Breaf introduction to 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

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.

Theory of imprecise probabilities and decision theory

Suppose that you are planning to spend your vacations and there are three possibilities to spend them at home, in Spain or Italy.
You can evaluate your decisions exactly if you know the information about weather.
If the weather will be bad, then it is preferable to stay at home, in case of good weather it is preferable to travel to Italy, and if the weather will be excellent, then you should choose the travel to Spain.
Assume that preferences can be measured by real numbers that show the benefit of each action.

	Bad Weather	Good Weather	Excellent Weather
Home	5	5	5
Italy	-5	10	10
Spain	-10	5	20

Table 1: Evaluation of actions

The mathematical model for describing decisions is in the following.
We have the space of the states of the world $S = \{s_1, ..., s_N\}$ and decisions $D = \{d_1, ..., d_K\}$ described by the functions $f_i : S \to \mathbb{R}$, $i = 1, ..., k$.

In our example,

$S = \{$Bad Weather, Good Weather, Excelent Weather$\}$,

$$D = \{Home, Italy, Spain\},$$

and functions are depicted below.

$$f_1(Home): \begin{array}{lr} S & \mathbb{R} \\ Bad\,Weather & \to 5 \\ Good\,Weather & \to 5 \\ Excelent\,Weather & \to 5 \end{array}$$

$$f_2(Italy): \quad \begin{array}{rl} S & \mathbb{R} \\ Bad\,Weather & \rightarrow -5 \\ Good\,Weather & \rightarrow 10 \\ Excelent\,Weather & \rightarrow 10 \end{array}$$

$$f_3(Spain): \quad \begin{array}{rl} S & \mathbb{R} \\ Bad\,Weather & \rightarrow -10 \\ Good\,Weather & \rightarrow 5 \\ Excelent\,Weather & \rightarrow 20 \end{array}$$

The value of function $f_i(s_k)$ can be interpreted as a gain (expressed as amount of money) if we make decision d_i and observe the state of the world s_k.

Obviously, states of the world can appear randomly and this can be described by a probability distribution $p(s_k)$, i.e. we assign to each state of the world s_k the probability $p(s_k)$ such that

1) $p(s_k) \geqslant 0, k = 1, ..., N,$

2) $\sum_{k=1}^{N} p(s_k) = 1.$

Let this distribution be known. Then the choice of the best decision is based on expected utility computed as follows

$$u(d_i) = \sum_{k=1}^{N} f_i(s_k)p(s_k).$$

Thus, $d_k \in D$ is the best decision if $u(d_k) \geqslant u(d_i)$ for all $i \in \{1, ..., K\}$. We denote the linear quasi-order on decisions by $\preccurlyeq$, i.e.

$$d_i \preccurlyeq d_j \text{ if } u(d_i) \leqslant u(d_j).$$

Let us consider the following problem. Assume that we have a linear quasi-order $\preccurlyeq$ on decisions. What are the conditions, when $\preccurlyeq$ coincides with the order based on expected utility.

Theorem 1. Let $\mathcal{K}$ be the closed cone of possible decisions, i.e. $f_1, f_2 \in \mathcal{K}$ implies $\lambda_1 f_1 + \lambda_2 f_2 \in \mathcal{K}$ for any $\lambda \in [0, +\infty)$, and let $\preccurlyeq$ be a linear quasi-order on $\mathcal{K}$. Then $\preccurlyeq$ coincides with the order based on expected utility with a certain probability distribution on S iff

1. $\lambda f_1 + (1-\lambda) f_3 \prec \lambda f_2 + (1-\lambda) f_3$ for any $f_1, f_2, f_3 \in \mathcal{K}$ such that $f_1 \prec f_2$ and $\lambda \in (0,1]$.
2. $f_1, f_2 \in \mathcal{K}$ and $f_1 \leqslant f_2$ implies $f_1 \preccurlyeq f_2$.
3. Let $\{f_i\}_{i=1}^{\infty} \subseteq \mathcal{K}$, $\lim\limits_{i\to\infty} f_i \to f$, and $g_1 \preccurlyeq f_i \preccurlyeq g_2$, $i = 1, 2, \ldots$. Then $g_1 \preccurlyeq f \preccurlyeq g_2$.

