Information Quantities and Possibility Measures

Tomáš Kroupa

[1] Laboratory for Intelligent Systems Prague, University of Economics, Winston Churchill Sq. 4, 130 67 Prague, Czech Rep.*
[2] Fac. of Electrical Engineering, Czech Tech. University, Technicka 2, 166 27 Prague.
tom.kroupa@tiscali.cz

Abstract. Information theory for non-additive measures has always attracted considerable attention. This effort resulted, among others, in defining information measures for Dempster-Shafer (D-S) theory. In the paper, the properties of information quantities called the aggregate uncertainty (AU) and the nonspecificity are discussed within the framework of possibility theory that may be viewed as a special part of D-S theory for this purpose. An axiomatic approach to possibility theory (formulated by de Cooman) is adopted. Fundamental properties of the AU and the nonspecificity are analyzed and summarized. Moreover, an inequality clarifying their mutual relationship is presented. A relation between possibilistic independence (parameterized by a continuous t-norm) and the additivity requirement frequently imposed on uncertainty measures is explored in detail.

1 Introduction

Information theory covers a variety of uncertainty formalizations today. In the field of non-additive measures, several information quantities were defined in D-S theory and possibility theory – we refer an interested reader to [4].

In this book, there are distinguished the three uncertainty types. *Fuzzines* is a type of uncertainty closely connected with fuzzy set theory. *Nonspecifity* is dependent on the cardinality of relevant sets of alternatives and *conflict* stems from assigning potentially conflicting degrees of the uncertainty to these alternatives.

In D-S theory, an *aggregate uncertainty* (AU) measures the 'total' uncertainty associated with a given belief measure whereas a *nonspecificity* quantifies a degree of the 'nonspecificity' of the belief measure. Because possibility theory can be considered as a special part of D-S theory, these information quantities may be employed in calculating the uncertainties for possibility measures as well. The aim of this paper is to study their properties in an axiomatic framework for possibility theory [1].

* This work was supported by the grant GA ČR no. 201/02/1269.

Mária Bieliková (Ed.): SOFSEM 2002 Student Research Forum, pp. 21–26, 2002.

2 Possibility Theory

2.1 Possibility Measures

Let us introduce the basic notions concerning possibility theory. A *universe* $\mathbf{X}$ is a set having at least two elements; $\mathbf{X}$ contains all values that a variable X may take. We deal only with finite universes. $2^{\mathbf{X}}$ denotes a power set of $\mathbf{X}$. A *possibility measure* Π on $\mathbf{X}$ is a set function $\Pi : 2^{\mathbf{X}} \to [0,1]$ such that for any system $\{A_j : j \in J\}$ of the elements of $2^{\mathbf{X}}$ the condition $\Pi(\bigcup_{j\in J} A_j) = \max_{j\in J} \Pi(A_j)$ is satisfied. The possibility measure Π is *normal* if $\Pi(\mathbf{X}) = 1$. Only normal possibility measures are considered in this paper.

A *possibility distribution* π is a pointwise function $\pi : \mathbf{X} \to [0,1]$ having the property $\Pi(A) = \max_{x\in A} \pi(x)$ for all $A \subseteq \mathbf{X}$. Every possibility measure uniquely determines the possibility distribution and vice versa. Assume that π_{XY} is a *joint possibility distribution* defined on a Cartesian product $\mathbf{X}\times\mathbf{Y}$. We define a *marginal possibility distribution* by the equation $\pi_X(x) = \max_{y\in\mathbf{Y}} \pi_{XY}(x,y)$ for any $x \in \mathbf{X}$.

2.2 Independence in Possibility Theory

An independence of possibilistic variables is parameterized by a t-norm. The *t-norm* T is a commutative semigroup operation on $[0,1]$ having a neutral element 1 and being isotonic, i.e. $\forall a_1, a_2, b \in [0,1] : a_1 \leq a_2 \Rightarrow T(a_1,b) \leq T(a_2,b)$.

There exist three significant examples of continuous t-norms: the *Gödel's t-norm* $T_G(a,b) = \min(a,b)$, the *product t-norm* $T_\times(a,b) = ab$ and the *Lukasziewicz' t-norm* $T_L(a,b) = \max(0, a+b-1)$. It is easy to show [3] that the following inequality holds true for all $(a,b) \in [0,1]^2$:

$$T_L(a,b) \leq T_\times(a,b) \leq T_G(a,b). \tag{1}$$

In addition, $T \leq T_G$ for any t-norm T.

