Control and Cybernetics

vol. **36** (2007) No. 2

Estimation of tolerance relation on the basis of multiple pairwise comparisons with random errors

by

Leszek Klukowski

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

Abstract: The methods of tolerance relation estimation on the basis of pairwise comparisons with random errors, in the case of multiple comparisons for each pair, are proposed in the paper. Each comparison expresses the number of common features of both elements or the number of their missing features. The assumptions made about distributions of comparison errors are very weak, in particular they may be unknown. Two approaches are discussed: the first one, based on averaging of comparisons for each pair and the second, based on the median from comparisons. The estimated form of the relation is determined (in both cases) on the basis of the appropriate discrete programming task. The properties of estimators are based on some probabilistic inequalities. An example of application of the estimators proposed is presented.

Keywords: tolerance relation estimation, multiple pairwise comparisons, comparisons expressing the number of common features

1. Introduction

The tolerance relation is a relaxed form of the equivalence relation, i.e. without transitivity property. It divides a set of elements into a family of subsets with at least one non-empty intersection. The relation is a model of many real-life phenomena, e.g. analysis of marketing data (purchasing patterns of customers, when comparisons are applied to some number of independent purchases of each customer and the number of patterns is unknown); another example - analysis of empirical function shapes - is presented in Klukowski (2006).

The methods of tolerance relation estimation, presented in the paper, are extensions of the approach introduced in Klukowski (2002), section 4, for the case of $N > 1$ independent comparisons. The methods exploit the idea of *nearest adjoining order* introduced by Slater (1961) for the preference relation (see also

David, 1988). Two approaches are examined in the paper: the first one - based on averaged comparisons for each pair and the second - based on the median from comparisons. In both cases two forms of comparisons are examined. The first one determines the number of subsets of an intersection, which comprises both elements (in other words number of common features of both elements), the second one - the number of subsets, which do not comprise both elements (a number of missing features of both elements). The estimated form of the relation is obtained on the basis of optimal solution of some discrete optimization problems. They result from the fact that the expected values of some random variables (statistics based on pairwise comparisons, defined in Sections 3 and 4 of the paper) corresponding to actual relation, are lower than the expected values corresponding to any other relation. The properties of the estimators proposed are based on well known probabilistic inequalities: Hoeffding (see Hoeffding, 1963), Chebyshev (for expected value) and properties of the order statistics (see David, 1970). They express the probability of the event that the random variables, corresponding to actual relation, assume values lower than those corresponding to any other relation. In case of the first estimator - based on the averaging approach - the probability converges exponentially to one, for $N \to \infty$. A useful feature of the median approach is the simple form of the optimization task. Empirical experience and some asymptotic properties of the sample median indicate that efficiency of the median approach is also satisfactory. For both approaches it is possible to obtain some approximations of the probability in the case of unknown distributions of comparison errors.

The paper consists of six sections. The second section presents basic definitions, assumptions and notations. In the third section the averaging approach is examined. The fourth section presents the median approach and an algorithm for determination of the probability function of the median. In the fifth section an example of application of both approaches is discussed; the example is based on stochastic simulations. Last section sums up the results presented.

2. Basic definitions, assumptions and notations

It is assumed that there exists an (unknown) tolerance relation (reflexive, symmetric) in a finite set $\mathbf{X} = \{x_1, \ldots, x_m\}$ $(m \geqslant 3)$; the relation divides the set $\mathbf{X}$ into a family of subsets $\chi_1^*, \ldots, \chi_n^*$, $1 < n < m$, with the following properties:

$$\bigcup_{q=1}^{n} \chi_q^* = \mathbf{X}, \quad \chi_q^* \neq \{\emptyset\}, \quad \exists\, q, s\ (q \neq s) : \chi_q^* \cap \chi_s^* \neq \{\emptyset\}. \tag{1}$$

Moreover, in order to avoid the “degenerate” form of the relation it is assumed, additionally, that in each subset $\chi_q^* \subset \mathbf{X}$ there exists an element x_i, which belongs to the subset χ_q^* only, i.e.: $x_i \in \chi_q^*$ and $x_i \notin \chi_s^*$ for $s \neq q$.

The basis for further considerations is constituted by two functions, $T_1(\cdot)$ and $T_2(\cdot)$, defined as follows $T_1 : \mathbf{X} \times \mathbf{X} \to D$, $T_2 : \mathbf{X} \times \mathbf{X} \to D$, $D = \{0, 1, \ldots, n\}$, where:

$$T_1(x_i, x_j) = \#(\Omega_i^* \cap \Omega_j^*), \tag{2}$$

$$T_2(x_i, x_j) = \#(\Psi_i^* \cap \Psi_j^*), \tag{3}$$

where:
Ω_i^* – the set of the form $\Omega_i^* = \{s \,|\, x_i \in \chi_s^*\}$,
Ψ_i^* – the set of the form $\Psi_i^* = \{1, \ldots, n\} - \Omega_i^*$,
$\#(\Xi)$ – the number of elements of the set Ξ.

Under the assumption of the "non-degeneracy" of the relation, each of the functions $T_1(\cdot)$ and $T_2(\cdot)$, characterizes the form of the relation.

If an element x_i is included in some subset χ_q^*, this fact can be interpreted so that it possesses some feature; if it is included in a conjunction $\bigcap\limits_{q \in R} \chi_q^*$, then the element possesses some set of features. Thus, the function $T_1(\cdot)$ expresses the number of common features of elements x_i and x_j, while the function $T_2(\cdot)$ expresses the number of lacking features of both elements, from the set of all features existing in the set $\mathbf{X}$.

It is assumed that the basis for estimation of the relation is constituted by the results of comparisons $g_k^{(1)}(x_i, x_j)$ or/and $g_k^{(2)}(x_i, x_j)$ $(1 \leqslant k \leqslant N;$ $(x_i, x_i) \in \mathbf{X} \times \mathbf{X};\ j \neq i)$, corresponding to the form of the functions $T_1(\cdot)$ and $T_2(\cdot)$, respectively. The comparisons $g_k^{(f)}(x_i, x_j)$, observed instead of the (unknown) values $T_f(x_i, x_j)$, are disturbed by random errors; they can be obtained as a result of application of statistical tests, expert opinions or other decision functions.

The comparisons are defined in the following way:

$$g_k^{(1)}(x_i, x_j) = d_{ijk}^{(1)}, \quad d_{ijk}^{(1)} \in D, \tag{4}$$

$$g_k^{(2)}(x_i, x_j) = d_{ijk}^{(2)}, \quad d_{ijk}^{(2)} \in D, \tag{5}$$

where: $d_{ijk}^{(f)}$ $(f = 1, 2)$ is the assessment the value of $T_f(x_i, x_i)$, obtained in the k-th comparison.

The probabilities of random errors of each comparison are determined with the use of the probability function:

$$P(T_f(x_i, x_j) - g_k^{(f)}(x_i, x_j) = l)) = \alpha_{ijk}^{(f)}(l)$$
$$((x_i, x_i) \in \mathbf{X} \times \mathbf{X};\ f = 1, 2; -n \leqslant l \leqslant n). \tag{6}$$

It is assumed that comparisons $g_\kappa^{(f)}(x_i, x_j)$ and $g_\iota^{(f)}(x_q, x_s)$ $(\kappa \neq \iota)$ are independent, i.e.:

$$P((g_\kappa^{(f)}(x_i, x_j) = d_{ij\kappa}^{(f)}) \cap (g_\iota^{(f)}(x_q, x_s) = d_{qs\iota}^{(f)})) =$$
$$= P(g_\kappa^{(f)}(x_i, x_j) = d_{ij\kappa}^{(f)}) P(g_\iota^{(f)}(x_q, x_s) = d_{qs\iota}^{(f)}) \tag{7}$$

and the probabilities $\alpha_{ijk}^{(f)}(l)$ satisfy the conditions:

$$\sum_{l\leqslant 0}\alpha_{ijk}^{(f)}(l) > \frac{1}{2}, \quad \sum_{l\geqslant 0}\alpha_{ijk}^{(f)}(l) > \frac{1}{2}, \tag{8}$$

$$\left.\begin{array}{ll} \alpha_{ijk}^{(f)}(l) \geqslant \alpha_{ijk}^{(f)}(l+1), & l \geqslant 0, \\ \alpha_{ijk}^{(f)}(l) \geqslant \alpha_{ijk}^{(f)}(l-1), & l \leqslant 0. \end{array}\right\} \tag{9}$$

The conditions (8)-(9) guarantee that: zero is the median of each distribution (on the basis of median's definition), each probability function is unimodal and assumes maximum in zero. The expected value of the comparison error $E(T_f(\cdot) - g_k^{(f)}(\cdot))$ can differ from zero; it is typical for $T_f(\cdot) = 0$ or $T_f(\cdot) = n$.

