

# EBC-Estimator of Multidimensional Bayesian Threshold in Case of Two Classes

Oksana Kubaychuk\*<sup>✉</sup>

*Institute of Physics and Technology, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», 37, Prosp. Peremohy, Kyiv, 03056, Ukraine*

## ARTICLE INFO

### Article History

Received 24 Apr 2020

Accepted 27 Jul 2020

### Keywords

Estimator

Multidimensional Bayesian

threshold

Mixture with varying concentrations

### 2000 Mathematics Subject

Classification: 62G05, 62G30.

## ABSTRACT

Some threshold-based classification rules in case of two classes are defined. In assumption, that a learning sample is obtained from a mixture with varying concentration, the empirical-Bayesian classification (EBC)-estimator of multidimensional Bayesian threshold is constructed. The conditions of convergence in probability of estimator are found.

© 2020 The Authors. Published by Atlantis Press B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

## 1. INTRODUCTION

The model of a mixture of several probability distributions was mentioned for the first time by Newcomb [1] and Pearson [2]. Such mixtures naturally arise in many areas. In particular, in the theory of reliability and time of life, mixtures of gamma distributions [3] are used. Examples of the use of mixtures of normal distributions in the processing of biological and physiological data are given in [4]. In Slud [5], a mixture of two exponential distributions is used to describe the debugging process of the software. Some applications of the model of mixtures in medical diagnostics were given in [6,7].

The technique of a nonparametric analysis of mixtures where concentrations changes from observation to observation develops, actively. The problem of distributions estimating in case at known concentrations is considered in the works of Maiboroda [8,9]. Estimates of concentrations in two-component mixtures in work [10]. Works by Sugakova [11] and Ivanko [12] are devoted to the evaluation of component distribution densities. The correction algorithms for weighted empirical distribution functions are proposed in [13].

For the theoretical study of problems of nonparametric regression the nonhomogeneous weighted empirical distribution functions used by Stoune [14]. These were applied by Maiboroda in the tasks of analyzing the mixture. In particular, in Maiboroda [8] found conditions under which the weighted empirical distribution functions are unbiased and minimal estimators of unknown distribution functions of components of the mixture.

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 of type

$$g_t(\xi) = \begin{cases} 1, & \xi \leq t, \\ 2, & \xi > t. \end{cases}$$

According to this rule, an object is classified to the first class if its characteristic does not exceed a threshold  $t = 37^\circ\text{C}$ ; otherwise, an object is classified to belong to the second class. The empirical-Bayesian classification (EBC) [15,16] and minimization of the empirical risk (MER) [17,18] 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 [19] and the asymptotic of both methods of estimating is investigated.

\*Email: [o.kubaichuk@kpi.ua](mailto:o.kubaichuk@kpi.ua)

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 than 36°C. Another example: the person is sick, if the level of its hemoglobin exceeds 84 units or lower than 72 units. Accordingly, one can apply the classifiers of type

$$g_{t_1, t_2}^1(\xi) = \begin{cases} 1, & \xi \in [t_1, t_2], \\ 2, & \xi \notin [t_1, t_2] \end{cases}$$

or

$$g_{t_1, t_2}^2(\xi) = \begin{cases} 1, & \xi \notin [t_1, t_2], \\ 2, & \xi \in [t_1, t_2]. \end{cases}$$

In particular, this problem is discussed in [20,21].

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. Accordingly, 6 classifiers of forms

$$g_{t_1, t_2}(\xi) = \begin{cases} 1, & \xi < t_1, \\ 2, & t_1 \leq \xi \leq t_2, \\ 3, & \xi > t_2 \end{cases}$$

can be applied. This partial case was studied in [22].

## 2. SETTING OF THE PROBLEM

The problem of classification of an object  $O$  from the observation of its numerical characteristic  $\xi = \xi(O)$  is studied. We assume that the object may belong to one of two prescribed classes. An unknown class number containing  $O$  is denoted by  $ind(O)$ . A classification rule (briefly, classifier) is a function  $g: \mathbb{R}^n \rightarrow \{1, 2\}$  that assigns a value to  $ind(O)$  by using characteristic  $\xi$ . 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 forms

$$g_{2m-1, \mathbf{t}}^1(\xi) = \begin{cases} 1, & \xi \in \left\{ \bigcup_{k=1}^{m-1, m>1} [t_{2k}, t_{2k+1}] \right\} \cup \{(-\infty, t_1]\}, \\ 2, & \xi \in \left\{ \bigcup_{k=1}^{m-1, m>1} (t_{2k-1}, t_{2k}) \right\} \cup \{(t_{2m-1}, +\infty)\}, \end{cases}$$

$$g_{2m-1, \mathbf{t}}^2(\xi) = \begin{cases} 1, & \xi \in \left\{ \bigcup_{k=1}^{m-1, m>1} (t_{2k-1}, t_{2k}) \right\} \cup \{(t_{2m-1}, +\infty)\}, \\ 2, & \xi \in \left\{ \bigcup_{k=1}^{m-1, m>1} [t_{2k}, t_{2k+1}] \right\} \cup \{(-\infty, t_1)\}, \end{cases}$$

