KEMENY'S MEDIAN ALGORITHM.

APPLICATION FOR DETERMINING GROUP JUDGEMENT

Hanna Bury, Dariusz Wagner

Systems Research Institute of the Polish Academy of Sciences

Newelska 6, 01-447 Warsaw, Poland

Basic elements of the process of group judgement

<table>
<tr><th>Input data</th><th>Task</th><th>Scale of ordering</th></tr>
<tr><td rowspan="2">Set of elements O
$O=\{O_1,\ldots, O_n\}$
$O_i, O_j \in O$</td><td rowspan="2">To determine preference order of a set of elements or pairwise comparisons</td><td>The order scale</td></tr>
<tr><td rowspan="2">The number scale</td></tr>
<tr><td>Set of criteria Q
$Q=\{Q_1, \ldots, Q_q\}$
$Q_s, Q_t \in Q$</td><td>To classify elements of a set</td></tr>
</table>

The problem considered is the following

There is a set of n alternatives $O = \{O_1, \ldots, O_n\}$ to be ranked by the group of K experts from the point of view of a given criterion or a set of criteria. It is assumed that their judgements may have the form of preference orders or pairwise comparisons and that ties can occur. One has to determine a group judgement being the aggregation of experts' opinions.

Basic notions and definitions

Pairwise comparisons

Given two elements O_i, $O_j \in \mathcal{O}$. The expert is asked to express his opinion on

(i) the alternative O_i is better (according to a given criterion $\mathcal{Q}$) than O_j; this condition is written as $O_i \underset{Q}{\succ} O_j$

(ii) the alternative O_i is worse (according to a given criterion $\mathcal{Q}$) than O_j; this condition is written as $O_i \underset{Q}{\prec} O_j$

(iii) the alternative O_i is equivalent (according to a given criterion $\mathcal{Q}$) to O_j; this condition is written as $O_i \underset{Q}{\approx} O_j$

Preference orders

Given set of elements $\mathcal{O}$, the criterion $\mathcal{Q}$ and K experts. Each of experts is asked to determine the preference order of elements of this set. Two cases are to be distinguished:

(i) for O_i, $O_j \in \mathcal{O}$ expert judgements may have the form $O_i \succ O_j$ or $O_j \succ O_i$ only

The preference order of elements is as follows $O_{i_1}, \ldots\ldots, O_{i_n}$ and an element placed at the first position is regarded as the best one and that placed at the last position is regarded as the worst one, in other words the element placed at the position i_l is better (according to the given criterion $\mathcal{Q}$) than that placed at the position i_{l+1}, i.e. $O_{i_l} \succ O_{i_{l+1}}$.

(ii) some elements may be considered as equivalent, i.e. judgements given in the form $O_i \approx O_j$ are accepted.

The preference order of elements is

$$\underbrace{O_{i_{11}}, \ldots\ldots, O_{i_{1 l_1}}}_{\text{the first position}}, \underbrace{O_{i_{21}}, \ldots\ldots, O_{i_{2 l_2}}}_{\text{the second position}}, \ldots, \underbrace{O_{i_{t1}}, \ldots\ldots, O_{i_{t l_t}}}_{\text{the t-th position}},$$

where the number of elements placed at the j-th position is equal to l_j and $\sum_{j=1}^{t} l_j = n$, $t \le n$. In the extreme case t=1 and all the elements of the set $\mathcal{O}$ are considered to be equivalent.

The Condorcet method

Given the experts' preference orders $\{P^{(k)}\} = \{P^1, ..., P^K\}$. For this set one can construct the so called outranking matrix L

$$
L = \begin{array}{c|c|c|c|c|c}
 & O_1 & O_2 & \cdots & O_n & WB_i \\ \hline
O_1 & - & l_{12} & \cdots & l_{1n} & \sum_{j=1}^{n} l_{1j} \\ \hline
O_2 & l_{21} & - & \cdots & l_{2n} & \sum_{j=1}^{n} l_{2j} \\ \hline
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline
O_n & l_{n1} & l_{n2} & \cdots & - & \sum_{j=1}^{n} l_{nj}
\end{array} ,
$$

where l_{ij} is the number of experts in whose opinion $O_i \succ O_j$

l_{ij} is the number of experts in whose opinion $O_j \succ O_i$ or $O_j \approx O_i$.

