77/2012

Raport Badawczy
Research Report

RB/16/2012

Estimation of: the relations of equivalence, tolerance and preference on the basis of pairwise comparisons in binary and multivalent form with random errors

L. Klukowski

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. zw. dr hab. inż. Zbigniew Nahorski

Warszawa 2012

Estimation of the relations of: equivalence, tolerance and preference on the basis of pairwise comparisons in binary and multivalent form with random errors

Leszek Klukowski
Systems Research Institute Polish Academy of Sciences,
e-mail: Leszek.Klukowski@ibspan.waw.pl

The paper presents the estimators of three relations: equivalence, tolerance and preference in a finite set on the basis of multiple pairwise comparisons, disturbed by random errors; they have been developed by the author. The estimators can rest on: binary (qualitative), multivalent (quantitative) and combined comparisons. The estimates are obtained on the basis of discrete programming tasks. The estimators require weak assumptions about distributions of comparisons errors, especially allow non-zero expected values. The estimators have good statistical properties, in particular consistency. Precision of the estimators can be determined also with the use of simulation methods. The estimates can be validated in versatile way. The approach allows verification of the relation type – equivalence or tolerance (using binary comparisons).

The paper summarizes briefly the results obtained by the author; the broader view is presented in Klukowski 2011a.

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

1. Introduction

Estimation of the relations of equivalence, tolerance, or preference, on the basis of multiple pairwise comparisons with random errors, is aimed at determination of an actual structure of data. It also provides the properties of estimates. The properties comprise: consistency, distributions of errors, efficiency of estimators, etc. They allow for the statistical validation of estimates - including the assumptions concerning the comparison errors and existence of a relation.

The approach applied in the work rests on a statistical paradigm: to determine the relation form, which minimizes the inconsistencies (differences) with a sample - in the form of multiple pairwise comparisons. Such comparisons can be obtained with the use of statistical tests, experts opinions or other procedures, prone to generating random errors. The estimates are obtained on the basis of optimization tasks. However, the approach enables also extraction of some knowledge about relation type, which is (a priori) unknown. The example is determination of the type of a relation - equivalence or tolerance. Moreover, the entire estimation process: identification of relation type, estimation and validation of estimates, can be computerized.

The approach presented here is an original contribution of the author to the subject. The main components comprise: determination of weak assumptions about distributions of comparison errors, definition of two types of estimators and two types of data – qualitative and quantitative, properties of the estimators, and validation of estimates. The assumptions allow for the extension of the application sphere of pairwise techniques. The estimators considered have different efficiency and computational cost – the results of the work allow for choosing the best approach. The estimators allow for combining of both types of comparisons. The results of the work provide a comprehensive solution to an important statistical problem.

The problems, which require estimation of relations of equivalence, tolerance, or preference, on the basis of pairwise comparisons with random errors, appear in many disciplines of knowledge: economy, finance, medicine, etc.

The idea of estimators is based on the concept of the nearest adjoining order (NAO – Slater, 1961, David, 1988): to minimize the inconsistencies with the given set of comparisons. The estimators proposed are based on:

- minimization of the sum of inconsistencies between relation form and (whole) set of comparisons

or

- minimization of the sum of inconsistencies between relation form and the medians from multiple comparisons of each pair.

The comparisons are assumed also in two basic forms:

binary – expressing qualitative features of a pair, e.g. the direction of preference, and

- multivalent – expressing quantitative features of a pair, e.g. the difference of ranks of elements.

The errors of pairwise comparisons are realizations of some random variables. The assumptions about distributions of errors are weaker than those commonly used in the literature (David, 1988); they are satisfied in the case of each rational scientific investigation. The estimators can be applied also in the case of unknown distributions of comparison errors.

The estimators have good statistical properties, obtained on the basis of: • properties of random variables expressing differences between the (actual) relation form and pairwise comparisons, • the probabilistic inequalities (Hoeffding 1963, Chebyshev), • properties of order statistics (David, 1970). The properties guarantee convergence of estimates to actual relation form for the number of independent comparisons of each pair approaching infinity.

Thus, the estimators are consistent. The analytical properties of the estimators has been complemented with the use of a simulation survey (Klukowski 2011a Chap. 9, Klukowski *Control and Cybernetics* – to appear). It confirms their high efficiency, for a finite number of comparisons, and allows for determination of parameters, especially number of comparisons, guaranteeing high precision of the estimates. The results of estimation can be verified with the use of statistical tests. Thus, the results of the work fill the gap between the methods, which require strong assumptions, and the methods based on heuristic rules, not vested with formal properties.

The literature on pairwise comparisons with random errors concerns mainly ranking problems – classical results are presented in: David (1988), Bradley (1976, 1984), Davidson (1976) (bibliography), Brunk (1960). The authors mentioned present and discuss a complete range of existing methods: assumptions, estimators and their properties, tests for validation of results. In general, the assumptions required by the methods impose significant restrictions on probabilistic properties of comparisons; these assumptions constrain the application sphere. In particular, the comparisons can assume only the binary form, indicating the direction of the preference; some methods do not allow ties (equivalent elements). The basic methods are based on the linear model and the combinatorial models (David, 1988 Ch. 2, 4).

The literature concerning classification methods, based on pairs of elements is extremely extensive (see e.g. Gordon 1999, Hand 1986, Kaufman, Rousseeuv 1990, Hastie, Tibshirani, Friedman 2002, Koronacki, Ćwik 2005, Hartigan 1975). However, it should be emphasized that existing approaches do not cover entirely the problems presented in the work.

The paper consists with 6 sections. The second section presents main ideas of estimation, in particular the form of estimators. The next section – presents properties of estimators obtained by the author. In the forth section are discussed tests for validation of estimates - they verify assumptions required. The next section discusses briefly optimization algorithms, which can be applied for determining of estimates. Last section summarizes results of the author in the area under consideration and shows problems for further researches.

2. Estimation of the relations – main ideas

2.1. Definitions, notations and formulation of the estimation problems

The problem of estimation of relation on the basis of pairwise comparisons can be stated as follows.