Both types of comparisons, $g_k^{(1)}(x_i, x_j)$ and $g_k^{(2)}(x_i, x_j)$, can be used as a basis of estimation of the relation form - separately or simultaneously. In the second case it is assumed that comparisons $g_k^{(1)}(x_i, x_j)$ and $g_k^{(2)}(x_r, x_s)$ $((x_i, x_j), (x_r, x_s) \in \mathbf{X}\times\mathbf{X})$ are not correlated, i.e. $Cov(g_k^{(1)}(x_i, x_j), g_k^{(2)}(x_r, x_s)) = 0$. Correlation of comparisons $Vg_k^{(1)}(x_i, x_j), g_k^{(2)}(x_i, x_j)$ means that their content is similar.

It should be emphasized that comparisons of different pairs $g_k^{(f)}(x_i, x_j)$ and $g_k^{(f)}(x_r, x_s)$ $(< i, j > \neq < r, s >,\ k,\ f$ - fixed) are not assumed independent (in stochastic sense).

For simplification of further considerations it is assumed that the distributions of comparisons $g_k^{(f)}(x_i, x_j)$ are the same for each k $(1 \leqslant k \leqslant N)$; an extension for the case of different distributions for individual k is not difficult.

Let us define for any tolerance relation $\chi_1, \ldots, \chi_r$ in the set $\mathbf{X}$, the following sets of indices $I(\chi_1, \ldots, \chi_r)$ and $J(\chi_1, \ldots, \chi_r)$:

$$\begin{aligned} I(\chi_1, \ldots, \chi_r) = &\{< i, j > \big|\, \exists\, q, s\ (q = s \text{ not excluded })\\ &\text{such that } x_i, x_j \in \chi_q \cap \chi_s; j > i\}, \end{aligned} \tag{10}$$

$$\begin{aligned} J(\chi_1, \ldots, \chi_r) = &\{< i, j > \big|\, \text{there do not exist } q, s\\ &\text{such that } x_i, x_j \in \chi_q \cap \chi_s; j > i\}. \end{aligned} \tag{11}$$

The set $I(\chi_1, \ldots, \chi_r)$ includes such pairs of indexes $< i, j >$ that there exists an intersection $\chi_q \cap \chi_s$ of some subsets comprising both elements (x_i, x_j); when $q = s$, both elements belong to the same subset. The set $J(\chi_1, \ldots, \chi_r)$ includes such pairs (x_i, x_j), that both elements belong to different subsets χ_q, χ_s and both elements do not belong to the intersection $\chi_q \cap \chi_s$.

It is obvious, that:

$$I(\chi_1, \ldots, \chi_r) \cap J(\chi_1, \ldots, \chi_r) = \{\emptyset\}$$

and

$$I(\chi_1, \ldots, \chi_r) \cup J(\chi_1, \ldots, \chi_r) = \{< i, j > \big|\, 1 \leqslant i,\ j \leqslant m;\ j > i\,\}. \tag{12}$$

For any relation $\chi_1, \ldots, \chi_r$ in the set $\mathbf{X}$ the functions $t_1(x_i, x_j)$ and $t_2(x_i, x_j)$, characterizing this relation, are defined ($T_f(\cdot)$ relates to the "true" relation $\chi_1^*, \ldots, \chi_n^*$):

$$t_1(x_i, x_j) = \#(\Omega_i \cap \Omega_j), \tag{13}$$

$$t_2(x_i, x_j) = \#(\Psi_i \cap \Psi_j), \tag{14}$$

where:

$$\Omega_i \quad = \{s \,|\, x_i \in \chi_s\} \text{ and } \Psi_i = \{1, \ldots, r\} - \Omega_i. \tag{15}$$

The properties of the estimators proposed below are based on the properties of random variables $U_{fij}^{(k)}(\chi_1, \ldots, \chi_r)$ and $W_f^{(k)}(\chi_1, \ldots, \chi_r)$ defined as follows:

$$U_{fij}^{(k)}(\chi_1, \ldots, \chi_r) = |t_f(x_i, x_j) - g_k^{(f)}(x_i, x_j)|, \tag{16}$$

$$W_f^{(k)}(\chi_1, \ldots, \chi_r) = \sum_{<i,j> \in I(\chi_1, \ldots, \chi_r) \cup J(\chi_1, \ldots, \chi_r)} U_{fij}^{(k)}(\chi_1, \ldots, \chi_r). \tag{17}$$

For simplification of the notation, the symbols corresponding to the relation $\chi_1^*, \ldots, \chi_n^*$ will be denoted with asterisks (i.e. $U_{fij}^{(k)*}$, I^* , J^*, etc.) while corresponding to any other relation $\widetilde{\chi}_1, \ldots, \widetilde{\chi}_r$ - with tildas, e.g.:

$$U_{fij}^{(k)*} = |T_f(x_i, x_j) - g_k^{(f)}(x_i, x_j)|, \tag{18}$$

$$\widetilde{U}_{fij}^{(k)} = |\widetilde{t}_f(x_i, x_j) - g_k^{(f)}(x_i, x_j)|. \tag{19}$$

It follows from (6), (16) and the identify of distributions $g_k^{(f)}(x_i, x_j)(k = 1, ..., N)$ that the distribution function of each comparison error satisfies the conditions (index k is omitted in symbols $\alpha_{ijk}^{(f)}(l)$:

$$P(U_{fij}^{(k)*} = l) = \alpha_{ij}^{(f)}(-l) + \alpha_{ij}^{(f)}(l) \quad (l > 0). \tag{20}$$

3. The averaging approach

In the case of the averaging approach, the basis for the problem of estimation of the relation are the averages of the random variables $U_{fij}^{(k)}(\chi_1, \ldots \chi_r)$, $U_{fij}^{(k)*}$, $\widetilde{U}_{fij}^{(k)}$, $W_f^{(k)}(\chi_1, \ldots, \chi_r)$, $W_f^{(k)*}$ and $\widetilde{W}_f^{(k)}$, i.e.: the variables:

$$\overline{U}_{fij}(\chi_1, \ldots \chi_r) = \frac{1}{N} \sum_{k=1}^{N} |\, t_f(x_i, x_j) - g_k^{(f)}(x_i, x_j) \,|, \tag{21}$$

$$\overline{U}_{fij}^* = \frac{1}{N} \sum_{k=1}^{N} |\, T_f(x_i, x_j) - g_k^{(f)}(x_i, x_j) \,|, \tag{22}$$

$$\widetilde{U}_{fij}^{(k)} = \frac{1}{N} \sum_{k=1}^{N} |\, \widetilde{t}_f(x_i, x_j) - g_k^{(f)}(x_i, x_j) \,|, \tag{23}$$

$$\overline{W}^*_f = \sum_{<i,j>\in I^*\cup J^*} \overline{U}^*_{fij}, \tag{24}$$

$$\widetilde{\overline{W}}_f = \sum_{<i,j>\in \widetilde{I}\cup \widetilde{J}} \widetilde{\overline{U}}_{fij}. \tag{25}$$

The probabilistic properties of the difference: $\overline{W}^*_f - \widetilde{\overline{W}}_f$ – the basis for the properties of estimation the results – are determined on the basis of the Hoeffding inequality (see Hoeffding, 1963):

$$P(\sum_{k=1}^{N} Y_i - \sum_{k=1}^{N} E(Y_i) \geqslant Nt) \leqslant \exp\{-2Nt^2/(b-a)^2\}, \tag{26}$$

where:
Y_i $(i = 1,\ldots,N)$ – independent random variables satisfying the conditions: $P(a \leqslant Y_i \leqslant b) = 1$,
a, b, t – constants satisfying the conditions: $t > 0$, $b > a$.

They are determined in the following

THEOREM 1 *The random variables $\overline{W}^*_f$ and $\widetilde{\overline{W}}_f$, defined in (24) and (25) respectively, satisfy the conditions:*

$$E(\overline{W}^*_f - \widetilde{\overline{W}}_f) \ < 0, \tag{27}$$

$$P(\overline{W}^*_f - \widetilde{\overline{W}}_f < 0) \geqslant$$
$$\geqslant 1 - \exp\left\{ - \frac{N(\sum_{T_f(\cdot)\neq\widetilde{t}_f(\cdot)} E(\,|T_f(\cdot) - g_1^{(f)}(\cdot)| - |\,\widetilde{t}_f(\cdot) - g_1^{(f)}(\cdot)|\,)\,)^2}{2\vartheta^2(m-1)^2} \right\}, \tag{28}$$

where:
$T_f(\cdot) \neq \widetilde{t}_f(\cdot)$ – *denotes the set* $\{< i,j > \ \big|\, T_f(x_i,x_j) \neq \widetilde{t}_f(x_i,x_j); (x_i,x_j) \in \mathbf{X}\times\mathbf{X};\ j > i\}$,
ϑ *the number of elements of the set* $\{< i,j > \ \big|\, T_f(x_i,x_j) \neq \widetilde{t}_f(x_i,x_j);\ (x_i,x_j) \in \mathbf{X}\times\mathbf{X};\ j > i\}$.

