1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June - 2 July 1999

A Non-specificity Measure for Convex Sets of Probability Distributions

Joaquín Abellán
Dpto. Lenguajes y Ciencias de la Computación
Campus de Teatinos, Universidad de Málaga
29079 - Málaga - Spain

Serafín Moral
Dpto. Ciencias de la Computación
ETSI Informática, Universidad de Granada
18071 - Granada - Spain

Abstract

In belief functions, there are two types of uncertainty which are due to lack of knowledge: randomness and non-specificity. In this paper, we present a non-specificity measure for convex sets of probability distributions that generalizes Dubois and Prade's non-specificity measure in the Dempster-Shafer theory of evidence.

Keywords. Imprecise probabilities, uncertainty, imprecision, non-specificity.

1 Introduction

The concept of uncertainty is intricately connected to the concept of information. The amount of information obtained by an action must be measured by a reduction in uncertainty.

Shannon's entropy [13] has been the tool for quantifying uncertainty in classic information theory. This function has some desirable properties and has been used as the starting point when looking for another function which can measure uncertainty in situations where a probabilistic representation is not adequate.

In Dempster-Shafer's theory, Yager [14] makes the distinction between two types of uncertainty. One is associated with cases where the information is focused on sets with empty intersections and the other is associated with cases where the information is focused on sets with a cardinal over one. They are called *randomness* and *non-specificity* respectively.

According to Maeda and Ichihashi [10] a function that measures the total uncertainty in a basic probability assignment, b.p.a., should satisfy the following fundamental properties: it coincides with Shannon's Entropy for probabilities, it attains its maximum for the total ignorance and it is monotonic with respect to the inclusion of b.p.a. In Abellan and Moral [1] we have studied Maeda and Ichihashi's measure of Total Uncertainty in D-S theory [10], and we have analyzed its method of quantifying the uncertainty.

There are some situations in which the available information is represented by a convex set of probability distributions, c.s.p.d., as in Walley [15]. Here, we intend to work with convex sets of probability distributions, a model that generalizes the D-S theory. Our starting point is that in imprecise probabilities, we also have two sources of uncertainty: randomness and non-specificity. In general, there are randomness measures which can be easily generalized to imprecise probability, but this is not the case for non-specificity measures. In this paper, we will define a measure of non-specificity that generalizes Dubois and Prade's measure of non-specificity in D-S theory [5].

In Section Two, we consider the fundamentals of the Dempster-Shafer theory in order to establish the basic concepts and the notation. In Section Three, we present a function for c.s.p.d. to measure a non-specificity that generalizes the Dubois and Prade's function, thereby studying its main properties.

2 Uncertainty in D-S Theory

Let X be a finite set considered as a set of possible situations, $|X| = n$, $\wp(X)$ the power set of X and x any element in X

Dempster-Shafer theory [4, 12] is based on the concept of mass assignment. A *mass assignment* is a mapping

$$m : \wp(X) \to [0, 1],$$

such that $m(\emptyset) = 0$ and $\sum_{A \subseteq X} m(A) = 1.$

The value $m(A)$ represents the degree of belief that a specific element of X belongs to set A, but not to any particular subset of A.

The elements A of X for which $m(A) \neq 0$ are called focal elements.

There are two functions associated with each b.p.a.: a belief function, Bel, and a plausibility function, Pl:

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

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

We may note that belief and plausibility are interrelated for all $A \in \wp(X)$

$$Pl(A) = 1 - Bel(\overline{A}),$$

where $\overline{A}$ denotes the complement of A. Furthermore,

$$Bel(A) \leq Pl(A).$$

The measurement of uncertainty was first conceived in terms of the classical set theory. Hartley [8] measured the uncertainty of set A as $\ln |A|$. Therefore, if we want that our measure of non-specificity is a generalization of Hartley's measure, then if m is a b.p.a. focusing on a single set, i.e. $m(A) = 1$ and $m(B) = 0$ if $B \neq A$, then the uncertainty contained in m must be equal to $\ln |A|$.

The classical measure of Entropy [13] is defined by the following continuous function:

$$H(p) = -\sum_{i=1}^{n} p_i \ln(p_i),$$

where $p(p_1, ..., p_n)$ is a probability distribution.

