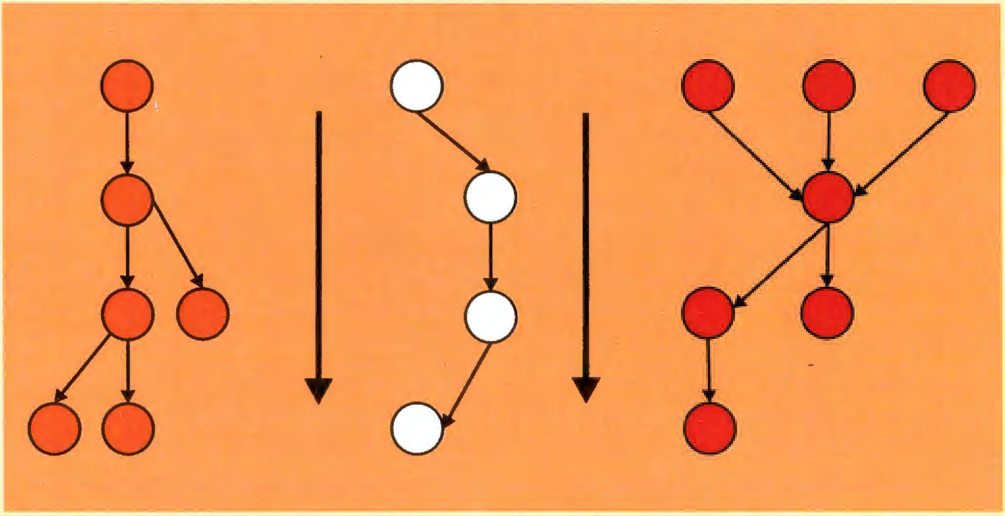
SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

MULTICRITERIA ORDERING AND RANKING:
PARTIAL ORDERS, AMBIGUITIES AND APPLIED ISSUES



Jan W. Owsiński and Rainer Brüggemann
Editors

Warsaw 2008

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

MULTICRITERIA ORDERING AND RANKING: PARTIAL ORDERS, AMBIGUITIES AND APPLIED ISSUES

Jan W.Owsiński and Rainer Brüggemann
Editors

Warsaw 2008

This book is the outcome of the international workshop held in Warsaw in October 2008 within the premises of the Systems Research Institute. All papers were refereed and underwent appropriate modification in order to appear in the volume. The views contained in the papers are, however, not necessarily those officially held by the respective institutions involved, especially the Systems Research Institute of the Polish Academy of Sciences.

ISBN 83-894-7521-9
EAN 9788389475213

Technical editing and typesetting:
Jan W.Owsiński, Anna Gostyńska, Aneta M.Pielak

Theoretical Developments

Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues
Jan W. Owsiński and Rainer Brüggemann, Editors

Estimation of the Preference Relation on the Basis of Medians From Pairwise Comparisons

Leszek Klukowski
Systems Research Institute Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland; (Leszek.Klukowski@ibspan.waw.pl)

The problem of estimation of the weak preference relation, with extension to partial order, on the basis of multiple pairwise comparisons with random errors is investigated. The estimators are based on the idea of the nearest adjoining order – NAO (see Slater, 1961). Two approaches are examined: comparisons indicating direction of preference and comparisons indicating difference of ranks. The properties of the estimators are based on distribution of median from comparisons and some probabilistic inequalities.

Keywords: estimation of the preference relation, multiple pairwise comparisons, nearest adjoining order method

1. Introduction

The paper presents some estimators of the weak preference relation (in a finite set) based on multiple pairwise comparisons with random errors. The estimators are based on the idea of the nearest adjoining order (NAO); it can be expressed in the following way – to determine the relation (estimate) with minimal number of inconsistencies with comparisons made (sample). Two types of comparisons are considered: the first one - indicating the direction of preference in each pair and the second - indicating the difference of ranks (positions in ranking). The assumptions about comparison errors are weaker than those commonly used in the literature (see, e.g., David, 1988; Kamishima and Akaho, 2006). In the case of direction of preference – each probability of correct comparison have to be greater than ½, in the case of difference of ranks – each distribution of comparison error have to be unimodal with median and maximum in zero. It is assumed also that comparisons of the same pair are independent (in stochastic sense). The estimate of the relation is obtained as a solution of some discrete programming problem, based on "aggregated" comparisons of each pair. Aggregation is typically done with the use of average or median. In the paper the median approach is examined, because it makes the optimisation problem easier.

The examples of the first approach (direction of preference) are discussed broadly in the literature (e.g. David, 1988). The second approach (differences of ranks) can be applied as a second step of estimation for rankings obtained from comparisons indicating direction of preference or to data in the form: the composer Y is better than Z for two classes.

The properties of distributions of comparison errors assumed in the paper may be verified with the use of statistical tests (for unimodality, mode, median, symmetry, etc.). The properties of the estimators are determined on the basis of well-known probabilistic inequalities and some properties of the order statistics (see David, 1970).

