Uncertainty measures in the theory of imprecise probabilities

Andrey G. Bronevich

JSC "Research, Development and Planning Institute for Railway Information Technology, Automation and Telecommunication"
Nizhegorodskaya 27, building 1, 109029, Moscow, Russia
brone@mail.ru

prepared for International Centre for Information and Uncertainty, Palacky University, Olomouc

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

Basic concepts of imprecise probabilities

- Classical probability theory works with single probability measures.
- The theory of imprecise probabilities works with sets of probability measures.

In this lecture we consider probability measures defined on the powerset 2^X of a finite set $X = \{x_1, ..., x_n\}$.

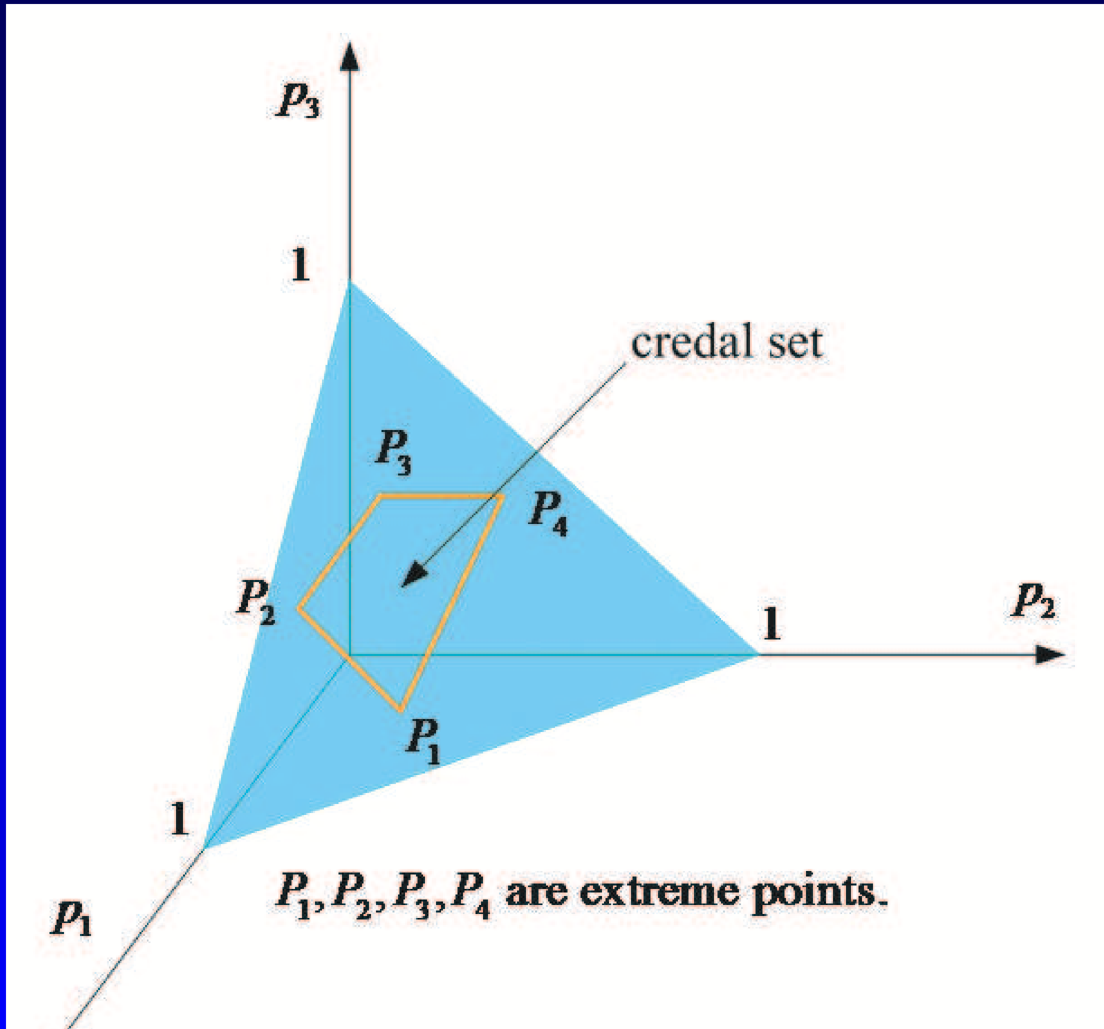
$M_{pr}(X)$ is the set of all probability measures on 2^X.

Credal sets

In this lecture a credal set is understood as a closed convex set of probability measures with a finite number of extreme points. If $\mathbf{P}$ is a credal set and $P_k \in M_{pr}(X)$, $k = 1, ..., m$, are its extreme points then

$$\mathbf{P} = \left\{ \sum_{k=1}^{m} a_i P_i | a_i \geqslant 0, \sum_{k=1}^{m} a_i = 1 \right\}.$$

Let $X = \{x_1, x_2, x_3\}$, then any credal set is convex subset of triangle consisting of points (p_1, p_2, p_3): $p_i \geqslant 0$, $p_1 + p_2 + p_3 = 1$.



Monotone measures

Let X be a finite set. A set function $\mu : 2^X \to [0,1]$ is called a monotone measure if

1. $\mu(\emptyset) = 0$, $\mu(X) = 1$ (norming);
2. $A \subseteq B$ implies $\mu(A) \leqslant \mu(B)$ (monotonicity).

Notation:

- $M_{mon}(X)$ is the set of all monotone measures on 2^X;
- $\mu_1 \leqslant \mu_2$ for $\mu_1, \mu_2 \in M_{mon}(X)$ if $\mu_1(A) \leqslant \mu_2(A)$ for all $A \in 2^X$.

Lower probabilities

A monotone measure μ is called a lower probability if there is a $P \in M_{pr}$ such that $\mu \leqslant P$.

Any lower probability μ defines a credal set

$$\mathbf{P}(\mu) = \{P \in M_{pr}(X) | P \geqslant \mu\}.$$

Let μ be a lower probability on 2^X, where $X = \{x_1, x_2, x_3\}$, then extreme points of $\mathbf{P}(\mu)$ can be found by solving the following inequalities:

$$\begin{cases} p_1 \geqslant \mu\left(\{x_1\}\right), \\ p_2 \geqslant \mu\left(\{x_2\}\right), \\ p_3 \geqslant \mu\left(\{x_3\}\right), \\ p_1 + p_2 \geqslant \mu\left(\{x_1, x_2\}\right), \\ p_1 + p_3 \geqslant \mu\left(\{x_1, x_3\}\right), \\ p_2 + p_3 \geqslant \mu\left(\{x_2, x_3\}\right), \\ p_1 + p_2 + p_3 = 1. \end{cases}$$

Clearly lower probabilities are less general than credal sets.

