Possibility Theory and Soft Data Analysis

In PRUF, the translation of \( q \) is expressed as the translation of the answer to \( q \), with a symbol of the form \( \alpha \) identifying the variable whose value is to be determined. As an illustration, for the example under discussion the proposition to be inferred from \( D \) may be expressed as

\[
p = \text{The average age of individuals in POPULATION is } \alpha f
\]

where \( f \) is a function of the entries in \( D \), say \( X_1, \ldots, X_m \).

Thus, to answer \( q \) we must compute the value of \( f(X_1, \ldots, X_m) \) from whatever information is available about \( D \).

To link the method under discussion to our earlier formulation of the problem of inference, we shall assume that the available information about \( D \) consists of the evidence

\[ E = \{ p_1, \ldots, p_n \} \]

in which the \( p_i \) are fuzzy propositions.

Our definition of translation in Section 3 implies that each of the \( p_i \) in \( E \) induces a possibility distribution over \( D \). Thus, letting \( \pi_1(X_1, \ldots, X_m) \) denote the possibility of \( D \) given \( p_1 \), we can assert that the possibility of \( D \) given \( p_1, \ldots, p_n \) is given by the conjunction

\[
\pi(D) = \pi_1(X_1, \ldots, X_m) \land \cdots \land \pi_n(X_1, \ldots, X_m)
\]

(4.18)

Thus, \( \pi(D) \), as expressed by (4.18), may be viewed as an elastic constraint on \( D \) which is induced by the evidence \( E = \{ p_1, \ldots, p_n \} \).

From the knowledge of \( \pi(D) \) we can infer the possibility distribution of the function

\[
z = f(X_1, \ldots, X_m)
\]

(4.19)

by invoking the extension principle, as shown in Section 3.

In this way, the determination of the possibility distribution of \( f \) reduces, in principle, to the solution of the following variational problem in mathematical programming.

\[
\mu(z) = \max_{X_1, \ldots, X_m} \pi_1(X_1, \ldots, X_m) \land \cdots \land \pi_n(X_1, \ldots, X_m)
\]

(4.20)
subject to
\[ z = f(x_1, \ldots, x_n) \, . \]

In terms of \( \mu(z) \), the possibility distribution of \( f \) may be expressed in the form
\[
\Pi_f = \int_{V} \mu(z)/z \tag{4.21}
\]
where \( V \) is the range of \( z \). An example illustrating the application of this technique will be discussed in Section 5.

As a further example, consider the proposition which occurs in Example (e), Section 1, namely:

\[ p \triangleq \text{Brian is much taller than most of his close friends} \]

For the purpose of representing the meaning of \( p \), it is expedient to assume that \( D \) is comprised of the relations

\[
\begin{array}{c|c|c}
\text{POPULATION} & \text{Name} & \text{Height} \\
\hline
\text{FRIENDS} & \text{Name1} & \text{Name2} & \mu \\
\hline
\text{MUCH TALLER} & \text{Height1} & \text{Height2} & \mu \\
\text{MOST} & p & \mu \\
\end{array}
\]

In the relation FRIENDS, \( \mu \) represents the degree to which an individual whose name is Name2 is a friend of Name1. Similarly, in the relation MUCH TALLER, \( \mu \) represents the degree to which an individual whose height is HEIGHT1 is much taller than one whose height is HEIGHT2. In MOST, \( \mu \) represents the degree to which a proportion, \( p \), fits the definition of MOST as a fuzzy subset of the unit interval.

To represent the meaning of \( p \) we shall translate \( p \)--in the spirit of (c) (Section 3)--into a procedure which computes the truth-value of \( p \) relative to a given \( D \). The procedure--as described below--may be viewed as a sequence of computations which, in combination, yield the truth-value of \( p \).

1. Obtain Brian's height from POPULATION. Thus,

\[ \text{Height}(\text{Brian}) = \text{Height}\ POPULATION[\text{Name} = \text{Brian}] \]
2. Determine the fuzzy set, MT, of individuals in POPULATION in relation to whom Brian is much taller.

Let \( \text{Name}_i \) be the name of the \( i \)th individual in POPULATION. The height of \( \text{Name}_i \) is given by

\[
\text{Height}(\text{Name}_i) = \text{Height}(\text{POPULATION}[\text{Name} = \text{Name}_i])
\]

Now the degree to which Brian is much taller than \( \text{Name}_i \) is given by

\[
\delta_i = \mu_{\text{MUCH TALLER}}(\text{Height}(\text{Brian}), \text{Height}(\text{Name}_i))
\]

and hence MT may be expressed as

\[
\text{MT} = \bigcup_i \delta_i/\text{Name}_i, \quad \text{Name}_i \in \text{POPULATION}
\]

where POPULATION is the list of names of individuals in POPULATION, \( \delta_i \) is the grade of membership of \( \text{Name}_i \) in MT, and \( \bigcup_i \) is the union of singletons \( \delta_i/\text{Name}_i \) (see footnote 3).

