STATISTIQUE ET ANALYSE DES DONNÉES

JAN MIKIEWICZ

Sample density as the function-estimate of population's distribution

Statistique et analyse des données, tome 7, nº 2 (1982), p. 57-70
<http://www.numdam.org/item?id=SAD_1982__7_2_57_0>

Statistiques et Analyse des Données
1982 - Vol. 7 n° 2 pp. 57-70

SAMPLE DENSITY AS THE FUNCTION-ESTIMATE OF POPULATION'S DISTRIBUTION

Jan MIKIEWICZ

Technical University of Wroclaw
ul. St. Wyspianskiego 27
50370 Wroclaw, POLAND

Resumé : *La méthode proposée constitue une certaine variété de la méthode des fonctions potentielles, cependant, en tant que fonction attribuée aux points qui représentent en* R^m *les éléments échantillons la répartition rotative de Cauchy a été appliquée, et au lieu de la somme, le produit normalisé de ces fonctions a été accepté. Parmi les estimateurs non paramétriques de la densité de probabilité le type discuté se caractérise par une simplicité particulière, et il a une interprétation gnoséologique. Un théorème concernant la convergence de cet estimateur a été présenté.*

Abstratc : *The method proposed is a certain type of the potential functions method, in wich the rotational Cauchy distribution is assumed as the function assigned to points, representing sample elements in* R^m *; instead of the sum, the normalized product of, these function is taken into account. Among the nonparametric estimates of probability density function, this is especially simple and has a gnosiological interpretation. A certain theorem concerning the convergence of this estimate is presented.*

Mots clés : *Probability density function. Nonparametric estimate.*

O - INTRODUCTION

In many problems dealt with in Nature's exploration the objects under study appear in the form of composed populations. Each class /i.e. subpopulation/ is a population by itself with its own natural probability distribution.

Such distributions of composed populations on $\mathbb{R}^m$-space are generally of irregular type with many maxima of probability density function. That may be easily shown in terms of the formula describing the density function/referred to as d.f. /of the composed population / $x \in R^m$ / :

$$(1) \qquad f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \ldots + \alpha_k f_k(x),$$

where α_i, for $i = 1,2,\ldots,k$, denote fractions which in the total mass of the population represent classes, and $f_i(x)$ indicate d. f. of the i-th subpopulation.

A case such that the division of the population is not known is especially difficult and important. This situation is pattern recognition problems [1] is called the "recognition without the help of a teacher". There should be distinguished some problems which have not been satisfactorily solved so far. They are the following :

1. Definition of "demarcation lines /hypersurfaces/ between territories in $\mathbb{R}^m$ deviding optimally the masses of representants of particular classes /the well known problem, see e.g. [1],[10]/.

2. Definition of fontiers /contours/ of classes, which is possible when the classes are not very dispersed /particularly important in pattern recognition problems/.

3. Estimation of gravity centres of classes /important in biology/.

All these problems may be also considered in terms of the sequential approach, associated with the search for an optimal sample size.

1 - THE BASIS OF THE SAMPLE DENSITY CONSTRUCTION

The method discussed here enables solving all of these problems. It requires the following assumptions : the component populations are of continuous type, defined on the real space $\mathbb{R}^m$; moreover, d. f. of subpopulations is differentiable Such assumptions are based on many empirical cases ; in some other cases they are admissible when treated approximately.

The well known method of obtaining the "picture of population" is the construction of "empirical distribution function", see [6]. There are some objections to this method. Let us consider them here :

1. Most of the populations considered in empirical sciences are of a continuous type. It follows that this estimate is an insufficient "picture" to represent these populations, especially for small samples.

2. The distribution function is usually a worse gnosiological tool for empirical investigations that the d. f.

3. In multidimensional cases the e.d.f. is defined, but practically it is useless. Already in a twodimensional case, in order to get acquainted with e.d.f., it is necessary to have a spatial model which gives to the empiric investigator not more than a spatial set of sample points. With a larger number of dimensions, a spatial model corresponding to intuition cannot be constructed and as an analytic instrument for discrimination and pattern recognition, the e.d.f. is of low use.

