OPERATIONS RESEARCH AND DECISIONS

No. 2 2012

DOI: 10.5277/ord120201

Leszek KLUKOWSKI*

ESTIMATORS OF THE RELATIONS OF EQUIVALENCE, TOLERANCE AND PREFERENCE BASED ON PAIRWISE COMPARISONS WITH RANDOM ERRORS

This paper presents a review of results of the author in the area of estimation of the relations of equivalence, tolerance and preference within a finite set based on multiple, independent (in a stochastic way) pairwise comparisons with random errors, in binary and multivalent forms. These estimators require weaker assumptions than those used in the literature on the subject. Estimates of the relations are obtained based on solutions to problems from discrete optimization. They allow application of both types of comparisons – binary and multivalent (this fact relates to the tolerance and preference relations). The estimates can be verified in a statistical way; in particular, it is possible to verify the type of the relation. The estimates have been applied by the author to problems regarding forecasting, financial engineering and bio-cybernetics.

Keywords: *estimation of relations, pairwise comparisons with random errors, nearest adjoining order*

1. Introduction

Estimation of the relations of (i) preference (an alternative of the equivalence relation, i.e. reflexive, transitive, symmetric, and strict preference relation, i.e. transitive, asymmetric), (ii) tolerance (reflexive, symmetric) and (iii) equivalence, within a finite set is applied to the ranking and classification of items – non-overlapping or overlapping – based on multiple pairwise comparisons with random errors. The necessity of determining such relations often appears in research on systems, especially in: financial engineering, prediction and biocybernetics. Many methods and algorithms, based

* Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland, e-mail: Leszek.Klukowski@ibspan.waw

on different methodological ideas, have been developed for these purposes. They are investigated as problems in mathematical statistics, cluster analysis, computational intelligence, rough sets, data mining and other fields (e.g. [6, 7, 11–15, 39, 40, 47]). Some of the methods and algorithms are based on heuristic ideas and, therefore, have weak formal properties. The methods which have good properties usually require strong assumptions about pairwise comparisons. The verification of such assumptions is not always possible (e.g. [2–4, 7, 8, 11]).

In terms of theory and practice, a desirable feature of methods of determining such relations is the conjunction of weak assumptions about errors in comparisons and good statistical properties of estimators. The purpose of this paper is to present methods of estimation and verification of the relations developed by the author which have such good theoretical and practical features. The approach used is general and homogeneous with the possibility of fully automating computations.

The method developed is based on the concept of nearest adjoining order (NAO), formulated by Slater [42] and developed by other authors (see [5, 6] and [7] Chap. 2). The concept of NAO estimators is consistent with the general idea of statistical estimation – to determine the form of the relation which exhibits the minimum difference to the data, i.e. comparisons. Very often, a set of comparisons is not a statistical sample because individual comparisons can have non-identical distributions and may be not independent.

The estimates are obtained based on solutions to discrete programming problems. Algorithms for solving such problems are presented in the literature (e.g. [7, 11, 13]).

This method can be applied in many areas – it is sufficient to be able to compare pairs of elements, with a random error. In particular, the comparisons can be made by statistical tests, experts or computer procedures for comparisons. The elements compared can be: random samples, empirical functions, faces, etc. (see e.g. [21, 24]). The author has recently applied such estimators in research concerning public debt optimization [34] – together with other procedures, e.g. Kohonen neuronal networks.

Problems of the form of estimating relations (classification, partitioning – overlapping or non-overlapping, ranking, etc.) appear constantly in the literature on the subject – monographs, articles, conference papers. The number of publications is huge – an exhaustive review of classical papers is presented in the monographs cited above; new results appear in various journals, e.g. Psychometrika, Journal of Classification and Journal of Marketing Research.

The paper consists of seven sections. The second section presents the main ideas of estimation and validation. The third section presents essential results concerning the form of the estimators and their properties. In the fourth section conclusions from a simulation experiment are discussed. The fifth and sixth sections briefly present methods of validating the estimates and algorithms for solving optimization problems. The last section discusses some original results (see the book by Klukowski [35]).

2. The idea of estimation and validation

The general idea of NAO estimators can be stated as follows: The relations under consideration can be expressed in the form of a family of subsets $\chi_1^{\ell *}, ..., \chi_n^{(\ell)*}, n > 1$, (the index ℓ denotes the relation type: p relates to the preference relation, τ – the tolerance relation, e – the equivalence relation), from a finite set $\mathbf{X} = \{x_1, ..., x_m\}$, $(m \geq 3)$. In the case of the preference relation, the family generates a sequence of subsets (the index of a subset is the rank of its elements). Pairwise comparisons $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ $(\upsilon \in \{b, \mu\}, k = 1, ..., N)$ are given for each pair of elements $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$ (the index υ denotes the kind of comparison, b – binary, μ – multivalent). Any comparison determines the appropriate relation for a pair (x_i, x_j) with the possibility of a random error. The kind of comparison (binary or multivalent) results from the procedure used (statistical test, expert, neuronal network). For example, in the case of the equivalence relation a comparison states whether two elements belong to the same subset $(g_{bk}^{(p)}(x_i, x_j) = 0)$ or to different subsets $(g_{bk}^{(p)}(x_i, x_j) = 1)$. A multivalent comparison can express, in the case of the preference relation, the difference between ranks (a distance within a ranking).

Any relation $\chi_1^{\ell *}, ..., \chi_n^{(\ell)*}$ can be expressed by the values (of a function) $T_\upsilon^{(\ell)}(x_i, x_j)$ which takes the values from the same set as the comparisons. The difference $g_{\upsilon k}^{(p)}(x_i, x_j) - T_\upsilon^{(p)}(x_i, x_j)$ expresses the random error in a comparison (a difference equal to zero means errorless comparison).

The assumptions made about the stochastic properties of the errors in comparisons are non-restrictive. In the case of binary comparisons, it is assumed that the probability of correct comparison is greater than the probability of an error in comparison. In the case of multivalent comparisons, it is assumed that the probability (density) function of errors in comparisons is unimodal with mode and median equal to zero. Additionally, in the case of multiple comparisons $(N > 1)$ of a pair, it is assumed that the individual comparisons $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ $(k = 1, ..., N; \langle i, j \rangle \in R_m)$ are stochastically independent. Thus, the expected values of errors in comparisons can be different from zero and comparisons of pairs including the same element (e.g. $(x_i, x_j), (x_j, x_k), (x_i, x_k)$) can be dependent. Moreover, the distributions of errors can be unknown.

