Measurement 46 (2013) 2927–2935

Contents lists available at SciVerse ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Ordinal measurement, preference aggregation and interlaboratory comparisons

Sergey V. Muravyov *

National Research Tomsk Polytechnic University, Department of Computer-aided Measurement Systems and Metrology, Pr. Lenina, 30, 634050 Tomsk, Russia

ARTICLE INFO

Article history:
Available online 9 May 2013

Keywords:
Rankings
Consensus relation
Preference aggregation
Kemeny ranking problem
Multiple optimal solutions
Interlaboratory comparisons

ABSTRACT

The classical problem of a single consensus ranking determination for m rankings of n alternatives has a potential of wide applications in information technologies, and particularly in measurement and instrumentation. The Kemeny rule is one of deeply justified ways to solve the problem allowing to find such a linear order (Kemeny ranking) of alternatives that a distance (defined in terms of a number of pair-wise disagreements between rankings) from it to the initial rankings is minimal. But the approach can result in considerably more than one optimal solutions what can reduce its applicability. By computational experiments outcomes, the paper demonstrates that a set of Kemeny rankings cardinality can be extremely large in small size cases ($m = 4, n = 15 \ldots 20$) and, consequently, special efforts to build an appropriate convoluting solution are needed. Application of the model to one of practical metrological problems, such as interlaboratory comparisons, is proposed and examined.

1. Introduction

In a series of earlier papers [1–6] by the author it was shown that a consideration of an ordinal scale measurement should involve notations of preference (particularly, in form of ranking or weak order) and consensus binary relations. In doing so, a measurement result on the ordinal scale should be the entire ranking of n objects and the ranking is one of elements of the weak order space. From the Representational Measurement Theory point of view, the preference aggregation problem could be seen as a particular case of the general conjoint measurement problem, see, for example [7,8]. This way of thinking is in accordance to the definition by Finkelstein [9–11]: "Measurement is, in the wide sense, an objective, empirical process of establishing a correspondence between properties of objects and events of the real world and a set of symbols and relations. The correspondence is such, that when a symbol is assigned to a manifestation of the property and another symbol is assigned to another manifestation of the same property, then the relation between the two symbols corresponds to a relation between the two manifestations of the property".

A single consensus ranking determination for m rankings (voters), possibly including ties, of n alternatives (candidates) is a classical problem that has been intensively investigated firstly as a Voting Problem in the framework of Social Choice Theory since the late XVIII century.

Condorcet in 1785, see [12], proposed a very natural rule for the consensus ranking determination: if some alternative obtains a majority of votes in pair-wise contests against every other alternative, the alternative is chosen as the winner in the consensus ranking. The Condorcet approach is widely recognized as the best rule for the consensus ranking determination, however, the binary relation defined by the *Condorcet rule* is not necessarily transitive, i.e. it can be for some consensus ranking β that $a_i \succ a_j$ and $a_j \succ a_k$ while $a_k \succ a_i$; $a_i, a_j, a_k \in \beta$. This *Condorcet paradox* may occur rather frequently, for example its chances are higher than 50% at $3 \leqslant m \leqslant \infty$ and $2 \leqslant n \leqslant 10$, if m is even; presence of ties reduces the probability, see, e.g. [13].

* Tel.: +7 3822 417527/420449; fax: +7 3822 420449.
E-mail address: muravyov@camsam.tpu.ru

0263-2241/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.measurement.2013.04.044

The *Kemeny rule* [14] is considered to be a reasonable way to get over the difficulty as it allows to find such a linear order (Kemeny ranking) β of alternatives that a distance (defined in terms of a number of pair-wise disagreements between rankings) from β to the initial rankings is minimal. But, in turn, the approach has two drawbacks:

- The Kemeny Ranking Problem (KRP) had been proven to be *NP*-hard [12,15–17].
- It may have considerably more than one optimal solutions.

The former is not so disturbing since, for reasonable problem sizes (up to $n < 30 \ldots 50$), there are exact algorithms for them to be effectively applied, see, for example [6,12,16–19]. Strangely enough, the latter blemish has been given short shrift by researchers despite its importance for the problem applicability. In fact, multiple optimal solutions may rank the alternatives in significantly different ways what can absolutely destroy a positive effect of a potential problem application.

Currently, the model has numerous interpretations and (or a potential for) applications in different domains, such as information retrieval, collaborative filtering [15], multiagent choice and multisensor fusion [15,20], hemometrics [21], digital image processing and pattern recognition [22,23], quality assessment and management [24], sport competitions judging [16,17], multiple criteria (or group) decision making [25], etc. However, the model being singularly fertile of deep measurement theoretical ideas does not have applications in real metrological practice.

The aim of this paper is first to demonstrate that a set of Kemeny rankings cardinality can be extremely large even in cases where $m = 4$ and $n = 15 \ldots 20$, and, consequently, special efforts to build some appropriate *convoluting solution* are needed. Second, it will be shown how the KRP-based model could be potentially applied to the interlaboratory comparisons problem. Interlaboratory comparisons need a reference value of the measurand to be assigned. It is necessary to have some procedure that allows to determine the reference value at a maximum number of participating laboratories results to be included into the determination and, at the same time, unreliable laboratory results must be disregarded.

The paper is organized as follows. In Section 2, after the KRP statement, an exact algorithm to find all Kemeny rankings for the given preference profile is briefly described. Some intriguing outcomes of computational experiments supporting the declared paper objective are reported and discussed. It is shown in Section 3 that a procedure of interlaboratory comparison can be implemented using the preference aggregation approach. Section 3 also provides a probabilistic way to justify the value m of the comparison participating laboratories.

