6th International Symposium on Imprecise Probability: Theories and Their Applications, Durham, United Kingdom, 2009

Approximation of coherent lower probabilities by 2-monotone measures

Andrey G. Bronevich
Technological Institute of Southern Federal University, Taganrog, RUSSIA
brone@mail.ru

Thomas Augustin
Department of Statistics, Ludwig-Maximilians University (LMU), Munich, GERMANY
thomas@stat.uni-muenchen.de

Abstract

The paper investigates outer approximations of coherent lower probabilities by 2-monotone measures. We characterize the set of (Pareto)-optimal outer approximations and provide powerful iterative algorithms to calculate such measures.

Keywords. Pareto optimal 2-monotone measure, additivity on lattices, simplex method, imprecision indices.

1 Introduction

Walley [21, p. 51] is often cited in saying that he does not "…know any 'rationality' argument for two-monotonicity, beyond its computational convenience." Of course, in particular in problems of larger scale, computational convenience, and even computational tractability, is still an issue, and so the problem of finding a suitable approximation of a coherent lower probability by 2-monotone measures arises naturally in many applications of imprecise probabilities (see also Section 3).

As analysis shows, the optimal choice of a 2-monotone measure can not be made uniquely, which may be understood from the fact that the minimum of two 2-monotone measures is not again a 2-monotone measure in general, and so we will characterize and derive Pareto optimal solutions to that problem.

The main idea of this paper consists of the following. For any coherent probability μ, we define a convex set $M_{2-mon\le\mu}$ of 2-monotone measures that are dominated by μ. Then any possible optimal choice of a 2-monotone measure in $M_{2-mon\le\mu}$ is produced by finding extreme points of $M_{2-mon\le\mu}$, which are not dominated by other measures in $M_{2-mon\le\mu}$, and any optimal measure is represented as a linear convex combination of such points. After some technical preliminaries (section 2) and a slightly more detailed look at the convenience of 2-monotonicity, we give in section 4 a necessary and sufficient condition for a 2-monotone measure to be an extreme point through lattices on which a 2-monotone measure is additive. In Section 5, we provide iterative algorithms for searching optimal extreme points, which then are illustrated by two examples. In the Appendix the reader can find some results on canonical sequences of monotone measures [5], which are used in the proofs.

2. Technical preliminaries

Let X be a measurable space and $\mathfrak{A}$ be a σ-algebra of its subsets. A set function $\mu:\mathfrak{A}\to[0,1]$ is called a *monotone measure* [14] if 1) $\mu(\varnothing)=0$, $\mu(X)=1$; and 2) $A,B\in\mathfrak{A}$, $A\subseteq B$ implies $\mu(A)\le\mu(B)$. We write $\mu_1\le\mu_2$ for monotone measures μ_1,μ_2 on $\mathfrak{A}$ if $\mu_1(A)\le\mu_2(A)$ for all $A\in\mathfrak{A}$. In this paper we consider the following families of monotone measures:

1) M_{mon} is the set of all monotone measures on $\mathfrak{A}$;

2) M_{pr} is the set of all finite additive probability measures on $\mathfrak{A}$, i.e. $M_{pr}\subseteq M_{mon}$ and additionally $\mu(A\cup B)=\mu(A)+\mu(B)$ for disjoint sets $A,B\in\mathfrak{A}$;

3) M_{low} is the set of all *lower probabilities* [22] on $\mathfrak{A}$, i.e. $M_{low}\subseteq M_{mon}$ and for any $\mu\in M_{low}$ there exists $P\in M_{pr}$ such that $\mu\le P$ (, and so $\mu\in M_{low}$ iff it satisfies the avoiding sure loss property [22]);

4) M_{coh} is the set of all *coherent lower probabilities* [22] on $\mathfrak{A}$, i.e. for any $\mu\in M_{coh}$ and $B\in\mathfrak{A}$ there exists $P\in M_{pr}$ such that $\mu\le P$ and $\mu(B)=P(B)$;

5) M_{2-mon} is the set of all 2-monotone measures [11] on $\mathfrak{A}$, i.e. $M_{2-mon}\subseteq M_{mon}$ and $\mu(A)+\mu(B)\le\mu(A\cup B)+\mu(A\cap B)$ for any $A,B\in\mathfrak{A}$.

6) M_{chain} is the set of all chain measures [14] on $\mathfrak{A}$, i.e. if $\mu\in M_{chain}$, then there is a chain $\Gamma\subseteq\mathfrak{A}$ such that $\varnothing\in\Gamma$, $X\in\Gamma$ and, for all B, $\mu(B)=\sup_{A\in\Gamma|A\subseteq B}\mu(A)$.

3. On the convenience of 2-monotonicity

As also discussed below, 2-monotone measures have some regular properties compared to coherent lower probabilities, which are very convenient from the computational point of view. Of particular importance is the property recalled in Remark 1 below, ensuring that for any chain of events there is a single classical probability in the core simultaneously attaining the lower probability for all elements of the chain. As a consequence, the enveloping lower and upper distribution functions define probabilities in the core, and so, for instance, a closed form for natural extension (calculating expectation of random variables) is available (repeated, e.g, in. [22, p. 30ff], where also some direct applications are given). By similar arguments a convenient closed form for calculating lower and upper conditional probabilities (in Walley's sense) can be derived (see, e.g., [22, p. 301, including the corresponding footnote]). Moreover, other common forms of conditioning, like Dempster's rule of conditioning ([13]), also called maximum likelihood updating ([15]), are then guaranteed to lead to a coherent, and indeed again 2-monotone, solution.

Our main motivation for the present study, however, has been the case of hypothesis testing, where one has to distinguish between two hypotheses described by imprecise probabilities, and decide which one is more likely to have produced the data. Similarly as in the case of calculating the conditional distribution or the natural extension, the testing problem can be expressed in terms of a single linear optimization problem (see [1, chapter 4]), but, even with the considerable improvement along the lines developed for decision problems in [20, section 3.2], the problem still increases exponentially in the sample size, and so still is, for the sample sizes usually common in statistics, simply computationally intractable.

A powerful way out is offered by Huber-Strassen theory ([18], and the work following it, see also [17, 3, 4] for reviews from different perspectives). The famous Huber-Strassen theorem (in [18, cf. also the finally obtained extension in [9]) ensures that 2-monotonicity is sufficient for the existence of a globally least favorable pair, i.e. a pair of classical probability distributions that

i) allow to represent the whole testing problem in determining the optimal test and

ii) can be calculated by considering sample size 1 only.