Credal sets

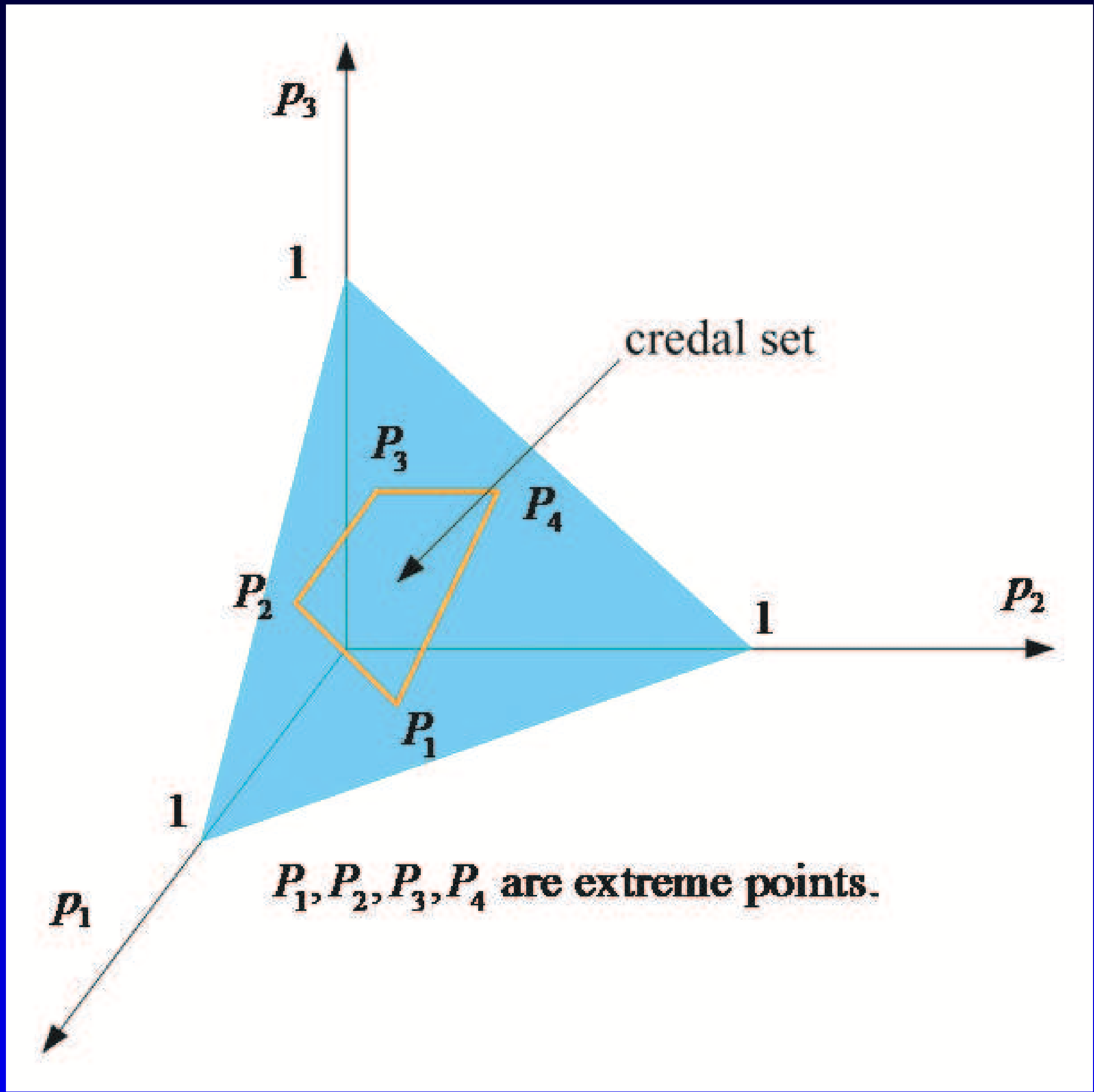
If we have incomparable decisions, then this can be modeled by credal sets. A credal set is closed, convex set of probability distributions (probability measures).

If $\mathbf{P}$ is a credal set with a finite number of extreme points $P_k \in M_{pr}(X)$, $k = 1, ..., m$, 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): $p_i \geqslant 0$, $p_1 + p_2 + p_3 = 1$.

The typical credal set



If we model uncertainty by a credal set $\mathbf{P}$, then for any decision $f \in \mathcal{K}$, we can compute upper and lower estimates of expected utility:

$$\bar{u}(f) = \sup_{p \in \mathbf{P}} \sum_{s_i \in S} f(s_i)p(s_i),$$

$$\underline{u}(f) = \inf_{p \in \mathbf{P}} \sum_{s_i \in S} f(s_i)p(s_i).$$

In this case the preference relation $\preccurlyeq$ is a partial quasi-order defined by $d_i \preccurlyeq d_j$ if $\underline{u}(f_i) \leqslant \underline{u}(f_j)$ and $\bar{u}(f_i) \leqslant \bar{u}(f_j)$.

Giron and Rios found the conditions, when a partial quasi-order on decisions can be described by upper and lower estimates of expected utility.

Theorem 2. Let $\mathcal{K}$ be a closed cone of possible decisions, i.e. $f_1, f_2 \in \mathcal{K}$ implies $\lambda_1 f_1 + \lambda_2 f_2 \in \mathcal{K}$ for any $\lambda \in [0, +\infty)$,

and let $\preccurlyeq$ be a partial quasi-order on $\mathcal{K}$.

Then $\preccurlyeq$ coincides with the order based on upper and lower estimates of expected utility with a certain credal set on S iff

1. $\lambda f_1 + (1-\lambda) f_3 \prec \lambda f_2 + (1-\lambda) f_3$ for any $f_1, f_2, f_3 \in \mathcal{K}$ such that $f_1 \prec f_2$ and $\lambda \in (0,1]$.
2. $f_1, f_2 \in \mathcal{K}$ and $f_1 \leqslant f_2$ implies $f_1 \preccurlyeq f_2$.
3. Let $\{f_i\}_{i=1}^{\infty} \subseteq \mathcal{K}$, $\lim\limits_{i\to\infty} f_i \to f$, and $g_1 \preccurlyeq f_i \preccurlyeq g_2$, $i = 1, 2, \ldots$. Then $g_1 \preccurlyeq f \preccurlyeq g_2$.

Upper and lower previsions

In decision theory functions from the set $\mathcal{K}$ can be called *decisions*, *acts*, or *gambles*. In the theory of imprecise probabilities the term "gamble" is adopted. The upper estimate $\bar{u}(f)$ of expected utility is called the *upper prevision* of $f \in \mathcal{K}$, and $\underline{u}(f)$ is called the *lower prevision* of $f \in \mathcal{K}$.

Values $\underline{u}(f)$ and $\bar{u}(f)$ have the following behavior interpretation.

Assume you would like to buy or sell shares to get a gain during some period.

Then values $\underline{u}(f)$ and $\bar{u}(f)$ can be interpreted as upper and lower bounds of expected gain.

Using this information you can sell shares with the price higher or equal to $\bar{u}(f)$ and buy them with the price equal or lower than $\underline{u}(f)$.

Suppose that values $\underline{u}(f)$ and $\bar{u}(f)$ are given by experts.

Then it is possible that there is a contradiction among their opinions. To detect a contradiction, it is used avoiding sure loss principle.

The estimates $\underline{u}(f)$ and $\bar{u}(f)$ on $\mathcal{K}$ *avoid sure loss* if there is a probability distribution p on S such that

$$\underline{u}(f) \leqslant \sum_{s_i \in S} f(s_i) p(s_i) \leqslant \bar{u}(f) \text{ for any } f \in \mathcal{K}.$$

In other words, the family of probability distributions

$$\mathbf{P} = \left\{ p | \underline{u}(f) \leqslant \sum_{s_i \in S} f(s_i) p(s_i) \leqslant \bar{u}(f), f \in \mathcal{K} \right\}$$

is not empty. Obviously, $\mathbf{P}$ is a credal set.