2. Kemeny ranking problem formulation and solution

This section results were first published in the conference paper [26]. In this section we will use the following symbols:

A	$\{a_1, a_2, \ldots, a_n\}$: a set of n alternatives
Λ	$\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$: a set of m rankings (preference profile)
R	$[r_{ij}]$: an $(n \times n)$ ranking matrix
P	$[p_{ij}]$: an $(n \times n)$ profile matrix
$d(\lambda_k, \lambda_l)$	a distance between two rankings λ_k and λ_l
$D(\lambda, \Lambda)$	a distance between arbitrary ranking λ and profile Λ (Kemeny distance)
Π	a set of all $n!$ linear (strict) order relations $\succ$ on A
β	Kemeny ranking (consensus relation), $\beta \in \Pi$
B	$\{\beta_1, \beta_2, \ldots, \beta_{N_{\text{kem}}}\}$: a set of Kemeny rankings, $B \subset \Pi$
N_{kem}	number of Kemeny rankings for the given profile Λ
D_{least}	a least distance from Λ to some linear order
$\mathbf{N_n}$	$\{1, 2, \ldots, n\}$: first n natural numbers
S	$\{s_1, s_2, \ldots, s_K\}$: a partial solution (leader) of the KRP
K	$0, \ldots, n-1$: a level of a search tree
N_{nds}	a total number of the search tree nodes generated
T	$\{t_1, t_2, \ldots, t_{K-n}\} = \mathbf{N_n} \setminus S$: a complement of S
D_{low}	an estimate of Kemeny distance for the ranking with leader S (lower bound)
D_{u}	a minimal value of Kemeny distance for generated to the moment complete solutions (upper bound)

2.1. Problem statement

Suppose we have m *rankings* provided by m experts (voters, focus groups, criteria, etc.) on set A of n *alternatives* (candidates). Then the preference profile Λ consists of m rankings (*weak orders*) $\lambda = \{a_1 \succ a_2 \succ \ldots \sim a_s \sim a_t \succ \ldots \sim a_n\}$, each may include a *strict preference* relation $\succ$ and an *indifference* relation (or *tie*) $\sim$.

The ranking λ can be represented by the *ranking matrix* R, rows and columns of which are labeled by the alternatives' numbers and $r_{ij} = 1$ if $a_i \succ a_j$; $r_{ij} = 0$ if $a_i \sim a_j$; $r_{ij} = -1$ if $a_i \prec a_j$. Then the symmetric difference distance function [14] between two rankings λ_k and λ_l is defined by formula

$$d(\lambda_k, \lambda_l) = \sum_{i<j} |r_{ij}^k - r_{ij}^l|, \tag{1}$$

where only elements of the upper triangle submatrix, r_{ij}, $i < j$, are summed up.

Then a distance between arbitrary ranking λ and profile Λ can be defined as follows:

$$D(\lambda, \Lambda) = \sum_{k=1}^{m} d(\lambda, \lambda_k) = \sum_{i<j} \sum_{k=1}^{m} |r_{ij}^k - r_{ij}| = \sum_{i<j} \sum_{k=1}^{m} d_{ij}^k, \tag{2}$$

Supposing $r_{ij} = 1$ for all $i < j$ that corresponds to the natural linear order $a_1 \succ a_2 \succ \ldots \succ a_n$, it is clear that for any $k = 1, \ldots, m$ we have $d_{ij}^k = |1 - 1| = 0$ if $a_i^k \succ a_j^k$; $d_{ij}^k = |0 - 1| = 1$ if $a_i^k \sim a_j^k$ and $d_{ij}^k = |-1 - 1| = 2$ if $a_i^k \prec a_j^k$.

Then the *profile matrix* P can be defined where

$$p_{ij} = \sum_{k=1}^{m} d_{ij}^{k}, \quad i,j = 1,\ldots,n, \tag{3}$$

and the KRP is formulated as follows:

$$\beta = \arg\min_{\lambda\in\Pi} D(\lambda,\Lambda) = \arg\min_{\lambda\in\Pi} \sum_{i<j} p_{ij}. \tag{4}$$

Every permutation of alternatives of A corresponds to transposition of the profile matrix rows and columns. Hence, the problem (4) means the determination of such a transposition of profile matrix rows and columns that the sum of elements of its upper triangle submatrix is minimal.

2.2. KRP solution algorithm

The algorithm described in this paragraph is a version of that proposed in [6], which used the *recursive B&B technique* and was intended to find only a first one of all possible Kemeny rankings. As opposed to that, this version finds *all possible* Kemeny rankings for the given profile.

Characteristic parameter of the matrix P is a least *distance* D_{least} from its preference profile Λ to some linear order. The parameter is calculated by summing up lesser elements of each pair (p_{ij}, p_{ji}), that is

$$D_{\text{least}} = \sum_{i<j} \min(p_{ij}, p_{ji}). \tag{5}$$

It is clear that if all initial rankings are consistent and, hence, matrix P is transitive, i.e. $p_{ik} \leqslant p_{ki}$ if $p_{ij} \leqslant p_{ji}$ and $p_{jk} \leqslant p_{kj}$, $i \neq j \neq k = 1,\ldots,n$, then the condition

$$D(\beta,\Lambda) = D_{\text{least}}, \tag{6}$$

is valid and D_{least} becomes an accessible value. An inverse proposition is also satisfied. If matrix P is intransitive, then $D(\beta,\Lambda) > D_{\text{least}}$ and D_{least} is inaccessible.

The algorithm investigates a tree-structured solutions space. Each node of the solution tree is in one-to-one correspondence with a set S, which is considered to be a representative (or leader) of all solutions containing it as a leading part. The tree root is the leader of absolutely all feasible solutions and for it $S = \varnothing$. At the next (first) tree level there are n leaders of cardinality 1. Each of the leaders has $n-1$ successors of cardinality 2 at the second level. Generally, each of Kth level leaders has $n-K$ successors of cardinality $|S| = K+1$, $K = 0,\ldots,n-1$. If $|S| < n-1$ then an appropriate solution S is to be called current *partial solution*. Given $|S| = n-1$ the set S is to be a current *complete solution* as, in this case, it defines an order of all elements of A.

As a leader is build up of elements of $\mathbf{N_n}$, for any leader S there exists its complement $T = \overline{S} = \mathbf{N_n} \setminus S$. Each leader S is build up by means of concatenation of its predecessor and first in order element t_l of T, i.e. $S = \{s_1,\ldots,s_{K-1}, s_K = t_l\}$, and at the same time t_l is removed from T.

Each leader (partial solution) has appropriate estimate D_{low} of a distance from profile Λ to the optimal linear order β, which is called a *low bound*. A minimal value of a distance function for generated to the moment complete solutions is termed an *upper bound* D_{u}.

