11

Entropy and Specificity in a Mathematical Theory of Evidence

Ronald R. Yager

Abstract. We review Shafer's theory of evidence. We then introduce the concepts of entropy and specificity in the framework of Shafer's theory. These become complementary aspects in the indication of the quality of evidence.

Key words: Entropy, Fuzzy sets, Specificity, Belief, Plausibility.

1 Introduction

In [1] Shafer presents a comprehensive theory of evidence. The problem of concern to Shafer is the location of some special element in a set X, called the frame of discernment or base set. In Shafer's framework he is provided with evidence as to the identity of this special element in terms of a mapping from the power set of X (set of all subsets of X) into the unit interval. This mapping which Shafer calls the basic assignment, associates with each subset A of X, the degree of belief that the special element is located in the set A with the understanding that he can't make any more precise statement with regards to the location of the element.

A significant aspect of Shafer's structure is the ability to represent in this common framework various different types of uncertainty, i.e. probabilistic uncertainty and possibilistic uncertainty. Our purpose here is to take some concepts developed in these individual frameworks and generalize them to the comprehensive framework of Shafer. In particular we shall generalize the idea of entropy from the probabilistic framework and specificity from the possibilistic framework. We shall find that these two measures of uncertainty provided complementary measures of the quality of a piece of evidence.

2 Shafer's Theory of Evidence

In Ref. 1 Shafer presents a comprehensive theory of evidence based on the concept of belief. The theory begins with the idea of using a number between zero and one to indicate the degree of support a body of evidence provides for a proposition. The fundamental concept in Shafer's theory is the basic assignment.[1]

Definition 1. *Assume m is a set mapping from subsets of the finite set X into the unit interval*

$$m : 2^X \to [0, 1]$$

such that

1) $m(\varnothing) = 0$
2) $\sum_{A \subset X} m(A) = 1$

m is then called a basic assignment.

The interpretation of m consistent with Shafer's theory is that there exists in the base set X some special unknown element u and $m(A)$ is the degree of belief that this element lies in the set A and nothing smaller than A. In order to help in the understanding of this concept I quote several attempts at clarification from Shafer [1].

"$m(A)$ is the belief that the smallest set that the outcome is in is A."

"$m(A)$ measures the total portion of belief that is confined to A yet none of which is confined to any proper subset of A."

"m(A) measures the belief mass that is confined to A but can move to every point of A."[2]

Note — The formulation of m leads us to the following observations:

1) $m(X)$ is *not* necessarily one.
2) $A \subset B$ does *not* necessarily imply $m(A) \leqq m(B)$.
3) It allows that belief not be committed to either A or not A.

Having introduced the idea of the basic assignment Shafer next introduces the concept of a belief function.

Definition 2. *Given a basic assignment m we can define a belief function*

$$\mathrm{Bel} : 2^X \to [0, 1]$$

such that for any $A \subset X$

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

---

[1] I have chosen to use the term basic assignment where Shafer uses the term basic probability assignment. I feel that the use of the word probability conjures up certain preconceived notions in the reader which I want to avoid.

[2] If the special element u is the age of some person, then $m(A)$ may measure the degree to which we believe that u is contained in the set *young*, where $A =$ young is defined as a subset of X.

Bel(A) *measures the belief that the special element is a member of* A*. Whereas* $m(A)$ *measures the amount of belief that one commits exactly to* A *alone, Bel*(A) *measures the total belief that the special element is in* A*.*

A subset A of X is called a focal element of a belief function Bel if $m(A) > 0$.

Shafer shows that $\text{Bel}(\varnothing) = 0$, $\text{Bel}(X) = 1$ and that for every collection $A_1, A_2, \ldots, A_n$ of subsets of X

$$\text{Bel}(A_1 \cup A_2 \ldots \cup A_n) \geqq \sum_{\substack{I \subset \{1,2,\ldots,n\} \\ I \neq \varnothing}} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{i \in I} A_i\right),$$

where $|I|$ denotes the cardinality of the set I.

Shafer also shows that a belief function uniquely determines an underlying basic assignment,

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} \text{Bel}(B),$$

$|A - B|$ indicates the cardinality of the elements in A not in B.

Shafer next defines the plausibility associated with A.

Definition 3. *Given a belief function* $\text{Bel} : 2^X \to [0,1]$ *we define a plausibility function Pl as,*

$$\text{Pl} : 2^X \to [0,1]$$

such that for any $A \subset X$

$$\text{Pl}(A) = 1 - \text{Bel}(\bar{A}).$$

Note — The following observations can be made with respect to P1:

1) $\text{Pl}(A)$ measures the degree to which one fails to doubt A, where dou $(A) = \text{Bel}(\bar{A})$
2) $\text{Pl}(A)$ measures the total belief mass that can move into A, whereas $\text{Bel}(A)$ measures the total belief mass that is constrained to A.
3) $\text{Pl}(A) = \sum_{B \cap A \neq \varnothing} m(B)$
4) $\text{Bel}(A) \leqq \text{Pl}(A)$

An important aspect of Shafer's theory involves the combination of belief functions to form a resulting belief function, that is, the combining of various sources of evidence. Shafer accomplishes this by use of Dempster's Rule of Combination. Zadeh [2] has raised some questions as to the appropriateness of this rule. Prade [3] has shown the relationship between Dempster's rule and the intersection of fuzzy sets. Smets [4] has used Shafer belief functions in medical diagnosis. Nguyen [5] has discussed the relationship between belief functions and random sets.

