UNCERTAINTY THEORIES: A UNIFIED VIEW

From imprecise probability to possibility theory

D. Dubois

IRIT-CNRS, Université Paul Sabatier
31062 TOULOUSE FRANCE

Work supported by the ANR project CRISCO2,
in cooperation with BRGM, Orléans

Outline

1. Variability vs incomplete information
2. Blending set-valued and probabilistic representations : uncertainty theories
3. Possibility theory in the landscape
4. Practical representations of incomplete probabilistic information
5. Contribution to risk analysis.

Origins of uncertainty

- The variability of observed repeatable natural phenomena : « **randomness ».**
  - Coins, dice…: what about the outcome of the next throw?
- The lack of information: **incompleteness**
  - because of information is often lacking, knowledge about issues of interest is generally not perfect.
- Conflicting testimonies or reports:**inconsistency**
  - The more sources, the more likely the inconsistency

Example

- **Variability**: daily quantity of rain in Belfast
  - May change every day
  - It can be estimated through statistical observed data.
  - Beliefs or prediction based on this data
- **Incomplete information** : Birth date of French President
  - It is not a variable: it is a constant!
  - You can get the correct info somewhere, but it is not available.
  - Most people may have a rough idea (an interval), a few know precisely, some have no idea: information is subjective.
  - Statistics on birth dates of other presidents do not help much.
- **Inconsistent information :** several sources of information conflict concerning the birth date (a book, a friend, a website).

The roles of probability

Probability theory is generally used for representing two aspects:

1. **Variability:** capturing (beliefs induced by) variability through repeated observations.
2. **Incompleteness (info gaps):** directly modeling beliefs via betting behavior observation.

These two situations are not mutually exclusive.

Using a single probability distribution to represent incomplete information is not entirely satisfactory:

The betting behavior setting of Bayesian subjective probability enforces a representation of partial ignorance based on single probability distributions.

1. **Ambiguity :** In the absence of information, how can a uniform distribution tell pure randomness and ignorance apart ?
2. **Instability :** A uniform prior on $x \in [a, b]$ induces a non-uniform prior on $f(x) \in [f(a), f(b)]$ if f is increasing and non-affine.
3. **Empirical doubt**: When information is missing, decision-makers do not always choose according to a single subjective probability (Ellsberg paradox).

Motivation for going beyond probability

- Have a language that distinguishes between uncertainty due to variability from uncertainty due to lack of knowledge or missing information.
- **The main tools to representing uncertainty are**
  - **Probability distributions :** good for expressing variability, but information demanding, and paradoxical for ignorance
  - **Sets**: good for representing incomplete information, but a very crude representation of uncertainty
- *Find representations that allow for both aspects of uncertainty*.

Set-Valued Representations of Partial Information

- An ill-known quantity x is represented as a *disjunctive* set, i.e. a subset E of *mutually exclusive* values, one of which is the real one.
- Pieces of information of the form « *all I know is that $x \in E$* »
  - **Intervals** E = [a, b]: good for representing incomplete numerical information
  - **Classical Logic**: good for representing incomplete symbolic (Boolean) information

    E = Models of a wff ϕ stated as true.

  but poorly expressive
- ***Such sets are as subjective as probabilities***

BOOLEAN POSSIBILITY THEORY

*If all **you** know is that $x \in E$ then*

- You judge **event A possible** if it is logically consistent with what you know : $A \cap E \neq \emptyset$

A Boolean possibility function : $\Pi(A) = 1$, and 0 otherwise

- **You believe event A (sure)** if it is a logical consequence of what we already know : $E \subseteq A$

A certainty (necessity) function : $N(A) = 1$, and 0 otherwise

*- This is a simple modal **epistemic** logic (KD45)*

$$N(A) = 1 - \Pi(A^c) \leq \Pi(A)$$

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)); \; N(A \cap B) = \min(N(A), N(B)).$$

WHY TWO SET-FUNCTIONS ?

- **Encoding 3 extreme epistemic states….**
  - Certainty of truth : $N(A) = 1$ (hence $\Pi(A) = 1$)
  - Certainty of falsity: $\Pi(A) = 0$ (hence $N(A) = 0$)
  - Ignorance : $\Pi(A) = 1$, $N(A) = 0$

….. **requires 2 Boolean variables!**

The Boolean counterpart of a subjective probability

With one function you can only say believe A or believe not-A.

Find an extended representation of uncertainty

- *Explicitly allowing for missing information (= that uses sets)*
- *More informative than pure intervals or classical logic,*
- *Less demanding and more expressive than single probability distributions*
- *Allows for addressing the issues dealt with by both standard probability, and logics for reasoning about knowledge.*

From sets to gradual possibility distributions

- *What about the birth date of the president?*
- **partial ignorance with ordinal preferences :** May have reasons to believe that 1933 > 1932 ≡ 1934 > 1931 ≡ 1935 > 1930 > 1936 > 1929
- **Linguistic information** described by fuzzy sets: " he is old " : membership μ_{OLD} induces a possibility distribution on possible birth dates.
- **Nested confidence intervals:**

$x \in [a_i, b_i]$, with certainty c_i

such that for the expert, $P(x \in [a_i, b_i]) \geq c_i$

Blending intervals and probability

- Representations that may account for variability, incomplete information, and belief must combine probability and sets.
  - Sets of probabilities : imprecise probability theory
  - Random(ised) sets : Dempster-Shafer theory
  - Fuzzy sets: numerical possibility theory
- **Relaxing the probability axioms** :
  - Each event has a degree of certainty and a degree of plausibility, instead of a single degree of probability
  - When plausibility = certainty, it yields probability

