Modelling Uncertainty

ANTONELLA BODINI

CNR-IMATI, Milano

Preface.

This paper is based on a talk given in Varenna, Lake Como, Italy, September 2003, at the Summer School TOWARDS ELECTRONIC DEMOCRACY: INTERNET-BASED DECISION SUPPORT organized by CNR-IMATI-Milano branch and supported by The European Science Foundation, establishing the programme TOWARDS ELECTRONIC DEMOCRACY (TED) with the objective of discussing and evaluating how advances in interactive decision analytic tools might help develop inclusive e-democratic systems which involve their electorates more fully in decision making.

The lectures and e-learning materials from the summer school are available at http://www.sal.hut.fi/TED/ .
1 Introduction.

The aim of this lecture is to show that there are several theories to both describe and measure uncertainty, which are different to the probabilistic approach. An outline for the paper is as follows. First, in Section 2 we give a definition of uncertainty, which will lead us to distinguish between uncertainty in the strict sense of the word and imprecision. Then we will move on to some different models used to describe them. In Section 3 we will consider the problem of measuring uncertainty. Section 4 provides some ideas on Fuzzy Sets Theory as a tool to describe imprecision, in particular in connection with the problem of Decision Making.
2 A definition of uncertainty.

When can we speak of uncertainty? Roughly speaking, when our information is not perfect. In general we can recognize two aspects of imperfection in the information: imprecision and uncertainty. We will look at the difference between these concepts by some examples. Let’s consider the following sentences:

a) “In the bottle there is about 1/4 l of water”

b) “I arrived in Varenna between 3.00 and 4.00 p.m.”

c) “Next Sunday there will be a match between Milan-Juventus: which team is going to win?”

d) “I’m going to toss a coin to decide if I go to the party or not”.

In a) and b) we deal with an imprecision, as the value of the data (quantity of water, time of arrival) isn’t exactly specified; in c) and d) we deal with uncertainty as the value of the data is exactly specified (heads/tails, Juventus/Milan) but we don’t know what it will be.

As imprecision and uncertainty are related to imperfect information, let me just start with a definition of information. Information is usually brought by a proposition: a telephone directory is not a piece of information, a piece of information is an answer to a question. A proposition associates to an object, $O$, an attribute, $A$, which, in turn, can take some value, $V$, of a scale established in advance. We can always associate to any proposition a number which indicates degree of confidence, $D$, in the proposition itself. For example:

Tomorrow $\underbrace{\text{the weather}}_{O}$ will $\underbrace{\text{almost surely}}_{D}$ be $\underbrace{\text{very}}_{V}$ $\underbrace{\text{hot}}_{A}$.

The four elements $(O, A, V, D)$ form an element of information: then we can say that imprecision is related to $V$ while uncertainty to $D$.

In the next part of this lecture let’s concentrate on uncertainty.
In order to deal with mathematical description and measurement of uncertainty, let me introduce a model for an element of information which doesn’t handle propositions but which is similar to the classical model identifying any proposition with a point, or a subset, of a suitable set.

Let \( \Omega \) be the set of all the elementary events and \( \mathcal{S} \) a family of subsets of \( \Omega \), the class of observable events. Let’s suppose that \( \mathcal{S} \) is an algebra, i.e.:

1. \( \emptyset \in \mathcal{S} \)
2. \( A \in \mathcal{S} \Rightarrow \overline{A} \in \mathcal{S} \), where \( \overline{A} = \{ \omega \in \Omega \mid \omega \notin A \} \)
3. \( A, B \in \mathcal{S} \Rightarrow A \cup B \in \mathcal{S} \)

At this point we can provide a description of the uncertainty pertaining \((\Omega, \mathcal{S})\) by means of a function

\[
g : \mathcal{S} \to \mathbb{R}^+\]

indicating our confidence in every observable event.

Traditionally, it was taken for granted that probability theory is an adequate model for the assessment of this confidence. However, it is more and more recognized that the concept of uncertainty is too broad to be captured by probability theory alone. With regard to this, let’s consider in particular some issues:

- in classical probability theory the degree of confidence in each event is required to be expressed precisely by a real number. This precision requirement is a consequence of the additivity axiom and is often difficult to satisfy, in particular when we are dependent on assessment based on subjective human judgements.

- the total absence of information is not captured by probability theory, because of the condition of normalization, for which if \( Pr(A) = 0 \) then \( Pr(\overline{A}) = 1 \). In the face of a total absence of information, a reasonable condition for the function \( g \) would be \( g(A) = g(\overline{A}) = constant \) for all \( A \in \mathcal{S} \).
An alternative approach to the probability theory was introduced by Sugeno (1977).

2.1 Monotone measures of Sugeno.

A monotone measure of Sugeno is a map

\[ g : S \rightarrow \mathbb{R}^+ \]

such that

1. \( g(\emptyset) = 0, \quad g(\Omega) = 1 \)
2. \( A \subset B \implies g(A) \leq g(B) \)
3. \( A_i \uparrow A, \) or \( A_i \downarrow A \implies \lim_{i \to \infty} g(A_i) = g(A) \)

So any probability measure is a monotone measure of Sugeno.

From the definition we have that

4. \( g(A) \in [0, 1] \)
5. \( g(A \cup B) \geq \max\{g(A), g(B)\} \)
6. \( g(A \cap B) \leq \min\{g(A), g(B)\} \)

Let’s consider some examples:

**Example 1.** Let \( \Omega \) be the set \( \{a, b, c\} \). The set function \( g_i \) defined in the following table

<table>
<thead>
<tr>
<th>( A )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
<th>( a, b )</th>
<th>( a, c )</th>
<th>( b, c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g_i(A) )</td>
<td>0.5</td>
<td>0.5</td>
<td>0.75</td>
<td>0.7</td>
<td>0.9</td>
<td>0.75</td>
</tr>
</tbody>
</table>

is a monotone measure of Sugeno which is not a probability measure, indeed:

\[ g_1(a) + g_1(b) + g_1(c) = 1.75 > 1 \]

**Example 2.** Let \( f \) be a function

\[ f : \Omega \rightarrow [0, 1] \]
such that
\[ \sup_{\omega \in \Omega} f(\omega) = 1 \]
then
\[ g_2(S) \rightarrow [0, 1] \]
\[ g_2(\emptyset) = 0, \quad g_2(A) = \sup_A f(\omega) \quad \forall A \neq \emptyset \]
defines a monotone measure of Sugeno which, in general, is not a probability measure as we can see, for example, by considering \( \Omega = \{a, b, c\} \) and the function \( f \) so that \( f(a) = f(b) = 1 \). From the definition we have that \( g_2(\{a\}) + g_2(\{b, c\}) = 1 + 1 = 2 \).
\[ \square \]
In the class of the monotone measures of Sugeno some are well known and very important: let me discuss them in slightly more detail.
Moreover let me restrict the treatment to finite sets. All concepts and majority of
the results which I’m introducing are valid for much more general spaces. But if we consider the more general spaces, the “ideas” would be buried under a lot of technical complications which here isn’t important to examine. So, from now on let \( \Omega \) be a finite space.

2.1.1 Possibility-Necessity measures.

We obtain an important class of Sugeno’s measures by considering all the functions \( g \) such that
\[ g(A \cup B) = \max \{g(A), g(B)\} \quad \text{for all events } A, B. \]
Such a measure is called possibility measure and is indicated by \( Po \). This definition reflects the use of the term “possibility” in the common parlance in the sense that:

a) an event can be slightly possible or quite possible and so on; b) the greater degree
between \( Po(A) \) and \( Po(B) \) guarantees the same degree of possibility to \( A \cup B \).

