008; Kubaychuk, 2010).

In all previous examples we have only two prescribed classes. The case of two thresholds and three prescribed classes deserves special attention. An example is the classification of the disease stages. Thus, during the diagnosis of breast cancer a tumor marker CA 15-3 is used. If the value is less than 22 IU/ml, then the person is healthy; if its level is in the range from 22 to 30 IU/ml –precancerous conditions can be diagnosed; if the index is above 30 IU/ml – patient has cancer. When solving some technical problems it is needed to consider the substance in its various aggregate forms: gaseous, liquid, solid. The transition from state to state occurs at a specific temperature. According to this, a boiling point and a melting point are used.

2. THE SETTING OF THE PROBLEM

The problem of the classification of an object O from the observation after its numerical characteristic $\xi=\xi(O)$ is studied. We assume that the object may belong to one of the three prescribed classes. An unknown number of a class containing O is denoted by $ind(O)$. A classification rule (briefly, classifier) is a function $g:\mathbb{R}\to\{1,2,3\}$ that assigns a value to $ind(O)$ by using characteristic ξ. In general, classification rule is defined as a general measurable function, but we restrict the consideration in this paper to the so-called threshold-based classification rules of the six forms

$$g^1_{t_1,t_2}(\xi)=\begin{cases}1, & \xi<t_1,\\ 2, & t_1\le\xi\le t_2,\\ 3, & \xi>t_2,\end{cases},\quad g^2_{t_1,t_2}(\xi)=\begin{cases}2, & \xi<t_1,\\ 1, & t_1\le\xi\le t_2,\\ 3, & \xi>t_2,\end{cases},\quad g^3_{t_1,t_2}(\xi)=\begin{cases}1, & \xi<t_1,\\ 3, & t_1\le\xi\le t_2,\\ 2, & \xi>t_2,\end{cases},$$

$$g^4_{t_1,t_2}(\xi)=\begin{cases}3, & \xi<t_1,\\ 2, & t_1\le\xi\le t_2,\\ 1, & \xi>t_2,\end{cases},\quad g^5_{t_1,t_2}(\xi)=\begin{cases}3, & \xi<t_1,\\ 1, & t_1\le\xi\le t_2,\\ 2, & \xi>t_2,\end{cases},\quad g^6_{t_1,t_2}(\xi)=\begin{cases}2, & \xi<t_1,\\ 3, & t_1\le\xi\le t_2,\\ 1, & \xi>t_2.\end{cases}$$

The a priori probabilities $p_i = P(ind(O)=i)$, $i=\overline{1,3}$ are assumed to be known. The characteristic ξ is assumed to be random, and its distribution depends on $ind(O): P(\xi(O)<x \mid ind(O)=i) = H_i(x)$, $i=\overline{1,3}$. The distributions H_i are unknown, but they have continuous densities h_i with respect to the Lebesgue measure. The family of classifiers is denoted by $G=\{g_{\mathbf{t}} : \mathbf{t}\in\mathbb{R}^2\}$. The probability of error of such a classification rules are given by

$$L(g^1_{\mathbf{t}}) = L^1(\mathbf{t}) = L^1(t_1,t_2) = P\{g^1_{\mathbf{t}}(\xi(O)) \ne ind(O)\} =$$

$$=(p_2+p_3)H_1(t_1)-(p_1+p_3)H_2(t_1)+(p_3+p_1)H_2(t_2)-(p_2+p_1)H_3(t_2)+p_2+p_1.$$

Analogically,

$$L(g^4_{\mathbf{t}})=(p_1+p_2)H_3(t_1)-(p_1+p_3)H_2(t_1)+(p_1+p_3)H_2(t_2)-(p_2+p_3)H_1(t_2)+p_2+p_3.$$

Furthermore, $L^1(t_2,t_1) = -L^4(t_1,t_2)+2p_2+p_3+p_1$.

Further, similarly

$$L(g^2_{\mathbf{t}})=(p_1+p_3)H_2(t_1)-(p_2+p_3)H_1(t_1)-(p_2+p_1)H_3(t_2)+(p_2+p_3)H_1(t_2)+p_2+p_1,$$

$$L(g^5_{\mathbf{t}})=(p_1+p_2)H_3(t_1)-(p_3+p_2)H_1(t_1)-(p_3+p_1)H_2(t_2)+(p_2+p_3)H_1(t_2)+p_3+p_1,$$

$$L^2(t_2,t_1) = -L^5(t_1,t_2)+2p_1+p_3+p_2,$$

$$L(g^3_{\mathbf{t}})=(p_2+p_3)H_1(t_1)-(p_3+p_1)H_2(t_2)+(p_2+p_1)(H_3(t_2)-H_3(t_1))+p_3+p_1,$$