Leader S defines a position of the matrix P rows and columns with indexes $s_1,\ldots,s_{K-1}$. Corresponding sum of elements of upper triangle submatrix of P is

$$D = \sum_{\substack{i=1,\ldots,K-1 \\ j=i+1,\ldots,K-1}} p_{s_i s_j} + \sum_{i,j=1,\ldots,n-K} p_{t_i t_j}. \tag{7}$$

Taking into account expanding the leader due to concatenation with element t_l gives

$$D_{\text{e}} = D + \sum_{i=1,\ldots,n-K} p_{s_K t_i}. \tag{8}$$

Finally, for the rest of matrix defined by elements of T, we use the same principle as for determination of the least distance (5). Then we have

$$D_{\text{low}} = D_{\text{e}} + \sum_{\substack{i=1,\ldots,n-K \\ j=i+1,\ldots,n-K}} \min(p_{t_i t_j}, p_{t_j t_i}). \tag{9}$$

A leader is considered to be promising in case where the condition $D_{\text{low}} \leqslant D_{\text{u}}$ is satisfied. If $D_{\text{low}} > D_{\text{u}}$ (notice that due to (9) D_{low} is the least possible value for the given leader S) then it is clear that all solutions including this leader are hopeless, i.e. cannot be optimal.

The algorithm presented in Fig. 1 contains two stages:

- *initialization*, where all necessary variables acquire initial values and the parameter D_{least} is calculated, and
- *call of recursive procedure LEADER* (K,D) that contains the main cycle by l, where l is a leader number at Kth level. At each cycle step the current partial solution S is generated that defines a new position of the matrix P rows and columns, for which D_{e} and D_{low} are calculated.

If $D_{\text{low}} \leqslant D_{\text{u}}$ and $K < n-1$ then the procedure *LEADER* $(K+1, D_{\text{e}})$ is called in order to check the next level of the search tree; this way *branching* is realized. After the current solution becomes complete one, it is memorized as a pair $\beta_1 = S$ and $D_{\text{u}} = D_{\text{low}}$. If $D_{\text{low}} > D_{\text{u}}$ then the corresponding leader and all its successors are considered to be hopeless and they are *pruned*. The search is continued until all hopeless solutions will be pruned. After the first optimal solution β_1 is found, the search is continued until all N_{kem} possible solutions with the same D_{u} will be determined. The algorithm is an exact one as it checks all the feasible incomplete solutions.

2.3. Computational experiment

The algorithm described was implemented in C++ language in the Microsoft Visual Studio development environment. The initial profile matrices were calculated by rankings obtained by uniting pseudo-random strict orders and ties generated separately on the basis of the C++ library function `randomize()`. Strict orders represented by $\mathbf{N_n}$ permutations were generated on the basis of the *uniform distribution* of integers in a specified range $1,\ldots,n$. Thereby, in our experimentation, we stick to the so called *impartial culture* condition implying just the uniform

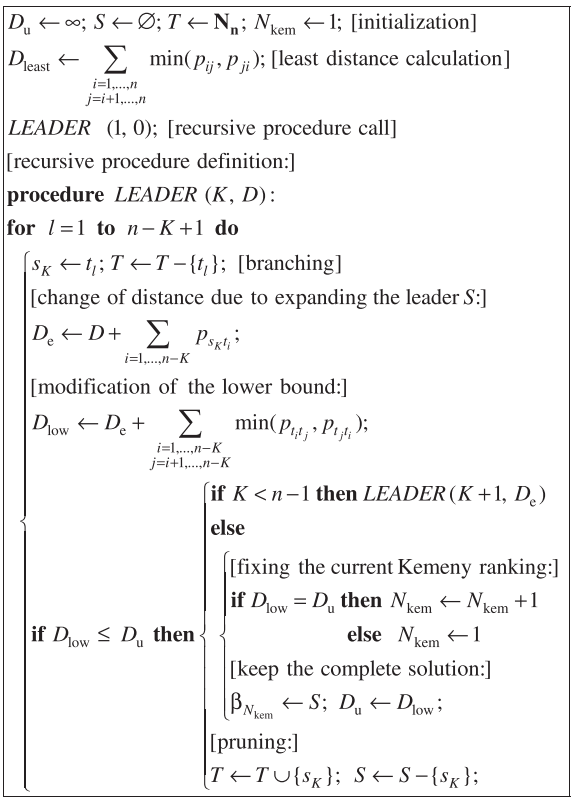

Fig. 1. The recursive B&B algorithm for all Kemeny rankings determination.

distribution of choices what is recognized to be a worst case for modeling preference profiles [13]. Ties were produced in similar way, and some additional measures were taken to reduce their density in each ranking. The function `randomize()` accepts some integer number, so called *starting point*, that served as an identifier of the generated profile (an corresponding individual KRP). Thus, using number of rankings m, number of ranking elements n and the starting point, anytime one can restore the particular individual profile.

This way, 3600 profiles were generated with different values of $m = 4 \ldots 15$ and $n = 10, 15, 20$, each served as input for the B&B algorithm. The detailed analysis of the experiment outcomes should be a subject of separate consideration. In this text, for the sake of conciseness, it can be reported that in most of the runnings, the algorithm outputs included relatively small number of optimal solutions: from 2 to 100. Relatively seldom (in nearly 10% of cases) N_{kem} was equal to several hundreds. However, in approx. 5% of cases, there were outstanding solution numbers approaching to one million and even more. Absolute record was 11,279,826 solutions for parameters $m = 4$, $n = 20$, starting point = 79, no ties.

One of the typical cases is shown in Table 1. The preference profile (which, in accordance to condition (6), is transitive as $D(\beta, \Lambda) = D_{least} = 472$), consisting of four rankings of 20 alternatives, having totally six ties, has 447,614 Kemeny rankings. The first seven optimal solutions for the profile are shown in Table 2. Notice that the B&B algorithm provided considerable reducing the solutions search space since it checked only 6,356,082 nodes while the cardinality of space of linear order relations is 20! = 2,432,902,008,176,640,000.