if  $n = 2m - 1$ ,  $m \in \mathbb{N}$  and

$$g_{2m, \mathbf{t}}^1(\xi) = \begin{cases} 1, & \xi \in \left\{ \bigcup_{k=1}^{m-1, m>1} (t_{2k}, t_{2k+1}) \right\} \cup \{(-\infty, t_1)\} \cup \{(t_{2m}, +\infty)\}, \\ 2, & \xi \in \left\{ \bigcup_{k=1}^m [t_{2k-1}, t_{2k}] \right\}, \end{cases}$$

$$g_{2m, \mathbf{t}}^2(\xi) = \begin{cases} 1, & \xi \in \left\{ \bigcup_{k=1}^m [t_{2k-1}, t_{2k}] \right\}, \\ 2, & \xi \in \left\{ \bigcup_{k=1}^{m-1, m>1} (t_{2k}, t_{2k+1}) \right\} \cup \{(-\infty, t_1)\} \cup \{(t_{2m}, +\infty)\}, \end{cases}$$

if  $n = 2m$ ,  $m \in \mathbb{N}$ , where  $\mathbf{t} = (t_1, t_2, \dots, t_n)$  is the multidimensional threshold.

The a priori probabilities  $p_i = P(\text{ind}(O) = i)$ ,  $i = 1, 2$  are assumed to be known. The characteristic  $\xi$  is assumed to be random, and its distribution depends on  $\text{ind}(O)$ :  $P(\xi(O) < x | \text{ind}(O) = i) = H_i(x)$ ,  $i = 1, 2$ . 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  $G = \{g_t : t \in \mathbb{R}^n\}$ .

Let,  $n = 2m - 1$ ,  $m \in \mathbb{N}$  then the probability of error of such a classification rules are given by

$$\begin{aligned} L(g_{2m-1,t}^1) &= L_{2m-1}^1(t) = P\{g_t^1(\xi(O)) \neq \text{ind}(O)\} = \sum_{i=1}^2 P\{\text{ind}(O) = i\} P\{g_t^1(\xi(O)) = 3 - i | \text{ind}(O) = i\} \\ &= p_1 \sum_{k=1}^m (H_1(t_{2k}) - H_1(t_{2k-1})) + p_1 (1 - H_1(t_{2m+1})) + p_2 \sum_{k=1}^m (H_2(t_{2k+1}) - H_2(t_{2k})) + p_2 H_2(t_1) \end{aligned}$$

$$\begin{aligned} L(g_{2m-1,t}^2) &= L_{2m-1}^2(t) = P\{g_t^2(\xi(O)) \neq \text{ind}(O)\} = \sum_{i=1}^2 P\{\text{ind}(O) = i\} P\{g_t^2(\xi(O)) = 3 - i | \text{ind}(O) = i\} \\ &= p_1 \sum_{k=1}^m (H_1(t_{2k+1}) - H_1(t_{2k})) + p_1 H_1(t_1) + p_2 \sum_{k=1}^m (H_2(t_{2k}) - H_2(t_{2k-1})) + p_2 (1 - H_2(t_{2m+1})). \end{aligned}$$

Analogically, for  $n = 2m$ ,  $m \in \mathbb{N}$ :

$$\begin{aligned} L(g_{2m,t}^1) &= L_{2m}^1(t) = P\{g_{2m,t}^1(\xi(O)) \neq \text{ind}(O)\} = \sum_{i=1}^2 P\{\text{ind}(O) = i\} P\{g_{2m,t}^1(\xi(O)) = 3 - i | \text{ind}(O) = i\} \\ &= p_1 \sum_{k=1}^m (H_1(t_{2k}) - H_1(t_{2k-1})) + p_2 (1 - H_2(t_{2m})) + p_2 \sum_{k=1}^{m-1, m>1} (H_2(t_{2k+1}) - H_2(t_{2k})) + p_2 H_2(t_1) \end{aligned}$$

$$\begin{aligned} L(g_{2m,t}^2) &= L_{2m}^2(t) = P\{g_{2m,t}^2(\xi(O)) \neq \text{ind}(O)\} = \sum_{i=1}^2 P\{\text{ind}(O) = i\} P\{g_{2m,t}^2(\xi(O)) = 3 - i | \text{ind}(O) = i\} \\ &= p_2 \sum_{k=1}^m (H_2(t_{2k}) - H_2(t_{2k-1})) + p_1 H_1(t_1) + p_1 \sum_{k=1}^m (H_1(t_{2k+1}) - H_1(t_{2k})) + p_1 (1 - H_1(t_{2m})) \end{aligned}$$

A classification rule  $g^B \in G$  is called a Bayesian classification rule in the class  $G$ , if  $L(g)$  attains its minimum at  $g^B$  ( $g^B = \arg \min_{g \in G} L(g_t)$ ).