$$L(g^6_{\mathbf{t}})=-(p_2+p_3)H_1(t_2)+(p_1+p_3)H_2(t_1)+(p_1+p_2)(H_3(t_2)-H_3(t_1))+p_2+p_3,$$

$$L^3(t_2,t_1) = -L^6(t_1,t_2)+2p_3+p_1+p_2.$$

A classification rule $g^B\in G$ is called a Bayes classification rule in the class G, if $L(g)$ attains its minimum at g^B ($g^B = \arg\min_{g\in G} L(g_{\mathbf{t}})$). The threshold $\mathbf{t}^B$ for a Bayes classification rule is called the Bayes threshold:

$$\mathbf{t}^B = \arg\min_{\mathbf{t}\in\mathbb{R}^2} L(\mathbf{t}) \qquad (1)$$

For $L^i_{\mathbf{t}}$, $i=\overline{1,6}$ we have: $\mathbf{t}^{iB} = \arg\min_{t_1\in\mathbb{R},t_2\in\mathbb{R}} L^i(t_1,t_2) = (\arg\min_{t_1\in\mathbb{R}} L^i_1(t_1), \arg\min_{t_2\in\mathbb{R}} L^i_2(t_2))$, and

$$L^1_1(t_1)=(p_2+p_3)H_1(t_1)-(p_1+p_3)H_2(t_1),$$

$$L^1_2(t_2)=(p_3+p_1)H_2(t_2)-(p_2+p_1)H_3(t_2)+p_1+p_2,$$

$$L^2_1(t_1)=-(p_2+p_3)H_1(t_1)+(p_1+p_3)H_2(t_1),$$

$$L^2_2(t_2)=(p_3+p_2)H_1(t_2)-(p_2+p_1)H_3(t_2)+p_1+p_2,$$

$$L^3_1(t_1)=-(p_1+p_2)H_3(t_1)+(p_3+p_2)H_1(t_1),$$

$$L_2^3(t_2) = -(p_3 + p_1)H_2(t_2) + (p_2 + p_1)H_3(t_2) + p_1 + p_3,$$

$$L_1^4(t_1) = -(p_1 + p_3)H_2(t_1) + (p_1 + p_2)H_3(t_1),$$

$$L_2^4(t_2) = (p_3 + p_1)H_2(t_2) - (p_2 + p_3)H_1(t_2) + p_3 + p_2,$$

$$L_1^5(t_1) = (p_1 + p_2)H_3(t_1) - (p_3 + p_2)H_1(t_1),$$

$$L_2^5(t_2) = -(p_3 + p_1)H_2(t_2) + (p_2 + p_3)H_1(t_2) + p_1 + p_3,$$

$$L_1^6(t_1) = (p_1 + p_3)H_2(t_1) - (p_1 + p_2)H_3(t_1),$$

$$L_2^6(t_2) = -(p_3 + p_2)H_1(t_2) + (p_2 + p_1)H_3(t_2) + p_3 + p_2.$$

Let us consider the threshold rule $g^1_{t_1,t_2}$. The functions H_i (and, hence h_i) are unknown. One can estimate these functions from the data $\{\xi_{j:N}\}_{j=1}^N$, being a sample from a mixture with varying concentrations, where $\xi_{j:N}$ are independent, if N is fixed and $P\{\xi_{j:N} < x\} = w^1_{j:N}H_1(x) + w^2_{j:N}H_2(x) + w^3_{j:N}H_3(x)$. Here $w^i_{j:N}, i = \overline{1,3}$ is a known concentration in the mixture of objects of the i-th class at the moment when an observation j is made (Maiboroda, 2003), $\sum_{i=1}^3 w^i_{j:N} = 1$.

To estimate the distribution function H_i, empirical distribution function

$$\hat{H}_i^N(x) = \frac{1}{N}\sum_{j=1}^N a^i_{j:N}\mathbf{I}\{\xi_j < x\}$$

is used, where $\mathbf{I}\{A\}$ is the indicator of an event A and $a^i_{j:N}$ are known weight coefficients (Maiboroda, 2003; Sugakova, 1998)

$$a^k_{j:N} = \frac{1}{\det\tilde{\mathbf{A}}_N}\sum_{i=1}^3 (-1)^{k+i}\gamma_{ki}w^k_{j:N}$$

defined if $\det\tilde{\mathbf{A}}_N \neq 0$, where $\tilde{\mathbf{A}}_N = \left(\left\langle w^k, w^l\right\rangle\right)_{k,l=1}^3$ is the Gramm matrix, where $\left\langle w^k, w^l\right\rangle = \frac{1}{N}\sum_{j=1}^N w^k_{j:N}w^l_{j:N}$ and γ_{ki} is the (k,i) main minor of $\tilde{\mathbf{A}}_N$.

