Raport Badawczy
Research Report

RB/17/2013

Estimation of the partial order on the basis of pairwise comparisons in binary multivalent form

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. dr hab. inż. Zbigniew Nahorski

Warszawa 2013

Estimation of the partial order on the basis of pairwise comparisons in binary and multivalent form

Leszek Klukowski

Systems Research Institute Polish Academy of Sciences,

Newelska 6, 02-792 Warsaw

e-mail: Leszek.Klukowski@ibspan.waw.pl

Abstract

The problem of estimation of the partial order on the basis of multiple pairwise comparisons with random errors, in binary and multivalent form is investigated. The estimators are based on the idea of the nearest adjoining order (see Slater 1961, Klukowski 2011). Two approaches are examined: comparisons indicating direction of preference (binary) and comparisons indicating difference of ranks (multivalent) - both with possibility of incomparable elements. The properties of estimators and optimization problems for estimates obtaining are similar as in the case of complete relation. The assumptions about distributions of comparisons errors are not the same – they comprise the case of incomparable elements.

Keywords: estimation of partial order, multiple pairwise comparisons with random errors, binary and multivalent comparisons.

1 Introduction

The problem of estimation of complete preference relation on the basis of multiple pairwise comparisons in binary and multivalent form with random errors has been considered in Klukowski 1994, 2011 Chapt. 7 – 11. The same approach can be applied to the partial order – the main difference is taking into account incomparable elements. This fact indicates the following modifications: equivalent elements are not allowed, distributions of comparisons errors include probabilities related to incomparable elements, an aggregation of comparisons of individual pairs with the use of median is not considered.

The idea of estimation is minimization of differences between relation form, expressed in a specified way, and comparisons (Slater 1961). Thus, the estimates are obtained as optimal solutions of the integer programming problems.

The approach rests on statistical paradigm; therefore, provides properties of estimates and possibility of verification of the results. The main property is consistency, for number of comparisons (of each pair) $N \to \infty$, under weak assumptions about comparison errors. In the case of binary comparisons it is assumed that probability of a correct comparison is greater than incorrect one. In the case of multivalent comparisons, expressing differences of ranks of comparable elements, it is assumed that distributions of comparisons are unimodal with mode and median equal zero. In the case of pairs including incomparable elements, it is assumed that the probability of correct recognition of incomparability is greater than ½. The estimators can be applied also in the case of unknown distributions of comparison errors, which have to satisfy the assumptions made

In earlier works of the author (Klukowski 1994, 2008, 2011) two kinds of estimators have been considered: the first one based on total sum of differences, between relation form and comparisons, and the second - based on sum of differences with medians of comparisons of each pair. The second estimator requires lower computational cost, what is important for large N. In the case of partial orders, such the estimator can be applied, in simple way, only for binary comparisons. Thus, the median case is omitted in the work.

The idea of the estimators, in the case of binary comparisons and complete relation, was presented in Slater (1961); some other ideas in the area of pairwise comparisons have been presented in: David (1988), Bradley (1984), Flinger, Verducci (eds.) (1993).

The paper consists of four sections and appendix with the proof of the theorem from section 3. The second section presents definitions, notations and assumptions about comparisons errors. In next section is considered the form of estimators, for both kinds of comparisons, and their properties. Last section summarizes the results.

2 Definitions, notations and assumptions about comparisons errors

2.1 Definitions and notations

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

We are given a finite set of elements $\mathbf{X} = \{x_1, ..., x_m\}$ ($3 \le m < \infty$). There exists in the set $\mathbf{X}$ the partial order relation $\mathbf{R}^{(p)}$. Each pair of elements (x_i, x_j) is ordered or incomparable; thus the set of indices:

$$R_{ij} = \{<i, j> \mid i = 1, ..., m, j = i+1, ..., m\} \qquad (1)$$

can be divided into two disjoint subsets, including comparable I_o and incomparable I_c pairs of indexes, i.e.: $R_m = I_o \cup I_n$.