We are given a finite set of elements $\mathbf{X}=\{x_1, ..., x_m\}$ $(3 \leq m < \infty)$. There exists in the set $\mathbf{X}$: the equivalence relation $\mathbf{R}^{(e)}$ (reflexive, transitive, symmetric), or the tolerance relation $\mathbf{R}^{(\tau)}$ (reflexive, symmetric), or the preference relation $\mathbf{R}^{(p)}$ (alternative of the equivalence relation and strict preference relation). 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 the family $\chi_1^{(e)*}, ..., \chi_n^{(e)*}$ having the following properties:

$$\bigcup_{q=1}^{n} \chi_q^{(e)*} = \mathbf{X}, \tag{2.1}$$

$$\chi_r^{(e)*} \cap \chi_s^{(e)*} = \{\mathbf{0}\}, \tag{2.2}$$

where:

$\mathbf{0}$ – the empty set,

$$x_i, x_j \in \chi_r^{(e)*} \equiv x_i, x_j - \text{equivalent elements}, \tag{2.3}$$

$$(x_i \in \chi_r^{(e)*}) \cap (x_j \in \chi_s^{(e)*}) \equiv x_i, x_j - \text{non-equivalent elements for } i \neq j, r \neq s. \tag{2.4}$$

The tolerance relation generates the family $\chi_1^{(\tau)*}, ..., \chi_n^{(\tau)*}$ with the property (2.1), i.e. $\bigcup_{q=1}^{n} \chi_q^{(\tau)*} = \mathbf{X}$, and the properties:

$\exists r, s\ (r \neq s)$ *such that* $\chi_r^{(\tau)*} \cap \chi_s^{(\tau)*} \neq \{\mathbf{0}\}$,

$$x_i, x_j \in \chi_r^{(\tau)*} \equiv x_i, x_j - \text{equivalent elements}, \tag{2.5}$$

$$(x_i \in \chi_r^{(\tau)*}) \cap (x_j \in \chi_s^{(\tau)*}) \equiv x_i, x_j - \text{non-equivalent elements for } i \neq j \text{ and } (x_i, x_j) \notin \chi_r^{(\tau)*} \cap \chi_s^{(\tau)*}, \tag{2.6}$$

each subset $\chi_r^{(\tau)*}$ $(1 \leq r \leq n)$ includes an element x_i such that $x_i \notin \chi_s^{(\tau)*}$ $(s \neq r)$. (2.7)

The preference relation generates the family $\chi_1^{(p)*}, ..., \chi_n^{(p)*}$ with the properties (2.1), (2.2) and the property:

$$(x_i \in \chi_r^{(p)*}) \cap (x_j \in \chi_s^{(p)*}) \equiv x_i \text{ is preferred to } x_j \text{ for } r<s. \tag{2.8}$$

The relations defined by the conditions (2.1) - (2.8) can be expressed, alternatively, by the values (functions) $T_\upsilon^{(\ell)}(x_i, x_j)$ $((x_i, x_j) \in \mathbf{X} \times \mathbf{X};\ \ell \in \{p, e, \tau\},\ \upsilon \in \{b, \mu\}$; symbols b, μ denote – respectively – the binary and multivalent comparisons), defined as follows:

$$T_b^{(e)}(x_i,x_j)=\begin{cases}0 \text{ if exists } r \text{ such that } (x_i,x_j)\in \chi_r^{(e)*},\\ 1 \text{ otherwise;}\end{cases} \tag{2.9}$$

• the function $T_b^{(e)}(x_i,x_j)$, describing the equivalence relation, assuming binary values, expresses the fact if a pair (x_i,x_j) belongs to a common subset or not;

$$T_b^{(r)}(x_i,x_j)=\begin{cases}0 \text{ if exists } r,s\ (r=s \text{ not excluded}) \text{ such that}\\ \quad (x_i,x_j)\in \chi_r^{(r)*}\cap\chi_s^{(r)*},\\ 1 \text{ otherwise;}\end{cases} \tag{2.10}$$

• the function $T_b^{(r)}(x_i,x_j)$, describing the tolerance relation, assuming binary values, expresses the fact if a pair (x_i,x_j) belongs to any conjunction of subsets (also to the same subset) or not; the condition (2.7) guarantees uniqueness of the description;

$$T_\mu^{(r)}(x_i,x_j)=\#(\Omega_i^*\cap\Omega_j^*), \tag{2.11}$$

where:

Ω_l^* - the set of the form $\Omega_l^*=\{s\mid x_l\in\chi_s^{(r)*}\}$,

$\#(\Xi)$ - the number of elements of the set Ξ;

• the function $T_\mu^{(r)}(x_i,x_j)$, describing the tolerance relation, assuming multivalent values, expresses the number of subsets of conjunction including both elements; condition (2.7) guarantees the uniqueness of the description;

$$T_b^{(p)}(x_i,x_j)=\begin{cases}0 \text{ if there exists } r \text{ such that } (x_i,x_j)\in\chi_r^{(p)*},\\ -1 \text{ if } x_i\in\chi_r^{(p)*},\ x_j\in\chi_s^{(p)*} \text{ and } r<s;\\ 1 \text{ if } x_i\in\chi_r^{(p)*},\ x_j\in\chi_s^{(p)*} \text{ and } r>s;\end{cases} \tag{2.12}$$

• the function $T_b^{(p)}(x_i,x_j)$, describing the preference relation, assuming binary values, expresses the direction of preference in a pair or the equivalence of its elements;

$$T_\mu^{(p)}(x_i,x_j)=d_{ij}\Leftrightarrow x_i\in\chi_r^{(p)*},\ x_j\in\chi_s^{(p)*},\ d_{ij}=r-s; \tag{2.13}$$

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

2.2. Assumptions about pairwise comparisons

The relation $\chi_1^{(\cdot)*},\ldots,\chi_n^{(\cdot)*}$ is to be determined (estimated) on the basis of N ($N\geq1$) comparisons of each pair $(x_i,x_j)\in\mathbf{X}\times\mathbf{X}$; any comparison $g_{\iota k}^{(\cdot)}(x_i,x_j)$ evaluates the actual value of $T_{\upsilon}^{(\ell)}(x_i, x_j)$ and can be disturbed by a random error. The following assumptions concerning the comparison errors are made:

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

A2. Any comparison $g_{\upsilon k}^{(\ell)}(x_i, x_j)$ $(\ell \in \{e, \tau, p\};\ \upsilon \in \{b, \mu\};\ k = 1, ..., N)$, is the evaluation of the value $T_{\upsilon}^{(\ell)}(x_i, x_j)$, disturbed 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:

$$P(g_{bk}^{(\ell)}(x_i, x_j) - T_b^{(\ell)}(x_i, x_j) = 0 \mid T_b^{(\ell)}(x_i, x_j) = \kappa_{bij}^{(\ell)}) \ge 1 - \delta$$
$$(\kappa_{bij}^{(\ell)} \in \{-1, 0, 1\},\ \ \delta \in (0, \tfrac{1}{2})), \tag{2.14}$$

$$\sum_{r \le 0} P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) > \tfrac{1}{2} \qquad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., \pm m\},\ r \text{ – zero or an integer number}), \tag{2.15}$$

$$\sum_{r \ge 0} P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = -r \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) > \tfrac{1}{2} \qquad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., \pm m\},\ r \text{ – zero or an integer number}), \tag{2.16}$$

$$P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r) \ge P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r + 1 \mid$$
$$T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) \quad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., m\},\ r > 0), \tag{2.17}$$

$$P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r) \ge P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r - 1 \mid$$
$$T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) \quad (\kappa_{\mu ij}^{(\ell)} \in \{0, ..., m\},\ r < 0), \tag{2.18}$$

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

The assumption A3 makes it possible to determine the distributions of estimation errors of estimators proposed in this work. However, determination of the exact distributions of the (multidimensional) errors, in an analytic way, is complicated and in practice unrealizable. The main properties of the estimators, especially their consistency, are valid without the assumption.

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

In the case of the preference relation including equivalent elements, the condition (2.14) can be relaxed to the form (2.15) – (2.16).

The assumptions A2 – A3 reflect the following properties of distributions of comparisons errors:

• the probability of correct comparison is greater than of the incorrect one - in the case of binary comparisons (inequality (2.14));

• zero is the median of each distribution of comparison error (inequalities (2.14) – (2.16));

• zero is the mode of each distribution of comparison error (inequalities (2.14) – (2.18));

• the set of all comparisons comprises the realizations of independent random variables;

• the expected value of any comparison error can differ from zero.

The assumptions about comparisons errors are not restricted. Especially, the errors can have non-zero expected values; the probabilities of errorless results have to satisfy the mode and median condition. These features guarantee broad spectrum of applications and protects against incorrect results.

2.3. **The form of estimators**

The main idea of the estimators proposed, i.e. minimization of differences between the relation and the pairwise comparisons, refers to a well-known principle. However, in the case under consideration, it does not indicate analytical properties, because it is not associated with minimization of the likelihood function (which requires distributions of comparison errors) or the sum of error squares. In our case, the properties of the estimators have been obtained on the basis of differences between the properties of the errorless estimate (actual form of the relation) and the estimates different from the errorless one. The properties have been proven by the author on the basis of the well-known probabilistic inequalities (see Hoeffding, 1963, Chebyshev - for variance), properties of order statistics (David, 1970), and convergence of variances. The theoretical properties have been verified through the simulation survey.

Two forms of estimators are examined.

The estimate based on the total sum of differences, denoted $\hat{\chi}_1^{(\ell)},...,\hat{\chi}_n^{(\ell)}$ (or $\hat{T}_\upsilon^{(\ell)}(x_i,x_j)$ $<i,j>\in R_m$), results from the minimization problem:

$$\min_{\chi_1^{(\ell)},...,\chi_r^{(\ell)}\in F_X^{(\ell)}} \{ \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \left| g_{\upsilon k}^{(\ell)}(x_i,x_j) - t_\upsilon^{(\ell)}(x_i,x_j) \right| \}, \qquad (2.19)$$

where:

$F_X^{(\ell)}$ - the feasible set, i.e. the family of all relations $\chi_1^{(\ell)},...,\chi_r^{(\ell)}$ of ℓ - th type in the set **X**,

$t_\upsilon^{(\ell)}(x_i,x_j)$ - the function describing any relation $\{\chi_1^{(\ell)},...,\chi_r^{(\ell)}\}$ of ℓ -th type,

R_m - the set of the form $R_m = \{<i,j> \; | 1\leq i,j\leq m;\; j>i\}$

(symbol $g_{\upsilon k}^{(\ell)}(x_i,x_j)$ is used for both random variables and realizations, because this does not lead to misunderstanding).

In the case of the preference relation and binary comparisons the following transformation is also applied:

$$\theta(g_{uk}^{(\ell)}(x_i,x_j)-t_{\upsilon}^{(\ell)}(x_i,x_j))=\begin{cases}0 \text{ if } g_{uk}^{(\ell)}(x_i,x_j)=t_{\upsilon}^{(\ell)}(x_i,x_j);\\ 1 \text{ if } g_{uk}^{(\ell)}(x_i,x_j)\neq t_{\upsilon}^{(\ell)}(x_i,x_j).\end{cases} \tag{2.19a}$$

The criterion function with the use of the transformation (2.19a), expresses the number of differences between the comparisons and the function $T_{\upsilon}^{(p)}(x_i,x_j)$. It is simpler from the computational point of view, because the variables $\theta(g_{uk}^{(\ell)}(x_i,x_j)-t_{\upsilon}^{(\ell)}(x_i,x_j))$ assume binary values (zero or one), while the difference $\left|g_{uk}^{(\ell)}(x_i,x_j)-t_{\upsilon}^{(\ell)}(x_i,x_j)\right|$ assumes values from the set $\{0,\pm 1,\pm 2\}$. The properties of both approaches are similar (Klukowski, 1990b).

The estimate based on medians, denoted $\tilde{\chi}_1^{(\ell)},\ldots,\tilde{\chi}_r^{(\ell)}$ (or $\tilde{T}_{\upsilon}^{(\ell)}(x_i,x_j)$), is obtained on the basis of the following minimization problem:

$$\min_{\chi_1^{(\ell)},\ldots,\chi_r^{(\ell)}\in F_X}\{\sum_{<i,j>\in R_m}\left|g_{\upsilon}^{(\ell,me)}(x_i,x_j)-t_{\upsilon}^{(\ell)}(x_i,x_j)\right|\}, \tag{2.20}$$

where:

$g_{\upsilon}^{(\ell,me)}(x_i,x_j)$ - the sample median (a middle value) in the set $\{g_{\upsilon,1}^{(\ell)}(x_i,x_j),\ldots,g_{\upsilon N}^{(\ell)}\}$.

The estimate, resulting from the criterion (2.19) or (2.19a) will be denoted with symbols $\hat{\chi}_1^{(\ell)},\ldots,\hat{\chi}_r^{(\ell)}$ or (equivalently) $\hat{T}_{\upsilon}^{(\ell)}(x_i,x_j)$, while the estimate resulting from the criterion (2.20) - with symbols $\tilde{\chi}_1^{(\ell)},\ldots,\tilde{\chi}_r^{(\ell)}$ or $\tilde{T}_{\upsilon}^{(\ell)}(x_i,x_j)$.

