Annales Academiae Paedagogicae Cracoviensis

Folia 60 Studia Philosophica V (2008)

Jan Mikiewicz

The synergism of taxonomy and mathematical statistics as the epistemological instruments

1

The exact (i.e. quantitative) research methods are necessary for development of empiric science. Taxonomy and mathematical statistics belong to this type of methods. But it is not known that just in the hard connection of these two branches of science it is possible to gain the new profitable empirical methods, and also it is possible to understand better some basic procedures, as theory of gauge and inductive logic.

In this paper this matter will be discussed concisely, with the recall to the earlier papers of the author, in which the synergism of taxonomy and statistics gave two new empirical methods. First of all, one of these methods will be confirmed by two proofs. It pertains to the sample densities method defined by formulae (1) and (2)[1] whose optimality is proved. This density does not only converge to the probability density in the population, but also indicates some types in it (the local maxima of density) – it is therefore also a taxonomic method. Parzen's method[2] is not optimal, because when in its constructions there appear free choices, it is not clear how failures are made in the approximation of density in population by this method. Moreover, this method is awkward in applications, especially in a multidimensional case, i.e. when in the examined objects many features must be taken into account.

2

Today, taxonomy, as an empirical method, is not a scientific discipline which has its own basis (the set of primordial notions) and methodological system. Nevertheless, in the author's opinion, such a system may be attained.

[1] J. Mikiewicz, *Sample density as the function-estimate of population's distribution,* „Statistiques et Analyse des Données" 1982, vol. 7, n° 2, pp. 57–70; idem, *On the optimality of sample-densities method*, „Questiones", vol. 1, Wrocław 2002, p. 207–228.

[2] E. Parzen, *An estimation of a probability density function and mode*, „Annals of Mathematical Statistics" 1962, vol. 33, Chapter 6.

The basic notion of taxonomy is, in my opinion, similarity (likeness), however this notion is at present only intuitive. The quantitative grasp of the real world contributes, nevertheless, to the precision of such notions. In this way it is often necessary to restrict the intuitive sense of the notion; e.g. in physics the primordial sense of notions: velocity, strength, power were restricted to quantified ones. (A new approach to quantified notions has been proposed by T. Grabińska i M. Zabierowski in 1978-81, Institute of Metrology I-21, Wrocław; cf Zabierowski's paper of this volume).

The same process must be performed for "similarity". The similarity measure must be based on some quantitative features which may be interpreted as coordinates in any Euclidean space. When the feature is univalent, it is to be assigned to number 1 when this feature is represented in the object, and number 0 when it is not represented. Hugo Steinhaus (Wrocław) used as the measure of the distance of two ecosystems the number of common species living in these ecosystems, related to the total number of species living in them. Similar distance may be presented in Euclidean space, when we consider the space in which the axes of coordinates correspond to species and ecosystems to points of this space. In this case we put the coordinate 1 when a given specie is represented in the ecosystem, and put 0 when it is not so.

It is also possible to consider the "inverted" space: Let each object from the considered set be assigned one axis of the coordinates, and each feature be assigned one point in this (Euclidean) space in such a manner that each coordinate has the value of this feature observed by the corresponding object. In this model the distance between two points (i.e. features) is equal 1 minus the (empiric) correlation coefficient between these two features. A similar formula has already appeared in Devroye's paper[3]. The author knows another connection between the correlation coefficient and the distance in Euclidean space. There appears here again the connection between taxonomy and statistics. The feature of objects may be e.g. a curve or a surface; the measure of similarity may be defined as the distance between two curves or two surfaces (Devroye).

In this model it may be stated that the objects of taxonomy may be considered as the points in space – Euclidean, eventually abstract. On this basis, the similarity must be defined as the distance of points in the conventional space. The intuitive sense of similarity is generally contained in this definition, but psychologically, in some cases, the perceired similarity may be not conformable to this definition.

There is especially the problem: when we accept the Euclidean space, does the triangle law oblige the similarity or not? The obligation of this law in Euclidean spaces is ensured by Minkowski's inequality, for all distances of the class called the Minkowski's type by the author[4]. For example, when the measure of similarity be-

[3] L. Devroye, *A Course in Density Estimation*, Birkhäuser, Boston–Basel–Stuttgard 1987.