Proof.
The proof of the inequality (27) for $f = 1$, under the assumption that the distributions of comparison errors (see (6)) are the same for each k $(k = 1,\ldots,N)$.

The difference: $U^{(k)*}_{1ij} - \widetilde{U}^{(k)}_{1ij}$ can be expressed in the following way:

$$U^{(k)*}_{1ij} - \widetilde{U}^{(k)}_{1ij} = |\,T_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,| - |\,\widetilde{t}_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,|. \tag{29}$$

The inequality $T_1(\cdot) \neq \widetilde{t}_1(\cdot)$ indicates that: $T_1(\cdot) > \widetilde{t}_1(\cdot)$ or $T_1(\cdot) < \widetilde{t}_1(\cdot)$. In the case when $T_1(\cdot) > \widetilde{t}_1(\cdot)$ each random variable $g_k^{(1)}(\cdot)$ can assume values, which satisfy the conditions:

(i) $g_k^{(1)}(\cdot) \geqslant T_1(\cdot)$;

(ii) $\widetilde{t}_1(\cdot) < g_k^{(1)}(\cdot) < T_1(\cdot)$;

(iii) $g_k^{(1)}(\cdot) \leqslant \widetilde{t}_1(\cdot)$.

For the values $g_k^{(1)}(\cdot) \geqslant T_1(\cdot)$ (case (i)) the difference $U_{1ij}^{(k)*} - \widetilde{U}_{1ij}^{(k)}$ equals: $-T_1(\cdot) + \widetilde{t}_1(\cdot)$; the last value is negative, its probability satisfies the inequality (see (8)): $\sum_{l \leqslant 0} P(T_1(\cdot) - g_k^{(1)}(\cdot) = l) > \frac{1}{2}$. In case (iii) the difference (29) is equal to: $T_1(\cdot) - \widetilde{t}_1(\cdot) > 0$ with probability (see (8) and (9)) $\sum_{l \geqslant T_1(\cdot) - \widetilde{t}_1(\cdot)} P(T_1(\cdot) - g_k^{(1)}(\cdot) = l) < \frac{1}{2}$. The inequality (ii) indicates $T_1(\cdot) - \widetilde{t}_1(\cdot) \geqslant 2$ and the difference (29) is equal to: $T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)}(\cdot)$. Moreover, the values $T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)}(\cdot)$ $(\widetilde{t}_1(\cdot) < g_k^{(1)}(\cdot) < T_1(\cdot))$ satisfy the condition:

$$-T_1(\cdot) + \widetilde{t}_1(\cdot) < T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)}(\cdot) < T_1(\cdot) - \widetilde{t}_1(\cdot) \tag{30}$$

and assume the values from the set $\{-T_1(\cdot)+\widetilde{t}_1(\cdot)+2, \ldots, T_1(\cdot)-\widetilde{t}_1(\cdot)-2\}$ with probabilities $P(T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)} = \iota) = P(g_k^{(1)} = (T_1(\cdot) + \widetilde{t}_1(\cdot) - \iota)/2)$. The expression $T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)}(\cdot)$ $(\widetilde{t}_1(\cdot) < g_k^{(1)}(\cdot) < T_1(\cdot))$ assumes values placed symmetrically around zero; their probabilities satisfy the conditions:

$$P(T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)} = -\iota) \geqslant P(T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)} = \iota) \quad (\iota > 0);$$

the last inequality results from the fact that in the case of $T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)}(\cdot) = -\iota$ the value of the difference $T_1(\cdot) - g_k^{(1)}(\cdot)$ is smaller (closer to zero), than in the case of $T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)}(\cdot) = \iota$. By assembling the facts concerning the case of $T_1(\cdot) > \widetilde{t}_1(\cdot)$, i.e.:

$$\sum_{l \leqslant 0} P(T_1(\cdot) - g_k^{(1)} = l) > \tfrac{1}{2} \tag{31}$$

$$\sum_{l \geqslant T_1(\cdot) - \widetilde{t}_1(\cdot)} P(T_1(\cdot) - g_k^{(1)} = l) < \tfrac{1}{2} \tag{32}$$

$$P(T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)} = -\iota) \geqslant P(T_1(\cdot) + \widetilde{t}_1(\cdot) - 2g_k^{(1)} = \iota) \quad (\iota > 0), \tag{33}$$

one can obtain:

$$E(U_{1ij}^{(k)*} - \widetilde{U}_{1ij}^{(k)} | T_1(\cdot) > \widetilde{t}_1(\cdot)) < 0. \tag{34}$$

The inequality:

$$E(U_{1ij}^{(k)*} - \widetilde{U}_{1ij}^{(k)} | T_1(\cdot) < \widetilde{t}_1(\cdot)) < 0 \tag{35}$$

corresponding to the case $T_1(\cdot) < \widetilde{t}_1(\cdot)$ is proved in a similar way.

The inequalities (34) and (35) imply – for each k ($k = 1, \ldots, N$) – the inequality:

$$E(U_{1ij}^{(k)*} - \widetilde{U}_{1ij}^{(k)}) < 0; \tag{36}$$

which is sufficient for (27).

Proof of the inequality (28).

The inequality (28) is proved on the basis of Hoeffding inequality (26). The difference: $\overline{W}_f^* - \widetilde{\overline{W}}_f$ can be expressed in the following way:

$$\overline{W}_1^* - \widetilde{\overline{W}}_1 = \frac{1}{N}\sum_{k=1}^{N} \sum_{T_1(x_i,x_j)\neq\widetilde{t}_1(x_i,x_j)} (\,|\,T_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,|- \\ -|\,\widetilde{t}_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,|\,). \tag{37}$$

The probability $P(\overline{W}_1^* - \widetilde{\overline{W}}_1 < 0)$ can be expressed in the form:

$$P(\overline{W}_1^* - \widetilde{\overline{W}}_1 < 0) = 1 - P(\overline{W}_1^* - \widetilde{\overline{W}}_1 \geqslant 0). \tag{38}$$

The probability $P(\overline{W}_1^* - \widetilde{\overline{W}}_1 \geqslant 0)$ can be evaluated in the following way. It follows from (29), that:

$$P(\overline{W}_1^* - \widetilde{\overline{W}}_1 \geqslant 0) = P(\frac{1}{N}\sum_{k=1}^{N} \sum_{T_1(x_i,x_j)\neq\widetilde{t}_1(x_i,x_j)} |\,T_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,|- \\ -|\,\widetilde{t}_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,| \geqslant 0). \tag{39}$$

Introducing the notations:

$$D_k^{(1)}(x_i,x_j) = |\,T_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,| - |\,\widetilde{t}_1(x_i,x_j) - g_k^{(1)}(x_i,x_j)\,| \tag{40}$$

one can express the probability (39) in the form:

$$P(\overline{W}_1^* - \widetilde{\overline{W}}_1 \geqslant 0) = P(\frac{1}{N}\sum_{k=1}^{N} \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_k^{(1)}(\cdot) \geqslant 0) = \\ = P(\frac{1}{N}\sum_{k=1}^{N} \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_k^{(1)}(\cdot) - \frac{1}{N}\sum_{k=1}^{N} \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_k^{(1)}(\cdot)) \geqslant \\ \geqslant -\frac{1}{N}\sum_{k=1}^{N} \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_k^{(1)}(\cdot))) =$$

$$= P(\sum_{k=1}^{N} \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_k^{(1)}(\cdot) - N \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_1^{(1)}(\cdot)) \geqslant$$
$$\geqslant N(- \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_1^{(1)}(\cdot)))). \tag{41}$$

The probability (41) can be evaluated on the basis of the inequality (26), in the following way:

$$P(\sum_{k=1}^{N} \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_k^{(1)}(\cdot) - N \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_1^{(1)}(\cdot)) \geqslant$$
$$\geqslant N(- \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_1^{(1)}(\cdot)))) \leqslant \exp\left\{ -\frac{2N(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_1^{(1)}(\cdot)))^2}{(2\vartheta(m-1))^2} \right\}. \tag{42}$$

The expression in the exponent results from the fact that: • each value $D_1^{(1)}(x_i, x_j)$ satisfies the condition $-(m-1) \leqslant D_1^{(1)}(x_i, x_j) \leqslant m-1$ (because $n < m$ and therefore the number of subsets generating any conjunction in the tolerance relation cannot exceed $m-1$), • the number of components of the sum equals ϑ, • all expected values $E(D_k^{(1)}(x_i, x_j))$ are the same for $1 \leqslant k \leqslant N$ and • their values are equal to $E(D_1^{(1)}(x_i, x_j))$. Moreover, the sum $\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_1^{(1)}(x_i, x_j))$ is negative and therefore the term:

$$- \sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_1^{(1)}(x_i, x_j))$$

is positive. The inequality (42) is equivalent to the proved inequality (28). The proof for $f = 2$ is similar. ■

The inequality (27) shows that the expected value of the random variable $\overline{W}_f^*$ is lower than the expected value of any other variable $\widetilde{\overline{W}}_f$. Moreover, the probability $P(\overline{W}_f^* < \widetilde{\overline{W}}_f)$ exceeds or is equal to the right hand side of the inequality (28). Thus, it is rational to estimate the relation $\chi_1^*, \ldots, \chi_n^*$ with the relation $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$, which minimizes the value of the random variable $\overline{W}_f(\chi_1, \ldots, \chi_r)$, for comparisons $g_k^{(1)}(x_i, x_j)$ $(k = 1, \ldots, N; (x_i, x_j) \in \mathbf{X} \times \mathbf{X})$. It is meaningful that the evaluation of the lower bound of the probability $P(\overline{W}_f^* < \widetilde{\overline{W}}_f)$ converges exponentially to zero, for $N \to \infty$. In the case of non-identical distributions of comparisons errors (for different k) the expected value $E(D_1^{(1)}(x_i, x_j))$ have to be replaced with $\min_k\{E(D_k^{(1)}(x_i, x_j))\}$. The probability $P(\overline{W}_1^* - \widetilde{\overline{W}}_1 < 0)$ can be also evaluated with the use of other probabilistic inequalities.