Determination of an estimate of the relation $\chi_1^{(\ell)*}, ..., \chi_n^{(\ell)*}$ is equivalent to finding the form of the relation with minimum difference to comparisons $g_{\upsilon k}^{(p)}(x_i, x_j)$ $(\langle i, j \rangle \in R_m, k = 1, ..., N)$. The estimate is obtained based on optimum solution to a problem from discrete optimization. The simplest form of such a problem, formulated by Slater (for $N = 1$), is equivalent to finding a Hamilton path in a graph [7].

A wide range of generalizations of the Slater formulation, developed by the author, has been presented in [35]. Two kinds of estimators and two kinds of comparisons are examined for each relation (the equivalence relation is an exception – only binary comparisons are considered). The first estimator is based on minimizing the sum of absolute differences between comparisons and the form of the relation (expressed by the function $T_v^{(\ell)}(x_i, x_j)$). The other estimator is based on the medians from the comparisons of each pair.

The statistical properties of the estimators have been derived (the majority of these are proved in earlier papers of the author). Consistency is a fundamental statistical property. In the case of the preference relation, a simulation experiment has been carried out, which shows the precision of the estimators investigated.

The results of the estimation can be thoroughly verified with the use of statistical tests. Investigating the properties of some tests requires carrying out simulations [35]. Such procedures allow, in particular, to test the existence of a relation (under the alternative hypothesis – absence of such a relation, randomness of data, some other data structure, etc.). Thus, the results of estimation, positively validated, are reliable and valuable.

3. The form of estimators and their properties

3.1. Assumptions about errors in comparisons

An equivalence relation $\mathbf{R}^{(e)}$ (reflexive, transitive, symmetric), tolerance relation $\mathbf{R}^{(\tau)}$ (reflexive, symmetric), or preference relation $\mathbf{R}^{(p)}$ (alternative to the equivalence relation together with the strict preference relation) is defined for the set $\mathbf{X}$. Each relation generates some family of subsets $\chi_1^{(\ell)*}, ..., \chi_n^{(\ell)*}$ ($\ell \in \{p, e, \tau\}$; $n \geq 2$).

The equivalence relation generates a family $\chi_1^{(e)*}, ..., \chi_n^{(e)*}$ with the following properties:

$$\bigcup_{q=1}^{n} \chi_q^{(e)*} = \mathbf{X} \tag{1}$$

$$\chi_r^{(e)*} \cap \chi_s^{(e)*} = \{\mathbf{0}\} \tag{2}$$

where:

$$\begin{array}{ll} \mathbf{0} & \text{– the empty set} \\ x_i, x_j \in \chi_r^{(e)*} \equiv x_i, x_j & \text{– equivalent elements} \end{array} \tag{3}$$

non-equivalent elements for $i \neq j, r \neq s$

$$(x_i \in \chi_r^{(e)*}) \cap (x_j \in \chi_s^{(e)*}) \equiv x_i, x_j \tag{4}$$

The tolerance relation generates a family $\chi_1^{(\tau)*}, ..., \chi_n^{(\tau)*}$ with the property (1), i.e. $\bigcup_{q=1}^{n} \chi_q^{(\tau)*} = \mathbf{X}$, and the properties:

$$\begin{gathered} \exists r, s\ (r \neq s) \text{ such that } \chi_r^{(\tau)*} \cap \chi_s^{(\tau)*} \neq \{\mathbf{0}\}, \\ x_i, x_j \in \chi_r^{(\tau)*} \equiv x_i, x_j - \text{ equivalent elements} \end{gathered} \tag{5}$$

$$(x_i \in \chi_r^{(\tau)*}) \cap (x_j \in \chi_s^{(\tau)*}) \equiv x_i, x_j$$

$$\text{non-equivalent elements for } i \neq j \text{ and } (x_i, x_j) \notin \chi_r^{(\tau)*} \cap \chi_s^{(\tau)*} \tag{6}$$

each subset $\chi_r^{(\tau)*}\ (1 \leq r \leq n)$ includes an element x_i such that

$$x_i \notin \chi_s^{(\tau)*}\ (s \neq r) \tag{7}$$

The preference relation generates a family $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$ with the properties (1), (2) and the property:

$$(x_i \in \chi_r^{(p)*}) \cap (x_j \in \chi_s^{(p)*}) \equiv x_i \text{ is preferred to } x_j \text{ for } r < s \tag{8}$$

The relations defined by the conditions (1)–(8) can be expressed alternatively by the values (functions) $T_{\upsilon}^{(\ell)}(x_i, x_j)$ $((x_i, x_j) \in \mathbf{X} \times \mathbf{X};\ \ell \in \{p, e, \tau\},\ \upsilon \in \{b, \mu\})$; the symbols b and μ denote binary and multivalent comparisons, respectively, defined as follows:

$$T_b^{(e)}(x_i, x_j) = \begin{cases} 0 \text{ if there exists } r \text{ such that } (x_i, x_j) \in \chi_r^{(e)*} \\ 1 \text{ otherwise} \end{cases} \tag{9}$$

• Assuming binary comparisons, the function $T_b^{(e)}(x_i, x_j)$, describing the equivalence relation, expresses whether a pair (x_i, x_j) belongs to a common subset or not

$$T_b^{(\tau)}(x_i, x_j) = \begin{cases} 0 \text{ if there exists } r, s \ (r = s \text{ not excluded}) \text{ such that} \\ (x_i, x_j) \in \chi_r^{(\tau)*} \cap \chi_s^{(\tau)*} \\ 1 \text{ otherwise} \end{cases} \tag{10}$$

• Assuming binary comparisons, the function $T_b^{(\tau)}(x_i, x_j)$, describing the tolerance relation, expresses whether a pair (x_i, x_j) belongs to any conjunction of subsets (or the same subset) or not; the condition (7) guarantees the uniqueness of the description

$$T_\mu^{(\tau)}(x_i, x_j) = \#(\Omega_i^* \cap \Omega_j^*) \tag{11}$$

where: Ω_l^* – a set of the form $\Omega_l^* = \{s \mid \ x_l \in \chi_s^{(\tau)*}\}$, $\#(\Xi)$ – the number of elements in the set Ξ.

• Assuming multivalent comparisons, the function $T_\mu^{(\tau)}(x_i, x_j)$, describing the tolerance relation, expresses the number of subsets of conjunction including both elements; condition (7) guarantees the uniqueness of the description

$$T_b^{(p)}(x_i, x_j) = \begin{cases} 0 \text{ if there exists } r \text{ such that } (x_i, x_j) \in \chi_r^{(p)*} \\ -1 \text{ if } \ x_i \in \chi_r^{(p)*},\ x_j \in \chi_s^{(p)*} \text{ and } r < s \\ 1 \ \text{ if } \ x_i \in \chi_r^{(p)*},\ x_j \in \chi_s^{(p)*} \text{ and } r > s \end{cases} \tag{12}$$

• Assuming binary comparisons, the function $T_b^{(p)}(x_i, x_j)$, describing the preference relation, expresses the direction of preference in a pair of elements or their equivalence

$$T_\mu^{(p)}(x_i, x_j) = d_{ij} \Leftrightarrow x_i \in \chi_r^{(p)*},\ \ x_j \in \chi_s^{(p)*},\ \ d_{ij} = r - s \tag{13}$$

• Assuming multivalent comparisons, the function $T_\mu^{(p)}(x_i, x_j)$, describing the preference relation, expresses the difference between the ranks of the elements x_i and x_j.