If $\mathcal{K}$ is a finite set, then checking whether estimates avoid sure loss or not is produced by solving linear programming problem.

It is easy to see that the estimate $\bar{u}(f) = a$ is equivalent to $\underline{u}(-f) = -a$, because the inequality

$$\sum_{s_i \in S} f(s_i) p(s_i) \leqslant \bar{u}(f)$$

is equivalent to inequality

$$-\bar{u}(f) \leqslant \sum_{s_i \in S} (-f(s_i))p(s_i).$$

This allows us in the sequel to consider lower previsions on $\mathcal{K}$ assuming that upper previsions are computed by $\bar{u}(f) = -\underline{u}(-f)$.

The following theorem allows us to express avoiding sure loss principle directly through the values of $\underline{u}(f)$.

Theorem. A lower prevision $\underline{u}$ avoids sure loss on $\mathcal{K}$ iff

$$\sup_{s \in S} \sum_{k=1}^{N} (\lambda_k f_k(s) - \lambda_k \underline{u}(f_k)) \geqslant 0$$

for any $f_1, ..., f_N \in \mathcal{K}$ and $\lambda_1, ..., \lambda_N \geqslant 0$.

Example. Suppose that you bet on the match between Barcelona and Bavaria, $S = \{s_1, s_2, s_3\}$, where $s_1 :=$"Barselona wins", $s_2 :=$"Bavaria wins", $s_3 :=$"Draw", and you have the following gambles

$$f_k(s) = \begin{cases} 1, & s = s_k, \\ 0, & otherwise. \end{cases}$$

Assume that $\underline{u}(f_1) = 0.6$, $\underline{u}(f_2) = 0.3$, and $\underline{u}(f_3) = 0.2$.
Then

$$\sup_{s \in S} \sum_{k=1}^{3} (f_k(s) - \underline{u}(f_k)) = 1 - 1.1 = -0.1,$$

i.e. the avoiding sure loss principle is not fulfilled.

This means that if you bet on all possible results, then you loss money no matter the outcome of the match.

Coherence

Let a lower prevision $\underline{u}$ on $\mathcal{K}$ avoids sure loss. Then it defines a credal set

$$\mathbf{P} = \left\{ p | \underline{u}(f) \leqslant \sum_{s_i \in S} f(s_i) p(s_i), f \in \mathcal{K} \right\}. \qquad (1)$$

It is possible that the estimate $\underline{u}(f)$ is not exact w.r.t. $\mathbf{P}$ and we can compute exact boundaries for expected utility by finding

$$\underline{P}(f) = \inf_{p \in \mathbf{P}} \sum_{s_i \in S} f(s_i) p(s_i), \; f \in \mathcal{K}.$$

$\underline{P}$ is called the *natural extension* of $\underline{u}$.

If $\underline{P}(f) = \underline{u}(f)$ for all $f \in \mathcal{K}$, then $\underline{u}$ is called a *coherent lower prevision*.

Obviously, any functional $\underline{P}$ on $\mathcal{K}$ defined by a credal set $\mathbf{P}$ by formula (1) is a coherent lower prevision.

The computation of natural extension can produced by solving the following linear programming problem:

$$\sum_{s_i \in S} f(s_i)p(s_i) \to \min$$

given

$$\begin{cases} \sum\limits_{s_i \in S} f_k(s_i)p(s_i) \geqslant \underline{u}(f_k),\ f_k \in \mathcal{K}, \\ \sum\limits_{s_i \in S} p(s_i) = 1, \\ p(s_i) \geqslant 0,\ s_i \in S. \end{cases}$$

The natural extension can be also computed as

$$\underline{P}(f) = \sup\left\{\sum_k \lambda_k \underline{u}(f_k) \middle| \sum_k \lambda_k f_k \leqslant f, \lambda_k \geqslant 0, f_k \in \mathcal{K}\right\}$$

Example. Assume that in previous example

$\underline{u}(f_1) = 0.3$, $\underline{u}(f_2) = 0.3$, $\underline{u}(f_3) = 0.2$,

$\bar{u}(f_1) = 0.6$, $\bar{u}(f_2) = 0.6$, $\bar{u}(f_3) = 0.4$.

