Generalized Hartley Measures on Credal Sets

Andrey G. Bronevich BRONE@MAIL.RU
Igor N. Rozenberg I.ROZENBERG@GISMPS.RU
JSC "Research and Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport", Orlikov per.5, building 1 107996 Moscow, Russia

Abstract

The paper considers various extensions of the Hartley measure on credal sets and their investigation based on a system of axioms.

Keywords: uncertainty measures, generalized Hartley measure, credal sets

1. Introduction

In the theory of imprecise probabilities [4, 23], there are many functionals for measuring uncertainty [5, 11, 19]. Among them we distinguish measures of conflict, non-specificity and total uncertainty. Conflict is related to modeling uncertainty by probability measures; non-specificity is connected to the choice of a probability model among possible ones. The total uncertainty is conceived as aggregated uncertainty of these two types. Historically, measures of conflict have been firstly introduced in probability theory, and they are known as entropies [22, 20]. The most popular of them are the Shannon entropy and Rényi entropies. The measure of non-specificity has been firstly introduced for analyzing information that can be described by non-graded possibility measures and called the Hartley measure [18]. There are several attempts to extend these measures to various models of imprecise probabilities [1, 5], or especially to belief functions [16]. As we can see from the literature [2, 3, 5, 11, 19] the most justified of them are the maximal (upper) entropy for evaluating total uncertainty, the minimal (lower) entropy for measuring conflict and the generalized Hartley measure as a measure of non-specificity.

Unfortunately, the generalized Hartley measure was fully accepted only for belief functions. Although there are several extensions of it to coherent lower probabilities and credal sets [1, 5], the thorough investigation of their properties was not produced yet. To close these gap, in this paper we consider three possible extensions of the Hartley measure to credal sets. The third one is new and based on an interpretation of the Hartley measure in decision theory as follows. An imprecise probability model can be viewed as a system of precise probability models. We call two probability models P_1 and P_2 fully contradictory if there is a decision f on the set of alternatives $X = \{x_1, \ldots, x_n\}$ whose utility is the highest and equal to $\max_{x \in X} f(x)$ for P_1 and the utility of f is the smallest and equal to $\min_{x \in X} f(x)$ for P_2. We show that the Hartley measure can be viewed as the logarithm of the maximal number of pairwise fully contradictory probability measures in the corresponding credal set. We also introduce the notion of ε-contradictory probability measures, and make hints of how to generalize the Hartley measure in this way.

The paper has the following structure. In Section 2, we recall the basic notions from the theory of belief functions and imprecise probabilities. Section 3 gives the axioms of classical uncertainty measures: the Shannon entropy and the Hartley measure. In Sections 4 and 5, we recall [5] the axioms for uncertainty measures on belief functions, the possible disaggregations of a measure of total uncertainty onto measures of conflict and non-specificity and the extension of these results on credal sets. In Section 6, we describe some properties of the linear extension of the generalized Hartley measure on coherent lower probabilities and credal sets. In Section 7, we introduce the Hartley measure on credal sets and investigate its properties. The paper is ended with the discussion of obtained results.

2. Monotone Measures and Credal Sets

Let X be a finite non-empty set and let 2^X denote the powerset of X. Then a set function $\mu : 2^X \to [0,1]$ is called a *monotone measure* or capacity [13, 15] if

1) $\mu(\emptyset) = 0$, $\mu(X) = 1$ (norming);

2) $\mu(A) \leqslant \mu(B)$ for every $A, B \in 2^X$ with $A \subseteq B$ (monotonicity).

Note that a probability measure is a special case of monotone measures. In this case, $\mu(A) + \mu(B) = \mu(A \cup B)$ for any disjoint $A, B \in 2^X$. We denote by $M_{mon}(X)$ the set of all monotone measures on 2^X, and by $M_{pr}(X)$ the set of all probability measures on 2^X. We will further use the following operations and relations on monotone measures. Let $\mu_1, \mu_2 \in M_{mon}(X)$, then

1) $\mu = a\mu_1 + (1-a)\mu_2$, where $a \in [0,1]$, if $\mu(A) = a\mu_1(A) + (1-a)\mu_2(A)$ for all $A \in 2^X$ (μ is called the convex sum of μ_1 and μ_2)[1];

1. Sometimes, we use the convex sum $\mu = a\mu_1 + (1-a)\mu_2$, where $\mu_i \in M_{mon}(X_i)$, $i = 1, 2$, and $X_1 \cap X_2 = \emptyset$. In this case, $\mu \in M_{mon}(X)$,

2) $\mu_1 \leqslant \mu_2$ if $\mu_1(A) \leqslant \mu_2(A)$ for all $A \in 2^X$;

3) $\mu_2 = \mu_1{}^d$ if $\mu_2(A) = 1 - \mu_1(A^c)$ for all $A \in 2^X$, where A^c denotes the complement of A (μ_2 is called the dual of μ_1).

μ is called a *lower probability* [4, 23] if the set $\mathbf{P}(\mu) = \{P \in M_{pr}(X) | P \geqslant \mu\}$is not empty, if, in addition, $\mu(A) = \inf_{P \in \mathbf{P}(\mu)} P(A)$ for all $A \in 2^X$, then μ is called a *coherent lower probability*. The family of probability measures $\mathbf{P}(\mu)$ is called the *credal set*. More generally, a credal set $\mathbf{P}$ is a non-empty convex and closed subset of $M_{pr}(X)$. The convexity means that $P_1, P_2 \in \mathbf{P}$ implies that $aP_1 + (1-a)P_2 \in \mathbf{P}$ for every $a \in [0,1]$. The closeness of $\mathbf{P}$ means that if we represent every $P \in M_{pr}(X)$, where $X = \{x_1,...,x_n\}$, as a point $P = (P(\{x_1\}),...,P(\{x_n\}))$ in $\mathbb{R}^n$, then $\mathbf{P}$ is a closed subset of $\mathbb{R}^n$. Instead of defining credal sets using lower probabilities, we can do it by upper probabilities. A $\mu \in M_{mon}(X)$ is called an upper probability if μ^d is a lower probability. It is easy to show that $\{P \in M_{pr}(X) | P \geqslant \mu\} = \{P \in M_{pr}(X) | P \leqslant \mu^d\}$ for every lower probability $\mu \in M_{mon}(X)$, i.e. imprecise probability models based on conjugate lower and upper probabilities are equivalent. Analogously if μ is a coherent lower probability, then μ^d is called a coherent upper probability.

If $\mu \in M_{mon}(X)$ has the property $\mu(A) + \mu(B) \le \mu(A \cap B) + \mu(A \cup B)$ for all $A, B \in 2^X$, then it is called *2-monotone* [13, 15]. It is possible to show that every 2-monotone measure is a coherent lower probability. The set of all 2-monotone measures on 2^X is denoted by $M_{2-mon}(X)$.

An important example of coherent lower probabilities (2-mononotone measures) are belief functions [21]. A $\mu \in M_{mon}(X)$ is called a *belief function* if there is $m : 2^X \to [0,1]$ called the *basic belief assignment* (bba) with $\sum_{A \in 2^X} m(A) = 1$ and $m(\emptyset) = 0$ such that $\mu(A) = \sum_{B \in 2^X | B \subseteq A} m(B)$. The set of all belief functions on 2^X is denoted by $M_{bel}(X)$. Assume that $\mu \in M_{bel}(X)$ with the bba m. A set $A \in 2^X$ is called a *focal element* for μ if $m(A) > 0$. The set of focal elements is called the *body of evidence*. A belief function with only one focal element B is called *categorical* and denoted by $\eta_{\langle B \rangle}$. Obviously,

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

Every $Bel \in M_{bel}(X)$ with the bba m is represented as a convex sum of categorical belief functions:

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

In the sequel, we will consider credal sets with a finite number of extreme points, and the set of all such credal sets on 2^X is denoted by $Cr(X)$. If $P_1,...,P_m \in \mathbf{P}$ are extreme points of $\mathbf{P} \in Cr(X)$, then every $P \in \mathbf{P}$ can be represented as a convex sum of $P_1,...,P_m$, i.e. $P = \sum_{i=1}^m a_i P_i$ for some $a_i \geqslant 0$, $i = 1,...,m$, with $\sum_{i=1}^m a_i = 1$.

