Imprecise probabilities and belief functions

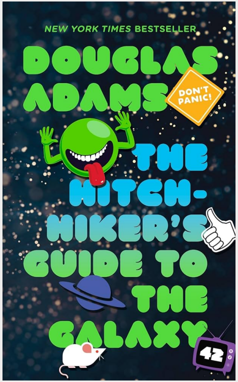
Enrique Miranda

University of Oviedo
mirandaenrique@uniovi.es

BFTA 2025-Granada, October 21, 2025

UNIVERSIDAD
OVIEDO

Why this lecture?



“We demand rigidly defined areas of doubt and uncertainty!”

Forward

Introduction

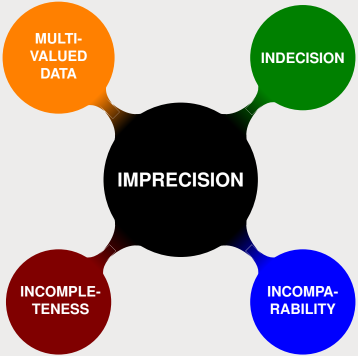
In many situations, the available information does not allow to determine the probability distribution associated with an experiment.

To deal with these cases, a multitude of mathematical models have been proposed. These are usually encompassed under the name imprecise probabilities.

In this lecture, we shall review the main models, signalling their strong and weak points:

1. Sets of probabilities.
2. Non-additive measures.
3. Incomplete preferences.

Imprecise probabilities: why?



IMPRECISION

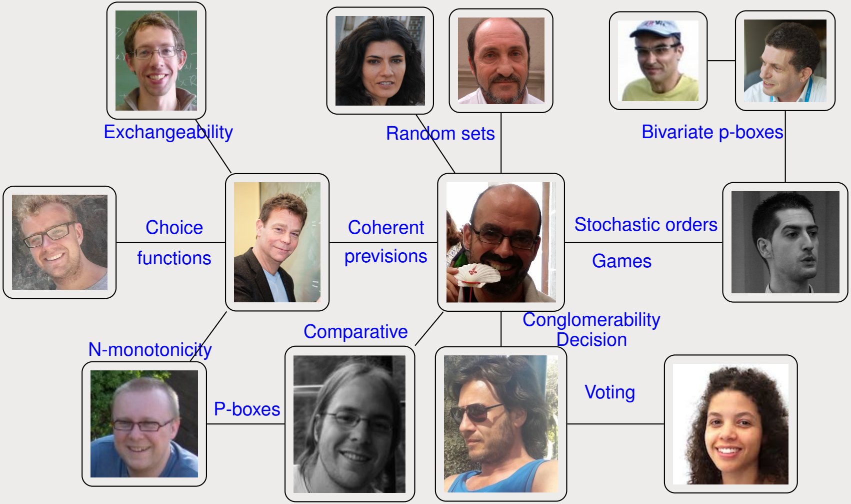
MULTI-VALUED DATA

INDECISION

INCOMPLETENESS

INCOMPARABILITY

What type of imprecision have I been working on?



Exchangeability

Random sets

Bivariate p-boxes

Choice functions

Coherent previsions

Stochastic orders

Games

Conglomerability

Decision

Comparative

N-monotonicity

P-boxes

Voting

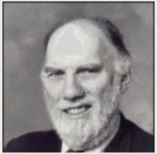
1. Robust models

The first approach we shall consider is based on robust models, where small variations of the initial model do not affect the results.

In this sense, we can consider sets of probability measures and also some particular cases where the set is determined by a few parameters.

This gives rise to models such as *p*-boxes, random sets or distortion models.

Credal sets



Given a measurable space $(\mathcal{X}, \mathcal{A})$, a set of probability measures $\mathcal{M}$ over $\mathcal{A}$ is called credal set.

Credal sets allow to model the available information with as much precision as we have.

Usually (*but not always*) we consider a convex set.

Why (not) use credal sets?

- Almost all models in the literature can be represented with a credal set. ✓
- For many inferences it suffices to work with their extreme points. ✓
- In the case of finite spaces, they are characterised via the notion of coherence. ✓
- There are not satisfactory characterisations of convex sets of σ-additive probabilities. ✗
- The elicitation of a credal set may be difficult. ✗

Particular cases

In order to reduce the computational complexity and ease the elicitation, there are several particular cases of credal sets that are often used in practice.

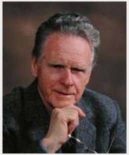
We can consider p-boxes, where the give a lower and an upper bound of the distribution function, and consider the credal set

$$\mathcal{M} = \{P \text{ probability} : \underline{F}(x) \leq F_P(x) \leq \overline{F}(x) \ \forall x\}$$

or probability intervals, where we only specify the lower and upper probabilities of the singletons:

$$\mathcal{M} = \{P \text{ probability} : \underline{P}(\{x\}) \leq P(\{x\}) \leq \overline{P}(\{x\}) \ \forall x\}.$$

Random sets



Given a random variable $U_0 : \Omega \to \mathcal{X}$, there are situations where we do not know its exact value:

- problems in data transmission;
- missing data;
- combination of values from several sources;
- $\ldots$

One option is to consider a multi-valued mapping $\Gamma : \Omega \to \mathcal{P}(\mathcal{X})$, so that $U_0(\omega) \in \Gamma(\omega)$ for every $\omega \in \Omega$.