The threshold  $t^B$  for a Bayesian classification rule is called the Bayesian threshold:

$$t^B = \arg \min_{t \in \mathbb{R}^n} L(t) \quad (1)$$

Let,

$$t_i = \arg \min_{t \in \mathbb{R}} L_{i,2m-1}^1(t), \quad t_1 < t_2 < \dots < t_{2m-1}$$

Denote Bayesian threshold for classifier  $g_{2m-1,t}^1$ :

$$t_{2m-1}^{1B} = \arg \min_{t \in \mathbb{R}^{2m-1}} L_{2m-1}^1(t) = (t_1, t_2, \dots, t_{2m-1}),$$

where

$$L_{i,2m-1}^1(t) = (-1)^i (p_1 H_1(t) - p_2 H_2(t)), \quad i = 1, \dots, 2m - 2, \quad m \in \mathbb{N}, \text{ and}$$

$$L_{2m-1,2m-1}^1(t) = p_1 - (p_1 H_1(t) - p_2 H_2(t)).$$

Analogically, let

$$t_i = \arg \min_{t \in \mathbb{R}} L_{i,2m-1}^2(t), \quad t_1 < t_2 < \dots < t_{2m-1}$$

For  $g_{2m-1,t}^2$ :

$$\mathbf{t}_{2m-1}^{2B} = \arg \min_{\mathbf{t} \in \mathbb{R}^{2m-1}} L_{2m-1}^2(\mathbf{t}) = (t_1, t_2, \dots, t_{2m-1}),$$

where

$$L_{i,2m-1}^2(t) = (-1)^i (-p_1 H_1(t) + p_2 H_2(t)), \quad i = 1, \dots, 2m-2, \quad m \in \mathbb{N}, \text{ and}$$

$$L_{2m-1,2m-1}^2(t) = p_2 + (p_1 H_1(t) - p_2 H_2(t)).$$

Let,

$$t_i = \arg \min_{t \in \mathbb{R}} L_{i,2m}^1(t), \quad t_1 < t_2 < \dots < t_{2m}$$

For  $g_{2m,t}^1$ :

$$\mathbf{t}_{2m}^{1B} = \arg \min_{\mathbf{t} \in \mathbb{R}^{2m}} L_{2m}^1(\mathbf{t}) = (t_1, t_2, \dots, t_{2m}),$$

where

$$L_{i,2m}^1(t) = (-1)^i (p_1 H_1(t) - p_2 H_2(t)), \quad i = 1, \dots, 2m-1, \quad m \in \mathbb{N}, \text{ and}$$

$$L_{2m,2m}^1(t_2) = p_2 + (p_1 H_1(t_{2m}) - p_2 H_2(t_{2m}))$$

Let,

$$t_i = \arg \min_{t \in \mathbb{R}} L_{i,2m}^2(t), \quad t_1 < t_2 < \dots < t_{2m}$$

For  $g_{2m,t}^2$ :

$$\mathbf{t}_{2m}^{2B} = \arg \min_{\mathbf{t} \in \mathbb{R}^{2m}} L_{2m}^2(\mathbf{t}) = (t_1, t_2, \dots, t_{2m}),$$

where

$$L_{i,2m}^2(t) = (-1)^i (-p_1 H_1(t) + p_2 H_2(t)), \quad i = 1, \dots, 2m-1, \quad m \in \mathbb{N}$$

$$L_{2m,2m}^2(t) = p_1 - (p_1 H_1(t) - p_2 H_2(t)).$$

Denote

$$L_n^1(\mathbf{t}) = L_{1,n}^1(t_1) + L_{2,n}^1(t_2) + \dots + L_{n,n}^1(t_n),$$

$$L_n^2(\mathbf{t}) = L_{1,n}^2(t_1) + L_{2,n}^2(t_2) + \dots + L_{n,n}^2(t_n),$$

where  $n = 2m-1$ ,  $m \in \mathbb{N}$  or  $n = 2m$ ,  $m \in \mathbb{N}$ .

When determining the best threshold, one faces the problem of estimating the threshold by using a learning sample, whose members are classified correctly. We consider the Bayesian empirical classification method, in assumption, that a learning sample is obtained from a mixture with varying concentration.

The distribution functions  $H_s$  (and, of course, densities  $h_s$ ) are assumed to be unknown. One can estimate these functions from the data  $\Xi_N = \{\xi_{j:N}\}_{j=1}^N$  being a sample from a mixture with varying concentration, where  $\xi_{j:N}$  are independent if  $N$  is fixed and

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

Here  $w_{j:N}$  is a known concentration in the mixture of objects of the first class at the moment when an observation  $j$  is made [23].

To estimate the distribution functions  $H_s$ , we use weighted empirical distribution functions

$$\widehat{H}_s^N(x) = \frac{1}{N} \sum_{j=1}^N a_{j:N}^s 1\{\xi_j < x\} \quad (2)$$

where  $1\{A\}$  is the indicator an event  $A$  and  $a_{j:N}^s$  are known weight coefficients:

$$a_{j:N}^1 = \frac{1}{\Delta_N} \left( (1 - S_N^1) w_{j:N} + (S_N^2 - S_N^1) \right), \quad a_{j:N}^2 = \frac{1}{\Delta_N} (S_N^2 - S_N^1 w_{j:N}),$$

$$S_N^k = \frac{1}{N} \sum_{j=1}^N (w_{j:N})^k, \quad k = 1, 2, \quad \Delta_N = S_N^2 - (S_N^1)^2$$

(see [23]).