Let variables X and Y have a joint possibility distribution π_{XY}. We call the variables X and Y *T-independent* if $\pi_{XY}(x,y) = T(\pi_X(x), \pi_Y(y))$ for any $(x,y) \in \mathbf{X}\times\mathbf{Y}$. The joint possibility distribution π_{XY} on $\mathbf{X}\times\mathbf{Y}$ is then called a *T-product extension* of π_X and π_Y.

2.3 Possibility Measure as an Upper Envelope

According to [2], every normal possibility measure Π on $\mathbf{X}$ can be assigned a system

$$\mathbb{P}(\Pi) = \{P : P(A) \leq \Pi(A), \text{ for all } A \subseteq \mathbf{X}\}$$

of probability measures on $\mathbf{X}$ dominated by Π.[3] Moreover, Π is an *upper envelope* of $\mathbb{P}(\Pi)$, i.e. for any $A \subseteq \mathbf{X}$ there exists $P \in \mathbb{P}(\Pi)$ such that $P(A) = \Pi(A)$. As we require every possibility measure Π to be normal, Π always dominates at least one probability measure P and $\mathbb{P}(\Pi)$ is therefore nonempty.

[3] For the sake of simplicity, we denote by $\mathbb{P}(\Pi)$ also the set of all probability distributions corresponding to the probability measures in $\mathbb{P}(\Pi)$.

3 Possibilistic Information Quantities

3.1 Aggregate Uncertainty

The AU is conceived as a measure of the uncertainty capturing both the non-specifity and the conflict in an aggregated fashion. Its definition [4] originated from more general efforts to introduce a meaningful measure of the 'total' uncertainty within D-S theory. In the following, $\mathcal{H}(p)$ denotes the Shannon entropy of a probability distribution p.

Definition 1. *Let π be a possibility distribution on $\mathbf{X}$. The* aggregate uncertainty $\mathcal{U}$ *of π is defined as the maximum entropy attainable within $\mathbb{P}(\Pi)$, that is $\mathcal{U}(\pi) = \max_{p \in \mathbb{P}(\Pi)} \mathcal{H}(p)$.*

Although a single computation of $\mathcal{U}$ involves finding a solution to a nonlinear optimization problem, a simple and an effective algorithm can be employed [4]. Moreover, due to the convexity of $\mathbb{P}(\Pi)$ and the concavity of the Shannon entropy, the probability distribution $\hat{p}$ maximizing the entropy within constraints specified by $\mathbb{P}(\Pi)$ is uniquely determined. These are some of the basic properties of aggregate uncertainty (for proof see [4]):

1. *range:* $\mathcal{U}(\pi) \in [0, \log_2 |\mathbf{X}|]$ for any π on $\mathbf{X}$;
2. *maximum:* $\mathcal{U}(\pi) = \log_2 |\mathbf{X}|$ if π is *non-informative*, i.e. $\pi(x) = 1$ for all $x \in \mathbf{X}$;
3. *minimum:* $\mathcal{U}(\pi) = 0$ iff π is *degenerated*, i.e. $\pi(x') = 1$ for a single $x' \in \mathbf{X}$ and $\pi(x) = 0$ for all $x \neq x'$;
4. *subadditivity:* the inequality $\mathcal{U}(\pi_{XY}) \leq \mathcal{U}(\pi_X) + \mathcal{U}(\pi_Y)$ holds for a joint possibility distribution π_{XY} on $\mathbf{X} \times \mathbf{Y}$ with the marginal possibility distributions π_X and π_Y.

The maximality condition from 2. can be further refined.

Proposition 1. $\mathcal{U}(\pi) = \log_2 |\mathbf{X}|$ *iff π satisfies the condition*

$$\forall A \subseteq \mathbf{X} : \frac{|A|}{|\mathbf{X}|} \leq \max_{x \in A} \pi(x). \tag{2}$$

Proof. $\mathcal{U}(\pi) = \log_2 |\mathbf{X}|$ if and only if $\mathbb{P}(\Pi)$ contains the uniform probability distribution $\hat{p}$. This is obviously equivalent with having satisfied the following inequality for every $A \in \mathbf{X}$:

$$\frac{|A|}{|\mathbf{X}|} = \sum_{x \in A} \hat{p}(x) = \hat{P}(A) \leq \Pi(A) = \max_{x \in A} \pi(x).$$

□

Possibility distributions can be partially ordered in this way: $\pi_1 \leq \pi_2$ whenever $\pi_1(x) \leq \pi_2(x)$ for all $x \in \mathbf{X}$. The following assertion is then straightforward to obtain.