The relation $\chi_1^{(\ell)*}, ..., \chi_n^{(\ell)*}$ is to be estimated based on N ($N \geq 1$) comparisons of each pair $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$; any comparison $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ evaluates the actual value of $T_\upsilon^{(\ell)}(x_i, x_j)$ and can be perturbed by a random error. The following assumptions are made about the errors in comparisons:

A1. The relation type, i.e.: equivalence, tolerance or preference, is known, the number of subsets n – unknown.

A2. Any comparison $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ $(\ell \in \{e, \tau, p\};\quad \upsilon \in \{b, \mu\};\quad k = 1, ..., N)$, is an evaluation of the value $T_{\upsilon}^{(\ell)}(x_i, x_j)$, perturbed by a random error. The probabilities of errors $g_{\upsilon k}^{(\ell)}(x_i, x_j) - T_{\upsilon}^{(\ell)}(x_i, x_j)$ have to satisfy the following assumptions:

$$P(g_{bk}^{(\ell)}(x_i, x_j) - T_b^{(\ell)}(x_i, x_j) = 0 \mid T_b^{(\ell)}(x_i, x_j) = \kappa_{bij}^{(\ell)}) \geq 1 - \delta \quad \left(\kappa_{bij}^{(\ell)} \in \{-1, 0, 1\},\quad \delta \in \left(0, \frac{1}{2}\right)\right) \tag{14}$$

$$\sum_{r \leq 0} P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) > \frac{1}{2} \qquad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., \pm m\} \tag{15}$$

r – an integer number,

$$\sum_{r \geq 0} P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = -r \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) > \frac{1}{2} \quad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., \pm m\} \tag{16}$$

r – an integer number,

$$P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r) \geq P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r + 1 \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) \quad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., m\},\ r > 0) \tag{17}$$

$$P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r) \geq P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r - 1 \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) \quad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., m\},\ r < 0) \tag{18}$$

A3. The comparisons $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ $(\ell \in \{e, \tau, p\}$; $\upsilon \in \{b, \mu\};\ k = 1, ..., N)$ are independent random variables.

Assumption A3 allows determination of the distributions of the errors in estimation (see the next section). However, determination of the exact distributions of the (multidimensional) errors is complicated and, in practice, unrealizable in an analytic way. The main properties of the estimators, in particular their consistency, are valid without this assumption.

Assumption A3 can be relaxed in the following way: the comparisons $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ and $g_{\upsilon l}^{(\ell)}(x_r, x_s)$ $(l \neq k; r \neq i, j; s \neq i, j)$, i.e. including different elements, have to be independent.

In the case of the weak preference relation (i.e. including the possibility of equivalent elements), condition (14) can be relaxed to the form (15), (16).

Assumptions A2, A3 reflect the following properties of the distributions of errors:

• In the case of binary comparisons, the probability of correct comparison is greater than that of incorrect comparison (Inequality (14)).

• Zero is a median of any distribution of the error in the comparison (Inequalities (14)–(16)),

• Zero is a mode of any distribution of the error in comparison (Inequalities (14)–(18)).

• The set of all comparisons consists of realizations of independent random variables.

• The expected value of any error in the comparison may differ from zero.

The assumptions about the distributions of errors in comparisons are not restrictive. In particular, the errors can have non-zero expected values; the distribution of errors in comparisons has only to satisfy the mode and median condition. These features guarantee a broad spectrum of applications and protect against incorrect results.

3.2. The form of estimators – the idea of minimization of the differences between comparisons and the form of the relation

As mentioned above, two kinds of estimators have been examined – the first one based on the total sum of absolute differences between the form of the relation expressed by the values $T_\upsilon^{(\ell)}(x_i, x_j)$ and comparisons, and the second, based on differences between the medians from comparisons of each pair. The properties of these estimators have been determined by the author based on probabilistic inequalities (in particular Hoeffding's inequality [16] and Chebyshev's inequality), properties of the sample median [6] and results indicating the convergence to zero of the variances of the random variables examined.

The estimator based on the number of absolute differences, denoted by the symbol $\hat{\chi}_1^{(\ell)}, \ldots, \hat{\chi}_{\hat{n}}^{(\ell)}$ (or $\hat{T}_\upsilon^{(\ell)}(x_i, x_j)$ $\langle i, j\rangle \in R_m$), assumes the form of the solution (or solutions) to the discrete optimization problem:

$$\min_{\chi_1^{(\ell)}, \ldots, \chi_r^{(\ell)} \in F_{\mathbf{X}}^{(\ell)}} \left\{ \sum_{\langle i,j\rangle \in R_m} \sum_{k=1}^{N} \left| g_{\upsilon k}^{(\ell)}(x_i, x_j) - t_\upsilon^{(\ell)}(x_i, x_j) \right| \right\} \tag{19}$$

where: $F_{\mathbf{X}}^{(\ell)}$ – the feasible set, i.e. all families $\chi_1^{(\ell)}, \ldots, \chi_r^{(\ell)}$ in the set $\mathbf{X}$, $t_\upsilon^{(\ell)}(x_i, x_j)$ – a function describing the relation $\{\chi_1^{(\ell)}, \ldots, \chi_r^{(\ell)}\}$, R_m – a set of the form $R_m = \{\langle i, j\rangle \,|\, 1 \le i, j \le m;\ j > i\}$ (note that $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ denotes both the random variable and its realization).

In the case of the preference relation and binary comparisons, the following transformation, which simplifies the optimization problem (19), can be applied:

$$\theta(g_{\upsilon k}^{(\ell)}(x_i,x_j)-t_\upsilon^{(\ell)}(x_i,x_j))=\begin{cases}0 \text{ if } g_{\upsilon k}^{(\ell)}(x_i,x_j)=t_\upsilon^{(\ell)}(x_i,x_j)\\ 1 \text{ if } g_{\upsilon k}^{(\ell)}(x_i,x_j)\neq t_\upsilon^{(\ell)}(x_i,x_j)\end{cases} \tag{19a}$$

After transformation (19a), the criterion function (19) expresses the number of inconsistencies between comparisons and the values $T_b^{(p)}(x_i,x_j)$; he transformation reduces the number of variables in the problem and does not change the properties of the estimates [20, 22].