While i) can be alleviated by a concept of local least favorability ([1, chapter 3), [2], [16]), property ii) can not be generalized appropriately (see the analysis of the proof in [1, p. 223ff.]). As a consequence, statistical models described by coherent, but not 2-monotone measures, often have to be approximated appropriately to be able to determine appropriate statistical testing procedures.

4 Approximation by 2-monotone measures (finite case)

In this case, we assume that X is a finite set and $\mathfrak{A}$ is the power set of X, i.e. $\mathfrak{A} = 2^X$. Let $\mu \in M_{low}$, then $\nu \in M_{mon}$ is defined as a Pareto optimal approximation of μ if $\nu \le \mu$ and $\nu \le \nu' \le \mu$ for $\nu' \in M_{mon}$ implies that $\nu' = \nu$. For any $\mu \in M_{low}$, we denote $M_{2-mon\le\mu} = \{\nu \in M_{2-mon} \mid \nu \le \mu\}$.

Lemma 1. *Any Pareto optimal 2-monotone measure for a $\mu \in M_{coh}$ can be represented as a convex linear combination of Pareto optimal extreme points of $M_{2-mon\le\mu}$.*

Proof. It is clear that the set $M_{2-mon\le\mu}$ has a finite set of extreme points $\{\mu_i\}$, because it can be described by a finite number of inequalities. Therefore any $\nu \in M_{2-mon\le\mu}$ can be represented as a linear convex combination of these points, i.e. $\nu = \sum_i a_i \mu_i$, where $a_i \ge 0$, $\sum_i a_i = 1$. Assume that in the above representation there is an extreme point $\mu_{i'}$ such that $a_{i'} > 0$ and $\mu_{i'}$ is not Pareto optimal, i.e. there is $\mu' \in M_{2-mon\le\mu}$ such that $\mu_{i'} < \mu'$ (i.e., $\mu_{i'} \le \mu'$ and $\mu_{i'} \ne \mu'$). Then we define $\nu' = \sum_{i|i\ne i'} a_i\mu_i \ + a_{i'}\mu'$. It is clear that $\nu' \in M_{2-mon\le\mu}$ and $\nu < \nu'$, therefore, ν is not Pareto optimal, which means that the coefficient a_i has to be equal to zero if the corresponding extreme measure μ_i is not Pareto optimal. This fact proves the lemma.■

The previous lemma says that the full description of Pareto optimal 2-monotone measures for $\mu \in M_{coh}$ can be given by knowing only its Pareto optimal extreme 2-monotone measures. Therefore, we have to answer the following question: what characteristics define extreme points uniquely? For this reason, we further involve some results concerning additivity properties of 2-monotone measures. We will consider lattices of the algebra $\mathfrak{A}$. A lattice is a subset of $\mathfrak{A}$ closed with respect to intersection and union. We say that $\mu \in M_{2-mon}$ is additive on a lattice $\mathfrak{L} \subseteq \mathfrak{A}$ if $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$ for any $A, B \in \mathfrak{L}$. Next straightforward result shows the way how we can describe additivity of 2-monotone measures.

Lemma 2. *Let $\mathfrak{S}$ be the set of all possible lattices in $\mathfrak{A}$, on which $\mu \in M_{2-mon}$ is additive. Then $\mathfrak{S}$ is a covering*[1] *of $\mathfrak{A}$.*

Proof. Let $X = \{x_1 ..., x_n\}$. Consider maximal chains in $\mathfrak{A} = 2^X$ of the type $\Gamma = \{B_0, B_1, ..., B_n\}$, $\varnothing = B_0 \subset B_1 \subset ... \subset B_n = X$, $|B_i \setminus B_{i-1}| = 1$, $i = 1,...,n$. It is clear that such chains are lattices and every monotone measure

[1] An arbitrary covering $\mathfrak{C}$ of $\mathfrak{A}$ is a family of non-empty subsets of $\mathfrak{A}$ such that $\bigcup_{\mathfrak{a}\in\mathfrak{C}} \mathfrak{a} = \mathfrak{A}$.

is additive on them, i.e., we get the required covering that consists of all these lattices.■

We denote by $\mathfrak{S}_\mu$ the covering of $\mathfrak{A}$ that consists of all maximal lattices, on which $\mu \in M_{2-mon}$ is additive. For example, if μ is a probability measure, then the covering is a singleton, which contains only one element $\mathfrak{A}$. If a $\mu \in M_{2-mon}$ is such that $\mu(A)+\mu(B)<\mu(A\cup B)+\mu(A\cap B)$ for any $A,B\in\mathfrak{A}$ with $A\not\subseteq B$ and $B\not\subseteq A$ then $\mathfrak{S}_\mu$ obviously consists of all maximal chains in $\mathfrak{A}$. It is important to emphasize that any $\Lambda\in\mathfrak{S}_\mu$ has to contain $\varnothing$ and X, since these sets are additive elements for any $\mu\in M_{2-mon}$.

Another convenient characterization of 2-monotone measures is recalled in

Remark 1. For any $\mu\in M_{2-mon}$, define the convex set $core(\mu)$ of probability measures, defined by $core(\mu)=\{P\in M_{pr} \mid P\geq\mu\}$. It is well-known that this set is non-empty and usually called the core of μ. Moreover, it is possible to describe all extreme points of this set [10]. To do this, we should consider all maximal chains of the algebra 2^X on $X=\{x_1,x_2,...,x_n\}$. Then any extreme point P_γ is generated by a maximal chain $\gamma=\{B_0,B_1,...,B_n\}$, where $\varnothing=B_0\subset B_1\subset...\subset B_n=X$ and $B_k=\{x_{i_1},...,x_{i_k}\}$, $k=1,...,n$, as $P_\gamma(\{x_{i_k}\})=P_\gamma(B_k\setminus B_{k-1})=\mu(B_k)-\mu(B_{k-1})$, i.e. P_γ is chosen such that $P_\gamma(B)=\mu(B)$ for all $B\in\gamma$.