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

$$\widehat{h}_s^N(x) = \frac{1}{Nk_N} \sum_{j=1}^N a_{j:N}^s K\left(\frac{x - \xi_{j:N}}{k_N}\right),$$

where  $K$  is a kernel (the density of some probability distribution) and  $k_N$  is a smoothing parameter [11,24].

The empirical-Bayesian estimator is constructed as follows. At first, one determines the set  $T_N$  of all solutions of the equation

$$\widehat{p}_1 h_1(t) - \widehat{p}_2 h_2(t) = 0, \quad t_1 < t_2 < \dots < t_n, \quad (3)$$

where  $n = 2m - 1$ ,  $m \in \mathbb{N}$  or  $n = 2m$ ,  $m \in \mathbb{N}$  for every  $N$ .

Second, one chooses

$$\widehat{\mathbf{t}}_{N,n}^{1EBC} = \left( \arg \min_{t \in T_N} L_{N,1,n}^1(t), \arg \min_{t \in T_N} L_{N,2,n}^1(t), \dots, \arg \min_{t \in T_N} L_{N,n,n}^1(t) \right)$$

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

$$L_{N,i,2m-1}^1(t) = (-1)^i (p_1 H_1^N(t) - p_2 H_2^N(t)), \quad i = 1, \dots, 2m - 2, \quad m \in \mathbb{N},$$

$$L_{N,2m-1,2m-1}^1(t) = p_1 - (p_1 H_1^N(t) - p_2 H_2^N(t));$$

$$L_{N,i,2m}^1(t) = (-1)^i (p_1 H_1^N(t) - p_2 H_2^N(t)), \quad i = 1, \dots, 2m - 1, \quad m \in \mathbb{N},$$

$$L_{N,2m,2m}^1(t) = p_2 + (p_1 H_1^N(t) - p_2 H_2^N(t)),$$

$$\widehat{t}_{N,i,n}^{1EBC} = \arg \min_{t \in T_N} L_{N,i,n}^{N,i,n}(t), \quad i = 1, \dots, n, \quad n = 2m(n = 2m - 1),$$

therefore

$$\widehat{\mathbf{t}}_{N,n}^{1EBC} = \left( \widehat{t}_{N,1,n}^{1EBC}, \widehat{t}_{N,2,n}^{1EBC}, \dots, \widehat{t}_{N,n,n}^{1EBC} \right);$$

$$L_{N,n}^1(\mathbf{t}) = L_{N,1,n}^1(t_1) + L_{N,2,n}^1(t_2) + \dots + L_{N,n,n}^1(t_n),$$

where  $n = 2m - 1$ ,  $m \in \mathbb{N}$  or  $n = 2m$ ,  $m \in \mathbb{N}$ ,

or

$$\widehat{\mathbf{t}}_{N,n}^{2EBC} = \left( \arg \min_{t \in T_N} L_{N,1,n}^2(t), \arg \min_{t \in T_N} L_{N,2,n}^2(t), \dots, \arg \min_{t \in T_N} L_{N,n,n}^2(t) \right),$$

where

$$L_{N,i,2m-1}^2(t) = (-1)^i (-p_1 H_1^N(t) + p_2 H_2^N(t)), \quad i = 1, \dots, 2m-2, \quad m \in \mathbb{N},$$

$$L_{N,2m-1,2m-1}^2(t) = p_2 + (p_1 H_1^N(t) - p_2 H_2^N(t));$$

$$L_{N,i,2m}^2(t) = (-1)^i (-p_1 H_1^N(t) + p_2 H_2^N(t)), \quad i = 1, \dots, 2m-1, \quad m \in \mathbb{N},$$

$$L_{N,2m,2m}^2(t) = p_1 - (p_1 H_1^N(t) - p_2 H_2^N(t)),$$

$$\hat{t}_{N,i,n}^{1EBC} = \arg \min_{t \in T_N} L_{N,i,n}^1(t), \quad i = 1, \dots, n, \quad n = 2m(n = 2m-1),$$

and

$$\hat{\mathbf{t}}_{N,n}^{1EBC} = \left( \hat{t}_{N,1,n}^{1EBC}, \hat{t}_{N,2,n}^{1EBC}, \dots, \hat{t}_{N,n,n}^{1EBC} \right).$$

$$L_{N,n}^2(\mathbf{t}) = L_{N,1,n}^2(t_1) + L_{N,2,n}^2(t_2) + \dots + L_{N,n,n}^2(t_n),$$

where  $n = 2m-1$ ,  $m \in \mathbb{N}$  or  $n = 2m$ ,  $m \in \mathbb{N}$ .

An example of multidimensional threshold in case of two classes is shown on Figures 1 and 2 (Mathcad v.13 was used).

