International Journal of General Systems, Vol. 35, No. 1, February 2006, 29–44

Disaggregated total uncertainty measure for credal sets

J. ABELLÁN†*, G.J. KLIR‡¶ and S. MORAL†*

†Department of Computation Science and A.I., University of Granada, Granada 18071, Spain
‡Department of Systems Science and Industrial Engineering, Binghamton University (SUNY), Binghamton, NY 13902 6000, USA

(Received 19 May 2005; in final form 8 November 2005)

We present a new approach to measure uncertainty/information applicable to theories based on convex sets of probability distributions, also called credal sets. A definition of a total disaggregated uncertainty measure on credal sets is proposed in this paper motivated by recent outcomes. This definition is based on the upper and lower values of Shannon's entropy for a credal set. We justify the use of the proposed total uncertainty measure and the parts into which it is divided: the maximum difference of entropies, which can be used as a non-specificity measure (imprecision), and the minimum of entropy, which represents a measure of conflict (contradiction).

Keywords: Imprecise probabilities; Credal sets; Lower probabilities; Order-2 capacities; Theory of evidence; Uncertainty based information

1. Introduction

There are two classical theories of uncertainty, both formalized in terms of classical set theory. The older one, which is also simpler and more fundamental, is based on the notion of possibility. The newer one, which has been considerably more visible, is based on the notion of probability (Klir 2006).

In classical possibility theory, uncertainty means that among all conceived and mutually exclusive alternatives in a given situation (predictions, retrodictions, diagnoses, crime suspects, etc.) more alternatives than one are possible. The larger the set of possible alternatives, the less specific is the identification of the true alternative. It is thus common to refer to this type of uncertainty as non-specificity.

Information is gained, and non-specificity reduced, by obtaining any evidence that allows us to reduce the set of possible alternatives. When only one alternative remains possible, we have no uncertainty. However, when none of the alternatives remains possible, then, clearly, the set of considered alternatives is not sufficient to identify the true alternative and must be properly extended.

*Corresponding author. Email: {jabellan,smc}@decsai.ugr.es
¶ Email: gklir@binghamton.edu

ISSN 0308-1079 print/ISSN 1563-5104 online
http://www.tandf.co.uk/journals
DOI: 10.1080/03081070500473490

The question of how to measure the amount of uncertainty (non-specificity) associated with a finite set of possible alternatives was addressed by Hartley (1928). He showed that the only sound way of measuring this amount is to use a logarithm of the number of possible alternatives. Denoting the finite set of alternatives that, according to all available evidence, are possible by A, the Hartley measure of non-specificity, H, is usually expressed by the formula $H(A) = \log_2|A|$, where $|A|$ denotes the cardinality of A. The choice of the logarithm base 2 is a consequence of choosing bits as measurement units. One bit of non-specificity is equivalent to uncertainty regarding the truth or falsity of one elementary proposition.

The Hartley measure is applicable only to finite sets. Although its counterpart for infinite sets (subsets of the n-dimensional Euclidean space for some $n \geq 1$), is now available as well (Klir and Yuan 1995) it is not of interest in this paper, which is restricted to finite sets.

The second classical uncertainty theory is based on the notion of classical (additive) probability measure (Halmos 1950). It is well known that the amount of uncertainty associated with a probability distribution function, p, on a finite set X of considered alternatives is measured (in bits) by the functional

$$S(p(x)|x \in X) = -\sum_{x \in X} p(x) \log_2 p(x),$$

which is usually referred to as Shannon entropy (Shannon 1948). Again, its counterpart for infinite sets is of no interest in this paper. In order to get insight into the probabilistic uncertainty, which is measured by the Shannon entropy, the previous equation can be rewritten as

$$S(p(x)|x \in X) = -\sum_{x \in X} p(x) \log_2 \left[1 - \sum_{y \neq x} p(y)\right].$$

For each given probability distribution function p, the value $p(x)$ expresses the degree (a fraction of the total value of one) to which alternative x is claimed to be the true alternative. Since the total value of one is distributed among the alternatives, clearly, these claims conflict with one another. One way of expressing the total conflict, $c_p(x)$, within a given probability distribution function p with respect to the claim $p(x)$ is by using the expression

$$c_p(x) = -\log_2 \left[1 - \sum_{y \neq x} p(y)\right].$$

The Shannon entropy can then be expressed as

$$S(p(x)|x \in X) = \sum_{x \in X} p(x) c_p(x).$$

It is clear now that the Shannon entropy measures a weighted average of conflicts among the claims allocated to individual alternatives in set X by function p. It is thus common to refer to the probabilistic type of uncertainty as conflict.

It is fair to say that the probability-based uncertainty theory has been far more visible than the one based on possibility. Moreover, the fundamental distinction between the two classical theories has been concealed in the vast literature on probability-based information theory, where the Hartley measure is almost routinely viewed as a special case of the Shannon entropy. This view, which results likely from the coincidence that the value of the Hartley measure for a set of possible alternatives is the same as the value of the Shannon entropy for the uniform probability distribution on the same set, is ill-conceived. Indeed, the Hartley measure is totally independent of any probabilistic assumptions. Furthermore, given evidence expressed by a set of possible alternatives, any probability distribution on the same set, not only the uniform one, is consistent with this evidence. It is now clear that the two classical theories of uncertainty are complementary. Evidence that can be formalized in one of them cannot be formalized in the other one. Using both of them as needed allows us to deal with a broader range of situations than the range captured by either of them. Unfortunately, this broader range of situations is still rather narrow. Broadening it by generalizing the classical uncertainty theories became possible by the emergence of two important generalizations in mathematics in the second half of the 20th century. One of them is the generalization of classical measure theory (Halmos 1950) to the theory of monotone measures. The other one is the generalization of classical set theory to fuzzy set theory. Classical measures are generalized by abandoning the requirement of additivity, as was suggested first by Choquet (1953–1954). Classical sets are generalized by abandoning the requirement of sharp boundaries between sets, as was suggested first by Zadeh (1965).