Upper probabilities

A monotone measure μ is called an upper probability if there is a $P \in M_{pr}$ such that $\mu \geqslant P$.

Any upper probability generate a credal set $\{P \in M_{pr}(X) | P \leqslant \mu\}$.

It is possible to consider only lower probabilities. Let μ be an upper probability. Introduce into consideration a measure $\mu^d(A) = 1 - \mu(A^c)$. The measure μ^d is called dual of μ. Clearly μ^d and μ generate the same credal set

Coherent lower probabilities

A lower probability μ is called a coherent lower probability if for any $A \in 2^X$ there is a $P \in M_{pr}$ such that $\mu \leqslant P$ and $\mu(A) = P(A)$.

Any coherent lower probability can be generated as follows: if $\mathbf{P}$ is a credal set then

$$\mu(A) = \min_{P \in \mathbf{P}} P(A),\, A \in 2^X,$$

is a coherent lower probability.

Coherent upper probabilities

An upper probability μ is called a coherent upper probability if for any $A \in 2^X$ there is a $P \in M_{pr}$ such that $\mu \geqslant P$ and $\mu(A) = P(A)$.

Any coherent upper probability can be generated as follows: if $\mathbf{P}$ is a credal set then

$$\mu(A) = \max_{P \in \mathbf{P}} P(A),\, A \in 2^X,$$

is a coherent upper probability.

2-monotone measures

A monotone measure is called 2-monotone if the following inequality holds:

$$\mu(A) + \mu(B) \leqslant \mu(A \cap B) + \mu(A \cup B).$$

for the dual measure the following inequality holds:

$$\mu^d(A) + \mu^d(B) \geqslant \mu^d(A \cap B) + \mu^d(A \cup B).$$

This measure is called 2-alternative. It is known that any 2-monotone measure is a coherent lower probability, and any 2-alternative measure is a coherent upper probability.

Belief and plausibility measures

Belief and plausibility measures are defined by means of a basic probability assignment. A basic probability assignment m is a non-negative set function on 2^X such that

1. $m(\emptyset) = 0$;
2. $\sum_{A \in 2^X} m(A) = 1$ (norming).

Then

$$Bel(A) = \sum_{B \subseteq A} m(B) \text{ and } Pl(B) = \sum_{B \cap A \neq \emptyset} m(A).$$

The set A is called focal for some m if $m(A) > 0$.

Some times, it is useful to represent belief functions using $\{0,1\}$-valued measures:

$$\eta_{\langle B\rangle}(A) = \begin{cases} 1, & B \subseteq A, \\ 0, & otherwise. \end{cases}$$

Then

$$Bel(A) = \sum_{B \in 2^X} m(B)\eta_{\langle B\rangle}(A).$$

The sense of $\eta_{\langle B\rangle}$ is the following. It describes situation when we know that the random variable definitely takes values from the set B, but we don't know any additional information.
Clearly, $Pl = Bel^d$.

Possibility and necessity measures

A possibility measure Pos is such that $Pos \in M_{mon}$,

$$Pos(A \cup B) = \max\{Pos(A), Pos(B)\} \; A, B \in 2^X.$$

A necessity measure Nec is such that $Nec \in M_{mon}$,

$$Nec(A \cap B) = \min\{Nec(A), Nec(B)\} \; A, B \in 2^X.$$

The dual of a necessity measure is a possibility measure. Any necessity measure is a belief measure. A belief measure is a necessity measure if focal elements form a chain.

Möbius transform

The set of all set functions on 2^X is a linear space and the system of set functions $\left\{\eta_{\langle B\rangle}\right\}_{B\in 2^X}$ is the basis of it. We can find the representation

$$\mu = \sum_{B\in 2^X} m(B)\eta_{\langle B\rangle}$$

of any $\mu : 2^X \to \mathbb{R}$ using the Möbius transform:

$$m(B) = \sum_{A\subseteq B} (-1)^{|B\setminus A|}\mu(A).$$

Projections of measures

Let $\mu \in M_{mon}(X)$ and $\varphi : X \to Y$, then μ^{φ} is a monotone measure on 2^Y defined by

$$\mu^{\varphi}(B) = \mu\left(\{x \in X | \varphi(x) \in B\right), \text{ where } B \in 2^Y.$$

Let $\mu \in M_{mon}(X \times Y)$ then marginal measures μ_X and μ_Y are defined by

1. $\mu_X(A) = \mu(A \times Y)$ for $A \in 2^X$;
2. $\mu_Y(A) = \mu(X \times A)$ for $A \in 2^Y$.

Projections of credal sets

The same operations are analogously defined for credal sets:

Let $\mathbf{P} \in Cr(X)$ and $\varphi : X \to Y$ then

$$\mathbf{P}^{\varphi} = \{P^{\varphi} | P \in \mathbf{P}\}.$$

Let $\mathbf{P} \in Cr(X \times Y)$ then marginal credal sets $\mathbf{P}_X$ and $\mathbf{P}_Y$ are defined by

$$\mathbf{P}_X = \{P_X | P \in \mathbf{P}\} \text{ and } \mathbf{P}_Y = \{P_Y | P \in \mathbf{P}\}.$$

Shannon entropy

Let P be a probability measure on 2^X then the Shannon entropy is defined by

$$S(P) = -c \sum_{x_i \in X} P(\{x_i\}) \ln P(\{x_i\}), \text{ where } c > 0.$$

If information is measured in bits, and the information of one bit is equal to 1 then

$$S(P) = -\sum_{x_i \in X} P(\{x_i\}) \lg_2 P(\{x_i\}).$$

The Shannon entropy measures conflict in the information.

Hartley measure

Let us assume that we have the information that the random variable takes definitely a value from an non-empty set $A \subseteq X$, The uncertainty of this information is measured by a Hartley measure:

$$H(A) = c \ln |A|.$$

If information is measured in bits, and the information of one bit is equal to 1 then

$$H(A) = \lg_2 |A|.$$

The Hartley measure reflects the non-specificity in the information.