3. Determine the fuzzy set, CF, of individuals in POPULATION who are close friends of Brian.

To form the relation CLOSE FRIENDS from FRIENDS we intensify FRIENDS by squaring it (i.e., by replacing \( \mu \) with \( \mu^2 \)). Then, the fuzzy set of close friends of Brian is given by

\[
\text{CF} = \mu_{\times \text{Name}_2}^{\text{FRIENDS}^2}[\text{Name}_1 = \text{Brian}]
\]

4. Form the count of elements of CF:

\[
\text{Count(CF)} = \sum_i \mu_{\text{CF}}(\text{Name}_i)
\]

where \( \mu_{\text{CF}}(\text{Name}_i) \) is the grade of membership of \( \text{Name}_i \) in CF and \( \sum_i \) is the arithmetic sum. More explicitly

\[
\text{Count(F)} = \sum_i \mu_{\text{FRIENDS}}^2(\text{Brian, Name}_i)
\]

5. Form the intersection of CF and MT, that is, the fuzzy set of those close friends of Brian in relation to whom he is much taller.
$$H \triangleq CF \cap MT$$

6. Form the count of elements of $H$.

$$\text{Count}(H) = \sum_i \mu_H(\text{Name}_i)$$

where $\mu_H(\text{Name}_i)$ is the grade of membership of $\text{Name}_i$ in $H$ and $\sum_i$ is the arithmetic sum.

7. Form the ratio

$$r = \frac{\text{Count}(MT \cap CF)}{\text{Count}(CF)}$$

which represents the proportion of close friends of Brian in relation to whom he is much taller.

8. Compute the grade of membership of $r$ in MOST

$$\tau = \mu_{\text{MOST}}[\rho=r]$$

The value of $\tau$ is the desired truth-value of $p$ with respect to $D$ and, equivalently, the possibility of $D$ given $p$.

In terms of the membership functions of FRIENDS, MUCH TALLER and MOST, the value of $\tau$ is given explicitly by the expression

$$\tau = \mu_{\text{MOST}} \left( \sum_i \mu_{MT}(\text{Height}(\text{Brian}), \text{Height}(\text{Name}_i)) \land \mu_{CF}(\text{Brian, Name}_i) \right) / \sum_i \mu_{CF}(\text{Brian, Name}_i)$$

(4.22)

In summary, the procedure in question serves to represent the meaning of $p$ by describing the operations that must be performed on $D$ in order to compute the truth-value of $p$ with respect to $D$. Thus, viewed as an expression in PRUF, (4.22) is in effect a mathematical description of a procedure which defines $\tau$ as a function of $D$. However, as was stressed in Section 3, the meaning of $p$ is the procedure itself rather than the value of $\tau$ which it returns for a given $D$.

5. Examples of Inference from Soft Data

To illustrate the application of some of the techniques
described in the preceding sections, we shall consider
several simple examples, including Examples (a), (b), (c) and
(e) of Section 2. As is generally the case in inference from
soft data, the chains of inference in these examples are
short.

Example 1 (Example (a), Section 1).

X is a large number

Y is much larger than X

How large is Y?

Solution. On applying the compositional rule of infer-
ence (4.15), we obtain the following expression for the
possibility distribution of Y

\[ \Pi_Y = \text{LARGE} \cdot \text{MUCH LARGER} \] (5.1)

or, more explicitly,

\[ \pi_Y(v) = \sup_u (\mu_{\text{LARGE}}(u) \land \mu_{\text{MUCH LARGER}}(u,v)) \] (5.2)

where LARGE and MUCH LARGER are the fuzzy denotations of
large and much larger, respectively.

Example 2.

X is small

Y is approximately equal to X

Z is much larger than both X and Y

How large is Z?

Solution. Proceeding as in Example 1, we obtain the
following expression for the possibility distribution of Z

\[ \Pi_Z = (\text{MUCH LARGER THAN} \cdot \text{APPROXIMATELY EQUAL} \cdot \text{SMALL}) \]

\[ \cap \text{MUCH LARGER THAN} \cdot \text{SMALL} \] (5.3)

in which the intersection implies that Z is much larger than
X, and Z is much larger than Y.
Example 3 (Example (b), Section 1).

Most Frenchmen are not tall
Elie is a Frenchman

How tall is Elie?

Solution. First, we interpret the question as follows:

Most Frenchmen are not tall
Elie is a Frenchman picked at random

What is the probability that Elie is tall?

Second, we assume that the database consists of a single relation of the form

\[
\text{POPULATION} \| \text{Name} \| \mu
\]

in which \( \mu_i \) is the degree to which \( \text{Name}_i \) is tall, and \( i \) ranges from 1 to \( N \).

Now, the constraint on the database induced by the proposition

\[ P \triangleq \text{Most Frenchmen are not tall} \]

gives rise to the possibility distribution expressed by

\[
\pi_P(\text{POPULATION}) = \mu_{\text{MOST}} \left( \frac{1}{N} \sum_i (1 - \mu_i) \right)
\]  \hspace{1cm} (5.4)

in which the argument of \( \mu_{\text{MOST}} \) represents the proportion of Frenchmen who are not tall.

Furthermore, if a Frenchman is chosen at random, then the probability that he is tall is given by (see (3.40))

\[
\text{Prob[Frenchman is tall]} = \frac{1}{N} \sum_i \mu_i.
\]  \hspace{1cm} (5.5)

Thus, the proposition (in which \( \lambda \) is a linguistic probability)

\[ Q \triangleq \text{The probability that a Frenchman is tall is } \lambda \]

induces the possibility distribution

\[
\pi_Q(\text{POPULATION}) = \mu_{\lambda} \left( \frac{1}{N} \sum_i \mu_i \right).
\]  \hspace{1cm} (5.6)
To apply the entailment principle to the problem in hand, we have to find a λ such that
\[ \mu_\lambda \left( \frac{1}{n} \sum_i \mu_i \right) \geq \mu_{\text{MOST}} \left( \frac{1}{n} \sum_i \mu_i \right). \] (5.7)

Furthermore, to be as informative as possible, the λ in q should be as small as possible in the sense that there should be no λ' such that
\[ \lambda'(v) \leq \lambda(v) \] (5.8)
for all v in [0,1] and λ'(v) < λ(v) for at least some v in [0,1].