In notations like $M_{pr}(X)$, $Cr(X)$, we can drop the reference set X, if its definition in the context is not necessary.

where $X = X_1 \cup X_2$, and $\mu(A) = a\mu_1(A \cap X_1) + (1-a)\mu_2(A \cap X_2)$ for all $A \in 2^X$.

3. Hartley Measure and Shannon Entropy

The *Hartley measure H* evaluates uncertainty if we only know that the unknown value of a random variable ξ is in a finite set A. Let us formulate the desirable properties of such non-specificity measure. We suppose that $H : \mathscr{P} \to [0, +\infty)$, where $\mathscr{P}$ is the family of all finite non-empty sets.

H1. Boundary condition: $H(A) = 0$ for $A \in \mathscr{P}$ iff $|A| = 1$.

H2. Monotonicity: $H(A) \leqslant H(B)$, $A, B \in \mathscr{P}$, if $A \subseteq B$.

H3. Label independency: let $\varphi : A \to B$ be a bijection between finite sets $A, B \in \mathscr{P}$, then $H(A) = H(B)$.

H4. Additivity: $H(A \times B) = H(A) + H(B)$ for $A, B \in \mathscr{P}$.

Let us discuss the above properties. Property H1 says that only in the case of exact information, when $|A| = 1$, the value of H is equal to zero. According to the Property H2, knowing that $\xi(\omega) \in A$ is more exact than $\xi(\omega) \in B$, therefore, $H(A) \leqslant H(B)$. The bijection φ in Property H3 can be conceived as giving new names to the elements of A. Thus, this renaming does not affect the uncertainty. Property H4 means the following. Assume that we have two random variables $\xi$and η, and we only know that $\xi(\omega) \in A$ and $\eta(\omega) \in B$, then we can conclude that $(\xi(\omega), \eta(\omega)) \in A \times B$. We assume that if we aggregate two independent (non-interacted) pieces of information, then uncertainty behaves additively. This implies Property 4.

Properties H2 and H4 imply the subadditivity property.

H5. Subadditivity: Let $A \in \mathscr{P}$ be such that $A \subseteq X \times Y$ and let A_X and A_Y be its projections on X and Y, respectively, i.e. $A_X = \{x \in X | \exists y \in Y : (x,y) \in A\}$ and $A_Y = \{y \in Y | \exists x \in X : (x,y) \in A\}$. Then $H(A_X) + H(A_Y) \geqslant H(A)$.

It is possible to show that every functional H with Properties H1-H4 can be represented in the form: $H(A) = c \ln |A|$, where $c > 0$. If we additionally require that $H(A) = 1$ if $|A| = 2$, then $H(A) = \log_2 |A|$.

The situation, when we only know that $\xi(\omega) \in A$, can be described by the set of all possible probability distributions of ξ that coincides with $M_{pr}(A)$, or by the categorical belief function $\eta_{\langle A \rangle}$, because, $M_{pr}(A) = \{P \in M_{pr}(X) | P \geqslant \eta_{\langle A \rangle}\}$.

Another special case is when a random variable ξ takes its values in a finite set X and we know the probability distribution of ξ. Before describing an uncertainty measure S for this case, called the Shannon entropy, we will introduce the following constructions:

a) let $\varphi : X \to Y$ and $\mu \in M_{mon}(X)$, then the image μ^φ of μ is defined by $\mu^\varphi(B) = \mu(\varphi^{-1}(B))$, where $\varphi^{-1}(B) = \{x \in X | \varphi(x) \in B\}$;

b) let $\mu \in M_{mon}(X \times Y)$, then projections μ_X and μ_Y on X and Y, respectively, are defined by $\mu_X(A) = \mu(A_X)$ and $\mu_Y(A) = \mu(A_Y)$ for every $A \in 2^{X \times Y}$.

The *Shannon entropy* is the functional $S : M_{pr} \to [0, +\infty)$ with the following properties:

S1. Boundary condition: $S(P) = 0$ for $P \in M_{pr}(X)$ iff $P = \eta_{\langle\{x\}\rangle}$ for some $x \in X$.

S2. Label independency: let $\varphi : X \to Y$ be a bijection between finite sets X and Y, and $P \in M_{pr}(X)$, then $S(P^{\varphi}) = S(P)$.

S3. Expansibility: let $\varphi : X \to Y$ be an injection such that $X \subseteq Y$ and $\varphi(x) = x$ for every $x \in X$, then $S(P^{\varphi}) = S(P)$ for every $P \in M_{pr}(X)$.

S4. Additivity: let $P \in M_{pr}(X \times Y)$, then $S(P) = \sum\limits_{y \in Y} P_Y(\{y\}) S(P_{|y}) + S(P_Y)$, where $P_{|y} \in M_{pr}(X)$ is defined by $P_{|y}(A) = P(A \times \{y\})/P_Y(\{y\})$ for every$A \in 2^X$.

S5. Subadditivity: let $P \in M_{pr}(X \times Y)$, then $S(P_X) + S(P_Y) \geqslant S(P)$.

Note that Properties S1-S2 have the same interpretation as for the Hartley measure. Property S3 means that adding dummy elements to X does not affect the value $S(P)$. Properties S2-S3 are equivalent to

S2-S3. Let $\varphi : X \to Y$ be an injection, then $S(P^{\varphi}) = S(P)$ for every $P \in M_{pr}(X)$.

Property S4 has the following interpretation. Let $P \in M_{pr}(X \times Y)$ describe the joint probability distribution of random variables ξ_X and ξ_Y with values in $X$and Y, respectively. Then Property S4 is equivalent to $H(\xi_X, \xi_Y) = H(\xi_X|\xi_Y) + H(\xi_Y)$, where $H(\xi_X, \xi_Y)$ is the entropy of the joint probability distribution of ξ_X and ξ_Y, and $H(\xi_X|\xi_Y)$ is the entropy of ξ_X given ξ_Y. The weak form of Property S4 is

S4*. $S(P_X \times P_Y) = S(P_X) + S(P_Y)$ for every $P_X \in M_{pr}(X)$ and $P_Y \in M_{pr}(Y)$, where $P = P_X \times P_Y$ on $2^{X \times Y}$is the product of P_X and P_Y defined by $P(\{(x, y)\}) = P_X(\{x\}) P_Y(\{y\})$ for all $x \in X$ and $y \in Y$.

Clearly, Property S4 has the form of Property S4* for independent random variables ξ_X and ξ_Y.

Property S5 means that the maximal uncertainty for the pair of random variables ξ_X and ξ_Y with known probability distributions is achieved, when ξ_X and ξ_Y are independent. It is possible to prove that Properties S1-S5 imply that S has the following form:

$$S(P) = -c \sum_{x \in X} P(\{x\}) \ln P(\{x\}),$$

where $P \in M_{pr}(X)$, $c > 0$, and $0 \ln 0 = 0$ by convention. This functional is uniquely defined if we define the norming condition. Taking $S(P_u) = 1$ for $P_u \in M_{pr}(X)$ with $X = \{x_1, x_2\}$ and $P_u(\{x_1\}) = P_u(\{x_2\}) = 0.5$, we get

$$S(P) = -\sum_{x \in X} P(\{x\}) \log_2 P(\{x\}).$$

4. Uncertainty Measures on Belief Functions

While modelling uncertainty by belief functions, we distinguish two types of uncertainty: *non-specificity* and *conflict*. Conflict is related to modelling uncertainty by probability measures; non-specificity comes from the possible choices of a "true" probability model among admissible ones. We also need to introduce a measure of total uncertainty that aggregates uncertainty of these two types. Therefore, we should define three functionals:

- a measure of conflict $U_C : M_{bel} \to [0, +\infty)$;
- a measure of non-specificity $U_N : M_{bel} \to [0, +\infty)$;
- a measure of total uncertainty $U_T : M_{bel} \to [0, +\infty)$.

In [5], the following system of axioms is proposed (see also a slightly different system of axioms for U_T in [17]).

B1. Boundary condition: $U_N(\mu) = 0$ for $\mu \in M_{pr}(X)$ and $U_C(\eta_{\langle B\rangle}) = 0$ for $B \in 2^X \backslash \{\emptyset\}$.

B2. Expansibility and label independency: let $\varphi : X \to Y$ be an injection and $\mu \in M_{bel}(X)$, then $U_T(\mu^{\varphi}) = U_T(\mu)$, $U_N(\mu^{\varphi}) = U_N(\mu)$, $U_C(\mu^{\varphi}) = U_C(\mu)$.