One can apply kernel estimators to estimate the densities of distributions h_i:

$$\hat{h}_i^N(x) = \frac{1}{Nk_N}\sum_{j=1}^N a^i_{j:N}K\left(\frac{x - \xi_{j:N}}{k_N}\right),$$

where K is a kernel (the density of some probability distribution), $k_N > 0$ is a smoothing parameter (Sugakova, 1998; Ivan'ko, 2003).

Let us construct the threshold estimator using EBC method (Kubaychuk, 2008). The empirical Bayes estimator is constructed as follows. First, one determines the sets T_{N_1} and T_{N_2} of all solutions of the equations

$$(p_2 + p_3)\hat{h}_1^N(t) - (p_1 + p_3)\hat{h}_2^N(t) = 0, \text{ and } (p_1 + p_3)\hat{h}_2^N(t) - (p_1 + p_2)\hat{h}_3^N(t) = 0$$

respectively. Second, one chooses

$$\hat{\mathbf{t}}^{EBC} = \operatorname*{arg\,min}_{t_1\in T_{N_1}, t_2\in T_{N_2}, t_1<t_2} L_N^1(t_1,t_2),$$

as an estimator for $\mathbf{t}^B$, where

$$L_N^1(t_1,t_2)=$$
$$=(p_2+p_3)\hat{H}_1^N(t_1)-(p_1+p_3)\hat{H}_2^N(t_1)+(p_3+p_1)\hat{H}_2^N(t_2)-(p_2+p_1)\hat{H}_2^N(t_2)+p_1+p_2$$

and $L_N^1(t_1,t_2)$ is the estimator for $L^1(t_1,t_2)$:

$$L_{N_1}^1(t_1)=(p_2+p_3)\hat{H}_1^N(t_1)-(p_1+p_3)\hat{H}_2^N(t_1),$$

$$L_{N_2}^1(t_2)=(p_3+p_1)\hat{H}_2^N(t_2)-(p_2+p_1)\hat{H}_2^N(t_2)+p_1+p_2,$$

$$\hat{t}_{N_1}^{EBC}=\arg\min_{t_1\in T_{N_1}} L_{N_1}^1(t_1),\ \hat{t}_{N_2}^{EBC}=\arg\min_{t_2\in T_{N_2}} L_{N_2}^1(t_2).$$

The sets T_{N_1} and T_{N_2} are constructed under condition $t_1<t_2$.

Let the densities h_i exist and be s times continuously differentiable in some neighborhood of the points t_1^B, t_2^B. Put

$$f_s^2(t)=(-1)^s\left((p_3+p_1)\frac{d^s h_2}{dt^s}-(p_2+p_3)\frac{d^s h_1}{dt^s}\right),$$

$$f_s^2(t)=(-1)^s\left((p_2+p_1)\frac{d^s h_1}{dt^s}-(p_3+p_1)\frac{d^s h_2}{dt^s}\right).$$

Let's assume, $\lim_{N\to\infty} r_{N_i}=r_i$, $i=1,2$ exist. Put

$$r_{N_i}=\left[N^{-1}\sum_{j=1}^{N}\left(b_{j:N}^i\right)^2\left[w_{j:N}^1 h_1\left(t_i^B\right)+w_{j:N}^2 h_2\left(t_i^B\right)+w_{j:N}^3 h_3\left(t_i^B\right)\right]\right]^{\frac{1}{2}},\ i=1,2,\ \text{where}$$

$$b_{j:N}^1=(p_2+p_3)a_{j:N}^1-(p_1+p_3)a_{j:N}^2,\ b_{j:N}^2=(p_1+p_3)a_{j:N}^2-(p_1+p_2)a_{j:N}^3.$$

Let's denote

$$W_{N_i}^1(\tau)=N^{\frac{2}{3}}\left(L_{N_i}^1\left(t_i^B+N^{-\frac{1}{3}}\tau\right)-L_{N_i}^1\left(t_i^B\right)-L_i^1\left(t_i^B+N^{-\frac{1}{3}}\tau\right)+L_i^1\left(t_i^B\right)\right),\ i=1,2.$$

3. MAIN RESULTS

In what follows we assume that:

(A) the threshold $\mathbf{t}^B$ defined by (1) exists and it is the unique point of the global minimum for $L^1(\mathbf{t})$ (t_1^B is the unique global minimum point for $L_1^1(t_1)$, t_2^B is the unique global minimum point for $L_2^1(t_2)$).

(B) The limits $\lim_{N\to\infty}\inf\det\mathbf{\Gamma}_N>c>0$ exist; $\lim_{N\to\infty}\sum_{r=1}^{M}\left\langle (a^k)^2,w^r\right\rangle_N\cdot h_r(x)$, $1\le k\le M$, $M=3$.

