Classical models of uncertainty

Workshop on belief functions

Thierry Denœux

April, 2016

Uncertainty

- This workshop is about the theory of belief functions, a formal framework for reasoning and making decisions under uncertainty.
- This framework originates from Arthur Dempster's seminal work on statistical inference with lower and upper probabilities. It was then further developed by Glenn Shafer, who showed that belief functions can be used as a general framework for representing and reasoning with uncertain information.
- The theory of belief functions, also referred to as Evidence theory or Dempster-Shafer theory, has been widely used in several areas such as Artificial Intelligence, Information Fusion and Risk Analysis. Recently, there has been a revived interested in its application to statistical inference.

This chapter

- In this introductory lecture, we will discuss the concept of uncertainty and review two popular formalisms for handling uncertainty: sets and probabilities.
- As we shall see, the theory of belief functions builds upon these two approaches: in a way, a belief function can be seen as the assignment of probabilities to sets.

Overview

Ubiquity of uncertainty

- Uncertainty is present in every area of human activity.
- Typically, we are interested in some question Q, such as:
  - What is the mean value of some variable in a population?
  - What will be the economic growth rate in the United States next year?
  - What was the amount of carbon dioxide emission in China in 2012?

  etc.

Notations

- In the following, we will denote by Ω the set of possible answers (one and only one is assumed to be true), and by ω the true answer.
- If we know the exact value of ω, this is a situation of complete certainty. If we know nothing at all (except that ω is in Ω), we have complete uncertainty.
- Actually, these two extreme situations are not frequent: usually, we have only partial knowledge of ω, based on limited evidence about the question of interest.
- The issue then arises of how to represent such partial information in such a way that it can be used for further reasoning, computation and rational decision making.

Aleatory vs. epistemic uncertainty

It has become customary in some areas (such as risk analysis) to distinguish between two main sources of uncertainty:

1. Randomness: the question of interest concerns some property of an object taken at random from a well-defined population (such as, e.g., the color of a ball to be drawn from an urn). We say that we have random, aleatory or physical uncertainty. Such uncertainty cannot be reduced because it depends on the physical property of the population and of the random experiment.
2. Lack of knowledge: for instance, the name of the next president of the US is unknown, but it is not random because there is no notion of random experiment. Such uncertainty is said to be epistemic. It can be reduced by acquiring further information related to the question of interest.

Classical models of uncertainty

Two classical models

1. Sets / propositional logic.
2. Probabilities.

Overview

Set-membership approaches

- Perhaps the simplest way of representing partial knowledge about some question is as a set $A \subseteq \Omega$ that certainly contains the true answer ω.
- There is a vast literature on set-membership approaches to uncertainty, with application, e.g., in computer science and automatic control.
- An important special case is interval arithmetics, which includes syntactic rules to compute with intervals, making it possible to produce rigorous enclosures of solutions to model equations.
- The fundamental operations on sets are simple forms of operations on the more complex representations that will be studied later.
- We will briefly review some of these operations, and show the link with propositional logic.

Overview

Conjunctive combination

- Assume that two sources provide two subsets A and B of Ω, assumed to contain the answer to the question of interest. How to combine these pieces of information?

Conjunctive combination

- Assume that two sources provide two subsets A and B of Ω, assumed to contain the answer to the question of interest. How to combine these pieces of information?
- If both sources can be trusted, then it is reasonable to consider that the true answer is in the intersection of A and B, denoted by $A \cap B$, which is the set containing the elements of Ω that belong to both A and B.
- This mode of fusing information is called conjunctive; it is reasonable when all information sources are assumed to be reliable.

Disjunctive combination

- However, when A and B are disjoint, i.e., $A \cap B = \emptyset$, this rule leads a contradiction. In that case, the assumption that the two sources can be trusted is no longer tenable.
- How to combine information in this case?

Disjunctive combination

- However, when A and B are disjoint, i.e., $A \cap B = \emptyset$, this rule leads a contradiction. In that case, the assumption that the two sources can be trusted is no longer tenable.
- How to combine information in this case?
- It is then more cautious to conclude that the true answer is in the union of A and B, denoted by $A \cup B$, which is the set containing the elements of Ω that belong to A or B.
- This is the simplest form of disjunctive rule for pooling information, which is suitable when at least one of the sources is assumed to be reliable.

