Ranking procedures by pairwise comparison using random sets and the imprecise Dirichlet model

Lev V. Utkin
Department of Computer Science
St.Petersburg State Forest Technical Academy
Institutski per. 5, 194021, St.Petersburg, Russia
e-mail: lev.utkin@mail.ru

Methods for ranking of alternatives or objects by pairwise comparisons using random set theory are proposed in the paper. Efficient algorithms weekly depending on the number of independent sources of data are considered. Methods using the imprecise Dirichlet model are used for obtaining cautious comparison measures when the number of expert judgments is rather small and standard methods of random set theory may give risky results. The methods allow us to overcome some difficulties concerning the conflicting or contradictory sources of data. Various numerical examples illustrate the proposed algorithms and methods.

*Keywords:* expert judgments, pairwise comparison, random set theory, belief and plausibility functions, Dirichlet distribution, imprecise probabilities, decision making

## 1. Introduction

Many application problems (multi-attribute decision making, data classification, etc.) deal with ranking of alternatives or objects. The ranking of $n$ alternatives $a_1, ..., a_n$ can be expressed as

$$a_{i_1} \prec a_{i_2} \prec ... \prec a_{i_n}, \tag{1}$$

and means that the alternative $a_{i_n}$ is strictly preferred to $a_{i_{n-1}}$, the alternative $a_{i_{n-1}}$ is strictly preferred to $a_{i_{n-2}}$, ... , the alternative $a_{i_2}$ is strictly preferred to $a_{i_1}$. We will denote the ranking of $n$ alternatives by the sequence of indices $(i_1, ..., i_n)$ corresponding to (1).

Table 1
Numbers of different comparative judgments

| | $p_1$ | $p_2$ | ... | $p_{n-1}$ | $p_n$ |
|---|---|---|---|---|---|
| $p_1$ | X | $c_{1,2}$ | ... | $c_{1,n-1}$ | $c_{1,n}$ |
| $p_2$ | $c_{2,1}$ | X | ... | $c_{2,n-1}$ | $c_{2,n}$ |
| ... | ... | ... | ... | ... | ... |
| $p_{n-1}$ | $c_{n-1,1}$ | $c_{n-1,2}$ | ... | X | $c_{n-1,n}$ |
| $p_n$ | $c_{n,1}$ | $c_{n,2}$ | ... | $c_{n,n-1}$ | X |

There are a lot of ranking procedures depending on initial data and elicitation techniques. An interesting and comprehensive review of ranking procedures and their comparison have been carried out by Hüllermeier and Fürnkranz [1,2]. An important class of elicitation techniques consists of the psychological scaling models that use the concept of paired comparisons. Therefore, one of the prevailing ways for getting initial data for ranking is pairwise comparisons. The popularity of the paired comparison methods can perhaps be contributed to the observation that experts are more comfortable making comparisons rather than directly assessing a quantity of interest [3]. There are various methods of pairwise comparisons. One of the well-known is a method used in the Analytic Hierarchy Process (AHP) [4] where experts supply the ratio of their preferences of one decision over another. However, in spite of possible simplifications of this method [5,6], this elicitation procedure may be rather difficult for expert sometimes. Therefore, we consider the simplest comparisons when the experts compare alternatives without providing some degree of their preferences.

Let $A = (a_1, ..., a_n)$ be a set of alternatives. The comparative preference of the form $a_i \prec a_j$ means that $a_j$ is preferred to $a_i$. There are different interpretations of the above preferences. By $a_i \prec a_j$ here, we mean that $a_j$ is strictly more probable than $a_i$ and there exists a finitely additive probability $P$ such that [7,8]

$$a_i \prec a_j \Rightarrow P(a_i) < P(a_j).$$

We will denote $P(a_i) < P(a_j)$ by $p_i < p_j$ for short.

Suppose that experts supply $N$ comparative judgments about probabilities of $n$ states of nature or events. These judgments can be represented in the form of Table 1, where $c_{i,j}$ means the number of judgments of the form $p_j > p_i$ and $N = \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} c_{i,j}$. In particular, $c_{i,j}$ may be 0 for some $i$ and $j$. This means that there are no judgments comparing the alternatives $a_i$ and $a_j$.

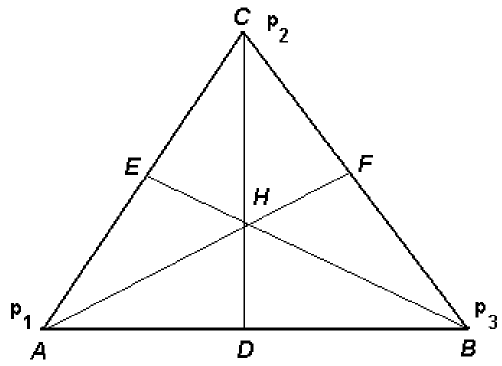


Figure 1. Polytopes produced by comparative judgments

For the case $n = 3$, these judgments restrict some polytopes (triangles) on a unit simplex $S(1, 3)$ of probabilities $p = (p_1, p_2, p_3)$ (see Fig.1). For example, the judgment $p_1 > p_2$ corresponds to the triangle $AEB$, the judgment $p_1 < p_2$ restricts a set of distributions $p$ by the triangle $BEC$. Every comparative probability judgment corresponds to some right-angled triangle of the considered simplex. Generally, if $n > 3$, then the judgments restrict polytopes on the $n$-dimensional simplex $S(1, n)$.

It is obvious that the number of possible different judgments formed by pairwise comparisons equals $L = n(n - 1)$. Since the judgment $p_j > p_i$ is supplied by experts $c_{i,j}$ times, then it can be said that some probability measure can be assigned to this judgment. This implies that every point $p$ of the simplex $S(1, n)$ can be regarded as a value of a random variable defined on the sample space $S(1, n)$ and having a probability density function or a set of density functions $\Phi = \{\varphi(p)\}$, $p \in S(1, n)$, which are defined by numbers $c_{ij}$ and some additional reasonable assumptions which will be considered below.

Denote every polytope produced by the judgment $p_j > p_i$ by $A_{ij}$. Then we have $c_{ij}$ polytopes $A_{ij}$, $i = 1, ..., n$, $j = 1, ..., n$, $j \neq i$. Consequently, by taking some frequency function defined for every $A_{ij}$ and depending on $c_{ij}$, we can find probabilities of arbitrary subsets of the simplex, which, in turn, will be a basis for computing the set of density functions $\Phi$. Since subsets $A_{ij}$ by different $i$ and $j$ are intersecting, one of the ways for choosing the frequency function is to use the so-called basic probability assignments from random set theory (or Dempster-

Shafer theory) [9,10]. So, we can use the random set theory for computing lower and upper probabilities (called belief and plausibility functions) of arbitrary events defined on the simplex. If we take events corresponding to different rankings of the form (1), then their belief and plausibility functions can be regarded as measures for comparing different rankings and for choosing the "optimal" one. Therefore, we study in the paper ranking procedures based on the comparison of belief and plausibility measures corresponding to each sequence of indices $(i_1, ..., i_n)$ and the first problem solved in the paper is to develop an algorithm for computing the belief and plausibility measures for every ranking.