These two generalizations in mathematics expanded substantially the framework for formalizing uncertainty. This expansion is two-dimensional. In one dimension, the classical theory of additive measures is expanded to the less restrictive theory of monotone measures, within which numerous branches are distinguished for dealing with monotone measures with various special properties. In the other dimension, the formalized language of classical set theory is expanded to the more expressive language of fuzzy set theory, where further distinctions are based on various special types of fuzzy sets. A research program whose objective is to study within this two-dimensional framework the dual concepts of information-based uncertainty (uncertainty resulting from information deficiency) and uncertainty-based information (information measured by reduction of uncertainty) was introduced in 1991 under the name generalized information theory (Klir 1991). A recent book (Klir 2006) is a comprehensive and up-to-date coverage of results that have emerged from this research program.

The type of uncertainty captured by classical possibility theory is non-specificity; the type of uncertainty captured by classical probability theory is conflict. When these classical theories are generalized, the two types of uncertainty may coexist. This means that we need justifiable measures of non-specificity and conflict in each generalized uncertainty theory. These measures may be obtained by generalizing the Hartley measure and the Shannon measure or they may be formulated in some other way.

The purpose of this paper is to investigate measures of non-specificity and conflict in a very general uncertainty theory that is formalized in terms of classical set theory. This is a theory that can deal with arbitrary closed and convex sets of probability distributions, which are often referred to as credal sets. All considerations in this paper are based on the assumption that we deal only with finite sets of alternatives.

The paper is organized as follows. An overview of previous efforts to generalize the Hartley measure and the Shannon entropy to some non-classical uncertainty theories is presented in Section 2. An aggregate measure of uncertainty, which captures both non-specificity and conflict, is discussed in Section 3. Its possible disaggregations are investigated in Section 4, and Section 5 is devoted to conclusions.

2. Generalized uncertainty measures

2.1 Uncertainty in Dempster–Shafer theory

Early efforts to generalize the classical uncertainty theories focused on a fairly popular non-classical theory of uncertainty that is usually referred to as Dempster–Shafer theory (DST) (Dempster 1967 and Shafer 1976). To introduce basic concepts associated with the theory, let X denote a finite set of considered alternatives (a universal set) and let $\wp(X)$ denote the power set of X.

One of the basic concepts of DST is a function $m:\wp(X) \rightarrow [0,1]$ for which $m(\emptyset) = 0$ and $\sum_{A \subseteq X} m(A) = 1$. This function is called a basic probability assignment. Any set for which $m(A) > 0$ is called a focal element of m, and the set of all focal elements with their values of m is called a body of evidence.

Two functions are associated in DST with function m: a belief function, Bel, and a plausibility function, Pl, which are defined for all $A \subseteq X$ by the formulas:

$$\mathrm{Bel}\,(A) = \sum_{B \subseteq A} m(B), \quad \mathrm{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B).$$

Clearly, $\mathrm{Bel}\,(A) \leq \mathrm{Pl}(A)$ for each $A \in \wp(X)$ and, hence, Bel and Pl can be interpreted as lower and upper probabilities, respectively. Moreover, the equation

$$\mathrm{Pl}\,(A) = 1 - \mathrm{Bel}\,(A^{\mathrm{c}}),$$

where A^{c} denotes the complement of A, holds for all $A \in \wp(X)$. In terms of Choquet capacities (Choquet 1953–1954), belief functions are Choquet capacities of order ∞ and plausibility functions are called alternate Choquet capacities of order ∞.

Consider now the product space of considered alternatives, $X \times Y$, of finite sets X and Y. Given a joint basic probability assignment m on $X \times Y$, its marginal counterpart, m_X and m_Y, are defined for all $A \in \wp(X)$ and all $B \in \wp(Y)$ by the formulas

$$m_X(A) = \sum_{R|A=R_X} m(R), \quad m_Y(B) = \sum_{R|B=R_Y} m(R),$$

where R_X and R_Y denote projections of R on X and Y, respectively.

Let us now summarize the various efforts to derive justifiable generalizations of the classical Hartley measure and Shannon entropy to DST. The Hartley measure was first generalized by Higashi and Klir (1983) to the theory of graded possibilities (a special branch of DST dealing with nested bodies of evidence) and, then, it was further generalized by Dubois and Prade (1984) to DST. This generalized Hartley measure, GH, is a functional defined for each basic probability assignment m by the formula

$$\mathrm{GH}\,(m) = \sum_{A \subseteq X} m(A) \log_2 |A|.$$

.

This functional clearly characterizes the expected value of the Hartley measure for all subsets of X. It attains its minimum zero, when m is a probability distribution function, and its maximum is obtained when $m(X) = 1$. It also satisfies all the properties (as formulated within DST) that are required for any measure of uncertainty. Its uniqueness in the theory of graded possibilities was proven by Klir and Mariano (1987) and in DST by Ramer (1987).

Efforts to generalize Shannon entropy to DST were not as successful. Several ways of measuring conflict among evidential claims expressed by the basic probability assignment were suggested in the literature, but they all failed in the essential requirement of subadditivity. An example is the functional

$$E(m) = -\sum_{A \subseteq X} m(A) \log_2 \mathrm{Pl}\, |A|,$$

which was suggested by Yager (1983). The failure of E to satisfy the subadditivity requirement would have been acceptable if this essential requirement were satisfied for the sum $\mathrm{GH} + E$. Unfortunately, this was neither the case for E nor for any of the other suggested candidates for generalized Shannon entropy. A summary of these unsuccessful efforts is given in Klir and Wierman (1998) and Klir (2006).