From the definition it follows that at least one between \( A \) and \( \overline{A} \) is certainly possible
\[ \max \{Po(A), Po(\overline{A})\} = 1 \]
but this fact doesn’t imply that $Po(A) = 0$ when $Po(A) = 1$.

The measure $g_2$ introduced in Example 2 is a possibility measure. In particular, let’s consider a dominated statistical model $\{(X, \mathcal{X}, P_\theta) : \theta \in \Theta\}$ where $\Theta$ is a finite set. Let $l$ be the corresponding likelihood function; define the function $l^*$ as the supremum on $\theta$ of the likelihood:

$$l^*(x) = \sup_{\theta} l(\theta; x).$$

Choose $\Omega = \Theta$ and $f$ equal to the normalized likelihood:

$$f(\theta) = \frac{l(\theta; x)}{l^*(x)}$$

Then $g_2$

$$(\Theta \supseteq \Lambda \mapsto \sup_{\Lambda} \frac{l(\theta; x)}{l^*(x)}$$
defines a possibility measure having a direct statistical interpretation in terms of maximum likelihood. □

We can construct a possibility measure in the following manner: firstly, we choose a nested sequence

$$\mathcal{F} = \{E_1, \ldots, E_I\}, \quad E_i \subset E_{i+1}$$

of observable events, the family of the focal elements; after that, we assign to each focal element $E_i$ a strictly positive real number $m(E_i)$ so that

$$\sum_{i=1}^{I} m(E_i) = 1$$

that is we introduce a basic assignment function, $m$; finally, we define the possibility of $A$ in the following manner:

$$Po(A) = \sum\{m(E_i) \mid i : E_i \cap A \neq \emptyset\}, \quad Po(\emptyset) = 0$$
For example:

![Diagram of body of evidence](image)

\[ Po(A) = m(E_2) + m(E_3) \]

The couple \((\mathcal{F}, m)\) is called **body of evidence** and represents the available knowledge on \((\Omega, \mathcal{S})\). So, the possibility of \(A\) is the sum of all piece of knowledge which agree with \(A\). □

We obtain a different class by considering the measures \(g\) such that

\[ g(A \cap B) = \min\{g(A), g(B)\} \quad \text{for all events } A, B. \]

Such a measure is called **necessity measure** and is indicated by \(N\). From the definition it follows that at least one between \(A\) and \(\overline{A}\) is not necessary, so

\[ N(A) + N(\overline{A}) \leq 1 \quad \forall A. \]

We can again define a necessity measure starting from a body of evidence in this way: the necessity of \(A\) is the sum of the basic assignments for the focal elements \(E_i\) which imply \(A\):

\[
N(A) = \sum \{m(E_i) \mid i : E_i \subseteq A\}, \quad N(\emptyset) = 0
\]
For example:

\[ N(A) = m(E_i) \]

Note that we can go from a Possibility measure to a Necessity measure by duality

\[ Po(A) = 1 - N(\overline{A}) \]

and viceversa

\[ N(A) = 1 - Po(\overline{A}). \]

Starting from the representation by focal elements and basic assignment functions we can immediately get to an important generalization of possibility/necessity measures. It is well known as the *Mathematical Theory of Evidence* and was introduced by Dempster (1967) and Shafer (1976).
2.2 Mathematical Theory of Evidence.

Let $m$ be a basic assignment function, that is

$$m : \mathcal{S} \rightarrow [0, 1]$$

such that

$$\sum_{A \in \mathcal{S}} m(A) = 1$$

Let me call any event $E$ such that $m(E) > 0$ a focal element. A Belief measure is a set function

$$Bel : \mathcal{S} \rightarrow [0, 1]$$

defined in this way:

$$Bel(A) = \sum \{m(E) \mid E \subseteq A\}$$

Observe that the difference between a Belief measure and a Necessity measure is that now no structure is imposed on $\mathcal{F}$. For example:

A Belief measure is a monotone measure of Sugeno satisfying the following properties:

1. $Bel(A) + Bel(\overline{A}) \leq 1$

2. $Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$
The dual measure of a Belief measure is called a **Plausibility measure**:

\[ Pl(A) = 1 - Bel(\overline{A}) \]

In terms of focal elements a Plausibility measure is defined in this way:

\[ Pl(A) = \sum \{m(E) \mid E \cap A \neq \emptyset \} \]

Of course, if the focal elements are nested then a Belief measure is a Necessity measure and a Plausibility measure is a Possibility measure.

A Belief measure is a Probability measure if and only if

\[ Bel(A \cup B) = Bel(A) + Bel(B) - Bel(A \cap B) \quad \forall A, B \]

In the case here considered of a finite set \( \Omega \), given a Belief measure \( Bel \) we can determine its basic assignment function (and the focal elements) in the following manner:

\[ m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B). \]

For more general spaces there aren’t, in my knowledge at least, more general results.

The following diagram shows the relationships among the classes of Sugeno’s monotone measures introduced here. For the sake of completeness we note that the measure \( g_1 \) of the Example 1 doesn’t belong to any of these classes.
So a probability measure is a monotone measure of Sugeno. Now, let \((\mathcal{F}, m)\) be a body of evidence and \((\text{Bel}, P_l)\) the conjugate monotone measures associated to \((\mathcal{F}, m)\). For any \(E\) in \(\mathcal{F}\) let’s fix an element \(x_E \in E\) and consider the map \(p:\)

\[ p : \Omega \to \mathbb{R}^+ \]

\[ p(x) = \sum \{ m(E) \mid E \in \mathcal{F} : x_E = x \} \]

This map is a probability density function. Indeed, \(0 \leq p(x) \leq 1\) because \(m\) is a nonnegative function and \(\sum_E m(E) = 1\), by the definition of \(m\). Moreover, if \(x\) doesn’t represent any focal element \(E\) then \(p(x) = 0\) while, if \(\mathcal{X}\) is the set of the focal element \(E\) such that \(x = x_E\), then \(p(x) = \sum_{E \in \mathcal{X}} m(E)\). By construction \(\mathcal{X} \cap \mathcal{X}' = \emptyset\) and \(\cup \mathcal{X} = \mathcal{F}\), then

\[ \sum_{x \in \Omega} p(x) = \sum_x \sum_{E \in \mathcal{X}} m(E) = \sum_E m(E) = 1. \]

This proof is quite cumbersome; it becomes very simple after considering a picture of the situation:
Here \( p(x_1) = m(E_1) + m(E_3) \) and \( p(x_2) = m(E_2) \) are the probability masses of \( p \).

Of course the corresponding probability \( P \) is defined as:

\[
P(A) = \sum \{m(B) \mid B \in \mathcal{F} : x_B \in A\}
\]

therefore

\[
P(A) \geq \sum \{m(B) \mid B \subseteq A\} = Bel(A)
\]

\[
P(A) \leq \sum \{m(B) \mid B \cap A \neq \emptyset\} = Pl(A)
\]

that is

\[(*) \quad Bel(A) \leq P(A) \leq Pl(A)\]

Then we can think to \( Bel \) and \( Pl \) as bounds for a probability measure.

That brings me to the introduction of the concept of **lower/upper probability** and, more in general, of **imprecise probability**.
2.3 Imprecise Probabilities.

It is recognized that imprecise probabilities of different types exist, for example:

- Sugeno’s measures
  - Beliefs functions
  - Possibility measures
- upper/lower probabilities
- Choquet capacities
- upper/lower previsions

and

- sets of probability measures (Levy, 1980; Berger, 1994)
- comparative probability orderings (Keynes, 1921; Koopman, 1940; Fine, 1973 and 1977; Fishburn, 1986)
- classificatory models (Fine, 1973, Walley & Fine, 1979)
- sets of desiderable gambles (Williams, 1976; Walley, 1991)
- intervals of measures (DeRobertis & Hartigan, 1981)
- Baconian probabilities (Cohen, 1977).