In the case of the preference relation and medians from comparisons, the following transformation is also applied:

$$\theta(g_{\upsilon}^{(\ell,me)}(x_i,x_j)-t_{\upsilon}^{(\ell)}(x_i,x_j))=\begin{cases}0 \text{ if } g_{\upsilon}^{(\ell,me)}(x_i,x_j)=t_{\upsilon}^{(\ell)}(x_i,x_j);\\ 1 \text{ if } g_{\upsilon}^{(\ell,me)}(x_i,x_j)\neq t_{\upsilon}^{(\ell)}(x_i,x_j),\end{cases} \tag{2.20a}$$

instead of the difference $\left|g_{\upsilon}^{(\ell,me)}(x_i,x_j)-t_{\upsilon}^{(\ell)}(x_i,x_j)\right|$.

The transformation (2.20a) sums up the number of inconsistencies between the comparisons and the relation form, while the difference $\left|g_{\upsilon}^{(\ell,me)}(x_i,x_j)-t_{\upsilon}^{(\ell)}(x_i,x_j)\right|$ takes also into account the opposite direction of preference in a comparison. The optimization based on transformation (2.20a) is simpler to solve; and both approaches have similar efficiency (see Klukowski, 1990b).

It is clear that the number of estimates, resulting from the criterion functions (2.19), (2.19a), (2.20), (2.20a) can exceed one; the unique estimate can be determined in a random way or as a result of validation. Multiple estimates can appear also in other methods (see David 1988, Ch. 2). The minimal values of the respective functions are equal zero.

The assumptions A1 – A3 allow for inference about distributions of errors of estimates. Let us discuss first the estimator based on of the criterion (2.19). For each relation type one can determine a finite set including all possible realizations of comparisons

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

and the probability of each realization. The use of the criterion (2.19) determines: the estimate, its probability and estimation error. The error has the form: $\{\hat{T}_{\upsilon}^{(\ell)}(x_i, x_j) - T_{\upsilon}^{(\ell)}(x_i, x_j);\ <i, j> \in R_m\}$, i.e. it is a multidimensional random variable. The analysis of such error is, in fact, unrealizable and it is suggested to replace it with one-dimension error:

$$\hat{\Delta}_{\upsilon}^{(\ell)} = \sum_{<i,j> \in R_m} \left| \hat{T}_{\upsilon}^{(\ell)}(x_i, x_j) - T_{\upsilon}^{(\ell)}(x_i, x_j) \right| . \qquad (2.21)$$

The estimate with the error $\hat{\Delta}_{\upsilon}^{(\ell)} = 0$ is the errorless estimate. The probability of such error can be determined in the analytic way – as a sum of probabilities of all realizations of comparisons indicating the errorless estimate. It is clear that its value (probability) depends on the number of comparisons N and the variance of comparison errors; increase of N decreases the probability of such error and decreases the variance of the estimator. The probabilities of errors different from zero can be determined in a similar way; all possible errors and their probabilities determine the distribution function of the estimation error. Determination of the probability function in the analytic manner is complicated and involves huge computational cost - even for moderate m. Therefore, simulation approach has to be used for this purpose. Simulation study provides complementary (to analytic results) knowledge about efficiency of estimators, especially useful in applications.

Similar considerations apply for the criteria (2.19a), (2.20), (2.20a).

3. Properties of estimators

The analytical properties of the estimators, established by the author, have mainly asymptotic character, i.e. they apply to the case $N \to \infty$. The properties guarantee the basic feature of the estimators - consistency. It is clear that errorless estimates can be also obtained for finite N, with probability close to one, because the number of variants (in optimization problems) is huge, but finite. In general, precision of estimates depends not only on N, but also on distributions of comparison errors and some features of the form of relation, e.g. the number of subsets n and the number of elements in each subset. The precision level is also not the same for both estimators considered. Simulation survey (Klukowski 2011a, 2012a, b) gives indications about the necessary number of N for given distributions of comparison errors.

The analytical properties of the estimators are based on properties of random variables expressing differences between pairwise comparisons and the relation form (expressed by $T_b^{(f)}(x_i, x_j)$). It has been demonstrated in the papers of the author that the variables corresponding to the actual relation form have different properties than the variables corresponding to any other relation. The following results have been obtained:

(i) the expected values of the variables, corresponding to actual relation form are lower than the expected values of variables corresponding to any other relation;

(ii) the variances of the variables expressing differences between comparisons and the relation form, both - actual and different than actual, divided by the number of comparisons N in the case of sum of differences, converge to zero for $N \to \infty$;

(iii) the probability of the event that the variable corresponding to actual relation assumes a value lower than the variable corresponding to a relation other than actual converges to one for $N \to \infty$; the speed of convergence guarantees good efficiency of the estimates.

Properties (i) - (iii) provide the basis for construction of estimators; these properties have been complemented with some additional features and a simulation study. An important result of the simulation survey consists in the fact that efficiency of the estimator based on the sum of inconsistencies is higher than of the median estimator; the latter estimator is, though, simpler from computational point of view and more robust with respect to outliers.

Let us illustrate these considerations by the simplest case, i.e. equivalence relation and the estimator resulting from the criterion (2.20). The differences between any comparison $g_{bk}^{(e)}(x_i, x_j)$ and the value $T_b^{(e)}(x_i, x_j)$ assume the form:

$$U_{bk}^{(e)*}(x_i, x_j) = \begin{cases} 0 \;\; if \;\; g_{bk}^{(e)}(x_i, x_j) = T_b^{(e)}(x_i, x_j); \; T_b^{(e)}(x_i, x_j) = 0; \\ 1 \;\; if \;\; g_{bk}^{(e)}(x_i, x_j) \neq T_b^{(e)}(x_i, x_j); \; T_b^{(e)}(x_i, x_j) = 0, \end{cases} \qquad (2.22)$$

$$V_{bk}^{(e)*}(x_i, x_j) = \begin{cases} 0 \;\; if \;\; g_{bk}^{(e)}(x_i, x_j) = T_b^{(e)}(x_i, x_j); \; T_b^{(e)}(x_i, x_j) = 1; \\ 1 \;\; if \;\; g_{bk}^{(e)}(x_i, x_j) \neq T_b^{(e)}(x_i, x_j); \; T_b^{(e)}(x_i, x_j) = 1. \end{cases} \qquad (2.23)$$

The sum of differences assumes, for any k $(1 \le k \le N)$, the form:

$$\sum_{<i,j> \in I^{(e)*}} U_{bk}^{(e)*}(x_i, x_j) + \sum_{<i,j> \in J^{(e)*}} V_{bk}^{(e)*}(x_i, x_j), \qquad (2.24)$$

where:

$I^{(e)*}$ - the set of pairs $\{<i, j> \;|\; T_b^{(e)*}(x_i, x_j) = 0\}$,

$J^{(e)*}$ - the set of pairs $\{<i,j> \;|T_b^{(e)*}(x_i,x_j)=1\}$.