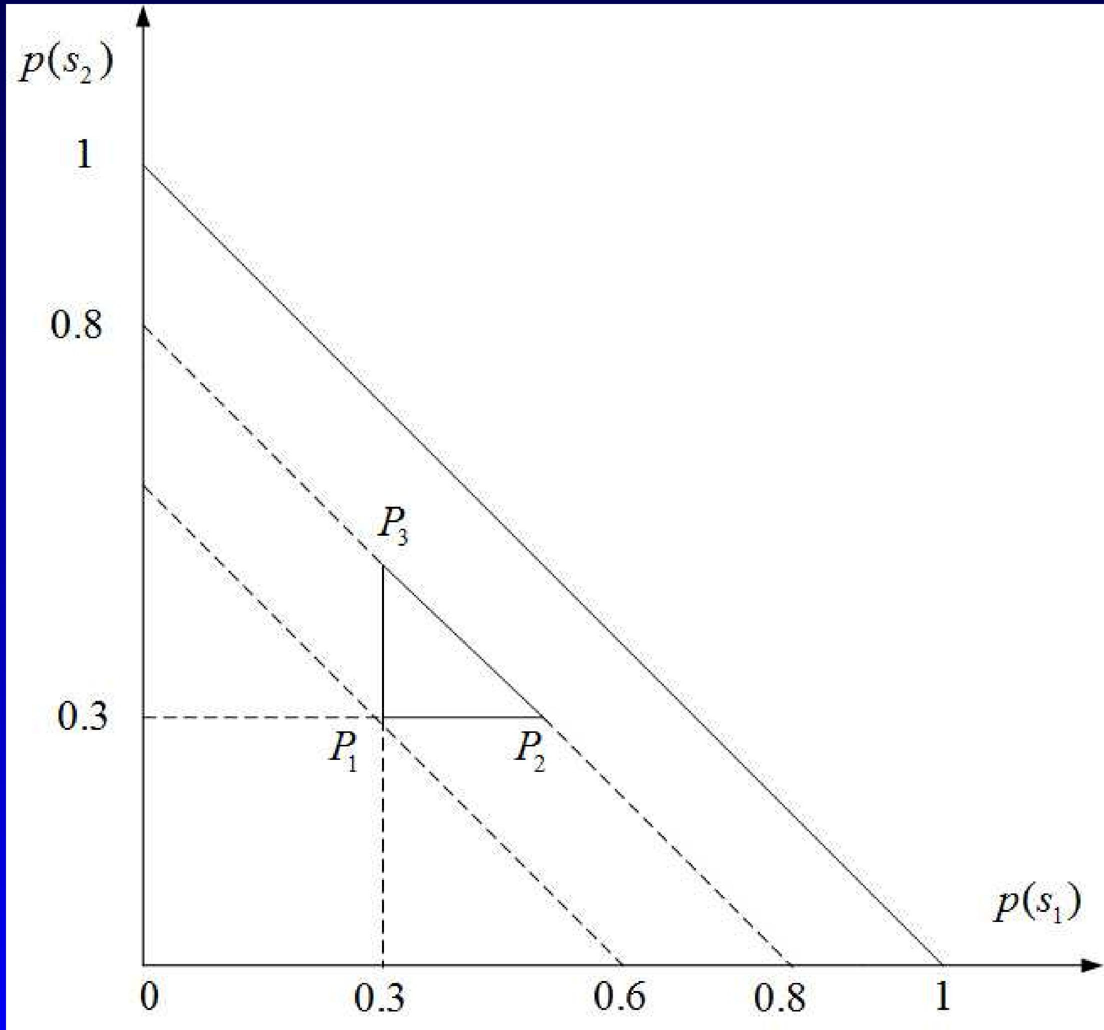
Then the corresponding credal set can be described by the following system of linear inequalities:

$$\begin{cases} 0.3 \leqslant p(s_1) \leqslant 0.6, \\ 0.3 \leqslant p(s_2) \leqslant 0.6, \\ 0.2 \leqslant p(s_3) \leqslant 0.4, \\ p(s_1) + p(s_2) + p(s_3) = 1. \end{cases}$$

If we express the value $p(s_3) = 1 - p(s_1) - p(s_2)$, and put it to the third inequality, we get

$$\begin{cases} 0.3 \leqslant p(s_1) \leqslant 0.6, \\ 0.3 \leqslant p(s_2) \leqslant 0.6, \\ 0.6 \leqslant p(s_1) + p(s_2) \leqslant 0.8. \end{cases}$$

The graphical solution of this system of inequalities is shown below.



We see that the extreme points of this credal set $\mathbf{P}$ are

$$P_1 = \left(\underbrace{0.3}_{p(s_1)}, \underbrace{0.3}_{p(s_2)}, \underbrace{0.4}_{p(s_3)} \right), P_2 = \left(\underbrace{0.5}_{p(s_1)}, \underbrace{0.3}_{p(s_2)}, \underbrace{0.2}_{p(s_3)} \right),$$

$$P_3 = \left(\underbrace{0.3}_{p(s_1)}, \underbrace{0.5}_{p(s_2)}, \underbrace{0.2}_{p(s_3)} \right).$$

We see that estimates $\underline{u}$ and $\bar{u}$ avoid sure loss, but they are not coherent. We can compute the natural extension of $\underline{u}$ and $\bar{u}$, using formulas:

$$\underline{P}(f_k) = \inf_{p \in \mathbf{P}} \sum_{s_i \in S} f_k(s_i) p(s_i),$$

$$\bar{P}(f_k) = \sup_{p \in \mathbf{P}} \sum_{s_i \in S} f_k(s_i) p(s_i).$$

Then

$\underline{P}(f_1) = 0.3$, $\underline{P}(f_2) = 0.3$, $\underline{P}(f_3) = 0.2$,

$\bar{P}(f_1) = 0.5$, $\bar{P}(f_2) = 0.5$, $\bar{P}(f_3) = 0.4$.

Using natural extension, we can extend a coherent prevision on the linear space of gambles. Therefore, in the sequel we assume that $\mathcal{K}$ is a linear space of functions on S.

Theorem. Let $\mathcal{K}$ be a linear space of functions on S. A function $\underline{P} : \mathcal{K} \to \mathbb{R}$ is a coherent lower prevision

iff

1. $\underline{P}(\lambda f + c) = \lambda \underline{P}(f) + c$ for all $\lambda \geqslant 0$, $c > 0$, and $f \in \mathcal{K}$.
2. $\underline{P}(f) \leqslant \underline{P}(g)$ if $f \leqslant g$ and $f, g \in \mathcal{K}$.
3. $\underline{P}(f + g) \geqslant \underline{P}(f) + \underline{P}(g)$ for all $f, g \in \mathcal{K}$.

Sets of desirable gambles

Let the imprecise probability model is given by a coherent lower prevision $\underline{P} : \mathcal{K} \to \mathbb{R}$. Then it can be equivalently described by the set of desirable gambles.

A gamble $f \in \mathcal{K}$ is called *desirable* if $\underline{P}(f) > 0$, i.e. the set of all desirable gambles that corresponds to $\underline{P}$ is

$$\mathcal{D} = \{f \in \mathcal{K} | \underline{P}(f) > 0\} .$$

It is easy to derive that $\mathcal{D}$ is a convex cone in $\mathcal{K}$ and it is characterized as follows:

a) if $f \leqslant 0$, then $f \notin \mathcal{D}$;

b) if $f > 0$, then $f \in \mathcal{D}$;

c) if $f, g \in \mathcal{D}$, then $f + g \in \mathcal{D}$;

d) if $f \in \mathcal{D}$ and $\lambda > 0$, then $\lambda f \in \mathcal{D}$.