The estimated form $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$ of the relation $\chi_1^*, \ldots, \chi_n^*$ can be obtained on the basis of the solution of optimization tasks:

$$\min_{F_x} \left[\sum_{k=1}^{N} \sum_{\mathbf{X}\times\mathbf{X}} |\, t_1^{(\iota)}(x_i, x_j) - g_k^{(f)}(x_i, x_j)\,| \right], \quad (f = 1 \text{ or } 2) \tag{43}$$

or

$$\min_{F_x} \left[\sum_{k=1}^{N} \sum_{\mathbf{X}\times\mathbf{X}} (\,|t_1^{(\iota)}(x_i, x_j) - g_k^{(1)}(x_i, x_j)\,| + |\, t_2^{(\iota)}(x_i, x_j) - g_k^{(2)}(x_i, x_j)\,|\,) \right], \tag{44}$$

where:

F_x – the feasible set of the problem (the set including all tolerance relations satisfying the conditions (1) and the "non-degeneracy" condition);

$t_f^{(\iota)}(\cdot)$ – the function characterizing the relation $\chi_1^{(\iota)}, \ldots, \chi_{r^{(\iota)}}^{(\iota)}$ from the set F_X.

The feasible set of each of the problems (43) and (44) is finite and the optimal solution always exists; however, the number of solutions of each task may exceed one. In the case of multiple solutions we obtain a family of solutions $\hat{\chi}_1^{(1)}, ..., \hat{\chi}_{\hat{n}(1)}^{(1)}, ..., \hat{\chi}_1^{(\upsilon)}, ..., \hat{\chi}_{\hat{n}(\upsilon)}^{(\upsilon)}$ $(\upsilon \geq 2)$; the inequality $\{\overline{W}^* < \widetilde{\overline{W}}\}$ does not hold, but the event $\{\overline{W}^* \leq \widetilde{\overline{W}}\}$ is not excluded and $P(\overline{W} < \widetilde{\overline{W}}) \leq P(\overline{W}^* \leq \widetilde{\overline{W}})$. In such a case we have the alternative $(\hat{\chi}_1^{(1)}, ..., \hat{\chi}_{\hat{n}(1)}^{(1)} \equiv \chi_1^*, ..., \chi_n^*) \cup (\hat{\chi}_1^{(\upsilon)}, ..., \hat{\chi}_{\hat{n}(\upsilon)}^{(\upsilon)} \equiv \chi_1^*, ..., \chi_n^*)$ and the evaluation (28) relates to the alternative. The unique solution can be determined randomly or with the use of additional criterion, e.g. minimal value of the function (43) or (44) on the set $I(\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}})$.

The evaluation of the probability (28) can be determined in the case of known probability distributions of the comparison errors. In the opposite case, it is possible to determine some approximations of the evaluation. As the basis of the approximation one can use:

- the estimated form of the relation $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$ (allowing to determine the estimates $\hat{T}_f(\cdot)$ and $\hat{n}$), the formulas (31)–(33) together with the conditions (8)–(9), or
- the estimated form of the probability functions $\alpha_{ij}^{(f)}(l)$ obtained on the basis of comparisons $g_1^{(f)}(\cdot), \ldots, g_N^{(f)}(\cdot)$.

The first approach can be used for any value of N. The second approach requires – for purpose of realistic estimates – an appropriate number of comparisons N $(N >> n)$.

Let us notice that the right-hand side of the inequality (28) is based on the constraint $-(m-1) \leqslant D_1^{(1)}(x_i, x_j) \leqslant (m-1)$. Typically, the value $\pm(m-1)$ is excessive (significantly greater than n); especially in the case of $m - 1 >$

$\max_{\mathbf{X}\times\mathbf{X}}\{T_1(x_i,x_j)\}$ the constraint $\pm(m-1)$ negatively influences (decreases) the evaluation (28). Therefore, it is rational to replace the value $m-1$ with the estimate $\hat{n}$ or $\max_{\mathbf{X}\times\mathbf{X}}\{\hat{T}_1(x_i,x_j)\}$.

4. The median approach

In the case of median approach the basis for estimation is provided by the median from comparisons of each pair and it is assumed that $N = 2\tau + 1$ ($\tau = 1,\ldots,$). More precisely, each set of comparisons $g_1^{(f)}(x_i,x_j),\ldots,g_N^{(f)}(x_i,x_j)$ ($(x_i,x_j) \in \mathbf{X}\times\mathbf{X}$) is replaced with their median $g_{me,N}^{(f)}(x_i,x_j)$ and the variables $U_{fij}^{(k)}(\chi_1,\ldots,\chi_r)$, $U_{fij}^{(k)*}$, $\widetilde{U}_{fij}^{(k)}$, $W_f(\chi_1,\ldots,\chi_r)$, W_f^*, $\widetilde{W}_f$ ($f = 1,2$) are replaced – respectively – with the variables:

$$U_{fij}^{(me,N)}(\chi_1,\ldots,\chi_r) = |\, t_f(x_i,x_j) - g_{me,N}^{(f)}(x_i,x_j)\,|, \tag{45}$$

$$U_{fij}^{(me,N)*} = |\, T_f(x_i,x_j) - g_{me,N}^{(f)}(x_i,x_j)\,|, \tag{46}$$

$$\widetilde{U}_{fij}^{(me,N)} = |\, \widetilde{t}_f(x_i,x_j) - g_{me,N}^{(f)}(x_i,x_j)\,|, \tag{47}$$

$$W_f^{(me,N)*} = \sum_{<i,j>\in I^*\cup J^*} U_{fij}^{(me,N)*}, \tag{48}$$

$$\widetilde{W}_f^{(me,N)} = \sum_{<i,j>\in \widetilde{I}\cup\widetilde{J}} \widetilde{U}_{fij}^{(me,N)}, \tag{49}$$

where: $g_{me,N}^{(f)}(x_i,x_j)$ – the median from comparisons $g_1^{(f)}(x_i,x_j),\ldots,g_N^{(f)}(x_i,x_j)$, i.e. the $\frac{N+1}{2}$-th order statistics $g_{((N+1)/2)}^{(f)}(x_i,x_j)$ ($g_{(1)}^{(f)}(x_i,x_j),\ldots,g_{(N)}^{(f)}(x_i,x_j)$ – non-decreasingly ordered results of comparisons).

4.1. The form of the estimator and its properties

The problem considered in this point is similar to the single comparison case. However, the probability function of the median $g_{me,N}^{(f)}(x_i,x_j)$ of comparisons $g_1^{(f)}(x_i,x_j),\ldots,g_N^{(f)}(x_i,x_j)$ ($(x_i,x_j) \in \mathbf{X}\times\mathbf{X}$) is not the same, as the probability function of individual comparison $g_k^{(f)}(x_i,x_j)$ ($1 \leqslant k \leqslant N$); therefore the properties of the tolerance relation estimated on the basis of the medians are also not the same, as in the single comparison case. The properties of the estimator based on medians are presented in the following

Theorem 2 *The random variables* $W_f^{(me,N)*}$ *and* $\widetilde{W}_f^{(me,N)}$ *defined in (48) and*

(49) satisfy the conditions:

$$E(W_f^{(me,N)*} - \widetilde{W}_f^{(me,N)}) < 0, \tag{50}$$

$$P(W_f^{(me,N)*} < \widetilde{W}_f^{(me,N)}) \geqslant$$
$$\geqslant \frac{1}{\vartheta(m-1)} E\Big(\sum_{T_1(\cdot) \neq \widetilde{t}_1(\cdot)} |\, T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \,| -$$
$$- |\, \widetilde{t}_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \,| \Big), \tag{51}$$

where:

ϑ *– the number of elements of the set* $\{< i, j > | T_1(x_i, x_j) \neq \widetilde{t}_1(x_i, x_j);\ (x_i, x_j) \in \mathbf{X} \times \mathbf{X};\ j > i\}$.