B3. Monotonicity w.r.t. mapping: let $\varphi : X \to Y$ be a mapping and $\mu \in M_{bel}(X)$, then $U_T(\mu^{\varphi}) \leqslant U_T(\mu)$.

B4. Monotonicity: let $\mu_1, \mu_2 \in M_{bel}(X)$ and $\mu_1 \leqslant \mu_2$, then $U_N(\mu_1) \geqslant U_N(\mu_2)$ and $U_T(\mu_1) \geqslant U_T(\mu_2)$.

B5. The first additivity property: let $\mu \in M_{bel}(X \times Y)$ be such that $\mu = \sum\limits_{A \in 2^X} m_X(A) \eta_{\langle A \times B\rangle}$, where $B \in 2^Y/\{\emptyset\}$, $\mu_X = \sum\limits_{A \in 2^X} m_X(A) \eta_{\langle A\rangle}$ and $\mu_Y = \eta_{\langle B\rangle}$, then

$$U_T(\mu) = U_T(\mu_X) + U_T(\mu_Y).$$

B6. The second additivity property: let $\mu \in M_{bel}(X \times Y)$ and $\mu_Y \in M_{pr}(Y)$, then

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

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

B7. Subadditivity: let $\mu \in M_{bel}(X \times Y)$, then $U_T(\mu_X) + U_T(\mu_Y) \geqslant U_T(\mu)$.

B8. Disaggregation: $U_C(\mu) + U_N(\mu) = U_T(\mu)$ for every $\mu \in M_{bel}(X)$.

Let us observe that if we consider the restriction of the above axioms for probability measures, then we get Properties S1-S5 of the Shannon entropy, analogously, the restriction of these axioms for categorical belief functions are Properties H1-H5 of the Hartley measure. Thus, $U_T(P) = U_C(P) = S(P)$ for every $P \in M_{pr}$ and $U_T(\mu) = U_N(\mu) = H(\mu)$ for every categorical belief function μ. Because the axioms B1-B8 have the same interpretation as the basic properties formulated for the Shannon entropy and the Hartley measure, we will explain only some of them.

Consider the following explanation of Axiom B3 given in [5]. Assume that $\mu \in M_{bel}(X)$ and $\varphi : X \to Y$ is such that $\varphi(x) = y_i$ if $x \in X_i$, where $\{X_1, ..., X_k\}$ is the partition of X. Assume also that $y_i \in X_i$, $i = 1, ..., k$. Thus, φ has the following interpretation: if the true alternative is in X_i, then it is y_i. Because any additional information reduces uncertainty, we should require that $U_T(\mu^\varphi) \leqslant U_T(\mu)$.

Since in Axiom B4 $\mathbf{P}(\mu_1) \supseteq \mathbf{P}(\mu_2)$, μ_1 looks as a model of uncertainty at least as with the same or higher non-specificity than μ_2. Therefore, we require that $U_N(\mu_1) \geqslant U_N(\mu_2)$ and $U_T(\mu_1) \geqslant U_T(\mu_2)$.

Axiom B5 is transformed to Property H4 for the Hartley measure, when μ_X is a categorical belief function. Formally, in Axiom B5 the *Möbuis product* $\times_M$ of μ_X and μ_Y is used, because by definition,

$$\mu = \mu_X \times_M \mu_Y = \sum_{A\in 2^X}\sum_{B\in 2^Y} m_X(A)m_Y(B)\eta_{\langle A\times B\rangle},$$

where m_X and m_Y are bbas of μ_X and μ_Y, respectively.

Axiom B6 is the generalization of Property H4. Actually, we get Property H4 taking $\mu_X \in M_{pr}(X)$. Note that the conditional belief function $\mu_{|y}$ is well justified in the theory of imprecise probabilities. It is possible equivalently to exchange Axiom B6 to

B6*. Let $\{X_1,...,X_m\}$ be a partition of X and $\mu_i \in M_{bel}(X_i)$, $i=1,...,m$. Consider $\mu = \sum_{i=1}^m a_i\mu_i$, where $\sum_{i=1}^m a_i = 1$ and $a_i \geqslant 0$, $i=1,...,m$. Then $U_T(\mu) = \sum_{i=1}^m a_i U_T(\mu_i) + U_T(P)$, where $P \in M_{pr}(Y)$ is such that $Y = \{1,...,m\}$ and $P(\{i\}) = a_i$, $i=1,...,m$.

There are several results [5], that we will use later. Axioms B1-B8 imply that

1) $U_T\left(\sum_{i=1}^m a_i\mu_i\right) \geqslant \sum_{i=1}^m a_i U_T(\mu_i)$ for every $\sum_{i=1}^m a_i = 1$, $a_i \geqslant 0$, $\mu_i \in M_{bel}(X)$, $i=1,...,m$.

2) The set $\mathfrak{F}$ of all functionals, satisfying Axioms B1-B8, is a convex cone, i.e. if $U_T^{(i)} \in \mathfrak{F}$, $i=1,2$, then $aU_T^{(1)} + bU_T^{(2)} \in \mathfrak{F}$ for any $a,b \geqslant 0$. We can impose norming conditions $U_T(\eta_{\langle X\rangle}) = a$ and $U_T(P_u) = b$, where $P_u \in M_{pr}(X)$ defines the uniform probability distribution on X and $|X| = 2$. Axiom B4 implies that $a \geqslant b$, and we need to take $a > 0$ providing U_T to be not identical zero. We denote the set of all possible total uncertainty measures with such norming conditions by $\mathfrak{F}_{a,b}$. There are two different total uncertainty measures satisfying Axioms B1-B8. It is possible to prove that $\mathfrak{F}_{a,0}$, $a>0$, is a singleton, i.e. such norming conditions uniquely define the total uncertainty measure that coincides with the generalized Hartley measure [16]:

$$GH(\mu) = \sum_{B\in 2^X} m(B)H(B),$$

where $\mu \in M_{bel}(X)$ and m is the bba of μ. The set $\mathfrak{F}_{a,a}$, $a>0$, is also non-empty, since $S_{\max} \in \mathfrak{F}_{a,a}$, where $S_{\max}$ is the maximal (upper) entropy, i.e. $S_{\max}(\mu) = \sup_{P\in\mathbf{P}(\mu)} S(P)$.

3) There are several admissible disaggregations of $U_T \in \mathfrak{F}$:

a) $U_C(\mu) = \inf_{P\in\mathbf{P}(\mu)} U_T(P)$, where U_T is the Shannon entropy on M_{pr}, $U_N = U_T - U_C$. This U_C is denoted by $S_{\min}$ and called the *minimal (lower) entropy*.

b) $U_N(\mu) = \sum_{B\in 2^X} m(B)U_T(\eta_{\langle B\rangle})$, where m is the bba of $\mu \in M_{bel}(X)$, and $U_C = U_T - U_N$. In this case, U_N is obviously the generalized Hartley measure.

5. Uncertainty Measures on Credal Sets

In [5], a reader can find a system of axioms for uncertainty measures on credal sets, that generalizes Axioms B1-B8. For describing them, we define the following operations on credal sets:

1) let $\varphi: X \to Y$ be a mapping and $\mathbf{P} \in Cr(X)$, then $\mathbf{P}^\varphi = \{P^\varphi | P \in \mathbf{P}\}$;

2) $\mathbf{P} = a\mathbf{P}_1 + (1-a)\mathbf{P}_2$ for $a \in [0,1]$ and $\mathbf{P}_1, \mathbf{P}_2 \in Cr(X)$ if $\mathbf{P} = \{aP_1 + (1-a)P_2 | P_1 \in \mathbf{P}_1, P_2 \in \mathbf{P}_2\}$;

3) let $\mathbf{P} \in Cr(X\times Y)$, then $\mathbf{P}_X = \{P_X|P \in \mathbf{P}\}$ and $\mathbf{P}_Y = \{P_Y|P \in \mathbf{P}\}$;

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

Using the above operations, we define axioms for U_T,U_C and U_N on Cr as follows [5]:

C1. Boundary condition: let $\mathbf{P} \in Cr(X)$, then $U_N(\mathbf{P}) = 0$ if $\mathbf{P} = \{P\}$, and $U_C(\mathbf{P}) = 0$ if $\mathbf{P} = \mathbf{P}(\eta_{\langle B\rangle})$ for some $B \in 2^X\backslash\{\emptyset\}$.

C2. Expansibility and label independency: 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 $\mathbf{P} \in Cr(X)$.