Lemma 3. *Any lattice in $\mathfrak{S}_\mu$ contains a maximal chain.*

Proof. Consider an arbitrary lattice $\Lambda\subseteq 2^X$, on which μ is additive. Let Γ be a sequence of sets with the following properties: 1) a minimal algebra that contains Γ coincides with 2^X; 2) first elements of Γ are all elements of Λ. Then the limit measure[2] μ_Γ, in the canonical sequence constructed by Γ is a probability measure, and also $\mu_\Gamma(A)=\mu(A)$ for all $A\in\Lambda$. Since any such sequence Γ is equivalent to some maximal chain $\gamma\subseteq 2^X$, we get $\mu_\Gamma(A)=\mu(A)$ for all $A\in\gamma$. Consider a lattice, on which μ and μ_Γ have the same values. It is clear that this lattice contains Λ and γ, and also μ is additive on it. It means that any lattice in $\mathfrak{S}_\mu$ contains a maximal chain. ■

[2] The explanation of terms: "limit measure", "canonical sequence of monotone measures", … are given in Appendix.

Proposition 1. *There is the one-to-one correspondence between maximal lattices in $\mathfrak{S}_\mu$ and extreme points of $core(\mu)$ for every $P\in core(\mu)$ defined by $\Lambda=\{A\in\mathfrak{A} \mid P(A)=\mu(A)\}$, where $\Lambda\in\mathfrak{S}_\mu$.*

Proof. Because any lattice $\Lambda\in\mathfrak{S}_\mu$ contains a maximal chain $\gamma\subseteq\Lambda$, we can define that P_γ corresponds to Λ. Using canonical sequences of 2-monotone measures, it easy to prove that $P_\gamma(B)=\mu(B)$ for all $B\in\Lambda$. This proves that if Λ contains two different maximal chains, then they generate the same probability measure, i.e. we show that such a construction generates the unique probability measure P_γ, where $\gamma\subseteq\Lambda$, with $P_\gamma(B)=\mu(B)$ for all $B\in\Lambda$. We finish the proof of the proposition by showing that for any maximal chain γ the set $\{B\in 2^X \mid P_\gamma(B)=\mu(B)\}\in\mathfrak{S}_\mu$. It is easy to check that this set is a lattice. Let $P_\gamma(A)=\mu(A)$ and $P_\gamma(B)=\mu(B)$ for some $A,B\in 2^X$. Then we have to prove that also $P_\gamma(A\cap B)=\mu(A\cap B)$ and $P_\gamma(A\cup B)=\mu(A\cup B)$. The above condition implies that

$$\mu(A)+\mu(B)\leq\mu(A\cap B)+\mu(A\cup B)\leq$$
$$P_\gamma(A\cap B)+P_\gamma(A\cup B)=P_\gamma(A)+P_\gamma(B)=\mu(A)+\mu(B),$$

i.e. $\mu(A)+\mu(B)=\mu(A\cap B)+\mu(A\cup B)$, $P_\gamma(A\cap B)=\mu(A\cap B)$ and $P_\gamma(A\cup B)=\mu(A\cup B)$. Using again canonical sequences of 2-monotone measures, it is easy to prove that such a lattice is maximal, i.e. we have the one-to-one correspondence between maximal lattices in $\mathfrak{S}_\mu$ and extreme points in $core(\mu)$.■

Proposition 2. *Let $\mu\in M_{coh}$, $\nu\in M_{2-mon\leq\mu}$, $S_{\nu=\mu}=\{A\in\mathfrak{A} \mid \nu(A)=\mu(A)\}$, $S_{\nu=0}=\{A\in\mathfrak{A} \mid \nu(A)=0\}$. Then ν is an extreme point of $M_{2-mon\leq\mu}$ iff its values are defined by the sets $S_{\nu=\mu}$, $S_{\nu=0}$, $\mathfrak{S}_\nu$ uniquely.*

Proof. A set function ν is in $M_{2-mon\leq\mu}$ iff it satisfies the following conditions:

1) $\nu(\varnothing)=0$, $\nu(X)=1$;

2) $\nu(A)\geq 0$ for all $A\in A$;

3) $\nu(A)\leq\nu(B)$ if $A\subseteq B$;

4) $\nu(A)+\nu(B)\leq\nu(A\cap B)+\nu(A\cup B)$ for all $A,B\in A$;

5) $\nu(A)\leq\mu(A)$ for all $A\in 2^X$.

These conditions can be considered as a system of linear inequalities on values $\nu(A)$, $A\in 2^X$. From the theory of linear inequalities, we know that any extreme point can be calculated by solving linear equalities, obtained by the subset of inequalities if we change "$\leq$" to "$=$". Show that we can confine ourselves to using equalities that are generated by 2), 4), and 5). It is not necessary to use 1) because $\nu(\varnothing)=\mu(\varnothing)=0$, $\nu(X)=\mu(X)=1$. We show further that any equality $\nu(C)=\nu(D)$ for $C\subset D$ ($C\neq D$), generated by 3), can be derived from the additivity of ν. In this case, we take $A=C$, $B=D\setminus C$. Then $A\cap B=\varnothing$, $\nu(B)=0$, $\nu(A)+\nu(B)=\nu(A\cap B)+\nu(A\cup B)$, and the last equality, $\nu(B)=0$, $\nu(A\cap B)=0$ implies that $\nu(C)=\nu(D)$. Therefore, we conclude that the proposition is true. ■

Consider some corollaries from Propositions 1 and 2:

Corollary 1. *Let the notation of Proposition 2 be used. Then ν is an extreme point of $M_{2-mon\le\mu}$ if for any $\Lambda\in\mathfrak{S}_\nu$ a probability measure P_Λ with $P_\Lambda(A)=\mu(A)$ for all $A\in S_{\nu=\mu}\cap\Lambda$ and $P_\Lambda(A)=0$ for all $A\in S_{\nu=0}\cap\Lambda$ is defined uniquely.*

Proof. It is easy to see that Corollary 1 is a direct consequence of Propositions 1 and 2. ■

Corollary 2. *Let notations of Proposition 2 be used. Then ν is an extreme point of $M_{2-mon\le\mu}$ if for any $\Lambda\in\mathfrak{S}_\nu$ the set $\Lambda\cap\left(S_{\nu=\mu}\cup S_{\nu=0}\right)$ contains a maximal chain. In addition, $\nu=\wedge_\gamma P_\gamma$, where the minimum in the right side of the last formula is taken over all possible probability measures P_γ, defined for each maximal chain $\gamma\subseteq S_{\nu=\mu}\cup S_{\nu=0}$ by $P_\gamma(A)=\mu(A)$ for $A\in\gamma\cap S_{\nu=\mu}$ and $P_\gamma(A)=0$ for $A\in\gamma\cap S_{\nu=0}$. Moreover, if $S_{\nu=0}\setminus S_{\nu=\mu}=\varnothing$, then such a ν is Pareto optimal.*