The paper consists of 6 Sections. In Sections 2 and 3, the problem formulation and basic definitions and notations are presented. The estimator corresponding to the case of direction of preference is discussed in Section 4, the estimator corresponding to difference of ranks - in Section 5. Last Section summarizes the results and suggests further research.

2. Formulation of the problem

It is assumed that there exists an unknown weak preference relation $\mathbf{R}$ in a finite set $\mathbf{X} = \{x_1, ..., x_m\}$ $(m \geq 3)$; the relation can be expressed in the form:

$$\mathbf{R} = \mathbf{I} \cup \mathbf{P}, \tag{1}$$

where:
$\mathbf{I}$ - the equivalence relation (reflexive, transitive, symmetric),
$\mathbf{P}$ - the strict preference relation (transitive, asymmetric).

The preference relation $\mathbf{R}$ generates from elements of the set $\mathbf{X}$ the family (sequence) of subsets $\chi_1^*, \ldots, \chi_n^*$ $(n \leq m)$, such that each element $x_i \in \chi_r^*$ is preferred to any element $x_j \in \chi_s^*$ $(r < s)$ and each subset χ_q^* $(1 \leq q \leq n)$ comprises equivalent elements only.

The relation $\mathbf{R}$ can be characterised by the function T_1: $\mathbf{X} \times \mathbf{X} \rightarrow D_1$, $D_1 = \{-1, 0, 1\}$ or $T_2 : \mathbf{X} \times \mathbf{X} \rightarrow D_2, D_2 = \{-(n-1), ..., 0, ..., n-1\}$, defined as follows:

$$T_1(x_i, x_j) = \begin{cases} -1 \; if \; x_i \in \chi_r^* \; and \; x_j \in \chi_s^*, r < s, \\ 0 \; if \; x_i, x_j \in \chi_q^* \; (1 \le q \le n), \\ 1 \; if \; x_i \in \chi_r^* \; and \; x_j \in \chi_s^*, r > s; \end{cases} \tag{2}$$

$$T_2(x_i, x_j) = d_{ij} \;\Leftrightarrow\; x_i \in \chi_r^*, \quad x_j \in \chi_s^*, \quad d_{ij} = r - s. \tag{3}$$

The function $T_1(x_i, x_j)$ expresses the direction of preference in a pair (x_i, x_j), the function $T_2(x_i, x_j)$ – the difference of ranks of elements x_i and x_j.

The relation $\chi_1^*, \ldots, \chi_n^*$ is to be determined (estimated) on the basis of pairwise comparisons, of elements of the set $\mathbf{X}$, disturbed by random errors. Each pair $(x_i, x_j) \in \mathbf{X}$ is compared independently (in stochastic sense) N times; the result of *k-th* comparison (k=1, …, N; $N=2\omega+1$; ω=1, …,) is the value of the function $g_{1k} : \mathbf{X} \times \mathbf{X} \rightarrow D_{1g}$, $D_{1g}=\{-1, 0, 1\}$ (direction of preference case) or $g_{2k} : \mathbf{X} \times \mathbf{X} \rightarrow D_{2g}$, $D_{2g}=\{-(m-1), -(m-1)+1, \ldots, m-1\}$ (difference of ranks case), defined as follows:

$$g_{1k}(x_i, x_j) = \begin{cases} -1 \; if \; an \; assesment \; of \; direction \; of \; the \; prefernce \; in \; a \; pair \; (x_i, x_j) \\ \quad indicates \; x_i \in \chi_r^* \; and \; x_j \in \chi_s^*, r < s, \\ 0 \; if \; an \; assesment \; of \; direction \; of \; the \; prefernce \; in \; a \; pair \; (x_i, x_j) \\ \quad indicates \; x_i, x_j \in \chi_q^* \; (1 \le q \le n), \\ 1 \; if \; an \; assesment \; of \; direction \; of \; the \; prefernce \; in \; a \; pair \; (x_i, x_j) \\ \quad indicates \; x_i \in \chi_r^* \; and \; x_j \in \chi_s^*, r > s; \end{cases} \tag{4}$$

$$g_{2k}(x_i, x_j) = c_{ijk}, \tag{5}$$

where:

c_{ijk} is an assessment of the difference of ranks in a pair (x_i, x_j), in k-th comparison.

The set D_{2g} can include integers from the range: $-(m-1), \ldots, m-1$ because the number n of relation subsets is assumed unknown.

It is assumed that the comparisons $g_{fk}(x_i, x_j)$ (f=1, 2; k=1, …, N) are disturbed with random errors. The following assumptions are made about comparison errors.

The case of direction of the preference:

$$P(g_{1k}(x_i,x_j)=T_1(x_i,x_j))\geq 1-\delta,\quad (k=1, \ldots, N),\quad \delta\in(0,\, 1/2). \tag{6}$$

$$\begin{aligned} &P((g_{1k}(x_i,x_j)=T_1(x_i,x_j))\cap(g_{1k}(x_r,x_s)=T_1(x_r,x_s)))= \\ &P(g_{1k}(x_i,x_j)=T_1(x_i,x_j))P(g_{1l}(x_r,x_s)=T_1(x_r,x_s)) \qquad (k\neq l); \end{aligned} \tag{7}$$

the distributions of the errors $T_1(x_i, x_j) - g_{1k}(x_i, x_j)$ – the same for each k (k=1, …, N), the value of δ - known.