While we shall not in this paper be concerned with the question of the combination of evidence, we shall use a concept developed by Shafer in his approach to combining evidence.

Definition 4. *Assume* Bel_1 *and* Bel_2 *are two belief functions over* 2^X *with their associated basic assignments* m_1 *and* m_2*. The weight of conflict between* Bel_1 *and* Bel_2*, denoted* $\mathrm{Con}(\mathrm{Bel}_1, \mathrm{Bel}_2)$*, is defined as*

$$\mathrm{Con}\,(\mathrm{Bel}_1, \mathrm{Bel}_2) = -\ln(1-k)$$

where

$$k = \sum_{\substack{i,j \\ A_i \cap B_j = \varnothing}} m_1(A_1) \cdot m_2(B_j).$$

The situation of no conflict occurs when $k = 0$ and hence $\mathrm{Con}(\mathrm{Bel}_1, \mathrm{Bel}_2) = 0$. If Bel_1 and Bel_2 are flatly contradictory $k = 1$ and $\mathrm{Con}(\mathrm{Bel}_1, \mathrm{Bel}_2) = \infty$. Thus $\mathrm{con}(\mathrm{Bel}_1, \mathrm{Bel}_2) \geqq 0$ and increases with increasing conflict.

3 Types of Belief Functions

Shafer introduces various classes of belief functions. We shall discuss some of these in the following.

Definition 5. *A belief function over* 2^X *is called a vacuous belief function if*

$$\mathrm{Bel}(X) = 1 \text{ and } \mathrm{Bel}(A) = 0 \text{ for } A \neq X.$$

Note

1) If Bel is a vacuous belief function, then $m(X) = 1$ and $m(A) = 0$ for $A \neq X$.
2) Vacuous belief functions are used in situations where there is no evidence.

Definition 6. *A belief function is called a simple support function focused at* A *if*

$$\mathrm{Bel}(B) = \begin{cases} 0 & \text{if} \quad A \not\subset B \\ 1 & \text{if} \quad B = X \\ s & \text{if} \quad A \subset B, B \neq X. \end{cases} \qquad \text{for} \quad 0 < s < 1$$

Note If Bel is a simple support function focussed at A, then its basic assignment function m is:

$$m(A) = \mathrm{Bel}(A) = s$$
$$m(X) = 1 - \mathrm{Bel}(A) = 1 - s$$
$$m(B) = 0 \text{ for all others.}$$

The simple support function focused at A is used to indicate the situation that we think the special outcome is in A with belief s.

We shall call the simple support function focused at A with $m(A) = 1$ the *certain support function focused* at A.

Definition 7. *A belief function on* 2^X *is said to be a Bayesian belief function if*

$$\text{Pl}(A) = \text{Bel}(A) \text{ for all } A \subset X.$$

Note The following are two equivalent formulations of a Bayesian belief function.

I) $\text{Bel}(\varnothing) = 0$
   $\text{Bel}(X) = 1$
   $\text{Bel}(A \cup B) = \text{Bel}(A) + \text{Bel}(B)$, whenever $A \cap B = \varnothing$

II) $\text{Bel}(A) + \text{Bel}(\bar{A}) = 1$

Theorem 1. *If Bel is a Bayesian belief function, then the basic assignment m is such that m takes non-zero values for only subsets of X that are singletons. Hence*

$$\sum_{x \in X} m(\{x\}) = 1$$

The Bayesian structure implies that none of the evidence mass has freedom of movement.

The Bayesian structure forms the prototype in Shafer's theory for probabilistic uncertainty in which the basic assignment function m plays the role of the probability distribution function p. That is, every probability distribution $p : 2^X \to [0, 1]$ can be associated with a Bayesian belief function in which $p(x) = m(\{x\})$.

We note that a Bayesian structure is fully defined by a point function of X equal to $m(\{x\})$.

Since

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

and since $\text{Pl}(A) = \text{Bel}(A)$ for Bayesian belief structure,

$$\text{Pl}(A) = \sum_{x \in A} m(\{x\}).$$

Furthermore,

$$\text{Bel}(\{x\}) = \text{Pl}(\{x\}) = m(\{x\}).$$

Hence

$$\text{Bel}(A) = \sum_{x \in A} \text{Bel}(\{x\}) = \text{Pl}(A) = \sum_{x \in A} \text{Pl}(\{x\}).$$

Definition 8. *A belief function Bel:* $2^X \to [0, 1]$ *is said to be consonant if*

1) $\text{Bel}(\varnothing) = 0$
2) $\text{Bel}(X) = 1$
3) $\text{Bel}(A \cap B) = \text{Min}(\text{Bel}(A),\ \text{Bel}(B))$ for all $A,\ B \subset C$

Note — The following are two equivalent formulations of a consonant belief function:

1) $\mathrm{Pl}(A \cup B) = \mathrm{Max}(\mathrm{Pl}(A),\ \mathrm{Pl}(B))$
2) $\mathrm{Pl}(A) = \mathrm{Max}_{x \in A}[\mathrm{Pl}(\{x\})]$ for all $A \neq \varnothing$

Note — Every simple support function is consonant.