It should be noted that initial data in the form of pairwise comparison can be elicited from different independent sources and the independence condition has to be taken into account. We propose to use the so-called Dempster rule of combination for combining independent sources of data. However, it is well known that this rule meets with huge computational difficulties, especially, when the number of sources is rather large. Therefore, the second problem solved in the paper is to develop an efficient algorithm for computing the belief and plausibility measures, whose complexity weekly depends on the number of independent sources of data.

It is obvious that the random sets can successfully be used if the number of expert judgments is large enough. Therefore, the third problem solved in the paper is to develop a method for computing the cautious belief and plausibility measures taking into account the possible small number of judgments. This method is based on using the imprecise Dirichlet model [11] and on extension of belief and plausibility measures [12]. It turns out that this method allows us to overcome some difficulties concerning the conflicting or contradictory sources of data. Various numerical examples illustrate the proposed algorithms and methods.

## 2. Belief functions and random sets

Let $U$ be a universal set under interest, usually referred to in evidence theory as the *frame of discernment*. Suppose $N$ observations were made of an element $u \in U$, each of which resulted in an imprecise (non-specific) measurement given by a set $A$ of values. Let $c_i$ denote the number of occurrences of the set $A_i \subseteq U$, and $\mathcal{P}(U)$ the set of all subsets of $U$ (power set of $U$). A frequency function $m$, called *basic probability assignment* (BPA), can be defined such that [9,10]:

$$m : \mathcal{P}(U) \to [0, 1],$$

$$m(\varnothing) = 1, \quad \sum_{A \in \mathcal{P}(U)} m(A) = 1.$$

Note that the domain of BPA, $\mathcal{P}(U)$, is different from the domain of a probability density function, which is $U$. According to [9], this function can be obtained as follows:

$$m(A_i) = c_i/N. \tag{2}$$

If $m(A_i) > 0$, i.e. $A_i$ has occurred at least once, then $A_i$ is called a *focal element*.

According to [10], the *belief* $Bel(A)$ and *plausibility* $Pl(A)$ measures of an event $A \subseteq \Omega$ can be defined as

$$Bel(A) = \sum_{A_i : A_i \subseteq A} m(A_i), \quad Pl(A) = \sum_{A_i : A_i \cap A \neq \varnothing} m(A_i). \tag{3}$$

As pointed out in [13], a belief function can formally be defined as a function satisfying axioms which can be viewed as a weakening of the Kolmogorov axioms that characterize probability functions. Therefore, it seems reasonable to understand a belief function as a generalized probability function [9] and the belief $Bel(A)$ and plausibility $Pl(A)$ measures can be regarded as lower and upper bounds for the probability of $A$, i.e., $Bel(A) \le \Pr(A) \le Pl(A)$.

Let us explain the belief and plausibility functions in terms of a *multivalued sampling process*. Consider a probability measure $P(\omega)$ defined on a universal set $\Omega$ (which can be thought of as the set of our observations) related to $U$ (the set of the values of our measurements) through a multivalued mapping $G : \Omega \to \mathcal{P}(U)$. Then the BPA is [9]:

$$m(A_i) = P(\omega_i) = c_i/N, \ \omega_i \in \Omega.$$

This multivalued mapping expresses the imprecision of the measurement experienced during each observation, i.e., our inability to attach a single number to each observation. So, for each set $A_i \in \mathcal{P}(U)$, the value $m(A_i)$ expresses the probability of $\omega_i = G^{-1}(A_i)$ ($\omega_i \in \Omega$). A *random set* is the pair $(\mathcal{F}, m)$, where $\mathcal{F}$ is the family of all $N$ focal elements.

Let $A$ be a subset of $U$. If we define $X_*$ as the subset of $\Omega$ whose elements must lead to $A$: $X_* = \{\omega \in \Omega : G(\omega) \subseteq A\}$, then the lower probability of $A$, according to Dempster's principle of inductive reasoning, is defined by $\underline{P}(A) = Bel(A) = P(X_*)$. If we define $X^*$ as the subset of $\Omega$ whose elements may lead to $A$: $X^* = \{\omega \in \Omega : G(\omega) \cap A \neq \varnothing\}$, then the upper probability of $A$ is given by $\overline{P}(A) = Pl(A) = P(X^*)$.

If there are $r$ independent different sources of evidence, then Dempster's rule of combination of evidence can be used for computing combined BPAs. The Dempster's rule combines multiple belief functions through their BPAs. Let $m_q(A_i^{(q)})$,

$i = 1, ..., n_q$, be the BPAs obtained from the $q$-th source of evidence. Then a combined BPA, $m(B)$, of a set $B$ is given by

$$m(B) = \frac{1}{1-K} \sum_{A_i^{(1)} \cap ... \cap A_j^{(q)} = B} \prod_{q=1}^{r} m_q(A_i^{(q)}), \tag{4}$$

where

$$K = \sum_{A_i^{(1)} \cap ... \cap A_j^{(q)} = \varnothing} \prod_{q=1}^{r} m_q(A_i^{(q)}). \tag{5}$$

$K$ represents basic probability mass associated with conflict. Note that Dempster's rule can not be used in case of $K = 1$, i.e., conflicting evidence can not be combined.

## 3. The first ranking procedure

Note that expression (1) can be written in probability terms as follows:

$$p_{i_1} < p_{i_2} < ... < p_{i_n}. \tag{6}$$

It follows from the above sections that every inequality $p_i < p_j$ in (6) forms the polytope $A_{ij}$ in the $n$-dimensional unit simplex $S(1, n)$ and can be characterized by the BPA $m(A_{ij}) = c_{ij}/N$. Condition (6) can be represented as a set of simple inequalities of the form $p_i < p_j$ as follows:

$$\begin{aligned} &p_{i_1} < p_{i_2},\ p_{i_1} < p_{i_3}, ...,\ p_{i_1} < p_{i_n}, \\ &p_{i_2} < p_{i_3},\ p_{i_2} < p_{i_4}, ...,\ p_{i_2} < p_{i_n}, \\ &\quad ... \\ &p_{i_{n-1}} < p_{i_n}. \end{aligned} \tag{7}$$

**Proposition 1** *Let $A_{ij}$ be a subset of the $n$-dimensional simplex $S(1, n)$ produced by the judgment $p_i < p_j$, $B(i_1, ..., i_n)$ a subset of the $n$-dimensional simplex $S(1, n)$ produced by the condition $p_{i_1} < p_{i_2} < ... < p_{i_n}$, i.e., $B(i_1, ..., i_n) = A_{i_1,i_2} \cap ... \cap A_{i_{n-1},i_n}$. The belief and plausibility measures (lower and upper probabilities) of $B(i_1, ..., i_n)$ can be computed as follows:*

$$Bel(B(i_1, ..., i_n)) = 0, \tag{8}$$

$$Pl(B(i_1, ..., i_n)) = \sum_{k=1}^{n-1} \sum_{j=1}^{k} m(A_{i_j, i_{k+1}}). \tag{9}$$

**Proof.** Note that

$$B(i_1,...,i_n) = A_{i_1,i_2} \cap A_{i_2,i_3} \cap ... \cap A_{i_{n-1},i_n}.$$

This implies that $B(i_1,...,i_n) \subset A_{i_v,i_w}$ and $A_{i_v,i_w} \subsetneq B(i_1,...,i_n)$ for all $i_v, i_w$. Consequently, $Bel(B) = 0$. On the other hand, the set $B$ belongs to all subsets produced by every inequality in (7). Consequently, there holds

$$\begin{aligned} Pl(B(i_1,...,i_n)) &= m(A_{i_1,i_2}) + m(A_{i_1,i_3}) + ... + m(A_{i_1,i_n}) \\ &+ m(A_{i_2,i_3}) + m(A_{i_2,i_4}) + ... + m(A_{i_2,i_n}) \\ &+ ... + m(A_{i_{n-1},i_n}) \\ &= \sum_{k=1}^{n-1} \sum_{j=1}^{k} m(A_{i_j,i_{k+1}}), \end{aligned}$$

as was to be proved. ■

Since the belief function is 0 for arbitrary sets of indices $(i_1,...,i_n)$, we compare different rankings $(i_1,...,i_n)$ only by comparing the plausibility functions. The rule for determining the "best" ranking is the following. The "best" ranking from all possible ones (6) should be chosen in such a way that makes $Pl(B(i_1,...,i_n))$ achieve its maximum.