The estimator $\hat{\chi}_1^{(\ell)},\ldots,\hat{\chi}_r^{(\ell)}$ ($\hat{T}_\upsilon^{(\ell)}(x_i,x_j)$) based on the medians from the comparisons of individual pairs, is obtained based on the discrete optimization problem:

$$\min \sum_{\chi_1^{(\ell)},\ldots,\chi_r^{(\ell)}\in F_{\mathbf{X}}} \left|g_\upsilon^{(\ell,me)}(x_i,x_j)-t_\upsilon^{(\ell)}(x_i,x_j)\right| \tag{20}$$

where: $g_\upsilon^{(\ell,me)}(x_i,x_j)$ – the median from comparisons $\{g_{\upsilon,1}^{(\ell)}(x_i,x_j),\ldots,g_{\upsilon N}^{(\ell)}(x_i,x_j)\}$. The binary transformation (19a) can also be applied in the case of the median estimator.

The number of solutions, obtained based on optimization problems (19), (20) can exceed one. A unique estimate can be determined from this set of solutions in a random way or as a result of validation (see Section 5). The minimum possible value of the criterion functions (19), (20) equals zero; such a value indicates a perfect goodness of fit (of the comparisons and the estimate obtained). This eliminates the necessity of validation (one exception is the case of multiple estimates).

Assumptions A1–A3 enable inference about the distributions of errors in estimates. Let us discuss firstly the estimator based on the criterion (19). For each relation type one can determine a finite set including all possible realizations of comparisons

$$g_{\upsilon k}^{(\ell)}(x_i,x_j),\ \ (\ell\in\{e,\tau,p\},\upsilon\in\{b,\mu\},k=1,\ldots,N;\ \langle i,j\rangle\in R_m)$$

and the probability of each realization. Therefore, the use of criterion (19) determines: the estimate, its probability and estimation error. The error has the form: $\{\hat{T}_\upsilon^{(\ell)}(x_i,x_j)-T_\upsilon^{(\ell)}(x_i,x_j);\ \langle i,j\rangle\in R_m\}$, i.e. it is a multidimensional random variable. Thus, the estimator has a distribution.

The analysis of a multidimensional error is, in practice, unrealizable and it is suggested to replace it with the one-dimension error:

$$\hat{\Delta}_\upsilon^{(\ell)}=\sum_{\langle i,j\rangle\in R_m}\left|\hat{T}_\upsilon^{(\ell)}(x_i,x_j)-T_\upsilon^{(\ell)}(x_i,x_j)\right| \tag{21}$$

When the error $\hat{\Delta}_{\upsilon}^{(\ell)} = 0$, this indicates an errorless estimate; low values of the error indicate an estimate near to the actual form of the relation. The distribution of the estimator based on medians is defined in a similar way.

The distributions of estimators can also be determined in the case of comparisons which are dependent when the form of such dependence is known. Determining the distributions of the estimators in an analytic way, even when the distributions of the errors in comparisons are known, is difficult due to the complex form of the set of elementary events; therefore such distributions can be analyzed using simulations. Similar considerations are valid in the case of the estimator based on the medians from comparisons.

3.3. Properties of estimators

Analytical properties of the estimators established by the author have mainly an asymptotic character, i.e. they apply to the case $N \to \infty$. These properties guarantee a basic feature of the estimators – consistency. It is clear that errorless estimates can also be obtained for finite *N* with probability close to one, because the feasible set in the optimization problems (19), (20) is huge, but finite. In general, the precision of the estimates depends not only on *N*, but also on the distributions of the errors in comparison and certain features of the relation, e.g. the number of subsets *n* and the number of elements in each subset. There is also a difference in the precision level of the two estimators considered.

The analytical properties of the estimators are based on the properties of the random variables expressing the differences between pairwise comparisons and the form of the relation (values of $T_{\upsilon}^{(\ell)}(x_i, x_j)$). It has been proved that the random variables expressing these differences, i.e.:

$$W_{\upsilon N}^{(\ell)*} = \sum_{\langle i,j\rangle \in R_m} \sum_{k=1}^{N} \left| g_{\upsilon k}^{(\ell)}(x_i, x_j) - T_{\upsilon}^{(\ell)}(x_i, x_j) \right|$$

or

$$W_{\upsilon N}^{(\ell,me)*} = \sum_{\langle i,j\rangle \in R_m} \left| g_{\upsilon}^{(\ell,me)}(x_i, x_j) - T_{\upsilon}^{(\ell)}(x_i, x_j) \right|$$

have different properties to the variables

$$\tilde{W}_{\upsilon N}^{(\ell)} = \sum_{\langle i,j\rangle \in R_m} \sum_{k=1}^{N} g_{\upsilon k}^{(\ell)}(x_i, x_j) - \tilde{T}_{\upsilon}^{(\ell)}(x_i, x_j) \Big|$$

and

$$\tilde{W}_{\upsilon N}^{(\ell,me)} = \sum_{\langle i,j\rangle \in R_m} \left| g_{\upsilon}^{(\ell,me)}(x_i,x_j) - \tilde{T}_{\upsilon}^{(\ell)}(x_i,x_j) \right|$$

expressing the differences between comparisons and any other relation. The differences defining these random variables are defined in the criterion functions (19), (20).

The following main results have been obtained:

I. The expected values of the random variables corresponding to the difference from the actual form of the relation are smaller than the expected values of the variables corresponding to the difference from any other relation, i.e. $E(W_{\upsilon N}^{(\ell)*}) < E(\tilde{W}_{\upsilon N}^{(\ell)})$ and $E(W_{\upsilon N}^{(\ell,me)*}) < E(\tilde{W}_{\upsilon N}^{(\ell,me)})$.

II. The variances of the variables expressing the sum of differences between comparisons and the form of a relation, both the actual or any other relation, divided by the number of comparisons, N, converge to zero as $N \to \infty$, i.e. $\lim_{N\to\infty} \mathrm{Var}((1/N)W_{\upsilon N}^{(\ell)*}) = 0$ and $\lim_{N\to\infty} \mathrm{Var}((1/N)\tilde{W}_{\upsilon N}^{(\ell)}) = 0$.

III. The variances of the variables expressing the sum of differences between the medians from comparisons of each pair and the form of the relation converge to zero as $N \to \infty$, i.e. $\lim_{N\to\infty} \mathrm{Var}(W_{\upsilon N}^{(\ell,me)*}) = 0$, $\lim_{N\to\infty} \mathrm{Var}(\tilde{W}_{\upsilon N}^{(\ell,me)}) = 0$.