For every cone of desirable games, we can recover the corresponding coherent lower and upper previsions by formulas:

$$\underline{P}(f) = \sup\left\{\alpha \in \mathbb{R} | f - \alpha \in \mathcal{D}\right\},$$

$$\bar{P}(f) = \inf\left\{\alpha \in \mathbb{R} | \alpha - f \in \mathcal{D}\right\}.$$

Obviously, these formulas imply that

$$\bar{P}(f) = -\underline{P}(-f).$$

Avoiding sure loss for sets of desirable gambles

Assume that an expert say that gambles from the set $\mathcal{D}_0$ are desirable. If expert's beliefs are contradictory, then there are desirable gambles $f_1, ..., f_N$ in $\mathcal{D}_0$ such that

$$\lambda_1 f_1 + ... + \lambda_N f_N \leqslant 0.$$

In other words, $\mathcal{D}_0$ avoids sure loss iff

$$\sup_{s \in S} \{\lambda_1 f_1(s) + ... + \lambda_N f_N(s) | \lambda_i > 0, f_i \in \mathcal{D}_0\} > 0.$$

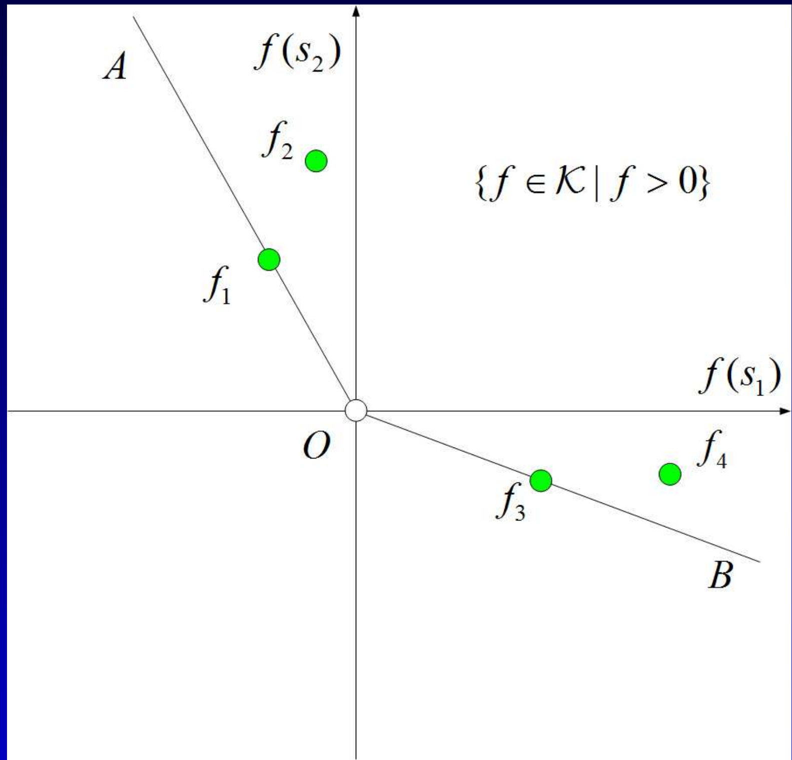
Natural extension for sets of desirable gambles

Let us assume that the set of desirable gambles $\mathcal{D}_0$ avoids sure loss and $\mathcal{D}_0 \supseteq \{f \in \mathcal{K} | f > 0\}$. Then we can describe the set of all desirable gambles in linear space $\mathcal{K}$ by

$$\mathcal{D} = \{\lambda_1 f_1(s) + ... + \lambda_N f_N(s) | \lambda_i > 0, f_i \in \mathcal{D}_0\},$$

i.e. $\mathcal{D}$ is a minimal cone in $\mathcal{K}$ that contains all linear combinations of gambles in $\mathcal{D}_0$ with positive coefficients. The set $\mathcal{D}$ can be considered as a natural extension of $\mathcal{D}_0$.

The graphical illustration of natural extension



$f_1, f_2, f_3, f_4 \in \mathcal{D}_0$,

$\mathcal{D} = \widehat{AOB} \backslash \{(0,0)\}$ is the natural extension.

Lower and upper probabilities

Imprecise probability models based on credal sets and coherent lower previsions are equivalent for the finite case, but the model based on the sets of desirable gambles is slightly general.

Less general models are based on lower and upper probabilities.

Consider gambles, defined by

$$1_A(s) = \begin{cases} 1, & s \in A, \\ 0, & s \notin A, \end{cases} \quad \text{where } A \subseteq S.$$

Let $\mathcal{K} = \{1_A\}_{A \in 2^S}$, where 2^S is the powerset of S, and $\underline{P}$ is a lower prevision on $\mathcal{K}$. Then we can describe values of $\underline{P}$ on $\mathcal{K}$ by a set function

$$\mu(A) = \underline{P}(1_A), \text{ where } A \in 2^S.$$

If $\underline{P}$ avoids sure loss on $\mathcal{K}$, then μ is called a *lower probability*.

Analogously, if $\bar{P}$ is an upper prevision and avoids sure loss, then a set function μ on 2^S defined by $\mu(A) = \bar{P}(1_A)$, where $A \in 2^S$, is called an *upper probability*.

Let μ be a lower probability on 2^S. Then the value $\mu(A)$ can be interpreted as a lower estimate of probability, therefore,

1. $\mu(\emptyset) = 0$ and $\mu(S) = 1$;
2. $\mu(A) \leqslant \mu(B)$ if $A \subseteq B$ (the larger set should have the larger probability).

Thus, lower and upper probabilities obey axioms 1 and 2 for monotone measures. Probability measures are also monotone measures, such that

$$\mu(A) + \mu(B) = \mu(A \cup B) \text{ if } A \cap B = \emptyset.$$

We denote the set of all probability measures on 2^S by M_{pr}.