**Example 1** *We turn to an example considered by Hüllermeier and Fürnkranz [2] concerning the difference between ranking methods. Suppose that there are $n = 4$ alternatives and the true ranking is given by $a_4 \prec a_3 \prec a_2 \prec a_1$. The matrix of pairwise comparisons is*

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $p_1$ | $X$ | 1 | 2 | 9 |
| $p_2$ | 9 | $X$ | 1 | 1 |
| $p_3$ | 8 | 9 | $X$ | 1 |
| $p_4$ | 1 | 9 | 9 | $X$ |

.

*The experts provided the judgment $p_4 > p_1$ have made an error, since they strongly prefers $a_4$ to $a_1$ ($c_{41} = 1$). By using the ranking procedure, we get from (8) and (9) $Bel(B(i_1,...,i_4)) = 0$ for all $(i_1,...,i_4)$, and*

$$\begin{aligned} Pl(B(4,3,2,1)) &= (9+9+1+9+8+9)/60 = 0.75, \\ Pl(B(4,3,1,2)) &= (9+9+1+9+8+1)/60 = 0.617, \\ Pl(B(4,1,3,2)) &= (1+9+9+2+1+9)/60 = 0.517, \\ &... \end{aligned}$$

*The maximum of the plausibility function is achieved at* $(i_1, ..., i_4) = (4, 3, 2, 1)$. *This implies that* $a_4 \prec a_3 \prec a_2 \prec a_1$. *This result coincides with the Slater-optimal ranking [14] and allows us to correct some possible incorrect judgments.*

Let us consider another example.

**Example 2** *Suppose that the pairwise comparisons are provided by two independent sources*

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 1 | 2 |
| $p_2$ | 0 | $X$ | 0 |
| $p_3$ | 0 | 1 | $X$ |

,

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 0 | 0 |
| $p_2$ | 1 | $X$ | 1 |
| $p_3$ | 2 | 0 | $X$ |

.

*By uniting both matrices, we get*

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 1 | 2 |
| $p_2$ | 1 | $X$ | 1 |
| $p_3$ | 2 | 1 | $X$ |

.

*Then for all vectors* $(i_1, i_2, i_3)$, *we get* $Bel(i_1, i_2, i_3) = 0$, $Pl(i_1, i_2, i_3) = 0.5$. *Hence, the optimal ranking can not be find. The main reasons of the impossibility to find optimal rankings are that the experts provides contradictory judgments and we do not take into account that these experts belong two different and independent groups (sources of evidence).*

The above example shows that the simple union of two or more sources might lead to incorrect results. Therefore, a combination rule taking into account independence of sources should be considered. We will use Dempster's rule of combination.

## 4. The second ranking procedure

Suppose that every expert supplies a set of comparative judgments and there are $r$ independent groups of experts (or experts). Then, by using Dempster's rule of combination, we can aggregate judgments of all experts.

Let $m_k(A_{ij}) = c_{ij}^{(k)}/N_k$ be the BPA computed on the basis of judgments provided by the $k$-th expert. Here $N_k$ is the total number of judgments provided by the $k$-th expert. By considering different groups of experts of different experts as independent sources of evidence, we can apply the Dempster 's combination rule for computing BPAs of the subset $B(i_1,...,i_n)$ or briefly $B$. Generally, Dempster's combination rule is a hard computational task. However, by using the specific structure of comparative judgments, we can significantly simplify the algorithm for computing the combined BPAs and, consequently, belief and plausibility functions.

The first direct way for computing $m(B)$ is to look over all possible products of BPAs of subsets $A_{ij}$ intersecting the subset $B$ and then to sum all products which satisfy conditions $\cap_{i,j}A_{ij}=B$ (for computing $(1-K)m(B)$) and $\cap_{i,j}A_{ij}=\varnothing$ (for computing $K$). Note that there exists $L=n\left(n-1\right)/2$ subsets $A_{ij}$ including $B$ and they are

$$A_{i_1,i_2}, A_{i_1,i_3}, ..., A_{i_1,i_n}, A_{i_2,i_3}, ..., A_{i_2,i_n}, ..., A_{i_{n-1},i_n}.$$

If we replace indices $i_v, i_{t+1}$ by $t+v$, $t=1,...,n-1$, $v=1,...,t$, then the above sequence of the subsets $A_{i_v,i_{t+1}}$ can be rewritten as $A_1, A_2, ..., A_L$. Then, according to the direct way of computation, we can write

$$m(B)=\frac{1}{1-K}\left[\sum_{i_1=1}^{L}\cdots\sum_{i_r=1}^{L}\prod_{q=1}^{r}m_q(A_{i_q})I(i_1,...,i_r)\right], \tag{10}$$

where $I(i_1,...,i_r)=1$ if $A_{i_1}\cap...\cap A_{i_r}=B$ and $I(i_1,...,i_r)=0$ otherwise.

The value $K$ can be obtained in the same way. Let us rewrite the set of non-intersecting pairs $A_{12}, A_{21}, ..., A_{n-1,n}, A_{n,n-1}$ as $A_1, A_2, ..., A_{2L}$. Then we can look over all possible products of BPAs of subsets $A_1, A_2, ..., A_{2L}$ under condition that there is at least one pair $A_{ij}, A_{ji}$ or $A_{i,i+1}, A_{i+1,i}$. Hence

$$K=\left[\sum_{i_1=1}^{2L}\cdots\sum_{i_r=1}^{2L}\prod_{q=1}^{r}m_q(A_{i_q})J(i_1,...,i_r)\right], \tag{11}$$

where $J(i_1,...,i_r)=1$ if $A_{i_1}\cap...\cap A_{i_r}=\varnothing$ and $J(i_1,...,i_r)=0$ otherwise.

The above way for computing $m(B)$ can simply be realized if the value $r$ is rather small. However, by increasing the number of independent sources of evidence, the computational problem becomes extremely hard because $L^r$ different products of BPAs (without $K$) and $(2L)^r$ different products of BPAs (by computing $K$) have to be looked over. One can see that $L^r+(2L)^r$ strongly increases with $r$. Therefore, we propose another way whose computational complexity weakly depends on the number $r$.