The partial order relation can be expressed in binary and multivalent way. The binary description $T_b(x_i, x_j)$ $(<i, j> \in R_m)$, expresses direction of preference in a pair of elements or their incomparability; it assumes the form:

$$T_b(x_i, x_j) = \begin{cases} -1 \; if \; x_i \; precedes \; x_j, \\ 1 \; if \; x_j \; precedes \; x_i; \\ 2 \; if \; x_i \; and \; x_j \; incomparable. \end{cases} \quad (2)$$

The multivalent description $T_\mu(x_i, x_j)$ expresses the difference of ranks of comparable elements, denoted d_{ij}, or their incomparability; $d_{ij} = r - s$ determines a distance between the elements: r is a rank of x_i, s is a rank of x_j. The distance can be presented at a digraph – it is a number of edges connecting elements of a pair; it has to be lower than m. This description assumes the form:

$$T_\mu(x_i, x_j) = \begin{cases} d_{ij} \; if \; elements \; x_i \; and \; x_j \; are \; comparable, \\ m \; if \; elements \; x_i \; and \; x_j \; are \; not \; comparable. \end{cases} \quad (3)$$

The values of the binary description are included in the set $\{-1, 1\} \cup \{2\}$, the values of multivalent description - in the set $\{-(m-1), ..., -1, 1, ..., m-1\} \cup \{m\}$. The sets including "comparable values" – binary and multivalent will be denoted – respectively $\wp_b$ and $\wp_\mu$:

$$\wp_b = \{-1, 1\}, \quad \wp_\mu = \{-(m-1), ..., -1, 1, ..., m-1\}. \quad (4)$$

Examples of the values $T_\upsilon(x_i, x_j)$ $(\upsilon \in \{b, \mu\})$.

The relation form – a partial order:

x_1 precedes x_2, x_1 precedes x_3, x_2 and x_3 incomparable, x_2 precedes x_4, x_3 precedes x_4, x_4 precedes x_5,

i.e.: $I_o = \{<1,2>, <1,3>, <1,4>, <1,5>, <2,4>, <2,5>, <3,4>, <3,5>, <4,5>\}$,
$I_n = \{<2,3>\}$.

The values of $T_\upsilon(x_i, x_j)$ $(\upsilon \in \{b, \mu\})$ assume the form:

$$T_b(x_i, x_j) = \begin{bmatrix} \times & -1 & -1 & -1 & -1 \\ & \times & 2 & -1 & -1 \\ & & \times & -1 & -1 \\ & & & \times & -1 \\ & & & & \times \end{bmatrix}, \quad T_\mu(x_i, x_j) = \begin{bmatrix} \times & -1 & -1 & -2 & -3 \\ & \times & 5 & -1 & -2 \\ & & \times & -1 & -2 \\ & & & \times & -1 \\ & & & & \times \end{bmatrix}.$$

2.2 Assumptions about comparison errors

The relation form, i.e. the function $T_b(x_i, x_j)$ or $T_\mu(x_i, x_j)$, has to be determined (estimated) on the basis of N $(N \geq 1)$ comparisons of each pair (x_i, x_j) $(<i, j> \in R_m)$ in binary form or multivalent form, disturbed by random errors. The form of the function $T_\upsilon(x_i,x_j)$ $(\upsilon \in \{b,\mu\})$ has to be compatible with comparisons; they will be denoted – respectively - $g_{bk}(x_i,x_j)$ and $g_{\mu k}(x_i,x_j)$ $(k=1,\dots,N)$. The comparison errors - respectively $\phi^*_{bk}(x_i,x_j)$ or $\phi^*_{\mu k}(x_i,x_j)$ can be expressed in the following form:

$$\phi^*_{bk}(x_i,x_j)=\begin{cases}0 \; if \; g_{bk}(x_i,x_j) \; and \; T_b(x_i,x_j) \; are \; the \; same, \\ 1 \; f \; g_{bk}(x_i,x_j) \; and \; T_b(x_i,x_j) \; are \; not \; the \; same,\end{cases} \quad (5)$$

$$\phi^*_{\mu k}(x_i,x_j)=\begin{cases}0 \; if \; g_{\mu k}(x_i,x_j)=m \; and \; T_\mu(x_i,x_j)=m, \\ g_{\mu k}(x_i,x_j)-T_\mu(x_i,x_j) \; if \; g_{\mu k}(x_i,x_j), T_\mu(x_i,x_j)\neq m, \\ 2m-1 \; in \; other \; cases.\end{cases} \quad (6)$$

The distributions of comparison errors have to satisfy the following assumptions.