Types of uncertainty in the theory of imprecise probabilities

Conflict. It refers to probability measures.

Non-specificity. It refers to the choice of a probability measure from the possible alternatives.

Types of uncertainty measures:

- U_N is a measure of non-specificity;
- U_C is a measure of conflict;
- U_T is a measure of total uncertainty.

Requirements for choosing uncertainty measures suggested by George J. Klir

Subaddivity: The amount of uncertainty in a joint representation of evidence (defined on Cartesian product) cannot be greater then the sum of amounts of uncertainty in the associated marginal representations of evidence.

Additivity: The amount of uncertainty in a joint representation of evidence is equal to the sum of the amounts of uncertainty in the associated marginal representations of evidence if and only if the marginal representations are non-interactive according to the rules of uncertainty calculus involved.

Monotonicity: When evidence can be ordered in the uncertainty theory employed (as in possibility theory), the relevant uncertainty measure must preserve this ordering.

Continuity: Any measure of uncertainty must be continuous functional.

Expansibility: Expanding the universal set by alternatives that are not supported by evidence must not affect the amount of uncertainty.

Symmetry: The amount of uncertainty does not change when elements of the universal set are rearranged.

Range: The range of uncertainty is $[0, M]$, where 0 must be assigned to the unique uncertainty function that describe full certainty and M depends on the size of the universal set involved and on the chosen unit of measurement (normalization).

Branching/Consistency: When uncertainty can be computed in multiple ways, all acceptable within within the calculus of the uncertainty theory involved, the results must be the same (consistent).

Normalization: A measurement unite defined by specifying what the amount should be for a particular (and usually very simple) uncertainty function.

Axioms for choosing an uncertanty measure on M_{pr}

Subaddivity: Let $P \in M_{pr}(X \times Y)$, then $U_T(P_X) + U_T(P_Y) \geqslant U_T(P)$.

Additivity: Let $P \in M_{pr}(X \times Y)$ and $P = P_X \times P_Y$, then $U_T(P_X) + U_T(P_Y) = U_T(P)$.

Continuity: U_T is a continuous functional.

Expansibility: Let $P \in M_{pr}(X)$ and let $\varphi : X \to Y$ be a injection such that $X \subseteq Y$ and $\varphi(x) = x$ for all $x \in X$. Then $U_T(P^{\varphi}) = U_T(P)$.

Symmetry: Let $P \in M_{pr}(X)$ and let $\varphi : X \to X$ be a bijection, then $U_T(P^{\varphi}) = U_T(P)$.

Range: $U_T : M_{pr} \to [0, +\infty)$ and $U_T(P) = 0$ iff P is a Dirac measure, i.e. there is $x \in X$ such that $P(\{x\}) = 1$.

Normalization: Let $X = \{x_1, x_2\}$ and $P \in M_{pr}(X)$ is such that $P(\{x_1\}) = P(\{x_2\}) = 0.5$. Then $U_T(P) = 1$.

Remarks

1. It is well known that the above requirements lead to the Shannon entropy functional:

$$S(P) = -\sum_{x_i \in X} P(\{x_i\}) \lg_2 P(\{x_i\}).$$

2. The additivity axiom has the following interpretation through random variables: if random variables ξ_X ξ_Y are independent, then

$$U_T(\xi_X, \xi_Y) = U_T(\xi_X) + U_T(\xi_Y).$$

3. The additivity property of Shannon entropy can be understood also as

$$S(\xi_X, \xi_Y) = S(\xi_X|\xi_Y) + S(\xi_Y),$$

and the last expression can be taken as an additivity axiom.

4. The additivity axiom for general theories of imprecise probabilities must be based on more general independence principles than in the classical probability theory.

5. It is hard to understand what continuity means for functionals on credal sets.

6. Some times monotonicity requirement can be formulated as: Additional information reduces uncertainty.

7. It is possible to introduce one axiom that includes symmetry and expansibility axioms:

Let $P \in M_{pr}(X)$ and let $\varphi : X \to Y$ be a injection. Then $U_T(P^\varphi) = U_T(P)$.

Independent principles in the theory of imprecise probabilities

Notation:

X is a finite non-empty set;

Here we consider all possible sets of probability measures on 2^X.

The set of all possible such objects is denoted by $S_{pr}(X)$.

General definition

Let $\mathbf{P} \in S_{pr}(X \times Y)$, where X and Y are finite nonempty sets. Assume that $\mathbf{P}$ is the joint description of two random variables, ξ_X and ξ_Y, with values in X and Y, respectively. We say that ξ_Y is *irrelevant* to ξ_X if knowing an exact discription of ξ_Y has no influence on the description of ξ_X. They are *independent* if ξ_X is irrelevant to ξ_Y and ξ_Y is irrelevant to ξ_X.

Question: How this general definition can be viewed through concieved types of uncertainty: conflict and nonspecificity?

Independence in probability theory

Let $P \in M_{pr}(X \times Y)$ be the joint description of ξ_X and ξ_Y, and let P_X and P_Y be marginal probability measures.

Assume ξ_Y takes the value $y \in Y$. Then the information about ξ_X is described by $P_{|y} \in M_{pr}(X)$, defined by

$$P_{|y}(A) = \frac{P(A \times \{y\})}{P(X \times \{y\})},$$

where $A \in 2^X$ and $P(X \times \{y\}) \neq 0$.

ξ_Y is irrelevant to ξ_X iff
$P_{|y} = P_X$ for any $y \in Y$ with $P_Y(\{y\}) \neq 0$.

Random variables ξ_X and ξ_Y are independent
if ξ_X is irrelevant to ξ_Y, and ξ_Y is irrelevant to ξ_X.

It is well known that in probability theory irrelevance implies independence, and $P = P_X \times P_Y$.

Two types of conditioning

1. Let we know the exact description $P_Y \in \mathbf{P}_Y$, of the random variable, ξ_Y. Then

$$\mathbf{P}_{|P_Y} = \{\mu \in \mathbf{P} | \mu_Y = P_Y\}$$

is the conditioning given P_Y.

2. Let we know both the probability distribution and the true value $y \in Y$ of ξ_Y in the experiment. Then for any $y \in Y$ with $P_Y(\{y\}) > 0$

$$\mathbf{P}_{|P_Y,y} = \left\{\mu_{|y} | \mu \in \mathbf{P}_{|P_Y}\right\}$$

is the conditioning given P_Y and $y \in Y$.