### 3. MAIN RESULTS

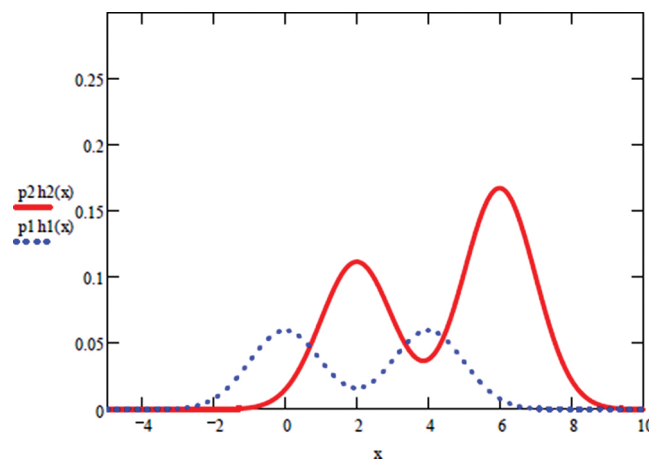
#### 3.1. Choice of Classifier

The choice of the classifier  $g_{2m-1,t}^1(g_{2m,t}^1)$  or  $g_{2m-1,t}^2(g_{2m,t}^2)$  depends on the smallest root of the equation

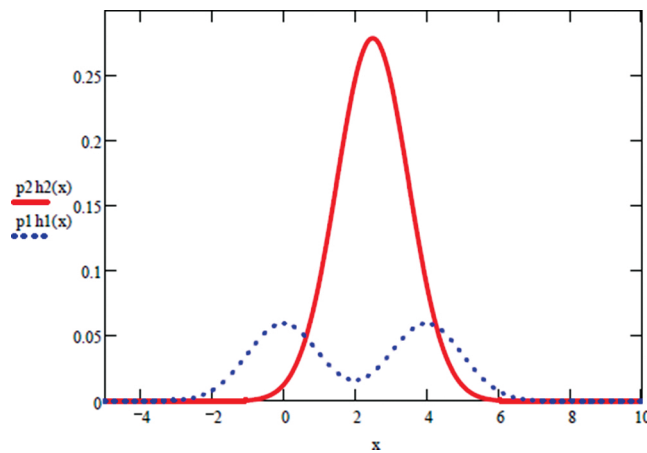
$$p_1 h_1 - p_2 h_2 = 0. \quad (4)$$

**Theorem 3.1.1.** *If root of (4) minimizes  $L_{1,2m-1}^1(L_{1,2m}^1)$ , then the classifier  $g_{2m-1,t}^1(g_{2m,t}^1)$  is selected, but if it minimizes  $L_{1,2m-1}^2(L_{1,2m}^2)$ , then it is selected  $g_{2m-1,t}^2(g_{2m,t}^2)$ .*

**Proof.** The statement follows from the properties  $L_{i,2m-1}^1 = -L_{i,2m-1}^2$ ,  $i = 1, \dots, 2m-2$  and  $L_{i,2m}^1 = -L_{i,2m}^2$ ,  $i = 1, \dots, 2m-1$ .



**Figure 1** | Three-dimensional threshold:  $h_1 = 0.5N(0,1) + 0.5N(4,1)$ ,  $h_2 = 0.4N(2,1) + 0.6N(6,1)$ ,  $p_1 = 0.3$ ,  $p_2 = 0.7$ .



**Figure 2** | Two-dimensional threshold:  $h_1 = 0.5N(0,1) + 0.5N(4,1)$ ,  $h_2 = N(2.5,1)$ ,  $p_1 = 0.3$ ,  $p_2 = 0.7$ .

**Remark 3.1.1.**  $L_{2m-1,2m-1}^1 = -L_{2m-1,2m-1}^2 + p_1 + p_2$  and  $L_{2m,2m}^1 = -L_{2m,2m}^2 + p_1 + p_2$ .

The next theorem can be proved analogically to Theorem 3.1.1.

**Theorem 3.1.2.** If root of (3) minimizes  $L_{N,1,2m-1}^1 \left( L_{N,1,2m}^1 \right)$ , then the  $\widehat{t}_{N,n}^{1EBC}$  is selected, but if it minimizes  $L_{N,1,2m-1}^2 \left( L_{N,1,2m}^2 \right)$ , then it is selected  $\widehat{t}_{N,n}^{2EBC}$ .

**Proof.** The statement follows from the properties  $L_{N,i,2m-1}^1 = -L_{N,i,2m-1}^2$ ,  $i = 1, \dots, 2m-2$  and  $L_{N,i,2m}^1 = -L_{N,i,2m}^2$ ,  $i = 1, \dots, 2m-1$ .

**Remark 3.1.2.**  $L_{N,2m-1,2m-1}^1 = -L_{N,2m-1,2m-1}^2 + p_1 + p_2$  and  $L_{N,2m,2m}^1 = -L_{N,2m,2m}^2 + p_1 + p_2$ .

**Remark 3.1.3.** As follows from [25–28], we can use the improved weighted distribution function if some coefficients are negative in (2).