From an other side there is a method of potential functions /see e.g. [1]/. These are sums of functions belonging to a very large class of functions, assigned to particular points of the considered /e.g. sample/ set. We may consider the class of continuous functions and then the continuity of a functional estimator is ensured. However, when the form of the function attributed to points is established, the functional estimator as their sum is not consistent and this can be shown easily. By the above mentioned consistence we understand the following property of the functional estimate $\varphi(x|S_n)$ based on a random sample S_n of n elements :

(2) $$\forall x \in \mathbb{R}^m \ \forall \varepsilon > 0 \ \lim_{n\to\infty} P\{|\varphi(x|S_n) - f(x)| < \varepsilon\} = 1,$$

where φ is an estimate of d.f. f(x) of population from which the sample was taken. In order to eliminate this defect of potential functions, a so called "kernels method" was used /due to [9] and [5]/, in which the potential functions $\Psi(x,x_i,h)$, of d.f. form, attributed to sample elements x_i, are called kernels, for which the parameter h /of scale/ changes /decreases/ with the growth of a random sample size, defined by n. Therefore h is the function of n and such function must have certain properties of convergence to zero together with n, if the function $\Sigma_i \Psi(x,x_i,h)$ put in the place of $\varphi(x|S_n)$ in the formula(2) has to meet this condition of convergence.

To kernels method we raise three objections here :

1. Arbitrary way to assume the type of function as a kernel. If there is an arbitrariness, then there is no optimization.

2. Lack of gnosiologic interpreation of the total dispersion dependent on "h", which except for some limitations in convergence, may be arbitrary.

3. The introduction of the scale parameter "h" and its dependance on the sample size, i.e. on its variation, cause complications ; especially as it is toilsome in algorithms adapted for computer calculations.

The sample d.f. /sdf/ proposed by the author seems to be free from the above presented objections relating to both the empirical d.f. as well as potential functions, based also on kernels. It is defined, basing on the sample $S_n = \{x_1,\ldots,x_n\}$ represented in $\mathbb{R}^m$, as follows :(*)

(3) $$\varphi(x|S_n) = C_m(S_n) \prod_{i=1}^{n} \bar{\varphi}(x|x_i),$$

where

(4) $$\bar{\varphi}(x|x_i) = c_m(1+\|x-x_i\|^2)^{-m},$$

and $C_m(S_n) > 0$ is the constant normalisation coefficient. The functions (4) attributed to the sample element x_i is a normalized, rational Cauchy distribution ; i.e. the scale parameter of this d.f. is $\lambda = 1$.

(*) The power m is conditioned by the character of integral convergence on the space $\mathbb{R}^m$ /necessary to make the function (4) a d.f./. If the sample consists at least of m elements, then it is enough to put in the formula (4), the numer 1 instead of m.

This is of course one of the possible forms of potential functions attributed to points. The admission of just this requires the following three asumptions :

1. Assumption of Cauchy distribution for each axis of the coordinates system.

2. Normalization of the scale parameter of this distribution i.e. the assumption that $\lambda = 1$.

3. Assumption of a rotational form of the multidimensional /multivariate/ distribution.

In [7] there assumptions and principles on which their introduction is based, were discussed broadly. There this is presented in outline.

The introduction of Cauchy distribution is justified by the Bayes postulate [4], which after all may be explained on the basis of statistical game theory /see e.g. [3]/. If we have only a sample element represented as a point in $\mathbb{R}^m$ and we do not have other information, then we must choose an element from class of distributions so as to have the highest probability in the said point and the largest dispersion /i.e. infinite/ of the distribution a posteriori.

The second assumption will be discussed broadly in the following paper of the author concerning the applications of this method. Here only the interpretation of the parameter λ will be given as the smallest distance of distinguishability. Let for i-th feature the accuracy of the best measuring instrument, i.e. its dispersion, be equal to λ_i. Then for two elements of the sample of which the results of measurements of the i-th feature do not differ more than by $2\lambda_i$ /after normalizing by 2/ i.e. generally for elements x_k and x_l $\delta(x_k,x_l) \le 2\lambda_i$, when considering only the i-th feature, we cannot state with high probability, that they differ really. In this case the sdf /proposed here/ has a property of "gluting", i.e. that between points $x_{k,i}$ and $x_{l,i}$ on the i-th axis there appears then only one maximum. If however, it is $\delta(x_k,x_l) > 2\lambda_i$ and in the vicinity of points $x_{k,i}$, $x_{l,i}$ there are no other sample elements, sdf has a local minimum between $x_{k,i}$ and $x_{l,i}$ /this property may be generalized on $\mathbb{R}^m$/.