A1. Any comparison $g_{\upsilon k}(x_i,x_j)$ $(\upsilon \in \{b,\mu\};\; k=1,\dots,N;\; <i,j>\in R_m)$, is an evaluation of the value $T_\upsilon(x_i,x_j)$; the probabilities of errors $P(\phi^*_{\upsilon k}(x_i,x_j)=l)$ $(l\in\{0,1\})$ have to satisfy the following assumptions:

$$P(\phi^*_{bk}(x_i,x_j)=0)\geq 1-\delta \quad (\delta\in(0,\tfrac{1}{2})), \quad (7)$$

$$P(\phi^*_{bk}(x_i,x_j)=0)+P(\phi^*_{bk}(x_i,x_j)=1)=1, \quad (8)$$

$$\sum_{l\leq 0} P(\phi^*_{\mu k}(x_i,x_j)=l \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)\neq m)\geq 1-\delta, \quad (\delta\in(0,\tfrac{1}{2})), (9)$$

$$\sum_{l\geq 0} P(\phi^*_{\mu k}(x_i,x_j)=l \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)\neq m)\geq 1-\delta, \quad (\delta\in(0,\tfrac{1}{2})),$$

(10)

$$P(\phi^*_{\mu k}(x_i,x_j)=l \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)\neq m, l\geq 0)\geq \\ \geq P(\phi_{\mu k}(x_i,x_j)=l+1 \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)\neq m, l\geq 0), \quad (11)$$

$$P(\phi^*_{\mu k}(x_i,x_j)=l \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)\neq m, l\leq 0)\geq \\ \geq P(\phi_{\mu k}(x_i,x_j)=l-1 \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)\neq m, l\leq 0), \quad (12)$$

$$\sum_{l\neq 2m-1} P(\phi^*_{\mu k}(x_i,x_j)=l \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)\neq m)+ \\ +P(\phi^*_{\mu k}(x_i,x_j)=2m-1 \mid T_\mu(x_i,x_j)\neq m)=1. \quad (13)$$

$$P(\phi^*_{\mu k}(x_i,x_j)=0 \mid g_{\mu k}(x_i,x_j),T_\mu(x_i,x_j)=m)\geq 1-\delta, \quad (\delta\in(0,\tfrac{1}{2})), (14)$$

$$P(\phi^*_{\mu k}(x_i,x_j)=0 \mid g_{\mu k}(x_i,x_j), T_\mu(x_i,x_j)=m)+$$
$$+P(\phi_{\mu k}(x_i,x_j)=2m-1 \mid g_{\mu k}(x_i,x_j)\neq m, T_\mu(x_i,x_j)=m)=1, \quad (15)$$

$$P(\phi^*_{\mu k}(x_i,x_j)=2m-1 \mid g_{\mu k}(x_i,x_j)=m, T_\mu(x_i,x_j)\in \wp_\mu)\leq$$
$$\leq P(\phi^*_{\mu k}(x_i,x_j)=l \mid l\neq 2m-1, T_\mu(x_i,x_j)\in \wp_\mu). \quad (16)$$

$$P(\phi^*_{\mu k}(x_i,x_j)=2m-1 \mid g_{\mu k}(x_i,x_j)=m, T_\mu(x_i,x_j)\in \wp_\mu)+$$
$$+\sum_{l\neq 2m-1}(\phi^*_{\mu k}(x_i,x_j)=l \mid T_\mu(x_i,x_j)\in \wp_\mu)=1. \quad (17)$$

A2. The comparisons: $g_{bk}(x_i,x_j)$ $(k=1,\ldots,N;\ <i,j>\in R_m)$ are independent random variables and the comparisons: $g_{\mu k}(x_i,x_j)$ $(k=1,\ldots,N;\ <i,j>\in R_m)$ are independent random variables.

The assumptions about comparisons reflect the following facts.

In the case of binary comparisons the probability of a correct comparison is greater than incorrect one (see (7), (8)).