If μ_1, μ_2 are monotone measures and $\mu_1(A) \leqslant \mu_2(A)$ for all $A \in 2^S$, then we write $\mu_1 \leqslant \mu_2$.

Interpretation of avoiding sure loss and natural extension for lower and upper probabilities

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

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

Obviously, a lower probability defines a credal set

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

and we can calculate the natural extension μ_{coh} of μ by

$$\mu_{coh}(A) = \inf_{P \in \mathbf{P}(\mu)} P(A).$$

If $\mu_{coh} = \mu$, then μ is called a coherent lower probability.

The equivalent definition: μ is a *coherent lower probability* if for any $A \in 2^S$ there is $P \in M_{pr}$ such that $\mu \leqslant P$ and $\mu(A) = P(A)$.

The dual relation $\bar{P}(f) = -\underline{P}(-f)$ for lower and upper previsions is transformed to the following dual relation on monotone measures. The set function μ^d is called the *dual* of μ if

$$\mu^d(A) = 1 - \mu(\bar{A}),\, A \in 2^S.$$

Assume that μ is a lower probability. Then obviously μ and μ^d define the same credal set

$$\mathbf{P}(\mu) = \{P \in M_{pr} | P \geqslant \mu\} = \left\{P \in M_{pr} | P \leqslant \mu^d\right\}.$$

This allows us to describe uncertainty using lower or upper probabilities.

2-monotone measures

A monotone measure μ is called 2-*monotone* if

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

for all $A, B \in 2^S$.

Any 2-monotone measure is a coherent lower probability, but 2-monotone measures is a special class of coherent lower probabilities.

Natural extension of a 2-monotone measure

The natural extension of a 2-monotone measure on the linear space of all gambles $\mathcal{K}$ is computed by the Choquet integral

$$\underline{P}(f) = \int_0^\infty \mu(\{s \in S | f(s) > t\}) dt$$

for any non-negative function $f \in \mathcal{K}$.

This integral can be extended on all functions in $\mathcal{K}$ using the property $\underline{P}(f + c) = \underline{P}(f) + c$, where $c \in \mathbb{R}$.

2-alternative measures

The dual measure μ^d of a 2-monotone measure μ is called 2-alternative. It obeys the following characteristic property:

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

for all $A, B \in 2^S$.

Belief and plausibility functions

A monotone measure $Bel : 2^S \to [0, 1]$ is called a *belief function* if it can be represented as

$$Bel(B) = \sum_{A \subseteq B} m(A),$$

where a non-negative set function m, called *the basic probability assignment*, has the following properties:

$$\sum_{A \in 2^S} m(A) = 1 \text{ and } m(\emptyset) = 0.$$

The dual measure of a belief function is called a plausibility function. It can be computed by m as follows

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

Any belief function can be represented as a convex sum of primitive belief functions of the type

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

as follows

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

Evidently any $\eta_{\langle\{A\}\rangle}$ describes uncertainty when we know that event A occurs but we don't have any additional information.

Any belief function is 2-monotone, i.e. belief functions are in the family of 2-monotone measures.

Example. Assume that experts predict the rate of euro to dollar. Each expert choose the interval of its possible values:

$$\left\{ \underbrace{[1.12, 1.32]}_{A_1}, \underbrace{[1.21, 1.41]}_{A_2}, \underbrace{[1.32, 1.51]}_{A_3}, \underbrace{[1.05, 1.32]}_{A_4} \right\}.$$

Then we can aggregate beliefs of experts using belief function

$$Bel = \sum_{i=1}^{4} m(A_i)\eta_{\langle\{A_i\}\rangle}.$$

In the last formula we can take into account the competence of experts by assigning values $m(A_i)$, $i = 1, ..., 4$. Assume that experts have the same competence, then $m(A_i) = 0.25$, $i = 1, ..., 4$ and, for example,

$$Bel([1.1, 1.5]) = 0.75;\ Pl([1.1, 1.5]) = 1;$$
$$Bel([1.4, 1.6]) = 0;\ Pl([1.1, 1.5]) = 0.5.$$

Possibility measures and necessity measures

A *possibility measure* Π is a monotone measure with the following property:

$\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}$ for every $A, B \in 2^S$.

A possibility measure is usually defined with the help of possibility distribution function

$$\pi(s) = \Pi(\{s\}), s \in S.$$

Then

$$\Pi(A) = \begin{cases} \max_{x \in A} \pi(x), & A \neq \emptyset, \\ 0, & A = \emptyset. \end{cases}$$

The dual of a possibility measure Π is a *necessity measure* N with the following characteristic property:

$$N(A \cap B) = \min\{N(A), N(B)\}$$

for every $A, B \in 2^S$.

Each necessity measure is a special belief function.

Let Bel be a belief function with the basic probability assignment m. Then a set A is called *focal* for Bel if $m(A) > 0$.

A belief function is a necessity measure iff the set of its focal elements is a chain with respect to $\subseteq$.

Conditioning

In the classical probability theory a probability of event A given B ($P(B) \neq 0$) is defined by

$$P(A|B) = P(A \cap B)/P(B).$$

Clearly, conditioning induces a probability measure on 2^S, which we denote P_B.

Let $\mathbf{P}$ be a credal set. Then $\mathbf{P}$ given B is the set of probability measures $\mathbf{P}_B$ defined by

$$\mathbf{P}_B = \{P_B | P \in \mathbf{P}, P(B) \neq 0\}.$$

Proposition. Let $\mathbf{P}$ be a credal set and let $B \in 2^S$ be such that $P(B) \neq 0$ for all $P \in \mathbf{P}$.

Then $\mathbf{P}_B$ is a credal set.

In addition, if $\mathbf{P}$ has a finite number of extreme points $P^{(1)}$, ..., $P^{(N)}$, then $P_B^{(1)}$, ..., $P_B^{(N)}$ are extreme points of $\mathbf{P}_B$.