After removing all six ties the same profile brought to almost doubling of N_{kem} that became equal to 811,918, $D(\beta, \Lambda) = D_{least} = 476$, $N_{nds} = 10{,}440{,}879$.

Two example profiles above were found to be transitive and, nevertheless, resulted in the paramount number of solutions, evidently due to essential inconsistency between initial rankings. It is interesting to study a case of intransitive profile. Corresponding example is shown in Tables 3 and 4. One can see that the set of solutions comprises a cycle (shown by gray background in Table 4) indicative of Condorcet paradox. However, the second optimal ranking saves the situation, and the final optimal solution could be {1,4,9,3,8,2,10,6,5,7}.

The last example has been obtained not from pseudo-random generation and from an application the KRP to analysis of uncertainty intervals provided by m laboratories for some reference value of a measurand, see Section 3. In corresponding profile, each ranking includes one pair of alternatives with the strict order relation and $n - 2$ pairs with the tie. Thus, Profile 3 presented by Table 5 has a high density of ties and rather minor differences in rankings, i.e., the condition of impartial culture is not valid here. This profile brought to 1440 optimal solutions, though they were all similar in positions of first through eighth places, namely: {10,11,9,8,12,2,3,1,...}.

3. Application to interlaboratory comparisons

When organizing interlaboratory comparisons for proficiency testing (see, for example [27]) the main task consists usually in determination of the *reference value* x_{ref} and its uncertainty range. Let the participating in comparisons laboratories be measuring nominally the same quantity Y, that is

$$Y = X_i, \ i = 1, \ldots, m, \tag{9}$$

where X_i is the quantity measured by the ith laboratory and m is a number of laboratories participating in a comparison. The aim of the comparisons is to determine an estimate y of Y and the associated uncertainty $u(y)$ in terms of estimates x_i of the X_i provided by the laboratories and their associated standard uncertainties $u(x_i)$.

The estimate y is typically calculated as a weighted mean by formula

$$y = \sum_{i=1}^{m'} \frac{x_i}{u(x_i)^2} \Bigg/ \sum_{i=1}^{m'} \frac{1}{u(x_i)^2}, \tag{10}$$

and corresponding uncertainty is

$$u(y) = \sqrt{1 \Bigg/ \sum_{i=1}^{m'} \frac{1}{u(x_i)^2}}, \tag{11}$$

where m' is the number of laboratories, results of which deem to be reliable.

It is follows from expressions (10) and (11) that a procedure of determination of the reference value must

Table 1
Profile 1 at $m = 4$, $n = 20$ and six ties.

1	11	4	14	10	8	5	∼12	16	9	2	∼15	6	18	13	20	17	19	7	3
20	9	3	1	17	6	8	14	13	7	2	15	4	18	16	12	5	11	10	19
18	14	20	16	∼9	3	7	∼4	11	15	8	10	17	19	∼13	2	5	1	6	12
8	2	13	1	17	14	4	15	20	12	9	7	18	19	5	∼3	10	6	11	16

$D_{least} = 472$ $D(\beta, \Lambda) = 472$ $N_{kem} = 447{,}614$ $N_{nds} = 6{,}356{,}082$

Table 2
A fragment of the optimal solutions set for Profile 1.

1	8	14	4	20	9	2	13	15	17	7	12	18	3	5	11	10	6	16	19
1	8	14	4	20	9	2	13	15	17	7	12	18	3	5	11	10	16	6	19
1	8	14	4	20	9	2	13	15	17	7	12	18	3	5	11	10	16	19	6
1	8	14	4	20	9	2	13	15	17	7	12	18	3	5	11	16	10	6	19
1	8	14	4	20	9	2	13	15	17	7	12	18	3	5	11	16	10	19	6
1	8	14	4	20	9	2	13	15	17	7	12	18	3	5	16	11	10	6	19
1	8	14	4	20	9	2	13	15	17	7	12	18	3	5	16	11	10	19	6

Table 3
Profile 2 at $m = 5$, $n = 10$, no ties.

1	9	6	4	8	3	2	5	10	7
2	8	4	1	6	3	9	5	7	10
4	9	5	2	3	10	1	8	7	6
3	10	8	1	4	9	6	7	2	5
9	3	1	10	8	4	7	2	6	5

$D_{least} = 120$ $D(\beta, \Lambda) = 128$ $N_{kem} = 4$ $N_{nds} = 1376$

Table 4
The optimal solutions set for Profile 2.

1	4	9	3	2	10	8	6	5	7
1	4	9	3	8	2	6	5	10	7
1	4	9	3	8	2	10	6	5	7
1	4	9	3	10	8	2	6	5	7

provide a highest possible consistency of the participating laboratories measurement results x_i, that is to allow shaping a subset of maximal possible power m' of laboratories providing reliable results (so called *largest consistent subset*, LCS). Therewith, this procedure must facilitate identifying unreliable results and subsequent elimination of corresponding laboratories from the set of comparison participants. This decision takes place, if the following condition is valid:

$$E_n = \frac{|x_i - x_{ref}|}{\sqrt{u(x_i)^2 + u(y)^2}} > 1. \tag{12}$$

Different ways of the problem solving are described in many publications. For example, in [28] statistical criteria testing the consistency assumption are analyzed in conjunction with full enumeration when building the LCS(s). In paper [29] an algebraic approach has been proposed giving the procedure LCS determination of polynomial complexity. In [30], it was proposed to consider the uncertainty range $u(x_i)$ as the rectangular distribution and to deem that each participant gives one vote to each value within its uncertainty range and no votes for values outside this range. This produces a robust algorithm that is insensitive to outliers, i.e. results with the uncertainty considerably lower than those of other participants. Examples of this approach successful application have been described in [31].

Below we consider how the problem could be solved in terms of *preference aggregation*. This proposal was firstly described in the conference paper [32].

3.1. Transformation of laboratories uncertainty ranges to rankings

Designate the uncertainty range $[-u(x_i), u(x_i)]$ obtained by ith laboratory through I_i. Define the *range of actual values* of measured quantity as algebraic union of the uncertainty ranges obtained by each laboratory:

$$U = \cup_{i=1}^{m}[-u(x_i), u(x_i)] = \cup_{i=1}^{m} I_i. \tag{13}$$

Let us partition this range into equal intervals in such a way that their number ensures *enough accuracy* of the measured quantity values representation. Then we have n values of the measured quantity $\{a_1, a_2, \ldots, a_n\} = A$, corresponding to the interval bounds that will play part of alternatives in the consensus ranking determination problem; see Fig. 2 where a range of actual values is divided into 11 equal intervals, bounds of which corresponds to 12 alternatives. The laboratories will play part of m voters in the problem.