The non-specificity function, introduced by Dubois and Prade [5], represents a measure of imprecision associated with a b.p.a. and is expressed as follows:

$$I(m) = \sum_{A \subseteq X} m(A) \ln(|A|).$$

$I(m)$ attains its minimum, zero, when m is a probability distribution. The maximum, $\ln(|X|)$, is obtained for a b.p.a., m, with $m(X) = 1$ and $m(A) = 0$, $\forall A \subset X$.

Maeda and Ichihashi [10] have proposed an uncertainty function. It quantifies the randomness and non-specificity contained in a b.p.a. on X as

$$UT(m) = I(m) + G(m),$$

where $I(m)$ is Dubois and Prade's non-specificity function and $G(m)$ is the solution of the problem:

$$Max \left\{ -\sum_{x \in X} p_x \ln p_x \right\},$$

where the maximum is taken over all the distribution on C_m, and C_m a closed convex set on $I\!R^{|X|}$, Harmanec and Klir [1994], that is defined as the set of probability distributions $\{(p_x) \mid x \in X\}$ satisfying the constraints:

$$(a)\; p_x \in [0,1] \;\; for \;\; all \;\; x \in X \;\; and \sum_{x \in X} p_x = 1;$$

$$(b)\; Bel(A) \leq \sum_{x \in A} p_x \leq 1 - Bel(\overline{A}) \;\; for \;\; all \;\; A \subseteq X.$$

In Abellan and Moral [1], we introduce a factor with some interesting properties, which can be used as a correction factor to modify Maeda and Ichihashi's measure. The basis is the cross entropy between two probability distributions as introduced by Kullback [9]

$$K(p,q) = \sum_{x \in X} p_x \ln \left(\frac{p_x}{q_x} \right),$$

where p and q are two probability distributions on a finite set X. This function is similar to an information measure and may be considered as a measure of *direct divergence,* [9]. It does not have all the properties of a distance.

We use this function in the following way. Let

$$R(m) = \underset{p \in Fr(C_m)}{Min} K(p, \widehat{q}),$$

where C_m is the c.s.p.d. associated to m, [4], $\widehat{q}$ is such that $G(m) = -\sum_{x \in X} \widehat{q}_x \ln(\widehat{q}_x)$, i.e. the probability distribution with maximum entropy inside C_m, and $Fr(C_m)$ is the frontier set of C_m. We call $R(m)$ the *Kullback Factor of m.*

In Abellan and Moral [1] we propose the following Total Uncertainty measure in D-S theory:

$$UTR(m) = I(m) + G(m) + R(m).$$

Both $G(m)$ and $R(m)$ can be easily generalized to the case of convex set of probability distributions. In fact they are expressed in terms of the convex set C_m associated to a mass assignment m. However, this is not the case of $I(m)$ which is calculated directly from m.

3 A General Function of Non-specificity

We want to measure the non-specificity contained in a general convex set of probability distributions, i.e. including the case of their faces not being parallel to the sides of the Probabilistic Polytope. In a first attempt, we used some functions to quantify the non-specificity in a c.s.p.d. based on the volume of a polytope in $\mathbb{R}^n$, because it is natural to think that the non-specificity is directly related to the volume. But it is difficult to obtain an expression based on volume of a polytope in R^n that is continuous and that is not equal to zero for degenerated polytopes (the non-specificity of a b.p.a. focused on a single set B should be equal to $\ln(|B|)$). For example, in $\mathbb{R}^3$, let C_1 and C_2 be the sets of all convex combinations of $\{(1,0,0);(0,0.5,0.5)\}$ and $\{(1,0,0);(0,0.5,0.5),(0,0.5-\partial,0.5+\partial)\}$, respectively. When $\partial \to 0$ then the non-specificity of C_1 and C_2 should be two similar quantities. However, since they have different dimensions, $\dim(C_1)=1$ and $\dim(C_2)=2$, this adjustment is complex. The volume of C_2 converges to 0 and the non-specificity of C_1 should be different from 0. Problems also arise when dealing with the monotonic property.

Finally, we try to generalize Dubois and Prade's non-specificity measure [5] to the general case of convex sets of probability distributions.

If we know the set of vertices in a convex set, then this set is completely determined. In order to solve the problem of computing all the vertices of a polytope. Mattheiss and Rubin [11], provide some methods for finding them.