The total sum of the differences between the relation form and the comparisons is equal:

$$W_{bN}^{(e)*}=\sum_{k=1}^{N}\Big(\sum_{<i,j>\in I^{(e)*}} U_{bk}^{(e)*}(x_i,x_j)+\sum_{<i,j>\in J^{(e)*}} V_{bk}^{(e)*}(x_i,x_j)\Big). \tag{2.25}$$

Under the assumptions A1, A2, A3, the expected values of the variables $U_{bk}^{(e)*}(x_i,x_j)$, $V_{bk}^{(e)*}(x_i,x_j)$ satisfy the inequalities: $E(U_{bk}^{(e)*}(x_i,x_j))\le\delta$, $E(V_{bk}^{(e)*}(x_i,x_j))\le\delta$. Therefore, the expected value of the variable $W_{bN}^{(e)*}$ satisfies the inequality $E(W_{bN}^{(e)*})\le\frac{Nm(m-1)}{2}\delta$. Assumptions A1 – A3 allow for determining the variance $Var(W_{bN}^{(e)*})$; its value is finite and satisfies the inequality $Var(W_{bN}^{(e)*})\le\frac{Nm(m-1)}{2}\delta(1-\delta)$.

Obviously:

$$E(\tfrac{1}{N}W_{bN}^{(e)*})\le\frac{m(m-1)}{2}\delta, \tag{2.26}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N}W_{bN}^{(e)*})=0. \tag{2.27}$$

Let us consider any relation $\tilde{\chi}_1^{(e)},...,\tilde{\chi}_{\tilde{n}}^{(e)}$ different than $\chi_1^{(e)*},...,\chi_n^{(e)*}$; this means that there exist pairs (x_i,x_j), such that $\tilde{T}_b^{(e)}(x_i,x_j)\neq T_b^{(e)}(x_i,x_j)$. Define the random variables $\tilde{U}_{bk}^{(e)}(x_i,x_j)$, $\tilde{V}_{bk}^{(e)}(x_i,x_j)$, corresponding to the such values $\tilde{T}_b^{(e)}(x_i,x_j)$:

$$\tilde{U}_{bk}^{(e)}(x_i,x_j)=\begin{cases}0 \;\; if \;\; g_{bk}^{(e)}(x_i,x_j)=\tilde{T}_b^{(e)}(x_i,x_j);\; \tilde{T}_b^{(e)}(x_i,x_j)=0;\\ 1 \;\; if \;\; g_{bk}^{(e)}(x_i,x_j)\neq\tilde{T}_b^{(e)}(x_i,x_j);\; \tilde{T}_b^{(e)}(x_i,x_j)=0,\end{cases} \tag{2.28}$$

$$\tilde{V}_{bk}^{(e)}(x_i,x_j)=\begin{cases}0 \;\; if \;\; g_{bk}^{(e)}(x_i,x_j)=\tilde{T}_b^{(e)}(x_i,x_j);\; \tilde{T}_b^{(e)}(x_i,x_j)=1;\\ 1 \;\; if \;\; g_{bk}^{(e)}(x_i,x_j)\neq\tilde{T}_b^{(e)}(x_i,x_j);\tilde{T}_b^{(e)}(x_i,x_j)=1.\end{cases} \tag{2.29}$$

The expected values $E(\tilde{U}_{bk}^{(e)}(x_i,x_j))$, $E(\tilde{V}_{bk}^{(e)}(x_i,x_j))$ assume the form:

$$\begin{aligned}E(\tilde{U}_{bk}^{(e)}(x_i,x_j))=0*P(g_{bk}^{(e)}(x_i,x_j)=0\;|T_b^{(e)}(x_i,x_j)=1)+\\ 1*P(g_{bk}^{(e)}(x_i,x_j)=1\;|T_b^{(e)}(x_i,x_j)=1)\ge 1-\delta,\end{aligned} \tag{2.30}$$

$$\begin{aligned}E(\tilde{V}_{bk}^{(e)}(x_i,x_j))=0\times P(g_{bk}^{(e)}(x_i,x_j)=0\;|T_b^{(e)}(x_i,x_j)=0)+\\ 1\times P(g_{bk}^{(e)}(x_i,x_j)=1\;|T_b^{(e)}(x_i,x_j)=0)\ge 1-\delta,\end{aligned} \tag{2.31}$$

and:

$$E(\tilde{W}_{bN}^{(e)})=\sum_{k=1}^{N}\Big(\sum_{\tilde{I}^{(e)}}E(\tilde{U}_{bk}^{(e)}(x_i,x_j))+\sum_{\tilde{J}^{(e)}}E(\tilde{V}_{bk}^{(e)}(x_i,x_j))\Big)>\frac{m(m-1)}{2}\delta. \tag{2.32}$$

The formulae (2.26)–(2.32) indicate that the expected value $E(\frac{1}{N}W_{bN}^{(p)*})$, corresponding to the actual relation $\chi_1^{(e)*},...,\chi_n^{(e)*}$, is lower than the expected value $E(\frac{1}{N}\tilde{W}_{bN}^{(p)})$, corresponding to any other relation $\tilde{\chi}_1^{(p)},...,\tilde{\chi}_{\tilde{n}}^{(p)}$. The variances of both variables converge to zero for $N \to \infty$. The variables $U_{bk}^{(e)*}(x_i,x_j)$, $V_{bk}^{(e)*}(x_i,x_j)$ assume values equal to $\left|g_{bk}^{(e)*}(x_i,x_j) - T_b^{(e)}(x_i,x_j)\right|$, used in the criterion function (2.19). Moreover, it can be also shown (see Klukowski, 1994), that:

$$P(W_{bN}^{(p)*} < \tilde{W}_{bN}^{(p)}) \geq 1 - \exp\{-2N(\tfrac{1}{2} - \delta)^2\}. \quad (2.33)$$

The above facts indicate that the estimator $\hat{\chi}_1^{(p)},...,\hat{\chi}_{\hat{n}}^{(p)}$, minimizing the number of inconsistencies with comparisons, guarantees the errorless estimate for $N \to \infty$. The inequality (2.33) shows that the errorless estimate can be obtained with the probability close to one for finite *N*. Moreover, the inequality indicates the influence of δ and N on the precision of the estimator. The distribution of an error of the estimator, for given parameters, has to be evaluated with the use of simulation approach.