Note — If Bel is a consonant belief function, then for all $A \subset X$ either $\mathrm{Bel}(A) = 0$ or $\mathrm{Bel}(\bar{A}) = 0$.

The characterization of a consonant belief function is expressed by the following theorem (Shafer).

Theorem 2. *A belief function is consonant if the focal elements of its basic assignment function m are nested. That is, if there exists a family of subsets of X, A_i, $i = 1, 2, \ldots, n$, such that $A_i \subset A_j$ for $i < j$ and $\Sigma_i m(A_i) = 1$.*

Note — A consonant belief structure is completely determined by a point function

$$f : X \to [0, 1]$$

such that $f(x) = \mathrm{Pl}(\{x\})$. At least one element $x \in X$, has $f(x) = 1$. This follows since for any $A \subset X$, $\mathrm{Pl}(A) = \mathrm{Max}_{x \in A}[\mathrm{Pl}(x)]$. Hence Pl is completely determined by Pl defined over the point set X. Since $\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\bar{A})$, $\mathrm{Bel}(A)$ is also uniquely determined. Since $\mathrm{Bel}(A)$ uniquely determines m we have completely defined the structure from this mapping.

This relationship can be made even clearer with the following construction suggested by Prade [2].

Assume we have a consonant belief structure. We can always build a nested sequence of sets

$$\{x_1\} \subset \{x_1, x_2\} \subset \{x_1, x_2, x_3\} \subset \ldots \subset X,$$

indicating these sets as $A_1 \subset A_2 \subset A_3 \ldots \subset A_n = X$ such that $\Sigma_{i=1}^{n} m(A_i) = 1$. Hence all the belief mass lies in this nested sequence. (Some of the elements in the sequence may have zero basic assignment but any subset not in the sequence definitely has zero basic assignment.)

Since

$$\mathrm{Pl}(B) = \sum_{B \cap A \neq \varnothing} m(A),$$

$$\mathrm{Pl}(\{x\}) = \sum_{\{x\} \cap A \neq \varnothing} m(A) = \sum_{\substack{i \\ \{x\} \cap A_{i \neq \varnothing}}} m(A_i) = \sum_{\substack{i \\ \text{such that} \\ x \in A_i}} m(A_l).$$

Therefore

$$\mathrm{Pl}(\{x_1\}) = m(A_1) + m(A_2) + \ldots m(A_{n-1}) + m(X)$$

$$\begin{array}{llll} \mathrm{Pl}(\{x_2\}) = & m(A_2) + & m(A_{n-1}) + m(X) & \\ \mathrm{Pl}(\{x_n\}) = m(X) & \vdots & \vdots & \vdots \end{array}$$

Conversely

$$m(A_i) = \mathrm{Pl}(\{x\}) - \mathrm{Pl}(\{x_{i+1}\})$$
$$m(X) = \mathrm{Pl}(\{x_n\})$$
$$m(A) = 0 \text{ for all else.}$$

The consonant belief structure forms the prototype for the possibilistic type of uncertainty introduced by Zadeh [6] in which the plausibility measure in Shafer's theory plays the role of the possibility measure π in Zadeh's theory. Furthermore, since $\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\bar{A})$, the belief function is analagous to Zadeh's measure of certainty [6].

The representations of both these common types of uncertainty in a similar format allows for a comparison of the two types of uncertainty. We see that in a certain respect possibilistic and probabilistic (consonant and Bayesian) uncertainty are opposite extremes. Whereas possibilistic uncertainty assigns its beliefs mass m to a nested sequence of sets, probabilistic uncertainty assigns its belief mass to a collection of disjoint sets. There exists only one type of belief structure which satisfies both structures.

Theorem 3. *The certain support function focused at $\{x\}$, i.e., such that $m(\{x\}) = 1$ for some $x \in X$ is the only belief function that is both a Bayesian and a consonant belief function.*

Note — This structure is a certainty structure in that we know that the special element is x.

4 Entropy Like Measure

An important concept in the theory of probability is Shannon's measure of entropy for a probability distribution. This is a measure of the discordance associated with a probability distribution. We shall introduce here a measure of entropy associated with a basic assignment function m.

Definition 9. *Assume that m is a basic assignment over 2^X with associated belief function Bel.*

We define the entropy of m as

$$Em = \sum_{A \subset X} m(A) \cdot \mathrm{Con}(\mathrm{Bel},\ \mathrm{Bel}_A)$$

where Bel_A is the certain support function focused at A. The next theorem justifies our use of the term entropy.