The unsuccessful attempts to find a justifiable generalization of the Shannon entropy in DST were eventually replaced with attempts to find an aggregated measure of both non-specificity and conflict (Harmanec and Klir 1994). Such a measure, which satisfies all the required properties, was found around the mid 1990s by several authors (see Klir (2006) for more details). This aggregate uncertainty measure is a functional $S*$ that for each belief function Bel in DST is defined by the maximum value of the Shannon entropy within the set of all probability distribution functions that dominate the belief function.

Formally,

$$S*(\mathrm{Bel}) = \max_{p \in C}\left[-\sum_{x \in X} p(x) \log_2 p(x)\right],$$

where C denotes the set of all probability distributions, p, such that $\mathrm{Bel}(A) \leq \sum_{x \in A} p(x)$ for all $A \subseteq X$.

Although functional $S*$ is acceptable on mathematical grounds as an aggregate measure of uncertainty in DST, it does not show explicitly the two coexisting components of uncertainty: non-specificity and conflict. As a result, it is highly insensitive to changes in evidence (Klir and Smith 2001). It is thus desirable to disaggregate it. Clearly, $S* = \mathrm{GH} + \mathrm{GS}$, where GH and GS denote appropriate generalizations of the Hartley measure (measuring non-specificity) and the Shannon entropy (measuring conflict). Since $S*$ and GH were well established in DST, it was suggestive to define GS indirectly as the difference $S* - \mathrm{GH}$. This disaggregation was suggested by Smith (2000) who showed that the range of GS defined in this way is $[0, \log_2 |X|]$. The disaggregated total uncertainty measure TU, is then defined as the pair

$$\mathrm{TU} = (\mathrm{GH}, \mathrm{GS}).$$

Now, it is guaranteed that $\mathrm{GH} + \mathrm{GS}$ satisfies all the required mathematical properties, and it does not matter whether each of the two components satisfies them as well. This is important since GS defined as $S* - \mathrm{GH}$ is not in general subadditive.

2.2 Uncertainty in credal sets

A credal set is a convex and closed set of probability distributions with a finite set of extreme points. A theory based on arbitrary credal sets is currently one of the most general theories dealing with imprecise probabilities. It subsumes DST and other well-defined uncertainty theories, such as the theory based on reachable probability intervals (Campos *et al.* 1994) or theories based on Choquet capacities of order k ($k \geq 2$, finite). Efforts to generalize uncertainty measures established in DST to credal sets are thus very important since any successful generalization will be applicable in all the theories subsumed under credal sets.

In order to understand how subadditivity and additivity is defined for uncertainty measures on credal sets, we need the following definition of marginal credal sets.

Definition 1. Let C be a credal set on the Cartesian product $X \times Y$. Then $C_X = \{p_X : \exists p \in C \text{ such that } p_X = \sum_{y \in Y} p(x, y)\}$ is called the marginal credal set of C on X. Analogously for C_Y.

A measure of non-specificity for credal sets was investigated by Abellán and Moral (2000). As is well known, each credal set C is associated with a unique capacity function (lower probability) f_C that is defined for all $A \in \wp(X)$ by the formula, $f_C(A) = \inf_{p \in C} p(A)$, $\forall A \in \wp(X)$.

This function can be converted to another set function, m_C, via the formula $m_C(A) = \sum_{B \subseteq A} (-1)^{|A-B|} f_C(B)$, $\forall A \in \wp(X)$, which is referred to as the Möbius transform (Grabisch 2000). Function m_C has similar properties as the basic probability assignment function in DST: $m_C(\emptyset) = 0$ and $\sum_{A \subseteq X} m_C(A) = 1$. However, contrary to the latter, m_C can be negative for some sets A. Therefore, m_C is usually called a mass assignment function, since the term "probability assignment" is not appropriate on this more general level. Using function m_C, Abellán and Moral (2000) defined a generalized Hartley measure for each credal set C by the formula

$$\mathrm{GH}(C) = \sum_{A \subseteq X} m_C(A) \log_2 |A|.$$

This functional does not distinguish between different credal sets that are associated with the same lower probability f_C, but, as is argued in Abellán and Moral (2000), this is not a problem for this type of measure. Unfortunately, it is now established that the generalized Hartley measure is not subadditive for all credal sets. We know that it is certainly not subadditive for those credal sets that cannot be represented by Choquet capacities of order two. Whether it is subadditive beyond DST is an open question (Abellán and Moral 2005a).

The aggregated uncertainty measure $S*$, which was initially introduced for DST, was also studied in the context of general credal sets by Abellán and Moral (2003). To examine this aggregate measure and its disaggregation on credal sets in more detail is the subject of the next section. In the rest of this paper, we refer to the maximum entropy measures $S*(C)$ on a credal set C as an upper entropy of C.

3. Upper entropy

Since the upper entropy $S*$, considered initially in DST, is defined in terms of a non-linear optimization problem, its practical utility was initially questioned. Fortunately, a relatively simple algorithm for computing $S*$ in DST was developed and its correctness proven (Meyerowitz *et al.* 1994 and Harmanec *et al.* 1996). It was initially assumed, but not proven, that the use of the algorithm can be extended for computing the upper entropy of any credal set (Klir 2003). However, it was shown by Abellán and Moral (2005b) that this assumption is not correct: they showed that the algorithm can be extended to credal sets associated with capacities of order two, but not to arbitrary credal sets associated with general probability intervals.

It was suggested by Abellán and Moral (2005b) and also by Klir and Smith (2001) that the upper entropy $S*$ can be used as an aggregate measures of uncertainty not only for credal sets associated with DST, but for all creedal sets. They also developed a special algorithm for credal sets based on interval-valued probability distributions (Abellán and Moral 2003). Moreover, it was suggested in Klir and Smith (2001) that the disaggregated total uncertainty

$$\mathrm{TU} = (\mathrm{GH}, S* - \mathrm{GH})$$

is also applicable to all credal sets. In this form, the lack of subadditivity of GH, which has been demonstrated for the most general credal sets, is of no consequence.