I’ll focus on the first group in particular because they are generalizations of Sugeno’s measures in some sense.
2.3.1 Upper/Lower Probability.

Upper/Lower probabilities were introduced by C. Smith (1961) starting from the idea – well known from Ramsey (1931) and de Finetti (1931) – that the most customary and obvious measure of personal belief is provided by betting. However, I’d prefer to follow an axiomatic approach rather than to introduce the problem of betting as a tool to measure personal beliefs.

To do this, let’s consider again (⋆) and observe that if we choose for some focal element \( B \) a different element \( x' \) representing it, then we obtain a probability measure \( P' \) different from \( P \) and having the same bounds. So, it is natural to consider, for a given subset \( P_1 \) of the set \( P \) of all probability measures defined on \( (\Omega, \mathcal{A}) \), where \( \mathcal{A} \) is the family of all subsets of \( \Omega \), the map \( v_* : \)

\[
v_*(A) = \inf \{ P(A) \mid P \in P_1 \}
\]

This set function is called lower probability and its conjugate function

\[
v^*(A) = \sup_{P_1} P(A) = 1 - v_*(\overline{A})
\]

upper probability.

The importance of considering objects like these, that is bounds of classes of probability measures, will come clear when, for example, we consider as \( \Omega \) a sampling space or a parameter space: in the classical framework, lower/upper probabilities on the sampling space might model uncertainty about the sampling distribution and in the Bayesian framework lower/upper probabilities on the parameter space might model uncertainty about the prior distribution. So, we can argue that we can apply these concepts in classical and Bayesian robustness analysis. Moreover, we might use lower/upper probabilities to model the disagreement among the different probability assessments provided by several experts on a given phenomenon.

As we can go from one to the other simply by duality, let’s consider lower probability only.
From the definition we have that

1. \( v_*(\emptyset) = 0, \quad v_*(\Omega) = 1 \)

2. \( A \subset B \Rightarrow v_*(A) \leq v_*(B) \)

3. \( A \cap B = \emptyset \Rightarrow v_*(A \cup B) \geq v_*(A) + v_*(B) \).

In general, these conditions are not sufficient to define a lower probability. Indeed, let \( \Omega \) be the set \( \{x_1, x_2, x_3, x_4\} \) and \( \underline{v} \) the function defined in this table

| \(|A|\) | 0 | 1 | 2 | 3 | 4 |
|-------|---|---|---|---|---|
| \( \underline{v} \) | 0 | 0 | 0.5 | 0.5 | 1 |
| \( P \)   | 0 | 0.25 | 0.5 | 0.75 | 1 |

where \( |A| \) denotes the cardinality of \( A \). We can easily see that \( \underline{v} \) satisfies 1.-3. and by solving a simple system of inequalities that only the uniform probability \( P \) dominates \( \underline{v} \), that is \( P(A) \geq \underline{v}(A) \), but \( P \) and \( \underline{v} \) don’t coincide so \( \underline{v} \) is not a lower probability. \( \square \)

Note that sometimes we can find the following definition: a lower probability is a set functions which satisfies 1.-3. But, in my knowledge at least, such a definition is so general that we immediately have to add further conditions to make it useful. In any way there exist some necessary and sufficient conditions so that a function which satisfies 1.-3. is a lower probability, see Huber (1981, Ch. 10) for some examples.

The following functionals, called lower/upper expectation respectively,

\[
E_*(X) = \inf_{P_1} \int X \, dP, \quad E^*(X) = \sup_{P_1} \int X \, dP
\]

are also relevant in Statistics and Decision Making: think to \( X \) as an utility function, for example.

Of course

\[
E_*(I_A) = v_*(A)
\]
but if $X$ isn’t an indicator function calculating $E_s(X)$ by the definition can be very difficult. The situation is simplified when $\mathbb{P}_1$ is $m$-closed:

$$\mathbb{P}_1 = \{ P \in \mathbb{P} : P(A) \geq v_s(A) \forall A \in \mathcal{A} \}$$

and $v_s$ is a 2-monotone Capacity because in this case $E_s(X)$ is a Choquet’s integral (Choquet, 1953).

2.3.2 Choquet’s capacities.

A 2-monotone capacity is a set function

$$v : \mathcal{A} \to \mathbb{R}^+$$

satisfying

1. $v(\emptyset) = 0$, $v(\Omega) = 1$

2. $A \subset B \Rightarrow v(A) \leq v(B)$

3. $v(A \cup B) \geq v(A) + v(B) - v(A \cap B)$ for all $A, B$

Note the difference between this condition 3. and the above property 3. of a lower probability:

$$A \cap B = \emptyset \Rightarrow v_s(A \cup B) \geq v_s(A) + v_s(B)$$

The conjugate function

$$u(A) := 1 - v(\overline{A})$$

is a 2-alternating capacity, that is

$$u(A \cup B) \leq u(A) + u(B) - u(A \cap B) \forall A, B$$

For some results in more general spaces (recall that here $\Omega$ is a finite set) see for example Huber & Strassen (1973) and Wasserman (1990).

Huber & Strassen show that the 2-alternating structure is necessary and sufficient for generalizing the Neyman-Pearson Lemma to sets of probabilities. Wasserman (1990)
considers a particular class \( \mathbb{P}_1 \) of probabilities on the parameter space with respect to which \( \nu_i \) is a \textit{Belief function} and he shows that this fact permits to have tractable formulas for the inference, in the sense of numerical calculations.

Now let's consider some examples of classes of probability widely used in robustness analysis which are characterized by Choquet capacities.

Let \( \{(X, \mathcal{X}, P_\theta) : \theta \in \Theta \} \) be a dominated model. Let \( \Theta \) be a finite space\(^1\) and introduce a prior distribution \( \pi_0 \) on \( \Theta \). Let \( \mathbb{P} = \mathbb{P}(\Theta) \) the space of all probability measures on \( \Theta \).

Robustness analysis is concerned with the sensitivity of the results of the inference with respect to the assumptions on the adopted model.

In particular, in Bayesian inference an interesting issue is checking if the inference is sensitive to the varying of the prior in a \textit{small neighborhood} of \( \pi_0 \).

Several of such neighborhoods can be characterized by means of capacities, first of all the well known class of the \textit{\( \varepsilon \)-contaminated priors}

\[
\mathcal{I} = \{ \pi = (1 - \varepsilon)\pi_0 + \varepsilon \alpha \mid \alpha \in \mathbb{P} \}
\]

which is characterized by the Plausibility function

\[
\nu^*(A) = \sup_{\pi \in \mathcal{I}} \pi(A) = \begin{cases} 
(1 - \varepsilon)\pi_0(A) + \varepsilon & \text{for } A \neq \emptyset \\
0 & \text{for } A = \emptyset
\end{cases}
\]

In Huber (1981) we find the bounds for the corresponding class of the posterior probabilities, \( \mathcal{I}(x) = \{ \pi(\bullet | x) : \pi \in \mathcal{I} \} \):

\[
\nu_i(A | x) = \sup_{\mathcal{I}(x)} = \frac{(1 - \varepsilon) \int_A l(\theta) \pi_0(d\theta)}{(1 - \varepsilon) \int_\Theta l(\theta) \pi_0(d\theta) + \varepsilon b}
\]

\[
a = \sup_{A} l(\theta) \quad b = \inf_{A^c} l(\theta)
\]

Some other important classes are:

\(^1\)or a Polish space
a variation neighborhood:

\[ \mathcal{I} = \{ \pi \in \mathbb{P} : \sup_A |\pi_0(A) - \pi(A)| \leq \varepsilon \} \]

characterized by

\[ v^*(A) = \min \{ \pi_0(A) + \varepsilon, 1 \} \]

and a Prohorov neighborhood:

\[ \mathcal{I} = \{ \pi \in \mathbb{P} : \pi(A) \leq \pi_0(A^\varepsilon) + \varepsilon \forall A \} \]

where \( A^\varepsilon = \{ x \in \Theta : \inf_{y \in \Theta} d_\Theta(x,y) \leq \varepsilon \} \), characterized by

\[ v^*(A) = \min \{ \pi_0(A^\varepsilon) + \varepsilon, 1 \}. \]

\[ \square \]

As concerns the application to modelling disagreement among several experts, You can refer to Walley’s papers (1982, 1997) \( \square \)

All these concepts (belief functions, upper/lower probabilities, Choquet capacities) and more besides are considered by Walley (1991) in his book, but in a more general framework.