Theorem 4. *Assume that m is a Bayesian structure. Then*

$$Em = -\sum_{x \in X} m(x) \cdot \ln m(x).$$

Proof.

$$Em = \sum_{A \subset X} m(A) \cdot \text{con}\left(\text{Bel},\ \text{Bel}_A\right).$$

Since for a Bayesian structure $m(A) = 0$ for all non-singletons,

$$Em = \sum_{x \in X} m(\{x\}) \cdot \text{con}\left(\text{Bel},\ \text{Bel}_A\right).$$

We shall denote the basic assignment function associated with the certain support function at $\{x\}$, by g_x. Then

$$g_x\left(\{x\}\right) = 1$$

$g_x(B) = 0$ for all other $B \subset X$, and Con $(\text{Bel},\ \text{Bel}_{\{x\}}) = -\ln(1-k)$, where

$$k = \sum_{\substack{i,j \\ \text{for } A_i \cap B_j = \varnothing}} m(A_i) \cdot g_x\left(B_j\right).$$

Since $g_x(B) = 0$ for $B \neq \{x\}$ and elsewhere equals 1,

$$k = \sum_{\substack{i \\ \text{for } A_i \cap \{x\} = \varnothing}} m(A_i)$$

Since m is Bayesian,

$$k = \sum_{\substack{i \\ \{x_i\} \cap \{x\} = \varnothing}} m\left(\{x_i\}\right) = \sum_{\substack{i \\ \text{for } x_i \neq x}} m\left(\{x_i\}\right) = 1 - m\left(\{x\}\right).$$

Thus

$$\begin{aligned}\text{Con}(\text{Bel}, \text{Bel}_{\{x\}}) &= -\ln\left(1 - \left(1 - m\left(\{x\}\right)\right)\right) \\ &= -\ln\left(m\left(\{x\}\right)\right),\end{aligned}$$

hence

$$Em = -\sum_{x \in X} m\left(\{x\}\right) . \ln m\left(\{x\}\right).$$

Thus this definition reduces to the Shannon entropy when the belief structure is Bayesian.

As a simplification for our further work we note that

$$\mathrm{Con}(\mathrm{Bel},\ \mathrm{Bel}_A) = -\ln(1-k)$$
$$\text{and } k = \sum_{\substack{i,j \\ \text{for } A_i \cap B_j = \varnothing}} m(A_i)\cdot m_A(B_j).$$

But since m_A is such that $m_A(A) = 1$ and elsewhere it is zero,

$$k = \sum_{\substack{i \\ A_i \cap A = \varnothing}} m(A_i).$$

However, since

$$1 = \sum_{A_i \subset A} m(A_i) = \sum_{A_i \cap A = \varnothing} m(A_i) + \sum_{A_i \cap A \neq \varnothing} m(A_i)$$

and since

$$\sum_{A_i \cap A \neq \varnothing} m(A_i) = \mathrm{Pl}(A)$$

it follows that

$$1 - k = \mathrm{Pl}(A),$$

where $\mathrm{Pl}(A)$ is the plausibility function associated with A under m. Thus

$$\mathrm{Con}(\mathrm{Bel},\ \mathrm{Bel}_A) = -\ln(\mathrm{Pl}(A)).$$

Hence

$$Em = -\sum_{A \subset X} m(A)\cdot \ln(\mathrm{Pl}(A)) - \sum_{A \subset X} \ln(\mathrm{Pl}(A))^{m(A)}$$

Thus we have proved the following.

Theorem 5. *For a belief structure with basic assignment m and plausibility Pl the entropy of this structure is*

$$Em = -\sum_{A \subset X} \ln\left(\mathrm{Pl}(A)^{m(A)}\right) = -\sum_{A \subset X} m(A)\cdot \ln \mathrm{Pl}(A):$$

Corollary 1.

$$e^{E_m} = \prod_{A \subset X} \left(\mathrm{Pl}(A)^{-m(A)}\right).$$

Proof.

$$e^{E_m} = e^{-\left(\Sigma \mathrm{Pl}(A)^{m(A)}\right)} = \prod_{A \subset X} e^{-\ln\left(\mathrm{Pl}(A)^{m(A)}\right)}$$
$$= \prod_{A \subset X} \left(\mathrm{Pl}(A)^{-m(A)}\right)$$

Since $\mathrm{Pl}(A) \in [0,1]$ for all $A \subset X$ then $\ln \mathrm{Pl}(A) \leqq 0$ and since $m(A) \in [0,1]$ then

$$Em = -\sum_{A \subset X} m(A) \cdot \ln\left(\mathrm{Pl}\left(A\right)\right) \geqq 0.$$

Thus Em assumes as its minimal value the value zero.

Let us look at the belief structures which take this minimal value for Em.