Remark 1. Condition (B) is sufficient for $\lim_{N\to\infty} r_{N_i}=r_i$, $i=1,2$ existence.

Theorem 1. Let conditions (A) and (B) hold. Assume that the densities h_i exist and are continuous, $k_N\to 0$ as $N_{k_N}\to\infty$, k is the continuous function, and

$$d^2 \overset{def}{=} \int_{-\infty}^{\infty} k^2(t)dt < \infty .$$

Then $\hat{\mathbf{t}}_N^{EBC} \to \mathbf{t}^B (\hat{t}_{N_1}^{EBC} \to t_1^B, \hat{t}_{N_2}^{EBC} \to t_2^B)$ in probability.

Proof. According to Theorem 1 of (Sugakova, 1998), the assumptions of the theorem imply that $\hat{h}_i^N(x) \to h_i(x)$ in probability at every point $x \in \mathbb{R}$. Therefore

$$u_{N_1}(x) = ((p_2 + p_3)\hat{h}_1^N(x) - (p_1 + p_3)\hat{h}_2^N(x)) \to u_1(x) = ((p_2 + p_3)h_1(x) - (p_1 + p_3)h_2(x)),$$

$$u_{N_2}(x) = ((p_1 + p_3)\hat{h}_2^N(x) - (p_1 + p_2)\hat{h}_3^N(x)) \to u_2(x) = ((p_1 + p_3)h_2(x) - (p_1 + p_2)h_3(x))$$

in probability.

Put $A_N(\delta_i) = \{\text{there exists } t_i : |t_i - t_i^B| \le \delta_i, u_{N_i}(t_i) = 0\}$ for $\delta_i > 0$. We can show that

$$P(A_N(\delta_i)) \to 1,\ N \to \infty \qquad (2)$$

Since t_1^B is the point of minimum $L_1^1(t)$, t_2^B is the point of minimum $L_2^1(t)$, $(L_1^1(t))' = u_{N_1}(t)$ and $(L_2^1(t))' = u_{N_2}(t)$ are continuous functions, it follows that $u_{N_i}(t)$ changes sign in the neighborhood of t_i^B. This means that there are t_i^- and t_i^+ such that

$$t_i^B - \delta_i < t_i^- < t_i^B < t_i^+ < t_i^B + \delta_i$$

and $u_i(t_i^-)u_i(t_i^+) < 0,\ i = 1,2$. Thus, $P(u_i(t_i^-)u_i(t_i^+) < 0) \to 1,\ N \to \infty$. Since $u_{N_i}(t)$ are continuous functions, $\{u_i(t_i^-)u_i(t_i^+) < 0\} \subseteq A_N(\delta_i)$. Therefore (2) is proved.

Let us fix an δ_i, $i = 1,2$. Hence, $L_1^1(t)$ and $L_2^1(t)$ are continuous functions on $\mathbb{R}$, $L_1^1(-\infty) = 0$, $L_1^1(+\infty) = p_2 - p_1$, $L_2^1(-\infty) = p_2 + p_1$, $L_2^1(+\infty) = p_3 + p_1$ and condition (A) is satisfied, then $\forall \delta_i > 0\ \exists \varepsilon_i\ \forall t_i : |t_i - t_i^B| > \delta_i$ it follows that $L_i^1(t_i) > L_i^1(t_i^B) + \varepsilon_i$. Let $0 < \delta_i' < \delta_i$ be such that $\forall t \in [t_i^B - \delta_i', t_i^B + \delta_i']$: $L_i^1(t_i) < L_i^1(t_i^B) + \varepsilon_i/4$. Put

$$B_{N_i} = \{ \inf_{t \notin [t_i^B - \delta_i', t_i^B + \delta_i']} L_{N_i}^1(t) > L_i^1(t_i^B) + \varepsilon_i/2 > \inf_{t \in [t_i^B - \delta_i', t_i^B + \delta_i']} L_{N_i}^1(t)\}.$$

Fix an arbitrary $\lambda_i > 0$. Using the uniform convergence $L_{N_i}^1$ to L_i^1, we obtain for sufficiently large N that $P(B_{N_i}) > 1 - \lambda_i/2$. From (2) it follows $P(A_N(\delta_i)) > 1 - \lambda_i/2$ for sufficiently large N. If the event $A_N(\delta_i')$ occurs, then there exists $t_i^* \in T_{N_i} \cap [t_i^B - \delta_i', t_i^B + \delta_i']$ such that $L_{N_i}^1(t_i^*) < L_{N_i}^1(t_i)$ for all $t_i \notin [t_i^B - \delta_i, t_i^B + \delta_i]$ given the event B_{N_i} occurs. Therefore, hence

$$P(A_N(\delta_i') \cap B_{N_i}) \ge P(A_N(\delta_i')) + P(B_{N_i}) - 1$$

it follows that

$$P\{|\hat{t}_i^{EBC} - t_i^B| < \delta\} \ge 1 - \lambda_i$$

for sufficiently large N. This completes the proof of the theorem, since λ_i, $i = 1,2$ is arbitrary.