[4] J. Mikiewicz, *Statistical selection method of the best objects*, edit. European Meeting of Statisticians, Transactions of the Conference, Prague 1977.

tween a daughter and a mother is a, and between a mother and a grandmother is b, it is the question of relation:

$$a + b \geq c \ ,$$

where c means the similarity between a daughter and a grandmother. In my opinion, it is the question of accommodating the intuitive feeling to the mathematical law, only. When we take into account the sets of genes as the bearers of features, the triangle law is valid. It is again such a model that when the gene of the defined type exists, we put 1, and when it is not so, we put 0; in the case of the set of 2 genes, we put number 2. In this model the distance is the measure of genetic similarity (dissimilarity).

Such a grasp of taxonomy (as considering the representation of objects as points) in the metric spaces enables us to examine the clusters of similar objects and it is in my opinion the main aim of taxonomy.

Let us consider additionally the well known notion of triangles similarity. The triangle is defined when two angles and one side are defined and consequently may be represented in three–dimensional Euclidean space. The geometric similarity of triangles is a singular case of our notion of similarity: the distance in features space is defined by the distance in the side dimensions only (difference in scale). Differences in angles are not taken into account. The kindred problem is with the similarity of a cat and a tiger - when the dimension of size (of scale) is omitted, these two animals are very similar.

J. Czekanowski[5] gave the first method of explaining the constellation of points in a multifeatures space on the plane. The diagram is identical with the correlation matrix, where the correlation coefficients represent distances between points in a multidimensional space, because these coefficients express the scalar products. The order of points in the diagram must be permuted so that big coefficients would be situated closest to the main diagonal. Such a method is effective when examined points create a chain. In another case this method gives a false picture. The next step was made in Wrocław. It is necessary here to discuss the term "Wrocław taxonomy". The proof of the theorem on the uniqueness of the shortest linkage, called in Poland "the dendrite"[6], two algorithms for its construction, and application of this graph to taxonomical problems in empiric sciences, were the important success in the methodology of science (it was also called "Florek's method").

This dendrite is from the graph-theory point of view an undirected graph, and there is an interesting relation between this graph and the directed graph known as the genealogical tree. In Wrocław the first graph spanned on points in multidimensional feature-space (Florek) was made where points were representatives of skulls of the well known fossil hominids from anthropoids to recent homo sapiens. The high conformity between this graph and the genealogical tree of homo sapiens made by anthropologist was shown.

[5] J. Czekanowski, *Zarys metod statystycznych w zastosowaniu do antropologii*, PTNW, Warszawa 1913.

[6] K. Florek and others, *Taksonomia wrocławska*, „Przegląd Antropologiczny" 1953.

The very important advantage of the shortest linkage (dendrite) is the fact that it may be displayed on the plane (chart of paper) any time, which enables displaying on a plane the kindred (similarity) of objects, in the formerly mentioned sense, represented as points in multidimensional space. It is therefore the basic method of – widespread in Anglo-Saxon countries – "the cluster analysis".

3

Similarly as taxonomy, mathematical statistics may be treated as the mathematised (not mathematical) model of reality elaborated as the gnoseological tool for research in empirical sciences. As the example of such models we may present here mechanics (i.e. Newton's model or Einstein's model)[7]. The modeling and problems connected with existing models in cosmology and physics are discussed e.g. in Kyburg, in Grabińska[8] (especially in appendix B) and in Zabierowski[9]. They are the mathematised models of physical reality and they are undoubtedly gnoseological tools for research.

In such models the primordial notions and also "axioms" (i.e. basic laws), which pertain obviously to the empirical reality must be presented. In the taxonomic model, the "axiom " is the formerly discussed triangle law in features spaces and the similarity as primordial notion. In my opinion, no attempt for "axiomatization" of mathematical statistics before my papers[10] is known. The great disorientation is generally observed concerning the foundations of mathematical statistics, which will be partly discussed later.

In the mentioned papers the primordial notions are: 1. population, 2. random choice from this population. The population is an arbitrarily defined set of objects with measurable or countable features on which the probability distribution, i.e. the normalized measure, is defined. The sets on which it is impossible to define the distribution cannot be populations. The criterion belongs to the so called Borel's sets.