C3. Monotonicity w.r.t. mapping: let $\varphi: X \to Y$ be a mapping, and $\mathbf{P} \in Cr(X)$, then $U_T(\mathbf{P}) \geqslant U_T(\mathbf{P}^\varphi)$.

C4. Monotonicity: let $\mathbf{P}_1, \mathbf{P}_2 \in Cr(X)$ and $\mathbf{P}_1 \supseteq \mathbf{P}_2$, then $U_N(\mathbf{P}_1) \geqslant U_N(\mathbf{P}_2)$ and $U_T(\mathbf{P}_1) \geqslant U_T(\mathbf{P}_2)$.

C5. The first additivity property: let X, Y be non-empty finite sets, $\mathbf{P}_X = \mathbf{P}\left(\eta_{\langle A\rangle}\right)$, $A \in 2^X\backslash\{\emptyset\}$, and $\mathbf{P}_Y \in Cr(Y)$, then $U_T(\mathbf{P}_X \times_N \mathbf{P}_Y) = U_T(\mathbf{P}_X) + U_T(\mathbf{P}_Y)$.

C6. The second additivity property: let $\mathbf{P} \in Cr(X\times Y)$ and $\mathbf{P}_Y = \{P_Y\}$, then $U_T(\mathbf{P}) = \sum_{y\in Y} P_Y(\{y\})U_T\left(\mathbf{P}_{|y}\right) + U_T(\mathbf{P}_Y)$, where $\mathbf{P}_{|y} = \left\{P_{|y} | P \in \mathbf{P}\right\}$.

C7. Subadditivity: let X, Y be non-empty finite sets and $\mathbf{P} \in Cr(X\times Y)$, then $U_T(\mathbf{P}_X) + U_T(\mathbf{P}_Y) \geqslant U_T(\mathbf{P})$.

C8. Disaggregation: $U_C(\mathbf{P}) + U_N(\mathbf{P}) = U_T(\mathbf{P})$ for every $\mathbf{P} \in Cr$.

Axiom C6 can be equivalently exchanged to

C6*. Let $\{X_1,...,X_k\}$ be a partition of the set X and $\mathbf{P} = \sum_{k=1}^m a_k\mathbf{P}_k$, where $\mathbf{P}_k \in Cr(X_k)$, $a_k \geqslant 0$, $k=1,...,m$, $\sum_{k=1}^m a_k = 1$. Then

$$U_T(\mathbf{P}) = \sum_{k=1}^m a_k U_T(\mathbf{P}_k) + U_T(P),$$

where $P \in M_{pr}(Y)$ is such that $Y = \{1,...,m\}$ and $P(\{i\}) = a_i$, $i=1,...,m$.

What do we know about the possible functionals U_T and its disaggregations? One known possible choice of U_T is the maximal entropy $S_{\max}(\mathbf{P}) = \sup_{P\in\mathbf{P}} S(P)$, where $\mathbf{P} \in Cr$; and the possible disaggregation can be based on the minimal entropy $S_{\min}(\mathbf{P}) = \inf_{P\in\mathbf{P}} S(P)$, where $\mathbf{P} \in Cr$, i.e. $U_T = S_{\max}$, $U_C = S_{\min}$, and $U_N = U_T - U_C$. There are several ways to extend the generalized Hartley measure to credal sets. It should be noted that any such extension does not satisfy Axioms C1-C8, as we can see by the next example.

Example 1 Assume that $\mathbf{P} \in Cr(X \times Y)$, where $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$. Assume also that any $P \in \mathbf{P}$ is described by a point $P = (p(x_1,y_1), p(x_1,y_2), p(x_2,y_1), p(x_2,y_2))$, where $p(x_i,y_j) = P(\{(x_i,y_j)\})$, $i,j = 1,2$, and $\mathbf{P}$ has two extreme points: $P_1 = (0, 0.5, 0.5, 0)$ and $P_2 = (0.5, 0, 0, 0.5)$, i.e. $\mathbf{P} = \{aP_1 + (1-a)P_2 | a \in [0,1]\}$. We see that $\mathbf{P}_X = \{P_X\}$, where $P_X = (\underbrace{0.5}_{x_1}, \underbrace{0.5}_{x_2})$, and $\mathbf{P}_Y = \{P_Y\}$, where $P_Y = (\underbrace{0.5}_{y_1}, \underbrace{0.5}_{y_2})$. Assume that U_N satisfies the subaddivity property, i.e. $U_N(\mathbf{P}_X) + U_N(\mathbf{P}_Y) \geqslant U_N(\mathbf{P})$. Because, in our example, $U_N(\mathbf{P}_X) = U_N(\mathbf{P}_Y) = 0$, we can conclude that $U_N(\mathbf{P}) = 0$.

Let us check whether the monotonicity w.r.t. mapping is fulfilled in this case. Consider the mapping $\varphi : X \times Y \to Z$ such that $\varphi(x_1,y_1) = \varphi(x_2,y_2) = z_1$, $\varphi(x_1,y_2) = \varphi(x_2,y_1) = z_2$. Then $\mathbf{P}^\varphi = \mathbf{P}(\eta_{\langle Z\rangle})$ and $U_N(P^\varphi) > 0$ for an non-trivial U_N, but this contradicts to the monotonicity w.r.t. mapping, since we see that $U_N(\mathbf{P}^\varphi) > U_N(\mathbf{P})$.

The first known extension of GH is proposed in [1] and based on the following construction. Let $\mathbf{P} \in Cr(X)$, then the corresponding coherent lower probability is defined as

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

After that we calculate the Möbius transform [12] m of μ

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

and finally,

$$GH_1(\mu) = \sum_{B\in 2^X} m(B)H(B).$$

The proof of GH_1 monotonicity can be found in [1, 7]. Formally, GH_1 can be seen as the linear extension of H to coherent lower probabilities, and if we use GH_1 on credal sets, then we have the same result if different credal sets generate the same coherent lower probability. GH_1 is not subadditive and monotone w.r.t. mapping as follows from the next example.

Example 2 Consider the credal set from Example 1. Let us compute values of the coherent lower probability μ. For this purpose, let us denote $u_1 = (x_1,y_1)$, $u_2 = (x_1,y_2)$, $u_3 = (x_2,y_1)$, $u_4 = (x_2,y_2)$. Then $\mu(\{u_1,u_2\}) = \mu(\{u_1,u_3\}) = \mu(\{u_2,u_4\}) = \mu(\{u_3,u_4\}) = 0.5$, $\mu(\{u_1,u_2,u_3\}) = \mu(\{u_1,u_2,u_4\}) = \mu(\{u_1,u_3,u_4\}) = \mu(\{u_2,u_3,u_4\}) = 0.5$, $\mu(\{u_1,u_2,u_3,u_4\}) = 1$. μ is equal to zero on other sets in $2^{X\times Y}$. The computation of m results in $m(\{u_1,u_2\}) = m(\{u_1,u_3\}) = m(\{u_2,u_4\}) = m(\{u_3,u_4\}) = 0.5$, $m(\{u_1,u_2,u_3\}) = m(\{u_1,u_2,u_4\}) = m(\{u_1,u_3,u_4\}) = m(\{u_2,u_3,u_4\}) = -0.5$, $m(\{u_1,u_2,u_3,u_4\}) = 1$. m is equal to zero on other sets in $2^{X\times Y}$. Assume that $H(B) = \log_2|B|$ for every $B \neq \emptyset$. Then $GH_1(\mu) = 2 - 2\log_2 3 + 2 \approx 0.83$. We see that both properties C3 and C7 are not fulfilled.

The second extension, introduced in [5], is based on the inner approximation of GH:

$$GH_2(\mathbf{P}) = \sup\{GH(\mu) | \mathbf{P}(\mu) \subseteq \mathbf{P}, \mu \in M_{bel}(X)\}.$$

In [5], a reader can find the proof that GH_2 can be used for disaggregation of U_T, and it is subadditive. Formally, we can also check the first and the second additivity properties defined as

The first additivity property: let X, Y be non-empty finite sets, $\mathbf{P}_X = \mathbf{P}(\eta_{\langle A\rangle})$, $A \in 2^X\setminus\{\emptyset\}$, and $\mathbf{P}_Y \in Cr(Y)$, then $U_N(\mathbf{P}_X \times_N \mathbf{P}_Y) = U_N(\mathbf{P}_X) + U_N(\mathbf{P}_Y)$.