**Proposition 2** *Let $A_{ij}$ be a subset of the $n$-dimensional simplex $S(1,n)$ produced by the judgment $p_i < p_j$, $B(i_1,...,i_n)$ a subset of the $n$-dimensional simplex $S(1,n)$ produced by the condition $p_{i_1} < p_{i_2} < ... < p_{i_n}$, i.e., $B(i_1,...,i_n) = A_{i_1,i_2} \cap ... \cap A_{i_{n-1},i_n}$, $\mathcal{I}$ the set of all binary vectors $\mathbf{I} = (I_1,...,I_{n-1})$, $I_k \in \{0,1\}$, $\mathcal{J}$ the set of all binary vectors $\mathbf{J} = (J_1,...,J_L)$, $J_k \in \{0,1\}$, and $w(\mathbf{I})$ and $w(\mathbf{J})$ weights of $\mathbf{I}$ and $\mathbf{J}$, respectively, i.e., $\sum_{v=1}^{n-1} I_v$ and $\sum_{v=1}^{L} J_v$. Denote $L = n(n-1)/2$. If there are $r < n-1$ sources of evidence, then BPA of the event $B(i_1,...,i_n)$ obtained by means of Dempster's rule and denoted $m(B)$ is 0. If there are $r \geq n-1$ sources of evidence, then BPA of the event $B(i_1,...,i_n)$ obtained by means of Dempster's rule is determined as follows:*

$$m(B) = \frac{1}{1-K}\left[\sum_{\mathbf{I}\in\mathcal{I}}(-1)^{n-1-w(\mathbf{I})}\prod_{q=1}^{r}\left(F_q + \sum_{v=1}^{n-1} I_v \cdot m_q(A_{i_v,i_{v+1}})\right)\right], \tag{12}$$

*where*

$$F_q = \sum_{t=1}^{n-1}\sum_{v=1}^{t} m_q(A_{i_v,i_{t+1}}) - \sum_{v=1}^{n-1} m_q(A_{i_v,i_{v+1}}),$$

$$1-K = \sum_{\mathbf{J}^{(1)},\mathbf{J}^{(2)}\in\mathcal{J}} (-1/2)^{L-w(\mathbf{J}_2)} \prod_{q=1}^{r} G_q(\mathbf{J}^{(1)},\mathbf{J}^{(2)}), \tag{13}$$

*where*

$$G_q(\mathbf{J}^{(1)},\mathbf{J}^{(2)}) = \sum_{k=1}^{n-1}\sum_{v=1}^{k} J_{v+k}^{(2)}\left[\left(1-J_{v+k}^{(1)}\right) m_q(A_{v,k+1}) + J_{v+k}^{(1)} \cdot m_q(A_{k+1,v})\right].$$

**Proof.** If $r < n-1$, then the set $B(i_1,...,i_n)$ can not be obtained because this set is produced by intersecting at least $n-1$ subsets $A_{ij}$. Consequently, $m(B) = 0$. Let $r \geq n-1$. The BPA of $B(i_1,...,i_n)$ is determined by subsets $A_{ij}$ whose intersection is $B(i_1,...,i_n)$. For every $B(i_1,...,i_n)$, there exists only one set $\Psi$ of subsets $A_{ij}$ such that

$$B(i_1,...,i_n) = A_{i_1,i_2} \cap A_{i_1,i_3} \cap ... \cap A_{i_{n-1},i_n}.$$

However, there is a set $\Phi$ of some subsets $A_{i,j}$ which do not belong to $\Psi$, but $B(i_1,...,i_n) \subset A_{ij}$. Let us denote without loss of generality $a_1,...,a_{n-1} \in \Phi$ and $a_n,...,a_L \in \Psi$, where $a_k = A_{i_k,i_{k+1}}$.

According to Dempster's rule of combination, the BPA $m(B)$ can be found as a sum of products

$$m_1(a_1) \cdot m_2(a_2) \cdot ... \cdot m_r(a_r),$$

such that every product contains $r$ BPAs $m_q(a_i)$, $q = 1, ..., r$, where $a_i \in \Phi \cup \Psi$, and at least $n-1$ of the subsets from $r$ ones constitute the set $\Psi$. Therefore, every term of the sum can be divided into two parts: the first part consists of BPAs of all subsets from $\Psi$, the second part may consists of BPAs of arbitrary subsets from $\Phi \cup \Psi$. On the other hand, we can find $m(B)$ by computing the sums of all possible products of arbitrary subsets from $\Phi \cup \Psi$ and by subtracting from these sums the products of BPAs of all subsets whose intersection does not give $B$. It is easy to prove that the sums of all possible products of arbitrary subsets from $\Phi$ can be represented as $\prod_{q=1}^{r}(D_q + F_q)$, where $D_q$ is the sum of BPAs of all subsets from $\Phi$, i.e.,

$$D_q = \sum_{v=1}^{n-1} m_q(a_v),$$

and $F_q$ is the sum of BPAs of all subsets from $\Psi$. Then the considered product produces all possible combinations of $r$ BPAs. However, the set of all corresponding products contains some products of BPAs whose subsets do not form $B$, i.e., their intersection is not $B$. In order to find $m(B)$, these products must be removed. For instance, these products can be produced by removing from all product at least one subset $a_v$ such that $a_v \in \Phi$. The set of these products is formed by

$$\prod_{q=1}^{r}\left(D_q - m_q(a_v) + F_q\right).$$

All the removed products contain BPAs of subsets

$$\Phi \backslash a_1 \cup \Psi, ..., \Phi \backslash a_{n-1} \cup \Psi.$$

Simultaneously, we have twice removed all products with subsets $\Phi \backslash \{a_j, ..., a_k\} \cup \Psi$, where the number of elements in $\{a_j, ..., a_k\}$ is 2. This implies that the products with subsets $\Phi \backslash \{a_j, ..., a_k\} \cup \Psi$ have to be added. However, by adding all possible subsets $\Phi \backslash \{a_j, ..., a_k\} \cup \Psi$, we introduce superfluous subsets $\{a_j, ..., a_k\} \cup \Psi$, where the number of elements in $\{a_j, ..., a_k\}$ is 3, which have to be removed. The above procedure is repeated before we obtain $n-1$ components in $\{a_j, ..., a_k\}$, i.e.,

$\Phi\backslash\{a_j,...,a_k\}=\varnothing$. The procedure supplemented by all possible products of arbitrary subsets from $\Phi$ (when $\{a_j,...,a_k\}=\varnothing$) can be carried out by looking over all possible binary vectors $\mathbf{I}$ with 0 non-zero elements ($w(\mathbf{I})=0$), 1 non-zero elements ($w(\mathbf{I})=1$), 2 non-zero elements ($w(\mathbf{I})=2$), etc. The corresponding products are produced by

$$\prod_{q=1}^{r}\left(\sum_{v=1}^{n-1} I_v m_q(A_{i_v,i_{v+1}})+F_q\right),$$

where $I_v$ is the element of $\mathbf{I}$.