The "axioms" of mathematical statistics are:

1. Each element chosen at random from the population may be assigned a random variable with the distribution identical with the distribution of the mentioned population.

[7] H. Kyburg, *Probability and inductive logic*, London 1970.

[8] T. Grabińska, *Poznanie i modelowanie*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 1994.

[9] M. Zabierowski, *Redshifts and Arplike Configurations in the Local Group*, [in:] *Progress in New Cosmologies. Beyond the big-bang*, E.C. Arp et al. (eds), New York 1993.

[10] J. Mikiewicz, *Zastosowanie statystycznej metody wyboru najlepszych obiektów w chemii*, Prace Naukowe Politechniki Wrocławskiej 1990, nr 24, „Studia i Materiały" nr 6, Wrocław 1990; idem, *Logika indukcji a statystyka matematyczna*, „Roczniki Filozoficzne KUL" 1994, tom XLII, z. 3.

2. Two arbitrary elements of defined population chosen at random from this population have the probability distributions stochastically independent on each other.

The author is convinced that these two "axioms" are sufficient for construction of mathematical statistics and enable us to introduce total probability theory, which gives great possibilities to this methodology. However, this junction must be clearly described. As it was mentioned, recent authors sometimes avoid the clear definition of the population and such "axioms", and this is the cause of some disorder in the statistical interpretation of some empirical procedures[11].

We shall take into account, as the example, the gauge procedure in empirical sciences. It is doubtless for specialists that each gauge procedure is connected with some failures which happen at random and therefore the probability theory must be applied here. However, it is not clear till this day in what manner this problem is connected with the mathematical statistics. The solution described in Mikiewicz[12] is based on the notion of potential populations. This problem is similar to the classical problem of coin or die tossing. When the circumstances of tossing are saved, we can imagine the infinite string of tosses as the potential population and some first (real) tosses, as the sample elements chosen at random from this population (the order of individual tosses is not intrinsic, because we have here the sequence of random variables identically distributed and independent from each other). The model for the string of measures of any real value is identical, because each operation of measure may be treated as burdened with the random failure which is the random variable with the mean zero and the variance defined by the character of the measure tool[13]. The tool of measure creates the potential population.

Another problem created the astonishingly rich literature concerning the logic of induction (inductive logic). In Polish there is a comprehensive book of H. Mortimer, which explains the main questions of this topic[14]. Among American authors the outstanding authors in this area are e.g. H. Kyburg[15] and J. Levi[16]. Though the main aim of this logic should be the research of objective truth, in this literature we observe the theories of subjective probabilities only. A distinguished author of **subjective** probability theory is a well known economist, J.M. Keynes, though he stated himself that he creates the **objective** probability theory.

In my opinion, the basic methodology, which enables the winning of objective science in empirical (quantitative) research is mathematical statistics, because it must

[11] J.R. Barra, *Mathematical foundations of statistics*, Paris 1982; Kyburg, *Probability and inductive logic*; J. Levi, *The enterprise of knowledge*, Cambridge 1980.

[12] J. Mikiewicz, *Logika indukcji a statystyka matematyczna...*

[13] J. Mikiewicz, *Sample density as the function-estimate of population's distribution...*; idem, *Zastosowanie statystycznej metody wyboru najlepszych obiektów w chemii...*

[14] H. Mortimer, *Logika indukcji. Wybrane problemy*, Warszawa 1982.

[15] H. Kyburg, *Probability and inductive logic...*

[16] J. Levi, *The enterprise of knowledge...*

base on the measure operations, which are stochastic, and in addition the Nature itself creates the stochastic mechanisms which give information by statistical procedures only. The logic of induction must be a part of epistemology which consists in applications of mathematical statistics and auxiliary procedures. An example of disorientation in this area is the well known definition of convergence to probability proposed by Mises. The author offered here the statistical problem. The inductive methods of J. St. Mill or other authors may be interpreted as statistical methods (e.g. some procedures of so called analysis of variance) or are bereft of sense. For example the fact that the sunrise appeared billions times does not give the certainty that next time there will also be the sunrise, because there may be an explosion of the sun. The empirical science is not deterministic; it is stochastic but for the human life it is sufficient to have a very high probability. As H. Steinhaus remarked, probability of the failure of the parachute is e.g. 1/10 000 and nevertheless people jump with the parachute.