In the case of multivalent comparisons the following properties hold true. The probability of correct detection of incomparable pair is greater than ½ (see (14), (15)). The distribution of the error, in than case of comparable pair, is unimodal with mode and median equal zero (see (10) – (13)). The probability of incorrect detection of incomparable pair is not greater than any probability of any incorrect difference of ranks (see (16)).

The assumption about independency of comparisons can be relaxed in such way that comparisons of the same pair are independent and comparisons of pairs comprising different elements are independent.

The random variables $\phi_{bk}(x_i,x_j)$ and $\phi_{\mu k}(x_i,x_j)$ for any partial order, denoted respectively by $t_b(x_i,x_j)$ and $t_\mu(x_i,x_j)$, can be expressed in the following form:

$$\phi_{bk}(x_i,x_j)=\begin{cases}0 \textit{ if } g_{bk}(x_i,x_j) \textit{ and } t_b(x_i,x_j) \textit{ are the same},\\ 1 \textit{ if } g_{bk}(x_i,x_j) \textit{ and } t_b(x_i,x_j) \textit{ are not the same},\end{cases}$$

$$\phi_{\mu k}(x_i,x_j)=\begin{cases}0 \textit{ if } g_{\mu k}(x_i,x_j)=m \textit{ and } t_\mu(x_i,x_j)=m,\\ g_{\mu k}(x_i,x_j)-t_\mu(x_i,x_j) \textit{ if } g_{\mu k}(x_i,x_j), t_\mu(x_i,x_j)\neq m,\\ 2m-1 \textit{ in other cases}.\end{cases}$$

3 Estimation problems and properties of estimates

The idea of estimation problems is to minimize differences between comparisons, in binary or multivalent form, and relation, expressed in compatible way. Thus, the estimates $\hat{T}_b(x_i,x_j)$ or

$\hat{T}_\mu(x_i,x_j)$ $(<i,j>\in R_m)$ are the optimal solutions of discrete programming problems – respectively:

$$\min_{F_X}\{\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\phi_{bk}(x_i,x_j)\}, \tag{18}$$

$$\min_{F_X}\{\sum_{<i,j>\in R_m}\sum_{k=1}^{N}\phi_{\mu k}(x_i,x_j)\}, \tag{19}$$

where:

F_X - feasible set, i.e. family of all partial orders in the set **X**,

$\phi_{\upsilon k}(x_i,x_j)$ $(\upsilon\in\{b,\mu\})$ - differences between comparisons and any relation from a family F_X.

In the book Klukowski (2011) has been proved consistency of such estimates, as $N\to\infty$, in the case of the complete preference relation. The proofs of the property are based on the following facts. Firstly the expected value of the random variables:

$$W_b^* = \sum_{<i,j>\in R_m}\sum_{k=1}^{N}\varphi_{bk}^*(x_i,x_j), \tag{20}$$

$$W_\mu^* = \sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\phi_{\mu k}^*(x_i,x_j)\right|, \tag{21}$$

expressing differences between comparisons and actual relation ($T_b(x_i,x_j)$ or $T_\mu(x_i,x_j)$ $(<i,j>\in R_m)$), are lower than expected values of the variables:

$$\widetilde{W}_b = \sum_{<i,j>\in R_m}\sum_{k=1}^{N}\widetilde{\phi}_{bk}(x_i,x_j), \tag{22}$$

$$\widetilde{W}_\mu = \sum_{<i,j>\in R_m}\sum_{k=1}^{N}\left|\widetilde{\phi}_{\mu k}(x_i,x_j)\right|, \tag{23}$$

expressing differences between comparisons and any other relation, expressed by $\widetilde{T}_b(x_i,x_j)$ or $\widetilde{T}_\mu(x_i,x_j)$. Secondly, the variances of these variables, i.e.: $Var(\frac{1}{N}W_b^*)$, $Var(\frac{1}{N}W_\mu^*)$, $Var(\frac{1}{N}\widetilde{W}_b)$, $Var(\frac{1}{N}\widetilde{W}_\mu)$, converges to zero, as $N\to\infty$. Thirdly, the probabilities: $P(W_b^*<\widetilde{W}_b)$ and $P(W_\mu^*<\widetilde{W}_\mu)$ converge to one, as $N\to\infty$; the speed of convergence is determined by exponential subtrahend. These relationships can be formulated shortly in the following