The properties of the median estimator are based on the fact that the random variables $\frac{1}{N}\sum_{k=1}^{N} U_{bk}^{(e)*}(x_i,x_j)$ and $\frac{1}{N}\sum_{k=1}^{N} V_{bk}^{(e)*}(x_i,x_j)$ converge, with probability one, to a limit equal or lower than δ, for $N \to \infty$. Therefore, the median $g_b^{(e,me)}(x_i,x_j)$ converges to the actual value $T_b^{(e)}(x_i,x_j)$. As a result, minimization of (2.20) guarantees that the estimate $\hat{\chi}_1^{(e)},...,\hat{\chi}_{\hat{n}}^{(e)}$ converges to $\chi_1^{(e)*},...,\chi_n^{(e)*}$. Moreover, it can be shown (see Klukowski, 1994) that:

$$P(W_{bN}^{(p,me)*} < \tilde{W}_{bN}^{(me,p)}) \geq 1 - 2\exp\{-2N(\tfrac{1}{2} - \delta)^2\}. \quad (2.34)$$

Inequality (2.34) gives some evaluation of precision and speed of convergence of the median estimator; the evaluation of error of the estimator has been obtained with the use of simulation approach (see Klukowski 2011a, Chap. 9).

The results presented in Klukowski (1994) include some additional inequalities and evaluations, especially for the case of single comparison for each pair. They are not repeated in this work, which concentrates on multiple comparisons. Moreover, simulation survey covers and completes some of these results.

The above considerations are valid also in the case of the tolerance and preference relations, estimated with the use of binary comparisons.

The case of multivalent comparisons, can be analyzed in a similar way. However, the considerations are more complicated from the analytical point of view – the details are presented in Klukowski 2011a, Chap. 6 and 8.

4. Validation of estimates

The estimators of relations are based on the assumptions A1–A3. The crucial assumption A1 states that the relations exist and their type is known, the assumptions A2 and A3 establish the properties of pairwise comparisons. These assumptions can be verified with the use of statistical tests; the positive result of verification validates the estimate obtained.

The first step of validation is to verify the assumptions about comparison errors. The assumptions A2 and A3 can be verified with the use of the well-known tests for independence, randomness, unimodality, and values of mode and median (see Daniel, 1990, Sheskin, 1997, Siegel and Castellan, 1988, Domański, 1990, Hollander, Wolfe, 1973, Randles, Wolfe, 1979, Sachs, 1978). Such hypotheses can be tested on the basis of comparisons:

$$g_{\upsilon,1}^{(\ell)}(x_i,x_j),\ \ldots,\ g_{\upsilon N}^{(\ell)}(x_i,x_j)\ \ (\upsilon\in\{b,\mu\},\ <i,j>\in R_m,\ \ell\in\{e,\tau,p\})$$

or differences:

$$g_{\upsilon k}^{(\ell)}(x_i,x_j)-\hat{T}_{\upsilon}^{(\ell)}(x_i,x_j),\ g_{\upsilon k}^{(\ell)}(x_i,x_j)-\tilde{T}_{\upsilon}^{(\ell)}(x_i,x_j)\ (k=1,\ldots,N);$$

with the details given in (Klukowski 2011a, Chap. 10, Klukowski 2011c).

The assumption of independence of the whole set of comparisons is difficult to verify; it seems more reliable to verify the assumption about independence of comparisons of individual pairs.

Verification of existence of a relation has to be done after the positive results of tests verifying the assumptions A2, A3 and has to be based on the estimates of the relation. Typical hypotheses verify the fact that the estimate is valid, i.e. the relation exists, under alternatives about the equivalency of all elements of the set **X** or randomness of comparisons or other data structure. Another basis for the verification is constituted by the optimal values of the functions (2.19), (2.19a) or (2.20), (2.20a); large values indicate significant differences with comparisons and suggest rejection of estimates. Critical values of such tests have to be obtained on the basis of simulations.

Some other features of estimates of relations can be used as the basis for verification, like, e.g., positive correlation of ranks of individual elements obtained on the basis of sequential subsets of comparisons:

$$g_{\upsilon,1}^{(\ell)}(x_i,x_j),\ \ldots,\ g_{\upsilon N}^{(\ell)}(x_i,x_j)\ \ (\ell\in\{e,\tau,p\},\ \upsilon\in\{b,\mu\},<i,j>\in R_m).$$

The tests for verification of relation type, i.e. equivalence or tolerance, and the weak or strong form of the preference relation have also been developed by the author (see Klukowski 2011a, Chap. 10).

5. Solving of optimization problems

Minimization of the functions (2.19), (2.20) is, in general, not an easy problem, because of the dimensions of the feasible set. Currently, the algorithms are available only for ranking problems based on binary single comparisons (see David, 1988, Ch. 2, Hansen P., et al 1994); they refer to the dynamic programming or branch-and-bound algorithms, some of them can be used for known n . The algorithms are efficient for the moderate number of elements m. In the case of large m, the problems can be also solved with the use of heuristic algorithms: genetic (Falkenauer, 1998), artificial neural networks, random search (Ripley, 2006), swarm intelligence (Abraham and Grosan, 2006), etc.

In the case of multivalent comparisons the exact algorithms are not available now. The problems with moderate number of elements m, i.e. 3–12, can be solved with the use of complete enumeration. Problems with higher number of elements can be solved using heuristic algorithms, mentioned above.

It is obvious that the estimators based on multivalent comparisons require more computations than those based on binary comparisons. However, speed of computers increases quickly and computational problems will disappear in a near future.

It seems that computers based on new quantum technology will allow for solving the problems without significant restrictions on the number of elements m. New optimization algorithms have to be developed for such computers.

6. Summary - achievements of the work and further researches

The work here contained constitutes the synthesis of main results of the author concerning estimation of three relations – equivalence, tolerance, and preference – on the basis of pairwise comparisons with random errors (see Klukowski in Literature). The problems of that type occur often in applications and have been investigated in statistical literature. Therefore, it appears reasonable to devote to them an entire individual work.