It is clear that $l_{ij} + l_{ji} = K, \quad i, j = 1,..., n, i \neq j$.

To avoid ambiguity it is assumed that the number of experts is odd.

The Condorcet winner is such an alternative O_i that $l_{ij} > \frac{K}{2}$, $j = 1, 2,..., n, i \neq j$. A difficulty with the Condorcet winner is that in general it need not exist.

The Borda method

Having assumptions as above it is worth to define so called Borda coefficient WB_i related to the alternative O_i

$$WB_i = \sum_{j=1,\ j\neq i}^{n} l_{ij}\ .$$

The Borda winner is such an alternative O_i that $WB_i = WB_{max}$. The Borda winner may be always (as opposite to the Condorcet winner) defined.

According to the Borda rule, the order of the set of alternatives $O = \{O_1, O_2, ..., O_n\}$ is determined by diminishing values of WB_i coefficients.

Example

Given following judgements of eleven experts

$P^1, \ldots, P^7 = \{O_1, O_2, O_3\}$

$P^8, \ldots, P^{11} = \{O_2, O_3, O_1\}$

The outranking matrix is as follows

$$L = \begin{array}{c|c|c|c|c} & O_1 & O_2 & O_3 & \sum s_{ij} \\ \hline O_1 & - & 7 & 7 & 14 \\ \hline O_2 & 4 & - & 11 & 15 \\ \hline O_3 & 4 & 0 & - & 4 \end{array} \quad \text{majority of experts} = 6 > \frac{11}{2}$$

Hence the alternative O_1 is the Condorcet winner and the alternative O_2 is the Borda winner.

The Arrow – Raynaud method

To avoid the disadvantages of the Borda method concerning the change of group opinion when the number of objects changed, Arrow i Raynaud proposed the following modification.

Given the outranking matrix L. For each row of the matrix L^n (the upper subscript has been added to denote the dimension of the matrix L), it means for each alternative O_i the maximum number of votes it has got as compared with other alternatives $l_{i\max} = \max_j l_{ij}$, $i = 1,\dots, n$ is determined. The alternative O_{i_n}, for which $l_{i_n} = \min_i l_{i\max}$ is placed at the last position. Next the row and column related to that alternative are removed from the L^n matrix and $l_{i\max}$ is determined again. This procedure is repeated unless the positions of all the alternatives are determined.

Example

Given outranking matrix L and the number of experts K=76.

$L^4 =$

	O_1	O_2	O_3	O_4	WB_i	$l_{i\max}$
O_1	–	40	62	48	150	62
O_2	36	–	76	62	174	76
O_3	14	0	–	40	54	40
O_4	28	14	36	–	78	**36**

majority of experts $> \frac{76}{2} = 38$.

The alternative O_4 will be placed at the last position in the preference order. After removing the fourth row and the fourth column the new matrix L^3 is as follows

$L^3 =$

	O_1	O_2	O_3	WB_i^3	$l_{i\max}$
O_1	–	40	62	102	62
O_2	36	–	76	112	76
O_3	14	0	–	14	**14**

.

The alternative O_3 will be placed at the last but one position on preference order. After removing the third row and the third column the new matrix L^2 is as follows

$$L^2 = \begin{array}{c|c|c|c|c} & O_1 & O_2 & WB_i^2 & l_{i\max} \\ \hline O_1 & - & 40 & 40 & 40 \\ \hline O_2 & 36 & - & 36 & \mathbf{36} \end{array} .$$

It can be seen that the alternative O_1 will be placed at the first position and O_2 at the second one. Hence the preference order has the form $O_1 \succ O_2 \succ O_3 \succ O_4$.

According to the Borda method the preference order is $O_2 \succ O_1 \succ O_4 \succ O_3$ and the Condorcet winner is the alternative O_1.