Let us consider $K$ now. Note that the intersection of $A_{ij}$ from different sources of evidence is empty if at least two sources provide disjoint sets $A_{ji}$ and $A_{ij}$. Denote without loss of generality $a_1^*=A_{12}$, $a_1^{**}=A_{21}$,..., $a_L^*=A_{n-1,n}$, $a_L^{**}=A_{n,n-1}$. Here $a_i^*\cap a_i^{**}=\varnothing$. Let us consider the set of combinations of subsets $a_i^*$ and $a_i^{**}$ produced by the following rule. If we take one of the binary vectors $\mathbf{J}^{(1)}\in\mathcal{J}$ and write a sequence $(a_1,...,a_L)$ such that $a_i=a_1^*$ by $J_i^{(1)}=0$ and $a_i=a_1^{**}$ by $J_i^{(1)}=1$, then every sequence $(a_1,...,a_L)$ does not contain at least one pair of non-intersecting subsets. Consequently, if we take $\mathbf{J}^{(1)}=(1,...,1)$, then every product $\prod_{q=1}^{r}\sum_{v=1}^{L}J_i^{(1)}m_q(a_i)$ does not belong to $K$ and can be added to $1-K$. However, after expansion these products taken for all vectors $\mathbf{J}^{(2)}\in\mathcal{J}$, we get terms containing terms with BPAs of subsets $\{a_1,...,a_L\}\backslash a_j$ two times, and one of identical terms has to be removed. For instance, BPAs of the set $\{a_1^*,a_2^*,...,a_{L-1}^*\}$ are computed twice by processing BPAs of the sets $\{a_1^*,a_2^*,...,a_{L-1}^*,a_L^*\}$ and $\{a_1^*,a_2^*,...,a_{L-1}^*,a_L^{**}\}$. Therefore, by taking all vectors $\mathbf{J}^{(2)}$ with the weight $L-1$ and dividing the obtained products into 2, we get the required number of the corresponding terms. However, by removing the terms, corresponding to the above vectors, from $1-K$, we loss products of BPAs, which correspond to subsets $\{a_1,...,a_L\}\backslash\{a_j,a_l\}$ with $L-2$ components, which are computed $2^2$ times. Therefore, by taking all vectors $\mathbf{J}^{(2)}$ with the weight $L-2$ and dividing the obtained products into $2^2$, we get the required number of the corresponding terms. By continuing the above procedure and by looking over all vectors $\mathbf{J}^{(2)}$, we obtain the products

$$2^{w(\mathbf{J}_2)-L}\cdot\prod_{q=1}^{r}G_q(\mathbf{J}^{(1)},\mathbf{J}^{(2)})$$

for every $\mathbf{J}^{(1)}\in\mathcal{J}$ and $\mathbf{J}^{(2)}\in\mathcal{J}$. By summing them over all $\mathbf{J}^{(1)}$ and $\mathbf{J}^{(2)}$, we get the value $1-K$, as was to be proved. ■

It follows from Proposition 2 that the combined BPAs of comparative judgments obtained by means of Dempster's rule can simply be computed. For computing $m(B)$, we do not need to know how the different subsets of the probability simplex $S(1,n)$ produced by comparative judgments interact with each other. Moreover, the complexity of computations weakly depends on the number $r$ of sources of evidence. One can see that $2^{n-1}$ different products of BPAs (without $K$) and $2^{2L}$ different products of BPAs (by computing $K$) have to be looked over. Therefore, the proposed expressions might be especially efficient by large values of $r$.

Proposition 2 can be generalized on arbitrary subsets of $S(1,n)$. Since every subset $C$ of $S(1,n)$ produced by a set of $w$ comparative judgments of the form $p_i < p_j$ can be represented as the intersection of $k$ subsets $A_{ij}$, then $m(C)$ can be computed by using (12)-(13) and by replacing $n$ by $w$ in these expressions. At that, the expression for $K$ remains without changes.

**Corollary 1** *Let $B(i_1,...,i_n)$ be a subset of the $n$-dimensional simplex $S(1,n)$ produced by the condition $p_{i_1} < p_{i_2} < ... < p_{i_n}$. Then the belief and plausibility functions of $B(i_1,...,i_n)$ under condition that there are $r$ sources of evidence are determined as follows:*

$$Bel(B) = m(B),$$

$$Pl(B) = \frac{1}{1-K}\left[\prod_{q=1}^{r}\left(\sum_{t=1}^{n-1}\sum_{v=1}^{t} m_q(A_{i_v,i_{t+1}})\right)\right].$$

**Proof.** There are no subsets of $S(1,n)$ included in $B$ except the set $B$. Therefore, the belief function is defined only by the BPA of $B$. Note that subsets intersecting the subsets $B$ are $A_{i_v,i_{t+1}}$, $t = 1,...,n-1$, $v = 1,...,t$. It is easy to see that all possible products of BPAs of these subsets are produced by

$$\prod_{q=1}^{r}\left(\sum_{t=1}^{n-1}\sum_{v=1}^{t} m_q(A_{i_v,i_{t+1}})\right).$$

Hence, we get the plausibility function. ■

**Example 3** *Let us return to Example 1 and suppose that the given pairwise comparisons are provided by three groups of experts (sources of evidence) such that we*

Table 2
Belief and plausibility function obtained by using the Dempster's rule for different rankings

| $(i_1, i_2, i_3, i_4)$ | $Bel$ | $Pl$ |
|---|---|---|
| $(4, 3, 2, 1)$ | 0.0262 | 0.524 |
| $(4, 3, 1, 2)$ | 0 | 0.267 |
| $(4, 1, 3, 2)$ | $4.37 \times 10^{-4}$ | 0.170 |
| ... | ... | ... |

*have $r = 3$ comparison matrices*

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $p_1$ | $X$ | 0 | 1 | 9 |
| $p_2$ | 3 | $X$ | 0 | 0 |
| $p_3$ | 0 | 6 | $X$ | 0 |
| $p_4$ | 0 | 1 | 4 | $X$ |

,

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $p_1$ | $X$ | 1 | 0 | 0 |
| $p_2$ | 0 | $X$ | 0 | 1 |
| $p_3$ | 8 | 3 | $X$ | 0 |
| $p_4$ | 0 | 0 | 5 | $X$ |

,

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $p_1$ | $X$ | 0 | 1 | 0 |
| $p_2$ | 6 | $X$ | 1 | 0 |
| $p_3$ | 0 | 0 | $X$ | 1 |
| $p_4$ | 1 | 8 | 0 | $X$ |

.

*It can be seen from the above matrices that the sum of judgments for every pairwise comparison coincides with the corresponding number of judgments in the matrix given in Example 1. Here $N_1 = 24$, $N_2 = 18$, $N_3 = 18$, $K = 0.152$. The results of computing the belief and plausibility functions are shown in Table 2.*

*The maximum of the belief and plausibility functions are achieved at $(i_1, ..., i_4) = (4, 3, 2, 1)$. This implies that $a_4 \prec a_3 \prec a_2 \prec a_1$.*

It can be seen from Example 3 that the optimal ranking does not explicitly depends on belief functions and coincides with the one considered in Example 1 without using the Dempster's rule of combination.

**Example 4** *Suppose that pairwise comparisons are provided by two groups of experts (sources of evidence) such that we have $r = 2$ comparison matrices*

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 28 | 2 |
| $p_2$ | 0 | $X$ | 0 |
| $p_3$ | 0 | 2 | $X$ |

,

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 0 | 0 |
| $p_2$ | 30 | $X$ | 1 |
| $p_3$ | 1 | 0 | $X$ |

.