The information about the dispersion of the measuring instrument appears clearly in sdf : when we have only one element of the sample, we have no information about this dispersion. Therefore sdf in this case has an infinite dispersion. If we have only two elements of identical values of the i-th feature /when considering only the i-th feature/, sdf has the smallest dispersion for this case, i.e. equal to λ_i. On the contrary, if we know the accuracy of the measuring instrument for each investigated feature of continuous type, we know thereby the constant value of parameter λ_i /$i = 1,2,\dots,m$/ for it, and we may normalize the adequate axis of the Cartesian coordinate system with their help. It enables the elimination of parameters in the normalized sdf formula (3)+(4).

The rotational form (4) of individual distribution on R^m was assumed because it seems to be natural and because was analytically convenient /it makes easier to prove the theorem ; compare the proof of lemma 1/.

In the formula (3) defining the distribution attributed to sample S_n /"sample alloy"/ the assumption of a product instead of a sum, which appears in the potential functions method, is essential and should be also explained. Because, as we mentioned already when analysing the assumption 2, the parameter λ has a gnosiological sense and therefore must be constant, the sum of the assumed distributions attribued to points of S_n is not a consistent estimate of f(x) /see formula (2)/. In the lemma 2 we show that in the case when $f(x) = 0$, the product defined by (3) and (4) is convergent to zero for $n \to \infty$, according to the criterion (2), while the sum potential functions /when remain positive on the whole space R^m also in the limit/ is then convergent to the positive number.

The introduction of a product of individual potential functions to sdf may be moreover motivated probabilistically : the probability defined by sample density on any elementary interval is in sdf proportional to the probabilities defined on this interval by distributions of all individual points.

2 - APPLICATIONS OF THE METHOD

The problem of calculation in practice of the normalizing coefficient $C_m(S_n)$, which must be so chosen as to make the function (3) the sdf, still remains. Because of analytical difficulties each time calculation of this coefficient is a serious problem of a data processing nature. However, for the discrimination and pattern recognition this may be neglected, as for these purposes it is enough to compare the ,values sdf defined on different points $x \in \mathbb{R}^m$, and the constant coefficient does not exert an influence on these comparisons.

The point of using alternatively the method of statistical discrimination and taxonomy depends on the interpreation of S_n set. If we consider this set as a random sample, then the problems presented at the beginning appear. If we do not have additional information, we consider this set deterministically and then the taxonomy and particularly the so called "cluster analysis" should be used [2]. One should emphasize here that the application of sdf defined by formulae (3) and (4) appear to be an improvement in both these domains ;

particularly the hypersurfaces of an equal density /aequidensae/ and also minimal /demarcation-/ hypersurfaces are analytically of a simplest form in this case, as they are polynomials of low degree and the coefficient may be then neglected.

In this paper because of the lack of place we shall focus our attention only on the problem of sdf convergence to f(x). The algorithmic part of presented method and some examples of its applications /to pattern recognition and biology/ will be the theme of a next paper.

3 - THE CONVERGENCE PROBLEM

Very general theorems are desired concerning the distribution of maximum sdf deviation from f(x), analogically to A.N. Kolmogoroff's theorem for empirical distribution functions /see e.g. [6]/. However, the sdf defined by formulae (3)+(4) is not receptive to the use of traditional methods of probability calculus as e.g. the method of characteristic functions. Therefore the author cannot present such theorem for sdf at this stage of investigations. The sdf derivative has however a convenient analytical form, which was used in the proof of the lemma 2.

The presented two lemmas are useful in proving theorems concerning the stochastical convergence of sdf. The lemma 2, independent on this meaning, is useful in pattern recognition problems. The theorem is limited only to the local convergence of the derivative's sign of one of two factors in which sdf may be factorised, to f'(x). This factor is based on the internal part of S_n, set, relative to the optional area A on which f'(x) has the constant sign. If however, the author's hypothesis based on the lemma 2, that the second factor of sdf (based on the external set $S_n - S_n^*$) does not exert an influence within the limits of A on the derivative's sign of sdf, then the thesis of the presented theorem is generally true, i.e. sign $\varphi'(x|S_n)$ converges stochastically to sign f'(x).