Precise general definition.

We say that ξ_Y is fully irrelevant (or irrelevant) to ξ_X iff

$$\mathbf{P}_{|P_Y,y} = \left(\mathbf{P}_{|P_Y}\right)_X = \mathbf{P}_X$$

for any $P_Y \in \mathbf{P}_Y$ and any $y \in Y$ with $P_Y(\{y\}) > 0$.

ξ_X and ξ_Y are called fully independent (or independent)if the full irrelevance is fulfilled in both directions.

Independence related to nonspecificity (marginal independence)

ξ_Y is marginally irrelevant to ξ_X if

$$\left(\mathbf{P}_{|P_Y}\right)_X = \mathbf{P}_X \text{ for any } P_Y \in \mathbf{P}_Y.$$

ξ_X and ξ_Y are called marginally independent if the marginal irrelevance is fulfilled in both directions.

Independence related to conflict (epistemical independence)

Let $\mathbf{P}_{|y} = \bigcup_{P_Y \in \mathbf{P}_Y | P_Y(\{y\})>0} \mathbf{P}_{|P_Y,y}$.

Then ξ_Y is epistemically irrelevant to ξ_X if

$$\mathbf{P}_{|y} = \mathbf{P}_X \text{ for any } y \in Y \text{ such that } \mathbf{P}_{|y} \neq \emptyset.$$

ξ_X and ξ_Y are called epistemically independent if the epistemical irrelevance is fulfilled in both directions.

Examles

Let random variables ξ_X and ξ_Y be described by a set $\mathbf{P} \in S_{pr}\left(X \times Y\right)$. Then

a) they are independent if

$$\mathbf{P} = \{P_X \times P_Y | P_X \in \mathbf{P}_X, P_Y \in \mathbf{P}_Y\};$$

b) ξ_Y is irrelevant to ξ_X if

$$\mathbf{P} = \{P \in M_{pr}\left(X \times Y\right) | P_X \in \mathbf{P}_X, P_Y \in \mathbf{P}_Y\}$$

and $\mathbf{P}_X = \mathbf{P}\left(\eta_{\langle B \rangle}\right)$ for some nonempty set $B \subseteq X$.

Main result

Theorem *Let random variables* ξ_X *and* ξ_Y *be jointly described by a credal set* $\mathbf{P} \in S_{pr}(X \times Y)$. *Then* ξ_Y *is fully irrelevant to* ξ_X *iff* ξ_Y *is marginally and epistemically irrelevant to* ξ_X.

It is possible to show by an example that there are cases when marginal and epistemical irrelevance does not imply full irrelevance in general.

Products

The inverse problem: How to define the joint description of independent sources of information using marginals?

The solution is based on the maximum uncertainty principle and on the following. If random variables ξ_X and ξ_Y are independent and described by sets $\mathbf{P}_X$ and $\mathbf{P}_Y$. Then among their possible joint descriptions there is a largest set defined by

$$\begin{aligned}\mathbf{P}_{\max} = \; & \{P \in M_{pr}(X \times Y) | \\ & \forall x \in X : P_{|x}, P_Y \in \mathbf{P}_Y; \\ & \forall y \in Y : P_{|y}, P_X \in \mathbf{P}_X \} \,.\end{aligned}$$

This set is called the product of $\mathbf{P}_X$ and $\mathbf{P}_Y$ and denoted by $\mathbf{P}_X \times \mathbf{P}_Y$.

Other products

The marginal independence implies the following product.

$$\mathbf{P}_X \times_N \mathbf{P}_Y = \{P \in M_{pr}(X \times Y) | P_X \in \mathbf{P}_X, P_Y \in \mathbf{P}_Y\}.$$

Under the assumption that ξ_Y is irrelevant to ξ_X, we get the following largest set:

$$\mathbf{P}_X \times_I \mathbf{P}_Y = \left\{ P \in M_{pr}(X \times Y) \middle| P_Y \in \mathbf{P}_Y; \forall y \in Y : P_{|y}, P_X \in \mathbf{P}_X \right\}.$$

Properties

if $\mathbf{P}_X$ and $\mathbf{P}_Y$ are credal sets, then the epistemic independence implies the introduced product $\mathbf{P}_X \times \mathbf{P}_Y$.

It is possible to show that if if $\mathbf{P}_X$ and $\mathbf{P}_Y$ are credal sets are credal sets, then $\mathbf{P}_X \times \mathbf{P}_Y$, $\mathbf{P}_X \times_N \mathbf{P}_Y$, $\mathbf{P}_X \times_I \mathbf{P}_Y$ are also credal sets, i.e. the introduced operations can be performed within credal sets.

Strong independence

Let $\mathbf{P}_X \in Cr(X)$ and $\mathbf{P}_Y \in Cr(Y)$. Then a credal set in $Cr(X \times Y)$, being a convex closure of the set $\{P_X \times P_Y | P_X \in \mathbf{P}_X, P_Y \in \mathbf{P}_Y\}$ describes *strong independence* of credal sets $\mathbf{P}_X$ and $\mathbf{P}_Y$. We denote this product by $\mathbf{P}_X \times_S \mathbf{P}_Y$.

The strong independence give us the smallest set of probability measures, for which independence is fulfilled. This implies from the next proposition.

Proposition. *Let independent random variables* ξ_X *and* ξ_Y *be described by a credal set* $\mathbf{P} \in Cr(X \times Y)$*. Then*

(i) ξ_Y *is irrelevant to* ξ_X *iff*
$\mathbf{P}_X \times_S \mathbf{P}_Y \subseteq \mathbf{P} \subseteq \mathbf{P}_X \times_I \mathbf{P}_Y$*;*

(ii) ξ_X *and* ξ_Y *are independent iff*
$\mathbf{P}_X \times_S \mathbf{P}_Y \subseteq \mathbf{P} \subseteq \mathbf{P}_X \times \mathbf{P}_Y$.

Möbius product

Let $\mu_X \in M_{bel}(X)$, $\mu_Y \in M_{bel}(Y)$ and let m_X, m_Y be their basic probability assingments.

Then the Möbius product of μ_X and μ_Y is a belief measure $\mu \in M_{bel}(X \times Y)$ with a basic probability assingment

$m(A \times B) = m_X(A)m_Y(B)$ (m is equal to 0 on other subsets of $X \times Y$).