Proof.

Proof of the inequality (50) for $f = 1$*, assuming the same distributions* $g_k^{(1)}(x_i, x_j)$ *for each* k *(*$k = 1, \ldots, N$*).*

The inequality (50) is true for $N = 1$ (it results from Theorem 1, for $N = 1$). For $N = 2\tau + 1$ ($\tau = 1, \ldots,$) it can be shown that the probability function $P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j)) = l)$ ($N = 2\tau + 1$; $\tau = 0, 1, \ldots,$) satisfies for each pair $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$ the inequalities:

$$P(T_1(x_i, x_j) - g_{me,N+2}^{(1)}(x_i, x_j) = 0) > P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = 0); \tag{52a}$$

$$P(T_1(x_i, x_j) - g_{me,N+2}^{(1)}(x_i, x_j) = l) < P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = l) \quad (l \neq 0). \tag{52b}$$

The inequalities (52a) and (52b) result from the following facts. The probabilities: $P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = l)$ can be expressed in the form (see David, 1970, Section 2.4):

$$P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = 0) =$$
$$= P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \leqslant 0) - P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \leqslant -1) =$$
$$= \frac{N!}{(((N-1)/2)!)^2} \int_{G(-1)}^{G(0)} t^{(N-1)/2} (1-t)^{(N-1)/2} dt, \tag{53a}$$

$$P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = l) =$$
$$= P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \leqslant l) - P(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \leqslant l-1) =$$
$$= \frac{N!}{(((N-1)/2)!)^2} \int_{G(l-1)}^{G(l)} t^{(N-1)/2} (1-t)^{(N-1)/2} dt, \tag{53b}$$

where:

$$G(l) = P(T_1(x_i, x_j) - g_k^{(1)}(x_i, x_j) \leqslant l).$$

The expressions (53a) and (53b) are determined on the basis of the beta distribution $B(p, q)$, with parameters $p = q = (N + 1)/2$. The expected value and variance of the distribution assume the form – respectively: $\frac{1}{2}$ and $((N + 1)/2)^2/((N + 1)^2(N + 2)) = (N + 2)/4$. The variance of the distribution converges to zero for $N \to \infty$ and the integrand in integrals (53a), (53b) is symmetric around $\frac{1}{2}$. These facts guarantee, that: the distributions of the random variables: $T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j)$ $((x_i, x_j) \in \mathbf{X} \times \mathbf{X})$ are for each N unimodal, their probability functions assume maximum in zero (i.e. for $T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = 0$) and satisfy the inequalities (52a), (52b). Last two conditions are sufficient (see assumptions (8), (9) and inequality (27) from Theorem 1) for the inequality (50).

Proof of the inequality (51).

Let us introduce the notations similar to those in Theorem 1:

$$D_{me}^{(1)}(x_i, x_j) = |\, T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j)\,| - |\, \widetilde{t}_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j)\,|. \quad (54)$$

Thus, the difference (54) can be expressed in the form:

$$P(W_1^{(me,N)*} < \widetilde{W}_1^{(me,N)}) = 1 - P(W_1^{(me,N)*} - \widetilde{W}_1^{(me,N)} \geqslant 0)$$

and the probability $P(W_1^{(me,N)*} - \widetilde{W}_1^{(me,N)} \geqslant 0)$ can be evaluated on the basis of Chebyshev inequality for expected value, in the following way:

$$P(W_1^{(me,N)*} - \widetilde{W}_1^{(me,N)} \geqslant 0) = P(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_{me}^{(1)}(\cdot) \geqslant 0) =$$

$$= P(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} (D_{me}^{(1)}(\cdot) + m - 1) \geqslant \vartheta(m-1)) \quad (55)$$

(ϑ – the number of components of the sum $\sum_{T_f(\cdot)\neq\widetilde{t}_f(\cdot)} D_{me}^{(1)}(\cdot)$) . The probability (55) can be evaluated with the use of Chebyshev inequality as follows:

$$P(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} (D_{me}^{(1)}(\cdot) + m - 1) \geqslant$$

$$\geqslant \vartheta(m-1)) \leqslant \frac{1}{\vartheta(m-1)} E(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} (D_{me}^{(1)}(\cdot) + (m-1))) =$$

$$= 1 + \frac{1}{\vartheta(m-1)} E(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_{me}^{(1)}(\cdot)). \quad (56)$$

The last expression in (56) is equal to the right-hand side of the inequality (51).

The proof for $f = 2$ is similar. ■

The expression $\frac{1}{\vartheta(m-1)} E(\sum_{T_1(\cdot) \neq \widetilde{t}_1(\cdot)} D^{(1)}_{me,N}(\cdot))$ (in the right-hand side of the equality (51)) is not positive, more precisely – it is included in the interval $(-1, 0)$. Its numerical value can be determined in the case of known distributions of comparison errors $P(T_1(\cdot) - g^{(1)}_{me,N}(\cdot))$. In the opposite case they can be approximated in some way. The approximation procedure based on the relationships (53a), (53b) (see David, 1970, section 2.4) and some additional assumptions (quasi-uniform distribution with equal values of the negative and positive tail) is proposed in Section 4.2 below.

Some asymptotic properties of the estimator based on the medians can be determined, too. They result from the properties of beta distribution for $N \to \infty$ (see relationships (53a), (53b)). They indicate that the median $g^{(1)}_{me,N}(\cdot)$ converges in stochastic sense to $T_1(\cdot)$, i.e. for any $\varepsilon > 0$ there is: $\lim_{N\to\infty} P(\mid g^{(1)}_{me,N}(\cdot) - T_1(\cdot) \mid > \varepsilon) = 0$ and the difference $E(W_1^{(me,N)*}) - E(\widetilde{W}_1^{(me,N)})$ converges to some negative value. The speed of convergence of the difference is the problem for future investigations.

The right-hand side of the inequality (51) is based on the fact that $-(m-1) \leqslant D^{(1)}_{me}(x_i, x_j) \leqslant m - 1$. Such constraint is typically (i.e. for $m - 1 > \max_{\mathbf{X}\times\mathbf{X}}\{T_1(x_i, x_j)\}$) excessive. Therefore, it is rational to replace the value $m-1$ (in the right-hand side of inequality (51)) with the estimate $\hat{n}$ or $\max_{\mathbf{X}\times\mathbf{X}}\{T_1(x_i, x_j)\}$.

The optimization problems for the median approach are similar to those formulated for the case of single comparison of each pair (see Klukowski, 2002), with difference that individual comparisons $g^{(f)}_k(x_i, x_j)$ $((x_i, x_j) \in \mathbf{X} \times \mathbf{X})$ are replaced with the medians $g^{(f)}_{me,N}(x_i, x_j)$ from N comparisons:

$$\min_{F_X} \left[\sum_{\mathbf{X}\times\mathbf{X}} \mid t^{(\iota)}_f(x_i, x_j) - g^{(f)}_{me,N}(x_i, x_j) \mid \right], \quad (f = 1 \text{ or } 2) \tag{57}$$

or

$$\min_{F_X} \left[\sum_{\mathbf{X}\times\mathbf{X}} (\mid t^{(\iota)}_1(x_i, x_j) - g^{(1)}_{me,N}(x_i, x_j) \mid + \mid t^{(\iota)}_2(x_i, x_j) - g^{(2)}_{me,N}(x_i, x_j) \mid) \right]; \tag{58}$$

F_X – the feasible set of the problem,

$t^{(\iota)}_f(\cdot)$ ($f = 1$ **or** 2) – the function characterizing the relation $\chi^{(\iota)}_1, \ldots, \chi^{(\iota)}_{r^{(\iota)}}$ from the set F_X.

The problems (57) and (58) are simpler to solve in comparison with the problems (43) and (44); the number of solutions may exceed one.

4.2. The procedure for approximation of the distribution function $P(T_1(x_i, x_j) - g^{(1)}_{me,N}(x_i, x_j) = l)$

The approximation procedure proposed in this point is especially useful for moderate N, namely $N = 5, 7, 9, 11$; for $N > 10$ the Gaussian approximation can also be used (see David, 1970, point 2.5).