Cartesian products, relations

- Let us now assume that we have two questions of interest, whose true answers are denoted by X and Y (X and Y may be called variables).
- Let Ω_X and Ω_Y be the sets of possible values for X and Y. To represent information about the values that X and Y may take jointly, we need to place ourselves in the Cartesian product $\Omega_X \times \Omega_Y$, denoted more concisely by Ω_{XY}, and defined as the set of ordered pairs (x, y) of an element of Ω_X and an element of Ω_y.
- A subset of R of Ω_{XY} is called a relation. It can be used to represent a constraint on the values that X and Y may take jointly.

Projection, cylindrical extension

- Let R be a relation on Ω_{XY}. The projection of R onto Ω_X, denoted by $R \downarrow \Omega_X$, is the subset of Ω_X defined by

$$R \downarrow \Omega_X = \{x \in \Omega_X | \exists y \in \Omega_Y, (x, y) \in R\}. \quad (1)$$

- Conversely, let A be a subset of Ω_X. Its cylindrical extension in Ω_{XY}, denoted by $A \uparrow \Omega_{XY}$, is the subset of Ω_{XY} defined as

$$A \uparrow \Omega_{XY} = A \times \Omega_Y = \{(x, y) \in \Omega_{XY} | x \in A\}. \quad (2)$$

Reasoning with relations

- To see how these notions can be used in a reasoning process, assume that we have
  - Evidence that X belongs to a subset A of Ω_X;
  - Evidence about the values that X and Y can take jointly, represented by a relation $R \subseteq \Omega_{XY}$.
- What can we deduce about Y? Let B denote the set of possible values for Y. It is clear that y belongs to B if and only of there is some x in A such that $(x, y) \in R$. Formally:

$$B = \{y \in \Omega_Y | \exists x \in A, (x, y) \in R\}, \quad (3)$$

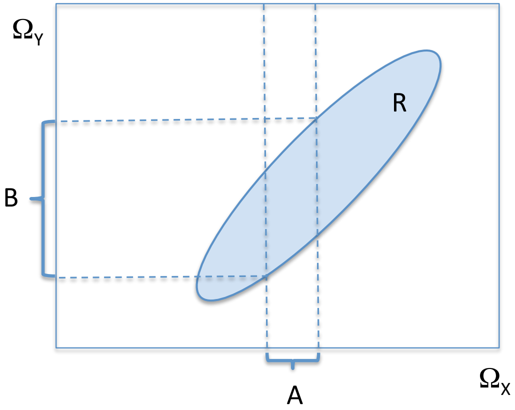
which can be written as

$$B = (R \cap (A \uparrow \Omega_{XY})) \downarrow \Omega_Y. \quad (4)$$

(see next slide)

- This kind of reasoning may straightforwardly be applied to any number of variables.

Reasoning with relations (continued)



Overview

Sources of uncertainty

Set-based representation
- Operations on sets
- Relationship with propositional logic
- Limitations of sets for representing uncertainty

Probabilistic representation
- Basic definitions
- Interpretations
- Cox's axioms
- Two paradoxes

Propositional variables, connectives

- Propositional logic is another formalism closely related to set theory. The basic constructs of that formalism are
  - propositional variables $p, q, r, \ldots$, which represent statements that can be true (T) or false (F), and
  - connectives $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$, which make it possible to build formulas expressing more complex propositions.
- The meaning of a connective is described by a truth table. For instance, the following table,

p	q	$p \rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

states that $p \rightarrow q$ is true if and only if p is false, or q is true.

Interpretations

- An interpretation is a mapping from the set of propositional variables to the set $\{T, F\}$ of truth values.
- To each formula ϕ corresponds the set $\mathcal{I}(\phi)$ of interpretations under which it is true. For instance, to $p \to q$ corresponds the set $\{(T, T), (F, T), (F, F)\}$.
- If ϕ and ψ are two formulas, then $\mathcal{I}(\phi \wedge \psi) = \mathcal{I}(\phi) \cap \mathcal{I}(\phi)$, $\mathcal{I}(\phi \vee \psi) = \mathcal{I}(\phi) \cup \mathcal{I}(\phi)$ and $\mathcal{I}(\neg\phi) = \overline{\mathcal{I}(\phi)}$, where $\overline{\mathcal{I}(\phi)}$ denotes the complement of $\mathcal{I}(\phi)$ in the set of all interpretations.
- Interpretations can be seen as representing states of the world, and a proposition can be identified to the set of states of the world under which it is true. Propositional logic and set theory thus have the same expressive power.