The principal aim of this section is to justify the use of the upper entropy as an aggregate measure of uncertainty (aggregate of non-specificity and conflict) for arbitrary credal sets. The issue of consistent disaggregations of the maximum entropy measure is addressed in Section 4.

In the particular case of belief functions, Harmanec and Klir (1994) have already considered that upper entropy is a measure of aggregated total uncertainty. They justify it by using an axiomatic approach. However, uniqueness of this measure has not been proven as yet. But perhaps the most compelling reason for this view is given in Walley (1991). It is based on the logarithmic scoring rule.

We start by explaining the case of a single probability distribution p. To be subject to this rule means that we are forced to select a probability distribution q on $\wp(X)$, and if the true value is x then we must pay $-\log_2(q(x))$. For example, if we say that $q(x)$ is very small and x is found to be the true value, we must pay a lot. If $q(x)$ is close to one, then we must pay a small amount. If our information about X is represented by a subjective probability p, then we should choose q so that $E_p[-\log_2(q(X))]$ is minimum, where E_p is the mathematical expectation with respect to p. This minimum is obtained when $q = p$ and the value of $E_p[-\log_2(q(X))]$ is the entropy of p: the expected loss or the minimum amount that we would require to be subject to the logarithmic scoring rule. This rule is widely used in statistics. The entropy is the negative of the expected logarithm of the likelihood under distribution p. The reason for using the logarithmic function is that if we make predictions in two independent experiments at the same time, then the payment should be the sum of the payments in the two experiments.

In the case of a credal set C, we can also apply the logarithmic scoring rule, but now we choose q in such a way that the upper expected loss $\bar{E}_p[-\log_2(q(X))]$ (the supremum of the expectations with respect to the probabilities in C) is minimum. Under fixed q, $\bar{E}_p[-\log_2(q(X))]$ is the maximum loss we can have (the minimum we should be given to accept this gamble). As we have freedom to choose q, we should select it, so that this amount $\bar{E}_p[-\log_2(q(X))]$ is minimized.

Walley shows that this minimum is obtained for the distribution $p_0 \in C$ with maximum entropy (it is proved on the basis of the Minimax theorem, which can be found in Appendix E of Walley 1991). Furthermore, $\bar{E}_p[-\log_2(p_0(X))]$ is equal to $S*(C)$, the upper entropy in C. This is the minimum payment that we should require before being subject to the logarithmic scoring rule. This argument is completely analogous with the probabilistic one, except that we change expectation to upper expectation. This is really a measure of uncertainty, as the better we know the true value of X, the less we should be paid to accept the logarithmic scoring rule.

Our approach is different from the principle of maximum entropy (Jaynes 1963). This principle always seeks a unique probability distribution, the one with the maximum entropy compatible with given constraints. But, here we are not saying that C can be replaced by the probability distribution with maximum entropy. We continue using the credal set to represent uncertainty. We only say that the uncertainty of the credal set can be measured by its upper entropy.

Perhaps the principal handicap to use the upper entropy as a total uncertainty are the cases where adding information does not reduce the amount of uncertainty measure by upper entropy, as we can see in the following example discussed by Kapur *et al.* (1995):

Example 1. **The dice problem**

Suppose that we have a six-faced dice and that we have no information about it. In this case, probabilities p_i, ($i = 1, 2, 3, 4, 5, 6$) are constrained only by axioms of probability theory:

$$\sum_{i=1}^{6} p_i = 1, \quad p_i \geq 0, \ \ \forall_i.$$

Then, the maximum entropy probability distribution (MEPD) of the convex set of probability distributions that satisfies these constraints is the uniform one, with $S^* = \log_2 6$. Now, if we add information that the mean is 7/2, the MEPD is again the uniform one, with $S^* = \log_2 6$. Assume now we add, in each step, information regarding the knowledge of the moment of order j, $\mathbf{M}_j = \sum_{i=1}^{6} i^j p_i$, with the following values:

Step 1 $M_1 = 7/2$,
Step 2 $M_2 = 91/6$,
Step 3 $M_3 = 441/6$,
Step 4 $M_4 = 2275/6$,
Step 5 $M_5 = 12201/6$.

In all of these steps, the MEPD of the associated credal set is still the uniform one, and $S^* = \log_2 6$.

Kapur *et al.* (1995) argue, in a natural way, that the "information supplied" in each step of this example produces equal "reduction of uncertainty". They use the difference $S^* - S_*$, where S_* denotes the lower entropy associated with each credal set, as a measure of information that produces reduction of uncertainty which is not captured by S^* alone. However, the difference $S^* - S_*$ alone does not capture total uncertainty since it is zero for each single probability distribution and, hence, it does not capture uncertainty in this case. In each step of the example, clearly, the credal set is reduced and, consequently, non-specificity is reduced as well. On the other hand, the lower entropy S_* is increased in each step. Since the total aggregated uncertainty S^* does not change, S_* may be viewed as measuring the increasing conflict. Its maximum would be reached when a credal set with a single probability distribution is obtained. Then, $S_* = S^*$ and non-specificity vanishes.

4. A new approach

It is argued in Klir (2005) that the total aggregated uncertainty $S*$ should be appropriately disaggregated to overcome its notorious insensitivity to changes in evidence. It is suggested that, in addition, to its disaggregation with respect to the generalized Hartley measure, first considered by Smith (2000), other disaggregations should be explored. One particular disaggregation is suggested in Klir (2005), which is based on the idea proposed by Abellán and Moral (2005a) that it is reasonable to use the difference $S* - S_*$ as a natural measure of non-specificity for credal sets. This alternative total disaggregated uncertainty, ${}^{a}\mathrm{TU}$, is defined as

$$ {}^{a}\mathrm{TU} = (S* - S_*, S_*), $$

where the first component is a measure of non-specificity and the second one is a measure of conflict. Properties of these two components are examined in the rest of this section.