4

One of aims of taxonomical research is looking for clusters in multidimensional spaces. The objects (points) in economy may be enterprises or customers, in biology-bacteria or other living individuals. In Wrocław the dendrite was made, in which the objects were autoengines.

We meet, however, interesting clusters, when the examined set is the random sample. Until today, the main object of statistical research was any category of regular populations, generally normally distributed (i.e. of Gaussian distribution). Although there are some rational foundations to expect the appearance of normal distributions in the world, in different areas, especially in the realm of the living beings, it is reasonable to expect the multimodal (i.e. irregular; with many maxima) distributions. When we consider e.g. the living population composed of some species or genera, we receive in one- or multi-dimensional space the multimodal distribution, as the picture of population, where we have the maxima which assign species or genera to feature values, but the total population is fuzzy because we observe the impact of random phenomena. The problem arises how to construct the estimator of this irregular, multimodal population, which would enable to obtain the opportune analysis of this unknown population; here appears the need of synergism of two methodologies: taxonomy and mathematical statistics. Some explorers in application of mathematics area acknowledged that for clusters exploration the shortest linkage (dendrite) method is sufficient[17]. This statement is true, when the clusters are "regular", i.e. are spherical and in total population are sufficiently separated from each

[17] J.A. Hartigan, *Consistency of single linkage for high density clusters*, JASA 1981, vol. 76, pp. 388–394.

other. In the other case the dendrite may give false picture of the explored set in multidimensional space, and when the set is the random sample, it may give the false picture of the population explored.

We have, important Wanke's method[18], which gave interesting results in statistical research of the composed populations, as taxonomic method has some other defect which will be explained. The method is based on dividing the features space on the cubes. Each axis representing one feature is divided for example in three intervals (in anthropology e.g. – the sculls: short, medium and long). When we consider n features, we have 3^n cubes in features space in this case. In biology one must use many features. When we use 10 features, we must examine 3^{10} cubes (a big number). We shall treat the cubes in which many elements (sample points) are contained as the clusters, i.e. the types, but other cubes, in which few elements are contained, must be omitted.

In some cases this method gives correct results, but in other cases – false. When the cluster, i.e. the real type, is located in tops of cubes, in this method it vanishes, because it is divided in many cubes. When the cluster is located on the wall of two cubes – the method will designate two clusters (types). In many cases the localization of types is assigned incorrectly by this method. It must be mentioned that both methods described here require more computing operations and comparisons than the method presented in Mikiewicz[19].

This method does not deform the real clusters (types), because the nonparametric estimation of probability density is convergent to the real density in population explored, which will be explained in the Appendix, and it is close to the real one with some probability. When the number of sample elements grows boundless, the sample density (1) becomes identical with the density in population with probability 1. The dendrite method, on the other hand, is dependent on the random situation of single points, which changes the shape of dendrite and therefore deforms the picture. Wanke's method is dependent on artificial (human) division of the space on cubes, which changes the number of clusters (types) and also the shapes of the picture types.

Additionally we shall give some information about Parzen's density method[20], which was developed independently from the approach given in the author's formulae (1) and (2) (in Appendix). This method offers the sample densities too, but these densities are optional (are not single – defined), because the form and dispersion of distributions assigned to alternate sample elements are not optimal and the author states only that the dispersion of these distributions must tend to zero, when the number n of sample elements tends to infinity. In these densities the sum of individual

[18] A. Wanke, *Metoda badania częstości występowania zespołów cech, czyli metoda stochastycznych korelacji wielorakich*, „Przegląd Antropologiczny" 1953, s. 106–147.

[19] J. Mikiewicz, *Sample density as the function – estimate of population's distribution...*; idem, *On the Optimality of Sample-densities Method...*

[20] E. Parzen, *An estimation of a probability density function and mode...*

(elementary) distributions is taken into consideration, while in the density (1) and (2) we have the product, which facilitates the calculations in taxonomic applications (the coefficients C^n and C' must not be computed in taxonomic problems). Moreover, here the individual distributions are constant (without parameters), while in Parzen's method[21] addition of new elements to the sample requires the change of the distribution formula of all sample elements. Obviously, the number of calculation units is in this case remarkably greater.