Kemeny's median

Given a preference order of n alternatives presented by the k-th expert (k=1,...,K)

$P^k = (O_{i_1}, O_{i_2}, \ldots, O_{i_n})$.

For such a preference order the following matrix of pairwise comparison can be constructed

$$\mathbf{A}^k = \begin{bmatrix} a^k_{11}, & \ldots, & a^k_{1n} \\ \vdots & \ddots & \vdots \\ a^k_{n1}, & \ldots, & a^k_{nn} \end{bmatrix} \text{ where } a^k_{ij} = \begin{cases} 1 & \text{for } O_i \succ O_j \\ 0 & \text{for } O_i \approx O_j \\ -1 & \text{for } O_i \prec O_j \end{cases}$$

Definition 1

Assume that two preference orders P^{k_1} and P^{k_2} are given. The distance between these two preference orders can be expressed as follows

$$d(P^{k_1}, P^{k_2}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ij}^{k_1} - a_{ij}^{k_2}\right|$$

Given a set of preference orders $\left\{P^{(k)}\right\} = \left\{P^1, \ldots, P^K\right\}$. The distance of some preference order P from this set is defined as follows

$$d\left(P, P^{(k)}\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K} d_{ij}\left(P, P^k\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K} d_{ij}^{k} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K}\left|a_{ij}^{k} - a_{ij}^{P}\right|$$

Assume that in the preference order P $O_i \succ O_j$, i.e. $a_{ij}^{P} = 1$. In order to determine the distance of this preference order from a given set $\left\{P^k\right\}$, one can make use of coefficients defined as follows

$$r_{ij} \underset{a_{ij}^P=1}{=} \sum_{k=1}^{K} d_{ij}\left(P, P^k\right) \underset{a_{ij}^P=1}{=} \sum_{k=1}^{K} \left|a_{ij}^k - a_{ij}^P\right| = \sum_{k=1}^{K} \left|a_{ij}^k - 1\right| \quad i, j = 1,\ldots n.$$

r_{ij} are called the loss coefficients and the matrix R=[r_{ij}] is called the loss matrix. It is assumed that r_{ij}=0 for all i=j. It should be noted that elements of the matrix R depend upon the form of preference orders P^k (k=1,...,K) only.

Making use of the coefficients r_{ij} (i,j=1,...,n) the distance of some preference order P from the set of experts' opinion can be rewritten as $d = \sum_{\substack{i \\ (i,j)\in I_P^{(1)}}} \sum_{j} r_{ij} + \sum_{\substack{i \\ (i,j)\in I_P^{(2)}}} \sum_{j} \sum_{k=1}^{K} \left|a_{ij}^k\right|$, where

$I_P^{(1)}$ - the set of indices (i,j) for which $O_i \succ O_j$ in the preference order P or - in other words - the set of indices for which $a_{ij}^P = 1$,

$I_P^{(2)}$ - the set of indices (i,j) for which $O_i \approx O_j$ in the preference order P or - in other words - the set of indices for which $a_{ij}^P = 0$.

Definition 2

A preference order P^M such that $M(P^1,\ldots,P^K) = \arg\min_P d(P, P^{(k)})$

is called the *median* of a set $(P^1,\ldots P^K)$.

In other words it is such a preference order that in sense of the distance is the "closest" one to all the preference orders of the set $\{P^{(k)}\}$.

To simplify considerations it is assumed that the set $I_P^{(2)}$ is empty. In other words it is assumed that equivalent alternatives do not occur in the median. Hence $d = \sum_{\substack{i \\ (i,j) \in I_P^{(1)}}} \sum_{j} r_{ij}$

One can show that the problem of determining Kemeny's median can be solved as an integer programming problem. The distance of a preference order P from the set of preference orders $\{P^{(k)}\}$ can be written as

$$d = \sum_{i=1}^{n} \sum_{j=1}^{n} y_{ij} r_{ij}, \text{ where } y_{ij} = \begin{cases} 1 & \text{if } (i, j) \in I_P^{(1)} \\ 0 & \text{if } (i, j) \notin I_P^{(1)} \end{cases}$$