Proof. Because P_Λ is defined uniquely if $\Lambda\cap\left(S_{\nu=\mu}\cup S_{\nu=0}\right)$ contains a maximal chain, we conclude that ν is an extreme point by Corollary 1. The formula $\nu=\wedge_\gamma P_\gamma$ is also true, since 2-monotonicity of ν implies that $P_\gamma\ge\nu$ for any $\gamma\subseteq S_{\nu=\mu}\cup S_{\nu=0}$. Observe also that, for any $\nu'\in M_{2-mon\le\mu}$ with $S_{\nu'=\mu}=S_{\nu=\mu}$ and $S_{\nu'=0}=S_{\nu=0}$, we have $\nu'\le\nu$, i.e. ν have the largest values for the fixed $S_{\nu=\mu}$ and $S_{\nu=0}$. Show that ν is Pareto optimal if $S_{\nu=0}\setminus S_{\nu=\mu}=\varnothing$. Suppose on the contrary that there is another $\nu'\in M_{2-mon\le\mu}$ such that $\nu'>\nu$. Then we should conclude that $S_{\nu=\mu}\subseteq S_{\nu'=\mu}$ and $S_{\nu=\mu}\neq S_{\nu'=\mu}$. We see that $\nu'\le\wedge_{\gamma\subseteq S_{\nu'=\mu}}P_\gamma\le\wedge_{\gamma\subseteq S_{\nu=\mu}}P=\nu$, and such a ν' does not exist, i.e. the corollary is proved in the whole. ■

Pareto optimal extreme points, described in Corollary 2, have desirable properties. They are uniquely defined by a chosen set $S_{\nu=\mu}$ and their values can be easily computed using explicit formulas. Therefore, it is desirable to study the conditions of existence of these extreme points, and to construct the algorithm for finding such sets $S_{\nu=\mu}$.

We see from Proposition 1 that any extreme point of $M_{2-mon\le\mu}$ is characterized by $S_{\nu=\mu}$, $S_{\nu=0}$, $\mathfrak{S}_\nu$. But we know that an arbitrary extreme point is not necessarily Pareto optimal. To investigate this situation, introduce so called elementary lattices in 2^X of two types. An elementary lattice Λ of the first type is given by $\Lambda=\left\{A,A\cup\{x_i\}\right\}$, where $A\in 2^X$ and $x_i\notin A$, and an elementary lattice of the second type by $\Lambda=\left\{A,A\cup\{x_i\},A\cup\{x_j\},A\cup\{x_i\}\cup\{x_j\}\right\}$, where $A\in 2^X$ and $x_i,x_j\notin A$. Using the above definition we can formulate the following necessary and sufficient feature of 2-monotonicity [7, 12].

Proposition 3. *A set function $\mu:2^X\to[0,1]$ is a 2-monotone measure iff*

1) $\mu(\varnothing)=0$, $\mu(X)=1$;

2) μ is monotone on all possible lattices in 2^X of the first type;

3) μ is 2-monotone on all possible lattices in 2^X of the second type.

Remark 2. Proposition 3 can be reformulated in the following simple way:

A set function $\mu:2^X\to[0,1]$ is a 2-monotone measure iff

1) $\mu(\varnothing)=0$, $\mu(X)=1$;

2) $\mu(A)\le\mu\left(A\cup\{x_i\}\right)$ for all possible $A\in 2^X$ and $x_i\notin A$;

3) $\mu\left(A\cup\{x_i\}\right)+\mu\left(A\cup\{x_j\}\right)\le\mu(A)+\mu\left(A\cup\{x_i\}\cup\{x_j\}\right)$ for all possible $A\in 2^X$ and $x_i,x_j\notin A$.

However, the consideration of elementary lattices is useful for characterizing Pareto optimal 2-monotone measures.

Proposition 4. *Let $\nu\in M_{2-mon\le\mu}$, $\mathcal{L}_1$ be the set of all elementary lattices of the first type on which ν is constant, and $\mathcal{L}_2$ be the set of all elementary lattices of the second type, on which ν is additive. Then ν is not Pareto optimal iff there is a non-identical zero, non-negative set function $\Delta\nu:2^X\to\mathbb{R}_+$ such that*

1) $\Delta\nu(A)=0$ if $A\in S_{\nu=\mu}$;

2) $\Delta\nu$ is monotone on all lattices in $\mathcal{L}_1$;

3) $\Delta\nu$ is 2-monotone on all lattices in $\mathcal{L}_2$.

Proof. *Necessity.* Let ν be not Pareto optimal. Then there is a $\nu' \in M_{2-mon\le\mu}$ such that $\nu' > \nu$. It is easy to check that $\Delta\nu = \nu' - \nu$ obeys all required properties.

Sufficiency. Let such a set function $\Delta\nu$ exist. Consider the following positive numbers:

$$\varepsilon_1 = \max\left\{h\left(\mu(A) - \nu(A)\right) \middle| A \in 2^X\right\},$$

$$\varepsilon_2 = \max\left\{h\left(\nu\left(A \cup \{x_i\}\right) - \nu(A)\right) \middle| A \in 2^X, x_i \notin A\right\},$$

$$\varepsilon_3 = \max\left\{w\left(A, x_i, x_j\right) \middle| A \in 2^X, x_i, x_j \notin A\right\},$$

where $h(t) = t$ if $t > 0$ and $t = 1$ else; $w\left(A, x_i, x_j\right) = h\left(\nu\left(A\right) + \nu\left(A \cup \{x_i\} \cup \{x_j\}\right) - \nu\left(A \cup \{x_i\}\right) - \nu\left(A \cup \{x_j\}\right)\right)$. Then choosing $\Delta\nu$ such that $\max\left\{\Delta\nu(A) \mid A \in 2^X\right\} \le \varepsilon$, where $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, 0.5\varepsilon_3,\}$, we get that the set function $\nu' = \nu + \Delta\nu$ is in $M_{2-mon\le\mu}$ and obviously $\nu' > \nu$, i.e. ν is not Pareto optimal. ■

5. Algorithms for searching Pareto optimal 2-monotone measures

In this section we present two algorithms. The first one improves a given approximation (two-monotone probability) to a Pareto-optimal approximation, the second one places the choice of a certain Pareto-optimal approximation on a certain linear imprecision index as an objective function.

Algorithm I

Input data: coherent lower probability μ on 2^X.

First step. Finding a 2-monotone measure ν_0 with $\nu_0 \le \mu$.

Second step. Finding a Pareto optimal 2-monotone measure ν with $\nu_0 \le \nu \le \mu$.

