# Comparison of Specificity and Information for Fuzzy Domains

Arthur Ramer
*Computer Science, University of New South Wales*
*PO Box 1, Kensington, NSW 2033, Australia*

May 1, 1992

## 1 Overview

Given a *universe of discourse* $X$—a domain of possible outcomes—an experiment may consist of selecting one of its elements, subject to operation of chance, or of observing the elements, subject to imprecision.

A priori uncertainty about the actual result of the experiment may be quantified, representing either the likelihood of the choice of $x \in X$ or the degree to which any such $x \in X$ would be suitable as a description of the outcome. The former case corresponds to probability distribution, while the latter gives a possibility assignment on $X$.

Study of such assignments and thier properties comes under the purview of *possibility theory* [1]. It, like probability theory, assigns values in between 0 and 1 to express likelihoods of outcomes. Here, however, similarity ends. Possibility theory uses *maximum* and *minimum* functions to combine uncertainty, where probability theory uses plus and times operations. This leads to a very dissimilar theory in its analytical framework, even though they share several semantic concepts.

One of them consists of expressing quantitatively the *uncertainty* associated with a given distribution [2, 3]. Its value corresponds to the gain of *information* that would result from conducting an experiment and ascertaining its actual result. This gain becomes simutaneously a decrease in *uncertainty* about the outcome of an experiment.

The other concept we consider in depth is one of *specificity.* Although it has been introduced previously in a few different forms, a closer analysis shows that they share main epistemic features. We follow here the presentation of Ramer and Yager [10].

Fuzzy set $(X, \mathbf{p})$ can be considered as a form of a likelihood function, with the elements of $X$ where $\mathbf{p}$ reaches its maximum playing privileged role. When selecting $x : \mathbf{p}(x_0) = \max \mathbf{p}(x)$ is important to ask how definite has been such decision, and whether another element would offer a close choice.

In this interpretation, *specificity* becomes an attribute of the complete set of possibilities, the attribute assuming either numeric or linguistic values. Here we develop a comprehensive model of such specificity, expressed as a *numerical function* of a possibility assignment.

## 2 Introduction

This paper demonstrates how an integrated theory can be built on the foundation of possibility theory. Information and uncertainty were cosidered in 'fuzzy' literature since 1982. Our departing point is the model proposed by Klir [4, 5] for the discrete case. It was elaborated axiomatically by Ramer [9], who also introduced the continuous model [7].

Specificity as a numerical function was considered mostly within Dempster-Shafer evidence theory. An explicit definition was given first by Yager [11], who has also introduced it in the context of possibility theory [12]. Axiomatic approach and the continuous model have been developed very recently by Ramer and Yager [10]. They also establish a close analytical correspondence between specificity and information.

In literature to date, specificity and uncertainty are defined only for the discrete finite domains, with a sole exception of [10]. Our presentation removes these limitations. We define specificity measures for *arbitrary measurable* domains. When discrete, they can be finite or infinite or, in general have $\mu(X) < \infty$ or $\mu(X) = \infty$. prespecified pattern. By abuse of the language we refer to this model as a *continuous* one.

We adopt the convention of avoiding, whenever possible, subscripts and indices. We do not specify explicitly basis of logarithms, as its change would simply amount to a multiplying all expressions by the same constant. Following tradition, binary logarithms—$log_2$—are assumed for the discrete distributions, and natural—$ln$— for the continuous cases. We use $(\tilde{p})$ for the decreasing rearrangement of the sequence $(p_i)$. For finite sequences, rearrangements are permutations of their elements. For infinite sequences and functions we construct rearrangements using *cuts*. To define $\tilde{f}$, given $f$ on $X$, we want all their $\alpha$-cuts to be of the same measure. We put

$$P(y) = \mu(\{x : f(x) \geq y\}),$$

$$\tilde{f}(x) = P^{-1}(x).$$

Now for the discrete rearrangements we associate with the sequence $(p) = (p_1, \ldots, p_n, \ldots)$ a step function $f : x \mapsto p_{\lceil x \rceil}$, where $\lceil x \rceil$ denotes the greatest integer no less than $x$. Then the descending rearrangement $\tilde{f}$ corresponds to $(\tilde{p})$.