The Möbius product of μ_X and μ_Y is denoted by $\mu = \mu_X \times_M \mu_Y$.

The probabilistic interpretation of Möbius product

Let $\mu \in M_{bel}(X)$ and let m be its basic probability assingment. Then μ can be concieved as a description of random value ξ with values in 2^X such that $\Pr(\xi = A) = m(A)$.

Then $\mu(A) = \Pr(\xi \subseteq A)$.

Let ξ be a random value with values in $2^{X \times Y}$ and ξ_X, ξ_Y be its projections on X and Y, respectively. Then

$\Pr(\xi_X = A) = \Pr\{pr_X \xi = A\}$,

$\Pr(\xi_Y = B) = \Pr\{pr_Y \xi = B\}$.

ξ_X and ξ_Y are independent according to the usual definition if for any $A \in 2^X$ and $B \in 2^Y$

$$\Pr(\xi_X = A)\Pr(\xi_Y = B) = \Pr\{pr_X\xi = A, pr_Y\xi = B\}.$$

If ξ_X and ξ_Y are independent, in addition, by known marginals $\mu_X \in M_{bel}(X)$ and $\mu_Y \in M_{bel}(Y)$ their joint description $\mu \in M_{bel}(X \times Y)$ according to the maximum uncertainty principle can be defined as

$$\mu = \mu_X \times_M \mu_Y.$$

Paper: A.G. Bronevich, G.J. Klir Axioms for Uncertainty Measures on Belief Functions and Credal Sets

Objectives for investigation: Introducing axioms for a total uncertainty measure and its disaggregation on belief functions and credal sets under the principle of uncertainty invariance.

Previous works

1. It was established that if we consider coherent lower probabilities, there are two types of uncertainty "conflict" and "nonspecificity". One can find other terms in the literature, for example, "conflict" = ŞrandomnessŞ, "nonspecificity" = "imprecision". 2.

There is an opinion that measures of uncertainty interact in additive manner, i.e. there is a measure of total uncertainty U_T that accumulates additively two types of uncertainty by

$$U_T = U_N + U_C,$$

where U_N is a measure of non-specificity and U_C is a measure of conflict.

3. Let U be an uncertainty measure. What kinds of properties it should possess? There is an opinion that these properties should generalize properties of the Shannon entropy and the Hartley measure.

Let us remind that the Shannon entropy S is the functional defined on the set of probability measures by

$$S(P) = -c \sum_{\omega \in \Omega} P(\{\omega\}) \ln P(\{\omega\}),$$

where $P \in M_{pr}$ and $c > 0$ is chosen by using the normalization (boundary) condition.

The Hartley measure H is used when we have the only information about random variable ξ that it takes value in a set A. This information can be described by a $\{0, 1\}$-valued necessity measure $\eta_{\langle A \rangle}$ and by definition

$$H\left(\eta_{\langle A \rangle}\right) = c \ln\left(|A|\right),$$

where $c > 0$ is chosen by using the normalization condition.

These measures have the following properties:

P1. Symmetry: $S\left(P^{\varphi}\right) = S(P)$,
$H\left(\eta^{\varphi}_{\langle A\rangle}\right) = H\left(\eta_{\langle A\rangle}\right)$ for any bijection $\varphi : \Omega_1 \to \Omega_2$.

P2. Label Independency: $S\left(P^{\varphi}\right) = S(P)$,
$H\left(\eta^{\varphi}_{\langle A\rangle}\right) = H\left(\eta_{\langle A\rangle}\right)$ for any bijection $\Omega_1 \to \Omega_2$.

P3. Expansibility: $S\left(P^{\varphi}\right) = S(P)$,
$H\left(\eta^{\varphi}_{\langle A\rangle}\right) = H\left(\eta_{\langle A\rangle}\right)$ for any injection $\varphi : \Omega_1 \to \Omega_2$
such that $\Omega_1 \subset \Omega_2$ and $\varphi(\omega) = \omega$ for each $\omega \in \Omega$.

P4. Additivity: $S\left(P_X \times P_Y\right) = S\left(P_X\right) + S\left(P_Y\right)$,
$H\left(\eta_{\langle A\times B\rangle}\right) = H\left(\eta_{\langle A\rangle}\right) + H\left(\eta_{\langle B\rangle}\right).$

P5. Subadditivity: Let $\Omega = X \times Y$, $P \in M_{pr}(\Omega)$ and $C \subseteq \Omega$. Then $S\left(P\right) \leqslant S\left(P_X\right) + S\left(P_Y\right)$,
$H\left(\eta_{\langle C\rangle}\right) \leqslant H\left(\eta_{\langle pr_X C\rangle}\right) + H\left(\eta_{\langle pr_Y C\rangle}\right)$.

Let us notice that P2 => P1 and P1, P2 and P3 can be equivalently changed to

P1 - P3. $S\left(P^{\varphi}\right) = S(P)$, $H\left(\eta^{\varphi}_{\langle A\rangle}\right) = H\left(\eta_{\langle A\rangle}\right)$ for any injection $\varphi : \Omega_1 \to \Omega_2$.

If we go to more general theories of imprecise probabilities, then there are questions: How to generalize these properties? What new properties can be considered as necessary ones?

Because of many approaches to independence in the theory of imprecise probabilities, it is not clear how define additivity properties of uncertainty measures.

Harmanec's axioms for a total uncertainty measure on M_{bel}

R0. Functionality. A measure of total uncertainty is a functional $U_T : M_{bel} \to [0, +\infty)$.

R1. Label Independency. Let X, Y be finite nonempty sets and $\varphi : X \to Y$ be a bijection. Then $U_T\left(\mu^{\varphi}\right) = U_T\left(\mu\right)$ for any $\mu \in M_{bel}(X)$.

R2. Continuity. Let $\mu \in M_{bel}(X)$, m be the Möbius transform of μ. Then the function $f(x) = U_T\left(\mu - x\eta_{\langle A \rangle} + x\eta_{\langle B \rangle}\right)$, which is defined for arbitrary nonempty sets $A, B \in 2^X$ and any $x \in [-m(B), m(A)]$, is continuous on $[-m(B), m(A)]$.