The second additivity property: let $\{X_1, ..., X_k\}$ be a partition of the set X and $\mathbf{P} = \sum_{i=1}^k a_i \mathbf{P}_i$, where $\mathbf{P}_i \in Cr(X_i)$, $a_i \geqslant 0$, $i = 1, ..., k$, $\sum_{i=1}^k a_i = 1$. Then

$$U_N(\mathbf{P}) = \sum_{i=1}^{k} a_i U_N(\mathbf{P}_i).$$

Clearly, the last property is the counterpart of C6*, in which we drop the term $U_N(P) = 0$, $P \in M_{pr}(Y)$.

Remark 1 In [5] a reader can find results that GH_2 obeys the first and second additivity properties. Obviously, it can be used for disaggregation of a measure of total uncertainty. In the next section, we will check these properties for GH_1.

6. Properties of GH_1

Theorem 1 *GH_1 obeys the first additivity property.*

Theorem 2 *GH_1 obeys the second additivity property.*

Lemma 1 *Let $P \in M_{pr}(X)$ and $P(\{x\}) > 0$ for all $x \in X$. Then the largest credal set $\mathbf{P} \in Cr(X)$ such that $P \in \mathbf{P}$ and $S_{\max}(\mathbf{P}) = S(P)$ is $\mathbf{P} = \{Q \in M_{pr}(X) | E_Q(f_S) \geqslant E_P(f_S)\}$, where $f_S(x) = \ln P(\{x\})$, $x \in X$.*

Remark 2 Consider how we can generalize Lemma 1 for the case, when P takes values equal to zero on some singletons. Assume that $P(\{x\}) = 0$ for some $x \in X$ and $Q(\{x\}) > 0$, then $\lim_{a\to+0} S(aQ + (1-a)P) = +\infty$, i.e. $P(\{x\}) = 0$ implies that $Q(\{x\}) = 0$ for every $Q \in \mathbf{P}$, i.e. we can reduce this problem considering only those elements of X, where $P(\{x\}) > 0$ like in Lemma 1.

Remark 3 Let $X = \{x_1,...,x_n\}$, then the special case of Lemma 1 is when $P(\{x_i\}) = 1/n$, $i = 1,...,n$. In this case, $E_Q(f) - E_P(f) = \sum_{x \in X} (Q(\{x\}) - P(\{x\}) \ln(1/n) = 0$, for every $Q \in M_{pr}(X)$, i.e. $\mathbf{P} = M_{pr}(X)$. Note that for other cases P is a boundary point of $\mathbf{P}$.

The next proposition shows the construction of coherent lower probabilities μ on 2^X with $P \in \mathbf{P}(\mu)$ and $S_{\max}(\mathbf{P}(\mu)) = S(P)$ for a given $P \in M_{pr}(X)$.

Proposition 1 *Let $P \in M_{pr}(X)$ such that $P(\{x\}) > 0$ for all $x \in X$. Consider a subset $\mathscr{A} \subseteq 2^X$ and a credal set defined by*

$$\mathbf{P} = \{Q \in M_{pr}(X) | \forall A \in \mathscr{A} : Q(A) \geqslant P(A)\}, \qquad (1)$$

Introduce the corresponding coherent lower probability μ defined by $\mu(A) = \inf\{Q(A) | Q \in \mathbf{P}\}$, $A \in 2^X$. Then $\mathbf{P}(\mu) = \mathbf{P}$, $P \in \mathbf{P}(\mu)$, and $S_{\max}(\mathbf{P}(\mu)) = S(P)$ iff there are $a_A \geqslant 0$, $A \in \mathscr{A}$, and $b \in \mathbb{R}$, such that $\sum_{A \in \mathscr{A}} a_A 1_A + b 1_X = f_S$, where f_S is defined like in Lemma 1.

Corollary 1 *Let $P \in M_{pr}(X)$, $P(\{x\}) > 0$ for all $x \in X$, and $\{P(\{x\}) | x \in X\} = \{a_1,...,a_k\}$, where values a_i are indexed such that $a_1 > a_2 > ... > a_k > 0$. Define a credal set $\mathbf{P}$ by formula (1), where and $A_i = \{x \in X | P(\{x\}) \geqslant a_i\}$, $i = 1,...,k$. Then $P \in \mathbf{P}$ and $S_{\max}(\mathbf{P}(\mu)) = S(P)$.*

Remark 4 It is well known that the credal set from Corollary 1 can be generated by the $\mu \in M_{bel}(X)$ whose body of evidence is $\mathscr{A}$ and the corresponding bba m is defined as $m(A_i) = a_i - a_{i+1}$, $i = 1,...,k$, where $a_{k+1} = 0$ by convention. Such a μ is a necessity measure, since focal elements of μ are linearly ordered w.r.t. the inclusion relation.

Example 3 Let $X = \{x_1,x_2,x_3,x_4\}$ and $P \in M_{pr}(X)$ is defined by $P(\{x_i\}) = q^i$, $i = 1,...,4$, where q is the positive root of the equation $q^4 + q^3 + q^2 + q - 1 = 0$ ($q = 0.518...$). Let us construct μ like in Corollary 1. Then $\mu \in M_{bel}(X)$ whose body of evidence is $\mathscr{A} = \{A_i\}_{i=1}^4$, where $A_i = \{x_1,...,x_i\}$, and $m(A_i) = q^i - q^{i+1}$, $i = 1,...,3$, $m(A_4) = q^4$. Let us construct μ_1 like in Proposition 1 by the system of sets $\mathscr{A}_1 = \{\{x_1,x_2\},\{x_1,x_3\}\}$. Let us show that there is a representation

$$y_1 1_{\{x_1,x_2\}} + y_2 1_{\{x_1,x_3\}} + y_3 1_{\{x_1,x_2,x_3,x_4\}} = f_S, \qquad (2)$$

where $y_1, y_2 \geqslant 0$, and $y_3 \in \mathbb{R}$. We see that (2) is equivalent to the following system of linear equations:

$$\begin{cases} y_1 + y_2 + y_3 = \ln q, \\ y_1 + y_3 = 2 \ln q, \\ y_2 + y_3 = 3 \ln q, \\ y_3 = 4 \ln q. \end{cases}$$

The solution of this system is $y_1 = -2\ln q$, $y_2 = -\ln q$, $y_3 = 4 \ln q$, and y_i, $i = 1,...,3$, satisfy the required conditions. The values of μ and μ_1 are given in Table 1.

Table 1: Monotone measures for Example 3

x_1	x_2	x_3	x_4	μ	μ_1	μ_2
1	0	0	0	q	$q-q^4$	$q-q^4$
1	1	0	0	$q+q^2$	$q+q^2$	$q+q^2$
1	0	1	0	q	$q+q^3$	q
1	1	1	0	$q+q^2+q^3$	$q+q^2$	$q+q^2$
1	0	0	1	q	$q-q^4$	$q-q^4$
1	1	0	1	$q+q^2$	$q+q^2$	$q+q^2$
1	0	1	1	q	$q+q^3$	q
1	1	1	1	1	1	1

By Proposition 1, $P \in \mathbf{P}(\mu)$, $P \in \mathbf{P}(\mu_1)$, and $S_{\max}(\mathbf{P}(\mu)) = S_{\max}(\mathbf{P}(\mu_1)) = S(P)$. Note that the set $\mathscr{M} = \{\mu \in M_{coh}(X) | P \in \mathbf{P}(\mu), S_{\max}(\mathbf{P}(\mu)) = S(P)\}$ does not contain the smallest element in general. To show this consider a coherent lower probability μ_2, defined by $\mu_2(A) = \min\{\mu(A), \mu_1(A)\}$, $A \in 2^X$. The values of μ_2 are also given in Table 1. Consider a probability measure P_1 with values $P_1(\{x_1\}) = q - q^4$, $P_1(\{x_2\}) = q^2 + q^4$, $P_1(\{x_3\}) = P_1(\{x_4\}) = (q^3 + q^4)/2$. It is easy to check that $P_1 \in \mathbf{P}(\mu_2)$,

$$S(P) = -(q + 2q^2 + 3q^3 + 4q^4) \ln q = 1.158...,$$
$$S(P_1) = -(q - q^4)\ln(q - q^4) - (q^2 + q^4)\ln(q^2 + q^4) - (q^3 + q^4)\ln \frac{q^3+q^4}{2} = 1.202....$$

therefore, $\mu_2 \notin \mathscr{M}$.