The case of difference of ranks:

$$\sum_{l\leq 0} P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l\ )>1/2, \tag{8}$$

$$\sum_{l\geq 0} P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l\ )>1/2, \tag{9}$$

$$P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l) \geq P(T_2\ (x_i, x_j) - g_{2k}(x_i, x_j) = l+1);\quad l\geq 0, \tag{10}$$

$$P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l) \geq P(T_2(x_i, x_j) - g_{2k}(x_i, x_j) = l-1);\quad l\leq 0, \tag{11}$$

$$P((g_{2k}(x_i, x_j)=c_{ijk})\cap(g_{2l}(x_r, x_s)=c_{rsl}))=P(g_{2k}(x_i, x_j)=c_{ijk})P(g_{2l}(x_r, x_s)=c_{rsl})\ (k\neq l); \tag{12}$$

the distributions of errors $T_2(x_i, x_j)-g_{2k}(x_i, x_j)$ - the same for each k (k=1, …, N).

The relationships (6) - (7) mean that: • correct comparison is more probable than incorrect, • comparisons $g_{1k}(x_i, x_j)$ and $g_{1l}(x_i, x_j)$ are independent. The relationships (8) - (12) mean that: • each distribution of comparison error is unimodal with median and mode in zero. Expected value $E(T_2(\cdot)-g_{2k}(\cdot)$ can differ from zero.

The problem of estimation of the preference relation can be formulated in the following way. To determine the sequence of subsets $\chi_1^*, \ldots, \chi_n^*$ or equivalently the form of the function $T_1(x_i, x_j)$ or $T_2(x_i, x_j)$ on the basis of pairwise comparisons $g_{1k}(x_i, x_j)$ or $g_{2l}(x_i, x_j)$ ((x_i, x_j)$\in\mathbf{X}\times\mathbf{X}$; k=1, …, N).

3. Basic definitions and notations

The following definitions and notations are necessary for further considerations:

- $t_f(x_i, x_j)$ (f=1, 2) - the function which determines the preference in each pair $(x_i, x_j)\in\mathbf{X}\times\mathbf{X}$, for any preference relation $\chi_1, \ldots, \chi_r$ in the set $\mathbf{X}$, i.e.:

$$t_1(x_i,x_j)=\begin{cases}-1\ if\ x_i\in\chi_r\ and\ x_j\in\chi_s,\ r<s,\\ 0\ if\ x_i,x_j\in\chi_q\ \ (1\le q\le n),\\ 1\ if\ x_i\in\chi_r\ and\ x_j\in\chi_s,\ r>s;\end{cases} \tag{13}$$

$$t_2(x_i, x_j) = d_{ij} \Leftrightarrow x_i \in \chi_k,\ \ x_j \in \chi_l;\ \ d_{ij}=k-l, \tag{14}$$

- $I(\chi_1, \ldots, \chi_r)$, $P_1(\chi_1, \ldots, \chi_r)$, $P_2(\chi_1, \ldots, \chi_r)$ - the sets of pairs of indices $<i,j>$, characterizing preferences in any relation $\chi_1, \ldots, \chi_r$, i.e.:

$$I(\chi_1, \ldots, \chi_r)=\{<i,j>\mid t_f(x_i, x_j) = 0; j>i\}, \tag{15}$$

$$P_1(\chi_1, \ldots, \chi_r)=\{<i,j>\mid t_f(x_i, x_j) < 0; j>i\}, \tag{16}$$

$$P_2(\chi_1, \ldots, \chi_r)=\{<i,j>\mid t_f(x_i, x_j) > 0; j>i\}; \tag{17}$$

- $R_m=I(\chi_1, \ldots, \chi_r)\cup P_1(\chi_1, \ldots, \chi_r)\cup P_2(\chi_1, \ldots, \chi_r)=\{<i,j>\mid 1\le i,j\le m; j>i\}$. (18)

The basis for construction of the estimators of the preference relation examined in the paper are the random variables $U_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)$, $V_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)$, $Z_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)$, $W_1^{(k)}(\chi_1,\ldots,\chi_r)$ defined, as follows:

$$U_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)=\begin{cases}0\ if\ g_{1k}(x_1,x_2)=0\ for<i,j>\in I(\chi_1,\ldots,\chi_r),\\ 1\ if\ g_{1k}(x_1,x_2)\neq 0\ for<i,j>\in I(\chi_1,\ldots,\chi_r),\end{cases} \tag{19}$$

$$V_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)=\begin{cases}0\ if\ g_{1k}(x_1,x_2)=-1\ for<i,j>\in P_1(\chi_1,\ldots,\chi_r),\\ 1\ if\ g_{1k}(x_1,x_2)\ge 0\ for<i,j>\in P_1(\chi_1,\ldots,\chi_r),\end{cases} \tag{20}$$