Some concepts first need to be defined: the lower capacity associated to a convex set and its Möbius inverse [3].

DEFINITION 1. Let C be a c.s.p.d. on a universal X. We define the following capacity function:

$$f(A) = \underset{P\in C}{Inf}\; P(A), \quad \forall A \in \wp(X),$$

where $\wp(X)$ is the power set of X.

DEFINITION 2. For any mapping $f : \wp(X) \to \mathrm{R}$ then the mapping $m : \wp(X) \to \mathrm{R}$, given by

$$m(A) = \sum_{B\subseteq A} (-1)^{|A-B|} f(B), \;\; \textit{for all } A \in \wp(X),$$

will be called the Möbius inverse of f.

This correspondence proves to be one-to-one, since conversely,

$$f(A) = \sum_{B\subseteq A} m(B), \;\; \textit{for all } A \in \wp(X),$$

as we can see in Shafer [12].

DEFINITION 3. Let C be a c.s.p.d. on a universal X, f its minimum lower probability as in Definition 1 and let m be its Möbius inverse. We say that m is the *assignment of masses* of C. And we call any $A \in X$ such that $m(A) \neq 0$, a focal element of m.

We can now define a general function of non-specificity.

DEFINITION 4. Let C be a c.s.p.d. on a universal X. Let m the assignment of masses associated to C. We define C the non-specificity of C as,

$$IG(C) = \sum_{A\subseteq X} m(A)\ln(|A|).$$

This extends to c.s.p.d. the non-specificity function defined by Dubois and Prade for belief functions.

This function takes, for a c.s.p.d., the non-specificity value of a larger set: the minimum capacity associated. However, this is not a problem because we do not add non-specificity. In an extreme case, like the one in the following example, we may see that the added probabilities do not decrease the specificity of the associated c.s.p.d.

Example 1. Let C be a c.s.p.d. on $X = \{x_1, x_2, x_3\}$ such that it is the set of convex combinations of the vertices $\{(0,0,1),(0.5,0.5,0)\}$. Then, let $f(A) = \underset{P\in C}{Inf}\; P(A), \;\; \forall A \in \wp(X)$ its associated capacity. Then, it coincides with the associated capacity of the set, C', of convex combinations of the vertices $\{(0.5,0.5,0),(0.5,0,0.5),(0,0.5,0.5),(0,0,1)\}$.

These can be seen in a simplex representation in Fig 1 and Fig 2, respectively. It is clear that $C \subset C'$ and that the inclusion is strict. However, we consider that there are no differences in the amount of imprecision in both cases. Having probability distributions $(0,0,1)$ and $(0.5,0.5,0)$, then the other distributions, $(0,0.5,0.5)$ and $(0.5,0,0.5)$, do not add imprecision. In convex set C, there is a mass of 1 which can be moved from x_3 to x_1 and x_2 (half of the mass to each one of them). With the new probability distributions the same mass can move from x_3 to these elements, but now it is possible that half of the mass moves first to one element and then to the other afterwards. That is, from $(0,0,1)$ we can obtain $(0,0.5,0.5)$ and then $(0.5,0.5,0)$ if we move first to x_2 and then to x_1, or from $(0,0,1)$ we can obtain $(0.5,0,0.5)$ and then $(0.5,0.5,0)$ if we move first to x_1 and then to

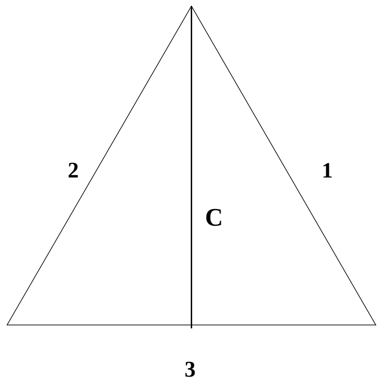

Fig. 1

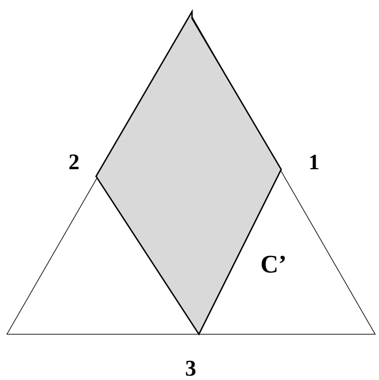

Fig. 2

x_2. Allowing to move the mass in two steps do not add imprecision: the moved mass and the initial and final points are the same.