Overview

Limitations

- The main limitation of set-based representations of uncertainty (and propositional logic) is that they do not allow the expression of doubt.
- As a consequence, they favor a conservative approach, in which the sets have to be chosen very large to contain the true value with full certainty. A lot of information is usually lost in such a representation.
- For instance, if an expert is asked to give an interval that surely contains the mean sea level in 2050, he will give a wide interval, even though he may actually believe that the mean sea level will be contained within narrower bounds.
- As we will show later, belief functions can be seen as extending the notion of set by allowing one to provide different sets with attached degrees of support.

Overview

Probability theory

- Probability theory is another classical formalism for representing and reasoning with uncertainty.
- After recalling some basic definitions, we will provide a brief review of interpretations and justifications of this approach.

Overview

Finitely additive probabilities

- Let Ω be a set and $\mathcal{A} \subseteq 2^{\Omega}$ an algebra of subsets of Ω, defined as non-empty collection of subsets of Ω (called events), closed under complementation and finite union, i.e., for all A and B in $\mathcal{A}$, $A \cup B \in \mathcal{A}$. We can remark that Ω necessarily belongs to $\mathcal{A}$.
- A finitely additive probability measure on $(\Omega, \mathcal{A})$ is a function P from $\mathcal{A}$ to $[0, 1]$ such that
  1. $P(\Omega) = 1$;
  2. For all elements A and B of $\mathcal{A}$ such that $A \cap B = \emptyset$,

$$P(A \cup B) = P(A) + P(B). \quad (5)$$

Inclusion-exclusion formula

- We can easily deduce from (5) that, for any elements A and B of $\mathcal{A}$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{6}$$

- More generally, we can prove by induction that, for any $k \geq 2$ and any collection $A_1, \ldots, A_k$ of elements of $\mathcal{A}$,

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} P\left(\bigcap_{i \in I} A_i\right). \tag{7}$$

- As we will see later, a weaker form of this property characterizes belief functions.

Countably additive probabilities

- The notion of finitely additive probability is often extended to allow probabilities to be assigned to the union or intersection of countable families or events.
- For this, we need to consider a non-empty collection $\mathcal{A}$ of subsets of Ω that is closed under complementation and countable union. Such a family is called a σ-algebra.
- A countably additive probability measure on $(\Omega, \mathcal{A})$ is a function P from $\mathcal{A}$ to $[0, 1]$ such that $P(\Omega) = 1$ and, for all countable collections (A_i), $i = 1, \ldots, \infty$, of pairwise disjoint elements of $\mathcal{A}$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{8}$$

  The triple $(\Omega, \mathcal{A}, P)$ is called a probability space.
- The notions of finitely additive and countably additive probability measures differ only when the space Ω is infinite.

Overview

Interpretations of probabilities

- The mathematical model briefly described above may be used to represent different aspects of the real world.
- In particular, it can be used to represent
  - objective properties of random experiments, or
  - subjective degrees of belief.
- These two interpretations will now be briefly reviewed.

Objective probabilities

- Probability theory is clearly suitable to represent aleatory uncertainty, in which case the probability $P(A)$ for an event $A \subseteq \Omega$ is interpreted either
  - as a frequency (actually, the limit of the frequency with which event A occurs, if the random experiment is repeated n times and $n \to +\infty$), or
  - as a propensity (Popper), i.e., the tendency of A to happen across a large number of repetitions of the random experiment.
- Since frequencies are additive, the additivity axiom (8) is well justified.
- Such probabilities can be considered as objective, because they describe physical properties of the chance setup. For instance, when tossing a coin, the probabilities $P(\mathrm{Heads}) = P(\mathrm{Tails}) = 1/2$ can be deduced from the symmetry of the coin.

Subjective probabilities

- The use of probability measures to represent epistemic uncertainty (as advocated by the Bayesian school) is more problematic, because in this case probabilities can clearly no longer be interpreted as frequencies.
- In this context, they are usually interpreted as subjective (or personal) degrees of belief.
- However, we need to define more precisely the meaning of this notion and to explain why degrees of belief should be additive. This can be done in, at least, two ways: using a constructivist or a behavioral approach.

Constructivist approach

- In the constructivist approach, we construct a probability measure P by comparing our evidence (i.e., what we know) about Ω to a random experiment with known chances.
- This allows us a construct a scale of degrees of belief, with canonical examples.
- For instance, in a coin tossing game, the chance for Heads is $1/2$, which is taken as our degree of belief that Heads will come up. If our beliefs about the truth of some proposition A (e.g., "There is life on Mars") is comparable to our belief that Heads will come up when tossing a coin, we can say that our personal probability for A is $1/2$.