$$Z_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)=\begin{cases}0\ if\ g_{1k}(x_1,x_2)=1\ for<i,j>\in P_2(\chi_1,\ldots,\chi_r),\\ 1\ if\ g_{1k}(x_1,x_2)\le 0\ for<i,j>\in P_2(\chi_1,\ldots,\chi_r),\end{cases} \tag{21}$$

$$W_1^{(k)}(\cdot)=\sum_{<i,j>\in I(\cdot)}U_{1ij}^{(k)}(\cdot)+\sum_{<i,j>\in P_1(\cdot)}V_{1ij}^{(k)}(\cdot)+\sum_{<i,j>\in P_2(\cdot)}Z_{1ij}^{(k)}(\cdot), \tag{22}$$

and the random variables $U_{2ij}^{(k)}(\chi_1, ..., \chi_r)$, $V_{2ij}^{(k)}(\chi_1, ..., \chi_r)$, $Z_{2ij}^{(k)}(\chi_1, ..., \chi_r)$, $W_2^{(k)}(\chi_1, ..., \chi_r)$ defined, as follows:

$$U_{2ij}^{(k)}(\chi_1, ..., \chi_r) = |g_{2k}(x_i, x_j)|; \qquad <i,j> \in I(\chi_1, ..., \chi_r), \tag{23}$$

$$V_{2ij}^{(k)}(\chi_1, ..., \chi_r) = |t_2(x_i, x_j) - g_{2k}(x_i, x_j)|; \qquad <i,j> \in P_1(\chi_1, ..., \chi_r), \tag{24}$$

$$Z_{2ij}^{(k)}(\chi_1, ..., \chi_r) = |t_2(x_i, x_j) - g_{2k}(x_i, x_j)|; \qquad <i,j> \in P_2(\chi_1, ..., \chi_r), \tag{25}$$

$$W_2^{(k)}(\cdot) = \sum_{<i,j>\in I(\cdot)} U_{2ij}^{(k)}(\cdot) + \sum_{<i,j>\in P_1(\cdot)} V_{2ij}^{(k)}(\cdot) + \sum_{<i,j>\in P_2(\cdot)} Z_{2ij}^{(k)}(\cdot). \tag{26}$$

The random variables and other symbols corresponding to the relation **R** (errorless result of the estimation problem) will be denoted: $U_{fij}^{(k)*}$, $V_{fij}^{(k)*}$, $Z_{fij}^{(k)*}$, I^*, P_1^*, P_2^*, $W_f^{(k)*}$, while symbols corresponding to any other relation $\tilde{\chi}_1, \ldots, \tilde{\chi}_{\tilde{n}}$, different than errorless one, will be denoted: $\tilde{U}_{fij}^{(k)}, \tilde{V}_{fij}^{(k)}, \tilde{Z}_{fij}^{(k)}, \tilde{I}$, $\tilde{P}_1, \tilde{P}_2, \tilde{W}_f^{(k)}$.

4. Estimator based on medians from comparisons indicating direction of preference

The basis for the estimation of the relation form are medians $U_{1ij,me}(\chi_1, ..., \chi_r)$, $V_{1ij,me}(\chi_1, ..., \chi_r)$, $Z_{1ij,me}(\chi_1, ..., \chi_r)$ from the comparisons (random variables) – respectively $U_{1ij}^{(k)}(\chi_1, ..., \chi_r)$, $V_{1ij}^{(k)}(\chi_1, ..., \chi_r)$, $Z_{1ij}^{(k)}(\chi_1, ..., \chi_r)$ (k=1, …, N; N=2ω+1), defined as follows:

$$U_{1ij,me}(\chi_1, ..., \chi_r) = \begin{cases} 0 \; if \; \sum_{k=1}^{N} U_{1ij}^{(k)}(\chi_1, ..., \chi_r) \le \omega, \\ 1 \; if \; \sum_{k=1}^{N} U_{1ij}^{(k)}(\chi_1, ..., \chi_r) > \omega; \end{cases} \tag{27}$$

$$V_{1ij,me}(\chi_1, ..., \chi_r) = \begin{cases} 0 \; if \; \sum_{k=1}^{N} V_{1ij}^{(k)}(\chi_1, ..., \chi_r) \le \omega, \\ 1 \; if \; \sum_{k=1}^{N} V_{1ij}^{(k)}(\chi_1, ..., \chi_r) > \omega; \end{cases} \tag{28}$$

$$Z_{1ij,me}(\chi_1,\ldots,\chi_r)=\begin{cases}0\ if\ \sum_{k=1}^{N} Z_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)\le\omega,\\ 1\ if\ \sum_{k=1}^{N} Z_{1ij}^{(k)}(\chi_1,\ldots,\chi_r)>\omega.\end{cases} \tag{29}$$

The random variables $U_{1ij}^{(k)}(\cdot)$, $V_{1ij}^{(k)}(\cdot)$, $Z_{1ij}^{(k)}(\cdot)$ can assume the values zero and one and, therefore, their medians (formulas (27) – (29)) are equivalent to majority in the sets including values zero and one.