Perhaps, if one thinks that the imprecision of both sets should be different, this could be due to the fact that they have really different dimensions and the area of C is 0, while the area of C' is greater than 0. But as we have said, the amount of imprecision is not equal to the volume (or area in the case of two dimensions). Very different volumes can have similar amount of imprecision. Something similar happens in the D-S theory. If we consider m and m' b.p.a.s on the same set X, such that:

$$m\{x_1, x_2\} = 1 \ and \ m(A) = 0 \ if \ A \neq \{x_1, x_2\}$$

$$m'\{x_i, x_j\} = \frac{1}{3}, \ \forall i, j, \ \ and$$
$$m'(A) = 0 \ if \ A \neq \{x_i, x_j\},$$

there is a noticeable difference between the dimensions of these sets and yet, $I(m) = I(m') = \ln(2)$. The variation of uncertainty between C and C' is of randomness type.

DEFINITION 5. Let C be a c.s.p.d. on a universal $X \times Y$. Then

$$C_X = \left\{ p_X : \exists p \in C \ such \ that \ p_X(x) = \sum_{y \in Y} p(x,y) \right\}$$

is called the marginal c.s.p.d. of C on X, where $\Pr_X(A)$ is the projection of the set $A \in X \times Y$ on X. Analogously for C_Y.

DEFINITION 6. Let C be a c.s.p.d. on a universal $X \times Y$, let m be an assignment of masses on C. Let C_X and C_Y be its marginals c.s.p.d. and m_X and m_Y its assignments of masses respectively. We say that there is random set independence under C iff $m(A \times B) = m_X(A) m_Y(B)$, with $A \in X$ and $B \in Y$.

3.1 Properties

With the above notation, function IG satisfies the same properties as function I, [6].

Property 1. It is zero for probability distributions.

Proof. It is immediate because $m(A) = 0 \ \forall A \subseteq X$ such that $|A| \geq 2$.

Property 2. It is monotonic, i.e. if C and C' are two c.s.p.d. on X such that $C \subseteq C'$ then $IG(C) \leq IG(C')$.

Proof. It is an immediate consequence of Lemma 4 in the Appendix.

Property 3. It is well defined, $IG(C) \geq 0 \ \forall C$ c.s.p.d. on X.

Proof. By Property 1 and Property 2.

Property 4. It is maximal for the total ignorance with a range in $[0, \ln(n)]$, where $n = |X|$.

Property 5. It is additive, i.e. C is a c.s.p.d. on a universal $X \times Y$ such that there is random set independence under C then $IG(C) = IG(C_X) + IG(C_Y)$.

Proof. It is the same proof as for I by Dubois and Prade [5].

Property 6. It is subadditive, i.e. if C is a c.s.p.d. on a universal $X \times Y$, then $IG(C) \leq IG(C_X) + IG(C_Y)$.

Proof. We introduce a function m' on $X \times Y$ such that $m'(A \times B) = m_X(A) m_Y(B)$. It is an assignment of masses for some C' c.s.p.d. on $X \times Y$.

We now define the set

$$C_X \times C_Y = \{p_X p_Y : p_X \in C_X \ and \ p_Y \in C_Y\}$$

of probability distributions on $X \times Y$, which is generally not a convex set.

Using Abellan and Moral [1] and Harmanec and Klir [7] we have

$$C \subseteq convex\ hull\ of\ (C_X \times C_Y) \subseteq C'.$$

And by Property 2 and Property 5,

$$IG(C) \leq IG(C') = IG(C_X) + IG(C_Y).$$

□

4 Conclusions

In this paper, we have shown that Dubois and Prades's non-specificity measure [5] may be extended to general convex sets and that it verifies similar properties. This process is based on assigning a lower probability, f, to a convex set of probability distributions, and then calculating the non-specificity using the inverse Möbius transformation of f.

This measure can be added with other factors of uncertainty in order to obtain a total uncertainty measure for convex sets of probability distributions.