A GENERAL SETTING FOR REPRESENTING GRADED CERTAINTY AND PLAUSIBILITY

- 2 set-functions Pl and Cr, with values in [0, 1], generalizing probability, possibility and necessity.
- **Conventions** :
  - $Pl(A) = 0$ "impossible" ;
  - $Cr(A) = 1$ "certain"
  - $Pl(A) = 1$; $Cr(A) = 0$ "ignorance" **(no information)**
  - $Pl(A) - Cr(A)$ quantifies ignorance about A
- **Postulates**
  - If $A \subseteq B$ then $Cr(A) \leq Cr(B)$ and $Pl(A) \leq Pl(B)$
  - $Cr(A) \leq Pl(A)$ "certain implies plausible"
  - $Pl(A) = 1 - Cr(A^c)$ duality certain/plausible

Imprecise probability theory

- A state of information is represented by a family $\mathcal{P}$ of probability distributions over a set X.
- To each event A is attached a probability interval $[P_*(A), P^*(A)]$ such that
  - $P_*(A) = \inf\{P(A), P \in \mathcal{P}\}$
  - $P^*(A) = \sup\{P(A), P \in \mathcal{P}\} = 1 - P_*(A^c)$
- $\{P(A), P \geq P_*\}$ is convex
- Usually $\mathcal{P}$ is strictly contained in $\{P(A), P \geq P_*\}$

Subjectivist view (Peter Walley)

- $P_{low}(A)$ is the highest acceptable price for buying a bet on event A winning 1 euro if A occurs
- $P^{high}(A) = 1 - P_{low}(A^c)$ is the least acceptable price for selling this bet and $P^{high}(A) \geq P_{low}(A)$
- Two rationality conditions:
  - **No sure loss:** $\{P(A), P \geq P_{low}\} \neq \emptyset$
  - **Coherence** condition

$$P_*(A) = \inf\{P(A), P \geq P_{low}\} = P_{low}(A)$$

- *A theory that handles convex probability sets :*
- Convex probability sets are usually characterized by lower expectations of real-valued functions (gambles), not just events.

Random sets and evidence theory

- A family $\mathcal{F}$ of « focal » (disjunctive) non-empty sets representing
  - A collection of incomplete observations (imprecise statistics).
  - Unreliable testimonies
  - Indirect information (induced by an incomplete mapping from a probability space)
- A positive weighting of focal sets (a random set) :

$$\sum_{E \in \mathcal{F}} m(E) = 1 \quad (\textit{mass function})$$

- It is a randomized incomplete information

Theory of evidence

- m(E) = probability that the most precise description of the available information is of the form "$x \in E$ "
- m(E) is attached to the event "getting the piece of evidence "$x \in E$ ", not to the event E

  = probability (only knowing "$x \in E$" and nothing else)

  – It is the portion of probability mass hanging over elements of E without being allocated.
- **DO NOT MIX UP m(E) (de dicto)**

  and P(E) (de re)

Theory of evidence

- **degree of certainty** (belief) :
  - $Bel(A) = \sum_{E_i \subseteq A,\ E_i \neq \emptyset} m(E_i)$
  - total mass of information <u>implying</u> the occurrence of A
  - *(probability of provability)*
- **degree of plausibility** :
  - $Pl(A) = \sum_{E_i \cap A \neq \emptyset} m(E_i) = 1 - Bel(A^c) \geq Bel(A)$
  - total mass of information <u>consistent with</u> A
  - *(probability of consistency)*

Theory of evidence vs. imprecise probabilities

- The set $\mathcal{P}_{\text{bel}} = \{P \geq \text{Bel}\}$ is coherent: Bel is a special case of lower probability
- Bel is ∞-monotone (super-additive at any order)
- The solution m to the set of equations $\forall\, A \subseteq X$

$$g(A) = \sum_{E_i \subseteq A} m(E_i)$$

is unique (Moebius transform)

– ***However it is positive iff g is a belief function***

Possibility Theory

(Shackle, 1961, Lewis, 1973, Zadeh, 1978)

- A piece of incomplete information "$x \in E$" admits of *degrees* of possibility.
- *E is mathematically a (normalized) fuzzy set.*
- $\mu_E(s)$ = Possibility(x = s) = $\pi_x(s)$
- **Conventions:**

  $\forall s$, $\pi_x(s)$ is the degree of plausibility of x = s

  $\pi_x(s) = 0$ iff x = s is impossible, totally surprising

  $\pi_x(s) = 1$ iff x = s is normal, fully plausible, unsurprising (but no certainty)

POSSIBILITY AND NECESSITY OF AN EVENT

How confident are we that $x \in A \subset S$? (*an event A occurs*)

given a possibility distribution π for x on S

- $\Pi(A) = \max_{s \in A} \pi(s)$:

  to what extent A is consistent with π

  (= some $x \in A$ is possible)

  The degree of possibility *that* $x \in A$

- $N(A) = 1 - \Pi(A^c) = \min_{s \notin A} 1 - \pi(s)$:

  to what extent no element outside A is possible

  = to what extent π implies A

  The degree of certainty (necessity) that $x \in A$

Basic properties

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B));$$

$$N(A \cap B) = \min(N(A), N(B)).$$

Mind that most of the time :

$$\Pi(A \cap B) < \min(\Pi(A), \Pi(B));$$

$$N(A \cup B) > \max(N(A), N(B)$$

Example: Total ignorance on A and B = A^c

Corollary $N(A) > 0 \Rightarrow \Pi(A) = 1$

Qualitative vs. quantitative possibility theories

- **Qualitative**:
  - **comparative**: A complete pre-ordering $\geq_\pi$ on U
    A well-ordered partition of U: E1 > E2 > … > En
  - **absolute:** $\pi_x(s) \in L$ = finite chain, complete lattice...
- **Quantitative**: $\pi_x(s) \in [0, 1]$, integers...

One must indicate where the numbers come from.