*It can be seen that the pairwise comparisons of different sources are rather conflicting and $K = 0.8$. The belief and plausibility functions for different rankings are shown in Table 3.*

Table 3
Belief and plausibility function obtained by using the Dempster's rule for different rankings

| $(i_1, i_2, i_3)$ | $Bel$ | $Pl$ |
|---|---|---|
| $(1, 2, 3)$ | 0.133 | 0.142 |
| $(1, 3, 2)$ | 0 | 0 |
| $(2, 1, 3)$ | 0.284 | 0.293 |
| $(2, 3, 1)$ | 0 | 0 |
| $(3, 1, 2)$ | 0.133 | 0.142 |
| $(3, 2, 1)$ | 0.284 | 0.293 |

Table 4
Belief and plausibility function for different rankings

| $(i_1, i_2, i_3)$ | $Bel$ | $Pl$ |
|---|---|---|
| $(1, 2, 3)$ | 0 | 0.484 |
| $(1, 3, 2)$ | 0 | 0.5 |
| $(2, 1, 3)$ | 0 | 0.515 |
| $(2, 3, 1)$ | 0 | 0.5 |
| $(3, 1, 2)$ | 0 | 0.484 |
| $(3, 2, 1)$ | 0 | 0.515 |

*It follows from the computation results that the optimal rankings are* $(2, 1, 3)$, $(3, 2, 1)$. *Moreover, we can not distinguish these two rankings. It is interesting to note that the rankings* $(1, 2, 3)$ *and* $(3, 1, 2)$ *are preferred to* $(1, 3, 2)$ *and* $(2, 3, 1)$. *If we unite both matrices*

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 28 | 2 |
| $p_2$ | 30 | $X$ | 1 |
| $p_3$ | 1 | 2 | $X$ |

,

*then the optimal rankings are the same* $(2, 1, 3)$ *and* $(3, 2, 1)$ *(see Table 4), but the rankings* $(1, 3, 2)$ *and* $(2, 3, 1)$ *are preferred to* $(1, 2, 3)$ *and* $(3, 1, 2)$.

The above example illustrates how the optimal rankings can be changed if we unite two independent sources. The following example shows another feature when

Table 5
Belief and plausibility function obtained by using the Dempster's rule for different rankings

| $(i_1, i_2, i_3)$ | $Bel$ | $Pl$ |
|---|---|---|
| $(1, 2, 3)$ | 0.1 | 0.3 |
| $(1, 3, 2)$ | 0 | 0 |
| $(2, 1, 3)$ | 0.2 | 0.4 |
| $(2, 3, 1)$ | 0 | 0 |
| $(3, 1, 2)$ | 0.2 | 0.4 |
| $(3, 2, 1)$ | 0.1 | 0.3 |

united set of pairwise comparisons (see Example 2) is divided into two independent sources.

**Example 5** *Let us return to Example 2, we have* $r = 2$ *sources of evidence. It can be seen that the pairwise comparisons of different sources are rather conflicting and* $K = 0.375$*. The belief and plausibility functions for different rankings are shown in Table 5.*

*It follows from the computation results that the optimal rankings are* $(2, 1, 3)$, $(3, 1, 2)$ *and we can not distinguish these two rankings. Nevertheless, by taking into account independence of two sources, we could choose two optimal rankings (see Example 2 for comparison).*

Now we have to define a rule for determining the optimal ranking when belief and plausibility functions are non-zero. In fact, this question is reduced to the problem how to compare two overlapping intervals. There exist a lot of methods for comparison of the overlapping intervals. The question of choosing the "best" method should be addressed to a decision maker. Nevertheless, we consider one of the most attractive and justified methods using the so-called caution parameter [15] or the parameter of pessimism $\eta$ which has the same meaning as the optimism parameter in Hurwicz criterion [16]. According to this method, the "best" ranking from all possible ones should be chosen in such a way that makes the convex combination $\eta \cdot Bel(B) + (1 - \eta)Pl(B)$ achieve its maximum. Here $\eta \in [0, 1]$ is the caution parameter. If $\eta = 1$, then we analyze only belief functions and make pessimistic decision. This type of decision is very often used [17,18]. If $\eta = 0$, then we analyze only plausibility functions and make optimistic decision.

## 5. Extended belief functions and the imprecise Dirichlet model

Definition (2) of BPAs can be used when the number of expert judgments $N$ is rather large. However, this condition may be violated in many real applications. Sometimes, we have single expert judgments from every source of evidence. If $N$ is small, inferences become too precise and incautious. In order to overcome this difficulty, the imprecise Dirichlet model [11] can be applied to extend belief and plausibility functions such that a lack of sufficient statistical data can be taken into account [19,12]. Another reason of using the imprecise Dirichlet model is a possible contradiction of different sources of evidence when $K$ in (4) is closed to 1. It turns out that the imprecise Dirichlet model allows us to avoid the contradiction.

Let $U = \{u_1, ..., u_M\}$ be a set of possible outcomes $u_j$. Assume the *standard multinomial model*: $N$ observations are independently chosen from $U$ with an identical probability distribution $\Pr\{u_j\} = \theta_j$ for $j = 1, ..., M$, where each $\theta_j \geq 0$ and $\sum_{j=1}^{M} \theta_j = 1$. Denote $\theta = (\theta_1, ..., \theta_M)$ and $\mathbf{n} = (n_1, ..., n_M)$, where $n_j$ is the number of observations of $u_j$ in $N$ trials, so that $n_j \geq 0$ and $\sum_{j=1}^{M} n_j = N$. Under the above assumptions the random variables $n_1, ..., n_M$ have a multinomial distribution.

The *Dirichlet* $(s, \mathbf{t})$ *prior distribution* for $\theta$, where $\mathbf{t} = (t_1, ..., t_M)$, has probability density function [20]

$$p(\theta) = \Gamma(s) \left( \prod_{j=1}^{M} \Gamma(st_j) \right)^{-1} \cdot \prod_{j=1}^{M} \theta_j^{st_j - 1}.$$

Here the parameter $t_i \in (0, 1)$ is the mean of $\theta_i$ under the Dirichlet prior; the hyperparameter $s > 0$ determines the influence of the prior distribution on posterior probabilities; the vector $\mathbf{t}$ belongs to the interior of the $M$-dimensional unit simplex denoted by $S(1, M)$; $\Gamma(\cdot)$ is the Gamma-function. When multiplied by the multinomial likelihood function, the Dirichlet $(s, \mathbf{t})$ prior density generates a posterior density function

$$p(\theta|\mathbf{n}) \propto \prod_{j=1}^{M} \theta_j^{n_j + st_j - 1},$$

which is seen to be the probability density function of a Dirichlet $(N + s, \mathbf{t}^*)$ distribution, where $t_j^* = (n_j + st_j)/(N + s)$.

The *imprecise Dirichlet model* (IDM) is defined by Walley [11] as the set of all Dirichlet $(s, \mathbf{t})$ distributions such that $\mathbf{t} \in S(1, M)$. For the IDM, the *hyperparameter* $s$ determines how quickly upper and lower probabilities of events converge as statistical data accumulate. Walley [11] defined $s$ as a number of observations needed to reduce the imprecision (difference between upper and lower probabilities) to half its initial value. Smaller values of $s$ produce faster convergence and stronger conclusions, whereas large values of $s$ produce more cautious inferences. At the same time, the value of $s$ must not depend on $M$ or a number of observations. The detailed discussion concerning the parameter $s$ and the IDM can be found in [21,11].