4.1 Maximum difference of entropies

The difference $S* - S_*$ suggested in Abellán and Moral (2005a) as non-specificity measure on credal sets, is fundamentally different from the generalized Hartley measure. There are examples in which $S* - S_* > \mathrm{GH}$ and examples in which the opposite happens. We have not studied the existence of a bound for the difference of these measures.

Measures $S* - S_*$ and GH have very different behaviours. Measure GH is a measure of absolute imprecision. However, $S* - S_*$ quantifies in a different way the same imprecision. Assume that we have a probability distribution p and that $\epsilon < \min_{x\in X} p(x)$ and consider the credal set C_ϵ given by all the probability distributions p' such that $\max_{x\in X}|p'(x) - p(x)| \leq \epsilon$. That is, in C_ϵ a maximum of variation of ϵ is allowed with respect to p. Then, $\mathrm{GH}(C_\epsilon)$ depends only on ϵ. On the contrary $S* - S_*$ depends also on p: If p is close to the uniform distribution, then $S* - S_*$ is smaller than if p is close to a degenerate distribution. This makes sense, as the same absolute difference of probability can be more important for probabilities close to 0, than for intermediate values of probability. The following example shows numerically a particular case of this fact.

Example 2. Consider $X = \{x_1, x_2\}$ and the following two basic probability assignments:

$$ \begin{aligned} &m_1(\{x_1\}) = 0.495, \quad m_1(\{x_2\}) = 0.495, \quad m_1(\{x_1, x_2\}) = 0.01, \\ &m_2(\{x_1\}) = 0.99, \qquad m_2(\{x_2\}) = 0, \qquad m_2(\{x_1, x_2\}) = 0.01. \end{aligned} $$

Clearly $\mathrm{GH}(m_1) = \mathrm{GH}(m_2) = 0.01 \log_2(2)$.

The credal sets associated with these masses are, $C_i = \{(p)|\mathrm{Bel}_i(A) \leq \sum_{x\in A} p(x) \leq \mathrm{Pl}_i(A), i = 1, 2\}$. We also have:

$$ \begin{aligned} &\mathrm{Bel}_1(\{x_1\}) = 0.495, \ \mathrm{Pl}_1(\{x_1\}) = 0.505; \quad \mathrm{Bel}_2(\{x_1\}) = 0.99, \ \mathrm{Pl}_2 = (\{x_1\}) = 1.00 \\ &\mathrm{Bel}_1(\{x_2\}) = 0.495, \ \mathrm{Pl}_1(\{x_2\}) = 0.505; \quad \mathrm{Bel}_2(\{x_2\}) = 0.0, \ \mathrm{Pl}_2(\{x_2\}) = 0.01 \end{aligned} $$

and

$$C_1 = \text{CH}(\{(0.495, 0.505), (0.505, 0.495)\}),$$

$$C_2 = \text{CH}(\{(0.99, 0.01), (1.00, 0.00)\}),$$

where CH stands for the convex hull.

Then, the range of entropy for m_1 is

$$(S^* - S_*)(C_1) = \log_2(2) + 0.495 \log_2(0.495) + 0.505 \log_2(0.505) = 0.0000721$$

and for m_2 is

$$(S^* - S_*)(C_2) = 0.99 \log_2(0.99) - 0.01 \log_2(0.01) - 0 = 0.0807909.$$

Credal sets C_1 and C_2 are examples of sets C_ϵ for $\epsilon = 0.005$ as defined above. They are associated with different probabilities. C_1 is obtained when p is the uniform distribution, whereas C_2 is obtained for the distribution p, given by $p(x_1) = 0.995$, $p(x_2) = 0.005$. The value of GH is the same for these credal sets, but the maximum difference of entropies is different. It is greater in C_2, the credal set with probabilities with values close to 0 and 1, than in C_1, the credal set with probabilities close to the uniform distribution.

In the above example, C_1 and C_2 are also very different from the point of view of their total uncertainty: C_1 is more uncertain than C_2. However, we also claim that C_2 may be considered as more imprecise than C_1. There is the same absolute difference between the extreme probabilities, but in the case of C_2 this difference is more important. Credal sets such as C_2 are obtained for rare events, as for example the probability of suffering an earthquake. It is not the same whether the probability is 0.01 or 0. The risks are very different. However, C_1 appears in situations in which an event has roughly the same probability than its complementary. It can be the case of a possibly biased coin. The values for the probability of heads of 0.50 or 0.49 are not so different. Our betting behaviour is more or less the same in both situations and, therefore, imprecision has less severe consequences. The difference between the two situations is not only that an earthquake is more important than the result of a coin. It is also in the differences in the betting behaviour associated with the two extreme situations. If the probability for x_1 is 0, we will never accept a bet for this event. However, for 0.01 we would accept bets for x_1 if we receive more than 100 units for the payment of 1. In this way, we could lose a lot of money if a probability of 0.01 is estimated by us as 0. The consequences of estimating 0.50 by 0.49 are not so important.

Table 1. Value of m, m' and Bel and Pl of m and m' in the Example 3.

Set	m	m'	Bel_m	Pl_m	$Bel_{m'}$	$Pl_{m'}$
$\{x_1\}$	1/4	1/12	1/4	1/2	1/12	5/12
$\{x_2\}$	1/4	1/3	1/4	1/2	1/3	7/12
$\{x_3\}$	1/4	1/3	1/4	1/2	1/3	7/12
$\{x_1,x_2\}$	0	0	1/2	3/4	5/12	2/3
$\{x_1,x_3\}$	0	0	1/2	3/4	5/12	2/3
$\{x_2,x_3\}$	0	0	1/2	3/4	2/3	11/12
$\{x_1,x_2,x_3\}$	1/4	1/4	1	1	1	1

If we want to use the maximum difference of entropies on credal sets alone as a measure of uncertainty, as suggested by Kapur *et al.* (1995), we can find some difficulties, as illustrated by the following example.

Example 3. Assume $X = \{x_1, x_2, x_3\}$ and let m, m' be basic probability assignments on X given in table 1, where also the associated belief and plausibility functions are shown.