Theorem 6. *For any simple support belief structure $Em = 0$.*

Proof. Assume our simple support structure is focused at B, with $m(B) = b$. Then since

$$Em = -\sum_{A \subset X} m(A) \cdot \ln \mathrm{Pl}\left(A\right),$$

and since for this type of belief function $m(B) = b,\ m(X) = 1 - b$ and for all sets A not equal to B or $X,\ m(A) = 0$, it then follows that

$$Em = -\left(b \cdot \ln\ \mathrm{Pl}\left(B\right)\right) + \left(\left(1 - b\right) \cdot \ln\ \mathrm{Pl}\left(X\right)\right).$$

Since

$$\begin{aligned} \mathrm{Pl}(A) &= 1 - \mathrm{Bel}(\bar{A}) \text{ we have} \\ \mathrm{Pl}(X) &= 1 - \mathrm{Bel}(\varnothing) = 1 - 0 = 1 \\ \mathrm{Pl}(B) &= 1 - \mathrm{Bel}(\bar{B}) = 1 - 0 = 1 \end{aligned}$$

from which we get $E_m = -(b \ln 1 + (1 - b) \ln 1) = 0$.

A more general classification of belief structures with zero entropy can be obtained.

Lemma 1. *Any belief structure for which the plausibility is one at all focal elements has $Em = 0$.*

Proof. This follows directly from

$$Em = -\sum_{A \subset X} m(A) \cdot \ln\ \mathrm{Pl}\left(A\right)$$

and the fact that $\ln 1 = 0$.

Lemma 2. *In a consonant belief structure the plausibility function is one at focal elements.*

Proof. Because of the nested nature of the focal elements of this structure there exists at least one $x \in X$ contained in all the focal elements, denote this x^*.

From the definition of plausibility it follows that

$$\mathrm{Pl}\left(\{x^*\}\right) = \sum_{A\cap\{x^*\}\neq\varnothing} m\left(A\right)$$

Since x^* is contained in all focal elements then $\mathrm{Pl}\{x^*\} = \sum_i m(A_i) = 1$, where Ai are all the focal elements.

We note that for any $A \subset X$

$$\mathrm{Pl}\left(A\right) = \underset{x\in A}{\mathrm{Max}}\left[\mathrm{Pl}\left\{x\right\}\right].$$

Hence if A_i is a focal element of m, then $x^* \in A_i$ and hence $\mathrm{Pl}(A_i) = 1$.
Thus we have shown the following theorem.

Theorem 7. *For every consonant believe structure $Em = 0$.*

Since consonant belief structures are isomorphic to possibility distributions and normalized fuzzy subsets, the concept of Shannon like entropy proves to be a meaningless or empty concept in a theory dealing with only normal fuzzy sets.

While it would be nice if only consonant belief structures had zero entropy this is not the case as seen from the following example [10].

Example 1. $X = \{x_1, x_2, x_3\}$

Let

$$A = \{x_1, x_2\} \quad B = \{x_2, x_3\}$$

Assume

$$m(A) = 1/2 \quad m\left(B\right) = 1/2$$

Since neither $A \subset B$ nor $B \subset A$, this is not a consonant belief structure. Our definition for entropy implies for this situation

$$Em = -\left[m\left(A\right)\cdot\ln\ \mathrm{Pl}\left(A\right) + m\left(B\right)\cdot\ln\ \mathrm{Pl}\left(B\right)\right].$$

But

$$\mathrm{Pl}\left(A\right) = \sum_{\substack{D \\ D\cap A\neq\varnothing}} m\left(D\right) = m\left(A\right) + m\left(B\right) = 1$$

and

$$\mathrm{Pl}\left(B\right) = \sum_{\substack{D \\ \cap B\neq\varnothing}} m\left(A\right) = m\left(A\right) + m\left(B\right) = 1$$

Hence $Em = 0$.

Actually the class of zero entropic belief structures can be classified as follows.

From our definition of Em, in order that $Em = 0$, any A where $m(A) \neq 0$ requires that $\ln \mathrm{Pl}(A) = 0$, which requires $\mathrm{Pl}(A) = 1$. Since

$$\mathrm{Pl}\,(A) = \sum_{\substack{B \\ B \cap A \neq \varnothing}} m\,(B)$$

this means that every pair of focal elements must have at least one element in common. Thus we have proved the following.

Theorem 8. *A belief structure has zero entropy if* $A_i \cap A_j \neq \varnothing$ *for each pair of focal elements.*

Thus we can see that this measure of entropy is related in some way to the disjointedness of the sets containing the evidence mass. We note that disjointedness in the focal elements is related to the discordance in the evidence.

We further note that Bayesian structures, while not the only ones, are prototypical examples of disjoint belief structures.

We now turn to belief structures which produce maximal type values for the entropy.

Theorem 9. *Em is finite.*

Proof. From our definition of Em and the fact that for non focal elements $m(A) = 0$, we get

$$Em = -\sum_{A_i} m\,(A_i) \cdot \ln\ \mathrm{Pl}\,(A_i)\,,$$

where A_i are the focal elements.

Since there are at most a finite number of focal elements, $Em = \infty$ iff $\ln \mathrm{Pl}(A_i) = -\infty$, for some i, hence $\mathrm{Pl}(A_i) = 0$ for some i. However, since $A_i \cap A_i \neq \varnothing \cdot m(A_i) > 0$ implies that $\mathrm{Pl}(A_i) > 0$.

Theorem 10. *Assume we have k focal elements with the values* $m(A_i) = a_i$. *Then Em is maximal if the focal sets* A_i *are disjoint, i.e., if* $A_i \cap A_j = \varnothing$ *for all* $i \neq j$.