Hence, the problem of determining the Kemeny median for a given set $\{P^{(k)}\}$ can be formulated as follows

$$\min_{y_{ij}} \ d \qquad \text{subject to}$$

$$y_{ij} + y_{ji} = 1, \quad \text{for } i \neq j \quad \sum_{i=1}^{n} \sum_{j=1}^{n} y_{ij} = \frac{n(n-1)}{2}, \quad i, j = 1, \ldots, n, \quad i \neq j$$

Moreover, the y_{ij} variables should be chosen in such a way that alternatives O_i (i=1,…,n) form a preference order. This can be written as

$$\text{there exist:}\begin{cases} i_1: & y_{i_1 j}=1, \quad j=1,\dots,n;\ j\neq i_1 \\ i_2\neq i_1: & y_{i_2 j}=1, \quad j=1,\dots,n,\ j\neq i_1,i_2 \\ i_a\neq i_1,i_2,\dots i_{a-1}: & y_{i_a j}=1, \quad j=1,\dots,n,\ j\neq i_1,i_2,\dots i_{a-1},i_a \\ & a=2,\dots,n-1 \end{cases}$$

Litvak introduced a modification of Kemeny's median which corresponds to the group of positional methods developed from the Borda method. He proposed a notion of so called preference vectors, formulated the modified definition of the distance and proved relevant theorems.

Heuristic algorithm for determining Kemeny's median

Given the loss matrix R and the number of experts K.

Step 1° The matrix $Q^{(n)}$ is to be determined

$$Q^{(n)} = E^{(n)} R^{(n)} E^{(n)}, \text{ where } E^{(n)} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}, \text{ dim E=n}$$

$R^{(n)} = R$, dim R=n.

Step 2° The smallest element excluding those of the main diagonal of the $Q^{(n)}$ matrix (denoted as $q^{n}_{ij\,min}$) is to be determined.

Next the values of $v^{n} = 2K\dfrac{(n-2)(n-3)}{2}$ as well as

$\tilde{d}^{n}_{ji\,min} = q^{n}_{ij\,min} - v^{n}$ are determined.

The alternatives that determine the row and the column of the $q^{n}_{ij\,min}$ element are denoted as O_{i_n} and O_{i_1} respectively.

Step 3° The rows and columns related to the alternatives O_{i_1} as well as O_{i_n} are removed from the $R^{(n)}$ matrix. The new matrix $R^{(n-2)}$ is obtained.

Step 4° The matrix $Q^{(n-2)}$ is to be determined

$$Q^{(n-2)} = E^{(n-2)} R^{(n-2)} E^{(n-2)}$$

Step 5° The smallest element excluding those of the main diagonal of the $Q^{(n-2)}$ matrix (denoted as $q_{ij\,min}^{n-2}$) is to be determined.

Next the values of $v^{n-2} = 2K\frac{(n-4)(n-5)}{2}$ as well as

$\tilde{d}_{ji\,min}^{n-2} = q_{ij\,min}^{n-2} - v^{n-2}$ are determined.

The alternatives that determine the row and the column of the $q_{ij\,min}^{n-2}$ element are denoted as $O_{i_{n-1}}$ and O_{i_2} respectively.

Step 6° The rows and columns related to the alternatives O_{i_2} as well as $O_{i_{n-1}}$ are removed from the $R^{(n-2)}$ matrix.

Steps 2 ÷ 4 are repeated unless the number of alternatives to be ordered is equal 4 – if n is an even number

3 – if n is an odd number.