With this as our objective, we first note that (5.4) may be rewritten as
\[ \pi_p(\text{POPULATION}) = \mu_{\text{MOST}} \left( 1 - \frac{1}{n} \sum_i \mu_i \right) \] (5.9)
where \text{ANT MOST} stands for the denotation of the antonym of most, i.e.,
\[ \mu_{\text{ANT MOST}}(v) = \mu_{\text{MOST}}(1-v), \quad v \in [0,1] \] (5.10)
which signifies that the membership function of \text{ANT MOST} is the mirror image of that of \text{MOST}.

At this juncture, then, we can assert that
\[ \pi_p \triangleq \text{Most Frenchmen are not tall} \] (5.11)
\[ \rightarrow \pi_p(\text{POPULATION}) = \mu_{\text{ANT MOST}} \left( \frac{1}{n} \sum_i \mu_i \right) \]
while
\[ r \triangleq \text{Prob\{Frenchman is tall\} is } \gamma \] (5.12)
\[ \rightarrow \pi_r(\text{POPULATION}) = \mu_\gamma \left( \frac{1}{n} \sum_i \mu_i \right) \]
where \( \gamma \) is a linguistic probability.

On comparing (5.11) with (5.12), we note that if the fuzzy set \text{LIKELY} is defined to be equal to \text{MOST}, i.e.,
\[ \mu_{\text{LIKELY}}(v) = \mu_{\text{MOST}}(v), \quad v \in [0,1] \] (5.13)
so that
\[ \mu_{\text{UNLIKELY}}(v) = \mu_{\text{ANT LIKELY}}(v) = \mu_{\text{ANT MOST}}(v) \]  

then we can infer from (5.11) and (5.12) the semantic equivalence (3.48)

\[ p \doteq \text{Most Frenchmen are not very tall} \iff r \doteq \text{Prob}\{\text{Frenchman is tall}\} \text{ is unlikely} \]

Consequently, as the answer to the posed question, we have

Most Frenchmen are not tall
Elie is a Frenchman

\[ \text{It is unlikely that Elie is tall} \]

In essence, then, what we have shown is that, under the assumption that the fuzzy sets MOST and LIKELY are equal, we can infer from the premise

\[ p \doteq \text{Most Frenchmen are not tall} \]

the semantically equivalent proposition

\[ r \doteq \text{It is unlikely that a Frenchman picked at random is tall} \]

from which it follows that "It is unlikely that Elie is tall."

Example 4.

Most Swedes are tall

\[ \text{How many Swedes are very tall?} \]

Solution. Suppose that the answer is of the form

\[ r \doteq Q \text{ Swedes are very tall} \]

where \( Q \) is a fuzzy quantifier. Then, proceeding as in Example 3, we have

\[ p \doteq \text{Most Swedes are tall} \to \Pi_p(\text{POPULATION}) = \mu_{\text{MOST}} \left( \frac{1}{N} \sum_i \mu_i \right) \]  

and

\[ (5.15) \]
\[ r \triangleq Q \text{ Swedes are very tall} \implies \pi_r(\text{POPULATION}) = \mu_Q(\frac{1}{N} \sum_i^N i \cdot \mu_i^2) \] (5.16)

Consequently, what we have to find is the "smallest" \( Q \) such that

\[ \mu_Q(\frac{1}{N} \sum_i^N \mu_i^2) \geq \mu_{\text{MOST}}(\frac{1}{N} \sum_i^N \mu_i) . \] (5.17)

It can easily be verified that such a \( Q \) is given by

\[ Q = \overset{2}{\text{MOST}} \] (5.18)

where the "left-square" of MOST is defined by

\[ \mu_{\overset{2}{\text{MOST}}} (F) = \mu_{\text{MOST}}(\sqrt{v}) , \quad v \in [0,1] . \] (5.19)

For, from the elementary inequality

\[ \sqrt{\frac{1}{N} \sum_i^N \mu_i^2} \geq \frac{1}{N} \sum_i^N \mu_i \] (5.20)

and the monotonicity of \( \mu_{\text{MOST}} \), it follows that

\[ \mu_{\text{MOST}}(\sqrt{\frac{1}{N} \sum_i^N \mu_i^2}) \geq \mu_{\text{MOST}}\left(\frac{1}{N} \sum_i^N \mu_i\right) \] (5.21)

which, in view of (5.19), implies that

\[ \mu_{\overset{2}{\text{MOST}}} \left(\frac{1}{N} \sum_i^N \mu_i^2\right) \geq \mu_{\text{MOST}}\left(\frac{1}{N} \sum_i^N \mu_i\right) \] (5.22)

and hence that the proposition

\[ p \triangleq \text{Most Swedes are tall} \]

entails

\[ q \triangleq \overset{2}{\text{Most Swedes are very tall}} \]

Example 5.

Naomi is not very tall is true

How true is it that Naomi is tall?

Solution. Suppose that the answer to the question is expressed as a proposition \( q \):

\[ q \triangleq \text{Naomi is tall is } \top \]

\[ \text{If MOST is interpreted as a fuzzy number } [90,18,20] \text{ then } \overset{2}{\text{MOST}} \text{ may be expressed as the product of MOST with itself.} \]
where $\tau$ is a linguistic truth-value, e.g., very true, more or less true, etc.