Let us compose the preference profile Λ which will consist of rankings describing the uncertainty ranges of each laboratory. Each kth rankings is the union of binary relations of strict order and equivalence possessing the following properties:

$$\begin{aligned} &a_i \succ a_j, \text{ if } a_i \in I_k \text{ and } a_j \notin I_k, \\ &a_i \sim a_j, \text{ if } a_i, a_j \in I_k \text{ or } a_i, a_j \notin I_k, \end{aligned} \tag{14}$$

whence it follows that each ranking includes *one pair* of alternatives with the strict order relation and $n - 2$ pairs with the equivalence relation.

Table 5
Profile 3 at $m = 10$, $n = 13$, 120 ties.

10	∼11	1	∼2	∼3	∼4	∼5	∼6	∼7	∼8	∼9	∼12	∼13
10	1	∼2	∼3	∼4	∼5	∼6	∼7	∼8	∼9	∼11	∼12	∼13
10	∼11	1	∼2	∼3	∼4	∼5	∼6	∼7	∼8	∼9	∼12	∼13
9	∼10	∼11	1	∼2	∼3	∼4	∼5	∼6	∼7	∼8	∼12	∼13
8	∼9	∼10	∼11	12	∼1	∼2	∼3	∼4	∼5	∼6	∼7	∼13
8	∼9	∼10	1	∼2	∼3	∼4	∼5	∼6	∼7	∼11	∼12	∼13
11	∼12	1	∼2	∼3	∼4	∼5	∼6	∼7	∼8	∼9	∼10	∼13
9	∼10	∼11	1	∼2	∼3	∼4	∼5	∼6	∼7	∼8	∼12	∼13
2	∼3	1	∼4	∼5	∼6	∼7	∼8	∼9	∼10	∼11	∼12	∼13
8	∼9	1	∼2	∼3	∼4	∼5	∼6	∼7	∼10	∼11	∼12	∼13
$D_{\text{least}} = 568$			$D(\beta, \Lambda) = 568$			$N_{\text{kem}} = 1440$				$N_{\text{nds}} = 83{,}237$		

Then, for an example of Fig. 2, we have ranking λ_1 of values for participating laboratory 1 as follows:

$$a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_{10} \sim a_{11} \sim a_{12}.$$

For ranking λ_2:

$$a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \sim a_{11} \succ a_1 \sim a_2 \sim a_{12},$$

and so on.

More obvious representation of the rankings can be as follows:

$$\{8,9\} \succ \{1,2,3,4,5,6,7,10,11,12\}, \{3,4,5,6,7,8,9,10,11\} \succ \{1,2,12\},$$

and so on.

Having constructed a preference profile containing measurement results of the all comparison participating laboratories, a consensus ranking can be found for it in the way described in Section 2. A winning alternative of the consensus ranking will be selected as the reference value x_{ref}.

3.2. An example

Consider how the proposed procedure works by an example taken from [30] that represents some model data where results of a reference object length measurements made by nine laboratories are shown in Fig. 3. The range of actual length values is reduced to Table 6.

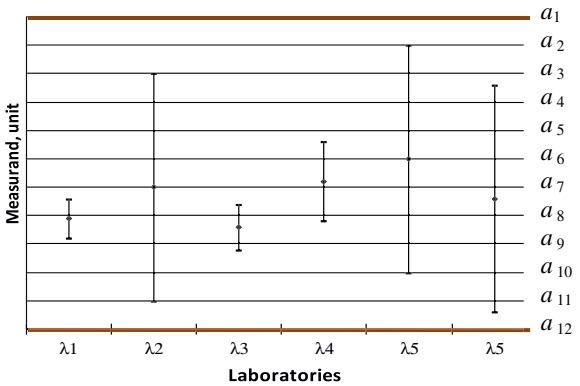

Fig. 2. An example of the range of actual values consisting of 11 equal intervals with corresponding 12 alternatives.

Rankings corresponding to uncertainty ranges of the nine laboratories produce the preference profile as follows:

$$\lambda_1 : a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_{10} \sim a_{11} \sim a_{12}$$
$$\lambda_2 : a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \sim a_{11} \succ a_1 \sim a_2 \sim a_{12}$$
$$\lambda_3 : a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_{10} \sim a_{11} \sim a_{12}$$
$$\lambda_4 : a_6 \sim a_7 \sim a_8 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_9 \sim a_{10} \sim a_{11} \sim a_{12}$$
$$\lambda_5 : a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \succ a_1 \sim a_{11} \sim a_{12}$$
$$\lambda_6 : a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \sim a_{11} \succ a_1 \sim a_2 \sim a_3 \sim a_{12}$$
$$\lambda_7 : a_8 \sim a_9 \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_{10} \sim a_{11} \sim a_{12}$$
$$\lambda_8 : a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \succ a_1 \sim a_2 \sim a_3 \sim a_4 \sim a_5 \sim a_{11} \sim a_{12}$$
$$\lambda_9 : a_3 \sim a_4 \sim a_5 \sim a_6 \sim a_7 \sim a_8 \sim a_9 \sim a_{10} \succ a_1 \sim a_2 \sim a_{11} \sim a_{12}.$$

For the given profile, the branch and bound algorithm found the following *eight* (that is $N_{\text{kem}} = 8$) optimal solutions of equal worth:

$$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$$
$$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_{12} \succ a_1$$
$$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_5 \succ a_4 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$$
$$a_8 \succ a_9 \succ a_6 \succ a_7 \succ a_{10} \succ a_5 \succ a_4 \succ a_3 \succ a_{11} \succ a_2 \succ a_{12} \succ a_1$$
$$a_8 \succ a_9 \succ a_7 \succ a_6 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$$
$$a_8 \succ a_9 \succ a_7 \succ a_6 \succ a_{10} \succ a_4 \succ a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_{12} \succ a_1$$
$$a_8 \succ a_9 \succ a_7 \succ a_6 \succ a_{10} \succ a_5 \succ a_4 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \succ a_{12}$$