If in any step of the algorithm an ambiguity occurs, i.e. the number of elements $q_{ij\,min}^{n}$ is greater than 1, all the next steps are to be repeated for each one of these elements

Example

Given the set of 7 alternatives O ={O_1, O_2, O_3, O_4, O_5, O_6, O_7} and the set of preference orders given by 11 experts for these alternatives

P^1 = {O_7, O_5, O_6, O_2, O_3, O_4, O_1}

P^2 = {O_4, O_1, O_5, O_6, O_2, O_7, O_3}

P^3 = {O_3, O_4, O_7, O_5, O_6, O_2, O_1}

P^4 = {O_5, O_7, O_6, O_1, O_4, O_3, O_2}

P^5 = {O_1, O_2, O_4, O_5, O_7, O_3, O_6}

P^6 = {O_5, O_1, O_2, O_6, O_4, O_7, O_3}

P^7 = {O_2, O_7, O_3, O_5, O_1, O_6, O_5}

P^8 = {O_3, O_1, O_7, O_4, O_5, O_6, O_2}

P^9 = {O_3, O_1, O_4, O_2, O_6, O_6, O_7}

P^{10} = {O_7, O_1, O_3, O_2, O_5, O_4, O_6}

P^{11} = {O_4, O_2, O_7, O_6, O_5, O_3, O_1}

The loss matrix R is as follows

$$R^{(7)} = \begin{bmatrix} 0 & 8 & 12 & 10 & 10 & 8 & 12 \\ 14 & 0 & 10 & 10 & 12 & 10 & 10 \\ 10 & 12 & 0 & 10 & 12 & 10 & 16 \\ 12 & 12 & 12 & 0 & 8 & 6 & 10 \\ 12 & 10 & 10 & 14 & 0 & 6 & 12 \\ 14 & 12 & 12 & 16 & 16 & 0 & 16 \\ 10 & 12 & 6 & 12 & 10 & 6 & 0 \end{bmatrix}$$

Step 1º $Q^{(7)}$ matrix is

$$Q^{(7)} = \begin{bmatrix} 330 & 344 & 352 & 340 & 344 & 364 & 338 \\ 338 & 330 & 344 & 334 & 340 & 360 & 330 \\ 330 & 338 & 330 & 330 & 336 & 356 & 332 \\ 342 & 348 & 352 & 330 & 342 & 362 & 336 \\ 338 & 342 & 346 & 340 & 330 & 358 & 334 \\ 318 & 322 & 326 & 320 & 324 & 330 & \mathbf{316} \\ 344 & 352 & 350 & 346 & 348 & 366 & 330 \end{bmatrix}$$

Step 2º $q^{7}_{ij\,min} = q_{67} = 316.$

Alternative O_7 is placed at the first position and O_6 at the last one.

The preference order to be considered is O_7,…,…,…,…,…,…, O_6.

v^7=220, $\tilde{d}^7_{ji\,min} = q^7_{ij\,min} - v^7$ = 316-220 = 96

Rows and columns related to the alternatives O_7 and O_6 are removed from the matrix R

Step 3º the new matrix $R^{(5)}$ is as follows

$$R^{(5)} = \begin{bmatrix} 0 & 8 & 12 & 10 & 10 \\ 14 & 0 & 10 & 10 & 12 \\ 10 & 12 & 0 & 10 & 12 \\ 12 & 12 & 12 & 0 & 8 \\ 12 & 10 & 10 & 14 & 0 \end{bmatrix}$$

Step 4º $Q^{(5)}$ matrix is

$$Q^{(5)} = \begin{bmatrix} 132 & 145 & 148 & 146 & 148 \\ 140 & 132 & 140 & 140 & 144 \\ \mathbf{138} & 146 & 132 & 142 & 146 \\ 140 & 146 & 144 & 132 & 142 \\ \mathbf{138} & 142 & 140 & 144 & 132 \end{bmatrix}$$

Step 5º $q^5_{ij\,min} = q_{31} = q_{51}$ = 138

v^5=66, $\tilde{d}^5_{ji\,min} = q^5_{ij\,min} - v^5$ = 138 – 66 = 72

Two orders are to be investigated.

1. For $q^{5}_{ij\,min} = q_{31}$ alternative O_1 is placed at the first position and O_3 at the last one. Hence the preference order to be considered is

$O_7, O_1, \ldots, \ldots, \ldots, O_3, O_6$.