The first step can be based on different approaches. For example, we can choose as ν_0 an arbitrary chain measure, generated by some maximal chain Γ of algebra 2^X. Then $\nu_0(B) = \sup_{A\in\Gamma|A\subseteq B} \mu(A)$ for all $B \in 2^X$. However, it is clear that the realization of the second step of the algorithm can be produced more effectively if the values ν_0 are close to the values of μ. In this sense, the following procedure is better than the first one.

1) Compute an auxiliary 2-monotone set function g on 2^X using the following formulas:

a) $g(A) = \mu(A)$ for all $A \in 2^X$ with $|A| \le 1$;

b) Let us compute all values of g on sets with cardinality less or equal to k. Then values of g on sets A with cardinality that is equal to $k+1$ are computed by

$$g(A) = \max\left\{\mu(A), \max_{x_i, x_j \in A} g\left(A \setminus \{x_i\}\right) + g\left(A \setminus \{x_j\}\right) - g\left(A \setminus \{x_i, x_j\}\right)\right\}.$$

Observe that in the last formula $g\left(A \setminus \{x_i\}\right) + g\left(A \setminus \{x_j\}\right) - g\left(A \setminus \{x_i, x_j\}\right) = g\left(A \setminus \{x_i\}\right)$ for $i = j$. Therefore, g is 2-monotone by Proposition 3. It is easy to see that $g \ge \mu$ and $g = \mu$ iff μ is 2-monotone, and also it is not necessarily $g(X) = 1$.

2) A 2-monotone measure $\nu_0 = \varphi \circ g$ is computed using a convex distortion function $\varphi : \left[0, g(X)\right] \to [0,1]$ that has to obey the following properties:

(i) $\varphi(0) = 0$, $\varphi(g(X)) = 1$;

(ii) $\varphi(g(A)) \le \mu(A)$ for all $A \in 2^X$.

According to, e.g., [14] ν_0 has to be also 2-monotone, i.e. $\nu_0 \in M_{2-mon} \le \mu$. The search of the mapping φ is also connected with solving the system of linear inequalities. It is clear that it is sufficient to know the values of φ only in the points in the set $Y = \left\{g(A) \mid A \in 2^X\right\}$. Let $Y = \left\{y_i\right\}_{i=0}^m$, where $0 = y_0 < y_1 < ... < y_m = g(X)$. Then the condition (ii) is transformed to $\varphi\left(y_i\right) \le \psi\left(y_i\right)$, where $\psi\left(y_i\right)$, $i = 1,...,m-1$, are corresponding values of μ in (ii), and convexity of φ means that $\varphi\left(y_i\right) \le \varphi\left(y_{i+1}\right)$, $i = 0,...,m-1$, and

$$\frac{\varphi\left(y_{i+1}\right) - \varphi\left(y_i\right)}{y_{i+1} - y_i} \le \frac{\varphi\left(y_i\right) - \varphi\left(y_{i-1}\right)}{y_i - y_{i-1}},\ i = 1,...,m.$$

Clearly the problem of searching φ is simpler than the initial problem, and we should try to choose φ with the largest values.

The second step can be performed iteratively by using procedures that are similar to the usual simplex method. Consider an algorithm that seems to be easily realizable and computationally effective. Let $\nu_k \in M_{2-mon\le\mu}$, and the following values

$$\Delta_1 = \mu(A) - \nu_k(A),$$

$$\Delta_2 = \min_{x_i \in X\setminus A}\left(\nu_k\left(A \cup \{x_i\}\right) - \nu_k(A)\right),$$

$$\Delta_3 = \min_{x_i \in X\setminus A, x_j \in A}\left(\nu_k\left(A \cup \{x_i\}\right) - \nu_k(A) - \nu_k\left(\left(A \setminus \{x_j\}\right) \cup \{x_i\}\right) + \nu_k\left(A \setminus \{x_j\}\right)\right)$$

are positive for a given $A \in 2^X$. Then, by Proposition 3, we can increase values of ν_k on the set A without any changes on other sets, and get a measure $\nu_{k+1} \in M_{2-mon\le\mu}$

by the rule $\nu_{k+1}(B)=\nu_k(B)+d$ if $B=A$ and $\nu_{k+1}(B)=\nu_k(B)$ otherwise, where $d=\min\{\Delta_1,\Delta_2,\Delta_3\}$. Thus, we can increase values by this rule until $d=0$ for any $A\in 2^X$. It easy to show that this procedure converges to a Pareto optimal 2-monotone measure after a finite number of iterations due to simplex method. Show that a measure ν_k is Pareto optimal if $d=0$ for any $A\in 2^X$. In this case, we have to show that a convex set $M=\{\nu\in M_{2-mon}\,|\,\nu_k\le\nu\le\mu\}$ is a singleton, i.e. $M=\{\nu_k\}$. Observe that values of ν_k can be considered as basic variables and the above condition ($d=0$ for any $A\in 2^X$) means that we cannot change them, i.e. the convex set M contains the only one extreme point ν_k, i.e. ν_k is Pareto optimal. Analogously, any iteration of the proposed procedure can be considered as an iterative step of the simplex method. This means that this procedure converges by a finite number of iterations.

Algorithm II. It is based on the usual application of simplex method. As a criterion a linear imprecision index can be used. By definition [8], a linear imprecision index f is a non-negative functional on M_{low} that satisfies the following properties:

1) $f(P)=0$ for any $P\in M_{pr}$;

2) $f\left(\eta_{\langle X\rangle}\right)=1$, where $\eta_{\langle X\rangle}$ describes the situation of complete ignorance, i.e. $\eta_{\langle X\rangle}(A)=1$ if $A=X$, $\eta_{\langle X\rangle}(A)=0$ otherwise;

3) $f(\nu_1)\le f(\nu_2)$ for any $\nu_1,\nu_2\in M_{low}$ such that $\nu_1\ge\nu_2$;

4) $f(a\nu_1+(1-a)\nu_2)=af(\nu_1)+(1-a)f(\nu_2)$ for arbitrary $a\in[0,1]$ and $\nu_1,\nu_2\in M_{low}$.