## 3.2. The Convergence in Probability of EBC-Estimator

In what follows we assume that

(A). The threshold  $t^B \in \mathbb{R}^n$ ,  $n = 2m-1$  ( $n = 2m$ ),  $m \in \mathbb{N}$  exists, is a unique of the global minimum of  $L_n^1(t)$ ,  $n = 2m-1$  ( $n = 2m$ ),  $m \in \mathbb{N}$  or  $L_n^2(t)$ ,  $n = 2m-1$  ( $n = 2m$ ),  $m \in \mathbb{N}$  ( $t_{i,n}^B$  is a global minimum point of  $L_{i,n}^1(t_i)$ ,  $i = 1, \dots, n$ ,  $n = 2m-1$  ( $n = 2m$ ),  $m \in \mathbb{N}$  or  $L_{i,n}^2(t_i)$ ,  $i = 1, \dots, n$ ,  $n = 2m-1$  ( $n = 2m$ ),  $m \in \mathbb{N}$ ).

(Bk). The limits  $S^l = \lim_{n \rightarrow \infty} S_{N,n}^l$ ,  $l = 1, 2, \dots, k$ , exists and  $\Delta = S^2 - (S^1)^2 > 0$ .

**Lemma 3.2.1.** Let conditions (A) and (Bk) hold. Assume that densities  $h_1$  and  $h_2$  exist and are continuous,  $k_N \rightarrow 0$ ,  $N_{k_N} \rightarrow \infty$ ,  $K$  is a continuous function, and

$$d^2 := \int_{-\infty}^{\infty} K^2(t) dt < \infty.$$

Then  $P(A_N(\delta_i)) \rightarrow 1$ ,  $N \rightarrow \infty$ , for  $\delta_i > 0$ , where  $A_N(\delta_i) = \{\exists t_i : |t_i - t_{i,n}^B| \leq \delta_i, u_N(t_i) = 0\}$  and  $u_N(x) = \widehat{p}_2 h_2(x) - \widehat{p}_1 h_1(x)$ ,  $i = 1, \dots, n$ .

**Proof.** According to Theorem 1 of [11], the assumptions of the theorem imply that  $\widehat{h}_s(x) \rightarrow h_s(x)$ ,  $s = 1, 2$  in probability at every point  $x \in \mathbb{R}$ . Therefore,

$$u_N(x) = \widehat{p}_2 \widehat{h}_2(x) - \widehat{p}_1 \widehat{h}_1(x) \rightarrow u(x) = p_2 h_2(x) - p_1 h_1(x)$$

in probability. For  $\delta > 0$ , let

$$A_N(\delta_i) = \{\exists t_i : |t_i - t_{i,n}^B| \leq \delta_i, u_N(t_i) = 0\}.$$

Now we shall that

$$P(A_N(\delta_i)) \rightarrow 1, N \rightarrow \infty. \quad (5)$$

Since,  $t_{i,n}^B$  is a point of minimum of  $L_{i,n}^1(t)$  (or  $L_{i,n}^2(t)$ ),  $i = 1, \dots, n$ ,  $n = 2m - 1$  ( $n = 2m$ ),  $m \in \mathbb{N}$ , and  $L_{i,n}^1(t) = u(t)$  or  $L_{i,n}^1(t) = -u(t)$  (similarly,  $L_{i,n}^2(t)$ ), depending on the parity or the oddness of  $i$ , is a continuous function,  $u(t)$  changes its sign in a neighborhood of the point  $t_{i,n}^B$ . This means that there are  $t_i^-$  and  $t_i^+$  such that

$$t_{i,n}^B - \delta_i < t_i^- < t_{i,n}^B < t_i^+ < t_{i,n}^B + \delta_i, i = 1, \dots, n, n = 2m - 1 (n = 2m), m \in \mathbb{N}$$

and  $u(t_i^-) u(t_i^+) < 0$ .

Thus,  $P(u_N(t_i^-) u_N(t_i^+) < 0) \rightarrow 1$ . Since  $u_N$  is a continuous function,  $\{u_N(t_i^-) u_N(t_i^+) < 0\} \subseteq A_N(\delta_i)$ . Therefore (5) is proved.

**Lemma 3.2.2.** Let assumptions of Lemma 3.2.1 for  $\delta'_i > 0$ ,  $0 < \delta'_i < \delta_i$  hold. Then

$$P(B_{i,N}) \rightarrow 1 \text{ as } N \rightarrow \infty,$$

where

$$B_{i,N} = \left\{ \inf_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^1(t) > L_{i,n}^1(t_{i,n}^B) + \frac{\varepsilon}{2} > \sup_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^1(t) \right\}, i = 1, \dots, n, n \in \mathbb{N}$$

for  $g_{2m-1,t}^1$  (or  $g_{2m,t}^1$ ) and

$$B_{i,N} = \left\{ \inf_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^2(t) > L_{i,n}^2(t_{i,n}^B) + \frac{\varepsilon}{2} > \sup_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^2(t) \right\}, i = 1, \dots, n, n \in \mathbb{N}$$

for  $g_{2m-1,t}^2$  (or  $g_{2m,t}^2$ ).