All theories agree on the fundamental maxitivity axiom

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$$

Theories diverge on the conditioning operation

POSSIBILITY AS UPPER PROBABILITY

- Given a numerical possibility distribution π, define
  $\boldsymbol{P}(\pi) = \{P \mid P(A) \leq \Pi(A) \text{ for all } A\}$

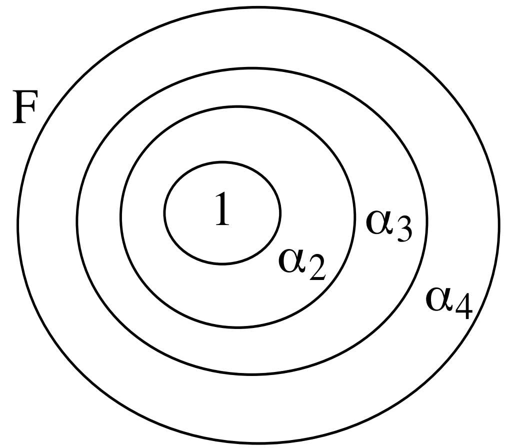
- Then, generally coherence holds:

$$\Pi(A) = \sup \{P(A) \mid P \in \boldsymbol{P}(\pi)\};$$

$$N(A) = \inf \{P(A) \mid P \in \boldsymbol{P}(\pi)\}$$

- So π is a faithful representation of a special family of probability measures

Random set view



possibility levels
$1 > \alpha_2 > \alpha_3 > \ldots > \alpha_n$

- *A basic probability assignment :*

  Let $m_i = \alpha_i - \alpha_{i+1}$ then $m_1 + \ldots + m_n = 1$, with focal sets = cuts
- $\pi(s) = \sum_{i:\ s \in Fi} m_i = Pl(\{s\})$.
- π is a one point-coverage function, or the contour function.
- *$Bel(A) = \sum_{Fi \subseteq A} m_i = N(A)$; $Pl(A) = \Pi(A)$*
- *Only in the consonant case can m be recalculated from π*

How to build possibility distributions

(not related to linguistic fuzzy sets!!!)

- *Nested* random sets (= *consonant belief functions*)
- *Likelihood functions* (in the absence of priors).
- *Probabilistic inequalities* (Chebyshev…)
- *Confidence intervals* (moving the confidence level between 0 and 1)
- *The cumulative PDF* of P **is** a possibility distribution (accounting for all probabilities stochastically dominated by P)

LANDSCAPE OF UNCERTAINTY THEORIES

BAYESIAN/STATISTICAL PROBABILITY

Randomized points

(extreme probabilities)

UPPER-LOWER PROBABILITIES

Disjunctive sets of probabilities

KAPPA FUNCTIONS (SPOHN)

DEMPSTER UPPER-LOWER PROBABILITIES

SHAFER-SMETS **BELIEF FUNCTIONS**

Random disjunctive sets

PLAUSIBIBILITY RANKING

Quantitative Possibility theory

Fuzzy (nested disjunctive) sets

Classical logic

Disjunctive sets

Practical representations of probability sets

1. Fuzzy intervals (possibility theory)
2. Probability intervals (restricting the probabilities of elementary events)
3. Probability boxes : pairs of PDF's
4. Generalized p-boxes : pairs of comonotonic possibility distributions (generalize 1 and 3)
5. Clouds (Neumaier): pairs of possibility distributions (generalize 4)

Some are special random sets other not.

(2: 2-monotone, 5 not even so)

From confidence sets to possibility distributions

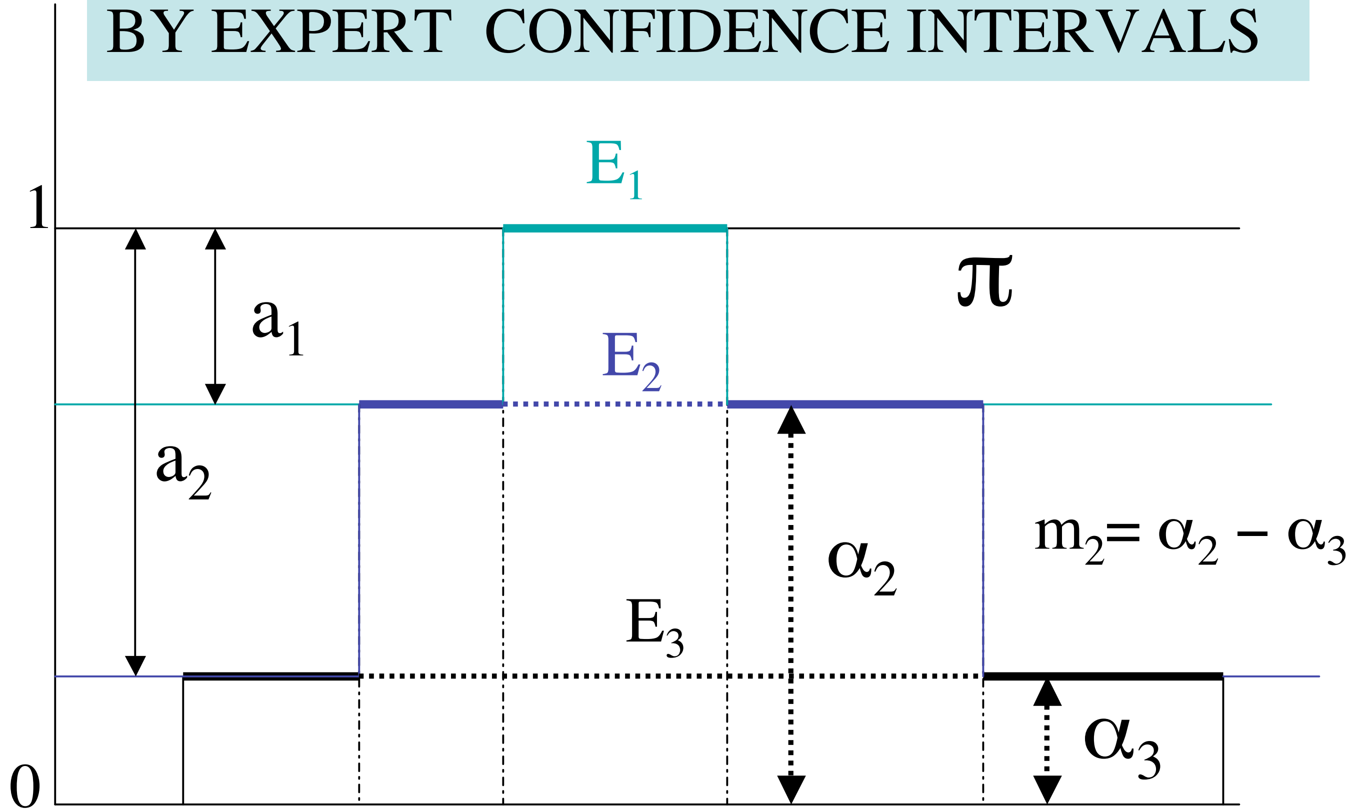
- Let $E_1, E_2, \ldots E_n$ be a nested family of sets
- A set of confidence levels $a_1, a_2, \ldots a_n$ in [0, 1]
- Consider the set of probabilities