The notable examples of such imprecision indices are the generalized Hartley measure [19] defined by

$$GH(\nu)=\frac{1}{\ln|X|}\sum_{A\in 2^X} m(A)\ln|A|,$$

where m is the Möbius transform [10] of the given $\nu\in M_{low}$, and an index f_{L_1} based on L_1 distance defined by

$$f_{L_1}(\nu)=\frac{1}{2^{|X|}-2}\sum_{A\in 2^X}\left|\bar{\nu}(A)-\nu(A)\right|,$$

where $\bar{\nu}$ is the dual of ν, i.e. $\bar{\nu}(A)=1-\nu(A^c)$. Notice that linear imprecision indices are linear functions w.r.t. values of a given $\nu\in M_{low}$. In particular, since $\bar{\nu}\ge\nu$ for any $\nu\in M_{low}$, we get

$$f_{L_1}(\nu)=\frac{1}{2^{|X|}-2}\sum_{A\in 2^X}\left(1-\nu(A)-\nu(A^c)\right)=$$
$$1-\frac{1}{2^{|X|-1}-1}\sum_{A\in 2^X\setminus\{\varnothing,X\}}\nu(A).$$

Notice that we can use also as a linear functional the L_1 distance between μ and its approximation ν, i.e. in this case

$$f(\nu)=\sum_{A\in 2^X}|\mu(A)-\nu(A)|.$$

Because $\nu\le\mu$, we obtain

$$f(\nu)=\sum_{A\in 2^X\setminus\{\varnothing,X\}}\mu(A)-\sum_{A\in 2^X\setminus\{\varnothing,X\}}\nu(A),$$

i.e. the criterion based on this metric is equivalent to the criterion f_{L_1}.

Therefore, the choice of Pareto optimal 2-monotone measure, based on a linear inclusion index, can be conceived as a linear programming problem, where we have a system of inequalities that describe a convex set $M_{2-mon\le\mu}$ and a linear criterion f.

6. Examples of the proposed algorithms working

To illustrate our method, we use examples of coherent lower probabilities from [6].

Example 1. Let $X=\{x_1,x_2,x_3,x_4\}$ and let $\mu\in M_{coh}$ be defined on 2^X by $\mu(A)=\min\{P_1(A),P_2(A)\}$, where $A\in 2^X$ and $P_1,P_2\in M_{pr}$ are defined through their values on singletons by $P_1(\{x_1\})=1/4$; $P_1(\{x_2\})=0$, $P_1(\{x_3\})=3/4$; $P_1(\{x_4\})=0$; $P_2(\{x_1\})=0$; $P_2(\{x_1\})=1/2$, $P_2(\{x_3\})=0$; $P_2(\{x_4\})=1/2$. The values of μ are given in Table 1. It is clear that $\mu\notin M_{2-mon}$, because, for example, $\mu(A)+\mu(B)>\mu(A\cup B)$ for $A=\{x_1,x_2\}$, $B=\{x_2,x_3\}$. Following the first step of Algorithm 1, we get an auxiliary 2-monotone set function g on 2^X with values also shown in Table 1. Then we need to find a convex distortion function φ, that is lower than function ψ (see Fig. 1). The found distortion function is also shown in Fig. 1 and can be given by the formula

$$\varphi(x)=\begin{cases}0.5x, & x\in[0,0.5],\\ 0.75x-0.125, & x\in(0.5,1.5].\end{cases}$$

It easy to check that ν_0 is not Pareto optimal in this case, because, for example, $d=1/8$ for the set $A=\{x_1,x_2,x_3\}$ and according to Algorithm 1, we obtain the next approximation $\nu_1\in M_{2-mon\le\mu}$ by the rule

$\nu_1(B) = \nu_0(B) + d$ if $B = A$ and $\nu_1(B) = \nu_0(B)$ otherwise. Producing in such a way iterations for sets $\{x_1, x_3, x_4\}$, $\{x_2, x_3, x_4\}$, $\{x_2, x_3\}$, $\{x_3, x_4\}$, we obtain a Pareto optimal measure $\nu \in M_{2-mon \le \mu}$ with values given in Table 1.

x_1	x_2	x_3	x_4	μ	g	ν_0	ν
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
1	1	0	0	1/4	1/4	1/8	1/8
0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0
0	1	1	0	1/2	1/2	1/4	3/8
1	1	1	0	1/2	3/4	7/16	½
0	0	0	1	0	0	0	0
1	0	0	1	1/4	1/4	1/8	1/8
0	1	0	1	0	0	0	0
1	1	0	1	1/4	1/2	1/4	¼
0	0	1	1	1/2	1/2	1/4	3/8
1	0	1	1	1/2	3/4	7/16	½
0	1	1	1	3/4	1	5/8	¾
1	1	1	1	1	3/2	1	1

Table 1. Results for Example 1.

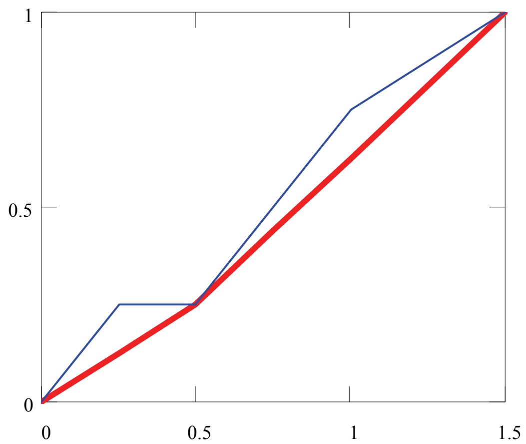

Figure 1: The distortion function for Example 1: φ - red line; ψ - blue line.

Example 2. Let $X = \{x_1, x_2, x_3, x_4\}$ and let $\mu \in M_{coh}$ have the values given in Table 2. We see that $\mu \notin M_{2-mon}$, since $\mu(A) + \mu(B) > \mu(A \cap B) + \mu(A \cup B)$ for $A = \{x_1, x_4\}$, $B = \{x_2, x_4\}$. Then, following the steps of Algorithm 1, we can get results that are shown in Table 2 and Fig. 2. The distortion function for this case can be defined by the formula

$$\varphi(x) = \begin{cases} 0.5x, & x \in [0, 2/3], \\ x - 1/3, & x \in (2/3, 4/3]. \end{cases}$$

x_1	x_2	x_3	x_4	μ	g	ν_0	ν_1
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0
0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0
0	1	1	0	0	0	0	0
1	1	1	0	2/3	2/3	1/3	2/3
0	0	0	1	0	0	0	0
1	0	0	1	1/3	1/3	1/6	1/6
0	1	0	1	1/3	1/3	1/6	1/6
1	1	0	1	1/3	2/3	1/3	1/3
0	0	1	1	1/3	1/3	1/6	1/6
1	0	1	1	1/3	2/3	1/3	1/3
0	1	1	1	1/3	2/3	1/3	1/3
1	1	1	1	1	4/3	1	1

Table 2. Results for Example 2.