Appendix (proof of Property 2)

First, we need some properties of the successive differences operator on a function real of variable real f, i.e.,

$$\Delta_h^{k+1} f(x) = \Delta_h^1(\Delta_h^k f(x)); \quad h \in \mathrm{R} \ \ and \ \ k \in \mathrm{N}$$

where $\Delta_h^1 f(x) = f(x+h) - f(x)$ and $\Delta_h^0 f(x) = f(x)$

Lemma 1. This operator is lineal, i.e.:

(1) $\Delta_h^k [f(x) + g(x)] = \Delta_h^k f(x) + \Delta_h^k g(x)$.

(2) $\Delta_h^k [\lambda f(x)] = \lambda \Delta_h^k f(x),\ \lambda \in I\!R$.

Lemma 2. It satisfies the following equality:

$$\Delta_h^k f(x) = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} f(x+ih)$$

Proof. By induction on k

$\Delta_h^2 f(x) = f(x+2h) - f(x+h) - (f(x+h) - f(x)) = f(x+2h) - 2f(x+h) + f(x)$.

$\Delta_h^3 f(x) = f(x+3h) - 2f(x+2h) + f(x+h) - [f(x+2h) - 2f(x+h) + f(x)] =$

$= f(x+3h) - 3f(x+2h) + 3f(x+h) - f(x)$.

Now,

$\Delta_h^{k+1} f(x) =$

$\Delta_h^1(\Delta_h^k f(x)) = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} f(x+(i+1)h) - \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} f(x+ih) =$

$= -(-1)^k \binom{k}{0} f(x) + \left[(-1)^k \binom{k}{0} f(x+h) - (-1)^{k-1} \binom{k}{1} f(x+h)\right] +$

$+ \left[(-1)^{k-1} \binom{k}{1} f(x+2h) - (-1)^{k-2} \binom{k}{2} f(x+2h)\right] + \ldots + (-1)^{k-k} \binom{k}{k} f(x+(k+1)h) =$

$= (-1)^{k+1} \binom{k}{0} f(x) + (-1)^{k+1-1} \binom{k+1}{1} f(x+h) + (-1)^{k+1-2} \binom{k+1}{2} f(x+2h) + \ldots +$

$+(-1)^{k+1-(k+1)} \binom{k+1}{k+1} f(x+(k+1)h) = \sum_{i=0}^{k+1} (-1)^{k+1-i} \binom{k+1}{i} f(x+ih)$. □

Lemma 3. Let $f(x) = \ln(x)$, $h = 1$ and $x \geq 1$, then $\Delta_1^{2k} f(x) \leq 0$ and $\Delta_1^{2k+1} f(x) \leq 0 \ \ \forall k$.

Proof. We know that the derivates of $f(x)$ satisfies that $f^{(2k)}(x) \leq 0$ and $f^{(2k+1)}(x) \geq 0$. Then $f^{(2k)}$ is always a concave function and $f^{(2k+1)}$ a convex function.

Let $g_1(x) = \Delta_1^2 f(x) = f(x+2) - 2f(x+1) + f(x)$. Since f'' is a concave function then

$$\frac{1}{2} f''(x+2) + \frac{1}{2} f''(x) \leq f''(x+1)$$

and

$$g_1''(x) = f''(x+2) - 2f''(x+1) + f''(x) \leq 0.$$

Hence $g_1(x)$ is a concave function.

Repeating the process, we have that

$$g_1^{(2k+2)}(x) =$$

$$f^{(2k+2)}(x+2) - 2f^{(2k+2)}(x+1) + f^{(2k+2)}(x) \leq 0, \ \ \forall k$$

and $g_1^{(2k)}(x)$ is a concave function.

Analogously, we can define $g_j(x) = \Delta_1^2 g_{j-1}(x)$ and then $g_j^{(2k)}(x)$ are concave functions $\forall k$ and $j = 1, 2, \ldots$, where we call $g_0(x) = f(x)$.

By the concavity property,

$\Delta_1^2 f(x) = f(x+2) - 2f(x+1) + f(x) \leq 0$

$\Delta_1^4 f(x) = \Delta_1^2 g_1(x) = g_1(x+2) - 2g_1(x+1) + g_1(x) \leq 0$

....