The random variables W_{1me}^{*} and $\widetilde{W}_{1me}$ (defined with the use of variables (27) – (29)), corresponding to the relations – respectively - $\chi_1^*,\ldots,\chi_n^*$ and $\widetilde{\chi}_1,\ldots,\widetilde{\chi}_{\tilde{n}}$ satisfy the following

Theorem 1

If the assumptions (6) – (7) and on identity of distributions of comparisons $g_{1k}(x_i, x_j)$ (k=1, …, N) for any pair $(x_i, x_j)\in \mathbf{X}\times\mathbf{X}$ are satisfied, then

$$E(W_{1me}^{*}-\widetilde{W}_{1me})<0, \tag{30}$$

$$P(W_{1me}^{*}<\widetilde{W}_{1me})>1\text{-}2\lambda_N, \tag{31}$$

where: $\lambda_N=\exp\{-2N(\frac{1}{2}-\delta)^2\}$.

Proof - see Klukowski (1994), point 5.2.

The inequalities (30) – (31) indicate that: • expected value of the variable W_{1me}^{*} corresponding to the relation $\chi_1^*,\ldots,\chi_n^*$ is lower than the variable $\widetilde{W}_{1me}$ corresponding to any other relation $\widetilde{\chi}_1,\ldots,\widetilde{\chi}_{\tilde{n}}$, • the probability of the event $\{W_{1me}^{*}<\widetilde{W}_{1me}\}$ converges exponentially to one for $N\to\infty$. The results indicate the form of the estimator – to determine the relation $\hat{\chi}_1,\ldots,\hat{\chi}_{\hat{n}}$, which minimize the value of the variable $W_{1me}(\chi_1,\ldots,\chi_r)$ for given comparisons $g_{11}(x_i,x_j),\ldots,g_{1N}(x_i,x_j)$ ($(x_i, x_j)\in \mathbf{X}\times\mathbf{X}$). The optimisation task for determining the estimate of the relation assumes the form:

$$\min_{\chi_1^{(t)},\dots,\chi_r^{(t)}\in F_X} \{ \sum_{<i,j>\in I(\chi_1^{(t)},\dots,\chi_r^{(t)})} U_{1ij,me}(\chi_1^{(t)},\dots,\chi_r^{(t)}) + \sum_{<i,j>\in P_1(\chi_1^{(t)},\dots,\chi_r^{(t)})} V_{1ij,me}(\chi_1^{(t)},\dots,\chi_r^{(t)}) + \sum_{<i,j>\in P_2(\chi_1^{(t)},\dots,\chi_r^{(t)})} Z_{1ij,me}(\chi_1^{(t)},\dots,\chi_r^{(t)})\}. \tag{32}$$

where:

F_X – the feasible set, i.e. the family of preference relations in the set $\mathbf{X}$,

$\chi_1^{(t)},\dots,\chi_r^{(t)}$ - t-th element of the feasible set.

The number of solutions of the problem (32) may exceed one; the unique solution can be determined randomly or with the use of additional criterion, e.g. minimization on the set of indices $<i,j>\in P_1(\cdot)\cup P_2(\cdot)$.

5. Estimator based on medians from differences of ranks

The basis for estimation of the relation form are random variables $U_{2ij,me}(\chi_1,\dots,\chi_r)$, $V_{2ij,me}(\chi_1,\dots,\chi_r)$, $Z_{2ij,me}(\chi_1,\dots,\chi_r)$, defined similarly as $U_{2ij}^{(k)}(\cdot)$, $V_{2ij}^{(k)}(\cdot)$, $Z_{2ij}^{(k)}(\cdot)$ (see (23) – (26)), but with replacement of individual comparisons $g_{2k}(\cdot)$ with their medians $g_{2me}(\cdot)$. More precisely, each random variable $g_{2me}(x_i, x_j)$ is the median in the set $\{g_{21}(x_i, x_j), \dots, g_{2N}(x_i, x_j)$ $((x_i, x_j)\in \mathbf{X}\times\mathbf{X}$; $N=2\omega+1$; $\omega=1, \dots,)$. The distribution of the median results from the properties of order statistics (see David, 1970, point 2.4). The properties of the estimator proposed below are based on random variables W_{2me}^{*} and $\widetilde{W}_{2me}$, defined with the use of the medians $g_{2me}(\cdot)$ (see (27) - (29)). The properties are presented in the following

Theorem 2

If the assumptions (8) – (12) and on identity of distributions of comparisons $g_{2k}(x_i, x_j)$ $(k=1, \dots, N)$ for any pair $(x_i, x_j)\in \mathbf{X}\times\mathbf{X}$ are satisfied, then:

$$E(W_{2me}^{*} - \widetilde{W}_{2me}) < 0, \tag{33}$$

$$P(W_{2me}^{*} < \widetilde{W}_{2me}) \geq -\frac{E(\sum_{v=1}^{8}\sum_{<i,j>\in S_v} Q_{2ij,me}^{(v)})}{\lambda_1(m-1)+2\lambda_2(m-1)+\lambda_3(m-2)}, \tag{34}$$

where:

$Q^{(1)}_{2ij,me} = U^{*}_{2ij,me} - \widetilde{V}_{2ij,me}$, $S_1 = \{<i,j> \mid <i,j> \in I^{*} \cap (\widetilde{P}_1 - P^{*}_1)\}$,

$Q^{(2)}_{2ij,me} = U^{*}_{2ij,me} - \widetilde{Z}_{2ij,me}$, $S_2 = \{<i,j> \mid <i,j> \in I^{*} \cap (\widetilde{P}_2 - P^{*}_2)\}$,

$Q^{(3)}_{2ij,me} = V^{*}_{2ij,me} - \widetilde{U}_{2ij,me}$, $S_3 = \{<i,j> \mid <i,j> \in P^{*}_1 \cap (\widetilde{I} - I^{*})\}$,

$Q^{(4)}_{2ij,me} = V^{*}_{2ij,me} - \widetilde{Z}_{2ij,me}$, $S_4 = \{<i,j> \mid <i,j> \in P^{*}_1 \cap (\widetilde{P}_2 - P^{*}_2)\}$,

$Q^{(5)}_{2ij,me} = V^{*}_{2ij,me} - \widetilde{V}_{2ij,me}$, $S_5 = \{<i,j> \mid <i,j> \in (P^{*}_1 \cap \widetilde{P}_1) \cap (T_2(x_i, x_j) \neq \widetilde{T}_2(x_i, x_j)\}$,

$Q^{(6)}_{2ij,me} = Z^{*}_{2ij,me} - \widetilde{U}_{2ij,me}$, $S_6 = \{<i,j> \mid <i,j> \in P^{*}_2 \cap (\widetilde{I} - I^{*})\}$,

$Q^{(7)}_{2ij,me} = Z^{*}_{2ij,me} - \widetilde{V}_{2ij,me}$, $S_7 = \{<i,j> \mid <i,j> \in P^{*}_2 \cap (\widetilde{P}_1 - P^{*}_1)\}$,

$Q^{(8)}_{2ij,me} = Z^{*}_{2ij,me} - \widetilde{Z}_{2ij,me}$, $S_8 = \{<i,j> \mid <i,j> \in P^{*}_2 \cap \widetilde{P}_2) \cap (T_2(x_i, x_j) \neq \widetilde{T}_2(x_i, x_j))\}$.

$\lambda_1 = \#(S_1 \cup S_2 \cup S_3 \cup S_6)$; $\lambda_2 = \#(S_4 \cup S_7)$; $\lambda_3 = \#(S_5 \cup S_8)$;

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

For the proof of the inequalities (33), (34) – see Klukowski (2007), Appendix.

Interpretation of the inequalities (33) – (34) is similar to those from Section 4. The optimisation problem for the case under consideration assumes the form:

$$\min_{\chi^{(t)}_1, \ldots, \chi^{(t)}_{r^{(t)}} \in F_X} \{ \sum_{<i,j> \in R_m} \left| t^{(t)}_2(x_i, x_j) - g_{2me}(x_i, x_j) \right| \}, \tag{35}$$

where:

F_X – the feasible set – a family of all preference relations in the set **X**;

$\chi^{(t)}_1, \ldots, \chi^{(t)}_{r^{(t)}}$ - an element of the set F_X;

$t^{(t)}_2(x_i, x_j)$ – the function characterizing the relation $\chi^{(t)}_1, \ldots, \chi^{(t)}_{r^{(t)}}$.

The number of solutions of the problem (35) also can exceed one.

Let us notice that the right-hand side of the inequality (34) does not guarantee the exponential convergence of the probability $P(W^{*}_{2me} < \widetilde{W}_{2me})$ to one, for $N \to \infty$. However, such type of convergence can be achieved in the following way. On the basis of the formulas expressing the distribution of the sample median (see Klukowski, 2007), Appendix) it can be determined a minimal value (integer) κ, ($\kappa \leq N$), which guarantee, for each pair $(x_i, x_i) \in \mathbf{X} \times \mathbf{X}$, the condition:

$$P(T_2(x_i, x_j) - g_{2,me(\kappa)}^{(\kappa)}(x_i, x_j) = 0) > ½, \tag{36}$$

where: $g_{2,me(\kappa)}^{(\kappa)}(\cdot)$ - the median in the subset of κ consecutive comparisons, e.g. $\{g_{21}(\cdot), \ldots, g_{2,\kappa}(\cdot)\}$.

Let us define the random variables $U_{2ij,\tau}(\chi_1, \ldots, \chi_r)$, $V_{2ij,\tau}(\chi_1, \ldots, \chi_r)$, $Z_{2ij,\tau}(\chi_1, \ldots, \chi_r)$ in the following way:

$$U_{2ij,\tau}(\chi_1, \ldots, \chi_r) = \begin{cases} 0; & g_{2me}^{(\tau)}(x_i, x_j) = t_2(x_i, x_j) \text{ for } <i,j> \in I(\chi_1, \ldots, \chi_r); \\ 1; & g_{2me}^{(\tau)}(x_i, x_j) \neq t_2(x_i, x_j) \text{ for } <i,j> \in I(\chi_1, \ldots, \chi_r), \end{cases} \tag{37}$$