  $$\mathcal{P} = \{P, P(E_i) \geq a_i, \text{ for } i = 1, \ldots n\}$$

- Then $\mathcal{P}$ is representable by means of a possibility measure with distribution

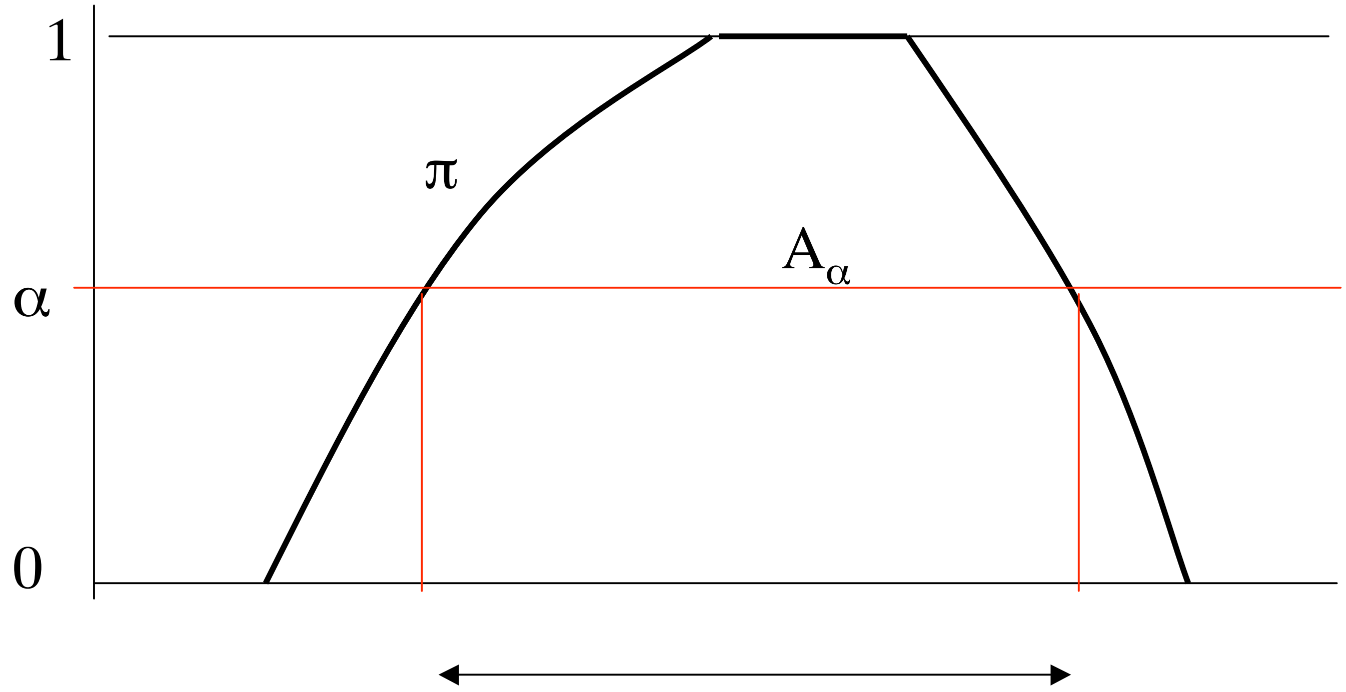
$$\pi(x) = \min_{i = 1, \ldots n} \max (\mu_{E_i}(x), 1 - a_i)$$

POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS



A possibility distribution can be obtained from any family of nested confidence sets :

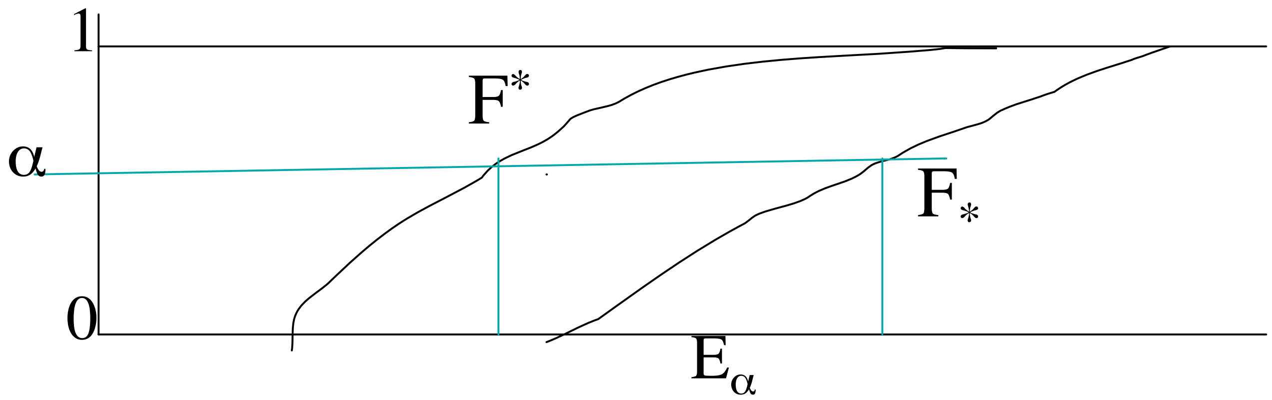
$P(A_\alpha) \geq 1 - \alpha, \alpha \in (0, 1]$