The procedure is based on: some of "upper bound" distribution, the formulas (53a, b) and the estimated form of the relation. The "upper bound" distribution (a kind of evaluation) is obtained on the basis of: the conditions (8)-(9), some quasi-uniform (discrete) distribution and the assumption that the values of positive and negative tail of the distribution are equal, i.e. $P(T_f(\cdot) - g^{(f)}_k(\cdot) < 0) = P(T_f(\cdot) - g^{(f)}_k(\cdot) > 0)$ – with exception of the extreme values of $T_f(\cdot)$ (minimum and maximum). The estimated form of the relation, i.e. $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$, allows to determine the values $\hat{T}_f(x_i, x_j)$ $((x_i, x_j) \in \mathbf{X} \times \mathbf{X})$ and $\hat{n}$. The estimates can be also used for determination of the extreme value $\max\{\hat{T}_f(x_i, x_j) \,|\, (x_i, x_j) \in \mathbf{X} \times \mathbf{X}\}$ and the set of admissible values (range) of each comparison $g^{(f)}_k(x_i, x_j)$ $((x_i, x_j) \in \mathbf{X}\times\mathbf{X})$. It is suggested to determine the range of each comparison in the following way: to assume the minimum equal to zero and the maximum equal to $\hat{n}$. The minimum is natural – because no result of comparison can be negative. The maximum can be assumed in many ways, e.g.: $\max\{\hat{T}_f(x_i, x_j) \,|\, (x_i, x_j) \in \mathbf{X}\times\mathbf{X}\}$ or $\hat{n}$ or $m-1$. The "compromise" value is the estimate $\hat{n}$, because $\max\{\hat{T}_f(x_i, x_j) \,|\, (x_i, x_j) \in \mathbf{X}\times\mathbf{X}\} \leqslant \hat{n} \leqslant m-1$. The assumptions about equal values of positive and negative tail and quasi–uniform distribution of each tail allow to determine the distributions completely. The relationships (53a, b) allow to determine the distribution functions of medians of comparison errors for $N > 1$.

The quasi-uniform distribution is constructed for $f = 1$ in the following way. The estimates $\hat{T}_1(\cdot)$ and $\hat{n}$ are used instead of the actual values $T_1(\cdot)$ and n (i.e. they are assumed to be constant, not realizations of the random variables). The probabilities $P(T_f(\cdot) - g^{(f)}_k(\cdot) < 0)$ and $P(T_f(\cdot) - g^{(f)}_k(\cdot) > 0)$ are assumed equal (for $\hat{T}_f(\cdot) \neq 0$ and $\hat{T}_f(\cdot) \neq \hat{n}$); the probabilities $P(T_f(\cdot) - g^{(f)}_k(\cdot) = -l)$ are assumed equal for each (integer) $l > 0$ and the probabilities $P(T_f(\cdot) - g^{(f)}_k(\cdot) = l)$ are assumed equal for each (integer) $l > 0$ (quasi-uniform distribution). For the case: $\hat{T}_1(\cdot) \neq 0$, $\hat{T}_1(\cdot) \neq \hat{n}$, $\hat{n} > 2$, $\hat{n}$ – odd and $\hat{T}_1(\cdot) < \hat{n}/2$ the "upper bound" distribution function $P_b(\cdot)$ of comparison errors is obtained for each pair (x_i, x_i) from the system of equations:

$$P_b(\hat{T}_1(\cdot) - g^{(1)}_k(\cdot) = \hat{T}_1(\cdot) - \hat{n}) = \ldots = P_b(\hat{T}_1(\cdot) - g^{(1)}_k(\cdot) = -1), \tag{59}$$

$$P_b(\hat{T}_1(\cdot) - g^{(1)}_k(\cdot) = 1) = \ldots = P_b(\hat{T}_1(\cdot) - g^{(1)}_k(\cdot) = \hat{T}_1(\cdot)), \tag{60}$$

$$P_b(\hat{T}_1(\cdot) - g^{(1)}_k(\cdot) < 0) = P_b(\hat{T}_1(\cdot) - g^{(1)}_k(\cdot) > 0),) \tag{61}$$

$$P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0) = P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 1), \tag{62}$$

$$\sum_{l=\hat{T}_1(\cdot)-\hat{n}}^{\hat{T}_1(\cdot)} P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = l) = 1. \tag{63}$$

In the case of $\hat{T}_1(\cdot) > \hat{n}/2$, equation (62) is replaced with the equation:

$$P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0) = P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = -1) \tag{64}$$

(the probability $P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0)$ is equal to $\max\{P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 1), P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = -1)\}$).

In the case of $\hat{T}_1(\cdot) = 0$ the system assumes the simple form:

$$\left.\begin{aligned} &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0) = \tfrac{1}{2} + \varepsilon, \\ &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = -l) = (\tfrac{1}{2} - \varepsilon)/\hat{n} \quad (l = 1, \ldots, \hat{n}), \end{aligned}\right\} \tag{65a}$$

while in the case of $\hat{T}_1(\cdot) = \hat{n}$, the second relationship in (65a) is replaced with:

$$P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = l) = (\tfrac{1}{2} - \varepsilon)/\hat{n} \quad (l = 1, \ldots, \hat{n}), \tag{65b}$$

where: ε – constant from the interval $(0, \frac{1}{2})$ (e.g. $\varepsilon = \frac{1}{2(\hat{n}+1)}$).

In the case of even $\hat{n}$ ($\hat{n} > 2$) it is necessary to take into account the equality $\hat{T}_1(\cdot) = \hat{n}/2$. In this case the distribution of comparison errors is assumed in the form of:

$$\left.\begin{aligned} &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0) = \tfrac{1}{\hat{n}+1} + \varepsilon, \\ &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = \pm l) = \tfrac{1}{\hat{n}+1} - \tfrac{\epsilon}{\hat{n}} \quad (l = 1, \ldots, \hat{n}/2), \end{aligned}\right\} \tag{66}$$

where: ε – constant from the interval $(0, 1 - \frac{1}{\hat{n}+1})$.

In the case of $\hat{n} = 2$, the system assumes the simplest form:

$$\left.\begin{aligned} &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0) = \tfrac{1}{2} + \varepsilon_1; \\ &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 1) = P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 2) = (\tfrac{1}{2} - \varepsilon_1)/2; \\ &\quad \text{for } \hat{T}_1(\cdot) = 2, \\ &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0) = \tfrac{1}{3} + \varepsilon_2; \\ &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = \pm 1) = (\tfrac{2}{3} - \varepsilon_2)/2; \text{ for } \hat{T}_1(\cdot) = 1, \\ &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = 0) = \tfrac{1}{2} + \varepsilon_1; \\ &P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = -1) = P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = -2) = (\tfrac{1}{2} - \varepsilon_1)/2; \\ &\quad \text{for } \hat{T}_1(\cdot) = 0 \end{aligned}\right\} \tag{67}$$

where: ε_1 – constant from the interval $(0, \frac{1}{2})$, ε_2 – constant from the interval $(0, \frac{2}{3})$.

The probability functions of comparison errors generated by the above systems of equations can be considered as a kind of a "conservative approximation" of the actual distribution function, because any other distribution (based on the estimated relation form $\hat{\chi}_1, \ldots, \hat{\chi}_{\hat{n}}$ and distribution functions with symmetric values of tails) is more concentrated (its variance is smaller). If there exists some knowledge about asymmetry of tails (e.g. the value of asymmetry coefficient), then the equation systems (59)-(67) ought to be modified, especially the equality $P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) < 0) = P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) > 0)$ have to be replaced with the equality $P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) < 0) = \gamma P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) > 0)$, where γ – a constant, which guarantees satisfying the conditions (8)-(9).

The distribution function obtained on the basis of the equation systems (59)-(67) allows for the use of the relationships (53a, b) for determination the of "upper bound" approximation of the probability function $P_b^{(me,N)}(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = l)$ $(l = -\hat{n}, \ldots, \hat{n})$ of the median in the following way:

$$\begin{aligned}
&P_b^{(me,N)}(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) = l) = \\
&= P_b^{(me,N)}(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \leqslant l) - \\
&- P_b^{(me,N)}(T_1(x_i, x_j) - g_{me,N}^{(1)}(x_i, x_j) \leqslant l-1) = \\
&= \frac{N!}{(((N-1)/2)!)^2} \int\limits_{G_b(l-1)}^{G_b(l)} t^{(N-1)/2}(1-t)^{(N-1)/2} dt, \qquad (68)
\end{aligned}$$

where: $G_b(l) = P_b(\hat{T}_1(x_i, x_j) - g_k^{(1)}(x_i, x_j) \leqslant l)$, $G_b(l-1) = P_b(\hat{T}_1(x_i, x_j) - g_k^{(1)}(x_i, x_j) \leqslant l-1)$.

The approach presented above allows to determine some approximation of the right-hand side of the inequality (51).

In the case of $N >> \hat{n}$ the upper bound distribution functions can be replaced with estimated distribution functions; especially nonparametric estimators can be used for this purpose.