Proposition 2 *Let $P \in M_{pr}(X)$ with $P(\{x\}) > 0$ for all $x \in X$ and*

$$\mathscr{M} = \{\nu \in M_{2-mon}(X) | P \in \mathbf{P}(\nu), S_{\max}(\mathbf{P}(\nu)) = S(P)\}.$$

Consider a $\mu \in \mathscr{M}$ constructed like in Corollary 1. Then $\nu \geqslant \mu$ for every $\nu \in \mathscr{M}$.

Remark 5 Note the result formulated in Proposition 2 does not contradict Example 3, where we construct μ_1 such that $\mu_1 \ngeqslant \mu$, since $\mu_1 \notin M_{2-mon}(X)$. This can be seen from the inequality:

$$\mu_1(\{x_1,x_2\}) + \mu_1(\{x_1,x_3\}) = 2q + q^2 + q^3 >$$
$$\mu_1(\{x_1\}) + \mu_1(\{x_1,x_2,x_3\}) = 2q + q^2 - q^4.$$

The above results can be seen as the investigation of the inner approximation of the credal set defined in Lemma 1 by coherent lower probabilities. The next proposition describes the upper approximation of this credal set.

Proposition 3 *Let $X = \{x_1,...,x_n\}$ and $P \in M_{pr}(X)$ such that $P(\{x_1\}) \geqslant P(\{x_2\}) \geqslant ... \geqslant P(\{x_n\}) > 0$. Consider the corresponding credal set $\mathbf{P} = \{Q \in M_{pr}(X) | E_Q(f_S) \geqslant E_P(f_S)\}$ from Lemma 1 and a coherent lower probability μ on 2^X defined by $\mu(A) = \inf\{P(A) | P \in \mathbf{P}\}$, where $A \in 2^X$. Then*

1) $\mu(A) = 0$ if $x_1 \notin A$;

2) let $x_1 \in A$, $A \neq X$, and $l = \min\{i | x_i \in A^c\}$, then $\mu(A) = (E_P(f_S) - f_S(x_l))/(f_S(x_1) - f_S(x_l))$ if $E_P(f_S) - f_S(x_l) > 0$ and $\mu(A) = 0$ otherwise.

Corollary 2 *Let $P \in M_{pr}(X)$, $P(\{x\}) > 0$ for all $x \in X$, and $\{P(\{x\}) | x \in X\} = \{a_1, ..., a_k\}$, where values a_i are indexed such that $a_1 > a_2 > ... > a_k > 0$. Let $\mathbf{P}$ be the credal set defined in Lemma 1 and a coherent lower probability μ on 2^X defined by $\mu(A) = \inf\{Q(A) | Q \in \mathbf{P}\}$. Then*

$$\mathbf{P} = \{Q \in M_{pr}(X) | \forall A \in \mathscr{A} : Q(A) \geqslant \mu(A)\},$$

where $\mathscr{A} = \{A_i\}_{i=1}^k$ and $A_i = \{x \in X | P(\{x\}) \geqslant a_i\}$, $i = 1, ..., k$.

Remark 6 Since the set $\mathscr{A}$ defined in Corollary 2 is linearly ordered by the inclusion relation, the measure μ is a necessity measure (a consonant belief function). Let us compute at first values of μ on $\mathscr{A}$:

1) $\mu(A_i) = 0$ if $E_P(f_S) - f_S(a_{i+1}) \leqslant 0$;

2) $\mu(A_i) = (E_P(f_S) - f_S(a_{i+1}))/(f_S(a_1) - f_S(a_{i+1}))$ if $E_P(f_S) - f_S(a_{i+1}) > 0$ and $i \neq k$

3) $\mu(A_k) = 1$, since $A_k = X$.

We can compute the corresponding bba m on by $m(A_i) = \mu(A_i) - \mu(A_{i-1})$, $i = 1, ..., k$, where $A_0 = \emptyset$ by convention. Using the above result, we can compute $GH(\mu)$ by the formula:

$$GH(\mu) = \sum_{i=1}^{k} (\mu(A_i) - \mu(A_{i-1})) \ln |A_i| =$$
$$\sum_{i=1}^{k-1} \mu(A_i) (\ln |A_i| - \ln |A_{i+1}|) + \mu(A_k) \ln |A_k| .$$

Let $m(A_k) \neq 1$ and $j = \min\{i \in \{1, ..., k-1\} | E_P(f_S) - f_S(a_{i+1}) > 0\}$, then

$$GH(\mu) = \sum_{i=j}^{k-1} \frac{(E_P(f_S) - \ln a_{i+1}) (\ln |A_i| - \ln |A_{i+1}|)}{\ln a_1 - \ln a_{i+1}} + \ln |A_k| .$$

Since $\frac{E_P(f_S) - \ln a_{i+1}}{\ln a_1 - \ln a_{i+1}} = 1 - \frac{\ln a_1 - E_P(f_S)}{\ln a_1 - \ln a_{i+1}}$ and $\sum_{i=j}^{k-1} (\ln |A_i| - \ln |A_{i+1}|) = \ln |A_j| - \ln |A_k|$, we get

$$GH(\mu) = \ln |A_j| + \sum_{i=j}^{k-1} \frac{(\ln a_1 - E_P(f_S)) (\ln |A_{i+1}| - \ln |A_i|)}{\ln a_1 - \ln a_{i+1}} .$$

In the sequel, we will represent $GH(\mu)$ as $GH(\mu) = \ln |A_j| + (\ln a_1 - E_P(f_S)) F(\mathbf{a})$, where

$$F(\mathbf{a}) = \sum_{i=j}^{k-1} \frac{\ln |A_{i+1}| - \ln |A_i|}{\ln a_1 - \ln a_{i+1}} .$$

Example 4 Let us construct an example of a $\mathbf{P} \in Cr(X)$ for which $GH_1(\mathbf{P}) > S_{\max}(\mathbf{P})$. Assume that $\mathbf{P}$ is constructed like in Lemma 1, i.e. there is a $P \in M_{pr}(X)$ with $P(\{x\}) > 0$ for all $x \in X$ and such that $\mathbf{P} = \{Q \in M_{pr}(X) | E_Q(f_S) \geqslant E_P(f_S)\}$. In our example we will assume that $A_1 = \{x_1\}$, $A_2 = \{x_1, ..., x_m\}$, $A_3 = X = \{x_1, ..., x_n\}$, $a_1 = q^2 a_3$ and $a_2 = q a_3$. In addition, these parameters are chosen such that $E_P(f_S) - \ln(a_2) = 0$. Then

$$E_P(f_S) - \ln(a_2) = a_1 \ln (a_1/a_2) + a_3 (n-m) \ln (a_3/a_2) = 0,$$

or $a_3 q^2 \ln q - (n-m) a_3 \ln q = 0$. Thus, we can choose $q = \sqrt{n-m}$ and the norming condition implies that $(n-m)a_3 + (m-1)\sqrt{n-m} a_3 + (n-m)a_3 = 1$, i.e.

$$a_3 = \frac{1}{2(n-m) + (m-1)\sqrt{n-m}},$$

$$F(\mathbf{a}) = \frac{\ln |A_3| - \ln |A_2|}{\ln a_1 - \ln a_3} = \frac{\ln n - \ln m}{\ln (n-m)},$$

$$GH(\mu) = \ln |A_2| + (\ln a_1 - E_P(f_S)) F(\mathbf{a}) =$$
$$\ln m + (\ln a_1 - \ln a_2) F(\mathbf{a}) =$$
$$\ln m + 0.5 \ln (n-m) \frac{\ln n - \ln m}{\ln (n-m)} = 0.5 (\ln n + \ln m),$$

$$a_2 = \frac{\sqrt{n-m}}{2(n-m) + (m-1)\sqrt{n-m}} = \frac{1}{2\sqrt{n-m} + (m-1)},$$

$$S(P) = -E_P(f_S) = -\ln a_2 = \ln (2\sqrt{n-m} + m - 1),$$

We see that

$$S(P) - GH(\mu) = \ln \left(\frac{2\sqrt{n-m} + m - 1}{\sqrt{mn}} \right).$$

Let us denote $x = m/n$, then

$$\lim_{n \to \infty} (S(P) - GH(\mu)) = \lim_{n \to \infty} \ln \left(\frac{2\sqrt{n}\sqrt{1-x} + nx - 1}{n\sqrt{x}} \right) =$$
$$0.5 \ln x < 0,$$

i.e. $S(P) - GH(\mu) < 0$ if n is sufficiently large. In particular, if $n = 100$ and $m = 36$, then $GH_1(\mathbf{P}) = GH(\mu) = \ln 60$, $S_{\max}(\mathbf{P}) = S(P) = \ln 51$, and we see that $GH_1(\mathbf{P}) > S_{\max}(\mathbf{P})$.