Conditioning for coherent lower previsions

Proposition. Let $\underline{P}$ be a coherent lower prevision and $\underline{P}(1_B) > 0$, where $B \in 2^S$. Then the updated information in case of event B occurs is described by a coherent lower prevision $\underline{P}_B$ that obeys the following generalized Bayesian rule:

$$\underline{P}\left(1_B(f - \underline{P}_B(f)\right) = 0.$$

Remark. By above Proposition for finding $\underline{P}_B(f)$ we need to solve the equation $\underline{P}\left(1_B(f - c\right) = 0$ w.r.t. c. We get the unique solution, since the function $\underline{P}\left(1_B(f - c\right)$ is strictly decreasing in c, when $\underline{P}(1_B) > 0$.

Conditioning for sets of desirable gambles

Proposition. Let $\mathcal{D}$ be a set of all desirable gambles in the linear space $\mathcal{K}$.

Then the updated information in case of event $B \in 2^S$ occurs is described by the set $\mathcal{D}_B$ of desirable gambles defined by

$$\mathcal{D}_B = \{f \in \mathcal{K} | 1_B f \in \mathcal{D}\}.$$

Example. Coin Tossing. Suppose that a fair coin is 'tossed' twice, in such a way that heads and tails are equally likely on each of the tosses but there can be arbitrary dependence between the tosses. For example, the coin may be tossed first in the usual way, but on the second 'toss' it may be placed to have the same outcome as the first toss, or it may be placed to have the opposite outcome from the first toss.
Let

$$\left\{ \underbrace{x_1}_{(Heads,Heads)}, \underbrace{x_2}_{(Heads,Tails)}, \underbrace{x_3}_{(Tails,Heads)}, \underbrace{x_4}_{(Tails,Tails)} \right\}$$

be the set of possible outcomes.

Then the corresponding credal set P is described by the following system of inequalities.

$$\begin{cases} p(x_1) + p(x_2) = 0.5, \\ p(x_1) + p(x_3) = 0.5, \\ p(x_1) + p(x_2) + p(x_3) + p(x_4) = 1, \\ p(x_i) \geqslant 0,\ i = 1, ..., 4. \end{cases}$$

Solving this system, we get

$$\begin{cases} p(x_2) = 0.5 - p(x_1), \\ p(x_3) = 0.5 - p(x_1), \\ p(x_4) = p(x_1), \\ p(x_i) \geqslant 0,\ i = 1, ..., 4. \end{cases}$$

Thus, extreme points of $\mathbf{P}$ are

$$\left\{ \underbrace{0}_{p(x_1)}, \underbrace{0.5}_{p(x_2)}, \underbrace{0.5}_{p(x_3)}, \underbrace{0}_{p(x_4)} \right\}, \left\{ \underbrace{0.5}_{p(x_1)}, \underbrace{0}_{p(x_2)}, \underbrace{0}_{p(x_3)}, \underbrace{0.5}_{p(x_4)} \right\}.$$

The corresponding coherent lower and upper previsions can be computed as

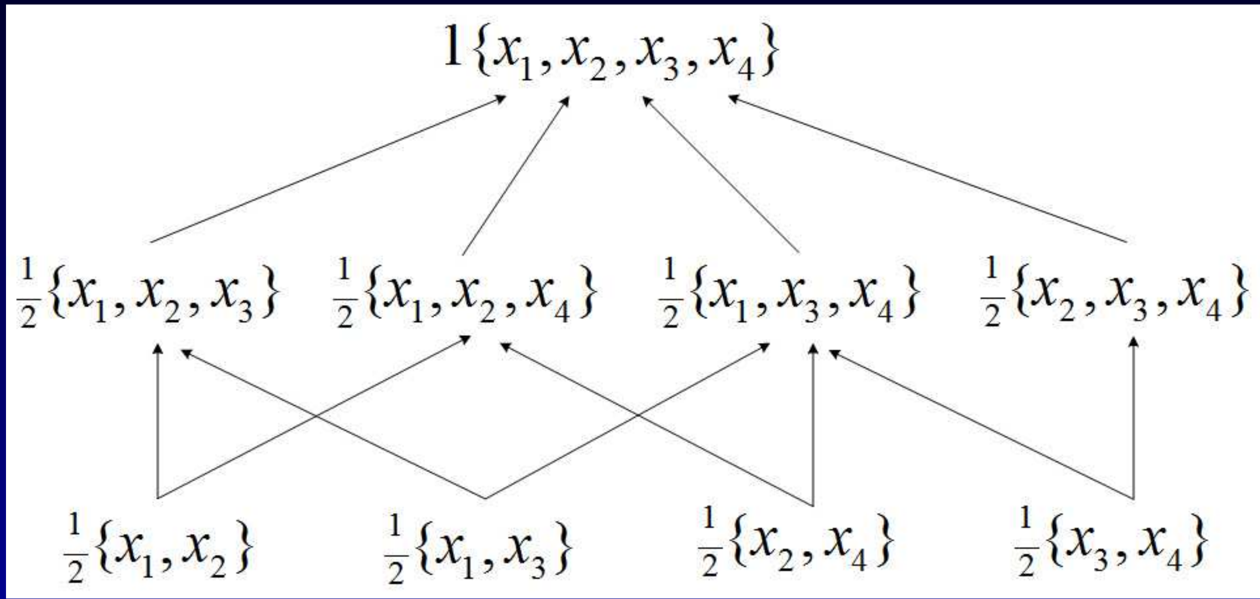
$$\underline{P}(f) = \min \left\{0.5f(x_1) + 0.5f(x_4), 0.5f(x_2) + 0.5f(x_3)\right\},$$

$$\bar{P}(f) = \max \left\{0.5f(x_1) + 0.5f(x_4), 0.5f(x_2) + 0.5f(x_3)\right\}.$$

The corresponding set of desirable gambles is

$$\mathcal{D} = \{f \in \mathcal{K} | f(x_1) + f(x_4) > 0\} \cap \\ \{f \in \mathcal{K} | f(x_2) + f(x_3) > 0\}.$$

We can also represent this uncertainty with the help of coherent lower probabilities. The values of the corresponding monotone measure μ are shown below.