Behavioral approach

- In the behavioral approach, we assume that the belief state of an agent can be deduced from observing its betting behavior.
- The following Dutch book argument, first put forward by Ramsey and de Finetti, shows that consistent betting behavior, in some sense, should be based on probabilities.
- A Dutch book is a set of odds and bets which guarantees a profit, regardless of the outcome of the gamble.

Dutch book argument

- Assume that you have to enter a game where there is a player and a banker.
- The player gives an amount of money $\$p$ to the banker and the banker gives the player \$1 if a proposition A is true, and 0 otherwise.
- You do not know if you will be the banker or the player, and you are asked to fix p.
- By definition, your fair betting rate $P(A) = p$ is equated to your personal probability of proposition A. It is postulated to measure your belief in A: the more you believe in A, the more money you will be willing to give to enter the game.
- Now, the main point is that an opponent can compile a book of bets from your offer that assures a net gain from you (a Dutch book) if and only if P fails to be a probability function.

Dutch book argument – continued

- To show this, consider two disjoint events A and B and the three following bets:
  1. Bet 1: the player gains \$1 if A is true and 0 otherwise.
  2. Bet 2: the players gain \$1 if B is true and 0 otherwise.
  3. Bet 3: the players gain \$1 if $A \cup B$ is true and 0 otherwise.
- Let $P(A)$, $P(B)$ and $P(A \cup B)$ be the fair prices you are willing to pay for the three tickets. Assume that $P(A \cup B) < P(A) + P(B)$.
- Then, the opponent can raise a Dutch book against you by deciding that you will be the player in the first two bets and the banker in the third bet.
- The only way to avoid sure loss is to set the three numbers $P(A)$, $P(B)$ and $P(A \cup B)$ such that $P(A \cup B) = P(A) + P(B)$.

Dutch book argument – continued

- Proof: you will have to pay $P(A) + P(B)$ to participate in the first two bets as a player and you will receive $P(A \cup B)$ as a banker in the third bet. The balance is thus $-P(A) - P(B) + P(A \cup B) < 0$.
- Now, as shown in the table below, you will not win any additional money, whatever the outcome.

A	B	Bet 1	Bet 2	Bet 3
0	0	0	0	0
1	0	1	0	-1
0	1	0	1	-1

- The only way to avoid sure loss is to set the three numbers $P(A)$, $P(B)$ and $P(A \cup B)$ such that $P(A \cup B) = P(A) + P(B)$.

Criticisms of the Dutch book argument

- If we interpret degrees of belief as betting rates, it can thus be argued that degrees of belief should be (finitely) additive and our state of knowledge should be represented by a probability measure.
- However, this point of view is open to criticism:
  1. First, the betting scheme just described is a highly idealized situation, and it is debatable if any situation of choice under uncertainty can fit this idealized picture (probabilities and utilities do not exist, they are a construction).
  2. Secondly, by slightly changing the story, we can arrive at different conclusions. For instance, assume that you are not obliged to enter the game and you are not required to accept to be the banker. Let $P_*(A)$ be the highest price you are willing to pay for the lottery ticket. Then, a Dutch book can be raised against you iff P_* fails to be a lower probability function, i.e., the lower envelope of a family of probability measures (Williams, 1976).

Overview

Axiomatic justification of probabilities

- Some scholars have attempted to justify the use of probabilities to represent degrees of belief using an axiomatic approach.
- In particular, the axioms of Cox (1956) and Savage (1954) are often invoked by Bayesians to argue that probability theory is the only "reasonable" formalism for reasoning with uncertainty.
- here, we will briefly discuss Cox's axioms. Savage's axioms will be discussed later.

Cox's axioms

- Let $Cr(A|B) \in \mathbb{R}$ be a measure of the "credibility" of proposition A, given that B is true, where A and B are non-empty subsets of Ω. Consider the following axioms:
  1. The credibility of the complement of A can be computed from the credibility of A,
  $$Cr(\overline{A}|B) = S[Cr(A|B)], \tag{9}$$
  where S is a twice differentiable function;
  2. The credibility of $A \cap A'$ given B is a function of the credibility of A given B, and the credibility or A', given $A \cap B$,
  $$Cr(A \cap A'|B) = F[Cr(A|B), Cr(A'|A \cap B))], \tag{10}$$
  where F is a twice differentiable function with a continuous derivative.

Meaning of Axiom 2

- Example to illustrate Axiom 2: if A is the proposition that some athlete can run to some point, given the conditions of the race expressed by B, and if A' denotes the proposition that he can come back, then the probability that he can run to the point and come back depends on the probability that he can reach the point, and the probability that he can come back, given that he has already reached the point.