R3. Expansibility. Let X and Y be finite nonempty sets, $X \subset Y$, and $\varphi : X \to Y$ be an injection, defined by $\varphi(x) = x$ for all $x \in X$. Then $U_T\left(\mu^{\varphi}\right) = U_T\left(\mu\right)$ for any $\mu \in M_{bel}(X)$.

R4. Subadditivity. Let $\mu \in M_{bel}\left(X \times Y\right)$, then $U_T\left(\mu_X\right) + U_T\left(\mu_Y\right) \geqslant U_T\left(\mu\right)$.

R5. Additivity. Let $\mu_X \in M_{bel}(X)$, $\mu_Y \in M_{bel}(Y)$, and let $\mu \in M_{bel}\left(X \times Y\right)$ be the Möbius product of μ_X and μ_Y. Then $U_T\left(\mu_X\right) + U_T\left(\mu_Y\right) = U_T\left(\mu\right)$.

R6. Monotone Dispensability. Let $\mu \in M_{bel}(X)$ and m be the Möbius transform of μ. If $\nu \in M_{bel}(X)$ can be represented as $\nu = \sum_{A \in 2^X \setminus \emptyset} m(A)\mu_A$, where $\mu_A \in M_{bel}(X)$ and $\mu_A \leqslant \eta_{\langle A \rangle}$ for all $A \in 2^X \setminus \emptyset$, then $U_T(\mu) \leqslant U_T(\nu)$.

R7. Probabilistic Normalization. If $X = \{x_1, x_2\}$, $P \in M_{pr}(X)$, and $P(\{x_1\}) = P(\{x_2\}) = 0.5$. Then $U_T(P) = 1$.

R8. Nonspecificity Normalization. If $X = \{x_1, x_2\}$, then $U_T\left(\eta_{\langle X \rangle}\right) = 1$.

Questions:

1. How to generalize continuity axiom R3 for credal sets?
2. How to generalize additivity axiom R6 for credal sets?
3. Why axiom R7 is presented in this form? May be it is better to use

R10. Strong Monotone Dispensability. Let $\mu, \nu \in M_{bel}(X)$ and $\mu \geqslant \nu$. Then $U_T(\mu) \leqslant U_T(\nu)$

4. Why it is required (see axioms R8 and R9) that $U_T\left(\eta_{\langle X\rangle}\right) = U_T(P)$ for $X = \{x_1, x_2\}$ and for the probability measure P defined in R8?

R1-R5 can be easily reformulated for credal sets.

D. Harmanec has proved that the upper entropy:

$$S^*(\mu) = \sup\{S(P)|P \in \mathcal{P}(\mu)\}$$

satisfies axioms R1-R9 and this is the smallest one among functionals obeying axioms R1-R9.

Possible disaggregations of S^*

1. $U_T = S^*, U_N = GH, U_C = S^* - GH,$

where GH is the generalized Hartley measure.

If $\mu = \sum_{A \in 2^X \setminus \emptyset} m(A)\eta_{\langle A \rangle}$, then

$$GH(\mu) = c \sum_{A \in 2^X \setminus \emptyset} m(A) \ln |A| \,.$$

2. $U_T = S^*, U_N = S^* - S_*, U_C = S_*,$
where S_* is the minimal entropy defined by

$$S_*(\mu) = \inf \{S(P) | P \in \mathbf{P}(\mu)\} \,.$$

Properties of uncertainty measures

	S^*	GH	$S^* - GH$	$S^* - S_*$	S_*
subadditivity	+	+	-	-	-
additivity w.r.t. Möbius product	+	+	+	-	-
addivity w.r.t. strong independence	+	-	-	+	+

Questions:

1. Is the property of subadditivity essential for measures of nonspecificity and measures of conflict?

2. How the generalized Hartley measure can be generalized for credal sets?

3. Does a justifiable subadditive measure of conflict exist or does not?

4. What additivity properties are essential for total uncertainty measures, measures of nonspecificity and measures of conflict?

5. Is a total uncertainty measure unique or is not?

To answer these questions, it is necessary

1. To introduce a system of axioms for uncertainty measures, which can be equivalently formulated for belief functions and credal sets.

2. To look critically at independence principles in the theory of imprecise probabilities through the problem of defining uncertainty measures with properties, which are similar to ones of the Shannon entropy.

Axioms for a total uncertainty measure and its disaggregation on belief functions

U_T is a measure of total uncertainty;
U_N is a measure of nonspecificity;
U_C is a measure of conflict.

Axiom 1. Let $\mu \in M_{bel}(X)$. Then $U_N(\mu) = 0$ if $\mu \in M_{pr}(X)$ and $U_C(\mu) = 0$ if $\mu = \eta_{\langle B \rangle}$, $B \in 2^X \backslash \emptyset$.

Axiom 2. Let $\varphi : X \to Y$ be an injection, i.e. $\varphi(x_1) \neq \varphi(x_2)$ if $x_1 \neq x_2$. Then $U_T(\mu^\varphi) = U_T(\mu)$, $U_N(\mu^\varphi) = U_N(\mu)$, $U_C(\mu^\varphi) = U_C(\mu)$ for any $\mu \in M_{bel}(X)$.

Partial cases of Axiom 2:

Symmetry Axiom if $Y = X$ and φ is a bijection;

Label Independency Axiom if φ is a bijection;

Expansibility Axiom if $X \subseteq Y$ is an injection such that $\varphi(x) = x$ for all $x \in X$.

Axiom 3. Let $\mu \in M_{bel}(X)$, $Y \subseteq X$, and $\varphi : X \to Y$ with $\varphi(x) \in \varphi^{-1}(\varphi(x))$ for any $x \in X$. Then $U_T(\mu) \geqslant U_T(\mu^{\varphi})$.

Axiom 4. If $\mu_1, \mu_2 \in M_{bel}(X)$ and $\mu_1 \leqslant \mu_2$, then $U_N(\mu_1) \geqslant U_N(\mu_2)$ and $U_T(\mu_1) \geqslant U_T(\mu_2)$.

Axiom 5. Let $\mu = \mu_X \times_M \mu_Y$, where $\mu_X \in M_{bel}(X)$, $\mu_Y \in M_{bel}(Y)$, and $\mu_X = \eta_{\langle A \rangle}$ for some $A \subseteq X$. Then $U_T(\mu) = U_T(\mu_X) + U_T(\mu_Y)$.