FUZZY INTERVAL: $N(A_\alpha) = 1 - \alpha$

Probability boxes

- A set $\mathcal{P} = \{P: F^* \geq P \geq F_*\}$ induced by two cumulative disribution functions is called a **probability box (p-box)**,
- **A p-box is a special random interval whose upper and bounds induce the same ordering.**
- **A continuous belief function….**



Possibility distributions vs. probability boxes

- A fuzzy interval M with mode m induces
  - An **upper distribution function** $F^*$is the increasing side of M:

    $\forall a, F^*(a) = \Pi_M((-\infty, a]) = M(a)$ if $a \leq m$ (= 1 otherwise).
  - A **lower distribution function** F_* (the decreasing side of M upside down):

    $\forall a, F_*(a) = N_M((-\infty, a]) = 1 - M(a)$ if $a > m$ (= 0 otherwise).
- Consider the p-box $\mathcal{P} = \{P: F^* \geq P \geq F_*\}$.
- **Claim:** $\mathcal{P}(\pi)$ is a proper subset of $\mathcal{P}$
  - Not all P in $\{P: F^* \geq P \geq F_*\}$ are such that $\Pi \geq P$
  - *Representing families of probabilities by fuzzy intervals is more informative than with the corresponding pairs of PDFs:*

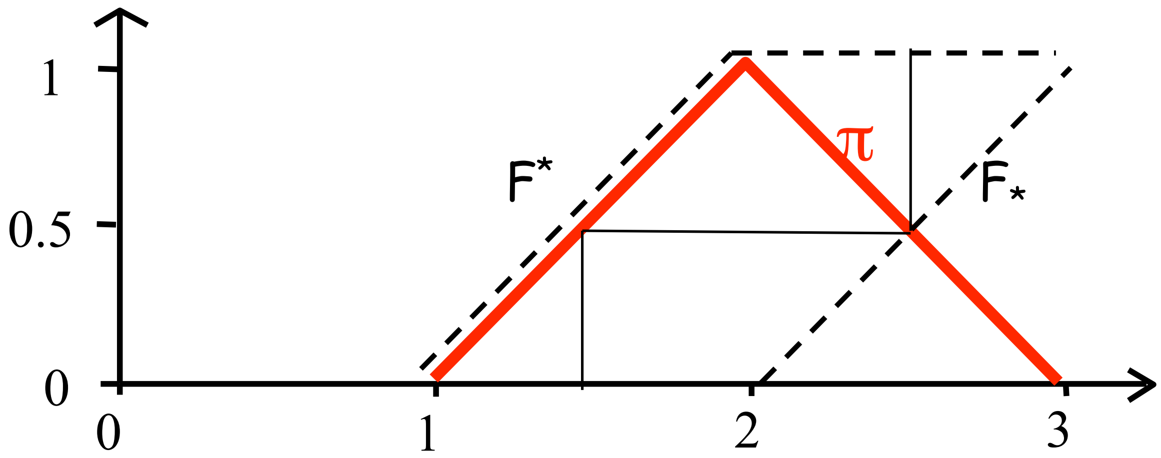
Counter-example:

A triangular fuzzy number with support [1, 3] and mode 2.

Let P be defined by P({1.5})=P({2.5})=0.5.

Then $F_* < F < F$ and $P \notin P(\pi)$ since

$P(\{1.5, 2.5\}) = 1 > \Pi(\{1.5, 2.5\}) = 0.5$



How useful are these practical representations:

- **Cutting complexity:**
  - Convex sets of probability are very complex representations
  - Random sets are potentially exponential
  - P-boxes, possibility distributions and other extensions are linear, but still encode convex probability set, often random sets.
- **Enriching the standard probability analysis** with meta-information and capabilities for reasoning about knowledge in the risk analysis process, while remaining tractable on modern computers.

Knowledge vs. evidence

- There are three kinds of information an agent can possess :
  - Generic knowledge
  - Singular evidence
  - Beliefs
- **Generic knowledge** pertains to a population of items, repeatable observables, …
- **Singular evidence** (observations) pertains to a single situation
- *Belief pertains to singular (not observed) events,*
  - *Either induced by statistical knowledge (Kyburg) based on some evidence on the case at hand.*
  - *Or directly assessed (betting interpretation, De Finetti)*
  - *Or defined by analogy (urn, bag of balls, Lindley)*

GENERIC vs. SINGULAR INFORMATION

- BACKGROUND KNOWLEDGE refers to a <u>class of situations</u> and summarizes a set of trends
  - Laws of physics,
  - Commonsense knowledge (birds fly)
  - Professional knowledge (of medical doctor),
  - Statistical knowledge
- PIECES OF EVIDENCE refer to a particular situation (testimonies) and are singular.
  - Measurement data
  - E.g. results of medical tests on a patient
  - Testimonies
  - Observations about the current state of facts