<u>Definition</u>

The function $\Psi(x,x_i)$, where $x \in \mathbb{R}^m$, will be called the <u>rotation-potential function</u>, if it is continuous and has the point, referred to as x_i /the index i will be used in sets/, which will be called the <u>centre</u>, and satisfies two following conditions :

For two optional points $x', x'' \in \mathbb{R}^m$ is satisfied :

1°) when $\delta(x',x_i) = \delta(x'',x_i)$, then $\Psi(x',x_i) = \Psi(x'',x_i)$,

2°) when $\delta(x',x_i) > \delta(x'',x_i)$, then $\Psi(x',x_i) < \Psi(x'',x_i)$,

where δ means the Pythagorean distance formula.

Lemma 1

Let the set of n points $S = \{x_1,x_2,\dots,x_n\}$, $x_i \in R^m$, be divided by the (m-1)-dimensional hyperplane H in two subsets : $S_1 = S \cap \Sigma_1$ and $S_2 = S \cap \Sigma_2$, where Σ_1, Σ_2 mean semispaces obtained by the section of $\mathbb{R}^m$ with H. S_1 has n_1 and S_2 n_2 elements /$n = n_1+n_2$/.

Let $n_1 > n_2$ and in S_1 n_2 elements are symmetric to all elements of S_2, relative to H /additional assumption/. Let to each x_i the rotation-potential function be assigned /see definition/, with the centre in x_i. For the set S the function ξ is defined :

(5) $$\xi(x,S) = c[\Psi(x,x_1) \oplus \Psi(x,x_2) \oplus\dots\oplus \Psi(x,x_n)],$$

in which $\oplus$ means addition or multiplication.

Then, for two optional areas : $\mu_1 \subset {}_1$ and $\mu_2 \subset \Sigma_2$, symmetric relative to H, the inequality is fulfiled :

(6) $$\int_{\mu_1} \xi(x,S)dx > \int_{\mu_2} \xi(x,S)dx.$$

When $\oplus$ means in this formula the multiplication, the constraint $\Psi_i > 0$ is necessary.

The proof :
Let S_1' be the subset of S symmetric, relative to H, to S_2 and let D be the set-difference : $D = S-S_1'-S_2$. It is easy to see that for each pair of points x',x'', where $x' \in \Sigma_1$ and $x'' \in \Sigma_2$, symmetric relative to H, due to the definition and (5), the inequality is fulfiled

(7) $\xi(x',D) > \xi(x'',D)$.

Consequently it is fulfiled

(8) $\xi(x',S) > \xi(x'',S)$,

because, as the consequence of (5), $\xi(x,S)$ may be decomposed in factors or sum elements (*), and due to the additional assumption it is $\xi(x',S_1') = \xi(x'',S_2)$. This gives, after integration, the thesis.

Lemma 2

Let S_n be the sample of n elements drawn from the population with the unknown d.f. $f(x)$, where $x \in \mathbb{R}^m$ and let $A \subset \mathbb{R}^m$ be the area of m-dimensional measure $\mu(A) \neq 0$, such that for $x \in A$ is $f(x) = 0$. Then the sdf defined by the formulae (3) and (4), for area $A^* \subset A$, such that for optional $\varepsilon > 0$, $\mu(A)-\mu(A^*) < \varepsilon$, fulfils the equality

(9) $$\forall_{x \in A^*} \ \forall_{\eta>0} \ \lim_{n\to\infty} P\{\varphi(x|S_n) < \eta\} = 1.$$

The proof

Let the hyperplane H_j be orthogonal to the j-th axis of m-coordinates system in $\mathbb{R}^m$. Let H_j cut this axis in the point with the coordinate x_o^j. We shall assume at the first stage of proving, that $f(x) = 0$ for all points $x \in \mathbb{R}^m$, such that theirs j-th coordinate $x^j > x_o^j$. It will be shown now, that for these points in this case is valid (9).

The derivative of $_{,}(x|S_n)$ for j-th variable is of the form :

(10) $$\varphi'(x|S_n) = -2 \ \varphi(x|S_n) \sum_{i=1}^{n} \frac{x^j - x_i^j}{1+\sum_{k=1}^{m} (x^k+x_i^k)^2}$$

Let $\{X_i^j\}$ be the matrix of random variables such, that for constant i (i.e. in the row) they have the joint distribution defined by $f(x)$, instead rows have identic and independent each other distributions. If we put for optional $x^j > x_o^j$:

(*) The constant coefficient c in (5) may be omitted here.

$$Y_i^j = x^j - X_i^j, \quad \text{and}$$

$$Y_i^* = [1 + \sum_{k=1}^{n} (x^k - X_i^k)^2]^{-1},$$

it results from assumptions that

$$P\{0 < Y_i^j\} = 1, \text{ and } \qquad 0 \le Y_i^* \le 1,$$

for $i = 1,2,\dots,n$.