Axiom 6. Let $\mu \in M_{bel}(X \times Y)$ and $\mu_Y \in M_{pr}(Y)$. Then

$$U_T(\mu) = \sum_{y \in Y} \mu_Y(\{y\}) \, U_T\left(\mu_{|y}\right) + U_T(\mu_Y),$$

where $\mu_{|y}(A) = \dfrac{\mu(A \times \{y\})}{\mu_Y(\{y\})}$, $A \in 2^X$.

Axiom 6 is the generalization of the property of Shannon entropy: $S(\xi, \eta) = S(\xi|\eta) + S(\eta)$, where ξ and η are random variables with values in X and Y.

Axiom 7. Let $\mu \in M_{bel}(X \times Y)$. Then $U_T(\mu) \leq U_T(\mu_X) + U_T(\mu_Y)$ (the subadditivity axiom).

Axiom 8. $U_C(\mu) + U_N(\mu) = U_T(\mu)$ for any $\mu \in M_{bel}$.

Corollaries from axioms

Corollary 1. Let $\mu_1, \mu_2 \in M_{bel}(X)$,

$$\mu = a\mu_1 + (1-a)\mu_2 \text{ for } a \in [0,1].$$

Then

$$aU_T(\mu_1) + (1-a)U_T(\mu_2) \leqslant U_T(\mu).$$

Corollary 2. Let $\mu = \sum_{k=1}^{m} a_k \mu_k$, where $\mu_k \in M_{bel}(X_k)$, $a_k \geqslant 0$, $k = 1, \ldots, m$, $\sum_{k=1}^{m} a_k = 1$, and X_k, $k = 1, \ldots, m$, be pairwise disjoint finite nonempty sets, i.e. $\{X_k\}_{k=1}^{m}$ is a partition of $X = \bigcup_{k=1}^{m} X_k$. Then

$$U_T(\mu) = \sum_{k=1}^{m} a_k U_T(\mu_k) + U_T(\mu^{\varphi}),$$

where $\varphi : X \to \{X_1, \ldots, X_m\}$ is such that $\varphi(x) = X_k$ if $x \in X_k$.

Corollary 3. *Let* $P \in M_{pr}(X)$. *Then* $U_T(P) = S(P)$, *where* S *is the Shannon entropy.*

Corollary 4. *Let* $\mu \in M_{bel}(\Omega)$ *and* $\mu = \eta_{\langle A \rangle}$, $A \in 2^{\Omega} \backslash \emptyset$. *Then* $U_T(\mu) = H(\mu)$, *where* H *is the Hartley measure.*

Corollary 5. *Let* $\mu = \sum_{k=1}^{m} a_k \mu_k$, *where* $\mu_k \in M_{bel}(X)$, $a_k \geqslant 0$, $k = 1, ..., m$, $\sum_{k=1}^{m} a_k = 1$, *and let* $P \in M_{pr}(\{1, ..., m\})$ *be such that* $P(\{k\}) = a_k$, $k = 1, ..., m$. *Then*

$$\sum_{k=1}^{m} a_k U_T(\mu_k) + U_T(P) \geqslant U_T(\mu).$$

P1. The maximal entropy S^* satisfies all the axioms for a total uncertainty measure on M_{bel}.

P2. Possible disaggregations of S^* on M_{bel}:

$U_T = S^*, U_C = S_*, U_N = S^* - S_*$, where S_* is the minimal entropy;

$U_T = S^*, U_N = GH, U_C = S^* - GH$, where GH is the generalized Hartley measure.

Axioms for uncertainty measures on credal sets

Axiom 1c. Let $\mathbf{P} \in Cr(X)$. Then $U_N(\mathbf{P}) = 0$ if $\mathbf{P}$ is a singleton and $U_C(\mathbf{P}) = 0$ if $\mathbf{P} = \mathbf{P}\left(\eta_{\langle B \rangle}\right)$, $B \subseteq X$.

Axiom 2c. Let $\varphi : X \to Y$ be an injection. Then $U_T\left(\mathbf{P}^{\varphi}\right) = U_T\left(\mathbf{P}\right)$, $U_N\left(\mathbf{P}^{\varphi}\right) = U_N\left(\mathbf{P}\right)$, $U_C\left(\mathbf{P}^{\varphi}\right) = U_C\left(\mathbf{P}\right)$ for any $\mathbf{P} \in Cr(X)$.

Axiom 3c. Let X, Y be finite sets, $\varphi : X \to Y$ and $\mathbf{P} \in Cr(X)$. Then $U_T\left(\mathbf{P}\right) \geqslant U_T\left(\mathbf{P}^{\varphi}\right)$.

Axiom 4c. If $\mathbf{P}_1, \mathbf{P}_2 \in Cr(X)$ and $\mathbf{P}_1 \supseteq \mathbf{P}_2$, then $U_N\left(\mathbf{P}_1\right) \geqslant U_N\left(\mathbf{P}_2\right)$ and $U_T\left(\mathbf{P}_1\right) \geqslant U_T\left(\mathbf{P}_2\right)$.

Axiom 5c. Let X, Y be finite sets, $\mathbf{P}_X = \mathbf{P}\left(\eta_{\langle A \rangle}\right)$, $A \subseteq X$, and $\mathbf{P}_Y \in Cr(Y)$. Consider a credal set $\mathbf{P}^* \in Cr(X \times Y)$, defined by $\mathbf{P}^* = \mathbf{P}_X \times_N \mathbf{P}_Y$. Then

$$U_T\left(\mathbf{P}^*\right) = U_T\left(\mathbf{P}_X\right) + U_T\left(\mathbf{P}_Y\right).$$

Axiom 6c. Let $\mathbf{P} \in Cr(X \times Y)$ and $\mathbf{P}_Y = \{P_Y\}$, where $P_Y \in M_{pr}(Y)$. Then

$$U_T(\mathbf{P}) = \sum_{y \in Y} P_Y(\{y\})\, U_T\left(\mathbf{P}_{|y}\right) + U_T\left(\mathbf{P}_Y\right),$$

where $\mathbf{P}_{|y} = \left\{P_{|y} | P \in \mathbf{P}\right\}$.

Axiom 7c. Let X, Y be finite sets and $\mathbf{P} \in Cr(X \times Y)$. Then $U_T(\mathbf{P}) \leq U_T(\mathbf{P}_X) + U_T(\mathbf{P}_Y)$ (the subadditivity axiom).