Proof.

$$Em = -\sum_{i=1}^{K} m\,(A_i) \cdot \mathrm{Pl}\,(A_i)$$

$$\mathrm{Pl}\,(A_i) = \sum_{\substack{j \\ \text{for } A_i \cap A_j \neq \varnothing}} m\,(A_j) = m\,(A_i)$$

$$+ \sum_{\substack{A_j \\ \text{for } A_i \cap A_j \neq \varnothing \\ i = j}} m(A_j) = a_i + d_i$$

$$Em = -\sum_{i=1}^{K} a_i \ln(a_i + d_i).$$

As d_i increases $\ln(a_i + d_i)$ increases and $-\sum_{i=1}^{K} a_i \ln(a_i + d_i)$ decreases hence Em is maximal when $d_i = 0$ for all i. This occurs when all the A_j are disjoint.

Theorem 11. *Assume we have k disjoint focal elements. Then Em is maximal if $m(A_i) = 1/K$ for all elements and in this case*

$$E_m = -\sum_{i=1}^{K} \frac{1}{K} \ln \frac{1}{K} = \ln\ K$$

Proof.

$$Em = -\sum_{i=1}^{K} a_i \ln a_i,$$

where

$$\sum_{i=1}^{n} a_i = 1, a_1 \geqq 0$$

A proof that this well known situation produces a maximal Em when $a_i = 1/k$ can be found in Ref. 7.

Assume that we have a belief structure defined over the set X with cardinality N. The maximal number of disjoint subsets of X consist of the N disjoint sets of singletons and this has a value of $\ln N$ when the belief mass is equally divided. It appears that this situation induces the largest entropy for a situation where the cardinality of X is N. We say “it appears” since to be certain that this is so, we must prove that there is no non-disjoint collection of RN focal elements which have more entropy than the best situation with N disjoint focal elements. We are not ready at this time to prove this theorem.

5 Specificity Like Measure

Yager[8, 9] has introduced a measure of specificity associated with a possibility distribution.

If $\Pi : X \to [0,1]$ is a possibility distribution over the finite set X, then Yager[8, 9] has defined the measure of specificity associated with Π as

$$S(\Pi) = \int_{0}^{\alpha_{\max}} \frac{1}{\text{card } \Pi_\alpha} d\alpha.$$

$\Pi_\alpha = \{x|\Pi(x) \geqq \alpha, x \in X\}$ is a crisp set called the α level set of Π, card Π_α is the number of elements in Π_α and $\alpha_{\max} = \mathrm{Max}_{x\in X}\Pi(x)$.

Yager[8, 9] has shown $S(\Pi)$ to have the following properties:

1) $0 \leqq S(\Pi) \leqq 1$.
2) $S(\Pi) = 1$ iff there exists one and only one $x \in X$ such that $\Pi(x) = 1$ and $\Pi(y) = 0$ for all $y \neq x$.
3) if Π and Π^* are such that $\mathrm{Max}_{x\in X}\Pi(X) = 1$ and $\Pi(x) \leqq \Pi^*(x)$ for all $x \in X$, when
$$S(\Pi) \geqq S(\Pi^*).$$

This measure is an indication of the specificity of a possibility distribution in the sense that it indicates the degree to which Π points to one and only one element as its manifestation.

Example 2. Let $X = \{a, b, c, d\}$ and let

$$\Pi(a) = 1$$
$$\Pi(b) = 0.7$$
$$\Pi(c) = 0.5$$
$$\Pi(d) = 0.2$$

$$\begin{array}{lll} 0 \leqq \alpha \leqq 0.2 & \Pi_\alpha = \{a, b, c, d\} & \text{card } \pi_\alpha = 4 \\ 0.2 < \alpha \leqq 0.5 & \Pi_\alpha = \{a, b, c\} & \text{card } \pi_\alpha = 3 \\ 0.5 < \alpha \leqq 0.7 & \Pi_\alpha = \{a, b\} & \text{card } \pi_\alpha = 2 \\ 0.7 < \alpha \leqq 1 & \Pi_\alpha = \{a\} & \text{card } \pi_\alpha = 1 \end{array}$$

$$S(\Pi) = \int_0^1 \frac{1}{\text{card } \Pi_\alpha} d\alpha$$

$$S(\Pi) = \int_0^{0.2} \frac{1}{4} d\alpha + \int_{0.2}^{0.5} \frac{1}{3} d\alpha + \int_{0.5}^{0.7} \frac{1}{2} d\alpha + \int_{0.7}^{1} d\alpha$$

$$S(\Pi) = (0.2)\frac{1}{4} + (0.3)\frac{1}{3} + (0.2)\,\frac{1}{2} + 0.3\,(1) = 0.55.$$

We now generalize this measure from possibilistic belief structures to any belief structure.

Definition 10. *Assume m is a belief structure defined over the set X the generalized specificity measure, denoted Sm, is defined as*

$$S_m = \sum_{\substack{A \subset X \\ A \neq \varnothing}} \frac{m(A)}{n_A}.$$

n_A is the number of elements in the set A, i.e., $n_A = \text{Card } A = |A|$.