To determine $\tau$, we set $q$ semantically equal to $p$ (see (3.49)), i.e., we assert that the possibility distributions induced by $p$ and $q$ are equal. Now, by (3.8) and (3.36), we have

$$\text{Naomi is not very tall is true} \rightarrow \Pi_{\text{Height(Naomi)}}^{F}$$

where

$$\mu_{F}(u) = \mu^{\text{TRUE}}(1 - \mu^{2}_{\text{TALL}}(u))$$

and

$$\text{Naomi is tall is } \tau \rightarrow \mu_{\tau}(\mu_{\text{TALL}}(u))$$

where $\mu_{\text{TALL}}$ and $\mu_{\text{TRUE}}$ are the membership functions of TALL and TRUE, respectively. Consequently, for all $u$ in the domain of the variable Height(Naomi), we have

$$\mu^{\text{TRUE}}(1 - \mu^{2}_{\text{TALL}}(u)) = \mu_{\tau}(\mu_{\text{TALL}}(u))$$

from which it follows that the membership function of $\tau$ is given by

$$\mu_{\tau}(v) = 1 - v^{2}, \quad v \in [0,1] .$$

Thus, if $\mu_{\text{TRUE}}$ is defined by

$$\mu_{\text{TRUE}}(v) = v^{2}$$

then

$$\mu_{\tau}(v) = 1 - \mu_{\text{TRUE}}(v)$$

and hence

$$\tau = \text{not true} .$$

On the other hand, if

$$\mu_{\text{TRUE}}(v) = v$$

then

$$\mu_{\tau}(v) = 1 - \mu_{\text{TRUE}}^{2}(v)$$

and

$$\tau = \text{not very true} .$$
Example 6  Marvin lives near MIT

Lucia lives near MIT

What is the distance between the residences of Marvin and Lucia?

Solution. Let \((X_M, Y_M)\) and \((X_L, Y_L)\) be the coordinates of the residences of Marvin and Lucia, respectively. Furthermore, let \(\Pi(X_M, Y_M)\) and \(\Pi(X_L, Y_L)\) be the possibility distributions induced by \(p\) and \(q\), that is, derived from the definition of the binary fuzzy relation NEAR.

Now, the distance between the residences of Marvin and Lucia is expressed by

\[
d = \sqrt{(X_M - X_L)^2 + (Y_M - Y_L)^2}.
\]  

Using (5.34) and applying the extension principle (2.34), the possibility distribution function of \(d\) is found to be given by

\[
\pi_d(w) = \sup_{u_1, v_1, u_2, v_2} \left( \pi(X_M, Y_M, u_1, v_1) \land \pi(X_L, Y_L, u_2, v_2) \right)
\]

subject to

\[
w = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}
\]  

(5.35)  

(5.36)

where the supremum is taken over all possible values of \(X_M\), \(Y_M\), \(X_L\) and \(Y_L\) subject to the constraint (5.36). Generally, \(\pi_d\) as defined by (5.35) will be a monotone decreasing function of \(w\), with \(\pi_d(w) = 1\) for sufficiently small values of \(w\).

Example 7  (Example (c), Section 1).

\(p_1 \triangleq \) It is unlikely that Andrea is very young

\(p_2 \triangleq \) It is likely that Andrea is young

\(p_3 \triangleq \) It is very unlikely that Andrea is old

\(q \triangleq \) How likely is it that Andrea is not old?

Solution. To find the answer to the posed question, we shall reduce the stated problem to the solution of a mathematical program, as described in Section 4.
First, each of the premises is translated into a constraint on the probability density, \( p \), of Andrea's age. Thus, using (3.8), (5.14) and (3.40), we have

\[
\begin{align*}
\pi_1(p) &= \mu_{\text{LIKELY}} (1 - \int_0^{100} \mu_{\text{YOUNG}}(u)p(u)du) \quad (5.37) \\
\pi_2(p) &= \mu_{\text{LIKELY}} \int_0^{100} \mu_{\text{YOUNG}}(u)p(u)du \quad (5.38) \\
\pi_3(p) &= \mu_{\text{LIKELY}}^2 (1 - \int_0^{100} \mu_{\text{OLD}}(u)p(u)du) \quad (5.39)
\end{align*}
\]

where \( \int_0^{100} \mu_{\text{YOUNG}}(u)p(u)du \) represents the probability of the fuzzy event "Andrea is young," with the understanding that the range of the variable \( \text{Age(Andrea)} \) is the interval \([0,100]\).

Next, we must translate the answer to the posed question, which we assume to be of the form "It is \( \lambda \) that Andrea is not old," where \( \lambda \) is a linguistic probability. Thus

\[
q \rightarrow \pi_q(p) = \mu_\lambda \left( \int_0^{100} (1 - \mu_{\text{OLD}}(u))p(u)du \right) \quad (5.40)
\]

where \( \mu_\lambda \) is the unknown membership function of \( \lambda \).

Finally, by using (4.20), the problem in question is reduced to the solution of the variational problem

\[
\mu_\lambda(\gamma) = \max_p \left\{ \left[ \mu_{\text{LIKELY}} (1 - \int_0^{100} \mu_{\text{YOUNG}}^2(u)p(u)du) \right] \\
\mu_{\text{LIKELY}} \int_0^{100} \mu_{\text{YOUNG}}(u)p(u)du \\
\mu_{\text{LIKELY}}^2 \left( 1 - \int_0^{100} \mu_{\text{OLD}}(u)p(u)du \right) \right\} \quad (5.41)
\]

subject to

\[
\gamma = \int_0^{100} (1 - \mu_{\text{OLD}}(u))p(u)du
\]

where \( \gamma \) is the numerical probability of the fuzzy event "Andrea is not old."
Example 8 (Example (e), Section 1).