- **Generic knowledge may be tainted with exceptions, incompleteness, variability**
  - *It is not absolutely true knowledge in the mathematical sense*: *tainted with exceptions,*
  - It all comes down to considering some propositions are generally more often the case than other ones.
  - *Generic knowledge induces a normality or plausibility relation on the states of the world.*
  - *numerical* (frequentist) or *ordinal* (plausibility ranking)
  - ***In the numerical case a credal set can account for incomplete generic knowledge***
- **Observed evidence** is often made of propositions known as true about the current world.
  - It is often incomplete and can be encoded as disjunctive sets, or wff in propositional logic.
  - It delimits a reference class of situations for the case under study.
  - It can be uncertain unreliable, (subjective probability, Shafer)
  - It can be irrelevant, wrong,

MODELLING GENERIC KNOWLEDGE, EVIDENCE, BELIEFS

1. **Generic information** ***(background knowledge)*** : it is modelled by a rule **base, a set of default rules, a set of conditional probabilities, a Bayesian network, a credal net work**
2. **Singular information on the current situation (*evidence*)** *Known facts (results of observations, tests, sensor measurement, testimonies) modeled by propositions in propositional logic, or variable instantiations.*
3. **Beliefs about the current situation are predictions in the form of propositions** derived from known facts and generic information, along with a degree of confidence

PLAUSIBLE REASONING

- Inferring **beliefs** (plausible conclusions) on the current situation from observed evidence, using generic knowledge
  - **Example : medical diagnosis**

  Medical knowledge + test results $\Rightarrow$ believed disease of the patient.
- *This mode of inference makes sense regardless of the representation*, but pure set-based representations are insufficient:
  - in a purely propositional setting, one cannot tell generic knowledge from singular evidence
  - in the first order logic setting there is no exception.
  - *Need more expressive settings for representing background knowledge, like non-monotonic reasoning, probability or credal sets*
- **The basic tool for exception-tolerant inference is conditioning** (not well-known in classical logic).

The belief construction problem

- *Beliefs of an agent about a situation are inferred from generic knowledge AND observed singular evidence about the case at hand.*
- *They are non-monotonically derived and can be questioned by new evidence.*
- **Example: Commonsense plausible inference**
  - Generic knowledge = birds fly, penguin are birds, penguins don't fly.
  - Singular observed fact = Tweety is a bird
  - Inferred belief = Tweety flies
  - Additional evidence = Tweety is a penguin
  - Inferred revised belief = Tweety does not fly

Belief construction

- *Beliefs of an agent about a situation are derived from generic knowledge and evidence about the case.*
- **Probabilistic beliefs**: Hacking principle again
  - Uncertain singular fact = a set C = what is known about the context of the current situation.
  - Generic knowledge = probability distribution P reflecting the trends in a population (of experiments) relevant to the current situation
  - A querying problem: is an uncertain proposition A true in the current situation?
  - $Bel_C(A) = P(A|C)$: equating belief and frequency
- **Assumption**: the current situation is typical of situations where C is true

Conditional Probability

- **Two concepts leading to 2 definitions**:

1. <u>derived</u> (Kolmogorov): $P(A \mid C) = \dfrac{P(A \cap C)}{P(C)}$

   requires $P(C) \neq 0$

2. <u>primitive</u> (de Finetti): $P(A|C)$ is directly assigned a value and P is derived such that

   $P(A \cap C) = P(A|C) \cdot P(C)$.

   - Makes sense even is $P(C) = 0$

Meaning : $P(A \mid C)$ is the probability of A

*if C represents **all** that is hypothetically known on the situation*

THE MEANING OF CONDITIONAL PROBABILITY

- P(A|C) : the probability of a conditional event « A in epistemic context C » (when C is all that is known about the situation).
- *It is the probability of A* ***knowing only C****, NOT the probability of A if C is true*.
- **Counter-example** :
  - Uniform Probability on {1, 2, 3, 4, 5}
  - P(Even |{1, 2, 3}) = P(Even |{3, 4, 5}) = 1/3
  - Under a classical logic interpretation :
    - From « if result ∈ {1, 2, 3} then P(Even) = 1/3 »
    - And« if result ∈ {3, 4, 5} then P(Even) = 1/3 »
    - Then (classical inference) : P(Even) = 1/3 unconditionally!!!!!
  - **But of course: P(Even) = 2/5.**
- So, conditional events A|C should be studied as single entities (De Finetti).

The nature of conditional probability

- In the frequentist settting a conditional probability $P(A|C)$ is a relative frequency.
- It can be used to represent the weight of rules of the form « generally, if C then A » understood as « Most C's are A's » with exceptions

In logic a rule « if C then A » is represented by material implication $C^c \cup A$ that rules out exceptions

- *But the probability of a material conditional is not a conditional probability!*
- *What is the entity* A|C *whose probability is a conditional probability???*

A conditional event!!!!

Material implication: the raven paradox

- Testing the rule « all ravens are black » viewed as $\forall x, \neg Raven(x) \vee Black(x)$
- Confirming the rule by finding situations where the rule is true.
  - Seeing a black raven confirms the rule
  - Seeing a white swan also confirms the rule.
  - But only the former is an example of the rule.

3-Valued Semantics of conditionals

- A rule « if C then A » shares the world into 3 parts
  - **Examples:** interpretations where $A \cap C$ is true
  - **Counterexamples:** interpretations where $A^c \cap C$ is true
  - **Irrelevant cases:** interpretations where C is false

Rules « all ravens are black » and « all non-black birds are not ravens » have the same exceptions (white ravens), but different examples (black ravens and white swans resp.)

- <u>Truth-table of « A|C » viewed as a connective</u>
  - Truth(A|C) = T if truth(A)= truth(C) = T
  - Truth(A|C) = F if truth(A)=T and truth(C) = F
  - Truth(A|C) = I if truth(C)= F

  Where I is a 3d truth value expressing « irrelevance »:

  I = T: $A \cup C^c$; I = F: $A \cap C$.