The following new results, presented here, should be emphasized.

1^0. 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 in a pair, while multivalent data – quantitative features, i.e. the number of subsets including both elements (tolerance relation) or distance between elements - in the form of difference of ranks (preference relation).

2^0. The assumptions concerning the comparison errors are weaker than those commonly used in the literature, especially:

a) expected values of comparison errors can differ from zero,

b) distributions of comparison errors may be unknown,

c) comparisons including the same element can be correlated.

Therefore, the algorithms proposed can be used in the cases, when the existing algorithms are not applicable (can produce incorrect results).

3^0. Two estimators have been examined; the first one is based on the sum of differences between the relation form and the comparison data, the second is based on differences between the relation form and the median from comparisons of each pair. The estimators have a simple intuitive form, i.e. optimization tasks, and 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 higher cost of computations. The median estimator requires a lower amount of computations in the case of application of optimization algorithms, and is more robust (robustness is important property in the case of multivalent comparisons).

4^0. The analytical properties of the estimators have been complemented with the results of simulation study. This allows for determining of parameters, especially the number of comparisons N, guaranteeing the required precision of estimates; a definite value of N provides for the frequency of errorless result close to one or equal one. The simulation approach allows for evaluation of the distribution of frequencies of errorless solution also in the case of unknown distributions of comparison errors. Such distributions are replaced by some boundary distributions – the quasi-uniform distributions, proposed by the author. The simulation study indicates an excellent efficiency of multivalent estimators – the original concept of the author; the errorless estimate can be obtained for moderate N and the probability of errorless comparison lower than ½.

5^0. The properties of estimates can be thoroughly validated; validation comprises the fact of existence of the relation and the assumptions as to the comparison errors. The assumptions can be verified with the use of known tests and the methods proposed by the author. The establishment of existence of relation can also be based on simulation approach. It is possible, as well, to choose the relation type – equivalence or tolerance, and the type of the preference relation – strict or weak. Therefore, the approach has the features of data mining techniques.

6^0. The precision of the estimators, examined in the simulation study (Chapter 9, Klukowski, 2011a, Klukowski – to appear), are based on measures proposed in the work: • frequency of the errorless estimate, • average absolute, one-dimensional error, and • distribution of the average absolute, one-dimensional error. The one-dimensional error is the sum of components of the multi-dimensional error. It is an adequate measure of difference between the estimate and the relation; however, multi-dimensional error can also be subject to analysis, especially in graphical form.

7^0. The approach proposed allows for combining of comparisons obtained from different sources, e.g. statistical tests, experts, neural networks. It is also possible to combine binary and multivalent data and to apply two-stage estimators, based, in the first stage, on binary comparisons, and in the second stage – on multivalent comparisons, obtained in the first stage.

8^0. The estimates are obtained on the basis results from optimization tasks. They can be solved with the use of complete enumeration of the feasible set or the heuristic algorithms. The first approach requires fast processors, which are available currently. Heuristic algorithms can be based on random search, genetic algorithms, swarm intelligence, or hierarchical agglomeration algorithms.

9^0. The approach presented will be developed in the following directions: statistical learning, estimation of more complex structures of data (e.g. hierarchical), multidimensional (multi-criteria) pairwise comparisons, etc. An important field is also constituted by application of the estimators and tests developed.

Literature

Abraham, A., Grosan C., eds., (2006) *Swarm Intelligence in Data Mining*. Springer

Brunk, H. D. (1960) *Mathematical models for ranking from paired comparisons*. JASA Vol. 55, 503 – 520.

Bradley, R. A. (1976) *Science, statistics and paired comparisons*. Biometrics vol. 32, 213 – 232.

Bradley, R. A. (1984) *Paired comparisons: some basic procedures and examples*. Handbook of Statistics Vol. 4, ed. P. R. Krishnaiah and P. K. Se. Amsterdam: North-Holland, 299 – 326.

Daniel, W. W. (1990) *Applied nonparametric Statistics* (2nd ed.). Boston: PWS-Kent Publishing Company.

David H. A. (1970) *Order Statistics*. J. Wiley.

David H. A. (1988) *The Method of Paired Comparisons*, 2nd ed. Ch. Griffin, London.

Davidson, R., R. (1976). *A bibliography on the Method of Paired Comparisons*. Biometrics Vol. 32, 241 – 251.

DeCani, J. S. (1972) *A branch and bound algorithm for maximum likeli-hood paired comparison ranking*. Biometrika 59, 131 – 136.

Domański C. (1990) *Testy statystyczne* (*Statistical tests*), in Polish. PWE, Warszawa.

Falkenauer, E. (1998) *Genetics Algorithms and Grouping Data*. J. Wiley.

Gordon, A., D. (1999) *Classification*, 2nd Edition, Chapman&Hall/CRC.

Hand, D. J. (1986) *Discrimination and Classification*. J. Wiley.

Hartigan, J. J. (1975) *Clustering Algorithms*. J. Wiley.

Hansen, P., Jaumard. B., Sanlaville E. (1994) *Partitioning Problems in Cluster Analysis: A Review of Mathematical Programming Approaches*. Studies in Classification, Data Analysis, And Knowledge Organization, Springer - Verlag.

Hastie, T., Tibshirani, R., Friedman, J. (2002) *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Springer

Hoeffding, W. (1963) *Probability inequalities for sums of bounded random variables*. JASA, Vol. 58, 13 – 30.

Hollander M., Wolfe D. A. (1973) *Nonparametric Statistical Methods*. J. Wiley.

Kaufman, L., Rousseeuw, P. J. (1990) *Findings Groups In Data: An Introduction to Cluster Analysis*. J. Wiley.

Klukowski L. (1990a) *Algorithm for classification of samples in the case of unknown number variables generating them* (in Polish). Przegląd Statystyczny XXXVII, 167 – 177.

Klukowski L. (1990b) *Ranking of items on the basis of pairwise comparisons in the case of random errors* (in Polish). In: Kulikowski, R., Kacprzyk, J. (eds.) SYSTEMS RESEARCH, Vol. 3, Omnitech Press, Warsaw, 212 - 268.

Klukowski L. (1991) *Classification of hormonal profiles* (in Polish). The II-nd Conference of Operational and Systems Research. Nahorski Z. (ed.), IBS PAN, Warsaw, 155 - 163.