## 3 Information and uncertainty

We use the model of possibility theory introduced by Zadeh [13]. We view mapping $\mathbf{p}$ as assigning a degree of assurance or certainty that an element of $X$ is the outcome of an experiment. A priori we know only the distribution $\mathbf{p}$; to determine $x \in X$ means to remove uncertainty about the result, thus entailing a gain of information. We would be particularly interested in quantifying that gain of information, which would also express the uncertainty inherent in the complete distribution $\mathbf{p}$.

Following established principles of information theory [3], we stipulate that such information function satisfies certain standard properties. For $\mathbf{p}_1$ on $X$ and $\mathbf{p}_2$ on $Y$ we define a *noninterracting, joint* distribution $\mathbf{p}_1 \otimes \mathbf{p}_2$ on $X \times Y$ as

$$\mathbf{p} \otimes \mathbf{p}_2 : (x, y) \mapsto \min(\mathbf{p}_1(x), \mathbf{p}_2(y)).$$

If $\mathbf{p}$ was already defined on a product domain $X \times Y$, we construct its *projections* (marginal distributions) using *maximum* operation

$$\mathbf{p}' : x \mapsto \max_y \mathbf{p}(x, y), \quad \mathbf{p}'' : y \mapsto \max_x \mathbf{p}(x, y).$$

There is often a need to consider a given assignment $\mathbf{p}$ as defined on on a larger domain, without, however, making any essential change to the possibility values it represents. We do so by defining $\mathbf{p}^Y$ for $Y \supset X$, as agreeing with $\mathbf{p}$ on the elements of $X$, and 0 otherwise. Lastly, the elements of the domain of discourse could be permuted; if $s : X \to X$ is one-to-one, we define

$$s(\mathbf{p}) : x \mapsto \mathbf{p}(s(x)).$$

We now postulate [5]

$$\begin{array}{lrl} \textit{additivity} & I(\mathbf{p}_1 \otimes \mathbf{p}_2) = & I(\mathbf{p}_1) + I(\mathbf{p}_2) \\ \textit{subadditivity} & I(\mathbf{p}) \leq & I(\mathbf{p}') + I(\mathbf{p}'') \\ \textit{symmetry} & I(s(\mathbf{p})) = & I(\mathbf{p}) \\ \textit{expansibility} & I(\mathbf{p}^Y) = & I(\mathbf{p}) \end{array}$$

It turns out that these properties essentially characterize the admissible information functions [6, 9]. Subject to the normalization of parameters, for the discrete case of $X = \{x_1, \ldots, x_n\}$

$$U(\mathbf{p}) = \sum (\tilde{p}_i - \tilde{p}_{i+1}) \log i$$

which can be also written using *finite differences* notation

$$U(\mathbf{p}) = \sum \tilde{p}_i \nabla \log i.$$

We observe that the distribution which carries the highest uncertainty value consists of assigning possibility 1 to all the events in $X$. It states that, a priori, every event is fully possible. This distribution, carrying no prior information, can be considered the *most uninformed* one.

We shall now extend previous definitions to arbitrary measurable domains [7]. To avoid technical complications, we consider only a typical case of the unit interval.

As a first step, the discrete formula $U(x) = \sum p_i \nabla \log i$ suggests forming $\int_0^1 \tilde{f}(x) d \ln x, = \int_0^1 \frac{\tilde{f}(x)}{x} dx$ as a candidate expression for the value of information. Unfortunately, $\tilde{f}(x)$ is equal to 1 at 0, and the integral above diverges. A solution can be found through a technique (used also in probability) of information distance between a given distribution and the most 'uninformed' one—where U-uncertainty attains its maximum. Our final formula becomes

$$I(f) = \int_0^1 \frac{1 - \tilde{f}(x)}{x} dx.$$

This integral is well defined and avoids the annoying singularity at 0. It can be used for a very wide class of functions, including all polynomials.