A conditional event is a pair of nested sets

- The solutions X of $A \cap C = X \cap C$ form the set
$$A|C = \{X: A \cap C \subseteq X \subseteq A \cup C^c\}$$
- It defines the symbolic Bayes-like equation:
$$A \cap C = (A|C) \cap C.$$
- The models of a conditional $A|C$ can be represented by the pair $(A \cap C, A \cup C^c)$, an interval in the Boolean algebra of subsets of S
- The set $A \cup C^c$ representing material implication contains the « non-exceptions » to the rule (the complement of $A \cap C^c$).

Semantics for three-valued logic of conditional events.

- Semantic entailment: $A|C \models B|D$ iff

  $A \cap C \subseteq B \cap D$ and $C^c \cup A \subseteq D^c \cup B$

B|D has more examples and less counterexamples than A|C.

In particular $A|C \models A|B \cap C$ is false.

- Quasi-conjunction (Ernest Adams):

$A|C \cap B|D = (C^c \cup A) \cap (D^c \cup B) | C \cup D$

Probability of conditionals

$P(A|C)$ is totally determined by

- $P(A \cap C)$ (proportion of examples)
- $P(A^c \cap C) = 1 - P(A \cup C^c)$ (proportion of counter-examples)

$$P(A|C) = \frac{P(A \cap C)}{P(A \cap C) + 1 - P(A \cup C^c)}$$

- $P(A|C)$ is increasing with $P(A \cap C)$ and decreasing with $P(A^c \cap C)$
- If $A|C \models B|D$ then $P(A|C) \leq P(B|D)$.

Probability and Non-Monotonic Reasoning

- Conditional probability is non-monotonic:
  - $P(A|C)$ can be close to 1 while $P(A|C \cap B)$ is close to 0: learning B makes A implausible.
- Already at the symbolic level :

  $A|C \models A|\ C \cap B$ is not valid (the latter has less examples)
- The three-valued symbolic logic of conditionals is NON monotonic.
- This is is necessary for coping with exceptions, and draw plausible conclusions under incomplete information.

CONDITIONING NON-ADDITIVE CONFIDENCE MEASURES

- **<u>Definition :</u>** A conditional confidence measure $g(A \mid C)$ is a mapping from conditional events $A \mid C \in S\times(S - \{\emptyset\})$ to $[0, 1]$ such that
  - $g(A \mid C) = g(A \cap C \mid C) = g(A^c \cup C \mid C)$
  - $g_C(\cdot) = g(. \mid C)$ is a confidence measure on $C \neq \emptyset$
- Two approaches:
  - <u>Bayes</u>-like $g(A \cap C) = g(A \mid_1 C) \cdot g(C)$
  - <u>Explicit</u> <u>Approach</u> $g(A \mid_2 C) = f(g(A \cap C), g(A \cup C^c))$
    Namely : $f(x, y) = x/(1+x-y)$

Conditioning a credal set

- *Let $\mathcal{P}$ be a credal set representing generic information and C an event*
- *Two types of processing :*

1. ***Querying** : C represents available singular facts: compute* the degree of belief in A in context C as

   Cr(A $|_1$ C) = Inf{P(A | C), P $\in \mathcal{P}$, P(C) > 0} (Walley).

2. ***Revision** : C represents a set of universal truths;*

   Add P(C) = 1 to the set of conditionals $\mathcal{P}$.

   Now we must compute Cr(A$|_2$C) =Inf{P(A) P $\in \mathcal{P}$, P(C) = 1}

   If P(C) = 1 is incompatible with $\mathcal{P}$, use maximum likelihood:

   Cr(A|C) =Inf{P(A|C) P $\in \mathcal{P}$, P(C) maximal }

Example : $A \rightleftarrows B \longrightarrow C$

- *$\mathcal{P}$ is the set of probabilities such that*
  - $P(B|A) \geq \alpha$ *Most A are B*
  - $P(C|B) \geq \beta$ *Most B are C*
  - $P(A|B) \geq \gamma$ *Most B are A*
- ***Querying on context A :*** *Find the most narrow interval for P(C|A) (Linear programming): we find*

$$P(C|A) \geq \alpha \cdot max(0, 1 - (1 - \beta)/\gamma)$$

  - *Note : if $\gamma = 0$, $P(C|A)$ is unknown even if $\alpha = 1$.*
- ***Revision:*** *Suppose $P(A) = 1$, then $P(C|A) \geq \alpha\cdot\beta$*
  - *Note: $\beta > max(0, 1 - (1 - \beta)/\gamma)$*
- **Revision improves generic knowledge, querying does not.**

CONDITIONING RANDOM SETS AS IMPRECISE PROBABILISTIC INFORMATION

- A disjunctive random set ($\mathcal{F}$, m) *representing background knowledge* is equivalent to a set of probabilities $\mathcal{P} = \{P: \forall A, P(A) \geq Bel(A)\}$ (*NOT conversely*).
- Querying this information based on evidence C comes down to performing a sensitivity analysis on the conditional probability $P(\cdot|C)$
  - $Bel_C(A) = \inf \{P(A|C): P \in \mathcal{P}, P(A) > 0\}$
  - $Pl_C(A) = \sup \{P(A|C): P \in \mathcal{P}, P(A) > 0\}$

- **Theorem:** functions $Bel_C(A)$ and $Pl_C(A)$ are belief and plausibility functions of the form

  $Bel_C(A) = Bel(C \cap A)/(Bel(C \cap A) + Pl(C \cap A^c))$

  $Pl_C(A) = Pl(C \cap A)/(Pl(C \cap A) + Bel(C \cap A^c))$

  where $Bel_C(A) = 1 - Pl_C(A^c)$
- *This conditioning does not add information*:
- If $E \cap C \neq \emptyset$ and $E \cap C^c \neq \emptyset$ for all $E \in \mathcal{F}$, then $m_C(C) = 1$ (the resulting mass function m_C expresses total ignorance on C)
  - **Example: If opinion poll yields:**
  - $m(\{a, b\}) = \alpha$, $m(\{c, d\}) = 1 - \alpha$,