Lemma 1. *If* $\pi_1 \leq \pi_2$*, then* $\mathcal{U}(\pi_1) \leq \mathcal{U}(\pi_2)$.

Further, taking into account this lemma and the t-norm inequality (1):

Lemma 2. *Assume that* π_X, π_Y *are possibility distributions on* $\mathbf{X}$ *and* $\mathbf{Y}$*, respectively. Let* $\pi_L, \pi_\times$ *and* π_G *be their* T_L*-,* $T_\times$*- and* T_G*-product extension, respectively. Then* $\mathcal{U}(\pi_L) \leq \mathcal{U}(\pi_\times) \leq \mathcal{U}(\pi_G)$.

The proposition stated above can be justified also on an intuitive ground: the less specific a T-product extension π_T is, the higher value $\mathcal{U}(\pi_T)$ should attain.

One might also ask a question if there is some analog of *additivity* of Shannon entropy, i.e. the equality $\mathcal{H}(p_{XY}) = \mathcal{H}(p_X) + \mathcal{H}(p_Y)$ which is satisfied for independent variables X, Y. However, in possibility theory, this property must be analyzed w.r.t. the independence parameterized by a t-norm: it can easily be demonstrated that the additivity property is generally violated in case of the three significant t-norms $T_G, T_\times, T_L$. In fact, we prove much more; the additivity is violated under *any* t-norm T.

Proposition 2. *There doesn't exist a t-norm* T *such that the equality*

$$\mathcal{U}(\pi_{XY}) = \mathcal{U}(\pi_X) + \mathcal{U}(\pi_Y) \tag{3}$$

is preserved for all possibility distributions π_X *and* π_Y *on* $\mathbf{X}$ *and* $\mathbf{Y}$*, respectively, where* π_{XY} *is a* T*-product extension of* π_X, π_Y.

Proof. Denote by π_G a T_G-product extension of π_X and π_Y. Due to the subadditivity of $\mathcal{U}$, we have $\mathcal{U}(\pi_X)+\mathcal{U}(\pi_Y) \geq \mathcal{U}(\pi_G)$. Since T_G is the maximal t-norm, we obtain further $\mathcal{U}(\pi_G) \geq \mathcal{U}(\pi_{XY})$ from Lemma 1. Realize that the additivity is generally violated under T_G-independence and π_{XY} is always equal or lower than π_G. In either case, the equality (3) can't be satisfied. □

Do there nevertheless exist some special examples of the possibilistic distributions π_X, π_Y preserving the additivity? An answer to this question is contained in the assertion below whose proof is rather technical and lengthy and hence we decided not to include it in this paper.

Proposition 3. *Let* π_X, π_Y *be possibility distributions on* $\mathbf{X}, \mathbf{Y}$*, respectively, and* π_{XY} *be their* T*-product extension where* T *is an arbitrary t-norm. The equality (3) holds true if at least one of the distributions* π_X *or* π_Y *takes only values from* $\{0, 1\}$.

The implication can't be reversed, i.e. the conditions imposed on the possibility distributions are merely sufficient.

3.2 Nonspecificity

The *nonspecificity* was proposed to quantify a degree of uncertainty resulting from both the values of possibility and the cardinality of the considered subsets of a universe $\mathbf{X}$. A point of resemblance between the AU and the nonspecificity is that it also was originally defined [4] in D-S theory. In order to introduce this information measure, we can't avoid defining the central notion of D-S theory, so called basic assignment. Given a universe $\mathbf{X}$ and a possibility measure Π, *basic assignment* (BA) m is a mapping $m : 2^{\mathbf{X}} \to [0, 1]$ such that for arbitrary $A \subseteq \mathbf{X}$ is $\Pi(A) = \sum_{B \subseteq \mathbf{X}: B \cap A \neq \emptyset} m(B)$. BA can also be easily calculated from any possibility measure and distribution [6]. We have $\sum_{A \subseteq \mathbf{X}} m(A) = 1$ and $m(\emptyset) = 0$.

Definition 2. *Let π be a possibility distribution on $\mathbf{X}$ and m be a corresponding BA. A* nonspecificity *$\mathcal{N}$ of π is then a weighted sum of Hartley functions, i.e. $\mathcal{N}(\pi) = \sum_{A \subseteq \mathbf{X}} m(A) \log_2 |A|$.*

Let us mention the properties of the nonspecificity in order to make a comparison with the aggregate uncertainty complete.