The other method based on the synergism of taxonomy and mathematical statistics is the method[22] of choice of the best object. This method also originates from the "Wrocław taxonomy", but in this case the statistical approach is applied.

The deterministic "choice of the best object" is based on the mentioned notion of the features space, which may here be the Euclidean space (possibly the Hilbert space), which must be metric, which, as we suggest, is the "axiom" of taxonomy. To the set of points, which represent in this space the set of considered objects, the additional point is enclosed, which represents "the ideal object". This object has features, which we accept as the best for this type (considered) of objects, and these features must be chosen arbitrarily – adequately to our needs. As the distance in our space is the measure of similarity, the object from the considered set, which is situated in this space close to the "ideal object", is most similar to this object and consequently is treated as the best.

The probabilistic approach to this method follows from the fact that the features of objects must be measured on real objects and, as it was mentioned, they become in this way random variables with dispersion depending on the measure tool[23]. The set of measured features of considered objects must be consequently normalized by dividing by suitable dispersions of features. It gives us the value $\sigma = 1$ of parameter in individual distributions (2). This procedure is demonstrated on a chemical example in Mikiewicz[24].

The aim of this procedure is to supply the explorers of theoretical or technical sciences with probabilistic method which enables us to express with high probability the statements designating the best object[25]. In the case presented in Mikiewicz[26], the probability assigned to the best object is (in the author's theorem[27]) 0.99.

[21] Ibidem.

[22] J. Mikiewicz, *Statistical selection method of the best objects...*; idem, *Zastosowanie statystycznej metody wyboru najlepszych obiektów w chemii...*

[23] J. Mikiewicz, *Sample density as the function-estimate of population's distribution...*; idem, *On the optimality of sample-densities method...*

[24] J. Mikiewicz, *Zastosowanie statystycznej metody wyboru najlepszych obiektów w chemii...*

[25] J. Mikiewicz, *Statistical selection method of the best objects...*

[26] J. Mikiewicz, *Zastosowanie statystycznej metody wyboru najlepszych obiektów w chemii...*

[27] Ibidem.

This method is based on the a posteriori probability theory of Student's distributions. Using here the method of confidence intervals (used in Mikiewicz[28]) gives the worse result. In this area the author has in mind further development of this theory, which will be published.

5

Resuming the previous reasoning, we must state that these two models: of taxonomy and of mathematical statistics are close to each other and must be applied jointly. In this cooperation different difficulties arise, and mathematical problems as well, which are a challenge for the human mind. Generally, in author's opinion, the taxonomy and mathematical statistics are the areas containing many undiscovered, but important laws. The basis for synergism of these methods is the features space, which obviously must be identical in both areas, because it is suitable for quantitative modeling of the objects and problems of Nature's research. This features space appears in two types:

1. Discrete (atomic) – when the feature appears in finite or denumerable number of states.
2. Continuous – when we are conscious of the fact that the precision of measuring the real existing value of the feature depends on our method (tool) of surveying[29].

In practice the feature may be of discrete type, but the density (number) of values is so high that the feature may be treated as continuous (example: the metal peas are composed of atoms, but for engineering are treated as continuous).

When an arbitrary set of objects is considered, the objects have some features of interest and the set is the object of taxonomy, but when features of the same value are repeating in objects – we have the population with probability distribution. When the separated maxima appear in the distribution, i.e. we have multimodal distribution, these local maxima are again the objects which may be examined by taxonomy. It shows the concatenation of taxonomic and probabilistic problems.

The calculations necessary in applications of the method of choice[30] are in many cases so limited that they may be performed on paper – without a computer[31], but the method of sample densities contributes very many computing units and therefore the computer aid is necessary. The author has a computer program which designates the

[28] J. Mikiewicz, *Dendrytowe obszary ufności*, „Zastosowania Matematyki" 1970, t. XI, z. 4, p. 391–421.