5. Example of application of the algorithms proposed

A simple (simulated) example of an application of the estimators proposed is considered below. The relation under examination assumes the form $\chi_1^* = \{x_1, x_2, x_3, x_4\}$, $\chi_2^* = \{x_3, x_4, x_5\}$, $\chi_3^* = \{x_4, x_6\}$, $\chi_4^* = \{x_7\}$. Each pair (x_i, x_j) is compared five times (comparisons of the same pair are independent); the results of comparisons (stochastic simulation) are presented in Table 1, while the distribution functions of the comparisons are presented in Table 2. The function $T_1(\cdot)$ assumes the following values:

$$T_1(x_1,x_5)=T_1(x_1,x_6)=T_1(x_1,x_7)=T_1(x_2,x_5)=T_1(x_2,x_6)=$$
$$=T_1(x_2,x_7)=T_1(x_5,x_6)=T_1(x_5,x_7)=T_1(x_6,x_7)=0;$$
$$T_1(x_1,x_2)=T_1(x_1,x_3)=T_1(x_1,x_4)=T_1(x_2,x_3)=T_1(x_2,x_4)=$$
$$=T_1(x_3,x_5)=T_1(x_4,x_5)=T_1(x_4,x_6)=1;\ T_1(x_3,x_4)=2.$$

Table 1. The results of comparisons (simulation)

Pair $< i,j >$	$g_{(1)}^{(1)}(\cdot)$	$g_{(2)}^{(1)}(\cdot)$	$g_{(3)}^{(1)}(\cdot)$	$g_{(4)}^{(1)}(\cdot)$	$g_{(5)}^{(1)}(\cdot)$	$g_{me}^{(1)}(\cdot)$	$\frac{1}{5}\sum_{k=1}^{5} g_k^{(1)}(\cdot)$
$< 1,2 >$	1	1	1	1	1	1	1
$< 1,3 >$	1	1	2	2	3	2	1.8
$< 1,4 >$	1	1	1	2	2	1	1.4
$< 1,5 >$	0	0	0	0	0	0	0
$< 1,6 >$	0	0	0	1	1	0	0.4
$< 1,7 >$	0	0	0	0	0	0	0
$< 2,3 >$	1	1	1	1	1	1	1
$< 2,4 >$	0	1	1	1	2	1	1
$< 2,5 >$	0	0	0	0	0	0	0
$< 2,6 >$	0	0	0	0	0	0	0
$< 2,7 >$	0	0	0	0	1	0	0.2
$< 3,4 >$	2	2	2	2	2	2	2
$< 3,5 >$	1	1	1	1	2	1	1.2
$< 3,6 >$	0	0	0	1	1	0	0.4
$< 3,7 >$	0	0	0	0	0	0	0
$< 4,5 >$	0	1	1	1	2	1	1
$< 4,6 >$	0	1	2	2	2	2	1.4
$< 4,7 >$	0	0	0	0	1	0	0.2
$< 5,6 >$	0	0	0	0	0	0	0
$< 5,7 >$	0	0	0	0	0	0	0
$< 6,7 >$	0	0	0	1	1	0	0.4

5.1. The algorithm based on the averaging approach

The estimated form of the relation $\chi_1^*,\ldots,\chi_4^*$ is obtained on the basis of the optimisation task (43), for $f=1$. It assumes the form $\hat{\chi}_1=\{x_1,x_2,x_3,x_4\}$, $\hat{\chi}_2=\{x_3,x_4,x_5\}$, $\hat{\chi}_3=\{x_4,x_6\}$, $\hat{\chi}_4=\{x_7\}$, i.e. is the same, as the relation $\chi_1^*,\ldots,\chi_4^*$; therefore $\hat{n}=n=4$. The minimal value of the function (43) equals 23, the solution is not multiple. The evaluation of the probability

Table 2. The probability distribution functions $P(g_k^{(1)}(x_i, x_j) = l)$ – the basis for simulations

Pair	$P(g_k^{(1)}(x_i, x_j) = l)$				
$< i, j >$	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$
$< 1, 2 >$	0.2	0.6	0.1	0.1	0.0
$< 1, 3 >$	0.2	0.5	0.2	0.1	0.0
$< 1, 4 >$	0.1	0.6	0.3	0.0	0.0
$< 1, 5 >$	0.7	0.2	0.1	0.0	0.0
$< 1, 6 >$	0.8	0.2	0.0	0.0	0.0
$< 1, 7 >$	0.9	0.1	0.0	0.0	0.0
$< 2, 3 >$	0.1	0.8	0.05	0.05	0.0
$< 2, 4 >$	0.2	0.75	0.05	0.0	0.0
$< 2, 5 >$	0.75	0.25	0.0	0.0	0.0
$< 2, 6 >$	0.65	0.35	0.0	0.0	0.0
$< 2, 7 >$	0.9	0.05	0.05	0.0	0.0
$< 3, 4 >$	0.0	0.1	0.7	0.1	0.1
$< 3, 5 >$	0.0	0.7	0.2	0.1	0.0
$< 3, 6 >$	0.8	0.2	0.0	0.0	0.0
$< 3, 7 >$	0.9	0.1	0.0	0.0	0.0
$< 4, 5 >$	0.3	0.6	0.1	0.0	0.0
$< 4, 6 >$	0.3	0.4	0.3	0.0	0.0
$< 4, 7 >$	0.9	0.1	0.0	0.0	0.0
$< 5, 6 >$	0.85	0.1	0.05	0.0	0.0
$< 5, 7 >$	0.95	0.05	0.0	0.0	0.0
$< 6, 7 >$	0.6	0.3	0.1	0.0	0.0

(28) is determined for the relation $\widetilde{\chi}_1 = \{x_1, x_2, x_3, x_4\}$, $\widetilde{\chi}_2 = \{x_1, x_3, x_4, x_5\}$, $\widetilde{\chi}_3 = \{x_1, x_4, x_6\}$, $\widetilde{\chi}_4 = \{x_7\}$ – similar to the relation $\chi_1^*, \ldots, \chi_4^*$. The difference between the relations $\chi_1^*, \ldots, \chi_4^*$ and $\widetilde{\chi}_1, \ldots, \widetilde{\chi}_4$ concerns the element x_1; in the relation $\chi_1^*, \ldots, \chi_4^*$ it belongs (exclusively) to the set χ_1^*, while in the relation $\widetilde{\chi}_1, \ldots, \widetilde{\chi}_4$ it belongs to the intersection $\bigcap_{r=1}^{3} \widetilde{\chi}_r$. The value of the function (43) corresponding to the relation $\widetilde{\chi}_1, \ldots, \widetilde{\chi}_4$ equals 41. The inequalities $T_1(\cdot) \neq \widetilde{t}_1(\cdot)$ appear for the pairs: (x_1, x_3), (x_1, x_4), (x_1, x_5), (x_1, x_6); the values $\widetilde{t}_1(\cdot)$ for these pairs are equal: $\widetilde{t}_1(x_1, x_3) = 2$, $\widetilde{t}_1(x_1, x_4) = 3$, $\widetilde{t}_1(x_1, x_5) = 1$, $\widetilde{t}_1(x_1, x_6) = 1$.

The evaluation (28) requires the probability functions of comparison errors and the values $T_1(x_i, x_j)$, $< i, j > \in \{T(x_i, x_j) \neq t_1(x_i, x_j)\}$. In the case of unknown distributions and $N = 5$ it is rational the use the approximation of probability functions $P_b(\hat{T}_1(\cdot) - g_k^{(1)}(\cdot) = l)$, described in section 4.2 (see (59)-

(67)). For the pair (x_1, x_3) the system of equations assumes the following form (the distribution functions $P_b(g_k^{(1)}(\cdot) = \iota)$ for all pairs satisfying the inequality $T_1(\cdot) \neq \widetilde{t}_1(\cdot)$ are presented in Table 3):

$$\left.\begin{array}{l} P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 1-4) = \\ \quad = P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 1-3) = \\ \quad = P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 1-2), \\ \sum_{l=1}^{3} P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = -l) = P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 1), \\ P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 0) = P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 1), \\ \sum_{l=-3}^{1} P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = l) = 1. \end{array}\right\} \quad (69a)$$

The solution of the above system assumes the form:

$$\left.\begin{array}{l} P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = -3) = \\ \qquad = P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = -2) = \\ \qquad = P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = -1) = \frac{1}{9}, \\ P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 0) = \\ \qquad = P_b(\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3) = 1) = \frac{1}{3}. \end{array}\right\} \quad (69b)$$

The expected value $E(D_{k,b}^{(1)}(x_1, x_3))$ corresponding to the above "upper bound" distribution assumes the form (after simple algebraic rearrangement):

$$\begin{array}{l} E(D_{k,b}^{(1)}(x_1,x_3)) = E_b(|\,\hat{T}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3)\,|\,) \\ -E_b(\,|\widetilde{t}_1(x_1,x_3) - g_k^{(1)}(x_1,x_3)\,|\,) = \\ \frac{1}{3}(|1-0| + |1-1|) + \frac{1}{9}(|1-2| + |1-3| + |1-4|) - \frac{1}{3}(|2-0| + |2-1|) - \\ -\frac{1}{9}(|2-2| + |2-3| + |2-4|) = -\frac{1}{3}, \end{array}$$

E_b - the expected value in the "upper bound" distribution.