Theorem

The following relationships hold true:

$$E(W_b^*) < E(\widetilde{W}_b), \tag{24}$$

$$E(W_\mu^*) < E(\widetilde{W}_\mu), \tag{25}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N} W_b^*) = 0, \quad \lim_{N\to\infty} Var(\tfrac{1}{N} \widetilde{W}_b) = 0, \tag{26}$$

$$\lim_{N\to\infty} Var(\tfrac{1}{N} W_\mu^*) = 0, \quad \lim_{N\to\infty} Var(\tfrac{1}{N} \widetilde{W}_\mu) = 0, \tag{27}$$

$$P(W_b^* < \widetilde{W}_b) \geq 1 - \exp\{-2N(\tfrac{1}{2} - \delta)^2\}, \tag{28}$$

$$P(W_\mu^* < \widetilde{W}_\mu) \geq 1 - \exp\{-2N\widetilde{\theta}^2\}, \tag{29}$$

where: $\widetilde{\theta}$ - positive constant, dependent on $\widetilde{T}_\mu(x_i, x_j)$.

Proof – Appendix.

The relationships (24) – (29) are the basis for the estimates $\hat{T}_b(x_i, x_j)$ and $\hat{T}_\mu(x_i, x_j)$ - indicate their consistency. It is so, because the actual relation generates random variables W_b^* or W_μ^* with minimal expected values in the family F_X and variances converging to zero. The optimal solutions of the problems (18) and (19), determining relations with minimal values of differences with comparisons, detect such the relation with probability converging to one. The approach can be applied in the case of unknown probabilities of comparison errors; it is especially important in the case of multivalent comparisons.

Minimization of the functions (18), (19) is not easy problem. For a low number *m*, of elements of the set **X**, i.e. several, they can be solved with a use of complete enumeration. For moderate *m* the problem with binary comparisons can be solved with the use of optimization software. In remaining cases heuristic algorithms are necessary (see also Hansen, Jaumard, Sanlaville 1994).

4 Concluding remarks

The paper presents estimators of the partial order relation on the basis of pairwise comparisons in binary and multivalent form. They have similar properties as estimators of complete relation – in particular consistency and speed of convergence. The case of binary comparisons is similar to the complete case, i.e. assumptions about comparisons errors and the form of the estimators. In the case of multivalent comparisons the assumptions about distributions of errors are more complex and lead to elimination the approach based on medians from comparisons. The estimators can be applied also to other structures of data, especially trees.

Appendix

Proof of the Theorem (relationships (24) – (29)).

The inequality (24), i.e. $E(W_b^*) < E(\widetilde{W}_b)$ can be proved in similar way as the inequality (32) in Klukowski (1994).

The expected value of the difference $W_b^* - \widetilde{W}_b$ assumes the form:

$$E(W_b^* - \widetilde{W}_b) =$$

$$E\Big(\sum_{<i,j>\in R_m} \sum_{k=1}^{N} \phi_{bk}^*(x_i,x_j) - \sum_{<i,j>\in R_m} \sum_{k=1}^{N} \widetilde{\phi}_{bk}(x_i,x_j)\Big) =,$$

$$\sum_{k=1}^{N} \sum_{<i,j>\in R_m} E(\phi_{bk}^*(x_i,x_j) - \widetilde{\phi}_{bk}(x_i,x_j)).$$

It is clear that each component $E(\phi_{bk}^*(x_i,x_j) - \widetilde{\phi}_{bk}(x_i,x_j))$ can be: zero or negative, because the probability of correct comparison is greater than 1/2; the value of zero corresponds to the case $T_b(x_i,x_j) = \widetilde{T}_b(x_i,x_j)$, negative – to the case $T_b(x_i,x_j) \neq \widetilde{T}_b(x_i,x_j)$. This fact is sufficient for the inequality (24).

The inequality (25) can be proved in similar way, however it is more cumbersome; the case of complete relation, is presented in Klukowski (2008)). Let us consider two possible cases: $T_\mu(x_i,x_j) = m$, $\widetilde{T}_\mu(x_i,x_j) \neq m$ and opposite $T_\mu(x_i,x_j) \neq m$, $\widetilde{T}_\mu(x_i,x_j) = m$.