IV. The probability of the event that the (random) variable expressing the sum of differences between comparisons and the actual form of the relation takes a smaller value than the variable expressing the sum of differences between comparisons and any other form of relation converges to one as $N \to \infty$; the speed of convergence, as determined by the coefficient of N in the corresponding exponential function, guarantees the good efficiency of the estimator i.e.:

$$P(W_{\upsilon N}^{(\ell)*} < \tilde{W}_{\upsilon N}^{(\ell)}) \geq 1 - \exp\{-2\theta N\} \tag{22}$$

θ – a positive constant.

In the case of binary comparisons we have:

$$P(W_{bN}^{(\ell,me)*} < \tilde{W}_{bN}^{(\ell,me)}) \geq 1 - \exp\left\{-2N\left(\frac{1}{2} - \delta\right)^2\right\} \quad \left(\ell \in \{e,p,\tau\},\ \ \delta \in \left(0, \frac{1}{2}\right)\right) \tag{23}$$

V. The probability of the event that the variable expressing the sum of differences between medians from comparisons and the actual form of the relation takes a lower value than the variable expressing the sum of differences between comparisons and any other form of relation converges to one as $N \to \infty$; the speed of convergence guarantees the good efficiency of the estimator. In the case of binary comparisons, the following relation is true:

$$P(W_{bN}^{(\ell,me)*} < \tilde{W}_{bN}^{(\ell,me)}) \geq 1 - 2\exp\left\{-2N\left(\frac{1}{2}-\delta\right)^2\right\} \quad \left(\ell \in \{e,p,\tau\}, \quad \delta \in \left(0,\frac{1}{2}\right)\right) \tag{24}$$

The properties (I)–(V) are the basis for the construction of the estimators and indicate their consistency (the inequalities: $E(W_{\upsilon N}^{(\ell)*}) < E(\tilde{W}_{\upsilon N}^{(\ell)})$ and $E(W_{\upsilon N}^{(\ell,me)*}) < E(\tilde{W}_{\upsilon N}^{(\ell,me)})$, convergence of the variances: $\lim_{N\to\infty}\mathrm{Var}(W_{\upsilon N}^{(\ell,me)*}) = 0$, $\lim_{N\to\infty}\mathrm{Var}(\tilde{W}_{\upsilon N}^{(\ell,me)}) = 0$, and the inequalities: $P(W_{\upsilon N}^{(\ell)*} < \tilde{W}_{\upsilon N}^{(\ell)})$, $P(W_{\upsilon N}^{(\ell,me)*} < \tilde{W}_{\upsilon N}^{(\ell,me)})$). Investigation of the precision of estimators and the speed of their convergence to the actual form of the relationship requires simulation. One important conclusion from such a simulation experiment is the confirmation of the higher efficiency of the estimator based on the total number of differences, for the values of N and δ examined.

Properties (I)–(V), proved for both estimators and both types of comparisons, are original, theoretical results of the author. Simulations are a valuable complement to the theoretical results.

4. Examination of the properties of the estimators of a preference relation using simulations

4.1. Purpose of the simulations

The purpose of the simulations is to determine the distributions of the errors of both estimators considered in the case of the preference relation. These distributions allow us not only to evaluate the precision of estimates but also to determine parameters for the comparisons, in particular their number N, guaranteeing the required precision of estimates. They are also necessary for the validation of estimates and testing hypotheses about the form of a relation.

4.2. Parameters of the simulations

The simulations were carried out for three forms of relation, with the use of the following parameters.

- The set **X** contains nine elements;
- three forms of relation:

– relation with nine subsets (linear order): $\{x_1\}, ..., \{x_9\}$ $(n = 9)$,

– relation with six subsets: $\{x_1\}, \{x_2, x_3\}, \{x_4\}, \{x_5, x_6\}, \{x_7\}, \{x_8, x_9\}$ (n = 6),
– relation with three subsets: $\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6, x_7, x_8, x_9\}$ (n = 3);
• binary comparisons with probability functions:

$$P(g_{bk}^{(p)}(x_i, x_j) = T_b^{(p)}(x_i, x_j)) = \alpha_{ij}, \;\; P(g_{bk}^{(p)}(x_i, x_j) \neq T_b^{(p)}(x_i, x_j)) = \frac{1-\alpha_{ij}}{2}$$

with three values of α_{ij}: 0.85, 0.90, 0.95 for all $\langle i, j\rangle$ (typical levels in statistical tests);
• multivalent comparisons with probability functions:

$$P(g_{\mu k}^{(p)}(x_i, x_j) = T_\mu^{(p)}(x_i, x_j)) = \alpha_{ij}$$

$$P(g_{\mu k}^{(p)}(x_i, x_j) - T_\mu^{(p)}(x_i, x_j) = -l) = \frac{1-\alpha_{ij}}{2L_{ij}^{(d)}}$$

$$(L_{ij}^{(d)} = T_\mu^{(p)}(x_i, x_j) + (m-1); \;\; l = -1, ..., -L_{ij}^{(d)})$$

$$P(g_{\mu k}^{(p)}(x_i, x_j) - T_\mu^{(p)}(x_i, x_j) = l) = \frac{1-\alpha_{ij}}{2L_{ij}^{(u)}}$$

$$(L_{ij}^{(u)} = m-1-T_\mu^{(p)}(x_i, x_j); \;\; l = 1, ..., L_{ij}^{(u)})$$

with three values of α_{ij}: 0.3334; 0.4167; 0.5000 (i.e. approximately 4/12, 5/12, 6/12) for all $\langle i, j\rangle$;
• number N of comparisons of each pair: 1, 3, 5, 7, 9.

The total number of cases analyzed equals 90 for each type of probability function, i.e.: three forms of relation, three forms of distributions of errors in comparisons, five values of the number of comparisons N, using both estimators. Each case has been simulated 100 or 200 times, using a random number generator. The book by Klukowski [35] also considers the case of error distributions with the property that increasing values of errors correspond to decreasing probabilities.

4.3. Results of the simulations. Conclusions

The results of the simulations have been analyzed in detail by Klukowski [35]. At this point, some examples of the results from the simulations (Tables 1, 2) and the basic conclusions are presented.