$\Delta_1^{2k} f(x) = \Delta_1^2 g_{k-1}(x) = g_{k-1}(x+2) - 2g_{k-1}(x+1) + g_{k-1}(x) \leq 0.$

Using a similar argument, $g_j^{(2k+1)}(x)$ are convex functions, $\forall k, j$, and we have that

$$g_j^{(2k+1)}(x) =$$

$$g_{j-1}^{(2k+1)}(x+2) - 2g_{j-1}^{(2k+1)}(x+1) + g_{j-1}^{(2k+1)}(x) \geq 0.$$

Hence $g_j^{(2k)}$ are non-decreasing functions, $\forall k, j$.

Obviously, if $w(x)$ is an non-decreasing function then $\Delta_1^1 w(x) \geq 0$. Now,

$$\Delta_1^{2k+1} f(x) = \Delta_1^1(\Delta_1^{2k} f(x)) = \Delta_1^1(g_{k-1}(x)) \geq 0 \qquad \square$$

Lemma 4. Let f, f' be two monotonous capacities on a universal X. Let m, m' be its Möbius inverses respectively. If exist $A \in X$ such that $f(A) - \epsilon = f'(A)$, $\epsilon > 0$, and $f(B) = f'(B)$ for $B \neq A$, then

$$\sum_{C \subseteq X} m(C) \ln(|C|) \leq \sum_{C \subseteq X} m'(C) \ln(|C|)$$

Proof. It is easy to prove that

$$\sum_{C \subseteq X} m'(C) \ln(|C|) - \sum_{C \subseteq X} m(C) \ln(|C|) =$$

$$-\epsilon \sum_{C \supseteq A} (-1)^{|C-A|} \ln(|C|).$$

If we denote $x = |A|$ and $N = |X| - x$, then we have:

$$\sum_{C \supseteq A} (-1)^{|C-A|} \ln(|C|) =$$

$$\sum_{i=0}^{N} (-1)^i \begin{pmatrix} N \\ i \end{pmatrix} \ln(x+i) =$$

$$(-1)^N \sum_{i=0}^{N} (-1)^{N-i} \begin{pmatrix} N \\ i \end{pmatrix} \ln(x+i) =$$

$$(-1)^N \Delta_1^N \ln(x),$$

by Lemma 2.

Now, by Lemma 3

$$\sum_{C \subseteq X} m'(C) \ln(|C|) - \sum_{C \subseteq X} m(C) \ln(|C|) =$$

$$(-\epsilon)(-1)^N \Delta_1^N \ln(x) \geq 0. \qquad \square$$

Now, Property 2 is an immediate consequence of this Lemma 4.

References

[1] J. Abellan and S. Moral. Completing a total uncertainty measure in D-S theory. *Int. J. General System*, 1999 (to appear).

[2] A. Chateauneuf and J.Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius Inversion. *Math. Soc. Sc.*, 17:263-283, 1989.

[3] G. Choquet. Théorie des capacités. *Ann. Inst. Fourier*, Grenoble, 5:131-292, 1953/54.

[4] A.P. Dempster. Upper and lower probabilities induced by a multivaluated mapping. *Ann. Math. Statistic* 38:325-339, 1967.

[5] D. Dubois and H. Prade. A note on measure of specificity for fuzzy sets. *BUSEFAL*, 19:83-89, 1984.

[6] D. Dubois and H. Prade. Properties and measures of information in evidence and possibility theories. *Fuzzy Sets and System*, 24:183-196, 1987.

[7] D. Harmanec and G.J. Klir. Measuring total uncertainty in Dempster-Shafer Theory: a novel approach. *Int. J. General System*, 22:405-419, 1994.

[8] R.V.L. Hartley. Transmission of information. *The Bell Systems Technical Journal* 7:535-563, 1928.

[9] S. Kullback. *Information Theory and Statistics.* Dover, 1968.

[10] Y. Maeda and H. Ichihashi. An uncertainty measure with monotonicity under the random set inclusion. *Int. J. General System*, 21:379-392, 1993.

[11] T.H. Mattheiss and D.S. Rubin. A survey and comparison of methods for finding all vertices of convex polyhedral sets. *Math. Oper. Res.* 5:167-185, 1980.

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

[13] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal* 27:379-423,623-656, 1948.

[14] R.R. Yager. Entropy and specificity in a mathematical theory of evidence. *Int. J. General System* 9:249-260, 1983.

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

[16] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* 1:3-28, 1978.