**Proof.** Fix  $\delta_i$ ,  $i = 1, \dots, n$ ,  $n \in \mathbb{N}$ . Since  $L_{i,n}^1$  ( $L_{i,n}^2$ ) are continuous functions on  $\mathbb{R}$ ,  $i = 1, \dots, n$ ,  $n \in \mathbb{N}$ ,

$$L_{i,2m-1}^1(-\infty) = 0, i = 1, \dots, 2m - 2 \quad (L_{i,2m-1}^2(-\infty) = 0, i = 1, \dots, 2m - 2),$$

$$L_{i,2m}^1(-\infty) = 0, i = 1, \dots, 2m - 1 \quad (L_{i,2m}^2(-\infty) = 0, i = 1, \dots, 2m - 1),$$

$$L_{i,2m-1}^1(+\infty) = (-1)^i (p_1 - p_2), i = 1, \dots, 2m - 2 \quad (L_{i,2m-1}^2(+\infty) = (-1)^i (-p_1 + p_2), i = 1, \dots, 2m - 2),$$

$$L_{i,2m}^1(+\infty) = (-1)^i (p_1 - p_2), i = 1, \dots, 2m - 1 \quad (L_{i,2m}^2(+\infty) = (-1)^i (-p_1 + p_2), i = 1, \dots, 2m - 1),$$

$$L_{2m-1,2m-1}^1(-\infty) = p_1, L_{2m-1,2m-1}^1(+\infty) = p_2, L_{2m,2m}^1(-\infty) = p_2, L_{2m,2m}^1(+\infty) = p_1,$$

$$L_{2m-1,2m-1}^2(-\infty) = p_2, L_{2m-1,2m-1}^2(+\infty) = p_1, L_{2m,2m}^2(-\infty) = p_1, L_{2m,2m}^2(+\infty) = p_2$$

and condition (A) holds. Moreover  $\forall \delta_i > 0 \exists \varepsilon_i$  such that  $L_{i,n}^1(t_i) > L_{i,n}^1(t_{i,n}^B) + \varepsilon_i$ ,  $i = 1, \dots, n$  (or  $L_{i,n}^2(t_i) > L_{i,n}^2(t_{i,n}^B) + \varepsilon_i$ ,  $i = 1, \dots, n$ ) for all  $t_i$  for which  $|t_i - t_{i,n}^B| > \delta_i$ .

Choose  $0 < \delta'_i < \delta_i$  so that

$$L_{i,n}^1(t_i) < L_{i,n}^1(t_{i,n}^B) + \frac{\varepsilon_i}{4} \quad (\text{or } L_{i,n}^2(t_i) < L_{i,n}^2(t_{i,n}^B) + \frac{\varepsilon_i}{4}), i = 1, \dots, n, n \in \mathbb{N}$$



for all  $t_i \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]$ . Denote

$$B_{i,N} = \left\{ \inf_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^1(t) > L_{i,n}^1(t_{i,n}^B) + \frac{\varepsilon}{2} > \sup_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^1(t) \right\}, i = 1, \dots, n, n \in \mathbb{N}$$

for  $g_{2m-1,t}^1(g_{2m,t}^1)$  and

$$B_{i,N} = \left\{ \inf_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^2(t) > L_{i,n}^2(t_{i,n}^B) + \frac{\varepsilon}{2} > \sup_{t \in [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]} L_{N,i,n}^2(t) \right\}, i = 1, \dots, n, n \in \mathbb{N}$$

for  $g_{2m-1,t}^2(g_{2m,t}^2)$ .

Fix an arbitrary  $\lambda_i > 0$ . Using the uniform convergence of  $L_{N,i,n}^1$  to  $L_{i,n}^1$  (or  $L_{N,i,n}^2$  to  $L_{i,n}^2$ ), we obtain the necessary statement for sufficiently large  $N$  for according  $B_{i,N}, p(B_{i,N}) > 1 - \lambda_i/2$ .

**Theorem 3.2.1.** Assume that conditions of Lemma 3.2.1 hold. Then

$\widehat{t}_{N,n}^{1EBC} \rightarrow t^B$  (or  $\widehat{t}_{N,n}^{2EBC} \rightarrow t^B$ ) in probability as  $N \rightarrow \infty$ , namely

$\widehat{t}_{N,i,n}^{1EBC} \rightarrow t_i^B, i = 1, \dots, n, n \in \mathbb{N}$  (or  $\widehat{t}_{N,i,n}^{2EBC} \rightarrow t_i^B, i = 1, \dots, n, n \in \mathbb{N}$ ) in probability as  $N \rightarrow \infty$ .