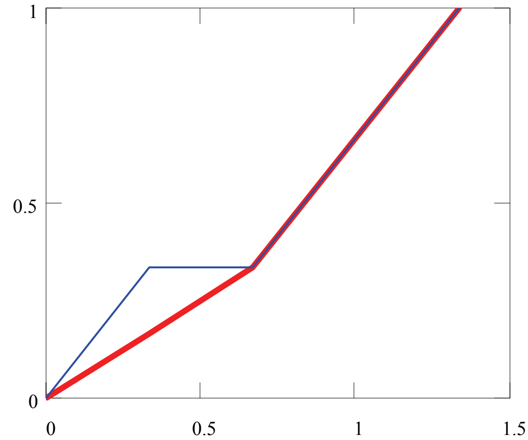

Figure 2: The distortion function for Example 2: φ - red line; ψ - blue line.

It is easy to check that ν_0 is not Pareto optimal in this case, because $d = 1/3$ for set $A = \{x_1, x_2, x_3\}$, and according to Algorithm 1, we obtain a Pareto optimal measure $\nu_1 \in M_{2-mon \le \mu}$ by the rule $\nu_1(B) = \nu_0(B) + d$ if $B = A$ and $\nu_1(B) = \nu_0(B)$ otherwise.

Notice that we can indeed apply the proposed algorithms to any monotone measure, i.e. μ need not be a coherent lower probability. This case is considered in the next example.

Example 3. Let $X = \{x_1, x_2, x_3, x_4\}$ and let $\mu \in M_{mon}$ be defined by $\mu(A) = 1$ if $|A| \geq 1$ and $\mu(A) = 0$. In this case the set of all Pareto optimal 2-monotone measures coincides with the set of all probability measures on 2^X, and, by Algorithm 1, we obtain a probability measure $\nu = \nu_0$ defined by $\nu(\{x_i\}) = 1/4$, where $i = 1,...,4$.

7. Concluding remarks

We have characterized and computed Pareto optimal outer approximations of coherent lower probabilities by 2-monotone measures. Further research includes obviously the study of the sensitivity of the results with respect to the choice of the approximation.

Also a closer investigation of some modifications of the algorithms is certainly rewarding, in particular in the following directions.

Because in principle the solution of the optimization problem is computationally very hard for large $n = |X|$, it is possible to solve it for some subalgebra $\mathfrak{B} \subseteq 2^X$. Let ν be Pareto optimal on $\mathfrak{B}$ for some μ on 2^X, then its inner extension $\underline{\nu}$ on 2^X defined by $\underline{\nu}(B) = \sup_{A \in \mathfrak{B} | A \subseteq B} \nu(A)$, $B \in 2^X$, is 2-monotone [14], and can be considered as an approximation of a Pareto optimal measure. The same approach can be used for a general infinite algebra $\mathfrak{A}$.

In light of the intended application to statistical hypotheses testing, it will also be interesting to replace the linear imprecision index in the objective function by the Kullback-Leibler distance, which has some close relation to the likelihood ratio underlying optimal hypotheses testing.

Notice that a Pareto optimal measure is not uniquely defined even in a case when we use a linear imprecision index in the linear programming problem. To get uniqueness, it seems to be possible to use the following approach: Let $\mathfrak{A} = 2^X$, where $|X| = n$, we have a linear order on $\mathfrak{A}$ defined by indexing its elements, i.e. $\mathfrak{A} = \{B_i\}_{i=1}^{2^n}$ and B_i is more preferable than B_j if $i < j$. Then we say that $\nu_1 \in M_{2-mon \leq \mu}$ is more preferable than $\nu_2 \in M_{2-mon \leq \mu}$ if there is an index k such that $\nu_1(B_i) = \nu_2(B_i)$ for $i = 1,...,k-1$, and $\nu_1(B_k) > \nu_2(B_k)$.

Another rewarding issue has been raised by one of the referees, looking at the so-to-say inverse problem: can every Pareto-optimal solution be obtained from a certain imprecision index? Irrespective of whether the answer is affirmative or not, in any way that would give a vivid natural characterization and classification of the Pareto optimal solutions.

Appendix: Canonical sequences of monotone measures: main results

Here we give a brief overview on results concerning canonical sequence of monotone measures. The detailed description with proofs can be found in [5].

Let μ_0 be a monotone measure on $\mathfrak{A}$, $\Gamma = \{B_k\}_{k=1}^{\infty}$ a sequence of sets in $\mathfrak{A}$. Then a sequence of monotone measures $\{\mu_k\}_{k=0}^{\infty}$, defined as

$$\mu_k(A) = \mu_{k-1}(A \cup B_k) - \mu_{k-1}(B_k) + \mu_{k-1}(A \cap B_k),$$

is called a *canonical sequence* of monotone measures, generated by Γ. It is easy to see that if μ_0 is 2-monotone, then the sequence $\{\mu_k\}_{k=0}^{\infty}$ is increasing, i.e. $\mu_0 \leq \mu_1 \leq ...$, and there is a limit $\mu_\Gamma(A) = \lim_{k \to \infty} \mu_k(A)$ for all $A \in \mathfrak{A}$, and $\mu_\Gamma \in M_{2-mon}$. If μ_0 is 2-alternating (submodular), the sequence $\{\mu_k\}_{k=0}^{\infty}$ is decreasing, i.e. $\mu_0 \geq \mu_1 \geq ...$, and the limit measure $\mu_\Gamma(A) = \lim_{k \to \infty} \mu_k(A)$, $A \in \mathfrak{A}$, is also 2-alternating. For our purpose, it is sufficient to consider the finite case where $\mathfrak{A} = 2^X$, $\Gamma = \{B_k\}_{k=1}^{m}$, and $\mu_\Gamma = \mu_m$.

Two sequences $\Gamma_1 = \{B_k\}_{k=1}^{n}$ and $\Gamma_2 = \{C_k\}_{k=1}^{m}$ in $\mathfrak{A}$ are called to be *equivalent* ($\Gamma_1 \sim \Gamma_2$) iff $\mu_{\Gamma_1} = \mu_{\Gamma_2}$ for any generating monotone measure μ_0.