In the first case:

$$E(|\phi_{\mu k}^*(x_i,x_j)| - |\widetilde{\phi}_{\mu k}(x_i,x_j)|\,; T_\mu(x_i,x_j) = m) =$$

$$E(|\phi_{\mu k}^*(x_i,x_j)|\ ; T_\mu(x_i,x_j) = m) - E(|\widetilde{\phi}_{\mu k}(x_i,x_j)|\ ; T_\mu(x_i,x_j) = m) =$$

$$E(|\phi_{\mu k}^*(x_i,x_j)|\ ; T_\mu(x_i,x_j) = m) = \qquad \text{(A1)}$$
$$(2m-1)P(\phi_{\mu k}^*(x_i,x_j) = 2m-1\,; T_\mu(x_i,x_j) = m),$$

$$E(|\widetilde{\phi}_{\mu k}(x_i,x_j)|\ ; T_\mu(x_i,x_j) = m) = E(|\widetilde{\phi}_{\mu k}(x_i,x_j)|\ ; \widetilde{T}_\mu(x_i,x_j) \neq m) =$$
$$(2m-1)P(|\widetilde{\phi}_{\mu k}(x_i,x_j)| = 2m-1\,; \widetilde{T}_\mu(x_i,x_j) \neq m) + \qquad \text{(A2)}$$
$$+ \sum_{l\in\wp_\mu} |l - \widetilde{T}_\mu(x_i,x_j)| (P(g_{\mu k}(x_i,x_j) = l).$$

The value of (A1) is lower than (A2); it is so because $P(\phi_{\mu k}^*(x_i,x_j) = 2m-1\,; T_\mu(x_i,x_j) = m) < 1/2$ and $P(\widetilde{\phi}_{\mu k}(x_i,x_j) = 2m-1 |\ \widetilde{T}_\mu(x_i,x_j) \neq m) > 1/2$.

In the second case:

$$E(|\phi^*_{\mu k}(x_i,x_j)|\ ;T_\mu(x_i,x_j)\neq m)=$$
$$(2m-1)P(\phi^*_{\mu k}(x_i,x_j)=2m-1\,;T_\mu(x_i,x_j)\neq m)+ \qquad \text{(A3)}$$
$$\sum_{l\in\wp_\mu}|\ l-T_\mu(x_i,x_j)|\,P(g_{\mu k}(x_i,x_j)=l;T_\mu(x_i,x_j)\neq m),$$

$$E(|\tilde{\phi}_{\mu k}(x_i,x_j)|\ ;\tilde{T}_\mu(x_i,x_j)=m)=$$
$$(2m-1)\sum_{l\in\wp_\mu}P(g_{\mu k}(x_i,x_j)=l\ |\ \tilde{T}_\mu(x_i,x_j)=m). \qquad \text{(A4)}$$

The value of (A3) is lower than (A4), because:

- in (A3) the component with maximal probability $P(g_{\mu k}(x_i,x_j)=T_\mu(x_i,x_j)\,;T_\mu(x_i,x_j)\neq m)$ equals zero, the component $(2m-1)P(g_{\mu k}(x_i,x_j)=m\,;T_\mu(x_i,x_j)\neq m)$ is a product including factor equal to minimal probability, remaining components are products including probabilities lower than maximal and values lower than $2m-1$;
- in (A4) the component with minimal probability $P(g_{\mu k}(x_i,x_j)=m\,;T_\mu(x_i,x_j)\neq m)$ equals zero, while remaining part of probability (i.e. $1-P(g_{\mu k}(x_i,x_j)=m\,;T_\mu(x_i,x_j)\neq m)$ is multiplied by $2m-1$.

The proof of the inequality (25), for remaining values of $T_\mu(x_i,x_j)$, $\tilde{T}_\mu(x_i,x_j)$ $(<i,j>\in R_m)$, is similar.