His main result is a demonstration that reasoning and decision making based on imprecise probabilities satisfy the principles of coherence and avoidance of sure loss, which are generally viewed as principles of rationality.

Walley introduces the more general concept of (coherent) upper/lower prevision in the framework of betting. It is impossible to show in short Walley’s theory – moreover it is a task beyond my knowledge and capabilities – so I’ll simply introduce some definitions and quickly show the connections between upper/lower probability as here introduced and upper/lower previsions.
2.3.3 Upper/Lower Previsions.

Let $\mathcal{L}(\Omega)$ be the set of all the bounded real functions defined on the measurable space $(\Omega, \mathcal{S})$ and $\mathcal{K}$ be a subset of $\mathcal{L}(\Omega)$. A **lower prevision** is any real function $P$ defined on $\mathcal{K}$ and an **upper prevision** is its conjugate function, $\overline{P}$:

$$P : \mathcal{K} \to \mathbb{R}$$

$$\overline{P}(X) = 1 - P(-X)$$

This definition is very general, so it doesn’t seem to have an evident meaning. Therefore, it has a *behavioural* interpretation: any element $X$ of $\mathcal{L}(\Omega)$ represents a *gamble*, that is an uncertain reward which will be paid after observing the value of $X$. Then $P(X)$ represents the supremum buying price for $X$, that is we accept to pay any price smaller than $P(X)$ for the uncertain reward $X$.

That behavioural interpretation plays a fundamental role in Walley’s theory: see Miranda et al. (2002) for some recent developments.

In fact, the interest lies in coherent previsions: in general, we can say that our beliefs are incoherent when we don’t accept some gambles whose outcome is better than the outcome of some combination of accepted gambles. Formally, our beliefs are coherent if for all $m \geq 0$, $n \geq 0$, $X_0, \ldots, X_n \in \mathcal{K}$

$$\sup \left( \sum_{j=1}^{n} G(X_j) - mG(X_0) \right) \geq 0$$

where $G(X) = X - P(X)$.

When $\mathcal{K}$ is a linear subspace then $P$ is coherent if and only if it satisfies the following properties$^2$

- $P(X) \geq \inf(X)$
- $P(\lambda \cdot X) = \lambda P(X)$, for $\lambda > 0$

$^2$Th. 2.5.5, Walley.
\[ P(X + Y) \geq P(X) + P(Y) \]

So \( P(X) = \inf(X) \) and \( \overline{P}(X) = \sup(X) \) are coherent previsions.

\( P \) avoids sure loss if there is no linear combination of acceptable gambles which certainly produces a net loss. Formally:

\[
\sup \sum_{j=1}^{n} G(X_j) \geq 0 \quad \forall \ n \geq 0, \ X_1, \ldots, X_n \in \mathcal{K}
\]

When \( \underline{P} \) and \( P \) coincide and are coherent we call them linear previsions.

Next results show that coherent lower/upper previsions generalize upper/lower probabilities.

Part b) of Th. 3.3.3., say that:

\( P \) is coherent if and only if it is the lower envelope of some class of linear previsions:

\[
P(X) = \inf \{ P_\gamma(X) : \gamma \in \Gamma \}
\]

Now let’s consider a class \( \mathcal{K} \) of indicator functions such that the class of the corresponding sets is an algebra and identify the events \( \bar{A}, A \cap B, A \cup B \) with the functions \( 1 - I_A, I_AI_B, I_A + I_B - I_AI_B \) in the usual way. In Walley a lower prevision \( \underline{P} \) defined on such a \( \mathcal{K} \) is called lower probability, but this is not the same definition which I have introduced above, as we can see by the same counterexample considered on page 16. However, Corollary 3.3.4 say that

\( \underline{P} \) is coherent iff it is the lower envelope of some class of additive probability

so, recalling that \( \Omega \) is a finite set and any additive probability is a \( \sigma \)-additive probability, a coherent lower prevision in Walley’s sense is a lower probability in the sense here introduced.

\( \square \)
At this point we can consider concluded the part of this lecture devoted to some models used to describe uncertainty.

Before going to the next part, let’s briefly sum up the situation.

The problem considered here is how describe a kind of lack of information we have called “uncertainty”. So, firstly we have defined an element of information as a set of 4 elements \((O, A, V, D)\): an object, an attribute of the object, a value for the attribute and a degree of confidence in the proposition \((O, A, V)\); then we have associated uncertainty to \(D\). In order to deal with mathematical description of uncertainty we have identified an element of information with a subset of a suitable set of elementary events.

Finally, we have described uncertainty by a set function \(g\) indicating the degree of confidence in every observable event. According to the different properties required for such a function, we have pointed out the monotone measures of Sugeno, including probability measures, possibility/necessity measures and the theory of evidence, and lower/upper probabilities. Moreover, we have briefly shown that lower/upper probabilities can be generalized to lower/upper previsions.

Next, we have to consider the problem of measuring uncertainty and to both describing and measuring imprecision.

I have organized the remaining part of this lecture in the following manner: firstly, I’ll introduce some measures of uncertainty based on the Dempster-Shafer theory. As these measures are not completely satisfying, I’ll move on to the introduction of an uncertainty measure defined in an axiomatic way.

At this point I will consider completed the part of this exposition devoted to uncertainty and I’ll move on to Fuzzy Sets Theory as a tool to describe imprecision.
3 Measures of uncertainty.

On the basis of the Dempster-Shafer theory several so-called “measures of uncertainty” have been proposed, for example

- **Nonspecificity**, Dempster-Shafer (1976)
  \[
  N(m) = \sum_{\mathcal{F}} m(A) \log_2 |A|
  \]

- **Dissonance**, Yager (1983)
  \[
  E(m) = -\sum_{\mathcal{F}} m(A) \log_2 P(A)
  \]

- **Confusion**, Höhle (1982)
  \[
  C(m) = -\sum_{\mathcal{F}} m(A) \log_2 Bel(A)
  \]

- **Discord**, Kliir-Ramer (1990)
  \[
  D(m) = -\sum_{\mathcal{F}} m(A) \log_2 \sum_{B \in \mathcal{F}} m(B) \frac{|A \cap B|}{|B|}
  \]

In fact, these measures can only capture some aspects of uncertainty depending on the structure of the body of evidence we are considering. Then other measures have been proposed like

- **Total Uncertainty**, Kliir-Ramer (1990)
  \[
  T(m) = D(m) + N(m)
  \]

\[ M(m) = \sum_{x} m(A) \log \frac{|A|}{m(A)} \]

These measures too are not completely satisfactory because, as we will see soon, they don’t meet a reasonable property for a measure of uncertainty. A different point of view, based on an axiomatic theory of information, was followed by Bruno Forte (1969). The presentation of the axiomatic theory of information goes beyond the aim of this lecture, so let me introduce directly the definition of uncertainty measure as presented by Forte rather than discuss the Information Theory.