We obtain the following values of $S^* - S_*$

$$(S^* - S_*)(C_m) = \log_2(3) - 1.5 = 0.08496,$$

$$(S^* - S_*)(C_{m'}) = \log_2(3) - 1.28067 = 0.30429.$$

If the total uncertainty is measured by $S^* - S_*$, then the uncertainty represented by m is smaller than the one represented by m':

$$(S^* - S_*)(C_{m'}) \simeq 4(S^* - S_*)(C_m).$$

This is contradictory. If we observe the values of Bel and Pl in table 1, we can think that there is more information in m' that in m. Clearly m' is pointed to the set $\{x_2, x_3\}$, while m is totally symmetrical. For each probability distribution in $C_{m'}$ there is a probability distribution in C_m whose entropy is greater than or equal, but not the other way around. Hence, we can say that m should not represent more information than m', or m' should not represent more uncertainty than m, but the use of $S^* - S_*$ quantifies it the wrong way. Using the total aggregated uncertainty S^* in DST we obtain the same value for m and m'. This is less contradictory than the values produced for $S^* - S_*$ function alone.

The following properties of the difference $S^* - S_*$ are established in Abellán and Moral (2005a):

(1) It has a range in $[0, \log_2|X|]$.
(2) It is an increasing monotone function. When $C \subseteq C'$, then

$$S^*(C) \leq S^*(C'), \quad S_*(C) \geq S_*(C')$$

and

$$S^*(C) - S_*(C) \leq S^*(C') - S_*(C').$$

(3) It is a continuous function, since it is a difference of two continuous functions.
(4) It is an additive function, since it is a difference of two additive functions.
(5) It is equal to zero for precise probabilities.

This function is not subadditive for credal sets, but this is acceptable when it is used as one component (measuring non-specificity) in the disaggregated total uncertainty ${}^{a}\mathrm{TU} = (S^* - S_*, S_*)$.

4.2 Lower entropy

The importance of minimum entropy probability distributions was earlier emphasized by Watanabe (1969, 1981) and Christensen (1981). Its use has often been rejected because it is rather difficult to compute it, due to the concavity of the entropy function. Procedures for computing the minimum entropy under some special constraints are presented in Kapur *et al.* (1995) and Yuan and Kesaven (1998). In Abellán and Moral (2000), a branch and bound algorithm is presented for computing the lower entropy for any credal set that is associated with capacities of order two.

Lower entropy S_* on credal sets is also analyzed in Abellán and Moral (2005a) as a measure of conflict. It represents a degree of internal contradiction similarly as function E in DST. The latter was introduced as a measure of conflict in DST (Yager 1983), and also as a part of a total uncertainty measure (Lamata and Moral 1987). The following is its basic property.

PROPERTY 1. Let m be a basic probability assignment on a finite set X. Then $E(m) = 0$ if and only if all focal sets have at least one common element.

This property implies that E is zero only for cases where there is no conflict. Similar property is verified for S_* but, contrary to E, S_* can be extended to more general theories than DST in a natural way. The extension of E to credal sets is problematic. If we want to extend it beyond DST, we need to replace the basic assignment function by the general mass assignment function and the plausibility function with the more general upper probability function. Then we can obtain examples where the values of E have no sense as we can see in the following example:

Example 4. Let C denote the credal set consisting of all probability distributions generated on set $X = \{x_1, x_2, x_3\}$ by convex combinations of the following set of probability distributions:

$$\{(0.5, 0.5, 0); \quad (0.5, 0, 0.5); \quad (0, 0.5, 0.5)\}.$$

This credal set corresponds to a set of reachable probability intervals (Campos *et al.* 1994). We readily obtain the following values:

$$m(\{x_1, x_2\}) = 0.5, \quad m(\{x_1, x_3\}) = 0.5, \quad m(\{x_2, x_3\}) = 0.5,$$

$$\bar{P}(\{x_1, x_2\}) = 1, \quad \bar{P} = (\{x_1, x_2\}) = 1, \quad \bar{P} = (\{x_2, x_3\}) = 1,$$

$$m(\{x_1, x_2, x_3\}) = -0.5, \quad \bar{P}(\{x_1, x_2, x_3\}) = 1,$$

$$E(C) = 0, \quad S_*(C) = \log_2(2).$$

According to functional E, there is no conflict in the given body of evidence. However, this is not correct since none of the elements of X is contained in all focal elements. The existence of conflict in this example is correctly recognized by S_*.

If we compare these functionals, we can prove that E is a lower bound of S_* in DST. To do that, we need the following Lemma.

LEMMA 1. (Dempster 1967)

Let X be a finite set, Bel a generalized belief function on X, i.e. a function that satisfies all the requirement of a belief function except the requirement that $\text{Bel}(X) = 1$. Let m be the corresponding generalized basic assignment; then, a tuple $\langle p(x) | x \in X \rangle$ satisfies the constraints

$$0 \leq p(x) \leq 1, \quad \forall x \in X,$$

$$\sum_{x \in X} p(x) = \mathrm{Bel}(X),$$

and

$$\mathrm{Bel}(A) \leq \sum_{x \in A} p(x), \quad \forall A \subseteq X,$$

if and only if there exist non-negative real numbers α_x^A for all non-empty sets $A \subseteq X$ and all $x \in A$ such that

$$p(x) = \sum_{A | x \in A} \alpha_x^A$$

and

$$\sum_{x | x \in A} \alpha_x^A = m(A).$$

PROPERTY 2. Let m be a basic probability assignment on subsets of a finite set X and let C_m denote the associated credal set. Then, $E(m) \leq S_*(C_m)$.