Table 1. The efficiency of estimators based on binary comparisons, $n = 9$ subsets

Number of comparisons *N*	Quantities	Probability of correct comparison					
		0.85		0.90		0.95	
		Sum.	Median.	Sum.	Median.	Sum.	Median.
1	% CR	20	20	29	29	49	49
	% of CRM	26	26	38	38	60	60
	AE	4.20	4.20	2.78	2.78	1.41	1.41
3	% CR	53	49	77	75	97	96
	% of CRM	56	55	78	80	97	97
	AE	0.88	1.13	0.38	0.45	0.03	0.03
5	% CR	82	82	92	92	99	99
	% of CRM	82	82	92	92	99	99
	AE	0.28	9.30	0.11	0.11	0.01	0.01
7	% CR	91	91	97	97	100	100
	% of CRM	91	91	97	97	100	100
	AE	0.10	0.10	0.03	0.03	0	0
9	% CR	95	95	100	100	100	100
	% of CRM	95	95	100	100	100	100
	AE	0.05	0.05	0	0	0	0

Computations by the author. % CR is the fraction of errorless singular estimates, % of CRM – the fraction of errorless estimates taking into account multiple solutions, AE – average estimation error, taking into account multiple solutions.

Table 2. The efficiency of estimators
based on multivalent comparisons, $n = 9$ subsets

Number of comparisons *N*	Quantity	Probability of correct comparison:					
		0.3334		0.4167		0.5000	
		Sum.	Median.	Sum.	Median.	Sum.	Median.
1	% of CR	17	17	31	31	60	60
	% of CRM	32	32	51	51	78	78
	AE	43.68	43.68	27.70	27.70	12.50	12.50
3	% of CR	78	58	91	79	97	92
	% of CRM	85	74	95	93	100	98
	AE	6.09	12.35	1.11	3.39	0.26	1.19
5	% of CR	95	71	99	97	100	100
	% of CRM	98	93	100	100	100	100
	AE	0.72	4.0	0.08	0.20	0	0
7	% of CR	98	92	100	100	100	100
	% of CRM	99	99	100	100	100	100
	AE	0.12	0.55	0	0	0	0
9	% of CR	100	100	100	100	100	100
	% of CRM	100	100	100	100	100	100
	AE	0	0	0	0	0	0

Computations by the author. Denotations the same as in Table 1.

General conclusions about the precision of estimators based on binary comparisons are as follows:

- Both estimators guarantee acceptable precision of estimation in the case when the probability δ does not exceed a typical significance level in statistical tests, i.e. $\delta \leq 0.10$, and the number of comparisons is greater or equal three, $N \geq 3$.
- An increase in the number of comparisons N indicates a rapid increase in the precision of the estimates, reflecting the exponential form of inequality (22).
- The average errors of the estimator based on the total number of differences are significantly smaller than for the estimator based on medians; these differences are not significant when both estimators guarantee errorless estimates with the probability greater than 0.95.
- The estimator based on medians requires at least $N + 2$ comparisons to obtain a precision similar to the estimator based on the total number of differences with N comparisons; it guarantees acceptable results in the cases: (i) $\alpha \geq 0.95$ and $N \geq 3$ and (ii) $\alpha \geq 0.85$ and $N \geq 5$; it also generates a greater number of multiple estimates.
- The best precision of estimation is achieved for the relation with three subsets.

General conclusions about the precision of estimators based on multivalent comparisons are as follows:

- Acceptable precision of estimation is achieved in the case when the probability of errorless comparisons is approximately equal to 1/2 of the probability of errorless binary comparison, i.e.: 4/12, 5/12, 6/12, in the case of multiple comparisons ($N > 1$).
- An estimator based on the sum of differences guarantees a significantly higher concentration of errors around zero than the estimator based on medians.
- An estimator based on medians requires at least two more comparisons than the one based on differences to achieve the same accuracy.
- The best precision is achieved for $n = m$ (for binary comparisons the worst results were obtained in that case).
- An increase in the number of comparisons results in a rapid increase in the precision of estimates and concentration of the frequencies of errors around zero.

The results of the simulation confirm good properties of the estimators proposed. The highest level of precision is guaranteed by the estimator based on the sum of differences between comparisons and the form of the relation based on multivalent comparisons. Examination of relations in sets with a greater number of elements ($m > 9$), shows that the above properties remain valid.

5. Validation of estimates

The estimators of the relations considered were constructed under assumptions A1–A3; the first of them states the existence of the relation in the set $\mathbf{X}$, the other two concern properties of the distributions of errors in comparisons. These assumptions can be verified with the use of statistical tests; a positive verification validates the estimates obtained and guarantees high reliability of the estimators.

Firstly, we checked the assumptions about the errors in comparisons, i.e. unimodality and independence of comparisons of the same pair and the mode and median being equal to zero. For this purpose, well known tests can be applied [9, 17, 43–45] They are based on comparisons $g_{\upsilon,1}^{(\ell)}(x_i,x_j)$, ..., $g_{\upsilon N}^{(\ell)}(x_i,x_j)$, $(\upsilon\in\{b,\mu\}$, $\langle i,j\rangle\in R_m$, $\ell\in\{e,\tau,p\})$ or differences: $g_{\upsilon k}^{(\ell)}(x_i,x_j)-\hat{T}_\upsilon^{(\ell)}(x_i,x_j)$ or $g_{\upsilon k}^{(\ell)}(x_i,x_j)-\widehat{T}_\upsilon^{(\ell)}(x_i,x_j)$, $(k=1,...,N)$.

The verification of the existence of the relation can only be made after positive verification of the assumptions about errors in comparisons. The hypothesis to be tested states the existence of a relation, the alternative states either the equivalence of all the elements, randomness of comparisons or some other data structure. Such tests can be based on the values of the criterion functions (19), (20); large realizations of these functions indicate a significant difference between the form of the relation and comparisons and thus imply rejecting the null-hypothesis. Critical values for such tests can be obtained based on simulations.

It should be noted that many various tests can be used to verify the existence of a particular type of relation (see also [7, 11]. For example, in the case of the preference relation, we can verify the consistency of the ranks of elements obtained based on sequential subsets of comparisons: $g_{\upsilon,1}^{(\ell)}(x_i,x_j)$, ..., $g_{\upsilon N}^{(\ell)}(x_i,x_j)$ $(\langle i,j\rangle\in R_m)$.

Tests for distinguishing between the tolerance and equivalence relations and tests for distinguishing between the strong and weak form of the preference relation [35, 36] can also be used to validate the results of estimation.

6. Algorithms for solving optimization problems

Minimization of the functions (19), (20) is a difficult problem because of huge (but known) computational cost. Currently, there exist algorithms for binary comparisons ([7] Chapt. 2, [12]) based on dynamic programming and branch and bound methods. Some of them require a known number of subsets n and, additionally, are efficient in the case of sets with a moderate number of elements m. In the case of sets which do not exceed several elements, complete enumeration can be used. Large sets require the application of heuristic algorithms, e.g. genetic [10], neuronal networks, random search [42], swarm intelligence [1].