Cox's theorem

- Under these assumptions, Cox showed that Cr is isomorphic to a probability distribution, in the sense that there exists a one-to-one mapping $g : \mathbb{R} \to \mathbb{R}$ such that $g \circ Cr$ is a probability measure, and

$$g[Cr(A|B)] \cdot g[Cr(B)] = g[Cr(A \cap B)] \tag{11}$$

for any A and non-empty B, with $Cr(B) = Cr(B|\Omega)$.

Criticisms of Cox axioms

- Significant as it may be, this result can hardly be considered as a final justification of probabilities for representing degrees of belief. Indeed, close inspection of the axioms shows that they can be seriously questioned.
- The first assumption is that the credibility of a proposition can be represented by a single number. This condition is not assumed in some alternative theories of uncertainty, such as the theory of belief functions.
- Axiom 1 is also debatable. If degrees of credibility are identified with degrees of support, the degree of support for some proposition is not a function of the degree of support for its negation (if A is not supported, $\overline{A}$ may be supported or not), and $Cr(A|\Omega)$ will not be determined by $Cr(\overline{A}|\Omega)$.
- Even admitting that $Cr(A \cap A'|B)$ should be a function of $Cr(A|B)$ and $Cr(A'|A \cap B)$, it is not obvious that the same function F should always be used.

Overview

Arguments against the exclusive use of probabilities

- As shown in the previous section, attempts to justify the use of probabilities to represent degrees of belief have not settled the question.
- In contrast, there appears to be some serious arguments against the use of probability theory as a model of epistemic uncertainty (Bayesian model).
- In particular,
  1. the use of a probability distribution to represent ignorance may lead to some inconsistencies, and
  2. probability theory does not seem to be a plausible model of how people make decisions based on weak information.
- These arguments are exemplified by the following two paradoxes.

The wine/water paradox

- Principle of Indifference (PI): in the absence of information about some quantity X, we should assign equal probability to any possible value of X
- The wine/water paradox

There is a certain quantity of liquids. All that we know about the liquid is that it is composed entirely of wine and water, and the ratio of wine to water is between 1/3 and 3.
What is the probability that the ratio of wine to water is less than or equal to 2?

The wine/water paradox – continued

- Let X denote the ratio of wine to water. All we know is that $X \in [1/3, 3]$. According to the PI, $X \sim \mathcal{U}_{[1/3,3]}$. Consequently

$$P(X \leq 2) = (2 - 1/3)/(3 - 1/3) = 5/8$$

- Now, let $Y = 1/X$ denote the ratio of water to wine. All we know is that $Y \in [1/3, 3]$. According to the PI, $Y \sim \mathcal{U}_{[1/3,3]}$. Consequently

$$P(Y \geq 1/2) = (3 - 1/2)/(3 - 1/3) = 15/16$$

- However, $P(X \leq 2) = P(Y \geq 1/2)$!

Ellsberg's paradox

- Suppose you have an urn containing 30 red balls and 60 balls, either black or yellow. You are given a choice between two gambles:
  - A: You receive 100 euros if you draw a red ball
  - B: You receive 100 euros if you draw a black ball

  Which one do you choose?

Ellsberg's paradox

- Suppose you have an urn containing 30 red balls and 60 balls, either black or yellow. You are given a choice between two gambles:
  - A: You receive 100 euros if you draw a red ball
  - B: You receive 100 euros if you draw a black ball

  Which one do you choose?
- Also, you are given a choice between these two gambles (about a different draw from the same urn):
  - C: You receive 100 euros if you draw a red or yellow ball
  - D: You receive 100 euros if you draw a black or yellow ball

  Which one do you choose?

Ellsberg's paradox

- Suppose you have an urn containing 30 red balls and 60 balls, either black or yellow. You are given a choice between two gambles:
  - A: You receive 100 euros if you draw a red ball
  - B: You receive 100 euros if you draw a black ball

  Which one do you choose?
- Also, you are given a choice between these two gambles (about a different draw from the same urn):
  - C: You receive 100 euros if you draw a red or yellow ball
  - D: You receive 100 euros if you draw a black or yellow ball

  Which one do you choose?
- Most people strictly prefer A to B, hence $P(red) > P(black)$, but they strictly prefer D to C, hence $P(black) > P(red)$

Conclusions

- The two main formalisms for representing uncertain information are based on sets and on probabilities, respectively.
- We have shown in this lecture that none of these two formalisms seems to be sufficient to represent all kinds of uncertainties.
- In the next lecture, we will introduce the theory belief functions, which can be seen as generalizing the two classical frameworks outlined above.