7. The Hartley Measure on Credal Sets

Assume that uncertainty is described by a probability measure $P \in M_{pr}(X)$ and the choice of the optimal decision is based on the expected utility, i.e. every decision is described by a function $f : X \to \mathbb{R}$ and the expected utility is defined as

$$E_P(f) = \sum_{x \in X} f(x) P(\{x\}).$$

If uncertainty is described by a credal set $\mathbf{P}$, then we only know the lower and upper bounds of expected utility defined by

$$\underline{E}_{\mathbf{P}}(f) = \inf_{P\in\mathbf{P}} E_P(f),\ \overline{E}_{\mathbf{P}}(f) = \sup_{P\in\mathbf{P}} E_P(f).$$

In the case of total uncertainty, when $\mathbf{P} = M_{pr}(X)$, we have

$$\underline{E}_{\mathbf{P}}(f) = \min_{x\in X} f(x),\ \overline{E}_{\mathbf{P}}(f) = \max_{x\in X} f(x).$$

We call probability measures $P_1, P_2 \in M_{pr}(X)$ fully contradictory (inconsistent) if there is a $f : X \to \mathbb{R}$ with $\min_{x\in X} f(x) < \max_{x\in X} f(x)$ such that $E_{P_1}(f) = \min_{x\in X} f(x)$ and $E_{P_2}(f) = \max_{x\in X} f(x)$.

Lemma 2 *$P_1, P_2 \in M_{pr}(X)$ are fully contradictory iff there is a $A \in 2^X$ such that $P_1(A) = P_2(A^c) = 0$.*

The next lemma shows that we can check the full contradiction of probability measures using the metric on $M_{pr}(X)$ defined by

$$d(P_1,P_2) = \sup_{A\in 2^X} |P_1(A) - P_2(A)|.$$

Lemma 3 *Probability measures $P_1, P_2 \in M_{pr}(X)$ are fully contradictory iff $d(P_1,P_2) = 1$.*

Remark 7 The distance d between probability measures plays an important role in measuring contradiction (conflict) between sources of information described by credal sets. A reader can find an important information about this in [6, 8, 9, 10, 11].

In this paper, we will compute the largest number of pairwise fully contradictory probability measures in a credal set, and this allows us to introduce the Hartley measure defined on credal sets.

Proposition 4 *Consider probability measures $P_1,\ldots,P_k \in M_{pr}(X)$ and sets $A_i = \{x \in X | P_i(\{x\}) > 0\}$, $i = 1,\ldots,k$. Then these probability measures are pairwise fully contradictory iff $A_i \cap A_j = \emptyset$ for any $i \neq j$.*

Proposition 5 *Let $\mathbf{P}$ be a closed and convex credal set with a finite number of extreme points. Then the largest system of pairwise fully contradictory probability measures can be chosen among extreme points of $\mathbf{P}$.*

Remark 8 Assume that $P_1,\ldots,P_k$ are the extreme points of a credal set $\mathbf{P}$ with the corresponding sets A_i, $i = 1,\ldots,k$ defined like in Proposition 4. Then the problem of finding the maximal system of pairwise fully contradictory probability measures in $\mathbf{P}$ can be reformulated in terms of graph theory. Consider the undirected graph G, whose vertices are sets A_i, $i = 1,\ldots,k$, and there is an edge between A_i and A_j, $i \neq j$, iff $A_i \cap A_j \neq \emptyset$. A set of vertices is called independent if every two vertices in it are not adjacent. Thus, the problem of finding the maximal system of pairwise fully contradictory probability measures is equivalent to the problem of finding the independent set in G with the largest cardinality.

Remark 9 We can conclude from Remark 8 that the search of the maximal system of pairwise fully contradictory probability measures is NP-hard, however, for special cases we have the solutions. For example, if $\mathbf{P} = \mathbf{P}(\eta_{\langle A\rangle})$, then the extreme points of this set are probability measures $\eta_{\langle\{x\}\rangle}$, $x \in A$, which are pairwise fully contradictory, and the cardinality of this system is $|A|$.

Based on Remark 9, we introduce the following definition.

Definition 1 Let $\mathbf{P} \in Cr(X)$, then the Hartley measure $H(\mathbf{P}) = \ln(n(\mathbf{P}))$, where $n(\mathbf{P})$ is the largest number of pairwise fully contradictory probability measures in $\mathbf{P}$.

Lemma 4 *Let $\mathbf{P} \in Cr(X)$, then there is a credal set $\mathbf{P}_1 \subseteq \mathbf{P}$ and a mapping $\varphi : X \to Y$ such that $\mathbf{P}_1^{\varphi} = \mathbf{P}(\eta_{\langle Y\rangle})$ and $H(\mathbf{P}) = \ln|Y|$.*

The main properties of H are given in the following proposition.

Proposition 6 *The Hartley measure H on Cr has the following properties:*

1) $H(\mathbf{P}(\eta_{\langle A\rangle})) = \ln|A|$ for every $A \neq \emptyset$;

2) $H(\mathbf{P}) = 0$ if $\mathbf{P} = \{P\}$, where $P \in M_{pr}$;

3) let $\mathbf{P} \in Cr(X)$ and $\varphi : X \to Y$, then $H(\mathbf{P}^{\varphi}) \leqslant H(\mathbf{P})$; in addition, $H(\mathbf{P}^{\varphi}) = H(\mathbf{P})$ if φ is an injection;

4) let $\mathbf{P}_1, \mathbf{P}_2 \in Cr(X)$ and $\mathbf{P}_1 \subseteq \mathbf{P}_2$, then $H(\mathbf{P}_1) \leqslant H(\mathbf{P}_2)$;

5) let X, Y be non-empty finite sets, $\mathbf{P}_X = \mathbf{P}\left(\eta_{\langle A\rangle}\right)$, $A \in 2^X \backslash \{\emptyset\}$, and $\mathbf{P}_Y \in Cr(Y)$, then $H(\mathbf{P}_X \times_N \mathbf{P}_Y) \geqslant H(\mathbf{P}_X) + H(\mathbf{P}_Y)$;

6) let $\{X_1, X_2\}$ be a partition of the set X and $\mathbf{P} = a\mathbf{P}_1 + (1-a)\mathbf{P}_2$, where $\mathbf{P}_i \in Cr(X_i)$, $i = 1,2$, and $a \in (0,1)$, then $U_N(\mathbf{P}) = \min\{U_N(\mathbf{P}_1), U_N(\mathbf{P}_2)\}$.

7) $H(\mathbf{P}) \leqslant S_{\max}(\mathbf{P})$ for every $\mathbf{P} \in Cr$.

Remark 10 Proposition 6 implies that H can be served as a non-specificity measure for $S_{\max}$ disaggregation, however, it does not obey the second additivity property according to the statement 6) of Proposition 6, and as follows from the next examples it is not subadditive, and it does not obey the first additivity property.

Example 5 Consider the credal set $\mathbf{P} \in Cr(X \times Y)$ from Example 1. We see that $\mathbf{P}_X = \{P_X\}$ and $\mathbf{P}_Y = \{P_Y\}$, i.e. $H(\mathbf{P}_X) = H(\mathbf{P}_Y) = 0$, and $\mathbf{P}$ has only two extreme points $P_1 = (0, 0.5, 0.5, 0)$ and $P_2 = (0.5, 0, 0, 0.5)$. Because P_1 and P_2 are fully contradictory, $H(\mathbf{P}) = \ln 2$, and $H(\mathbf{P}) > H(\mathbf{P}_X) + H(\mathbf{P}_Y)$.