Brian is much taller than most of his close friends

How tall is Brian?

Solution. Let $x$ denote Brian's height. In section 4, we have found that, relative to a given database $D$, the truth of $p$ is given by

$$
\tau = \mu_{\text{MOST}}\left(\frac{\sum_{i} \mu_{\text{MT}}(x, \text{Height}(\text{Name}_i)) \land \mu_{\text{CF}}(\text{Brian}, \text{Name}_i)}{\sum_{i} \mu_{\text{CF}}(\text{Brian}, \text{Name}_i)}\right)
$$

(5.42)

where $\mu_{\text{MT}}(x, \text{Height}(\text{Name}_i))$ is the degree to which Brian is much taller than $\text{Name}_i$ and $\mu_{\text{F}}$ is the degree to which $\text{Name}_i$ is Brian's close friend.

Now, for a given value of $x$ and a given $D$, the value of $\tau$ may be interpreted as the possibility of $x$ given $D$. Thus, the possibility distribution function of Brian's height is given by the same expression as $\tau$, and hence

$$
\text{Poss}\{\text{Height}(\text{Brian}) = x\} = \mu_{\text{MOST}}\left(\frac{\sum_{i} \mu_{\text{MT}}(x, \text{Height}(\text{Name}_i)) \land \mu_{\text{CF}}(\text{Brian}, \text{Name}_i)}{\sum_{i} \mu_{\text{CF}}(\text{Brian}, \text{Name}_i)}\right)
$$

(5.43)

Example 9. Find the consistency of the proposition

$p \equiv$ Sharon has more than a few good friends

with the database

$$
\text{GF}_{\text{Sharon}} = \text{Mary} + 0.9\text{Valya} + 0.9\text{Doris} + 0.8\text{John} + 0.7\text{Chris} + 0.6\text{Pat} + 0.5\text{Denise} + \cdots
$$

$$
\text{FEW} = 0.8/1 + 0.9/2 + 1/3 + 1/4 + 0.8/5
$$

(5.44)

$$
+ 0.5/6 + 0.2/7
$$

(5.45)

where $\text{GF}_{\text{Sharon}}$ is the fuzzy set of Sharon's good friends (arranged in order of decreasing degree of friendship) and $\text{FEW}$ is the fuzzy denotation of few.
Lotfi A. Zadeh

Solution. If FEW is defined by (5.45), then at least few is expressed by

\[ \triangleright \text{FEW} = 0.8/1 + 0.9/2 + 1/3 + 1/4 + \cdots \] (5.46)

where \( \triangleright \text{FEW} \) is the composition of the binary relation \( \triangleright \) with the unary relation FEW.

The FG cardinality of the fuzzy set \( GF_{Sharon} \) is given by

\[
\text{FGCount}(GF_{Sharon}) = 1/1 + 0.9/2 + 0.9/3 + 0.8/4 + 0.7/5 + 0.6/6 + 0.5/7 + \cdots
\] (5.47)

and hence the degree of consistency of \( p \) with the database is given by

\[
\gamma = \sup(\text{FGCount}(GF_{Sharon}) \cap \triangleright \text{FEW})
\]

\[
= \sup(0.8/1 + 0.9/2 + 0.9/3 + 0.8/4 + \cdots)
\]

\[
= 0.9
\]

6. Evidence, Certainty and Possibility

An important issue that arises in the analysis of soft data relates to the need for a way of assessing the degree of credibility of a conclusion which is inferred from a body of evidence.

For our purposes, it will be convenient to regard a body of evidence—or simply evidence, \( E \)—as a collection of fuzzy propositions, \( E = \{g_1, \ldots, g_n\} \). Furthermore, we shall assume that the evidence is granular in nature, that is, each \( g_i \), \( i = 1, \ldots, n \), is a granule of the form\(^7\)

\[ g_i \triangleq Y \text{ is } G_i \text{ is } \lambda_i \] (6.1)

and/or

\[ g_i \triangleq \text{If } X \text{ is } F_i \text{ then } Y \text{ is } G_i \] (6.2)

\(^6\)A fuller discussion of problems of this type may be found in [11].

\(^7\)A more detailed discussion of the concept of information granularity may be found in [94].
and/or

(c) \( q_i \triangleq \text{If } X \text{ is } F_i \text{ then } Y \text{ is } G_j \text{ is } \lambda_j \), \hspace{1cm} (6.3)

and/or

\( j = 1, \ldots, m \)

(d) \( q_i \triangleq X \text{ is } F_i \text{ is } p_i \) \hspace{1cm} (6.4)

where \( X \) and \( Y \) are variables taking values in \( U \) and \( V \), respectively; \( F_i, i = 1, \ldots, n \) and \( G_j, j = 1, \ldots, m \), are fuzzy subsets of \( U \) and \( V \); and \( p_i \) and \( \lambda_j \) are linguistic probabilities.

Although \( E \) may comprise a mixture of granules of the form (a), (b), (c) and (d), there are two special cases which are typical of the problems encountered in practice. In one, which we shall label Type I, all of the granules in \( E \) are of the form (a), and \( E \) may be regarded as the conjunction of \( q_1, \ldots, q_n \). In the other, all of the granules in \( E \) are of the form (b) and (d), and the evidence is said to be of Type II. In the latter case, we shall assume for simplicity that \( X \) ranges over a finite set which for convenience may be taken to be the set of integers \( \{1, \ldots, n\} \).