[29] J. Mikiewicz, *Sample density as the function-estimate of population's distribution...*; idem, *On the optimality of sample-densities method...*

[30] J. Mikiewicz, *Statistical selection method of the best objects...*

[31] J. Mikiewicz, *Zastosowanie statystycznej metody wyboru najlepszych obiektów w chemii...*

clusters in multidimensional space by relatively simple procedures based on[32](System WROTAX). It was an annex to author's report for Economic Institute of the University of Technology in Wrocław, in 1989.

APPENDIX

In the paper Mikiewicz[33], the reasoning showing the optimality of sample density (written further sd), defined by (1) and (2), was presented, but performed on the condition of not so big, constant n, i.e. the number of sample elements. The formulae, for $x \in \mathbb{R}^m$ and $x_1 \ldots x_n$ as elements of sample S_n, are next:

$$\phi(x \mid S_n) = C_m(S_n) \prod_{i=1}^{n} \tilde{\phi}(x \mid x_i) \tag{1}$$

where

$$\tilde{\phi}(x \mid x_i) = C'_m (1 + \|x - x_i\|^2)^{-p} \tag{2}$$

When m relative to n is small, the power p may be reduced to 1. The constant C'_m in the case m=1 is equal π^{-1} (Cauchy distribution; $\lambda = 1$).

The general character of convergence (1) and (2) to the population density (for $n \to \infty$) was omitted[34], and therefore we attempt here to show the taxonomic opportunity of this density for $n \to \infty$. For the necessary simplification we shall limit ourselves to one dimension (i.e. for $x \in \mathbb{R}^m$), but the generalization on $\mathbb{R}^m$ is possible.

The convergence of the well known empiric distribution function (edf) is sufficiently described by Kolmogoroff's distribution[35]. This function consists of segments of horizontal straight lines and has the incontinuity points in sample points x_i only. In Mikiewicz[36] the "empiric continuous distribution function" (ecdf) was proposed, which we define here in the following way: In each segment of edf we put a point in the middle. Let each neighbour pair of these points be conjoined by a straight segment, which gives us a continuous line. We perform the "ends" of this line as the following: The first point is situated conventionally on the x-axis e.g. in the distance equal to 2 distances between the first and second sample points (x_1, x_2) left from x_1,

[32] J. Mikiewicz, *Sample density as the function-estimate of population's distribution...*; idem, *On the optimality of sample - densities method...*

[33] Ibidem.

[34] Ibidem.

[35] A.N. Kolmogoroff, *Confidence limits for an unknown distribution function*, „Annals Mathem. Statistics" 1941, vol. 12, p. 461–463.

[36] J. Mikiewicz, *On the optimality of sample-densities method...*

and the last – similarly – on the horizontal straight line p=1 in the distance equal to 2 distances between last sample points($x_{n-1}x_n$). These two points will be conjoined with our ecdf by segments too, but conjoining the mentioned middle points.

This ecdf is a continuous line and is differentiated everywhere, without the set of the formerly discussed points. The differentiation of ecdf gives us the ”empirical density“ (esd), which is composed of rectangular poles with the surface – each equal $1/n$ (such diagrams are used by statistical “practicians“).

The definite integral of esd is exactly ecdf (it is the “Newton’s integral“) and reciprocally – esd is exactly defined by ecdf.

Because ecdf is convergent to population df, i.e. $\int_{-\infty}^{x} f(y)\,dy$, with the edf,

on the strenght of Kolmogoroff’s theorem[37], what is easy to see, esd is convergent to f(x) (see Annex II).

Now, we must consider the shifts of probability masses on the x-axis, arising when we put the individual probability densities (of Cauchy) instead of sample points, as it is made both[38]. This manipulation is necessary for “smoothing“ the discrete edf when the number n is not so big. Now, we shall show the character of convergence sd to f(x) when $n\to\infty$.

Putting the densities $\tilde{\Phi}(x \mid x_i)$ instead of the sample points (probability a posteriori), by constant but big n, causes the levelling of esd, but in small degree: From the theory of sd[39] it follows that the local minima must be situated in the distance Δ from the maxima, where $\Delta > 2\sigma$ and σ is the basic dispersion arising in the feature surveying. When the space is normalized, there must be $\Delta > 2$. Let the minimum be contained in the segment A and the maximum in the segment B of the same small length and the distance Δ between them. It is obvious that the probabilities on A for $\tilde{\Phi}(x \mid x_i)$ where $x_i \in$ B (when they are summarized) are remarkably greater than the probabilities on B for $\tilde{\Phi}(x \mid x_i)$ where $x_i \in$ A. All these probabilities are small when $\Delta > 2$.