In the case of multivalent comparisons, there are no algorithms to derive the exact solution, except complete enumeration. In general, approximate algorithms have to be used unless m is small. One may expect that the application of new, efficient quantum computers will enable us to find exact solutions of problems when sets are large.

7. Conclusions – new results of the work

Original results have been presented by the author in the subject of estimation of tsuch relations as equivalence, tolerance and preference over a finite set based on multiple pairwise comparisons with random errors.

The main results may be summarized as follows [42]:

• Two types of data have been taken into account: binary and multivalent. Binary data reflect qualitative features of the compared pairs of elements, i.e. equivalence or direction of preference within a pair, while multivalent data reflect quantitative features, i.e. the number of subsets including both elements (tolerance relation) or distance between elements in the form of difference between ranks (preference relation).

• The assumptions concerning errors in comparisons are weaker than those commonly used in the literature, in particular:

– the expected values of errors in comparisons can differ from zero,

– distributions of errors in comparisons may be unknown,

– comparisons including the same element can be correlated.

Therefore, the algorithms proposed can be used when other algorithms are not applicable (may produce incorrect results).

• Two estimators have been examined; the former one is based on the sum of differences between the form of the relation and the data from the comparisons, the latter is based on differences between the form of the relation and the median from comparisons of each pair. The estimators have a simple intuitive form, i.e. are solutions to the appropriate optimization problem, and have analytical properties guaranteeing good efficiency, especially in the case of multiple comparisons of each pair. The properties indicate, in particular, that the efficiency of the first estimator is better, but involves a higher computational cost. The median estimator requires lower computational cost when applying optimization algorithms, and is more robust (robustness is an important property in the case of multivalent comparisons).

• The analytical properties of the estimators have been complemented with the results of a simulation study. This allows the determination of the values of parameters, especially the number of comparisons N, guaranteeing the required precision of estimates; a sufficiently large N guarantees that the probability of an errorless result is close to one. The simulation approach also allows estimation of the probability of an errorless solution in the case when the distribution of errors in comparisons is unknown. Such distributions are replaced by a class of boundary distributions – quasiuniform distributions, proposed by the author. The simulation study indicates the excellent efficiency of multivalent estimators – an original concept of the author; errorless estimates can be obtained with probability close to one for moderate N when the probability of errorless comparison is lower than 1/2.

• The properties of estimates can be thoroughly validated; validation encompasses verification of the existence of the relation and the assumptions regarding the errors in comparisons. These assumptions can be verified with the use of known tests and the methods proposed by the author. The establishment of the existence of the relation can also be based on a simulation approach. It is possible, as well, to choose the relation type – equivalence or tolerance, and the type of preference relation – strict or weak. Therefore, the approach has features of data mining techniques.

• The precision of the estimators, examined in the simulation study [35, 36] is based on measures described in the paper: (i) frequency of errorless estimation, (ii) average absolute, one-dimensional error, and (iii) distribution of the average absolute, one-dimensional error. One-dimensional error is an appropriate measure of the difference between the estimate and the relation; however, multi-dimensional error can also be the subject of analysis, especially in a graphical form.

• The approach proposed allows combining comparisons obtained from various sources, e.g. statistical tests, experts, neural networks. It is also possible to combine binary and multivalent comparisons and to apply two-stage estimators, based, in the first stage, on binary comparisons, and in the second stage – on multivalent comparisons obtained from the first stage.

• The estimates are obtained based on solutions to optimization problems. They can be solved with the use of complete enumeration of the feasible set or heuristic algorithms. The first approach requires fast processors, which are currently available. Heuristic algorithms can be based on random search, genetic algorithms, swarm intelligence, or hierarchical agglomeration algorithms.

• The approach presented will be developed in the following directions: statistical learning, estimation of more complex data structures (e.g. hierarchical), multidimensional (multi-criteria) pairwise comparisons, etc. An important field is also constituted by application of the estimators and tests developed.

To summarize, a comprehensive method has been presented for estimation and validation of the results based on algorithms from: statistics, optimization and simulation. It is a significant contribution to the areas of: mathematical statistics, statistical computer systems and data mining. It provides results with good practical properties, i.e. non-restrictive assumptions, high precision and reliability. The methodology can be developed in many directions, in particular: more complex data structures (e.g. hierarchical), multidimensional comparisons and statistical learning [15, 40]. Another important area are applications, especially in the field of financial engineering.

References

[1] *Swarm Intelligence in Data Mining*, A. Abraham, C. Grosan (Eds.), Springer, Berlin 2006.
[2] BRADLEY R.A., *Science, Statistics and Paired Comparisons*, Biometrics, 1976, 32, 213–232.
[3] BRADLEY R.A., *Paired comparisons: some basic procedures and examples*, [in:] *Handbook of Statistics 4*, P.R. Krishnaiah, P.K. Se (Eds.), North-Holland, Amsterdam 1984, 299–326.
[4] BRUNK H.D., *Mathematical models for ranking from paired comparisons*, JASA, 1960, 55, 503–520.
[5] DANIEL W.W., *Applied Nonparametric Statistics* (2nd Ed.), PWS-Kent Publishing Company, Boston 1990.
[6] DAVID H.A., *Order Statistics*, Wiley, New York 1970.
[7] DAVID H.A., *The Method of Paired Comparisons*, 2nd ed. C. Griffin, London 1988.
[8] DAVIDSON R.R., *A bibliography on the Method of Paired Comparisons*, Biometrics, 1976, 32, 241–251.
[9] DOMAŃSKI C., *Statistical Tests*, PWE, Warszawa 1990 (in Polish).
[10] FALKENAUER E., *Genetic Algorithms and Grouping Data*, Wiley, New York 1998.
[11] GORDON A.D., *Classification*, 2nd Ed., Chapman & Hall/CRC, Boca Raton, 1999.
[12] HAND D.J., *Discrimination and Classification*, Wiley, New York 1986.
[13] HANSEN P., JAUMARD B., SANLAVILLE E., *Partitioning Problems in Cluster Analysis: A Review of Mathematical Programming Approaches*, *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, Berlin 1994.
[14] HARTIGAN J.J., *Clustering Algorithms*, Wiley, New York 1975.
[15] HASTIE T., TIBSHIRANI R., FRIEDMAN J., *The Elements of Statistical Learning. Data Mining, Inference and Prediction*, Springer, Berlin 2002.
[16] HOEFFDING W., *Probability inequalities for sums of bounded random variables*. JASA, 1963, 58, 13–30.
[17] HOLLANDER M., WOLFE D.A., *Nonparametric Statistical Methods*, Wiley, 1973.
[18] KLUKOWSKI L., *A Method of Forecasting Time Series Including a Piecewise Linear Component*, (unpublished Ph.D. dissertation), Systems Research Institute, Polish Academy of Sciences, Warsaw 1986 (in Polish).
[19] KLUKOWSKI L., *Algorithm for the classification of samples in the case of an unknown number of random variables generating them*, Przegląd Statystyczny, 1990, 37, 167–177 (in Polish).
[20] KLUKOWSKI L., *Ranking items on the basis of pairwise comparisons in the case of random errors*, [in:] Systems Research, 3, R. Kulikowski, J. Kacprzyk (Eds.), Omnitech Press, Warsaw 1990, 212–268 (in Polish).
[21] KLUKOWSKI L., *Classification of hormonal profiles*, II Conf. on Operational and Systems Research, Z. Nahorski (Ed.), IBS PAN, Warsaw 1991, 155–163 (in Polish).
[22] KLUKOWSKI L., *Some probabilistic properties of the nearest adjoining order method and its extensions*, Annals of Operational Research, 1994, 51, 241–261.
[23] KLUKOWSKI L., *The nearest adjoining order method for pairwise comparisons in the form of difference of ranks*, Annals of Operations Research, 2000, 97, 357–378.
[24] KLUKOWSKI L., *Tests for relation type – equivalence or tolerance – in a finite set of elements*, Control and Cybernetics, 2006, 35, 369–384.
[25] KLUKOWSKI L., *Estimation of the preference relation on the basis of medians from pairwise comparisons in the form of difference of ranks*, [in:] *Computer Recognition Systems 2*, M. Kurzynski, E. Puchala, M. Wozniak, A. Zolnierek (Eds.), Advances in Soft Computing, Springer, 2007, 45, 232–241.
[26] KLUKOWSKI L., *Estimation of tolerance relation on the basis of multiple pairwise comparisons with random errors*, Control and Cybernetics, 2007, 36, 443–466.
[27] KLUKOWSKI L., *Estimation of the preference relation on the basis of multiple pairwise comparisons with random errors*, [in:] *Modelling of Preferences vs. Risk*, T. Trzaskalik (Ed.), University of Economics in Katowice, Katowice 2008, 61–76 (in Polish).