### 3.1 Uncertainty Theory.

In §2 we pointed out that we deal with uncertainty when we have to start an experiment of some kind, we know exactly all the possible outcomes (heads/tails, Juventus/Milan) but we don’t know what will occur. So, the first step in defining a measure of uncertainty is the introduction of the definition of experiment or *experience*.

We define *complete experience* any finite partition \( \Pi = \{A_1, \ldots, A_n\} \) on \((\Omega, \mathcal{S})\), that is any finite family of events such that:

1) \( A_i \in \mathcal{S} \quad \forall \ i = 1, \ldots, n \)

2) \( A_i \cap A_j = \emptyset \quad \forall \ i \neq j \)

3) \( \bigcup_{i=1}^{n} A_i = \Omega. \)

If

3') \( \bigcup_{i=1}^{n} A_i = A \subset \Omega \)

we denote \( \Pi \) by \( \Pi_A \) and we call \( \Pi_A \) an *incomplete experience*.

Now, we introduce a suitable structure as a basis for definition of uncertainty measure.
Let $\mathcal{E}$ be a family of experiences on $(\Omega, \mathcal{S})$ and $\mathcal{E}_1 = \{\Pi = \{A\} : A \in \mathcal{S}\}$ be the subfamily of $\mathcal{E}$ whose elements are the incomplete experiences consisting of only one event: $\Pi = \{A\}$. Let’s suppose that for each observable event $A$ the corresponding incomplete experience belongs to $\mathcal{E}_1$. Let me denote such an experience simply by the corresponding event.

On $\mathcal{E}$ we define a partial order

$$\Pi_A \preceq \Pi_B \Leftrightarrow A = B, \forall A_i \in \Pi_A \exists B_j \in \Pi_B : A_i \subset B_j$$

which is the usual refinement of partitions, and two inner operations

$$\Pi_A \wedge \Pi_B = \{ C = A_i \cap B_j | \ A_i \in \Pi_A, \ B_j \in \Pi_B \}$$

If $A \cap B = \emptyset$ \quad $\Pi_A \vee \Pi_B = \{A_1, \ldots, A_n, B_1, \ldots, B_m\}$

corresponding to the union and intersection of two partitions. The following figures show a refinement and an intersection:

\[
\begin{array}{c}
\Pi_A = \{A_1, A_2, A_3\} \preceq \Pi_B = \{B_1, B_2\} \\
\Pi_A \wedge \Pi_B = \Pi_{C=A\cap B} = \{A_i \cap B_j | \ A_i \in \Pi_A, \ B_j \in \Pi_B \} \\
\end{array}
\]
Finally we introduce a class $\mathcal{K}$ consisting of all the couples of events that we can consider independent both in an algebraic (if $A \cap B \neq \emptyset$) and a semantic sense. For example, if $A$ is the class of people smoking and $B$ the class of people with lung cancer then $A$ and $B$ are independent in an algebraic sense as they are not disjoint or nested but, on the basis of personal knowledge and opinion about the relation between smoking and disease, we can conclude that these events are not independent in a semantic sense so we can’t include this couple in $\mathcal{K}$.

An uncertainty measure associated with the structure $(\Omega, \mathcal{S}, \mathcal{K}, \mathcal{E})$ is a function $H$ such that

1. $H_{\mathcal{E}_1}$ is an information measure, that is:
   
   1.1 $H(\Omega) = 0$, $H(\emptyset) = +\infty$
   
   1.2 $A \subset B \Rightarrow H(A) \geq H(B)$
   
   1.3 $(A, B) \in \mathcal{K} \Rightarrow H(A \land B) \geq H(A) + H(B)$

2. $\Pi_A \preceq \Pi_B \Rightarrow H(\Pi_A) \geq H(\Pi_B)$

3. If $(A_i, B_j) \in \mathcal{K}$ $\forall$ $i, j$ $\Rightarrow$

   \[ H(\Pi_A \land \Pi_B) = H(\Pi_A) + H(\Pi_B) \]

**Example 1.** If $(\Omega, \mathcal{S})$ is a probabilistic space, the well known Shannon entropy

\[ H(\Pi_A) = \frac{\sum_i P(A_i) \log P(A_i)}{\sum_i P(A_i)} \]

and Rényi entropy

\[ H_\alpha(A) = \frac{1}{1 - \alpha} \log \left( \frac{\sum_i P(A_i)^\alpha}{\sum_i P(A_i)} \right), \quad \alpha < 0 \]

satisfie these axioms. $\blacksquare$

However we can construct some meaning uncertainty measure even without introducing a probability measure, as shown in the following example.
Example 2. Let $J$ an information measure: both the following set functions

$$H(\Pi_A) = \sum_i J(A_i)$$

$$H(\Pi_A) = \inf_i J(A_i)$$

are uncertainty measures in the sense of Forte (and they have a very clear meaning).

Let me consider again the axiom 2. It says that if some uncertainty is associated to the experience whose outcomes are $\{B_1, B_2\}$ and someone even specifies that really the possible outcomes are $\{A_1, A_2, A_3 = B_2\}$, then we will assign to $\Pi_A = \{A_1, A_2, A_3\}$ a greater degree of uncertainty as the number of possible outcomes has increased. The following picture represents the situation using the above notation:

$$\Pi_A \leq \Pi_B \Rightarrow H(\Pi_A) \geq H(\Pi_B).$$

The property of monotonicity with respect to further specifications in the available knowledge can be stated even if we consider bodies of evidence rather than partitions. As we can see by some simple examples, that property is not fulfilled neither from the total uncertainty $T$ nor from the mean total uncertainty $M$: this is one of the main reasons for which that measures are not satisfactory as uncertainty measures. Let’s briefly look to these examples.

Example. Let’s consider the bodies of evidence elements

$$(\mathcal{F}, m) : \mathcal{F} = \{A_1, A_2, A_3, A_4\}$$

$$(\mathcal{F}', m') : \mathcal{F}' = \{A_1' \subset A_1, A_2' = A_1, A_3' = A_2, A_4' = A_3, A_5' = A_4\}$$

where $m, m'$ are defined as
\[(\mathcal{F}', m')\] is refinement of \((\mathcal{F}, m)\) in the sense that each element of \(\mathcal{F}'\) is included in some element of \(\mathcal{F}\), as shown in the above figure, and the basic assignment \(m'\) distributes the knowledge about each element of \(\mathcal{F}\) among the elements of \(\mathcal{F}'\) included in them. Here,

\[m'(A'_1) + m'(A_1) = m(A_1).\]

By simple calculations we obtain that \(T(m) - T(m') = 0.1(\log_2 49 - \log_2 41) > 0\). Moreover we can easily construct an example for which \(T(m) < T(m')\). Indeed, let consider the following bodies of evidence:

\[
(\mathcal{F}, m) : \mathcal{F} = \{A_1, A_2, A_3\}
\]

\[
(\mathcal{F}', m') : \mathcal{F}' = \{A'_1 = A_1, A'_2 \subset A_2, A'_3 = A_2, A'_4 = A_3\}
\]

that is

<table>
<thead>
<tr>
<th></th>
<th>(A'_1)</th>
<th>(A_1)</th>
<th>(A_2)</th>
<th>(A_3)</th>
<th>(A_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>| | | | | |</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\bullet)</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>(m(\bullet))</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>(m'(\bullet))</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>
and

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th>$A'_2$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\bullet</td>
<td>$</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$m(\bullet)$</td>
<td>0.5</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$m'(\bullet)$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

with

$$m'(A_1) + m'(A'_2) + m'(A_3) = m(A_1) + m(A_3)$$

In this case $T(m) - T(m') = -2.25907 < 0$.

So $T$ is not monotone in any sense. We can easily find similar examples for $M$ too, even when the body of evidence is not nested. In the latter case the definition of refinement is slightly more complicated but in the notation only (see Appendix). □

We have to note that the problem of determining an uncertainty measure based on the Dempster-Shafer theory, which satisfies all the natural requirements for such a measure is still today an open question.