As a simple illustration of evidence of Type I, assume that we are interested in Penny's age and that the available evidence about her age is comprised of the following soft data granules:

\( q_1 \triangleq \text{Penny is very young is unlikely} \)
\( q_2 \triangleq \text{Penny is young is very likely} \)
\( q_3 \triangleq \text{Penny is not young is unlikely} \)

As an illustration of evidence of Type II, we may have, as in Example (f) in Section 1:

(b) \( q_1 \triangleq \text{If Penny is an undergraduate student, then she is very young} \)
\( q_2 \triangleq \text{If Penny is a graduate student, then she is young} \)
g_3 \triangleq \text{If Penny is a doctor then she is not very young}

\begin{align*}
g_4 & \triangleq \text{Penny is an undergraduate student is unlikely} \\
g_5 & \triangleq \text{Penny is a graduate student is likely} \\
g_6 & \triangleq \text{Penny is a doctor is not likely}
\end{align*}

Given a collection of data granules such as those appearing in (a) and (b), we wish to infer from E an answer to a question of the general form:

$$q \triangleq \text{V is } Q \text{ is } ?\alpha$$

(6.5)

where Q is a specified fuzzy subset of V and ?\alpha is the desired linguistic probability. For example:

$$q \triangleq \text{Penny is not very young is } ?\alpha$$

(6.6)

to which the answer might be, say,

$$?\alpha \triangleq \text{not very likely}$$

In the case of evidence of Type I, an answer to a question of the form (6.5) may be obtained, in principle, by using the mathematical programming technique employed in Example 7, Section 5. In the case of evidence of Type II, however, we shall use a different approach involving a replacement of the posed question with a surrogate question, q_s, that is, a question which, unlike q, may be answerable based on the information contained in E. Such a question in the case of (6.6), for example, might be

$$q_s \triangleq \text{What is the degree of certainty that Penny is not very young?}$$

or

$$q_s \triangleq \text{What is the degree of possibility that Penny is not very young?}$$

The approach described in the sequel is based on a generalization of the concepts of upper and lower probabili-
ties [17, 29] which serve as a point of departure for Shafer's theory of evidence [67]. Viewed from the perspective of our approach, the latter theory is concerned with the special case where (a) the evidence is of Type II; (b) the $G_i$ and $Q$ are nonfuzzy sets; and (c) the $p_i$ are numerical probabilities.

Assuming, first, that the $G_i$ are fuzzy sets but the $p_i$ are numerical probabilities, we define the conditional possibility and the conditional certainty of the proposition "$Y$ is $Q$" (or, equivalently, the event "$Y$ is $Q$") given that "$Y$ is $G_i$" as follows:

$$\text{Poss}\{Y \text{ is } Q \mid Y \text{ is } G_i\} = \sup(Q \cap G_i) \quad (6.7)$$

$$\text{Cert}\{Y \text{ is } Q \mid Y \text{ is } G_i\} = \inf(G_i \Rightarrow Q) \quad (6.8)$$

where

$$\sup(Q \cap G_i) = \sup_v (\mu_Q(v) \wedge \mu_{G_i}(v)) \quad (6.9)$$

$$\inf(G_i \Rightarrow Q) = \inf_v ((1 - \mu_{G_i}(v)) \cdot \mu_Q(v)) \quad (6.10)$$

and $\mu_Q$ and $\mu_{G_i}$ are the membership functions of $Q$ and $G_i$, respectively.

In effect, the right-hand members of (6.7) and (6.8) serve as measures of the degree to which the proposition "$Y$ is $G_i$" influences one's belief in the proposition "$Y$ is $Q$." In particular, (6.7) serves as a measure of the degree of possibility while (6.8) plays the same role in relation to the degree of certainty. Note that when $Q$ and $G_i$ are nonfuzzy, we have

$$\sup(Q \cap G_i) = 1 \text{ if } Q \cap G_i \text{ is nonempty} \quad (6.11)$$

$$= 0 \text{ if } Q \cap G_i = \emptyset$$

and

$$\inf(G_i \Rightarrow Q) = 1 \text{ if } G_i \subseteq Q \quad (6.12)$$

$$= 0 \text{ otherwise}$$

Now since $X$ is assumed to be a random variable which takes the values $1, \ldots, n$ with respective probabilities
p_1, \ldots, p_n$, the conditional possibility and conditional certainty of the proposition "Y is Q" are also random variables whose respective expectations are given by

$$\text{E}P(Q) = \sum_i p_i \sup (Q \cap G_i)$$

$$= \sum_i p_i \sup_v (\mu_Q(v) \vee \mu_{G_i}(v))$$

(6.13)

$$\text{EC}(Q) = \sum_i p_i \inf (G_i \Rightarrow Q)$$

$$= \sum_i p_i \inf_v ((1 - \mu_{G_i}(v)) \vee \mu_Q(v))$$

$$= 1 - \text{E}P(Q')$$

(6.14)

We shall refer to $E_P(Q)$ and $EC(Q)$ as the expected possibility and the expected certainty, respectively, of the proposition "Y is Q." When Q and $G_1, \ldots, G_n$ are nonfuzzy, $EC(Q)$ and $E_P(Q)$ reduce to the Shafer's *degree of belief* and *degree of plausibility*, respectively, which correspond to the lower and upper probabilities in Dempster's work [17].