Remark 2. The estimator H_k (obtained by construction) is unbiased iff

$$\left\langle a^k w^m \right\rangle_N = \mathbf{I}\left\{m=k\right\}$$, for all $m=1,\dots,M$, $N>M$.

Then, it is easy to see that $\left\langle a \right\rangle_N = 1$.

Remark 3. Often, $\hat{H}_k$ is not a probability distribution, but it is not important. To estimate H_k you can use the corrected weighted empirical distribution function, if necessary. (Kubaychuk, 2003;Maiboroda and Kubaichuk, 2003; Maiboroda and Kubaichuk, 2004).

For the proof next theorem we need some auxiliary result on the asymptotic behavior of the processes $W^1_{N_i}$, $i=1,2$.

Lemma 1. Let condition (A) hold and $\tau_1 < \tau_2$. Put

$$A_N = A_N(\tau_1,\tau_2) = [N^{-1/3}\tau_1, N^{-1/3}\tau_2].$$

Then

$$W^1_{N_1}(\tau_2) - W^1_{N_1}(\tau_1) = N^{-1/3}\sum_{j=1}^{N} b^1_{j:N}(\mathbf{I}\{\xi_{j:N}\in A_N\} - P\{\xi_{j:N}\in A_N\}),$$

$$W^1_{N_2}(\tau_2) - W^1_{N_2}(\tau_1) = N^{-1/3}\sum_{j=1}^{N} b^2_{j:N}(\mathbf{I}\{\xi_{j:N}\in A_N\} - P\{\xi_{j:N}\in A_N\}).$$

Proof.

$$W^1_{N_1}(\tau_2) - W^1_{N_1}(\tau_1) =$$
$$= N^{2/3}[L^1_{N_1}(t^B_1 + N^{-1/3}\tau_2) - L^1_1(t^B_1 + N^{-1/3}\tau_2) - L^1_{N_1}(t^B_1 + N^{-1/3}\tau_1) + L^1_1(t^B_1 + N^{-1/3}\tau_1)] =$$

$$= N^{2/3}[p_2\hat{H}^N_1(t^B_1 + N^{-1/3}\tau_2) - p_1\hat{H}^N_2(t^B_1 + N^{-1/3}\tau_2) +$$
$$+ p_3(\hat{H}^N_1(t^B_1 + N^{-1/3}\tau_2) - \hat{H}^N_2(t^B_1 + N^{-1/3}\tau_2)) - p_2H_1(t^B_1 + N^{-1/3}\tau_2) + p_1H_2(t^B_1 + N^{-1/3}\tau_2) -$$
$$- p_3(H_1(t^B_1 + N^{-1/3}\tau_2) - H_2(t^B_1 + N^{-1/3}\tau_2)) - p_2\hat{H}^N_1(t^B_1 + N^{-1/3}\tau_1) + p_1\hat{H}^N_2(t^B_1 + N^{-1/3}\tau_1) -$$
$$- p_3(\hat{H}^N_1(t^B_1 + N^{-1/3}\tau_1) - \hat{H}^N_2(t^B_1 + N^{-1/3}\tau_1)) + p_2H_1(t^B_1 + N^{-1/3}\tau_1) - p_1H_2(t^B_1 + N^{-1/3}\tau_1) +$$
$$+ p_3(H_1(t^B_1 + N^{-1/3}\tau_1) - H_2(t^B_1 + N^{-1/3}\tau_1))] =$$

$$= N^{-1/3}\sum_{j=1}^{N}[p_2a^1_{j:N}\mathbf{I}\{\xi_{j:N}\in A_N\} - p_1a^2_{j:N}\mathbf{I}\{\xi_{j:N}\in A_N\} +$$
$$+ p_3(a^1_{j:N}\mathbf{I}\{\xi_{j:N}\in A_N\} - a^2_{j:N}\mathbf{I}\{\xi_{j:N}\in A_N\}) + p_1a^2_{j:N}P\{\xi_{j:N}\in A_N\} - p_2a^1_{j:N}P\{\xi_{j:N}\in A_N\} -$$
$$- p_3(a^1_{j:N}P\{\xi_{j:N}\in A_N\} - a^2_{j:N}P\{\xi_{j:N}\in A_N\})] =$$