Thus, we have a set of multiple optimal solutions $\beta = \{\beta_1, \beta_2, \ldots, \beta_8\}$ and must convolute them into a single one. Fortunately, in this case, the problem can be resolved on the basis of simple rational considerations. As, in the obtained solutions, the relations $a_4 \succ a_5$ and $a_5 \succ a_4$ occur the same number of times we conclude that $a_5 \sim a_4$. Similar considerations give $a_6 \sim a_7$ and $a_1 \sim a_{12}$. Then the final consensus relation is

$$\beta_{\text{final}} = \{a_8 \succ a_9 \succ a_6 \sim a_7 \succ a_{10} \succ a_4 \sim a_5 \succ a_3 \succ a_{11} \succ a_2 \succ a_1 \sim a_{12}\}.$$

The winner is the alternative a_8 which corresponds to the value 1.77395 inches (red[1] line in Fig. 3). It is selected as the reference value. It is clear from Fig. 3 that this value provides *maximal consistency* of the laboratory results in such a way that all results are consistent in this particular

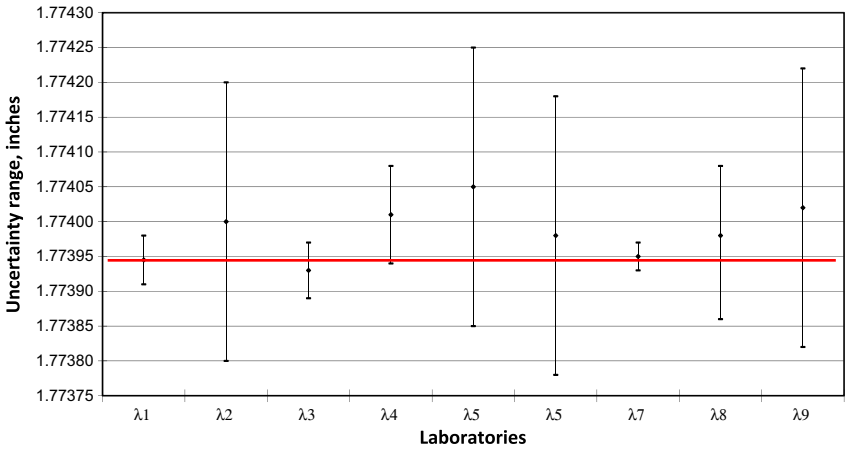

Fig. 3. Measurement results of nine laboratories for proficiency testing [30].

Table 6
Twelve alternatives of the actual length values range, inches.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
1.77430	1.77425	1.77420	1.77415	1.77410	1.77405	1.77400	1.77395	1.77390	1.77385	1.77380	1.77375

case. This result coincides with the outcome of Nielsen's analysis implemented by him using so called "Value Voted Most Likely To Be Correct" algorithm [30].

Experimental verification of the approach described above has been also carried out for other interlaboratory comparison data given in publications [27,29,31]. The obtained experimental outcomes confirm its correctness; for more examples, see [32].

3.3. Reasonable number of participating laboratories

When planning interlaboratory comparisons, it is necessary to assign a number m of participating laboratories. In this section, we are trying to apply simple probabilistic considerations to estimation of an upper bound for the number m, see also [24].

Let p be the probability of detecting the reference value by a single laboratory (we will call it *elementary probability*). Then the probability P of that the reference value is detected by m laboratories is defined as follows:

$$P = 1 - (1 - p)^m. \tag{15}$$

Clearly, the number of laboratories can be easily obtained from the formula (15), i.e.

$$m = \frac{\ln(1 - P)}{\ln(1 - p)}. \tag{16}$$

The graph plotted by formula (15) (see Fig. 4) shows that there is some critical value m_c of m such that any $m > m_c$ does not give an essential increase of number of attributes found. For example, at $p = 0.6$, there is no necessity to have more than five sensors as these five sensors have detected practically all attributes.

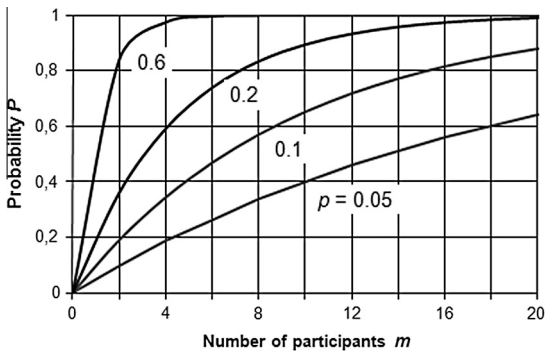

Fig. 4. Probability P depending on m for different values of p, see expression (15).

Let us now investigate how the probability P will increase after adding one more laboratory to the comparison participants. The following formula shows how many times the probability $P(m + 1) = P_1$ is greater than the probability $P(m) = P$:

$$\frac{P_1}{P} = \frac{1 - (1 - p)^{m+1}}{1 - (1 - p)^m} = 1 + \frac{p(1 - p)}{1 - (1 - p)^m}. \tag{17}$$

It can be seen from Fig. 5 that the increase of participants number by one results in a minor gain of the probability P. This gain becomes especially insignificant for all numbers $m > m_c = 4$. And the more probability p the more this insignificance is.

It is worth to estimate this gain in explicit and more general form. Let α be the relative probability growth

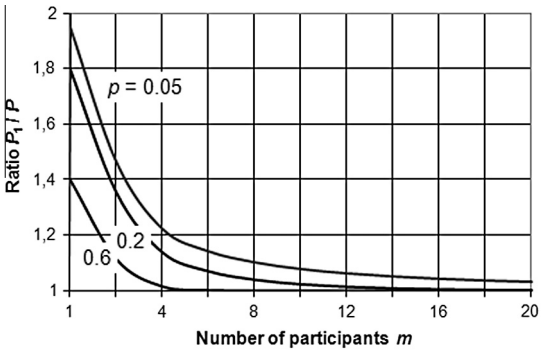

Fig. 5. The ratio P_1/P depending on m for different values of p, see expression (17).

resulting from inclusion of k additional laboratories into the group of m comparison participants, i.e.

$$\alpha = \frac{P(m+k) - P(m)}{P(m)} = \frac{P_k - P}{P}, \tag{18}$$

where

$$P_k = 1 - (1-p)^m (1-p)^k, \tag{19}$$

From (15), (20), and (21) we have

$$\alpha = (1-p)^m \frac{1-(1-p)^k}{1-(1-p)^m}. \tag{20}$$

Calculations of α are graphically presented in Fig. 6.