Example 6 Let $\mathbf{P} = \mathbf{P}_X \times_N \mathbf{P}_Y$, where $X = \{x_1, x_2\}$, $Y = \{y_1, y_2, y_3\}$, $\mathbf{P}_X = \mathbf{P}(\eta_{\langle X\rangle})$ and $\mathbf{P}_Y \in Cr(Y)$, defined by extreme points $P_Y^{(i)}$, $i = 1,2,3$, with probabilities: $P_Y^{(1)}(\{y_1\}) = P_Y^{(1)}(\{y_2\}) = 0.5$, $P_Y^{(2)}(\{y_2\}) = P_Y^{(2)}(\{y_3\}) = 0.5$, $P_Y^{(3)}(\{y_1\}) = P_Y^{(3)}(\{y_3\}) = 0.5$. Consider probability measures $P^{(i)} \in \mathbf{P}$, $i = 1,2,3$, defined by

Table 2: Properties of non-specificity measures.

Properties $\setminus U_N$	$S_{\max}-S_{\min}$	GH_1	GH_2	H
Boundary conditions	+	+	+	+
Monotonicity	+	+	+	+
Monotonicity w.r.t. mapping	-	-	-	+
The first additivity property	+	+	+	-
The second additivity property	+	+	+	-
Subadditivity	-	-	+	-
Weak sensitivity	+	+	-	-
Strong sensitivity	-	-	-	-
Disaggregation	+	-	+	+

$$P^{(1)}(\{(x_1,y_1)\}) = P^{(1)}(\{(x_1,y_2)\}) = 0.5,$$
$$P^{(2)}(\{(x_2,y_2)\}) = P^{(2)}(\{(x_1,y_3)\}) = 0.5,$$
$$P^{(3)}(\{(x_2,y_1)\}) = P^{(3)}(\{(x_2,y_3)\}) = 0.5.$$

Since probability measures $P^{(i)} \in \mathbf{P}$, $i = 1,2,3$, are pairwise fully contradictory, we conclude that $H(\mathbf{P}) \geqslant \ln 3$. We see that $H(\mathbf{P}_X) = \ln 2$ and $H(\mathbf{P}_Y) = \ln 1 = 0$. Therefore, $H(\mathbf{P}) > H(\mathbf{P}_X) + H(\mathbf{P}_Y)$, i.e. the first additivity property is not fulfilled.

8. Discussion and Conclusion

In previous sections, we have analyzed properties of non-specificity measures. Now we are ready to present the obtained results in Table 2. In Table 2, you can see also properties that characterize sensitivity of non-specificity measures formulated as follows:

Weak sensitivity: let $\mathbf{P} \in Cr(X)$, then $U_N(\mathbf{P}) = 0$ iff $\mathbf{P} = \{P\}$, where $P \in M_{pr}(X)$.

Strong sensitivity: let $\mathbf{P}_1, \mathbf{P}_2 \in Cr(X)$ and $\mathbf{P}_1 \subset \mathbf{P}_2$, then $U_N(\mathbf{P}_1) < U_N(\mathbf{P}_2)$.

A reader can check that $U_N = S_{\max} - S_{\min}$ is not subadditive using the credal set from Example 1. The monotonicity w.r.t. mapping is not fulfilled for $U_N = S_{\max} - S_{\min}$ as shown in the next example.

Example 7 Assume that $X = \{x_1,x_2,x_3\}$ and $\mu = 0.5\eta_{\langle\{x_1,x_2\}\rangle} + 0.5\eta_{\langle\{x_3\}\rangle}$. Clearly, $S_{\max}(\mu) = -0.5\ln 0.25 - 0.5\ln 0.5 = 1.5\ln 2$, $S_{\min}(\mu) = \ln 2$. Therefore, $S_{\max}(\mu) - S_{\min}(\mu) = 0.5\ln 2$. Consider a mapping $\varphi : X \to X$ defined by $\varphi(x_i) = \begin{cases} x_i, & i = 1,2, \\ x_2, & i = 3. \end{cases}$ Then $\mu^{\varphi} = 0.5\eta_{\langle\{x_1,x_2\}\rangle} + 0.5\eta_{\langle\{x_2\}\rangle}$, $S_{\max}(\mu^{\varphi}) = \ln 2$, $S_{\min}(\mu^{\varphi}) = 0$, i.e. $U_N(\mu^{\varphi}) > U_N(\mu)$ if $U_N = S_{\max} - S_{\min}$.

Based on Example 1, we see that all desirable properties of non-specificity measures cannot be fulfilled simultaneously, for example, if U_N is subadditive, then it is not monotone w.r.t. mapping and weakly sensitive. In opinion, the sensitivity and monotonicity w.r.t. mapping of a non-specificity measure has a higher importance, than its subadditivity. It is possible to increase sensitivity of the Hartley measure on credal sets introducing ε-Hartley measures. The idea consists in the following. Probability measures $P_1, P_2 \in M_{pr}(X)$ are called ε-contradictory for $\varepsilon \in (0,1]$ if $d(P_1,P_2) \geqslant \varepsilon$. Then $H_{\varepsilon}(\mathbf{P})$, where $\mathbf{P} \in Cr(X)$ is the logarithm of the largest number of pairwise ε-contradictory measures in $\mathbf{P}$. However, the investigation of these measures is not included in this paper.

It is possible to use the above non-specificity measures on coherent lower probabilities or 2-monotone measures. At the first glance, GH_1 on 2-monotone measures looks optimal. This may be the topic for further research.

References

[1] J. Abellán, S. Moral. A non-specificity measure for convex sets of probability distributions. *Int. J. Uncertain. Fuzz. Knowl.-Based Systems*, 8 (3):357–367, 2000.

[2] J. Abellán, S. Moral. Maximum entropy for credal sets. *Int. J. Uncertain. Fuzz. Knowl.-Based Systems*, 11 (5): 587—597, 2003.

[3] J. Abellán, S. Moral. Difference of entropies as a non-specificity function on credal sets. *Int. J. Gen. Systems*, 34 (3): 203–217, 2005.

[4] T. Augustin, F.P.A. Coolen, G. de Cooman, M. C. M. Troffaes (Eds.). *Introduction to Imprecise Probabilities*, Wiley, New York, 2014.

[5] A.G. Bronevich, G. J. Klir. Measures of uncertainty for imprecise probabilities: an axiomatic approach. *Int. J. Approx. Reason.*, 51(4): 365–390, 2010.

[6] A.G. Bronevich, I.N. Rozenberg. The choice of generalized Dempster-Shafer rules for aggregating belief functions. *Int. J. Approx. Reason.*, 56: 122–136, 2015.

[7] A.G. Bronevich, A.E. Lepskiy. Imprecision indices: axiomatic, properties and applications. *Int. J. Gen. Systems*, 44(7-8): 812–832, 2015.

[8] A.G. Bronevich, I.N. Rozenberg. Incoherence correction and decision making based on generalized credal sets. In: A. Antonucci, L. Cholvy, O. Papini (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 14th European Conference, ECSQARU 2017, Lugano, Switzerland, in: Proc., Lecture Notes in Artificial Intelligence, vol.10369, Springer International Publishing, Cham, pp.271–281, 2017.

[9] A.G. Bronevich, I.N. Rozenberg. Modelling uncertainty with generalized credal sets: application to conjunction and decision.*Int. J. Gen. Syst.*, 47 (1): 67—96, 2018.

[10] A.G. Bronevich, I.N. Rozenberg. The contradiction between belief functions: Its description, measurement, and correction based on generalized credal sets. *Int. J. Approx. Reason.*, 112: 119–139, 2019.

[11] A.G. Bronevich, I.N. Rozenberg. Metrical approach to measuring uncertainty. In: M.-J. Lesot, S. Vieira, M.Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier, R.R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer International Publishing, Cham, pp. 124–136, 2020.

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

[13] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5: 131–295, 1953.

[14] A. P. Dempster. Upper and lower probabilities induced by multivalued mapping. *Ann. Math. Statist.*, 38: 325–339, 1967.

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

[16] D. Dubois, H. Prade. A note on measures of specificity for fuzzy sets. *Int. J. Gen. Systems*, 10 (4): 279–283, 1985.

[17] D. Harmanec. Toward a characterization of uncertainty measure for the Dempster–Shafer theory. In: Proc. of the Eleventh International, Conference on Uncertainty in Artificial Intelligence, Montreal, Canada, pp. 255–261, 1995.

[18] R.V.L. Hartley. Transmission of information. *Bell System Technical J.*, 7 (3): 535—563, 1928.

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

[20] A. Rényi. On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, Contributions to the Theory of Statistics, University of California Press, Berkeley, Calif., pp. 547–561, 1961.

[21] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, N.J., 1976.

[22] E. Shannon. The mathematical theory of communication. *Bell System Technical J.*, 27 (3,4): 379–423, 623–656, 1948.

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