It is easy to see, that for optional distribution of Y_i^j fulfiling given condition, is valid (*) :

(11) $$\forall_{\omega<\infty} \lim_{n\to\infty} P\{\sum_{i=1}^{n} Y_i^j < \omega\} = 0.$$

For the product of random variables $Y_i^j Y_i^*$ is also fulfiled :

$$P\{Y_i^j Y_i^* \le 0\} = 0,$$

and when the formula (11) for the random variable $Y_i^j Y_i^*$ is used it results, that for $x^j > x_o^j$ is valid :

(12) $$\forall_{\omega<\infty} \lim_{n\to\infty} P\{\sum_{i=1}^{n} Y_i^j Y_i^* < \omega\} = 0,$$

which shows that in this case the second factor (the sum) in (10) is unconstrained, when $n \to \infty$, with probability 1.

Because φ is the product of functions, which for assumed x-es are /with probability 1/ decreasing, limited from below by 0, and also differentiable, φ must be in this interval in the limit decreasing or everywhere equal to zero. Therefore $|\varphi'(x|S_n)|$ must be in the limit constrained and by virtue of (10) it must be so, that for $x^j > x_o^j$ and $n \to \infty$, $\varphi \to 0$.

When none hyperplane such as H_j $(1 \le j \le m)$ exists, there exists the area $A^* \subset A$, surrounded with the set of $(m-1)$-dimensional hyperplanes $H_1, H_2, \dots, H_l$

(*) For pattern recognition problems the Cauchy distribution with the cutted out "window" (the area A on which $f(x) = 0$) is of interest. By virtue of lemma 2, this "window" may be reproduced using sdf.

(1 > 1), of which measure $\mu(A^*)$ is optionally close to the measure $\mu(A)$. In this case the complementary area to A^*, relative to $\mathbb{R}^m$, may be divided in parts (e.g. by aid of other hyperplanes) where each part is adjacent to one 1 hyperplanes so, that for H_j all $x_i \in S_n$, belonging to such part of $\mathbb{R}^m$, create the set S_n^j. Then occurs :

$$(13) \qquad \Sigma_{j=1}^{1} \quad S_n^j = S_n.$$

For each S_n^j, the proper rotation of coordinates system may be performed so, that for all $x_i \in S_n^j$, $x_i^j < x_o^j$ with probability 1, as in the case described at the beginning. Then also occurs :

$$(14) \qquad \varphi(x|S_n) = c \prod_{j=1}^{1} \varphi(x|S_n^j),$$

and for each factor in (14), with the adequate H_j, for $n \to \infty$, it is possible to carry out the previous presented proving.

Wich the product (14) corresponds the set-product of semispaces equal to A^* and for all S_n^j the formula (12) is fulfiled, which shows that for the product (14) $\varphi \to 0$, for $n \to \infty$.

Theorem

Let f(x), for $x \in \mathbb{R}^m$, be the differentiable d.f. of population and let $A \subset \mathbb{R}^m$ be the optional area on which the derivative f'(x), taken in the defined direction, has everywhere the same sign /not equal to zero/. Let S_n be the random sample drawn from this population and H be the optional, (m-1)-dimensional hyperplane orthogonal to defined derivative's direction, cutting the interior of A. For the area $A^* \subseteq A$ symmetric relative to H and for the set $S_n^* = S_n \cap A^*$ is then valid :

$$(15) \qquad \forall_{x \in H} \lim_{n \to \infty} P\{\text{sign } \varphi'(x|S_n^*) = \text{sign } f'(x)\} = 1,$$

where prime means the derivative in the assumed direction.