The expected values of the remaining pairs are determined in a similar way; the distribution functions for the pairs (x_1, x_5), (x_1, x_6) are determined on the basis of the system (65a), with $\varepsilon = 1/10$ and their sum is equal:

$$\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_{k,b}^{(1)}(x_1,x_3)) = -1.5727.$$

Thus, the evaluation of the right-hand side of inequality (28) equals:

$$\exp\left\{ -\frac{2N(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} E(D_{k,b}^{(1)}(\cdot)))^2}{(2\vartheta(\hat{n}-1))^2} \right\} = \exp\{-0.0405\} = 0.9603$$

and the evaluation of the probability corresponding to $\hat{n}-1$ (denoted $P^{(av)}_{b,\hat{n}-1}(\overline{W}^*_1 - \widetilde{\widetilde{W}}_1 < 0)$) assumes the form:

$$P^{(av)}_{b,\hat{n}-1}(\overline{W}^*_1 - \widetilde{\widetilde{W}}_1 < 0) \geqslant 1 - 0.9603 = 0.0397 \tag{70}$$

(the value $\hat{n}-1$ is used instead of $\hat{n}$, because the subset $\widetilde{\chi}_4$ includes one element only).

The evaluations obtained on the basis of the actual probability functions (see Table 2) assumes the form:

$$P^{(av)}_{b,\hat{n}-1}(\overline{W}^*_f - \widetilde{\widetilde{W}}_f < 0) \geqslant 1 - \exp\{-0.5444\} = 0.4198. \tag{71}$$

The evaluation (70) assumes a low value (close to zero), but the relations $\chi^*_1, \ldots, \chi^*_4$ and $\widetilde{\chi}_1, \ldots, \widetilde{\chi}_4$ are similar and differences between the “upper bound” and actual distributions are not negligible (see Tables 2 and 3). The evaluation (71), based on actual probability functions, is much better than the “conservative” one.

Table 3. The “upper bound” distributions functions $P_b(g^{(1)}_k(\cdot) = \iota)$

Pair $< i, j >$	$P_b(g^{(1)}_k(\cdot) = \iota)$				
	$\iota = 0$	$\iota = 1$	$\iota = 2$	$\iota = 3$	$\iota = 4$
$< 1, 3 >, < 1, 4 >$	1/3	1/3	1/9	1/9	1/9
$< 1, 5 >, < 1, 6 >$	6/10	1/10	1/10	1/10	1/10

5.2. The algorithm based on the median approach

The medians of the comparisons $g^{(1)}_{(1)}(x_i, x_j), \ldots, g^{(1)}_{(5)}(x_i, x_j)$ $((x_i, x_j) \in \mathbf{X} \times \mathbf{X})$ are presented in Table 1. The optimal solution of the task (57), for $f = 1$, is the same as the relation $\chi^*_1, \ldots, \chi^*_n$ and those based on the averaging approach. The minimal value of the function (57) equals 2, the solution is not multiple. The approximation of the right-hand side of the inequality (51) is determined with the use of the algorithm described in Section 4.2 (see (59)-(67) and (68)) and for actual distributions. The first step of the median approach – determination of the probabilities $P_b(\hat{T}_1(\cdot) - g^{(1)}_k(\cdot) = l)$ – is described in Section 5.1. The second step - determination the values of the formula (68) – is performed as follows. The expression for the distribution of the median (for the “conservative distributions") of comparison errors assumes the form (see (53a, b)):

$$P_b^{(me,5)}(T_1(x_i,x_j) - g_{me,5}^{(1)}(x_i,x_j) = l) =$$
$$= \frac{5!}{(((5-1)/2)!)^2} \int_{G_b(l-1)}^{G_b(l)} t^{(5-1)/2}(1-t)^{(5-1)/2} dt =$$
$$= 30 \int_{G_b(l-1)}^{G_b(l)} t^2(1-t)^2 dt = 30t^3\ (\tfrac{1}{3} - t/2 + t^2/5)\Big|_{G_b(l-1)}^{G_b(l)}. \tag{72}$$

For the pair (x_1, x_3) the distribution of the median of comparison errors obtained on the basis of: the probability function resulting from the relationships (69a), the estimate $\hat{T}_1(x_1, x_3)$, and the expression (72), assumes the following form:

$$P_b^{(me,5)}(\hat{T}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3) = -3) = 0.0112,$$
$$P_b^{(me,5)}(\hat{T}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3) = -2) = 0.0632,$$

$$P_b^{(me,5)}(\hat{T}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3) = -1) = 0.1305,$$
$$P_b^{(me,5)}(\hat{T}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3) = 0) = 0.5901,$$
$$P_b^{(me,5)}(\hat{T}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3) = 1) = 0.2050.$$

Thus, the expected value

$$E_b(\,|\hat{T}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3)\,| - |\,\widetilde{t}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3)\,|\,)$$

assumes the form:

$$E_b(\,|\,\hat{T}_1(x_1,x_3) g_{me,5}^{(1)}(x_1,x_3)\,| - |\,\widetilde{t}_1(x_1,x_3) - g_{me,5}^{(1)}(x_1,x_3)\,|\,) =$$
$$= 0.0112(|1-4| - |2-4|) + 0.0632(|1-3| - |2-3|)+$$
$$+0.1305(|1-2| - |2-2|) + 0.5901(|1-1| - |2-1|)+$$
$$+0.2050(|1-0| - |2-0|) = -0.5901.$$

The remaining components of the sum $E(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_{me,b}^{(1)}(\cdot))$ are determined in a similar way and:

$$E(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D_{me,b}^{(1)}(\cdot)) = -1.9104.$$

The evaluation of the probability, corresponding to the value $\hat{n}-1$ (denoted $P^{(me)}_{b,\hat{n}-1}(W_1^{(me,5)*} - \widetilde{W}_1^{(me,5)} < 0)$) assumes the form:

$$P^{(me,5)}_{b,\hat{n}-1}(W_1^{(me,5)*} < \widetilde{W}_1^{(me,5)}) \geqslant -\frac{1}{\nu(\hat{n}-1)} E_b(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D^{(1)}_{me,b}(\cdot)) =$$

$$-(1/(4*3))*(-1.9104) = 0.1592. \tag{73}$$

The evaluation based on actual distributions assumes the form:

$$P^{(me,5)}_{\hat{n}-1}(W_1^{(me,5)*} < \widetilde{W}_1^{(me,5)}) \geqslant -\frac{1}{\nu(\hat{n}-1)} E(\sum_{T_1(\cdot)\neq\widetilde{t}_1(\cdot)} D^{(1)}_{me}(\cdot)) =$$

$$= -(1/4*3)*(-2.9168) = 0.2421. \tag{74}$$

Both evaluations, (73) and (74), are rather poor, but they are based on rough probabilistic inequality (51). However, in the example under consideration, both approaches (averaging and median) indicate the same estimation result and therefore the evaluation of the probability (71) obtained for the averaging approach is valid also in the median case.

6. Summary and conclusions

The methods of the tolerance relation estimation presented in the paper are often essential for practice, but seldom discussed in the literature of the subject. The idea of the methods proposed is the same as in the earlier author's papers in this area (Klukowski, 1990, 1994, 2000, 2002, 2006). The results obtained are especially meaningful in the case of averaging approach, when $N \to \infty$; they indicate that the probability (28) converges exponentially to one. The estimator based on the median of comparisons also possesses some asymptotic stochastic properties and is simpler from the computational point of view. The range of statistical properties of both estimators can be extended.

The important features of the estimators proposed are weak assumptions about stochastic properties of the comparisons. Especially, the distributions function of comparison errors may be unknown, the comparisons of different pairs may be not independent in stochastic sense, and the specification of the number of the subsets in the relation is not required. Such features of comparisons are typical, when they are obtained with the use of statistical tests or other decision functions, which may involve generated random errors.

The estimated form of the relation is obtained on the basis of the optimal solution of appropriate discrete programming tasks. Therefore, the number of solutions may exceed one; each of them can be regarded as the estimated form of the relation. It is not a negative feature of the methods proposed; the unique estimate can be selected randomly or with the use of additional criteria.

Empirical experience confirms usefulness of both estimators proposed. However, some properties of the estimators are difficult to determine in analytic way, e.g. the probability that the estimate $\hat{\chi}_1 \ldots, \hat{\chi}_{\hat{n}}$ is equivalent to $\chi_1^*, \ldots, \chi_n^*$ for all relations $\widetilde{\chi}_1, \ldots, \widetilde{\chi}_r$ from the feasible set F_X. They can be examined with the use of simulation approach.

References

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

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

Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

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

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

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

Klukowski, L. (2002) Estimation of the tolerance relation on the basis of pairwise comparisons with random errors (in Polish). In.: Z. Bubnicki, O. Hryniewicz, R. Kulikowski eds. *Methods and Techniques of Data Analysis and Decision Support.* Academic Publishing 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.

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