Theorem 1. *Let $\Gamma_A = \{A_k\}_{k=1}^{n} \subseteq \mathfrak{A}$. Then there is an increasing sequence of sets $\Gamma_B = \{B_k\}_{k=1}^{m} \subseteq \mathfrak{A}$, $B_1 \subseteq B_2 \subseteq ... \subseteq B_m$, such that $\Gamma_A \sim \Gamma_B$. Minimal algebras $\mathfrak{A}_A$ and $\mathfrak{A}_B$, generated by Γ_A and Γ_B respectively, coincide, i.e. $\mathfrak{A}_A = \mathfrak{A}_B$.*

Let $\mu \in M_{mon}$ be a monotone measure on $\mathfrak{A}$. We call a set $B \in \mathfrak{A}$ an *additive element* w.r.t. μ iff $\mu(A) = \mu(A \cup B) - \mu(B) + \mu(A \cap B)$ for all $A \in \mathfrak{A}$. It is easy to check that $\varnothing, X$ are additive elements w.r.t. any $\mu \in M_{mon}$ and the set of all additive elements w.r.t. a monotone measure μ is an algebra.

Theorem 2. *Let $\{\mu_n\}_{n=0}^{\infty}$ be a canonical sequence of monotone measures, generated by $\{B_n\}_{n=1}^{\infty} \subseteq \mathfrak{A}$. Denote by $\mathfrak{M}_n$ the algebra, consisting of all additive elements w.r.t. μ_n. Then*

1) $\mathfrak{M}_0 \subseteq \mathfrak{M}_1 \subseteq ... \subseteq \mathfrak{M}_n \subseteq ....$;

2) μ_n is additive on $\mathfrak{M}_n$;

3) if $C \in \mathfrak{M}_n$ *, then* $\mu_n(C) = \mu_k(C)$ *for* $k \geq n$ *;*

4) $\{B_1, B_2, .., B_n\} \subseteq \mathfrak{M}_n$.

Notice that the above results imply several important consequences, which are used in this paper. In particular, let $X = \{x_1, ..., x_n\}$. $\mathfrak{A} = 2^X$, $\mu_0 \in M_{2-mon}$. Consider a canonical sequence of 2-monotone measures, generated by $\Gamma_A = \{A_k\}_{k=1}^m \subseteq \mathfrak{A}$, assuming that the minimal algebra containing Γ_A coincides with $\mathfrak{A}$. Then, by Theorem 2, $\mu_\Gamma \geq \mu$, μ_Γ is additive on $\mathfrak{A}$, i.e. μ_Γ is a probability measure, and by Theorem 1, there is a maximal chain $\Gamma_B = \{B_k\}_{k=0}^n \subseteq \mathfrak{A}$ such that $\varnothing = B_0 \subset B_1 \subset ... \subset B_n = X$, $|B_k \setminus B_{k-1}| = 1$, $k = 1,..,n$; μ_Γ is uniquely defined by $\mu_\Gamma(B_k) = \mu_0(B_k)$, $k = 1,..,n$.

Acknowledgement. We are very grateful to the referees for many very helpful and detailed remarks.

Andrey Bronevich from his side expresses his sincere thanks to the German Academic Exchange Service (DAAD), Russian Ministry of Education, Ludwig-Maximilians University and Prof. Dr. Thomas Augustin for the research opportunity provided.

References

[1] T. Augustin. *Optimale Tests bei Intervallwahrscheinlichkeit.* Vandenhoeck & Ruprecht, Göttingen, 1998.

[2] T. Augustin. On data-based checking of hypotheses in the presence of uncertain knowledge. In: Gaul, W., Locarek-Junge, H. (eds.). Classification in the Information Age. Springer. Heidelberg, 1999, pp. 127 – 135.

[3] T. Augustin. Neyman–Pearson testing under interval probability by globally least favorable pairs. Reviewing Huber–Strassen theory and extending it to general interval probability. *Journal of Statistical Planning and Inference* 105: 149 – 173, 2002.

[4] T. Augustin and R. Hable. On the impact of robust statistics on imprecise probability models: a review. To appear in: *Proc. of the the 10th International Conference on Structural Safety and Reliability*, Osaka, 2009.

[5] A.G. Bronevich. Canonical sequences of fuzzy measures. In Proc. of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-2004), Perugia-Italy, 2004, 8 pp.

[6] A.G. Bronevich. An investigation of ideals in the set of fuzzy measures. *Fuzzy Sets and Systems* 152: 271 – 288, 2005.

[7] A.G. Bronevich. On the closure of families of fuzzy measures under eventwise aggregations. *Fuzzy Sets and Systems* 153: 45 – 70, 2005.

[8] A.G. Bronevich and A.E. Lepskiy. Measuring uncertainty with imprecision indices. In de Cooman, G., Vejnarova. J., Zaffalon, M. (eds.). Proc. of the Fifth International Symposium on imprecise probability: Theory and Applications, Prague, Czech Republic, 2007, pp. 47 – 56.

[9] A. Buja. On the Huber-Strassen theorem. *Probability Theory and Related Fields* 73: 149 – 152, 1986.

[10] A. Chateauneuf and J.Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences* 17: 263 – 283, 1989.

[11] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, 5: 131 – 295, 1954.

[12] V.I. Danilov. *Lectures on Game Theory*. Russian Economical School, Moscow, 2002. (In Russian)

[13] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics* 37: 325 – 339, 1967.

[14] Denneberg D. *Non-additive Measure and Integral*. Dordrecht, Kluwer, 1997.

[15] I. Gilboa and D. Schmeidler. Updating ambiguous beliefs. *Journal of Economic Theory* 59: 33 – 49, 1993.

[16] R. Hable. Data-based decisions under imprecise probability and least favorable models. *International Journal of Approximate Reasoning* 50: 642 – 654, 2009.

[17] R. Hafner. Konstruktion robuster Teststatistiken. In: Schach, S., Trenkler, G. (eds.). Data Analysis and Statistical Inference. Festschrift in Honour of Prof. Dr. Friedhelm Eicker. Eul. Bergisch Gladbach, 1992, pp. 145-160.

[18] P.J. Huber, V. Strassen. Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* 1: 251–263, 1973.

[19] G. J. Klir. *Uncertainty and Information: Foundations of Generalized Information Theory*, Hoboken, NJ: Wiley-Interscience, 2006.

[20] L.V. Utkin and T. Augustin. Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In Cozman, F.G., Nau, R., Seidenfeld, T. (eds.). Proc. of the Fourth International Symposium on Imprecise Probabilities and their Applications, Pittsburgh (Carnegie Mellon), SIPTA, Manno, 2005, pp. 349-358.

[21] P. Walley. *Coherent lower (and upper) probabilities.* Research Report, University of Warwick, Department of Statistics, Warwick, 1981.

[22] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London, 1991.