It is a different situation when we introduce, instead of the summation, the multiplying (as it is presented in (1) and (2) also in Mikiewicz[40]), and when n is big. The multiplying hinders here the formerly showed leveling and causes the deepening of minima, which is necessary for taxonomic aims (distinguishing the types). The next theorem shows this property of density (1), (2), when $n\to\infty$.

[37] A.N. Kolmogoroff, *Confidence limits for an unknown distribution function...*

[38] J. Mikiewicz, *Sample density as the function-estimate of population's distribution...*; idem, *On the optimality of sample-densities method...*

[39] Ibidem.

[40] Ibidem.

The theorem

Let us study the population density f(x), distributed on the normalized axis (see above), continuous and having at least one maximum and local minimum close to it. Let the area (segment) A contains the minimum and area B of the same length contains the maximum. The distance between A and B, $\Delta > 2$. Between A and B f(x) is monotonic.

From this population the random sample S_n is drawn with big number of elements n. The ratio of probabilities for sd (defined by (1) and (2)) is the following:

$$r_n = \int_A \phi(x \mid S_n)\,dx / \int_B \phi(x \mid S_n)\,dx \tag{3}$$

If the inequality (5) is fulfilled, the ratio is r_n convergent, and if in the inequality (4) β is close to 1, if $n \to \infty$:

$$r_n \to 0$$

Proof:

We shall not take into account the elements x_i situated apart from the areas A and B, because they do not have impact on the result. Let the probability defined by f(x) on A will be a, and defined on B, will be b; in A there is n_a elements of S_n, in B n_b elements and $n_a/n_b \to a / b$, $(a << b)$.

For the individual distributions $\tilde{\phi}(x \mid x_i)$ let be:

when $x_i \in$ A, $\int_A \tilde{\phi}(x \mid x_i)\,dx = p_a^i$, and $\int_B \tilde{\phi}(x \mid x_i)\,dx = q_a^i$.

Similarly when $x_i \in$ B , we define p^i_b and q^i_b. It is obvious that for all $x_i \in$ A∪B, $p^i > 0.3$ and $q^i < 0.15$ (for Cauchy distribution).

Consequently the ratio (3) may be written:

$$r_n = \bigwedge_i \left[\left(\prod_{x_i \in A} p_a^i \prod_{x_i \in B} q_b^i \right) / \left(\prod_{x_i \in B} p_b^i \prod_{x_i \in A} q_a^i \right) \right] = \frac{\bar{p}_a^{n_a}}{\bar{p}_b^{n_b}} \cdot \frac{\bar{q}_b^{n_b}}{\bar{q}_a^{n_a}} = \left(\frac{\bar{q}_b}{\bar{p}_b} \right)^{n_b - n_a} \left(\frac{\bar{q}_b \bar{p}_a}{\bar{p}_b \bar{q}_a} \right)^{n_a}$$

where $\bar{p}_a$, $\bar{p}_b$, $\bar{q}_a$, $\bar{q}_b$ mean the medium values of adequate products of probability elements.

The above expressed ratio r_n is convergent under some conditions. Let the first factor be signed α^{t-u}, and the second β^u. It is obvious that $\alpha < 1$ and $t >> u$. When there exists such N that for $u > N$ the inequality occurs:

$$\alpha^{t-u} \beta^u < 1, \tag{4}$$

the ratio r_n is convergent, because the left hand side of (4) is monotonic. It is necessary to know how big the base β may be.

Another form of (4) is

$$\beta < \alpha^{1 - \frac{t}{u}}$$

Because $\alpha < ½$ and $t / u \rightarrow b / a$, for $b / a < 2$ we receive the upper bound for β:

$$\beta < 2. \quad (5)$$

When the length of segments A and B is very small, the ratios $\bar{q}_b \ / \ \bar{p}_b$ and $\bar{q}_a \ / \ \bar{p}_a$ are quite equal, and in such a case β is close to 1, and then it is obvious that

$$r_n \rightarrow \infty$$

It ends the proof.