Let $A$ be any non-trivial subset of $U$, i.e., $A$ is not empty and $A \neq U$, and let $n(A)$ denote the observed number of occurrences of $A$ in the $N$ trials, $n(A) = \sum_{u_j \in A} n_j$. Then, according to [11], the predictive probability $P(A|s)$ under the Dirichlet posterior distribution is

$$P(A|s) = (n(A) + st(A)) \,/\, (N + s) \,,$$

where $t(A) = \sum_{u_j \in A} t_j$.

It should be noted that $P(A|s) = 0$ if $A$ is empty and $P(A|s) = 1$ if $A = U$. By maximizing and minimizing $P(A|s)$ over $\mathbf{t} \in S(1, M)$, we obtain the posterior upper and lower probabilities of $A$:

$$\underline{P}(A|s) = n(A) / \,(N + s) \,, \; \overline{P}(A|s) = (n(A) + s) / \,(N + s) \,.$$

Before making any observations, $n(A) = N = 0$, so that $\underline{P}(A|s) = 0$ and $\overline{P}(A|s) = 1$ for all non-trivial events $A$. Therefore, by using the IDM, we do not need to choose one specific prior.

Now suppose that the outcomes are some subsets $A_j$ of the set $U$, but not only its points $u_j$. On one hand, this implies that these subsets can be considered in terms of belief function. On the other hand, we will show that these subsets can be considered in the framework of the IDM and this leads to the so-called extended belief and plausibility functions. A detailed description of the extended belief and plausibility functions was given by Utkin [12]. Here we shortly obtain these functions in terms of the multivalued sampling process (see section "Belief functions"). Suppose that the set $\Omega$ (the set of observations) consists of $M$ points $\omega_1, ..., \omega_M$. Every observed subset $A_j$, corresponds to one point $\omega_j$ with the probability $P(\omega_j) = \theta_j$, $j = 1, ..., M$. By having $N$ observations of $\omega_1, ..., \omega_M$ independently chosen from $\Omega$ with probabilities $P\{\omega_j\} = \theta_j$, $j = 1, ..., M$, we deal with the multinomial model. By assuming that the probabilities $\theta = (\theta_1, ..., \theta_M)$ have the Dirichlet $(s, \mathbf{t})$ distribution, we obtain the lower probability of $A$ as follows:

$$\underline{P}(A|s) = \frac{n(X_*) + st(X_*)}{N + s}.$$

where $t(X_*) = \sum_{\omega_j \in X_*} t_j$, $n(X_*) = \sum_{\omega_j \in X_*} c_j$.

By using the IDM and denoting $\varkappa = N/(N+s)$, we get for $A \subset U$

$$\begin{aligned}\underline{P}(A|s) &= \min_{\mathbf{t}\in S(1,M)} \frac{n(X_*) + st(X_*)}{N+s} = (N+s)^{-1} \sum_{\omega_j \in X_*} c_j \\ &= N \cdot Bel(A)/(N+s) = \varkappa \cdot Bel(A). \qquad (14)\end{aligned}$$

If $A = U$, then $\underline{P}(A|s) = 1$. The upper probability of $A$ can be obtained in the same way:

$$\overline{P}(A|s) = (N \cdot Pl(A) + s) \,/\, (N+s) = 1 - \varkappa(1 - Pl(A)). \qquad (15)$$

It should be noted that $\underline{P}(A|s)$ and $\overline{P}(A|s)$ are belief and plausibility functions with the BPA $m^*(A_i) = c_i/(N+s)$ for every $A_i$ and the additional BPA $m^*(U) = s/(N+s)$, i.e., $\underline{P}(A|s)$ and $\overline{P}(A|s)$ can be obtained as standard belief and plausibility functions under condition that there are $s$ additional observations $A_{n+1} = U$. If we denote $m(A_i) = c_i/N$, then $m^*(A_i) = m(A_i) \cdot N/(N+s) = \varkappa \cdot m(A_i)$, and

$$\underline{P}(A|s) = \sum_{A_i : A_i \subseteq A} m^*(A_i), \;\; \overline{P}(A|s) = m^*(U) + \sum_{A_i : A_i \cap A \neq \varnothing} m^*(A_i).$$

The above also follows from an interpretation of the hyperparameter $s$ as the number of *hidden* observations [11]. At the same time, the value $1 - \varkappa$ can be regarded as discount rate [10] characterizing the reliability of a source of data. This implies that the application of Walley's IDM leads to a discounting scheme with discounting rates strongly defined by the number of observations $N$ and by the hyperparameter $s$.

## 6. Cautious ranking procedure

Let us consider the first ranking procedure from the beginning. If there are $N$ comparative judgments, then the belief and plausibility functions given in (8) and (9) can be extended by using the IDM with hyperparameter $s$. If we denote extended belief and plausibility functions by $Bel_s$ and $Pl_s$, respectively, then, according to (14) and (15), there hold

$$Bel_s(B) = \varkappa \cdot Bel(B) = \frac{N \cdot Bel(B)}{N+s} = 0,$$

$$Pl_s(B) = 1 - \varkappa(1 - Pl(B)) = \frac{N \cdot Pl(B) + s}{N+s}.$$

It can be seen from the above expressions that the inequality $Pl(B_1) \leq Pl(B_2)$ implies the inequality $Pl_s(B_1) \leq Pl_s(B_2)$. Consequently, the extension of the belief and plausibility functions by using the IDM does not impact on the first procedure of ranking.

Now we consider how to extend the belief and plausibility functions in the second ranking procedures when there are $r$ sources of evidence. Suppose that there are $N_q$ comparative judgments from the $q$-th source of evidence. Denote $\varkappa_q = N_q/(N_q + s)$. The following proposition and corollary determine the extended belief and plausibility functions of the event $B(i_1, ..., i_n)$.

**Proposition 3** *If there are $r \geq n-1$ sources of evidence, then, BPA of the event $B(i_1, ..., i_n)$ obtained by means of Dempster's rule and by using the IDM under conditions and by notations of Proposition 2 is determined as follows:*

$$m^*(B) = \frac{1}{1-K^*}\left[\sum_{\mathbf{I}\in\mathcal{I}}(-1)^{n-1-w(\mathbf{I})}\prod_{q=1}^{r}\left(F_q^* + \sum_{v=1}^{n-1} I_v \cdot m_q(A_{i_v,i_{v+1}})\right)\right],$$

*where*

$$F_q^* = \varkappa_q \cdot F_q + (1-\varkappa_q),$$

$$K^* = K \cdot \prod_{q=1}^{r} \varkappa_q.$$

*Here $K$ and $F_q$ are determined in Proposition 2.*

**Proof.** The IDM produces the BPAs $m_q^*(A_i) = \varkappa_q \cdot m_q(A_i)$ for every $A_i$ and the additional BPA $m_q^*(U) = s/(N_q + s) = 1 - \varkappa_q$. Here $U = S(1, n)$. Then it follows from Proposition 2 that

$$\begin{aligned} F_q^* &= \sum_{t=1}^{n-1}\sum_{v=1}^{t} \varkappa_q m_q(A_{i_v,i_{t+1}}) - \sum_{v=1}^{n-1} \varkappa_q m_q(A_{i_v,i_{v+1}}) + m_q^*(U) \\ &= \varkappa_q \cdot F_q + (1-\varkappa_q). \end{aligned}$$