As Fig. 6 indicate an essential gain of the probability of detecting the reference value due to attraction of additional k participants exists only if the elementary probability p is low (see Fig. 6, $p = 0.05$). In this case, the dependency $\alpha(k)$ has almost linear character. However, already at $m = 7$, doubling of chances to find the reference value ($\alpha = 100\%$) happens only at $k = 10$.

At $p = 0.5$, if the comparison group has included 4 laboratories, addition of a new participant is useless as it gives out no new information on the reference value. In case of $p > 0.5$, one can see the loss of necessity in new participants as early as $k = 5$, though the comparison group consists of a single member.

Looking at Fig. 6 one can see that *it is important in what fashion a comparison group was setup*. Indeed, if $p = 0.05$, $m = 2$ and $k = 8$ give the relative growth $\alpha = 3.11$, whereas $m = 4$ and $k = 6$ produce only $\alpha = 1.16$, and at that the total number of participants is the same: $m + k = 10$. Thus, a combination of m and k with their fixed sum results in greater growth α, if $m < k$.

The number k can be easily determined in explicit form using expression (19). That is

$$k = \frac{\ln(1-P_k)}{\ln(1-p)} - m = m\frac{\ln(1-P_k)}{\ln(1-P)} - m, \tag{21}$$

In practice, the number k can be calculated on the assumption of desirable or critical value of P_k known. Surely, the elementary probability p should also be given, estimated or assigned.

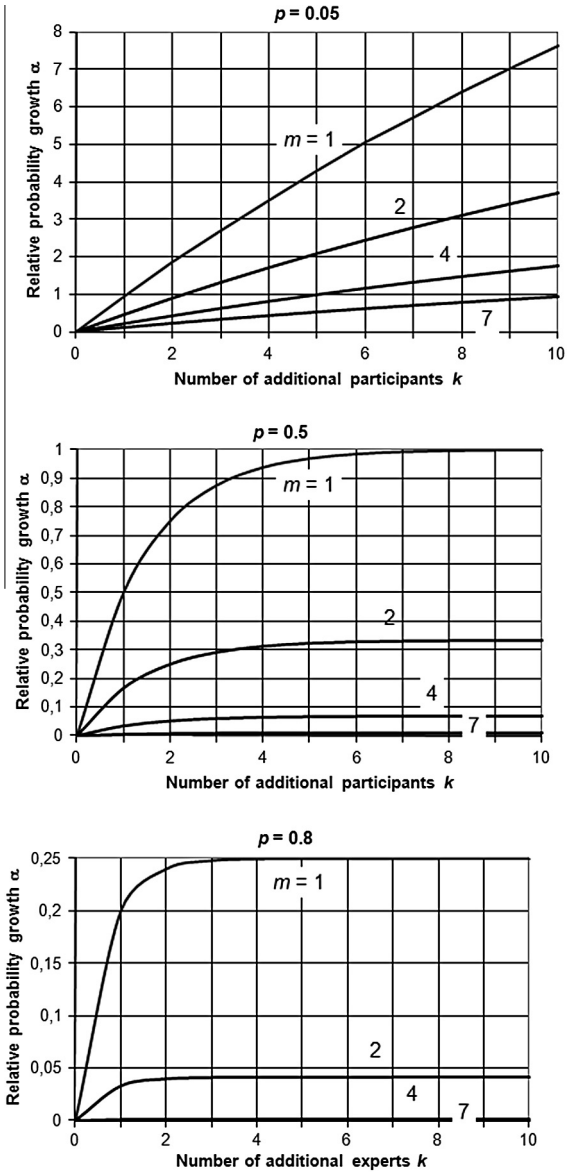

Fig. 6. Graphs (for $p = 0.05$, 0.5 and 0.8) of the relative probability growth α depending on number of additional laboratories k for different numbers m, see expression (20).

4. Conclusion

In the paper a preference aggregation model treated as ordinal measurement model has been considered. Determination of the ordinal scale measurement result in form of Kemeny ranking using an exact recursive B&B algorithm enables to draw at least two preliminary conclusions.

First, even if a preference profile is transitive (that is, all rankings in it are consistent) a number of multiple solutions of the KRP may be extremely large in spite of small amounts of m and n. This "chaotic" behavior of the KRP should be a subject of future research, see also [33].

Second, the multiple solutions require to develop special measures to build some appropriate final convoluting

solution. Though some appropriate ideas had been proposed in [16,17] the issue needs further investigations.

On the basis of preference aggregation approach a procedure of the largest consistent subset determination for interlaboratory comparisons has been proposed. Experimental verification of this method confirms its correctness. In contrast to other approaches, it provides additional information about interrelations of values in the actual values range. For instance, from the solution in paragraph 3.2 one can conclude that values a_6 and a_7 are equivalent to each other from the view point of their contribution into the final reference value and, at the same time, they both are less important for the reference value than the value a_8. This information, somehow or other, can be taken into account when analyzing the comparison results. The paper describes a simple probabilistic method for justification of reasonable number of interlaboratory comparison participants.

Obtained outcomes demonstrate the possibility of useful application of preference aggregation methods in metrological practice.

Acknowledgments

The author shall never forget Prof. L. Finkelstein's encouraging support and advertence and would like to thank Prof. R. Morawski for the honour to be invited to prepare a paper for this special issue of Measurement.