# 4 Principles of specificity

The discussion will be conducted in terms of a discrete countable distribution $(p_i)$, with finite distributions viewed as the initial segments. Our objective is to capture formally the informal intuition about specificity. The main premise is the principle of *juxtaposition:*

> $Sp(\mathbf{p})$ expresses the preference for a certain maximal $p_0$ over any and all the remaining $p_i$.

Now let us consider how, having selected $p_0 = \max(p)$, its *informal specificity* is estimated. We look first for the next largest $p_i$ and estimate how its presence diminishes the specificity. The process is then iterated in the order of decreasing values of $p_i$, every next value lowering the estimated specificity. We can picture it as a sequential process, its input the decreasing rearrangement ($\tilde{p}_i$). We may also surmise that, for a given $i$, the drop in specificity caused by $\tilde{p}_i$ will not depend on the earlier inputs $\tilde{p}_1, \ldots, \tilde{p}_{i-1}$. This assumption of *independent influence* is consistent with the *juxtaposition* interpretation of specificity.

Let us consider the effect of a uniform modification of $(p)$. For a scaling $\alpha\mathbf{p} = (\alpha p_1, \ldots, \alpha p_n, \ldots), 0 \leq \alpha \leq 1$, we may assume that the relative specificities remain unchanged, while with a shift of values $\mathbf{p} - \beta = (p_1 - \beta, \ldots, p_n - \beta, \ldots)$ no change should occur.

Last item considered will be the effect of offering yet another choice, identical in value to several choices already provided. The common perception of *specificity* is that the change due to such $n$-th choice will be ever less as $n$ increases—a *diminishing return*. For its *relative* effect, we can postulate taking away the same proportion of the specificity still available. After all, we consider yet another *identical* choice; only we consider it at stage $n$ and not sooner.

We can extract an analytical representation from the rules elaborated above. The result is a linear formula

$$Sp(\mathbf{p}) = \tilde{p}_1 - \sum_{i \geq 2} w_i \tilde{p}_i$$

with $\sum_{i \geq 2} w_i = 1$. ¿From here we can conclude that $\lim_{i \to \infty} w_i = 0$, and $1 > w_2 > w_3 > \cdots$, in agreement with the 'diminishing returns'.

We shall consider the linear form of $Sp(\mathbf{p})$ as *general* specificity function. It is general enough to fit most applications and, if $w_i$ are supplied, it offers a comparison scale among the distributions.

Coefficients $w_i$ can be established precisely if we assume the rule of constant influence of equal choices. After more calculations

$$Sp(\mathbf{p}) = \tilde{p}_1 - \sum_{i \geq 2} (\omega^{i-1} - \omega^i) \tilde{p}_i$$

for some $\omega$, $0 < \omega < 1$, producing a *definite* form of specificity. Choosing $\omega = \frac{1}{2}$ (in spirit of binary logarithms) gives $Sp(\mathbf{p}) = \tilde{p}_1 - \sum \frac{\tilde{p}_i}{2^{i-1}}$. In the above formulas the role of $\tilde{p}_1$ is manifestly different from that of $\tilde{p}_i, i \geq 2$. A more symmetric expression can be obtained defining $W_i = 1 - w_2 - \ldots - w_i$, resulting in a *general expression*

$$Sp(\mathbf{p}) = \sum_{i \geq 1} W_i (\tilde{p}_i - \tilde{p}_{i+1})$$

and the *definite one*

$$Sp(\mathbf{p}) = \sum_{i \geq 1} \omega^{i-1} (\tilde{p}_i - \tilde{p}_{i-1}).$$