Proof. By Lemma 1,$\forall x \in X$ and $\forall A \subseteq X$, $\exists \alpha_x^A \geq 0$ such that $\forall p \in C_m$ it is verified that:

$$p(x) = \sum_{A|x \in A} \alpha_x^A, \quad m(A) = \sum_{x|x \in A} \alpha_x^A.$$

Now,

$$S(p) = -\sum_x p(x)\log_2(p(x)) \quad = -\sum_x \sum_{A|x \in A} \alpha_x^A \log_2 p(x)$$

$$= -\sum_A \sum_{x|x \in A} \alpha_x^A \log_2(p(x)) \geq -\sum_A \sum_{x|x \in A} \alpha_x^A \log_2(p(A)) \geq -\sum_A m(A)\log_2(\mathrm{Pl}(A)). \quad \square$$

In Abellán and Moral (2005a), it is proved that S_* verifies the following properties:

(1) $S_*(C) \in [0, \log_2|X|]$.
It attains the interval extremes in a degenerate probability distribution and in the uniform probability distribution, respectively.
(2) It is a monotone decreasing function: $C \subseteq C' \Rightarrow S_*(C) \geq S_*(C')$.
(3) It is continuous.
(4) It is additive, i.e. if C is a credal set on a universal $X \times Y$ such that there is independence under C, $C = \mathrm{CH}(C_X \times C_Y)$, with CH the convex hull, then $S_*(C) = S_*(C_X) + S_*(C_Y)$.
(5) It is not subaditive, i.e. the inequality $S_*(C) \leq S_*(C_X) + S_*(C_Y)$ is not always satisfied.
(6) It is not superadditive, i.e. the inequality $S_*(C) \geq S_*(C_X) + S_*(C_Y)$ is not always satisfied.

5. Conclusions

The various uncertainty measures for credal sets are examined in this paper. The main theme of the paper is the issue of how the well-established aggregate uncertainty measure, which for each credal set is expressed by its upper entropy, can be disaggregated to overcome its insensitivity to changes in evidence. Each disaggregation breaks the aggregate uncertainty into two components, a measure of non-specificity and a measure of conflict. The paper focuses on a novel disaggregation, in which non-specificity is measured for each credal set by the difference between its upper and lower entropies, while conflict is measured by the lower entropy. It is shown that this disaggregation, contrary to other conceivable disaggregations, is applicable to arbitrary credal sets. It is argued that the two components of this disaggregated total uncertainty possess properties that qualify them well for measuring non-specificity and conflict. The conclusion is that this novel measure of total disaggregated uncertainty, ${}^{a}\mathrm{TU} = (S^* - S_*, S_*)$, is at this time the best justified measure of uncertainty for credal sets.

It is conceivable that other measures of uncertainty are more appropriate within the various special theories of uncertainty. In the theory of graded possibilities, for example, the disaggregation with respect to the generalized Hartley measure appears to be more sensitive to changes in evidence than ${}^{a}\mathrm{TU}$. In DST, disaggregations of the form ${}^{b}\mathrm{TU} = (S^* - \mathrm{GS}, \mathrm{GS})$, where GS denotes any of the candidates for a generalized Shannon entropy investigated in the 1980s and 1990s and referred to as dissonance, confusion, discord and strife (Klir and Wierman 1998), should be examined and compared among themselves as well as with ${}^{a}\mathrm{TU}$. These are interesting questions, but they are beyond the scope of this paper.

The overall conclusion of this paper can now be summarized as follows. While various disaggregations of the aggregate measure S^* may be preferable in various special theories of uncertainty, the disaggregation ${}^{a}\mathrm{TU} = (S^* - S_*, S_*)$ introduced in this paper is the best justified one when working within the most general theory of uncertainty, a theory based on arbitrary credal sets.

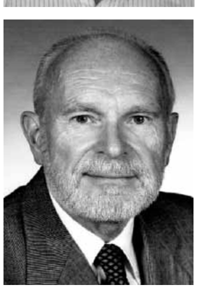
Acknowledgements
This work has been supported by the Spanish Ministry of Science and Technology under project Algra (TIN2004-06204-C03-02).

References

J. Abellán and S. Moral, “A non-specificity measure for convex sets of probability distributions”, *Int. J. Unc. Fuzz. Knowl. Based Syst.*, 8(3), pp. 357–367, 2000.
J. Abellán and S. Moral, “Maximum of entropy for credal sets”, *Int. J. Unc. Fuzz. Knowl. Based Syst.*, 11(5), pp. 1215–1225, 2003.
J. Abellán and S. Moral, “Maximum difference of entropies as a non-specificity measure for credal sets”, *Int. J. Gen. Syst.*, 34(3), pp. 201–214, 2005a.
J. Abellán and S. Moral, “An algorithm that computes the upper entropy for order-2 capacities”, submitted to *Int. J. Unc. Fuzz. Knowl. Based Syst.*, 2005b.
L.M. de Campos, J.F. Huete and S. Moral, “Probability intervals: a tool for uncertainty reasoning”, *Int. J. Unc. Fuzz. Knowl. Based Syst.*, 2(2), pp. 167–196, 1994.
G. Choquet, “Theory of capacities”, *Ann. Inst. Fourier*, 5, pp. 131–292, 1953–1954.
R. Christensen, *Entropy Minimax Source Book*, Lincoln, MA: Entropy Ltd, 1981, Vols 1–4.
A.P. Dempster, “Upper and lower probabilities induced by a multivaluated mapping”, *Ann. Math. Stat.*, 38(2), pp. 325–339, 1967.
D. Dubois and H. Prade, “A note on measure of specificity for fuzzy sets”, *BUSEFAL*, 19, pp. 83–89, 1984; also published in the *Int. J. Gen. Syst.*, 10(4), 279–285, 1985.