Our feeling is that Shafer's identification of "degree of belief" with the lower rather than the upper probability (or, more generally, with $E(Q)$ rather than $E_P(Q)$) is open to question, since there is no particular reason for singling out $EC(Q)$ or $E_P(Q)$ or, for that matter, any convex combination of them as a universal measure of the degree of belief.

Having defined the concepts of expected certainty and expected possibility, we are in a position to see the rationale for employing the technique of surrogate questions in the case of evidence of Type II. Taking for simplicity the case where the $G_i$ and Q are nonfuzzy and the $p_i$ are numerical probabilities, the evidence can be expressed in the form

---

8 It should be remarked that $EC(Q)$ and $E_P(Q)$ are not normalized—as are the lower and upper probabilities in the work of Dempster and Shafer. As is pointed out in [95], the normalization in question leads to counterintuitive results in application to combination of bodies of evidence.
Possibility Theory and Soft Data Analysis

\[ g_1 \triangleq \gamma \in G_1 \text{ or } g_2 \triangleq \gamma \in G_2 \text{ or } \ldots \text{ or } g_n \triangleq \gamma \in G_n \]

\[ \text{Prob}(g_1) = p_1 \text{ and } \text{Prob}(g_2) = p_2 \text{ and } \ldots \text{ and } \text{Prob}(g_n) = p_n \]

Now let us assume that the original question is: What is the numerical probability that \( \gamma \in Q \)? It is easy to see that the granularity of available evidence makes it infeasible to answer questions of this type for arbitrary \( Q \). Thus, we are compelled to replace the original unanswerable question with a surrogate answerable question which in some sense is close to the original question. In the case under discussion, such questions would be:

(a) What is the expected certainty (or, equivalently, the degree of belief (Shafer) or the lower probability (Dempster)) that \( \gamma \in Q \)?

(b) What is the expected possibility (or, equivalently, the degree of plausibility (Shafer) or the upper probability (Dempster)) that \( \gamma \in Q \)?

Based on the available evidence, the answers to (a) and (b) are:

\[ EC(Q) = \sum_i p_i \inf(G_i \Rightarrow Q), \quad i = 1, \ldots, n \]

and

\[ EP(Q) = \sum_i p_i \sup(G_i \cap Q) \]

where (see (6.11) and (6.12))

\[ \inf(G_i \Rightarrow Q) = 1 \text{ if } G_i \subseteq Q \]

\[ = 0 \text{ otherwise} \]

and

\[ \sup(G_i \cap Q) = 1 \text{ if } G_i \cap Q = \emptyset \]

\[ = 0 \text{ otherwise} \]

A serious shortcoming of the Shafer-Dempster approach is that if \( G_i \) and \( Q \) are nonfuzzy and the condition

\[ G_i \subseteq Q \]

is not satisfied exactly, then no matter how small the error
might be the contribution of the term \( p_i \inf(G_i \Rightarrow Q) \) to the value of \( EC(Q) \) in the summation

\[
EC(Q) = \sum_i p_i \inf(G_i \Rightarrow Q)
\]

would be zero. In intuitive terms, what this means is that a piece of evidence will be disregarded so long as there is the slightest doubt about its perfect validity. We avoid this extreme degree of conservatism in our approach by (a) allowing the \( G_i \) and \( Q \) to be fuzzy; and (b) fuzzifying the concept of containment, with the expression \( \inf(G_i \Rightarrow Q) \) in (6.14) representing, in effect, the degree to which \( G_i \) is contained in \( Q \). Thus, if \( G \) is regarded as a random variable which takes the values \( G_1, \ldots, G_n \) with respective probabilities \( p_1, \ldots, p_n \), then we can write

\[
EC(Q) = \text{Prob}\{G \subset Q\}
\]  

(6.15)

with the understanding that \( G \subset Q \) is a fuzzy event \([86]\) and that the degree to which \( G \subset Q \) is satisfied is expressed by

\[
\text{degree}(G \subset Q) = \inf(G \Rightarrow Q)
\]

Viewed in this perspective, (6.15) may be regarded as a natural generalization of Dempster's lower probability and Shafer's degree of belief.

For the purpose of illustration, we shall conclude this section by describing the application of our approach to Example (b). In this example, the \( G_i \) and \( Q \) are fuzzy and the \( p_i \) are linguistic probabilities. More specifically, we have

\[
\begin{align*}
G_1 & \triangleq \text{YOUNG}^2 \\
G_2 & \triangleq \text{YOUNG} \\
G_2' & \triangleq (\text{YOUNG}^2)' \\
Q & \triangleq \text{YOUNG}' \\
p_1 & \triangleq \text{UNLIKELY} = \text{ANT LIKELY} \\
p_2 & \triangleq \text{LIKELY} \\
p_3 & \triangleq (\text{LIKELY})'
\end{align*}
\]
where YOUNG is the denotation of young, YOUNG$^2$ is the denotation of very young, ANT is the antonym, i.e., (see (3.40))

$$\mu_{\text{ANT LIKELY}}(v) = \mu_{\text{LIKELY}}(1-v), \quad v \in [0,1]$$  (6.16)

and the prime represents the complement.

Now let

$$\alpha_1 = \sup(\text{YOUNG}^2 \cap \text{YOUNG}')$$  (6.17)

$$\alpha_2 = \sup(\text{YOUNG} \cap \text{YOUNG}')$$  (6.18)

$$\alpha_3 = \sup((\text{YOUNG}^2)' \cap \text{YOUNG}')$$  (6.19)

where the $\alpha_i$ are numbers in the interval $[0,1]$. (From (6.18) it follows that $\alpha_2 = 0.5$ but we shall not make use of this fact.) Then, using (6.13) we can express $E\Pi(Q)$ as

$$E\Pi(Q) = \alpha_1 \text{UNLIKELY} \circ \alpha_2 \text{LIKELY} \circ \alpha_3 \text{LIKELY}'$$  (6.20)

where $\circ$ denotes the sum of fuzzy numbers $[90,50,18]$.