$$V_{2ij,\tau}(\chi_1, \ldots, \chi_r) = \begin{cases} 0; & g_{2me}^{(\tau)}(x_i, x_j) = t_2(x_i, x_j) \text{ for } <i,j> \in P_1(\chi_1, \ldots, \chi_r); \\ 1; & g_{2me}^{(\tau)}(x_i, x_j) \neq t_2(x_i, x_j) \text{ for } <i,j> \in P_1(\chi_1, \ldots, \chi_r), \end{cases} \tag{38}$$

$$Z_{2ij,\tau}(\chi_1, \ldots, \chi_r) = \begin{cases} 0; & g_{2me}^{(\tau)}(x_i, x_j) = t_2(x_i, x_j) \text{ for } <i,j> \in P_2(\chi_1, \ldots, \chi_r); \\ 1; & g_{2me}^{(\tau)}(x_i, x_j) \neq t_2(x_i, x_j) \text{ for } <i,j> \in P_2(\chi_1, \ldots, \chi_r), \end{cases} \tag{39}$$

where:

$g_{2me}^{(\tau)}(x_i, x_j)$ - the median in the subset of comparisons $\{g_{2,(\tau-1)\cdot\kappa+1}(\cdot), \ldots, g_{2,\tau\cdot\kappa}(\cdot)\}$ ($\tau=1, \ldots, \vartheta$); the expression $a\cdot b$ - (index in $g_{2,a\cdot b}(\cdot)$) means the product of a and b,

ϑ - integer part of the quotient N/κ (odd number), i.e. $\vartheta = \text{ent}(N/\kappa)$.

Now, the approach, presented in Section 4, can be applied to the random variables $U_{2ij,\tau}(\chi_1, \ldots, \chi_r)$, $V_{2ij,\tau}(\chi_1, \ldots, \chi_r)$, $Z_{2ij,\tau}(\chi_1, \ldots, \chi_r)$ ($\tau=1, \ldots, \vartheta$). As a result, one can obtain the variables $U_{2ij,\vartheta}^{(me)}$, $V_{2ij,\vartheta}^{(me)}$, $Z_{2ij,\vartheta}^{(me)}$ defined as follows:

$$U_{2ij,\vartheta}^{(me)}(\chi_1, \ldots, \chi_r) = \begin{cases} 0; & \sum_{\tau=1}^{\vartheta} U_{2ij,\tau}(\chi_1, \ldots, \chi_r) < \vartheta/2 \quad \text{for } <i,j> \in I(\chi_1, \ldots, \chi_r); \\ 1; & \sum_{\tau=1}^{\vartheta} U_{2ij,\tau}(\chi_1, \ldots, \chi_r) > \vartheta/2 \quad \text{for } <i,j> \in I(\chi_1, \ldots, \chi_r), \end{cases} \tag{40}$$

$$V_{2ij,\vartheta}^{(me)}(\chi_1, ..., \chi_r) = \begin{cases} 0; & \sum_{\tau=1}^{\vartheta} V_{2ij,\tau}(\chi_1, ..., \chi_r) < \vartheta/2 \quad \text{for } <i,j> \in P_1(\chi_1, ..., \chi_r); \\ 1; & \sum_{\tau=1}^{\vartheta} V_{2ij,\tau}(\chi_1, ..., \chi_r) > \vartheta/2 \quad \text{for } <i,j> \in P_1(\chi_1, ..., \chi_r), \end{cases} \tag{41}$$

$$Z_{2ij,\vartheta}^{(me)}(\chi_1, ..., \chi_r) = \begin{cases} 0; & \sum_{\tau=1}^{\vartheta} Z_{2ij,\tau}(\chi_1, ..., \chi_r) < \vartheta/2 \quad \text{for } <i,j> \in P_2(\chi_1, ..., \chi_r); \\ 1; & \sum_{\tau=1}^{\vartheta} Z_{2ij,\tau}(\chi_1, ..., \chi_r) > \vartheta/2 \quad \text{for } <i,j> \in P_2(\chi_1, ..., \chi_r). \end{cases} \tag{42}$$

Let us apply the convention used in the previous Sections to the variables: $U_{2ij,\vartheta}^{(me)}(\cdot)$, $V_{2ij,\vartheta}^{(me)}(\cdot)$, $Z_{2ij,\vartheta}^{(me)}(\cdot)$, i.e. the symbols corresponding to the actual relation $\chi_1^*, \ldots, \chi_n^*$ will be marked with asterisks: $U_{2ij,\vartheta}^{(me)*}$, $V_{2ij,\vartheta}^{(me)*}$, $Z_{2ij,\vartheta}^{(me)*}$, while the symbols corresponding to any other relation $\widetilde{\chi}_1, \ldots, \widetilde{\chi}_r$ – with tildas: $\widetilde{U}_{2ij,\vartheta}^{(me)}$, $\widetilde{V}_{2ij,\vartheta}^{(me)}$, $\widetilde{Z}_{2ij,\vartheta}^{(me)}$.