In this case models described by P and μ are equivalent, but μ is not 2-monotone, because for sets $A = \{x_1, x_2\}$ and $B = \{x_1, x_3\}$

$$\mu(A) + \mu(B) = 1 \geqslant 0.5 = \mu(A \cup B) + \mu(A \cap B).$$

Let us consider on this example how conditioning is produced. Let us assume that at least one outcome is heads. Then we should find $\mathbf{P}_B$, where $B = \{x_1, x_2, x_3\}$. After updating each extreme point of $\mathbf{P}$, we get

$$\left\{ \underbrace{0}_{p(x_1)}, \underbrace{0.5}_{p(x_2)}, \underbrace{0.5}_{p(x_3)}, \underbrace{0}_{p(x_4)} \right\} \xrightarrow{given\ B} \left\{ \underbrace{0}_{p(x_1)}, \underbrace{0.5}_{p(x_2)}, \underbrace{0.5}_{p(x_3)}, \underbrace{0}_{p(x_4)} \right\}$$

$$\left\{ \underbrace{0.5}_{p(x_1)}, \underbrace{0}_{p(x_2)}, \underbrace{0}_{p(x_3)}, \underbrace{0.5}_{p(x_4)} \right\} \xrightarrow{given\ B} \left\{ \underbrace{1}_{p(x_1)}, \underbrace{0}_{p(x_2)}, \underbrace{0}_{p(x_3)}, \underbrace{0}_{p(x_4)} \right\}$$

Therefore,

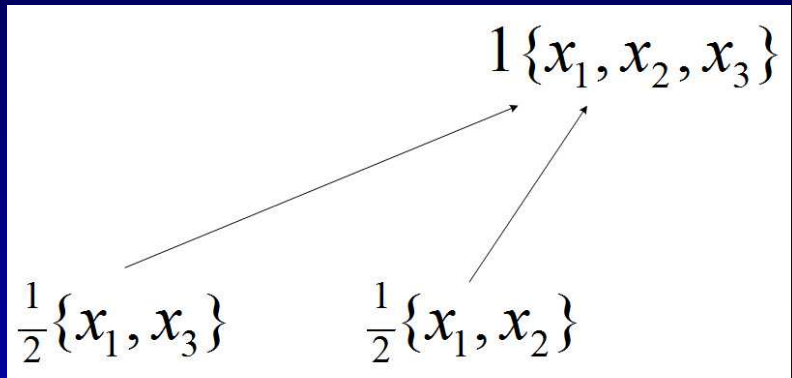
$$\underline{P}_B(f) = \min\left\{f(x_1), 0.5f(x_2) + 0.5f(x_3)\right\},$$

$$\bar{P}_B(f) = \max\left\{f(x_1), 0.5f(x_2) + 0.5f(x_3)\right\}.$$

We can represent this information as the set of desirable gambles as

$$\mathcal{D}_B = \{f \in \mathcal{K} | f(x_1) > 0\} \cap \{f \in \mathcal{K} | f(x_2) + f(x_3) > 0\}.$$

If we represent this information with coherent lower probabilities we get a monotone measure μ_B shown below.



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 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* $\mathrm{P} \in S_{pr}(X \times Y)$*. 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 P_X and P_Y are credal sets, then the epistemic independence implies the introduced product $\mathrm{P}_X \times \mathrm{P}_Y$.

It is possible to show that if if P_X and P_Y are credal sets, then $\mathrm{P}_X \times \mathrm{P}_Y$, $\mathrm{P}_X \times_N \mathrm{P}_Y$, $\mathrm{P}_X \times_I \mathrm{P}_Y$ 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" = "randomness", "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: $S\left(P^{\varphi}\right) = S(P)$,
$H\left(\eta_{\langle A\rangle}^{\varphi}\right) = H\left(\eta_{\langle A\rangle}\right)$ for any bijection $\varphi : \Omega_1 \to \Omega_2$.

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

P3. Expansibility: $S\left(P^{\varphi}\right) = S(P)$,
$H\left(\eta_{\langle A\rangle}^{\varphi}\right) = H\left(\eta_{\langle A\rangle}\right)$ 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 $\mu \in M_{bel}(X \times Y)$ be the Möbius product of μ_X and μ_Y. Then $U_T(\mu_X) + U_T(\mu_Y) = U_T(\mu)$.

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 $P\left(\{x_1\}\right) = P\left(\{x_2\}\right) = 0.5$. 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 $\varphi : X \to Y$. 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 $\mu_k \in M_{bel}\left(X_k\right)$, $a_k \geqslant 0$, $k = 1, ..., m$, $\sum_{k=1}^{m} a_k = 1$, and X_k, $k = 1, ..., m$, 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

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

where $\varphi : X \to \{X_1, ..., X_m\}$ 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* $\mu \in M_{bel}\left(\Omega\right)$ *and* $\mu = \eta_{\langle A \rangle}$, $A \in 2^{\Omega} \backslash \emptyset$. *Then* $U_T\left(\mu\right) = H\left(\mu\right)$, *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* $P \in M_{pr}\left(\{1, ..., m\}\right)$ *be such that* $P\left(\{k\}\right) = a_k$, $k = 1, ..., m$. *Then*

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

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 $U_T(\mathbf{P}^{\varphi}) = U_T(\mathbf{P})$, $U_N(\mathbf{P}^{\varphi}) = U_N(\mathbf{P})$, $U_C(\mathbf{P}^{\varphi}) = U_C(\mathbf{P})$ 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 $U_T(\mathbf{P}) \geqslant U_T(\mathbf{P}^{\varphi})$.

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

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 $P_Y \in M_{pr}\left(Y\right)$. Then

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

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 $f \neq 0$ 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 $U_T \in \mathfrak{F}_{a,b}\,(M_{bel})$:
$\bar{U}_T^{a,b}\,(\mu) = \inf \left\{ U_T(\nu) | \nu \in M_{bel|d}, \nu \preccurlyeq \mu \right\}$;.

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

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?