$$= N^{-1/3}\sum\nolimits_{j=1}^{N}[(p_2a^1_{j:N} - p_1a^2_{j:N}) + p_3(a^1_{j:N} - a^2_{j:N})]\cdot$$
$$\cdot\left(\mathbf{I}\left\{\xi_{j:N}\in A_N\right\} - P\left\{\xi_{j:N}\in A_N\right\}\right) = N^{-1/3}\sum\nolimits_{j=1}^{N} b^1_{j:N}(\mathbf{I}\{\xi_{j:N}\in A_N\} - P\{\xi_{j:N}\in A_N\}).$$

$$W^1_{N_2}(\tau_2) - W^1_{N_2}(\tau_1) =$$
$$= N^{2/3}[L^1_{N_2}(t^B_2 + N^{-1/3}\tau_2) - L^1_2(t^B_2 + N^{-1/3}\tau_2) - L^1_{N_2}(t^B_2 + N^{-1/3}\tau_1) + L^1_2(t^B_2 + N^{-1/3}\tau_1)] =$$

$$= N^{2/3}[(p_3+p_1)\hat{H}_2^N(t_2^B+N^{-1/3}\tau_2)-(p_2+p_1)\hat{H}_3^N(t_2^B+N^{-1/3}\tau_2)-$$
$$-(p_3+p_1)\hat{H}_2^N(t_2^B+N^{-1/3}\tau_1)+(p_2+p_1)\hat{H}_3^N(t_2^B+N^{-1/3}\tau_1)-$$
$$-(p_3+p_1)H_2(t_2^B+N^{-1/3}\tau_2)+(p_2+p_1)H_3(t_2^B+N^{-1/3}\tau_2)+$$
$$+(p_3+p_1)H_2(t_2^B+N^{-1/3}\tau_1)-(p_2+p_1)H_3(t_2^B+N^{-1/3}\tau_1)]=$$
$$= N^{-1/3}\sum_{j=1}^N[(p_3+p_1)a_{j:N}^2-(p_2+p_1)a_{j:N}^3]\cdot$$
$$\cdot(\mathbf{I}\{\xi_{j:N}\in A_N\}-P\{\xi_{j:N}\in A_N\}) = N^{-1/3}\sum_{j=1}^N b_{j:N}^2(\mathbf{I}\{\xi_{j:N}\in A_N\}-P\{\xi_{j:N}\in A_N\}).$$

This completes the proof of the lemma.

In what follows, the symbol $\Rightarrow$ stands for weak convergence.

Theorem 2. Let $D(u_i)$ be the space of functions without discontinuity of the second kind equipped with the uniform metric, W be the two sided standard Wiener process, (B) holds. Then, stochastic processes $W_{N_i}^1,\ i=1,2$ weakly converge as $N\to\infty$ to the process r_iW in the space $D(u_i)$ on an arbitrary finite interval $u_i=[\tau_{i-},\tau_{i+}]$.

Proof.

The trajectories of $W_{N_i}^1,\ i=1,2$ are continuous. It is enough to prove: the finite dimensional distributions of $W_{N_i}^1,\ i=1,2$ are asymptotically Gaussian, the second moments of increments converge and the distributions of $W_{N_i}^1,\ i=1,2$ are tight in $D(u_i)$. See (Billingsley, 1968).

We first compute $\mathbf{E}(W_{N_i}^1(\tau_2)-W_{N_i}^1(\tau_1))^2,\ i=1,2$. Let $\tau_1<\tau_2$, by using Lemma 1:

$$\mathbf{E}(W_{N_2}^1(\tau_2)-W_{N_2}^1(\tau_1))^2 = N^{-2/3}\sum_{j=1}^N(b_{j:N}^2)^2\mathbf{E}(\mathbf{I}\{\xi_{j:N}\in A_N\}-P\{\xi_{j:N}\in A_N\})^2=$$
$$= N^{-2/3}\sum_{j=1}^N(b_{j:N}^2)^2[(w_{j:N}^1H_1(A_N)+w_{j:N}^2H_2(A_N)\times$$
$$\times w_{j:N}^3H_3(A_N))-(w_{j:N}^1H_1(A_N)+w_{j:N}^2H_2(A_N)+w_{j:N}^3H_3(A_N))^2].$$
$$\mathbf{E}(W_{N_1}^1(\tau_2)-W_{N_1}^1(\tau_1))^2 = N^{-2/3}\sum_{j=1}^N(b_{j:N}^1)^2\mathbf{E}(\mathbf{I}\{\xi_{j:N}\in A_N\}-P\{\xi_{j:N}\in A_N\})^2=$$
$$= N^{-2/3}\sum_{j=1}^N(b_{j:N}^1)^2[(w_{j:N}^1H_1(A_N)+w_{j:N}^2H_2(A_N)\times$$
$$\times w_{j:N}^3H_3(A_N))-(w_{j:N}^1H_1(A_N)+w_{j:N}^2H_2(A_N)+w_{j:N}^3H_3(A_N))^2].$$