Finally, I'll move on to the last argument of this lecture: the Fuzzy Sets Theory as a tool to deal with imprecision.
4 Fuzzy Sets Theory.

Let’s consider the following element of information:

“David is a tall boy”.

In the notation of §2 we have

\[ O = David, \quad A = boy, \quad V = tall, \quad D = 1. \]

What is the meaning of the value “tall”? An usual way to translate this concept in the mathematical language is to identify it with a subset of a universal set, \( \mathbb{X} \): for example, “tall” may be the interval \([175, 185]\)cm, and \( \mathbb{R}^+ \) the universal set.

Note that here we deal with imprecision, as the value of the data “tall” is not exactly specified, in the sense that there isn’t a single value corresponding to “tall”. Moreover, note that if Brian’s height is 174 cm, then Brian is not tall! Of course, someone can suggest identifying “tall” with \([174, 185]\)cm, but then why not with \([173, 185]\)cm? In fact, what happens is that we are disposed to accept a lower left-hand bound but at the same time our belief in the goodness of the correspondence between “tall” and the new interval weakens.

In order to represent this situation Zadeh (1965) introduced a generalization of the indicator function: he considered a function

\[ \tilde{A} : \mathbb{X} \to [0, 1] \]

indicating the membership grade of each element of \( \mathbb{X} \) in the set in question. In other words, \( \tilde{A}(x) \) describes the compatibility between \( x \) and the considered concept. Such a function is called a membership function and the corresponding set a fuzzy set, but usually the latter term is used to indicate the membership function too. In our example we might consider the function

\[ \tilde{A}(x) = \begin{cases} 
0 & \text{if } x \leq 160 \text{ or } x \geq 200 \\
\frac{x-160}{15} & \text{if } 160 \leq x \leq 175 \\
1 & \text{if } 175 \leq x \leq 185 \\
\frac{200-x}{15} & \text{if } 185 \leq x \leq 200 
\end{cases} \]
to represent the fuzzy set corresponding to the concept ‘tall’: the following picture represents it together with the indicator function.

\[ \text{How can we handle Fuzzy sets?} \]

The original theory of fuzzy sets was formulated in terms of the following operators of set complement, union and intersection:

\[ (1) \quad \tilde{A}(x) = 1 - \tilde{\tilde{A}}(x) \]
\[ (2) \quad (\tilde{A} \cup \tilde{B})(x) = \max[\tilde{\tilde{A}}(x), \tilde{B}(x)] \]
\[ (3) \quad (\tilde{A} \cap \tilde{B})(x) = \min[\tilde{\tilde{A}}(x), \tilde{B}(x)] \]

These operators are the natural extensions of the usual operators on crisp sets, but they are not the only possible generalizations: several alternatives have been proposed, for example by Yager (1980):

\[ (1.1) \quad \tilde{A}^w(x) = \left(1 - \tilde{\tilde{A}}^w(x)\right)^{1/w} \]
\[ (1.2) \quad (\tilde{A} \cup \tilde{B})(x) = \min \left(1, \tilde{\tilde{A}}^w(x) + \tilde{B}^w(x)\right)^{1/w} \]
\[ (1.3) \quad (\tilde{A} \cap \tilde{B})(x) = \]
\[ = 1 - \min \left(1, [1 - \tilde{\tilde{A}}^w(x)]^w + [1 - \tilde{\tilde{B}}^w(x)]^w\right)^{1/w} \]

Usually any such class includes Zadeh’s operators: here

(1.1) = (1) for \( w = 1 \)
(1.2), (1.3) = (2), (3) for \( w \to \infty \)
However the original functions possess particular significance, for example with respect to the spreading of any error: if any error $\varepsilon$ is associated to the membership functions $\tilde{A}$ and $\tilde{B}$, then the maximum error associated to $\tilde{A}^c$, $\tilde{A} \cup \tilde{B}$, $\tilde{A} \cap \tilde{B}$ remains $\varepsilon$. In the remaining part of this exposition I’ll only consider the *standard* operators (1)-(3). For a simple but complete overview on the Fuzzy Sets Theory You might refer to the book of Klir & Folger (1989) for example.

Before illustrating a very simple application of Fuzzy Sets Theory, let me note that there exists some kind of connection between fuzzy sets and possibility theory. Indeed, if we consider a fuzzy set $\tilde{F}$ such that

$$\sup_x \tilde{F}(x) = 1$$

then

$$X \supset A \mapsto \sup_A \tilde{F}(x)$$

is a possibility measure, as we have noted in §2.1, Example 2.

Despite this common mathematical expression we have to note that the underlying concepts are very different. See Dubois & Prade (1980) pp. 136-137.

A very interesting chapter about Fuzzy Sets theory concerns the fuzzy logic and its applications, for example to control theory; however, as a major topic of TED is Decision Making I’ll illustrate a very simple application to this field. Unfortunately, I only have a basic knowledge of this topic so I apologize for the elementary level of what I’m going to say.
4.1 Fuzzy Decision Making.

Classical decision making generally deals with a set of alternatives, $X$, comprising the
decision space, a set of states of nature, $N$, comprising the state space, a relation,
$R : X \rightarrow N$, indicating the state or outcome to be expected from each alternative
action and, finally, a utility or objective function, $U : N \rightarrow \mathbb{R}^+$ which arranges these
outcomes according to an order of preferences.

The decision may involve the simple optimization of $U$, an optimization under con-
straints or an optimization given multiple-criteria.

A decision is said to be made under conditions of *certainty* when the outcome for each
action can be determined and ordered precisely. In this case, the alternative that leads
to the outcome having the highest utility is chosen. A decision is made under condi-
tions of *risk* when the only available knowledge concerning the outcome states is their
probability distributions. Again, this information can be used to optimize the utility
function. When the probabilities of the outcome states are unknown, decisions must
be made under conditions of *uncertainty*. In this case, fuzzy decision theories may be
used to accomodate this vagueness.

Bellman and Zadeh (1970) suggest a fuzzy model of decisions where both constraints
and goals are treated as fuzzy sets. We will show this model by a simple example.

**Example.**

Suppose we must choose one of four possible jobs $a$, $b$, $c$, and $d$, whose salaries (\$) are

\[
s(a) = 30,000, \quad s(b) = 25,000, \quad s(c) = 20,000, \quad s(d) = 15,000
\]

Our goal is to choose the job that will give us a high salary given the constraints
that the job is *interesting* and within close *driving distance*. The first step is that
representing the constraints and the goal by fuzzy sets. Let suppose that

\[
\begin{array}{c|cccc}
   x & a & b & c & d \\
   \tilde{C}_1(x) & 0.4 & 0.6 & 0.8 & 0.6 \\
\end{array}
\]

describes the interest of each job and
\[ \begin{array}{c|cccc} x & a & b & c & d \\ \hline \tilde{C}_2(x) & 0.1 & 0.7 & 0.9 & 1 \end{array} \]

its closeness. As concerns the fuzzy set corresponding to the concept of high salary, it might be more natural to assess the degree of compatibility \( \tilde{G}(x) \) by considering the space, say \( \mathbb{R}^+ \), of all the salaries, for example in this way:

\[
\tilde{G}(x) = \begin{cases} 
0 & 0 \leq x < 13,000 \\
-\frac{1}{800} \left( \frac{x}{1000} - 40 \right)^2 + 1 & 13,000 \leq x \leq 40,000 \\
1 & x > 40,000 
\end{cases}
\]

so the fuzzy set describing the compatibility of \( a, b, c, d \) with our concept of “high salary” comes out using the function \( s \):

\[
\tilde{G}^s(x) = \tilde{G}(s(x)) \Rightarrow
\]

\[ \begin{array}{c|cccc} x & a & b & c & d \\ \hline \tilde{C}^s(x) & 0.875 & 0.7 & 0.5 & 0.2 \end{array} \]