It follows from (11) that

$$K^* = \left[\sum_{i_1=1}^{2L}\cdots\sum_{i_r=1}^{2L}\prod_{q=1}^{r}\varkappa_q m_q(A_{i_q})J(i_1,...,i_r)\right] = K \cdot \prod_{q=1}^{r}\varkappa_q,$$

as was to be proved. ■

**Corollary 2** *Let $B(i_1, ..., i_n)$ be a subset of the $n$-dimensional simplex $S(1, n)$ produced by the condition $p_{i_1} < p_{i_2} < ... < p_{i_n}$. Then the belief and plausibility functions of $B(i_1, ..., i_n)$ under condition that there are $r$ sources of evidence by using the IDM are determined as follows:*

$$Bel_s(B) = m^*(B),$$

$$Pl_s(B) = \frac{1}{1 - K^*} \left[ \prod_{q=1}^{r} \left( \sum_{t=1}^{n-1} \sum_{v=1}^{t} \varkappa_q m_q(A_{i_v, i_{t+1}}) + (1 - \varkappa_q) \right) \right].$$

**Proof.** The prof is obvious. ∎

**Example 6** *Suppose that pairwise comparisons are provided by two groups of experts (sources of evidence) such that we have $r = 2$ comparison matrices*

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 10 | 0 |
| $p_2$ | 0 | $X$ | 0 |
| $p_3$ | 0 | 0 | $X$ |

,

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 0 | 0 |
| $p_2$ | 1 | $X$ | 0 |
| $p_3$ | 0 | 0 | $X$ |

.

*It can be seen that the pairwise comparisons of different sources are conflicting and $K = 1$. In other words, we can not make any decision by using Dempster's rule of combination. It is interesting to note that, by adding new non-contradictory sources of evidence, we do not get any solution and $K = 1$ because two sources are conflicting. At the same time, one can intuitively conclude that the second source deserves no credit because 10 experts from the first source supply the judgment $p_1 < p_2$ compared to one expert from the second source supplying the judgment $p_2 < p_1$. To overcome this difficulty, we use the IDM with $s = 1$. The calculation results are shown in Table 6. One can see from Table 6 that the rankings with $p_2 < p_1$ have the larger plausibility 0.92.*

*One can see from the results that it is difficult to find the optimal ranking in this example even by using the IDM. However, by adding new sources of evidence this situation might be changed. Suppose that we have the following additional judgments*

| | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| $p_1$ | $X$ | 2 | 4 |
| $p_2$ | 3 | $X$ | 1 |
| $p_3$ | 2 | 5 | $X$ |

.

Table 6
Belief and plausibility function obtained by using the IDM for different rankings

| $(i_1, i_2, i_3)$ | $Bel_1$ | $Pl_1$ |
|---|---|---|
| $(1, 2, 3)$ | 0 | 0.92 |
| $(1, 3, 2)$ | 0 | 0.92 |
| $(2, 1, 3)$ | 0 | 0.17 |
| $(2, 3, 1)$ | 0 | 0.17 |
| $(3, 1, 2)$ | 0 | 0.92 |
| $(3, 2, 1)$ | 0 | 0.17 |

Table 7
Updated belief and plausibility function obtained by using the IDM for different rankings

| $(i_1, i_2, i_3)$ | $Bel_1$ | $Pl_1$ |
|---|---|---|
| $(1, 2, 3)$ | 0.05 | 0.39 |
| $(1, 3, 2)$ | 0 | 0.58 |
| $(2, 1, 3)$ | 0.04 | 0.08 |
| $(2, 3, 1)$ | 0 | 0.06 |
| $(3, 1, 2)$ | 0.1 | 0.49 |
| $(3, 2, 1)$ | 0.05 | 0.10 |

*Without using the IDM, the additional judgments do not help us and this fact has been indicated above. However, if we take $s = 1$, then the optimal ranking can be found. The calculation results are shown in Table 7.*

The numerical example shows that the IDM allows us to find the optimal ranking when at least two sources of evidence are contradictory. Moreover, the IDM gives a more cautious solution when we have a small number of comparative judgments.

## 7. Conclusion

Methods for ranking of alternatives or objects based on using random set theory and the imprecise Dirichlet model have been proposed in the paper. The following virtues of the methods can be pointed out:

1. The methods allow us to correct some incorrect judgments (see Example 1).

2. The methods use the simplest type of pairwise comparisons.
3. The methods take into account the possible independence of sources of data.
4. The computational complexity of the methods weakly depends on the number of sources of data.
5. The methods give cautious decisions when the number of expert judgments is rather small.
6. The methods result interval-valued measures for comparison and different decision procedures (pessimistic, optimistic or their combination) can be used for choosing the "best" ranking.

## REFERENCES

1. E. Hullermeier and J. Furnkranz. Ranking by pairwise comparison: A note on risk minimization. In *Proc. of the IEEE International Conference on Fuzzy Systems, FUZZ-IEEE-04*, Budapest, Hungary, July 2004.
2. E. Hullermeier and J. Furnkranz. Comparison of ranking procedures in pairwise preference learning. In *Proc. of the 10th International Conference on Information Processing and Management of Uncertainty in Knoweldge-Based Systems, IPMU-04*, Perugua, Italy, 2004.
3. R.M. Cook. *Experts in Uncertainty. Opinion and Subjective Probability in Science*. Oxford University Press, New York, 1991.
4. T.L. Saaty. *Multicriteria Decision Making: The Analytic Hierarchy Process.* McGraw Hill, New York, 1980.
5. M. Beynon, B. Curry, and P. Morgan. The Dempster-Shafer theory of evidence: An alternative approach to multicriteria decision modelling. *Omega*, 28:37–50, 2000.
6. T. Tervonen, R. Lahdelma, and P. Salminen. A method for elicitating and combining group preferences for stochastic multicriteria acceptability analysis. TUCS Technical Report 638, Turku Centre for Computer Science, Turku, Finland, November 2004.
7. S.K.M. Wong, Y.Y. Yao, and P. Lingras. Comparative beliefs and their measurement. *International Journal of General Systems*, 22:69–90, 1993.
8. G. Regoli. Comparative probability orderings. In *Documentation Section on the website of the Society for Imprecise Probability Theory and Applications.* SIPTA, http://www.sipta.org, 1999.

9. A.P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annales of Mathematical Statistics*, 38:325–339, 1967.
10. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
11. P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996. with discussion.
12. L.V. Utkin. Extensions of belief functions and possibility distributions by using the imprecise Dirichlet model. *Fuzzy Sets and Systems*, 154(3):413–431, 2005.
13. J.Y. Halpern and R. Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.
14. P. Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48:303–312, 1961.
15. K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung*, volume I Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg, 2001.
16. R.D. Luce and H. Raiffa. *Games and decisions*. Wiley, New York, 1957.
17. T. Augustin. Expected utility within a generalized concept of probability - a comprehensive framework for decision making under ambiguity. *Statistical Papers*, 43:5–22, 2002.
18. K.-M. Osei-Bryson. Supporting knowledge elicitation and consensus building for Dempster-Shafer decision models. *International Journal of Intelligent Systems*, 18:129–148, 2003.
19. L.V. Utkin. Probabilities of judgments provided by unknown experts by using the imprecise Dirichlet model. *Risk, Decision and Policy*, 9(4):391 – 400, 2004.
20. S. Wilks. *Mathematical Statistics*. Wiley and Sons, New York, 1962.
21. J.-M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2-5):123–150, 2005.