[28] KLUKOWSKI L., *Determination of tolerance relation – alternative approach to intuitionistic and fuzzy sets*, [in:] *Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets an Related Topics*, Vol. 2, K. Atanassov, O. Hryniewicz, J. Kacprzyk, M. Krawczak, Z. Nahorski, S. Zadrożny (Eds.), Academic Publishing House Exit, Warsaw 2008.
[29] KLUKOWSKI L., *Estimation of the preference relation on the basis of multiple pairwise comparisons in the form of differences of ranks*, Control and Cybernetics, 2008, 37, 711–729.
[30] KLUKOWSKI L., *Tests for Relation Type – Equivalence or Tolerance – Based on Averaging of Multiple Binary Comparisons*, [in:] *Development in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Foundations*, Vol. 1, K. Atanassov K. Bustince, O. Hryniewicz, J. Kacprzyk, M. Krawczak, B. Riecan, E. Szmidt (Eds.), Academic Publishing House Exit, Warsaw 2008, 175–189.
[31] KLUKOWSKI L., *Estimation of the preference relation on the basis of medians from pairwise comparisons*, [in:] *Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues*, J. Owsiński, W. Bruggemann (Eds.), IBS PAN, Warsaw 2008, 109–120.
[32] KLUKOWSKI L., *Estimation of the relations: preference, equivalence and tolerance on the basis of pairwise comparisons*, [in:] *Development in Fuzzy Sets, Intuitionistic Fuzzy Sets, I. Foundations*, K.T. Atanassov, M. Baczyński, J. Drewniak, J. Kacprzyk, M. Krawczak, E. Szmidt, M. Wygralak, S. Zadrożny (Eds.), IBS PAN, SRI PAS, Warsaw 2010, 109–119.
[33] KLUKOWSKI L., *Tests for relation type – equivalence or tolerance based on medians from multiple pairwise comparisons*, Materiały i Studia Polskiego Stowarzyszenia Zarządzania Wiedzą, W. Bojar, J.W. Owsiński (Eds.), 31, 338–348, Bydgoszcz 2010.
[34] KLUKOWSKI L., *Optimization of public debt management in the case of stochastic budgetary constraints*, [in:] *Multiple criteria decisions making'09*, T. Trzaskalik, T. Wachowicz (Eds.), The University of Economics in Katowice, Katowice 2010.
[35] KLUKOWSKI L., *Methods of Estimation of Relations of: Equivalence, Tolerance, and Preference in a Finite Set*, IBS PAN, Series: Systems Research, Vol. 69, Warsaw 2011.
[36] KLUKOWSKI L., *Estimation of Tolerance Relation on the Basis of Pairwise Comparisons*, [in:] *Computer Recognition Systems 4, Advances in Intelligent and Soft Computing 95*, R. Burduk, M. Kurzyński, M. Woźniak, A. Żołnierek (Eds.), Springer, Berlin 2011, 91–98.
[37] KLUKOWSKI L., *Properties of estimators of the preference relation in the form of difference of ranks – simulation survey*, [in:] *Modelowanie Preferencji a Ryzyko'12*, T. Trzaskalik (Ed.), Zeszyty Naukowe Wydziałowe, 97, 119–132, Economic University in Katowice, Katowice 2012 (in Polish).
[38] KLUKOWSKI L., *Properties of estimators of the preference relation based on pairwise comparisons: binary and difference of ranks – simulation survey*, Submitted to Control and Cybernetics, 2012.
[39] KOHONEN T., *Self-Organizing Maps*, Springer, Berlin 1995.
[40] KORONACKI J., ĆWIK J., *Statistical Learning Systems*, WNT, Warsaw 2005 (in Polish).
[41] RANDLES R.H., WOLFE D.A., *Introduction to the Theory of Nonparametric Statistics*, Wiley, New York 1979.
[42] RIPLEY B.D., *Stochastic Simulation*, Wiley, New York 2006.
[43] SACHS L., *Applied Statistics*, Springer, New York 1978.
[44] SHESKIN D.J., *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, Boca Raton, 1997.
[45] SIEGEL S., CASTELLAN N.J. Jr., *Nonparametric Statistics for the Behavioral Sciences* (2nd Ed.), McGraw-Hill Book Company, New York 1988.
[46] SLATER P., *Inconsistencies in a schedule of paired comparisons*, Biometrika, 1961, 48, 303–312.
[47] WITTEN I.H., FRANK E., *Data Mining: Practical Machine Learning Tools and Techniques* (2nd Ed.), Morgan Kaufmann Publishers, Amsterdam 2005.