Taking into account that $H_i(A_N) \square\ h_i(t_1^B)N^{-1/3}(\tau_2-\tau_1),\ i=1,2,3$, we obtain:

$$\mathbf{E}(W_{N_1}^1(\tau_2)-W_{N_1}^1(\tau_1))^2=$$
$$= N^{-2/3}\sum_{j=1}^N N^{-1/3}(b_{j:N}^1)^2[w_{j:N}^1h_1(t_1^B)+w_{j:N}^2h_2(t_1^B)+w_{j:N}^3h_3(t_1^B)](\tau_2-\tau_1)\cdot$$
$$\cdot[1-N^{-1/3}(w_{j:N}^1h_1(t_1^B)+w_{j:N}^2h_2(t_1^B)+w_{j:N}^3h_3(t_1^B))(\tau_2-\tau_1)]\to$$
$$\to r_1^2(\tau_2-\tau_1)=\mathbf{E}(r_1W(\tau_2)-r_1W(\tau_1))^2 \text{ as } N\to\infty,$$

where

$$r_1=\lim_{N\to\infty}r_{N_1},\ r_{N_1}=[N^{-1}\sum_{j=1}^N(b_{j:N}^1)^2[w_{j:N}^1h_1(t_1^B)+w_{j:N}^2h_2(t_1^B)+w_{j:N}^3h_3(t_1^B)]]^{1/2}$$

Similarly, taking into account that $H_i(A_N) \approx h_i(t_2^B)N^{-1/3}(\tau_2-\tau_1), i=1,2,3$, we obtain $\mathbf{E}(W_{N_2}^1(\tau_2)-W_{N_2}^1(\tau_1))^2 \to r_2^2(\tau_2-\tau_1)=\mathbf{E}(r_2W(\tau_2)-r_2W(\tau_1))^2$ as $N\to\infty$,

where

$$r_2=\lim_{N\to\infty} r_{N_2},\ r_{N_2}=[N^{-1}\sum_{j=1}^N (b_{j:N}^2)^2[w_{j:N}^1 h_1(t_1^B)+w_{j:N}^2 h_2(t_1^B)+w_{j:N}^3 h_3(t_1^B)]]^{1/2}$$

The condition (B) holds, than all terms at sum from lemma 1 are uniformly bounded. Therefore, the finite dimensional distributions of processes $W_{N_i}^1,\ i=1,2$ are asymptotically Gaussian in view of the central limit theorem under the Lindeberg condition. The tightness of family of distributions $W_{N_i}^1,\ i=1,2$ is proving analogically to (Ivan'ko and Maiboroda, 2006). This completes the proof.

Theorem 3. Let conditions (A) and (B) hold. Assume that:

(i) the derivatives $h_k''(t)=d^2h_k(t)/dt^2$ exist and are bounded in a neighborhood of t_1^B,t_2^B and $f(t_i^B)\neq 0$, $i=1,2$;

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

(iii) $k_N=c/N^{1/5}$ for some nonrandom $c>0$.

Then $N^{2/5}(\hat{t}_i^{EBC}-t_i^B)\Rightarrow A_i+B_i\eta_i$, where

$$A_i=D^2c^{2/5}f_2^i(t_i^B)/(2f_1^i(t_i^B)),\ B_i=dr_i/(c^{1/10}f_1^i(t_i^B)),$$

and η_i is a standard Gaussian random variable, $i=1,2$.

Proof. Let

$$u_{N_1}(t)=(p_2+p_3)\hat{h}_1^N(t)-(p_1+p_3)\hat{h}_2^N(t),$$

$$u_{N_2}(t)=(p_1+p_3)\hat{h}_1^N(t)-(p_1+p_2)\hat{h}_3^N(t).$$

By the definition of $t_{N_i}^{EBC}$ we have $u_{N_i}(t_{N_i}^{EBC})=0$. Put $\delta_{N_i}=t_{N_i}^{EBC}-t_i^B$, $i=1,2$. Theorem 1 implies that $\delta_{N_i}\to 0$ in probability. Hence

$$\delta_{N_1}=-\frac{u_{N_1}(t_1^B)}{u_{N_1}'(t_1^B)}\approx-\frac{(p_2+p_3)(\hat{h}_1^N(t_1^B)-h_1(t_1^B))-(p_1+p_3)(\hat{h}_2^N(t_1^B)-h_2(t_1^B))}{f_1^1(t_1^B)},$$

$$\delta_{N_2}=-\frac{u_{N_2}(t_2^B)}{u_{N_2}'(t_2^B)}\approx-\frac{(p_1+p_3)(\hat{h}_2^N(t_2^B)-h_2(t_2^B))-(p_1+p_2)(\hat{h}_3^N(t_2^B)-h_3(t_2^B))}{f_1^2(t_2^B)}.$$