Summing up:

\[ \begin{array}{c|cccc} x & a & b & c & d \\ \hline \tilde{C}_1(x) & 0.4 & 0.6 & 0.8 & 0.6 \\
\tilde{C}_2(x) & 0.1 & 0.7 & 0.9 & 1 \\
\tilde{G}^s(x) & 0.875 & 0.7 & 0.5 & 0.2 \end{array} \]
Then, it is natural to choose the job which best satisfies both constraints and goal: the **fuzzy decision** is the fuzzy set obtained by intersection

\[
\hat{D}(x) = (\hat{C}_1 \cap \hat{C}_2 \cap \hat{C}')(x) = \min\{\hat{C}_1(x), \hat{C}_2(x), \hat{C}'(x)\}
\]

so in our example

<table>
<thead>
<tr>
<th>x</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>\hat{D}(x)</td>
<td>0.1</td>
<td>0.6</td>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

and the best job is \(b\). \(\Box\)

This method was applied in Brazil to choose the best vaccination strategy from 8 possibilities, see Massad & al. (1999). The authors maintain that the strategy adopted resulted in significant control of the epidemic.

Note that in this definition of fuzzy decision the two concepts, constraints and goal, are symmetric. In view of this symmetry we can consider the constraints as further goals. Indeed the decision \(\hat{D}\) of the previous example simply combines the different goals like high salary, interest and closeness. Next, we will consider this situation in slightly more detail.

### 4.1.1 Multi-criteria Decision Making.

Let’s consider \(n\) criteria and let’s \(\hat{A}_i(x)\) the fuzzy set indicating how much the alternative \(x\) satisfies the criteria \(i\). Then, a fuzzy decision is any fuzzy set combining the criteria

\[
\tilde{D}(x) = F(\hat{A}_1(x), \ldots, \hat{A}_n(x))
\]

such that

1. \(\tilde{A}_j(x) \geq \tilde{A}_j(y) \forall j \Rightarrow \tilde{D}(x) \geq \tilde{D}(y)\)

2. \(F(\tilde{A}_1(x), \ldots, \tilde{A}_n(x)) = F(\tilde{A}_{\sigma(1)}(x), \ldots, \tilde{A}_{\sigma(n)}(x))\) for all permutation \(\sigma\) of \(\{1, \ldots, n\}\)

35
that is, if \( x \) is always better than \( y \) then the decision supports \( x \) and \( F \) is symmetric.

For example,

\[
F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n)
\]
corresponds to the requirement that all criteria are satisfied, while

\[
F(a_1, \ldots, a_n) = \max(a_1, \ldots, a_n)
\]
corresponds to the requirement that at least one criterion is satisfied.

Yager (1988) introduces an important class of operators \( F \), the \textbf{ordered weighted averaging} operators, OWA for short. \( F \) is an OWA operator if there exist a vector

\[
(w_1, \ldots, w_n) \in [0, 1]^n \quad \text{with} \quad \sum_{i=1}^{n} w_i = 1
\]
such that

\[
F(a_1, \ldots, a_n) = w_1 a^1 + \ldots + w_n a^n
\]
where \((a^1, \ldots, a^n)\) is decreasing rearrangement of \((a_1, \ldots, a_n)\).

Note that both \( \min \) and \( \max \) are OWA operators:

\[
\min \iff w = (0, \ldots, 0, 1)
\]

\[
\max \iff w = (1, 0, \ldots, 0)
\]

and

\[
\min(a_1, \ldots, a_n) \leq w_1 a^1 + \ldots + w_n a^n \leq \max(a_1, \ldots, a_n)
\]

A way to fix the weights is to consider \( w_i \) as the increment in the satisfaction going from the situation in which \( i - 1 \) criteria are satisfied to the one where \( i \) criteria are satisfied, that is it is possible to obtain the weights starting from a \textbf{fuzzy quantifier} (regular and non-decreasing), see Yager (1994). A fuzzy quantifier is a fuzzy set

\[
\tilde{Q} : [0, 1] \to [0, 1]
\]
representing the degree of satisfaction on the basis of the proportion of satisfied criteria, so we require that:

\[ \tilde{Q}(0) = 0, \quad \tilde{Q}(1) = 1 \]

and

\[ r > t \Rightarrow \tilde{Q}(r) \geq \tilde{Q}(t). \]

For example, we can use the following fuzzy set in order to represent our degree of satisfaction when at least half of the criteria is satisfied:

\[ \tilde{Q} = \text{“at least half”} \]

Then, starting from a fuzzy quantifier, we define the weights \( w_i \) as

\[ w_i = Q \left( \frac{i}{n} \right) - Q \left( \frac{i - 1}{n} \right) \quad i = 1, \ldots, n \]

Using the quantifier “at least half” in our previous example we obtain:

\[ w_1 = 0, \quad w_2 = 0.74, \quad w_3 = 0.26 \]

and the best job with respect the corresponding OWA operator is \( c \):

<table>
<thead>
<tr>
<th>( x )</th>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{C}_1(x) )</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>( \tilde{C}_2(x) )</td>
<td>0.1</td>
<td>0.7</td>
<td>0.9</td>
<td>1</td>
</tr>
<tr>
<td>( \tilde{C}(x) )</td>
<td>0.875</td>
<td>0.7</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>( \tilde{C}'(x) )</td>
<td>0.322</td>
<td>0.67</td>
<td>0.722</td>
<td>0.496</td>
</tr>
</tbody>
</table>

37
Note that in this setting all the criteria are equally important, but often this is not the case. We can incorporate the importance of the criteria in this way. Let $I_j$ the importance associated to $\tilde{A}_j$. For any $x$ let’s consider the permutation $\sigma = \sigma(x)$ which arranges the vector $(\tilde{A}_1(x), \ldots, \tilde{A}_n(x))$ in decreasing order and use it to arrange the vector $(I_1, \ldots, I_n)$ too in the same sense: $\sigma I = (I_{\sigma(1)}, \ldots, I_{\sigma(n)})$. Then let’s define

$$w_{i, \sigma} = Q \left( \frac{1}{e} \sum_{j=1}^{i} I_{\sigma(j)} \right) - Q \left( \frac{1}{e} \sum_{j=1}^{i-1} I_{\sigma(j)} \right) \quad i = 1, \ldots, n$$

where $e = \sum_{j=1}^{n} I_j$. Note that now the weights depend on $x$.

For example, in the choice of the best job, if we consider the interest and the salary as equally important, say $I_1 = I_3 = 0.8$, while driving distance as less important, say $I_2 = 0.3$, we obtain

$$
\begin{array}{c|c|cccc}
I & x & a & b & c & d \\
\hline
0.8 & \tilde{C}_1(x) & 0.4 & 0.6 & 0.8 & 0.6 \\
0.3 & \tilde{C}_2(x) & 0.1 & 0.7 & 0.9 & 1 \\
0.8 & \tilde{C}^\ast(x) & 0.875 & 0.7 & 0.5 & 0.2 \\
\hline
\bar{D}(x) & & 0.40 & 0.66 & 0.68 & 0.45 \\
\end{array}
$$

and the best job now is $c$. □

I’ll conclude this example by noting that the $\tilde{A}_i(x)$ may represent the judgement of the expert $i$ about the action $x$, so we can use these very same arguments to model the consensus among experts called to identify a best action on the set $X$. With regard to this see, for example Bordogna, Fedrizzi and Pasì (1997).

38
4.2 Measure of Fuzziness.