**Proof.** Since (5)  $P(A_N(\delta'_i)) > 1 - \frac{\lambda_i}{2}$  for sufficiently large  $N$ . If the event  $A_N(\delta'_i)$  occurs, then there exists

$$t_i^* \in T_N \cap [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]$$

such that  $L_{N,i,n}^1(t_i^*) < L_{N,i,n}^1(t_i)$  for all  $t_i \notin [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]$  given the event  $B_{i,N}$  occurs (or  $L_{N,i,n}^2(t_i^*) < L_{N,i,n}^2(t_i)$  for all  $t_i \notin [t_{i,n}^B - \delta'_i, t_{i,n}^B + \delta'_i]$ ). Therefore,

$$P\left\{\widehat{t}_{N,i,n}^{1EBC} - t_{i,n}^B < \delta\right\} \geq P(A_N(\delta'_i) \cap B_{i,N}) \geq 1 - \frac{\lambda_i}{2} + 1 - \frac{\lambda_i}{2} - 1 = 1 - \lambda_i$$

and

$$P\left\{\widehat{t}_{N,i,n}^{2EBC} - t_{i,n}^B < \delta\right\} \geq P(A_N(\delta'_i) \cap B_{i,N}) \geq 1 - \frac{\lambda_i}{2} + 1 - \frac{\lambda_i}{2} - 1 = 1 - \lambda_i$$

given the event  $B_{i,N}$  occurs for sufficiently large  $N, i = 1, \dots, n, n \in \mathbb{N}$ , taking into account that

$$P(A_N(\delta'_i) \cap B_{i,N}) = P(A_N(\delta'_i)) + P(B_{i,N}) - P(A_N(\delta'_i) \cup B_{i,N}) \geq P(A_N(\delta'_i)) + P(B_{i,N}) - 1.$$

This completes the proof of the theorem, since  $\lambda_i, i = 1, \dots, n, n \in \mathbb{N}$  are arbitrary.

## 4. CONCLUSION

In this paper, we found the conditions of convergence in probability of the estimator for the Bayesian threshold constructed by the method of empirical-Bayesian classification for a sample from a mixture with variable concentrations.

## CONFLICT OF INTEREST

The author has no conflicts of interest to declare.

## ACKNOWLEDGMENTS

I thank the reviewers whose insightful comments helped me to improve this paper.

## REFERENCES

1. S. Newcomb, *Am. J. Math.* 8 (1886), 343–366.
2. K. Pearson, *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* 185 (1894), 71–110.
3. W. Nelson, *Applied Life Data Analysis: Nelson/Applied Life Data Analysis*, John Wiley & Sons, Inc., New York, NY, USA, 1982.
4. S.J. Roberts, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (1998), 1133–1142.
5. E. Slud, *Scand. J. Stat.* 24 (1997), 555–572.
6. S.L. Hui, X.H. Zhou, *Stat. Methods Med. Res.* 7 (1998), 354–370.
7. D. Rindskopf, W. Rindskopf, *Stat. Med.* 5 (1986), 21–27.
8. R.E. Maiboroda, *Ukr. Math. J.* 48 (1996), 618–622.
9. R.E. Maiboroda, *Theory Probab. Math. Stat.* 59 (1999), 121–128.
10. R.E. Maiboroda, *Theory Probab. Math. Stat.* 46 (1993), 71–75.
11. O.V. Sugakova, *Theory Probab. Math. Stat.* 59 (1999), 161–171.
12. Yu.O. Ivan'ko, *Visn. Mat. Mekh. Kyiv. Univ. Im. Tarasa Shevchenka.* 9 (2003), 29–35.
13. O. Kubaychuk, *Adv. Appl. Stat.* 54 (2019), 315–326.
14. C.J. Stone, *Ann. Stat.* 5 (1977), 595–620.
15. L. Devroye, L. Györfi, *Nonparametric Density Estimation, The L1 View*, John Wiley & Sons, Inc., New York, NY, USA, 1985.
16. Yu.O. Ivan'ko, R.E. Maiboroda, *Ukr. Math. J.* 54 (2002), 1722–1731.
17. V.N. Vapnik, in: *Pattern Recognition, Classification, Prediction*, Yu.I. Zhuravlev, eds, vol. 1, Nauka, Moscow, Russia, 1989, pp. 17–81.
18. V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1996.
19. Yu.O. Ivan'ko, R.E. Maiboroda, *Theory Probab. Math. Stat.* 74 (2007), 37–47.
20. O. Kubajchuk, *Visn. Mat. Mekh. Kyiv. Univ. Im. Tarasa Shevchenka.* 19 (2008), 47–51.
21. O.O. Kubaychuk, *Res. Bull. NTU Ukr. Kyiv Politechnic Inst.* 4 (2010), 78–85.
22. O. Kubaychuk, *J. Adv. Math.* 12 (2016), 6261–6269.
23. R.E. Maiboroda, *Statistical Analysis of Mixtures, a Course of Lectures*, Kyiv University, Kyiv, Ukraine, 2003.
24. Yu.O. Ivan'ko, *Visnyk KNU Ser. Matematika. Mekhanika.* 9 (2003), 29–35.
25. O. Kubaychuk, *Theory Stoch. Process.* 8 (2002), 226–231.
26. O.O. Kubajchuk, *Visn. Mat. Mekh. Kyiv. Univ. Im. Tarasa Shevchenka.* 9 (2003), 48–52.
27. R. Maiboroda, O. Kubaichuk, *Theory Probab. Math. Stat.* 69 (2004), 95–102.
28. R. Maiboroda, O. Kubaichuk, *Theory Probab. Math. Stat.* 70 (2005), 83–92.