Klukowski L. (1994) *Some probabilistic properties of the nearest adjoining order method and its extensions*. Annals of Operational Research, **51**, 241 – 261.

Klukowski L. (1997): *The nearest adjoining order method for pairwise comparisons in the form of difference of ranks*. International Conference TRANSITION to Advanced Market Institutions and Economies, Systems Research Institute, Polish Ac. of Sciences, 1997, INFORMS, IFORS, IFAC, IIASA, Warsaw, June 18-21, 220 -225.

Klukowski L. (2000): *The nearest adjoining order method for pairwise comparisons in the form of difference of ranks*. Annals of Operations Research vol. 97, 357-378.

Klukowski, L. (2002) *Estimation of the tolerance relation on the basis of pairwise comparisons with random errors* (in Polish). In: Bubnicki, Z., Hryniewicz, O., Kulikowski, R. (eds.) Methods and Techniques of Data Analysis and Decision Support. Ac. Publ. House EXIT, Warsaw, V-21 – V-35.

Klukowski L. (2006) *Tests for relation type - equivalence or tolerance - in finite set of elements*. Control and Cybernetics, **35**, 369 - 384.

Klukowski, L. (2007a) *Estimation of the Preference Relation on the Basis of Medians from Pairwise Comparisons in the Form of Difference of Ranks*. In Kurzynski (et al) Computer Recognition Systems 2, Advances in Soft Computing 45, 232 – 241, Springer.

Klukowski L. (2007b) *Estimation of tolerance relation the basis of multiple pairwise comparisons with random errors*. Control and Cybernetics, **36**, 443 - 466.

Klukowski L. (2007c) *Estimation of the preference relation on the basis of multiple pairwise comparisons with random errors*. In: Trzaskalik T. (eds.) MODELOWANIE PREFERENCJI A RYZYKO'08, AE Katowice 2008.

Klukowski L. (2008a) *Determination of tolerance relation – alternative approach to intuitionistic and fuzzy sets*. Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets an Related Topics. Vol. II., 85 – 94, K. Atanassov (et al), eds. Ac. Publ. House EXIT, Warsaw.

Klukowski L. (2008b) *Estimation of the preference relation the basis of multiple pairwise comparisons in the form of differences of ranks*. Control and Cybernetics, **37**, 711-729.

Klukowski L. (2008c) *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. I. K. Atanassov (et al), eds. Ac. Publ. House EXIT, Warsaw, 175 – 189.

Klukowski L. (2008d) *Estimation of the preference relation on the basis of medians from pairwise comparisons*. In: Owsiński, J., W., Bruggemann (eds.) Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues, IBS PAN, Warsaw, 109 - 120.

Klukowski L. (2010a) *Estimation of the relations: preference, equivalence and tolerance on the basis of pairwise comparisons.* Eighth Workshop on Intuitionistic Fuzzy Sets and generalized Nets. In: Development In Fuzzy Sets, Intuitionistic Fuzzy Sets, Vol. I: Foundations. IBS PAN, SRI PAS, 109 - 120.

Klukowski L. (2010b) *Tests for relation type – equivalence or tolerance based on medians from multiple pairwise comparisons.* In: Bojar W., Owsiński J., W. (eds.) MATERIAŁY I STUDIA POLSKIEGO STOWARZYSZENIA ZARZĄDZANIA WIEDZĄ. nr 31, 338 – 348, Bydgoszcz.

Klukowski L. (2010c) *Estimation of the preference relation on the basis of pairwise comparisons indicating direction of preference with random errors* (in Polish). In: Owsiński J., W. (eds.) MATERIAŁY I STUDIA POLSKIEGO STOWARZYSZENIA ZARZĄDZANIA WIEDZĄ, Nr 31, 57 – 64, Bydgoszcz.

Klukowski, L. (2011a) *Methods of Estimation of Relations of: Equivalence, Tolerance, and Preference in a Finite Set.* IBS PAN, Series: Systems Research, Volume 69, Warsaw.

Klukowski, L. (2011b) *Estimation of Tolerance Relation on the Basis of Pairwise Comparisons.* In: Burduk, R., (et al) Computer Recognition Systems 4, Advances in Intelligent and Soft Computing 95, 232 – 241, Springer.

Klukowski L. (2011c) *Tests for validation of estimates obtained on the basis of pairwise comparisons with random errors.* In Atanassov (at al): Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets and Related Topics, Vol. II: Applications, 75 – 91. SRI PAS, Warsaw.

Klukowski L. (2012a) *Properties of ostimators of the preference relation based on pairwise comparisons – simulation survey.* In Atanassov K. T. (at al, eds.) New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, 75 - 90. Volume II: Applications, SRI PAS, Warsaw.

Klukowski L. (2012b) *Properties of estimators of the preference relation in the form of difference of ranks – simulation survey* (in Polish). In Trzaskalik T.: Modelowanie Preferencji a Ryzyko'12. Studia Ekonomiczne, vol. 97, 119 – 132, Economic University, Katowice.

Klukowski, L. (to appear) *Properties of estimators of the preference relation based on pairwise comparisons: binary and difference of ranks - simulation survey.* Submitted to Control and Cybernetics.

Kohonen T. (1995) *Self-Organizing Maps.* Springer-Verlag.

Koronacki, J., Ćwik J. (2005) *Statystyczne systemy uczące się* (*Statistical Learning Systems*), in Polish. WNT, Warszawa.

Randles R. H., Wolfe D., A. (1979) *Introduction to The Theory of Nonparametrics Statistics.* J. Wiley.

Remage, R., Jr. And Thompson, W. A. Jr. (1966) *Maximum-likelihood paired comparison rankings.* Biometrika 53, 143 – 149.

Ripley B. D. (2006) *Stochastic Simulation.* J. Wiley.

Sachs L. (1978) *Applied Statistics,* Springer-Verlag, New York

Serfling, R., J. (1980): Approximation Theorems of Mathematical Statistics. J. Wiley.

Sheskin, D. J. (1997) *Handbook of Parametric and Nonparametric Statistical Procedures.* CRC PressBoca Raton, Boston, London, New York, Washington, D. C.

Siegel, S. and Castellan, N. J. Jr. (1988) *Nonparametric Statistics for the Behavioral Sciences* (2nd ed.). New York: McGraw-Hill Book Company.

Slater P. (1961) *Inconsistencies in a schedule of paired comparisons.* Biometrika, 48, 303 - 312.