First we show that this generalized measure reduces to the particular measure suggested by Yager for possibility distributions, i.e., for consonant belief structures.

Assume that X has n elements with membership grades

$$a_n \leqq a_{n-1} \leqq a_{n-2} \ldots \leqq a_1 = 1$$

Then

$$S(\Pi) = \int_0^{a_n} \frac{1}{n} d\alpha + \int_{a_n}^{a_{n-1}} \frac{1}{n-1} d\alpha + \int_{a_{n-1}}^{a_{n-2}} \frac{1}{n-2} d\alpha + \ldots \int_{a_2}^{a_1=1} \Big| \, 1 d\alpha$$

hence

$$S(\Pi) = \frac{1}{n} a_n + \frac{1}{n-1}(a_{n-1} - a_n) + \frac{1}{n-2}(a_n - a_{n-1}) + \ldots (a_1 - a_2)$$

More generally

$$S(\Pi) = \sum_{i=1}^{n} \frac{1}{i}(a_i - a_{i+1}),$$

with $a_{n+1} = 0$ by definition.

Now assume that m is a consonant belief structure.

As Prade [3] has shown, if m is consonant, then there exists a nested family of subsets $A_i \subset X$ such that card $A_i = i$ and $\sum_{i=1}^{n} m(A_i) = 1$, where n is the cardinality of X.

Thus

$$S_m = \sum_{\substack{A \subset X \\ A \neq \varnothing}} \frac{m(A)}{n_A} = \sum_{i=1}^{n} \frac{m(A_i)}{i}$$

Furthermore, it was shown by Prade [3] that if $a_n \leqq a_{n-1} \leqq \cdots \leqq a_1$ are the plausibilities of the singletons, the possibilities of the individual elements, then $m(A_i) = a_i - a_{i+1}$. Thus

$$S_m = \sum_{i=1}^{n} \frac{a_i - a_{i+1}}{i}$$

in the consonant case. This shows that our generalized definition captures the original case.

Theorem 12. *Assume that m is a belief structure over X, where the cardinality of X is n. Then*

$$\frac{1}{n} \leqq S_m \leqq 1.$$

Proof. (1) For any A, $n_A \leqq n$, hence

$$S_m \geqq \frac{1}{n}\sum_A m(A)$$

and since $\sum m(A) = 1$, then $S_m \geqq 1/n$.

2) For any $A \neq \varnothing$, $n_A \geqq 1$, hence

$$S_m \leqq \sum_{A \subset X} m\,(A) \leqq 1$$

Let us look at the situations which attain these extremal values for S_m

Theorem 13. *S_m assumes its minimal value for a given X iff m is a vacuous belief structure. This minimal value is 1/n where n is the cardinality of X.*

Proof. (1) If m is vacuous $m(X) = 1$ hence $S_m = 1/n$

2) If m is not vacuous then there exists some A, such that $m(A) > 0$ and $n_A < n$ hence

$$S_m \geqq \frac{1}{n}.$$

Theorem 14. *S_m assumes its maximal value of 1 iff m is a Bayesian belief structure.*

Proof. (1) Assume that m is Bayesian. Then the sets having $m(A) > 0$ are only the singletons. Thus

$$S_m = \sum_{i=1}^{n} m[\{x_i\}] = 1$$

2) Assume that m is not Bayesian. Then there exists some A such that $m(A) > 0$ and $n_A > 1$ hence $S_m < 1$.

Thus whereas the *entropy* measure is *minimized* for *consonant* belief structures the *specificity* is *maximized* for *Bayesian* belief structures.

To get further insight into this measure we consider its evaluation on simple support structures.

Theorem 15. *Assume that m is a simple support structure focused at B, with $m(B) = b$. Then*

$$S_m = \frac{b}{n_B} + \frac{1-b}{n}.$$

Proof. For a simple support structure

$$m(B) = b$$
$$m(X) = 1 - b$$
$$S_m = \frac{b}{n_B} + \frac{1-b}{n}.$$

If b increases S_m increases. Furthermore as n_B decreases, without becoming vacuous, S_m increases.

Let us now examine the workings of this measure on consonant belief structures.

Theorem 16. *Assume that m_1 and m_2 are consonant belief structures generating plausibility measures Pl_1 and Pl_2 such that, for each $x \in X$,*

$$\mathrm{Pl}_1(x) \leqq \mathrm{Pl}_2(x)$$

Then.

$$S_{m_1} \geqq S_{m_2}.$$

Proof. For consonant belief structures

$$S_m = \int_0^1 \frac{1}{\mathrm{Card}\ \Pi_\alpha} d\alpha$$

Since $\mathrm{Pl}_2(x) > \mathrm{Pl}_1(x)$, card $\prod_{2_\alpha} \geqq$ card $\prod_{1_\alpha}$.