To compute $E\Pi(Q)$ as a fuzzy number, we have to take into consideration the fact that the numerical probabilities must sum up to unity. Thus, on denoting these probabilities by $v_1'$, $v_2'$, and $v_3'$, and applying the extension principle (4.20), the determination of the membership function of $E\Pi(Q)$ is reduced to the solution of the following variational problem:

$$\mu(z) \triangleq \max_{v_1', v_2', v_3'} (\mu_{\text{LIKELY}}(1-v_1) \circ \mu_{\text{LIKELY}}(v_2)$$

$$\quad \circ (1-\mu_{\text{LIKELY}}(v_3)))$$

subject to

$$z = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$$  (6.22)

$$1 = v_1 + v_2 + v_3$$

Thus, expressed as a fuzzy set, we have

$$E\Pi(Q) = \int_{[0,1]} \mu(z)/z$$  (6.23)

where $\mu(z)$ is given by (6.21). To compute $E\Pi(Q)$, then, we can make use of the identity (6.14)

$$E\Pi(Q) = 1 - E\Pi(Q')$$  (6.24)
From our definitions of $\mathbb{E}(Q)$ and $\mathbb{C}(Q)$ it is a simple matter to derive a basic rule of conditioning which may be regarded as a generalization of those given by Dempster and Shafer. Specifically, assume that the evidence has the form:

If $X = i$ then $Y$ is $G_i$, $i = 1, \ldots, n$

$\mathbb{P}(X = i) = p_i$

and, in addition, we know that

$g_0 \overset{\Delta}{=} Y$ is $G_0$

where $G_0$ is a given fuzzy subset of $V$.

Clearly, the available evidence may be expressed in the equivalent form:

If $X = i$ then $Y$ is $G_i \cap G_0$, $i = 1, \ldots, n$

$\mathbb{P}(X = i) = p_i$

which implies that

$\mathbb{E}(Q)$ conditioned on "$Y$ is $G_0$" = $\mathbb{E}(Q \cap G_0)$ \hspace{1cm} (6.25)

and correspondingly

$\mathbb{C}(Q)$ conditioned on "$Y$ is $G_0$" = $1 - \mathbb{E}(Q' \cup G_0')$ \hspace{1cm} (6.26)

Remark. The connection between the definition of expected possibility—as expressed by (6.13)—with that of the upper probability in [17] and [67]—may be made more transparent by interpreting $\mathbb{E}(Q)$ as the probability of a fuzzy event—in the manner of (6.15). More specifically, if $\text{sup}(G \cap Q)$ is regarded as the degree of occurrence of the fuzzy event $G \cap Q?$, in which the question mark serves to signify that we are concerned with the degree to which $G$ intersects $Q$ rather than with the intersection of $G$ and $Q$, then we can write

$\mathbb{E}(Q) = \mathbb{P}(G \cap Q?)$ \hspace{1cm} (6.27)

with the understanding that $G$ is a random variable which
takes the values \( G_1, \ldots, G_n \) with respective probabilities \( p_1, \ldots, p_n \).

In summary, then, the expected possibility and expected certainty may be expressed in the form

\[
E\Pi(Q) = \text{Prob}\{G \cap Q?\} \tag{6.28}
\]

and

\[
EC(Q) = \text{Prob}\{G \subset Q\} \tag{6.29}
\]

which clarifies the sense in which \( E\Pi(Q) \) and \( EC(Q) \) may be viewed, respectively, as generalizations of the concepts of upper and lower probabilities—concepts which are defined in [17] and [67] under the assumption that the \( G_i \) and \( Q \) are nonfuzzy sets.

7. **Concluding Remark**

The approach to the analysis of soft data described in this paper represents a substantive departure from the conventional probability-based methods.

The main thesis underlying our approach is that, in general, the uncertainty which is intrinsic in soft data is a mixture of probabilistic and possibilistic constituents and, as such, must be dealt with by a combination of probabilistic and possibilistic methods. We have indicated, in general terms, how this can be done through the use of the concept of a possibility distribution and the related concepts of a linguistic variable, semantic entailment, semantic equivalence, and the extension principle. Finally, we have shown how the concepts of expected possibility and expected certainty relate to the important issue of credibility analysis, and indicated a way of reducing many of the problems in inference from soft data to the solution of nonlinear programs.

The issues associated with soft data analysis are varied and complex. Clearly, we have— at this juncture—only a partial understanding of the basic problem of inference from
soft data and the associated problem of credibility assessment. What is likely, however, is that, in the years to come, our understanding of these and related problems will be enhanced through a further development of possibility-based methods for the representation and manipulation of soft data.

8. References and Related Papers

128 Lotfi A. Zadeh

60. N. Rescher, Plausible Reasoning (Van Gorcum, Amsterdam, 1976).
63. E. Sanchez, Fuzzy relations, University of Marseille (1974).
64. E. Sanchez, M77/28, Univ. of Calif. Berkeley (1977).
85. F. M. Williams, Bayesian conditionalization and the principle of minimum information, University of Sussex (1978).
90. L. A. Zadeh, Information Sciences 8, 199 (1975); 8, 301 (1975); 9, 43 (1975).