Finally let us define the random variables $W_{2\vartheta}^{(me)*}$ and $\widetilde{W}_{2\vartheta}^{(me)}$:

$$W_{2\vartheta}^{(me)*} = \sum_{<i,j>\in I^*} U_{2ij,\vartheta}^{(me)*} + \sum_{<i,j>\in P_1^*} V_{2ij,\vartheta}^{(me)*} + \sum_{<i,j>\in P_2^*} Z_{2ij,\vartheta}^{(me)*}, \tag{43}$$

$$\widetilde{W}_{2\vartheta}^{(me)} = \sum_{<i,j>\in \widetilde{I}} \widetilde{U}_{2ij,\vartheta}^{(me)} + \sum_{<i,j>\in \widetilde{P}_1} \widetilde{V}_{2ij,\vartheta}^{(me)} + \sum_{<i,j>\in \widetilde{P}_2} \widetilde{Z}_{2ij,\vartheta}^{(me)}. \tag{44}$$

On the basis of the results presented in Theorem 1, Section 4, it is clear that:

$$P(W_{2\vartheta}^{(me)*} - \widetilde{W}_{2\vartheta}^{(me)} < 0) > 1 - 2\lambda_\vartheta, \tag{45}$$

where:

$$\lambda_\vartheta = \exp\{-2\vartheta(1/2 - \delta_{\max}^{(\kappa)})^2\} \tag{46}$$

$$\delta_{\max}^{(\kappa)} = \max_{(x_i,x_j)\in \mathbf{X}\times\mathbf{X}} \{P(T_2(x_i, x_j) \neq g_{2,me(\kappa)}^{(\kappa)}(x_i, x_j)\}.$$

If $\kappa>1$, then the convergence obtained as a result of the zero-one transformation is weaker, than in Section 4, because $\vartheta<N$ in the equality (46) (in other words the exponent in the right-hand side of relationship (45) "decreases with the step κ").

The case $\kappa=1$ is not excluded, in general, but it is satisfied only in the case $P(T_2(\cdot)\text{-}g_{2k}(\cdot)=0)>1/2$ for each pair $(x_i, x_j)\in\mathbf{X}\times\mathbf{X}$.

It seems viable to prove, that efficiency of the median approach in the case of difference of ranks is not worse than those based on the transformations (37) – (42), i.e. exponential "with the step κ"; the problem needs further investigations.

7. Summary

The paper presents two approaches to estimation of the preference relation on the basis of medians from multiple pairwise comparisons. The results obtained are based on weak assumptions about distributions of comparison errors; they can be verified with the use of statistical tests. The important property of the estimator based on "the direction of preference approach" is exponential convergence of the probability $P(W^*_{1me}<\widetilde{W}_{1me})$ to one, for $N\to\infty$. The convergence of the estimator based on differences of ranks is probably similar, but needs further investigations.

The approach presented above is applicable for complete relations with ties (represented by a complete graph). However, it can be easily extended for the case of partial order. For comparisons expressing direction of preference it can be achieved in simple way, i.e. assuming additional result of comparison – a pair of elements without connection. In analytic way such situation can be represented by a value of the function $T_1(x_i, x_j)=\Upsilon$ ($\Upsilon\notin\{-1, 0, 1\}$, $T_1(\cdot)\in D_{1\Upsilon}$, $D_{1\Upsilon}=\{-1, 0, 1, \Upsilon\}$, corresponding to non-connected elements (i.e. without arc or path), and a random variable:

$$Y_{1ij}^{(k)}=\begin{cases}0 \text{ if } g_{1k}(x_i,x_j)=\Upsilon \text{ and } T_1(x_i,x_j)=\Upsilon,\\ 1 \text{ if } g_{1k}(x_i,x_j)=\Upsilon \text{ and } T_1(x_i,x_j)\neq\Upsilon;\end{cases}$$

such that $P(T_1(x_i, x_j)=g_{1k}(x_i, x_j))\geq 1\text{-}\delta$. The variable $Y_{1ij}^{(k)}$ have to be incorporated into the formulas (22) and (32).

In the case of differences of ranks the formalization is more complex, especially assumptions about probability distributions of comparison errors have to be modified; the way of solving such problems will be the subject for further research.

References

David H.A. (1970) *Order Statistics*. John Wiley, New York.

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

Kamishima T., Akaho S. (2006) Supervised Ordering by Regression combined with Thurstone's Model. *Artificial Intelligence Review*, **25**, 231-246.

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

Klukowski L. (2007) *Estimation of the preference relation the basis of medians from pairwise comparisons in the form of difference of ranks*. In: *Computer Recognition Systems,* 2, M. Kurzynski et al. (Eds.), Springer-Verlag, Berlin Heidelberg, 232-241.

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

This book is a collection of papers, prepared in connection with the 8th International Workshop on partial orders, their theoretical and applied developments, which took place in Warsaw, at the Systems Research Institute, in October 2008. The papers deal with software developments (PYHASSE and other existing software), theoretical problems of ranking and ordering under various assumed analytic and decision-making-oriented conditions, as well as experimental studies and down-to-earth pragmatic questions.

ISBN 83-894-7521-9
EAN 9788389475213