1. *range:* $\mathcal{N}(\pi) \in [0, \log_2 |\mathbf{X}|]$ for any π on $\mathbf{X}$;
2. *maximum:* $\mathcal{N}(\pi) = \log_2 |\mathbf{X}|$ iff π is non-informative;
3. *minimum:* $\mathcal{N}(\pi) = 0$ iff π is degenerated;
4. *subadditivity:* the inequality $\mathcal{N}(\pi_{XY}) \leq \mathcal{N}(\pi_X) + \mathcal{N}(\pi_Y)$ holds for a joint possibility distribution π_{XY} on $\mathbf{X} \times \mathbf{Y}$ with the marginal possibility distributions π_X and π_Y.

Moreover, one can show that $\pi_1 \leq \pi_2$ implies $\mathcal{N}(\pi_1) \leq \mathcal{N}(\pi_2)$. Let us consider the case of T-independent possibilistic variables X, Y and make an effort to derive some results regarding this T-independence and additivity property. First and foremost, the proposition below is straightforward to deduce from the similar discussion in D-S theory [4].

Proposition 4. *Assume that π_G is a T_G-product extension of possibilistic distributions π_X a π_Y. Then*

$$\mathcal{N}(\pi_G) = \mathcal{N}(\pi_X) + \mathcal{N}(\pi_Y). \tag{4}$$

Moreover, in [5], we have proven that T_G is the only t-norm preserving (4). The following assertion is analogous to Prop. 3.

Proposition 5. *Let π_X, π_Y be possibility distributions on $\mathbf{X}, \mathbf{Y}$, respectively, and π_{XY} be their T-product extension where T is an arbitrary t-norm. The equality $\mathcal{N}(\pi_{XY}) = \mathcal{N}(\pi_X) + \mathcal{N}(\pi_Y)$ holds true if at least one of the distributions π_X or π_Y takes only the values from $\{0, 1\}$.*

3.3 Relation Between $\mathcal{U}$ and $\mathcal{N}$

Does the designation 'aggregate uncertainty' of the information measure $\mathcal{U}$ encompass all of its significant properties? Stated more precisely, if $\mathcal{U}$ aggregates the uncertainty of the type 'conflict' as well as the uncertainty of the type 'nonspecificity', we would expect the values of $\mathcal{U}$ to be always higher than those of $\mathcal{N}$.

The proposition below is thus confirming our conjecture. For proof see [5].

Proposition 6. *For any possibility distribution π on $\mathbf{X}$, $\mathcal{U}(\pi) \geq \mathcal{N}(\pi)$.*

4 Conclusions

In the paper, new properties concerning the AU and the nonspecificity were explored; they are contained in Section 3 in the form of propositions. The main result is the inequality between $\mathcal{U}$ and $\mathcal{N}$.

Further space is devoted to an analysis of additivity with respect to a possibilistic T-independence. Rather unsatisfactory results were achieved by adopting the classical additivity concept. We therefore propose to formulate the problem as follows: given a t-norm T and some information measure $\mathcal{I}$, does there exist a binary operation $\oplus_T : [0,\infty)^2 \to [0,\infty)$ with appropriate algebraic properties (for example, commutativity, associativity and neutral element 0) such that the equality

$$\mathcal{I}(\pi_{\mathcal{T}}) = \mathcal{I}(\pi_{\mathcal{X}}) \oplus \mathcal{I}(\pi_{\mathcal{X}})$$

is satisfied for a T-product extension of any possibility distributions π_X and π_Y? An answer to this question in still unknown, further research seems to be neccesary.

References

1. G. de Cooman. Possibility theory I-III. *Int. Journal of General Systems*, (25):291–371, 1997.
2. G. de Cooman and D. Aeyels. Supremum preserving upper probabilities. *Information Sciences*, (118):173–212, 1999.
3. E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms.* Kluwer Academic Publishers, Dordrecht, 2000.
4. G. J. Klir and M. J. Wierman. *Uncertainty-Based Information.* Studies in Fuzziness and Soft Computing. Springer-Verlag, 1998.
5. T. Kroupa. Information measures in possibility theory. In *Proc. of the 5th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, Koyasan, Japan, 2002.
6. Z. Wang and G. J. Klir. *Fuzzy Measure Theory.* Plenum Press, New York, 1992.