Axiom 8c. $U_C(\mathbf{P}) + U_N(\mathbf{P}) = U_T(\mathbf{P})$, $\mathbf{P} \in Cr$.

The set of all possible total uncertainty measures and its structure

$\mathfrak{F}(M_{bel})$ is the set of all total uncertainty measures on M_{bel}.

P1. $\mathfrak{F}(M_{bel})$ is a convex cone, i.e. $f_i \in \mathfrak{F}(M_{bel})$, $c_i \geqslant 0$, $i = 1, 2$, implies $c_1 f_1 + c_2 f_2 \in \mathfrak{F}(M_{bel})$, and $-f \notin \mathfrak{F}(M_{bel})$ for any $f \neq 0$ in $\mathfrak{F}(M_{bel})$.

Normalization conditions:

Let $X = \{x_1, x_2\}$ and $P \in M_{pr}(X)$ such that $P(\{x_1\}) = 0.5$. Then

$$\mathfrak{F}_{a,b}(M_{bel}) = \left\{ f \in \mathfrak{F}(M_{bel}) \,|\, f\left(\eta_{\langle X \rangle}\right) = a, f(P) = b \right\}.$$

Axiom 4 implies that $a \geqslant b \geqslant 0$. Any $f \not\equiv 0$ in $\mathfrak{F}_{a,b}(M_{bel})$ if $a > 0$.

Proposition. *For any* $a > 0$, $\mathfrak{F}_{a,0}(M_{bel}) = \{GH\}$, *where* GH *is the generalized Hartley measure with* $GH\left(\eta_{\langle X \rangle}\right) = a$, $|X| = 2$.

$\mathcal{F}(\mu)$ is the set of focal elements of $\mu \in M_{bel}$.

$M_{bel|d}(X)$ is the set of all possible belief measures on 2^X with disjoint focal elements.

Proposition. *Let* $f \in \mathfrak{F}_{a,b}\left(M_{bel}\right)$*,* $\mu \in M_{bel|d}(X)$*, and let* m *be the Möbius transform of* μ*. Then*

$$f(\mu) = a \sum_{B \in \mathcal{F}(\mu)} m(B) \lg_2 |B| - b \sum_{B \in \mathcal{F}(\mu)} m(B) \lg_2 m(B).$$

$\preccurlyeq$ is nonstrict order on M_{bel} defined by $\mu_1 \preccurlyeq \mu_2$ for $\mu_1 \in M_{bel}(X)$ and $\mu_2 \in M_{bel}(Y)$ if there is a mapping $\varphi : Y \to X$ such that $\mu_1^{\varphi} \leqslant \mu_2$.

P2. $\preccurlyeq$ is transitive on M_{bel} and $U_T(\mu_1) \geq U_T(\mu_2)$ if $\mu_1 \preccurlyeq \mu_2$.

An upper bound of an arbitrary $U_T \in \mathfrak{F}_{a,b}\left(M_{bel}\right)$:
$\bar{U}_T^{a,b}\left(\mu\right) = \inf\left\{U_T(\nu) | \nu \in M_{bel|d}, \nu \preccurlyeq \mu\right\};$.

A lower bound of an arbitrary $U_T \in \mathfrak{F}_{a,b}\left(M_{bel}\right)$:
$\underline{U}_T^{a,b}\left(\mu\right) = \sup\left\{U_T(\nu) | \nu \in M_{bel|d}, \mu \preccurlyeq \nu\right\}.$

P3. $\bar{U}_T^{a,b}(\mu)$, $\underline{U}_T^{a,b}(\mu)$ do not depend on a chosen $U_T \in \mathfrak{F}_{a,b}(M_{bel})$ by Proposition 3.

Proposition. *The following statements are true:*

1) $\underline{U}_T^{a,b} \leqslant U_T \leqslant \bar{U}_T^{a,b}$ *for any* $U_T \in \mathfrak{F}_{a,b}(M_{bel})$;
2) $\underline{U}_T^{a,a} = S^*$;
3) $\bar{U}_T^{a,0} = GH$.

P4. $\underline{U}_T^{a,b} \notin \mathfrak{F}_{a,b}(M_{bel})$ if $a > 0$ and $b = 0$.

Question: whether $\bar{U}_T^{a,b}$ is a total uncertainty measure or not?

Providing the uniqueness of a total uncertainty measure under the law of conflict-nonspecificity transformation

A measure of nonspecificity consists of 2 parts:

$U_N^{(1)}(\mu) = \sup\{U_C(g) | g \in M_{bel}(X), g \geqslant \mu\} - U_C(\mu)$ is the amount of nonspecificity, which can be transformed to conflict;

$U_N^{(2)}(\mu) = U_N(\mu) - U_N^{(1)}(\mu)$ is the amount of nonspecificity, which cannot be transformed to conflict.

Suppose that $U_N^{(1)}(\mu)$ can be transformed to pure conflict. Then

$$U_N^{(1)}(\mu) = \sup\{U_C(P) | P \in M_{pr}(X), P \geqslant \mu\} - U_C(\mu).$$

We have that $U_T = U_N^{(1)} + U_N^{(2)} + U_C$, where $S^* = U_N^{(1)} + U_C$ is a total uncertainty measure.

Assume that $U_N^{(2)}$ is a total uncertainty measure.

Then $U_T \in \mathfrak{F}_{a,b}\left(M_{bel}\right)$ is defined uniquely and it is represented by

$$U_T = S^* + GH,$$

where

$S^* \in \mathfrak{F}_{b,b}\left(M_{bel}\right)$ is the upper entropy:

$GH \in \mathfrak{F}_{a-b,0}\left(M_{bel}\right)$ is the generalized Hartley measure.

In particular, if $a = b$, then $U_T = S^*$.

Open problems

1. Are sets $\mathfrak{F}_{a,0}\left(M_{2-mon}\right)$, $\mathfrak{F}_{a,0}\left(Cr\right)$ empty? It is likely that $\mathfrak{F}_{a,0}\left(M_{2-mon}\right) \neq \emptyset$, i.e. the generalized Hartley measure can be linearly extended to the set of 2-monotone measures.

2. Are sets $\mathfrak{F}_{a,b}\left(M_{bel}\right)$, $a > 0$, $a \geqslant b \geqslant 0$, singletons?

3. What kind of additional justifiable properties should measures of nonspecificity and conflict possess?