M. Grabisch, "The interaction and Mäobius representations of fuzzy measures on finite speces, k-additive measures: a survey", in *Fuzzy Measures and Integrals: Theory and Applications*, M. Grabisch et al., Ed., New York: Springer-Verlag, 2000.
P.R. Halmos, *Measure Theory*, Princeton, NJ: D. Van Nostrand, 1950.
R.V.L. Hartley, "Transmission of information", *Bell Syst. Tech. J.*, 7(3), pp. 535–563, 1928.
D. Harmanec and G.J. Klir, "Measuring total uncertainty in Dempster–Shafer theory: a novel approach", *Int. J. Gen. Syst.*, 22(4), pp. 405–419, 1994.
D. Harmanec, G. Resconi, G.J. Klir and Y. Pan, "On the computation of uncertainty measure in Dempster–Shafer theory", *Int. J. Gen. Syst.*, 25(2), pp. 153–163, 1996.
M. Higashi and G.J. Klir, "Measures of uncertainty and information based on possibility distributions", *Int. J. Gen. Syst.*, 9(1), pp. 43–58, 1983.
E.T. Jaynes, "Information theory and statistical mechanics", in *Statistical Physics*, K. Ford, Ed., New York: Benjamin, 1963, pp. 182–218.
J.N. Kapur, G. Baciu and H.K. Kesavan, "The minmax information measure", *Int. J. Syst. Sci.*, 26(1), pp. 1–12, 1995.
G.J. Klir, "Generalized information theory", *Fuzzy Sets Syst.*, 40(1), pp. 127–142, 1991.
G.J. Klir, "An update on generalized information theory", *Proceeding of the Third International Symposium on Imprecise Probabilities and their Applications, ISIPTA '03*, Lugano 2003, pp. 321–334.
G.J. Klir, "Measuring uncertainty associated with convex sets of probability distributions: a new approach", *Proc. NAFIPS'05*, Ann Arbor, MI (only in CD) 2005.
G.J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*, Hoboken, NJ: John Wiley, 2006.
G.J. Klir and M. Mariano, "On the uniqueness of possibilistic measure of uncertainty and information", *Fuzzy Sets Syst.*, 24(2), pp. 197–219, 1987.
G.J. Klir and R.M. Smith, "On measuring uncertainty and uncertainty based information: recent developments", *Ann. Math. Artif. Intell.*, 32(1–4), pp. 5–33, 2001.
G.J. Klir and M.J. Wierman, *Uncertainty-Based Information*, Heidelberg: Physica-Verlag, 1998.
G.J. Klir and B. Yuan, "On nonspecificity of fuzzy sets with continuous membership functions", *Proc. 1995 Int. Conf. Syst. Man Cybern.*, Vancouver 1995.
M.T. Lamata and S. Moral, "Measures of entropy in the theory of evidence", *Fuzzy Sets Syst.*, 12, pp. 193–226, 1987.
A. Meyerowitz, F. Richman and E.A. Walker, "Calculating maximum-entropy probabilities densities for belief functions", *Int. J. Unc. Fuzz. Knowl. Based Syst.*, 2(4), pp. 377–389, 1994.
A. Ramer, "Uniqueness of information measure in the theory of evidence", *Fuzzy Sets Syst.*, 24(2), pp. 183–196, 1987.
G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ: Princeton University Press, 1976.
C.E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.*, 27(3–4), pp. 379–423, 623–656, 1948.
R.M. Smith, "Generalized information theory: resolving some old questions and opening some new ones", PhD dissertation, Binghamton University—SUNY, Binghamton (2000).
P. Walley, *Statistical Reasoning with Imprecise Probabilities*, London and New York: Chapman and Hall, 1991.
S. Watanabe, *Knowing and Guessing*, New York: John Wiley, 1969.
S. Watanabe, "Pattern recognition as a quest for minimum entropy", *Pattern Recognit.*, 13(5), pp. 381–387, 1981.
R.R. Yager, "Entropy and specificity in a mathematical theory of evidence", *Int. J. Gen. Syst.*, 9(4), pp. 249–260, 1983.
L. Yuan and H.K. Kesaven, "Minimum entropy and information measure", *IEEE Trans. Syst. Man Cybern. C*, 28(3), 1998.
L.A. Zadeh, "Fuzzy sets", *Inf. Control*, 8(3), pp. 338–353, 1965.

Joaquín Abellán received his PhD degree in January of 2003 from the University of Granada. He is an assistant professor of Computer Science and Artificial Intelligence at the University of Granada. His current research interests are representation of the uncertainty through convex sets of probability distributions and its applications to classification.



George J. Klir is currently a Distinguished Professor of Systems Science at Binghamton University—SUNY. He has been with SUNY since 1969. His earlier work was in the areas of systems modeling and simulation, logic design, computer architecture, and discrete mathematics. His current research interests include the areas of intelligent systems, generalized information theory, fuzzy set theory and fuzzy logic, theory of generalized measures, and soft computing. He is the author of over three hundred articles and 16 books. He has also edited 10 books and has been Editor of the *International Journal*

of General Systems since 1974 and the *International Book Series on Systems Science and Systems Engineering* since 1985. He was President of SGSR (1981–82), IFSR (1980–84), NAFIPS (1988–1991), and IFSA (1993–1995). He is a Fellow of IEEE and IFSA, and has received numerous awards and honors, including 5 honorary doctoral degrees, the Gold Medal of Bernard Bolzano, Lotfi A. Zadeh Best Paper Award, the Kaufmann's Gold Medal, SUNY Chancellor's Award for Excellence in Research and IFSA Award for Outstanding Achievement. His biography is included in many biographical sources, including *Who's Who in America, Who's Who in the World, American Men and Women of Science, Outstanding Educators of America, Contemporary Authors*, etc. His research has been supported for more than 20 years by grants from NSF, ONR, Air Force, NASA, NATO, Sandia Laboratories, and some industries.

Serafín Moral received his PhD degree in 1984 from the University of Granada. He is a professor of Computer Science and Artificial Intelligence and member of the research group on Uncertainty Treatment on Artificial Intelligence (UTAI) at the University of Granada. His current research interests are imprecise probabilities, propagation algorithms in dependence graphs and uncertainty and defeasible reasoning in general.