References

[1] S.V. Muravyov, V. Savolainen, Discrete-mathematical approach to formal description of measurement procedure, Measurement 18 (2) (1996) 71–80.
[2] S.V. Muravyov, V. Savolainen, Representation theory treatment of measurement semantics for ratio, ordinal and nominal scales, Measurement 22 (1/2) (1997) 37–46.
[3] S.V. Muravyov, Approach to mathematical structure concerning measurement science, in: K. Kariya, L. Finkelstein (Eds.), Measurement Science – A Discussion, Ohmsha & IOS Press, Tokyo, Amsterdam, 2000, pp. 37–49.
[4] S.V. Muravyov, V. Savolainen, Special interpretation of formal measurement scales for the case of multiple heterogeneous properties, Measurement 29 (2001) 209–223.
[5] S.V. Muravyov, Rankings: are they useful for measurement practitioners and are they in the scope of metrology and/or measurement science? TC7&TC21 round table "Logical and mathematical fundamentals of measurement science", in: Proceedings of the XVIII IMEKO World Congress, Rio de Janeiro, Brazil, <http://www.imeko.org/publications/wc-2006/PWC-2006-TC7-032u.pdf>, 2006.
[6] S.V. Muravyov, Rankings as ordinal scale measurement results, Metrology and Measurement Systems 13 (1) (2007) 9–24.
[7] F. Roberts, Discrete mathematical models with applications to social, Biological and Environmental Problems, Prince Hall Inc., Englewood Cliffs, 1976.
[8] F. Aleskerov, D. Bouyssou, B. Monjardet, Utility Maximization, Choice and Preference, second ed., Springer-Verlag, Berlin, 2007.
[9] L. Finkelstein, M.S. Leaning, A review of the fundamental concepts of measurement, Measurement 2 (1) (1984) 25–34.
[10] L. Finkelstein, Widely, strongly and weakly defined measurement, Measurement 34 (2003) 39–48.
[11] L. Finkelstein, Strength in numbers, Measurement and Control 41 (1) (2008) 21–24.
[12] J.P. Barthélemy, A. Guenoche, O. Hudry, Median linear orders: heuristics and a branch and bound algorithm, European Journal of Operational Research 42 (1989) 313–325.
[13] A.M.A. Van Deemen, The probability of the paradox of voting for weak preference orderings, Social Choice and Welfare 16 (1999) 171–182.
[14] J.G. Kemeny, J.L. Snell, Mathematical Models in the Social Sciences, Ginn, New York, 1962.
[15] C. Dwork, R. Kumar, M. Naor, D. Sivakumar, Rank aggregation methods for the Web, in: Proceedings of 10th WWW Conference, Hong Kong, 2001, pp. 613–622.
[16] M. Truchon, Aggregation of Rankings in Figure Skating, Working Paper 04-14, Universiŕe Laval, Canada, 2004.
[17] A. Davenport, D. Lovell, Ranking Pilots in Aerobatic Flight Competitions, IBM Research Report RC23631 (W0506-079), T.J. Watson Research Center, NY, 2005.
[18] V. Conitzer, A. Davenport, J. Kalagnanam, Improved bounds for computing Kemeny rankings, in: Proceedings of 21st AAAI Conference, Boston, MA, 2006, pp. 620–626.
[19] N. Betzler, M.R. Fellows, J. Guo, R. Niedermeier, F.A. Rosamond, Fixed-parameter algorithms for Kemeny rankings, Theoretical Computer Science 410 (45) (2009) 4554–4570.
[20] Shao Tao, M.C. Chan, S.V. Muravyov, E.V. Tarakanov, Pritrans: a prioritized converge-cast scheme using consensus ranking in wireless sensor networks, in: Proceedings of SAS- Limerick, Ireland, 2010, pp. 251–256.
[21] M.M.R. Mostert, G.A. Ayoko, S. Kokot, Application of chemometrics to analysis of soil pollutants, Trends in Analytical Chemistry 29 (5) (2010) 430–445.
[22] J.B. Luo, S.P. Etz, R.T. Gray, Normalized Kemeny and Snell distance: a novel metric for quantitative evaluation of rank-order similarity of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (8) (2002) 1147–1151.
[23] L. Lam, Ch.Y. Suen, Application of majority voting to pattern recognition: an analysis of its behaviour and performance, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 27 (5) (1997) 553–568.
[24] S.V. Muravyov, About reasonable number of rankings in preference profile when measuring quality, in: Proceedings of the 9th IMEKO Symposium on Measurement and Quality Control (ISMQC), Chennai, India, 2007, pp. 331–334.
[25] N. Nahar, K. Lyytinen, N. Huda, S.V. Muravyov, Critical success factors for information technology supported international technology transfer: finding expert consensus, Information and Management 43 (5) (2006) 663–677.
[26] S.V. Muravyov, E.V. Tarakanov, Multiple solutions of an exact algorithm for determination of all Kemeny rankings: preliminary experimental results, Proceedings of the International Conference on Instrumentation, Measurement, Circuits and Systems, ICIMCS 2011, Hong Kong, vol. 1, ASME Press, New York, 2011, pp. 17–20.
[27] M.A. Collett, M.G. Cox, T.J. Esward, P.M. Harris, J.A. Sousa, Aggregating measurement data influenced by common effects, Metrologia 44 (2007) 308–318.
[28] H.K. Iyer, C.M. Wang, D.F. Vecchia, Consistency tests for key comparison data, Metrologia 41 (2004) 223–230.
[29] M.G. Cox, The evaluation of key comparison data: determining the largest consistent subset, Metrologia 44 (2007) 187–200.
[30] H.S. Nielsen, Determining consensus values in interlaboratory comparisons and proficiency testing, in: Proceedings of NCSL International Workshop and Symposium, Tampa, Florida, USA, 2003, pp. 1–16.
[31] L. Brunetti, L. Oberto, M. Sellone, P. Terzi, Establishing reference value in high frequency power comparisons, Measurement 42 (2009) 1318–1323.
[32] S.V. Muravyov, I.A. Marinushkina, Largest consistent subsets in interlaboratory comparisons: preference aggregation approach, in: Proceeding of IMEKO TC1+TC7+TC13 Symposium, Jena, Germany, 2011, (urn:nbn:de:gbv:ilm1-2011imeko-036:0).
[33] T. Schwartz, From arrow to cycles, instability, and chaos by untying alternatives, Social Choice and Welfare 18 (2001) 1–22.