The proof :

The area A^* may be divided into elementary cubes :

$$(16) \qquad A^* = \sum_{j=1}^{1} A_1^j + \sum_{j=1}^{1} A_2^j,$$

so that it is possible the form the pairs : A_1^j, A_2^j, situated symmetrically relative to H. Let us consider initially one, optional pair of adjacent cubes. Let e.g. $f'(x) < 0$ in the direction from A_1^j to A_2^j. Then

$$\int_{A_1^j} f(x)dx > \int_{A_2^j} f(x)dx. \tag{17}$$

Probabilities in (17) will be signed ε_1, ε_2 $(\varepsilon_1 > \varepsilon_2)$. Let ${}^jS_n^1$ be the set of elements $x_i \in S_n$ for which $x_i \in A_1^j$, having n_1 elements, and ${}^jS_n^2$ be the set of $x_i \in S_n$ for which $x_i \in A_2^j$, having n_2 elements $(i = 1,2,\ldots,n)$. When n is growing unconstrainable it is easy to see, on the basis of Borel's theorem (see e.g. [8], p. 270) in respect to the Bernouilli's scheme, that in this case :

$$P\{\frac{n_1}{n_1+n_2} \to \frac{\varepsilon_1}{\varepsilon_1+\varepsilon_2}, \frac{n_2}{n_1+n_2} \to \frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}\} = 1, \tag{18}$$

and in consequence with probability 1 the inequality $n_1 > n_2$ is valid. Then, due to the lemma 1, for optional pair of symmetric, relative to H, areas B_1, $B_2 \in \mathbb{R}^m$, with probability 1 :

$$\int_{B_1} \xi(x, {}^jS_n^1 + {}^jS_n^2)dx > \int_{B_2} \xi(x, {}^jS_n^1 + {}^jS_n^2)dx, \tag{19}$$

by additional assumption concerning the symmetry of ${}^jS_n^2$ and some subset of ${}^jS_n^1$, agreeable to the lemma 1 (as we shall show this assumption appears to the overfluous).

When B_1, B_2 are adherent through points $x \in H$, (19) gives, for $\rho(B_1+B_2) \to 0$, where ρ means the diameter of the set, the inequality

$$\forall_{x\in H} \quad \xi'(x, {}^jS_n^1 + {}^kS_n^2) < 0, \tag{20}$$

where the derivative's direction is orthogonal to H.

Let now $n \to \infty$ and $1 \to \infty$ simultaneously, so that

$$\max_{1\le j\le 1} \rho(A_\cdot^j) \to 0. \tag{21}$$

In this case the additional assumption may be omitted, because the symmetry inside the elementary cubes lost the importance. Then for all $j = 1,2,\ldots,l$ the inequality (20) is valid, for optional established $x \in H$, with probability 1, and therefore it is also valid with probability 1 for the total subset S_n^*, which gives the thesis in this case :

$$(22) \qquad \forall_{x \in H} \lim_{n\to\infty} \varphi'(x|S_n) < 0.$$

This ends the proof in the case when $f'(x) < 0$ in the assumed direction. In the reciprocal case, it is sufficient to change the turn of the derivative's direction vector and the further proof runs identically. In the formula (15) appears, under the probability sign, the equivalence, because negation of the sentence on the right side causes negation on the left side too. This ends the proof.

REFERENCE WORKS

[1] Aizerman M.A., Braverman E.M., Rozonoer L.J. - The probability problem of pattern recognition learning and the method of potential functions. Automation and Remote Control, 25, 1964, p. 1175-1190.

[2] Bijen E.J. - Cluster Analysis. Survey and evaluation of techniques. Tilburg University Press. The Netherlands 1973.

[3] Blackwell D., Girshick M.A. - Theory of Games and Statistical Decisions, New-York, J. Wiley 1954.

[4] Box G.E.P., Tiao G.C. Bayesian Inference in Statistical Analysis. Addision - Wesley Publishing Company, Reading (Mass.) Menlo Park (Calif.) London Don Mills (Ont.) 1973.

[5] Cacoullos T. Estimation of a Multivariate Density. Technical Report n° 40. Department of Statistics. University of Minesota 1964.

[6] Kolmogoroff A.N. Confidence limits for an unknown distribution function. Annals Math. Statist. Vol. 12 1941, p. 461-463.

[7] Mikiewicz J. Zarys podstaw statystycznej metody gestosci probkowych. Prace Naukowe Instytutu Nauk Spolecznych Politechniki Wroclawskiej nr 15. Studia in Materialy Nr 8, 1979.

[8] Papoulis A. Probability Random Variables and Stochastic Processes. Mc Graw-Hill, Inc. 1965.

[9] Parzen E. An estimation of a probability density function and mode. Annals Math. Statist. Vol. 33, 1962 Chapter 6.

[10] Bean S.J., Tsokos C.P. Developments in Nonparametric Density Estimation. Intern. Statist. Review, 48, 1980, p. 267-287.