For the sake of completeness I’ll conclude this lecture by introducing some ideas about the problem of measuring imprecision, that is, in the framework we are considering, about the problem of measuring how fuzzy a fuzzy set is. This question is today still an open question, because while it is clear what we want there is no single way to obtain it.

In order to qualify as a meaningful measure of fuzziness, \( f \) must satisfy certain natural requirements:

1. \( f(\tilde{A}) = 0 \) iff \( \tilde{A} \) is a crisp subset;

2. If \( \tilde{A} \) is “less fuzzy than” \( \tilde{B} \) then \( f(\tilde{A}) \leq f(\tilde{B}) \);

3. \( f \) assumes the maximum value iff \( \tilde{A} \) is “maximally fuzzy”.

Of course, for each particular conception of the degree of fuzziness, the expressions “less fuzzy than” and “maximally fuzzy” can have different meaning, so they can give rise to different measures. One of them, perhaps the best known, is based on the following concept: the fuzzy set maximally fuzzy corresponds to the constant function 0.5 and \( \tilde{A} \) is less fuzzy than \( \tilde{B} \) if

\[
\begin{align*}
\tilde{A}(x) & \geq \tilde{B}(x) \quad \text{if} \quad \tilde{B}(x) \geq 0.5 \\
\tilde{A}(x) & \leq \tilde{B}(x) \quad \text{if} \quad \tilde{B}(x) \leq 0.5
\end{align*}
\]

that is

![Diagram](image-url)
The corresponding function \( f \) is

\[
f(\hat{A}) = -\sum_x \hat{A}(x) \log_2 \hat{A}(x) + (1 - \hat{A}(x)) \log_2(1 - \hat{A}(x)).
\]

Different measures can be obtained by comparing a fuzzy set \( \hat{A}(x) \) and the corresponding crisp set \( A(x) \) defined as

\[
A(x) = \begin{cases} 
1, & \hat{A}(x) > 0.5 \\
0, & \hat{A}(x) \leq 0.5
\end{cases}
\]

by some distance. If \( X \) is finite we can consider, for example

\[
f(\hat{A}) = \sum_x |\hat{A}(x) - A(x)|
\]

\[
f(\hat{A}) = \sqrt{\sum_x (\hat{A}(x) - A(x))^2}
\]

\( \square \)
Appendix.

In the examples considered in §3.1 we have informally introduced the definition of refinement of a nested body of evidence. Now we will give a more general definition\textsuperscript{3} and some examples illustrating the non monotonicity of the Total Uncertainty and the Mean Total Uncertainty (see §3) with respect to any body of evidence.

Let $(\mathcal{F}, m)$, $(\mathcal{F}', m')$ bodies of evidence on the same space $(\Omega, \mathcal{S})$. We say that $(\mathcal{F}', m')$ is a refinement of $(\mathcal{F}, m)$ if for each focal element $E \in \mathcal{F}$ there exists a refinement of $E$ in $\mathcal{F}'$, that is a family $\mathcal{R}(E) = \{E'_1, \ldots, E'_r\} \subset \mathcal{F}'$ such that

- $\{E'_1, \ldots, E'_r\}$ is a partition of $E$ or
- $E'_1 \subset \ldots \subset E'_r \subset E$ and

\[
\sum_{B \in \mathcal{C}(E)} m(B) = \sum \{m'(E') \mid E' \in \bigcup_{B \in \mathcal{C}(E)} \mathcal{R}(E)\}
\]

where $\mathcal{C}(E) = \{B \in \mathcal{F} : \mathcal{R}(B) \cap \mathcal{R}(E) \neq \emptyset\}$. In other words, for an element $E' \in \mathcal{R}(E)$ the value $m'(E')$ depends on the basic assignment $m(E)$ for each $E$ involved in the refinement yielding $E'$ so that the total knowledge $E'$ CONSERVATA!! We note that if $E_1$ and $E_2$ are different focal elements on $\mathcal{F}$ we don’t require that the corresponding refinements will be disjoint. So, for example a refinement of

\[E_1 \subset E_2 \subset E_3 \subset E_4\]

\[\text{\textsuperscript{3}Bodini, 1995.}\]
Example 1. Let’s consider the body of evidence \((\mathcal{F}, m)\):

and its refinement \((\mathcal{F}', m')\):

By some algebraic calculations it easy to see that \(T(m) - T(m') = 0.375 - 0.125 \log_2 3 > 0\). On the other hand, if we consider:
\[ |E_1| = |E_2| = 32, \ |E_1 \cap A_2| = 16 \]
\[ m(E_1) = 0.5, \ m(E_2) = 0.5 \]

and its refinement:

\[ E'_1 = E_1 \setminus E_2, \ E'_3 = E_1 \cap E_2, \ E'_3 = E_2 \setminus E_1, \]
\[ m'(E'_1) = 0.25, \ m'(E'_2) = 0.25, \ m'(E'_3) = 0.5. \]

In this case we have \( T(m) - T(m') = 1.5 - \log_2 3 < 0. \)

**Example 2.** Let’s consider again the body of evidence \((\mathcal{F}, m)\) and its refinement \((\mathcal{F}', m')\) introduced in the first part of Example 1: we obtain \( M(m) - M(m') = -0.125 + -0.375 \log_2 3 < 0. \) On the other hand, if we consider a body of evidence \((\mathcal{F}, m)\) and its refinement \((\mathcal{F}', m')\)

\[ m(E_1) = 0.21 \quad E'_1 = E_1 \setminus E_2, \ E'_3 = E_1 \cap E_2, \ E'_3 = E_2 \Rightarrow \]
\[ m(E_2) = 0.75 \quad m'(E'_1) = 0.25, \ m'(E'_2) = 0.25, \ m'(E'_3) = 0.5. \]

we obtain \( M(m) - M(m') = 0.75 - 0.5 \log_2 3 < 0. \)
REFERENCES


Acknowledgements

Vorrei sinceramente ringraziare il prof. Carlo Bertoluzza, che mi ha introdotta a gran parte degli argomenti qui trattati (ed il cui materiale mi è stato molto utile) e che, soprattutto, è sempre per me un preciso punto di riferimento, nonché di incoraggiamento.

I would sincerely to thank prof. Carlo Bertoluzza who was the first to introduce me to most of the topics presented here. His notes have been very useful in preparing this work. His constant support even more.
# INDEX

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>1</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>2</td>
</tr>
<tr>
<td>2. A definition of uncertainty</td>
<td>3</td>
</tr>
<tr>
<td>2.1 Monotone measures of Sugeno</td>
<td>5</td>
</tr>
<tr>
<td>2.1.1 Possibility-Necessity measures</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Mathematical Theory of Evidence</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Imprecise Probabilities</td>
<td>14</td>
</tr>
<tr>
<td>2.3.1 Upper-Lower Probabilities</td>
<td>15</td>
</tr>
<tr>
<td>2.3.2 Choquet’s Capacities</td>
<td>17</td>
</tr>
<tr>
<td>2.3.3 Upper-Lower Previsions</td>
<td>20</td>
</tr>
<tr>
<td>3. Measures of uncertainty</td>
<td>23</td>
</tr>
<tr>
<td>3.1 Uncertainty Theory</td>
<td>24</td>
</tr>
<tr>
<td>4. Fuzzy sets Theory</td>
<td>30</td>
</tr>
<tr>
<td>4.1 Fuzzy Decision Making</td>
<td>33</td>
</tr>
<tr>
<td>4.1.1 Multi-Criteria Decision Making</td>
<td>35</td>
</tr>
<tr>
<td>4.2 Measures of Fuzziness</td>
<td>39</td>
</tr>
<tr>
<td>Appendix</td>
<td>41</td>
</tr>
<tr>
<td>References</td>
<td>44</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>48</td>
</tr>
</tbody>
</table>