As a special case of this situation consider two consonant belief structures m_1 and m_2 defined over the same nested sets $A_1 \subseteq A_2 \subset \cdots \subset A_n$ where

$$\mathrm{Pl}_1(x) \leqq \mathrm{Pl}_2(x)$$
$$m_2(A_n) = \mathrm{Pl}_2(x_n) > \mathrm{Pl}_1(x_n) = m_1(A_n)$$

so $m_2(A_n) \geqq m_1(A_n)$.

In the same manner for all $j > 1$,

$$m_2(A_n) + m_2(A_{n-1}) + \ldots m_2(A_j) > m_1(A_j) + \ldots m_1(A_n).$$

But

$$\sum_{i=1}^{n} m_2(A_i) = \sum_{i=1}^{n} m_1(A_i),$$

hence $m_1(A_1) \geqq m_2(A_1)$ and $m_2(A_n) \geqq m_1(A_n)$.

Thus the higher the specificity the more of the evidence mass lies in the one element set and the less in the set X.

The meaning of the measure S_m appears to relate to the degree to which the evidence is pointing to a one element realization. When one considers that the total amount of plausibility assigned to the elements in X is

$$\sum_{x_i \in X} \mathrm{Pl}(X_i) = \sum_{A \subset X} n_A \cdot m(A)$$

it appears that S_m is a measure of the reduction of excess plausibility. We can also see that as the total plausibility value, which is always greater than the belief, gets closer to the belief value than S_m increases. Hence S_m appears inversely related to excess of plausibility over belief. In bringing the plausibility in a structure closer to the belief ascertained in the structure we are getting more specific in our allocation of evidence. This interpretation is reinforced by the fact that for Bayesian structures in which the plausibility always equals the belief, the value of S_m is maximum.

Since obtaining evidence involves a process of reducing possibilities, specificity thus seems to be measuring the effect of the evidence in that direction.

6 Using Both Measures

We feel that the two measures developed herein provide a complementary approach to measuring the certainty with which a belief structure is pointing to a unique outcome.

As noted, the entropy measure provides a measure of the dissonance of the evidence. This is illustrated by the fact that consonant belief structures have lowest entropic measures, while the highly dissonant type of Bayesian structures have high entropic measures.

The specificity measure provides an indication of the dispersion of the belief. We note that in this situation the Bayesian structure gets the highest grades, while the vacuous case gets the lowest.

As we noted earlier the only structure that is Bayesian, specific and consonant is the structure which $m(x) = 1$ for some $x \in X$. However this structure corresponds to the certain situation where the evidence points precisely to x as the special element.

Thus we see the following: the lower the E_m, the more consistent the evidence; and the higher S_m, the less diverse. Ideally we want low E_m and high S_m for certainty. Thus by using a combination of the two measures we feel that we can have a good indication of the quality of a belief structure with respect to suggesting one element as the outcome.

In particular the measure E_m indicates the success of the structure in reducing plausibilities, which is a desired quality in a belief structure up to a point. This point will be that where the reduction is so great that everything appears not possible, which implies an inconsistency in the evidence. The entropy measure thus indicates the success of the belief structure in being consistent. On the other hand, consistency is also desirable up to a point, this being where we leave everything as possible in order to obtain this consistency. The success with which we are able to satisfy both these criteria therefore provides a good procedure for judging the quality of evidence.

We here suggest as a measure of quality of a belief structure the two tuple (S_m, E_m). As we have noted, the ideal situation, certain knowledge, occurs only when $(S_m, E_m) = (1, 0)$. The closer a belief structure is to this point, the better quality of evidence it is supplying.

7 Conclusion

We have extended Shafer's theory of evidence to include a measure of entropy and specificity to be associated with a belief structure. These measures taken together provide an indication of the quality of the evidence supplied by a belief structure.

8 Acknowledgement

I would like to acknowledge my debt to Didier Dubois whose personal tutorial opened my eyes to Shafer's theory.

References

1. G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.
2. L. A. Zadeh, "On the validity of Dempster's rule of combination of evidence." Memo No. ERL M79/24, U. of California, Berkeley, 1979.
3. H. Prade, "On the link between Dempster's rule of combination of evidence and fuzzy set intersection." *Busefal*, **8**, 1981, pp. 60–64.
4. P. Smets, "Medical diagnosis: fuzzy sets and degrees of belief." *Fuzzy Sets and Systems*, **5**, 1981, pp. 259–266.
5. H. T. Nguyen, "On random sets and belief structures." *J. Math. Anal. and Appl.*, **65**, 1978, pp. 531–542.
6. L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems*, **1**, 1978, pp. 3–28.
7. A. I. Khinchin, *Mathematical Foundations of Information Theory*. Dover Publications, New York, 1957.
8. R. R. Yager, "Measuring tranquility and anxiety in decision making: An application of fuzzy sets." *Int. J. of General Systems*, **8**, 1982, pp. 139–146.
9. R. R. Yager, "Measurement of properties of fuzzy sets and possibility distributions." *Proc. Third International Seminar on Fuzzy Sets*, Linz, 1981, pp. 211–222.
10. M. Higashi and G. J. Klir, "Measures of uncertainty and information based on possibility distributions." *Int. J. of General Systems*, **9**, No. 1, 1982, pp. 43–58.