Similarly to the proof of Lemma 2 of (Ivan'ko, 2003), we obtain

$$N^{2/5}(-[(p_2+p_3)(\hat{h}_1^N(t_1^B)-h_1(t_1^B))-(p_1+p_3)(\hat{h}_2^N(t_1^B)-h_2(t_1^B))])$$
$$\Rightarrow D^2c^{2/5}f_2^1(t_1^B)/2+(dr_1/c^{1/10})\eta_1,$$

$$N^{2/5}(-[(p_1+p_3)(\hat{h}_2^N(t_2^B)-h_2(t_2^B))-(p_1+p_2)(\hat{h}_3^N(t_2^B)-h_3(t_2^B))])$$
$$\Rightarrow D^2c^{2/5}f_2^2(t_2^B)/2+(dr_2/c^{1/10})\eta_2$$

For $k_N=c/N^{1/5}$, where η_i is a standard Gaussian random variable, $i=1,2$.

This completes the proof.

SUMMARY AND CONCLUSIONS

The results obtained in this paper allow one to see the asymptotic behaviour of threshold-based classification rules constructed from a sample from a mixture with varying concentrations in case of three prescribed classes. This is another important step to solving the problem of the classification of an object from the observation after its numerical characteristic. Future research will be devoted to the situation with an arbitrary number of classes.

ACKNOWLEDGEMENTS

The author would like to thank the referees for their valuable comments.

REFERENCES

[1] Billingsley, P. (1968): Convergence of Probability Measures. New York: John Wiley & Sons, Inc.

[2] Devroye, L. and Gyorfi, L. (1985): Nonparametric Density Estimation. The L1 View. New York: John Wiley &Sons, Inc.

[3] Ivan'ko, Yu. O. (2003): The asymptotic behavior of kernel estimators and their derivatives constructed from observations from a mixture with varying concentrations. Visnyk KNU, Ser. Matematika.Mekhanika, 9, 29–35. (Ukrainian).

[4] Ivan'ko, Yu.O. and Maiboroda, R.E. (2002): Exponential estimates for the empirical Bayes risk in the classification of a mixture with varying concentrations. Ukrain. Mat. Zh., 54, no. 10, 1421–1428; English transl. in Ukrainian Math. J., 54, no. 10, 1722–1731.

[5] Ivan'ko, Yu.O. and Maiboroda, R.E. (2006): The asymptotic behaviour of threshold-based classification rules constructed from a sample from a mixture with varying concentrations. Teor. Imovirnost. Matem. Statist., 74, 34–43; English transl. in Theor. Probability Math. Statist., 74, 37–47.

[6] Kubaychuk, O.O. (2003): Estimation of moments from mixtures using the corrected weighted empirical distribution functions. Visnyk KNU, Ser. Matematika. Mekhanika, 9, 48–52. (Ukrainian).

[7] Kubaychuk, O.O. (2008): The asymptotic behavior estimator of Bayes threshold. Visnyk KNU, Ser. Matematika.Mekhanika, 19, 47–50. (Ukrainian).

[8] Kubaychuk, O.O. (2010): The estimator asymptotic behavior of the empirical risk minimization method for Bayesian border. Research Bulletin NTU KPI Ser. Physics and Mathematics, 4, 78–85. (Ukrainian).

[9] Maiboroda, R. and Kubaichuk, O. (2003): Asymptotic normality of improved weighted empirical distribution functions. Teor. Imovirnost. Matem. Statist., 69, 89–95; English transl. in Theor. Probability Math. Statist., 69, 95–102.

[10] Maiboroda, R. and Kubaichuk, O. (2004): Improved estimators for moments constructed from observations of a mixture. Teor. Imovirnost. Matem. Statist., 70, 74–81; English transl. in Theor. Probability Math. Statist., 70, 83–92.

[11] Maiboroda, R. E. (2003): Statistical Analysis of Mixtures. Kyiv: "Kyiv University" (Ukrainian).

[12] Sugakova, O. V. (1998): Asymptotics of a kernel estimate for distribution density constructed fromobservations of a mixture with varying concentrations. Teor. Imovirnost. Matem. Statist., 59, 156–166; English transl. in Theor. Probability Math. Statist., 59, 161–171.

[13] Vapnik, V. N. (1996): The Nature of Statistical Learning Theory. New York: Springer.

[14] Vapnik, V.N. (1989): Inductive principles for the search for empirical laws. Pattern. Recognition. Classification. Prediction, 1, 17–81. (Russian)