  The proportion of voters for a candidate in $C = \{b, c\}$ is unknown.
  - *However if we hear a and d resign ($Pl(\{a, d\} = 0)$ then $m(\{b\}) = \alpha$, $m(\{c\}) = 1 - \alpha$* (Dempster conditioning, see further on)

CONDITIONING UNCERTAIN SINGULAR EVIDENCE

- A mass function *m on S, represents uncertain evidence*
- A new **sure** piece of evidence is viewed as a conditioning event C

1. *Mass transfer* : for all $E \in \mathcal{F}$, m(E) moves to $C \cap E \subseteq C$
   - The mass function after the transfer is $m_t(B) = \Sigma_{E : C \cap E = B}\, m(E)$
   - But the mass transferred to the empty set may not be zero!
   - $m_t(\emptyset) = Bel(C^c) = \Sigma_{E : C \cap E = \emptyset}\, m(E)$ is the degree of conflict with evidence C
2. *Normalisation*: $m_t(B)$ should be divided by Pl(C)

   $= 1 - Bel(C^c) = \Sigma_{E : C \cap E \neq \emptyset}\, m(E)$

- ***This is revision of an unreliable testimony by a sure fact***

DEMPSTER RULE OF CONDITIONING = PRIORITIZED MERGING

The conditional plausibility function Pl(.|C) is

$$Pl(A|C) = \frac{Pl(A \cap C)}{Pl(C)} \quad ; \; Bel(A|C) = 1 - Pl(A^c|C)$$

- C surely contains the value of the unknown quantity described by m. **So $Pl(C^c) = 0$**
  - *The new information is interpreted as asserting the impossibility of C^c: Since C^c is impossible you can change $x \in E$ into $x \in E \cap C$ and transfer the mass of focal set E to $E \cap C$.*
- *The new information improves the precision of the evidence*

- **This conditioning is different from Bayesian (Walley) conditioning**

EXAMPLE OF REVISION OF EVIDENCE : The criminal case

- **Evidence 1** : three suspects : Peter Paul Mary
- **Evidence 2** : The killer was randomly selected man vs.woman by coin tossing.
  - So, S = { Peter, Paul, Mary}
- TBM modeling: The masses are m({Peter, Paul}) = 1/2 ; m({Mary}) = 1/2
  - Bel(Paul) = Bel(Peter) = 0. Pl(Paul) = Pl(Peter) = 1/2
  - Bel(Mary) = Pl(Mary) = 1/2
- **Bayesian Modeling:** A prior probability
  - P(Paul) = P(Peter) = 1/4; P(Mary) = 1/2

- **Evidence 3** : Peter was seen elsewhere at the time of the killing.
- **TBM**: So Pl(Peter) = 0.
  - m({Peter, Paul}) = 1/2; m_t({Paul}) = 1/2
  - *A uniform probability on {Paul, Mary} results*.
- **Bayesian Modeling:**
  - P(Paul | not Peter) = 1/3; P(Mary | not Peter) = 2/3.
  - A very debatable result that depends on where the story starts. *Starting with i males and j females:*
    - P(Paul | Paul OR Mary) = j/(i + j);
    - P(Mary | Paul OR Mary) = i/(i + j)
- *Walley conditioning:*
  - Bel(Paul) = 0; Pl(Paul) = 1/2
  - Bel(Mary) = 1/2; Pl(Mary) = 1

Important pending issues

- **Statistical inference tools** for imprecise probability models
- **Elicitation methods** for belief functions and imprecise probabilities
- **Information measures** beyond entropy, variance, etc.
- **Conditioning** : several definitions for several purposes.
- **Independence**: distinguish between epistemic and objective notions.
- Find a general setting for **information fusion** operations (e.g. beyond Dempster rule of combination).
- Find a consensual approach to **decision-making** under partial ignorance.

4 roles of conditioning

- **Prediction** : given evidence C and generic knowledge (a probability model P), predict observation A with belief degree P(A|C).
- **Fusion** : Merging uncertain evidence (a subjective probability P) with a sure fact C
- **Revision of generic knowledge** : Revise generic knowledge P by absorbing a new piece of information P(C) = 1, minimising change (probability kinematics)
- **Learning** : Given a probabilistic model that depends on a parameter θ interpreted as a conditional probability $P(A|\theta)$, improve the knowledge about θ based on a series of observations $C_1 \ldots C_n$ (e.g. by computing $P(\theta| C_1 \ldots C_n)$).
- Each task may require a specific form of conditioning in uncertaity theories generalizing probabilities

Conclusion

- *There exists a coherent range of uncertainty theories combining interval and probability representations.*
  - Imprecise probability is the proper theoretical umbrella
  - The choice between subtheories depends on how expressive it is necessary to be in a given application.
  - There exists simple practical representations of imprecise probability
- *Many open problems, theoretical, and computational, remain.*
- *How to get this general non-dogmatic approach to uncertainty accepted by traditional statisticians?*

Final quotes Lindley (2000, The Statistician)

- *« Probability is the only satisfactory expression of uncertainty »*
- *« Other rules, like those of fuzzy logic and possibility theory dependent on maxima and minima rather than sums and products, are out »*
- *« The last sentence is not strictly true…. A fine critique is Walley who went on to construct a system with a pair of numbers… instead of the single probability. The result is a more complicated system. My position is that the complication seems unnecessary. »*
- MY CONCLUSION : So possibility theory, simple support functions, random sets, p-boxes being as simple as probability, are back in!