ANNEX

(I) Two words – "taxonomy" and "statistics" are generally known as cutsh-words, but their suitable sense remains unknown. Moreover, the synergism of their methods remains unknown also for specialists of both methods. It may be surprising that, to my knowledge, the first who made an attempt to join these two methods was Jan Czekanowski[41]. As both methods in synergism (in conjunction) are, in my opinion, the basis for empirical sciences, I worked intensively in this area to present the problem in a possibly clear form.

The notion "taxonomy" was known, at the first, by biologists and agriculture specialists and is associated with the classification of living objects. For scientists connected in some way with Wrocław, it is associated with the cutsh-word "taksonomia wrocławska". Indeed, the taxonomical methods may be applied in many branches of science and practice and it is necessary to mention that the first application of taxonomy was perhaps in the antique studying of star constellations. The author of the important and, perhaps, basic taxonomical method – the shortest linkage, was Slovakian - Borůvka (he was an engineer). The second time this method was discovered independently in Wrocław, as the "Wrocław taxonomy" and applied to different scientific and practical problems[42]. Later this method was developed in western countries, but generally the authors of the method were forgotten. Understanding of the term "mathematical statistics" is for many people completely inaccessible and for others is misunderstood, mainly because the word "statistics" has, in popular using, many meanings. Generally, it means "the set (file) of numbers collected from some objects", in detail, connected with some state or social affairs, and for many people it is not important in what way (method) these numbers were collected. When the collection of numbers is the random sample, it is suitable to use the term "statistical data".

It is important to know that the term "statistics" was introduced at the half of XIX century, mainly by the physicist and astronomer L. A. J. Quetelet. Quetelet

[41] J. Czekanowski, *Zarys metod statystycznych w zastosowaniu do antropologii...*

[42] J. Mikiewicz, *Zastosowanie statystycznej metody wyboru najlepszych obiektów w chemii...*

used methods of probability theory to social and anthropological research, because he also worked for state administration (the word "statistics" originates from the Latin word "status" = state). The basic statistical term "population" also originates from anthropological and social area. In statistics "population" is an abstract term and means the set of objects ordered after one or more quantified features (in one or more – dimensional spaces), from which the elements are to be drawn at random. This set of random elements is called "random sample". The term "population" is a basic notion in mathematical statistics and may not be omitted, though now such a tendency is observed (e.g. [43]). We observe in Nature many material populations and therefore the statistics may be treated as the research method in empirical sciences. "Geometrical probabilities" use material notions as well, e.g. Buffon's needle or probability of waiting for the bus, which shows that essence of probability is related to the material world.

(II) Let us design ecdf by $C_n(x)$, and the mentioned empirical distribution function by $S_n(x)$, both based on the n-elements sample drawn from the population defined by $F(x)$. It is easy to see that for each x and each n $|(Cn(x) - Sn(x)|$ is valid. Because $S_n(x)$ converges monotonously to $F(x)$ when $n \to \infty$ (Glivenko's theorem, V. Glivenko 1933), also $C_n(x)$ did so.

(III) The Lwow shool of Czekanowski was transferred after 1945 year to Wrocław. Latter years were a bad time for "Wrocław taxonomy", and also for the cubes method of A. Wanke[44]. These two methods were elaborated simultaneously with synergism. At first, these two methods were applied mainly in Wrocław to anthropology which was the continuation of research of Lwow anthropological school of J. Czekanowski between the two world wars; this school in Wrocław was attacked and almost annihilated by the communist administration. Moreover, Wrocław mathematicians (it means: the application group) separated themselves from mathematical statistics in taxonomic research, though, as we have shown, the synergism of these methods is very important (compare the difficulties of author in this way[45]).

[43] J.R. Barra, *Mathematical foundations of statistics...*

[44] A. Wanke, *Metoda badania częstości występowania zespołów cech...*

[45] J. Mikiewicz, *Dendrytowe obszary ufności...*; idem, *O poziomach ufności w taksonomii wrocławskiej*, „Zastosowania Matematyki", t. VII, p. 1–40.