The rows and columns related to alternatives O_3 and O_1 are removed from the matrix R. The new matrix $R^{(3)}$ is

$$
\begin{array}{c c}
 & \begin{array}{ccc} O_2 & O_4 & O_5 \end{array} \\
R^{(3)} = & \begin{bmatrix} 0 & 10 & 12 \\ 12 & 0 & 8 \\ 10 & 14 & 0 \end{bmatrix} \begin{array}{c} O_2 \\ O_4 \\ O_5 \end{array}
\end{array}
$$

There are two equivalent orders of remaining alternatives:

O_2, O_4, O_5 or O_4, O_5, O_2

The resulting preference orders are

$O_7, O_1, O_2, O_4, O_5, O_3, O_6$ or $O_7, O_1, O_4, O_5, O_2, O_3, O_6$

The distance of these orders from the given set of experts' orders is

d = 96 + 72 + 10 + 12 + 8 = 198.

2. For $q^{5}_{ij\,min} = q_{51}$ alternative O_1 is placed at the first position and O_5 at the last one.

The preference order to be considered is O_7, O_1,…,…,…, O_5, O_6.

The rows and columns related to alternatives O_5 and O_1 are removed from the matrix R.

The new matrix $R^{(3)}$ is

$$
\begin{array}{c} \\ R^{(3)} = \end{array}
\begin{array}{c} \begin{array}{ccc} O_2 & O_3 & O_4 \end{array} \\ \begin{bmatrix} 0 & 10 & 10 \\ 12 & 0 & 10 \\ 12 & 12 & 0 \end{bmatrix} \end{array}
\begin{array}{l} \\ O_2 \\ O_3 \\ O_4 \end{array}
$$

The remaining alternatives should be ordered as follows O_2, O_3, O_4.

The resulting preference order is O_7, O_1, O_2, O_3, O_4, O_5, O_6.

The distance of this order from the given set of experts' orders is

d = 96 + 72 + 10 + 10 + 10 = 198.

The loss matrix R may be rewritten as

	O_7	O_1	O_2	O_4	O_5	O_3	O_6	Σr_{ij}
O_7	0	10	12	12	10	6	6	56
O_1	12	0	8	10	10	12	8	48
O_2	10	14	0	10	12	10	10	42
O_4	10	12	12	0	8	12	6	26
O_5	12	12	10	14	0	10	6	16
O_3	16	10	12	10	12	0	10	10
O_6	16	14	12	16	16	12	0	**d=198**

References

[1] de Borda J-C.: Memoire sur les elections au Scrutin Histoire de l'Academie Royale des Sciences 1781.

[2] de Condorcet M.: Essai sur l'Application de l'Analyse à la Probabilité des Rendues à la Pluralité des Voix. Paris 1785.

[3] Bury H., Petriczek G., Wagner D. Determining of group opinion with the Kemeny's median method (in Polish). In: Preference Modelling and Risk' 99, ed. T. Trzaskalik, Katowice 1999.

[4] Bury H., Wagner D. The use of Kemeny's median for group decision making. Integer programming approach. Proceedings of the Sixth International Conference on Methods and Models in Automation and Robotics, Miedzyzdroje, 28-31 August 2000.

[5] Bury H., Wagner D. Computer Group Support Systems (in Polish). In: Information Technologies in Management. Decision Support Systems, (eds.) J.Studzinski, L.Drelichowski, O.Hryniewicz, J.Kacprzyk , Warszawa, 2000.

[6] Kemeny J. Mathematics without numbers, Daedalus 88, 1959.

[7] Kemeny J., Snell L. Mathematical Models in the Social Sciences. Ginn. Boston 1960.

[8] Litvak B.G. Expert Information. Methods of acquisition and analysis (in Russian). Radio and Communication, Moscow 1982.

[9] Nurmi H. Comparing voting systems, Dordrecht, D.Reidel, 1987.

[10] Saari D.G., Merlin V.R. A geometric examination of Kemeny's rule. Social Choice and Welfare, vol. 17, 2000.