The validity of the relationships (26) results from following facts:

- each random variable: $\sum_{<i,j>\in R_m}\phi^*_{bk}(x_i,x_j)$ and $\sum_{<i,j>\in R_m}\tilde{\phi}_{bk}(x_i,x_j)$ $(k=1,\ldots,N)$ has finite, bounded expected value and variance,
- the variances of the variables: $\frac{1}{N}\sum_{k=1}^{N}\sum_{<i,j>\in R_m}\phi^*_{bk}(x_i,x_j)$, $\frac{1}{N}\sum_{k=1}^{N}\sum_{<i,j>\in R_m}\tilde{\phi}_{bk}(x_i,x_j)$ are bounded; their values will be denoted - respectively: $\frac{1}{N}V^*_b$ and $\frac{1}{N}\tilde{V}_b$, where: V^*_b and V_b - maximal variances of the variables $\sum_{k=1}^{N}\sum_{<i,j>\in R_m}\phi^*_{bk}(x_i,x_j)$ $(k=1,\ldots,N)$ and $\sum_{k=1}^{N}\sum_{<i,j>\in R_m}\tilde{\phi}_{bk}(x_i,x_j)$ respectively;
- the values $\frac{1}{N}V^*_b$ and $\frac{1}{N}\tilde{V}_b$ converges to zero as $N\to\infty$.

The validity of the relationships (27) can be proved in similar way.

The validity of the inequalities (28), (29) can be proved on the basis of Hoeffding's (1963) inequalities for a sum of independent bounded random variables. The inequality assumes the form:

$$P(\sum_{k=1}^{N}Y_k-\sum_{k=1}^{N}E(Y_k)\geq Nt)\leq\exp(-2Nt^2/(b-a)^2), \qquad (*)$$

where:

$Y_1, ..., Y_N$ - independent random variables satisfying $P(a \le Y_k \le b) = 1, \quad a < b$,

t – positive constant.

The inequality (*) can be applied to the random variables $\sum_{k=1}^{N} \sum_{<i,j>\in R_m} (\phi^*_{\upsilon k}(x_i, x_j) - \widetilde{\phi}_{\upsilon k}(x_i, x_j))$ $(\upsilon \in \{b, \mu\})$, after a following transformation:

$$P(\sum_{k=1}^{N} \sum_{<i,j>\in R_m} (\phi^*_{\upsilon k}(x_i, x_j) - \widetilde{\phi}_{\upsilon k}(x_i, x_j)) < 0) =$$

$$1 - P(\sum_{k=1}^{N} \sum_{<i,j>\in R_m} (\phi^*_{\upsilon k}(x_i, x_j) - \widetilde{\phi}_{\upsilon k}(x_i, x_j)) \ge 0) =$$

$$1 - P(\sum_{k=1}^{N} \sum_{<i,j>\in R_m} (\phi^*_{\upsilon k}(x_i, x_j) - \widetilde{\phi}_{\upsilon k}(x_i, x_j)) -$$

$$E\sum_{k=1}^{N} \sum_{<i,j>\in R_m} (\phi^*_{\upsilon k}(x_i, x_j) - \widetilde{\phi}_{\upsilon k}(x_i, x_j)) \ge$$

$$\ge -E\sum_{k=1}^{N} \sum_{<i,j>\in R_m} (\phi^*_{\upsilon k}(x_i, x_j) - \widetilde{\phi}_{\upsilon k}(x_i, x_j))).$$

In the case of binary comparisons the value $(b-a)^2$, in inequality (*), equals one and (after simple transformations) $t = (1/2 - \delta)$; moreover the component: $-E\sum_{k=1}^{N} \sum_{<i,j>\in R_m} (\phi^*_{\upsilon k}(x_i, x_j) - \widetilde{\phi}_{\upsilon k}(x_i, x_j)$ is negative. These facts proves the inequality (28).

The proof of the inequality (29) is similar with such a difference that the value $(b-a)^2$ is different than one and the value of t cannot be expressed on the basis of δ; more precisely, the value of t depends on distributions of comparison errors and a value of $\widetilde{T}(x_i, x_j)$. The value of θ^2 in (29) is determined in similar way as in the case of complete relation (Klukowski (2008)).

References

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

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

[2] Flinger A. F., Verducci J. S. (eds.) (1993) *Probability Models and Statistical Analyses for Ranking Data*. Lecture Notes in Statistics 80, Springer-Verlag, New York, Berlin, Heidelberg.

[3] 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.

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

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

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

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

[8] Klukowski L. (2012) *Properties of estimators 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.

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