Under certain measurability conditions, Γ is called a random set.

Credal sets and random sets

Given a random set, the information about the distribution of U_0 is given by

$$\mathcal{P}(\Gamma) = \{P_U : U \in S(\Gamma)\},$$

where

$$S(\Gamma) = \{U \text{ r.v. } : U(\omega) \in \Gamma(\omega) \forall \omega\}$$

is the set of measurable selections of the random set.

In particular, for any measurable set A in $\mathcal{X}$ we can bound

$$P(\{\omega : \Gamma(\omega) \subseteq A\} \leq P_{U_0}(A) \leq P(\{\omega : \Gamma(\omega) \cap A \neq \emptyset\}.$$

Credal sets and random sets

Given a random set, the information about the distribution of U_0 is given by

$$\mathcal{P}(\Gamma) = \{P_U : U \in S(\Gamma)\},$$

where

$$S(\Gamma) = \{U \text{ r.v.} : U(\omega) \in \Gamma(\omega) \forall \omega\}$$

is the set of measurable selections of the random set.

In particular, for any measurable set A in $\mathcal{X}$ we can bound

$$P(\{\omega : \Gamma(\omega) \subseteq A\} \leq P_{U_0}(A) \leq P(\{\omega : \Gamma(\omega) \cap A \neq \emptyset\}.$$

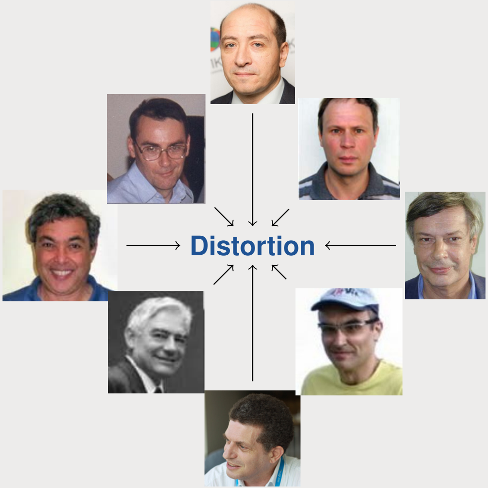
Back
Forward

Distortion models

Distortion models modify *somehow* a probability measure P_0 by a factor $\delta > 0$, that may represent:

- the proportion of contaminated data;
- the fee taken by the betting house;
- the distance to the original model we want to be robust;
- ...

Does anybody care?



Distortion

Examples of distortion models (I)

Some distortion models are based in considering credal sets determined by the values of P_0. We have for instance:

1. Contamination models, associated with the credal set

$$\mathcal{M} = \{(1-\delta)P_0 + \delta Q : Q \text{ probability}\}.$$

2. Constant odds ratio models, where

$$\mathcal{M} = \left\{ Q : \frac{Q(A)}{Q(B)} \geq (1-\delta)\frac{P(A)}{P(B)} \ \forall A, B \right\}.$$

Examples of distortion models (II)

Given a distance d between probability measures, we can consider

$$\mathcal{M}(P_0, d, \delta) = \{P \in \mathbb{P}(X) \mid d(P, P_0) \leq \delta\}.$$

We have for instance the distances:

3. total variation

$$d_{TV}(P, Q) = \sup_{A \subset \mathcal{X}} |P(A) - Q(A)|.$$

4. Kolmogorov

$$d_K(P, Q) = \sup_{x \in \mathcal{X}} |F_P(x) - F_Q(x)|,$$

assuming we have a total order in $\mathcal{X}$.

5. L_1

$$d_{L_1}(P, Q) = \sum_{A \subseteq \mathcal{X}} |P(A) - Q(A)|.$$

Comparison between the models

In International Journal of General Systems 49(6),602-674, (2020) we compared these and other distortion models in terms of a number of criteria:

- The size of the credal set in terms of the initial model P_0 and the distortion factor δ.
- The number of extreme points of the credal set.
- The properties of the distortion function d.
- The properties of the *non-additive measure* it induces.

Moving away from probability

"The laws of probability, so true in general, so fallacious in particular". (E. Gibbon)

2. Non-additive measures

The second approach is based on relaxing the axioms from probability theory, considering more general models.

Usually, we keep the conditions of monotonicity and normalization, and relax the property of additivity.

This leads to a number of different models, that are usually called capacities, non-additive measures or fuzzy measures.

Capacities: definition

Given a space $\mathcal{X}$, a capacity is a monotone and normalised function $\underline{P}: \mathcal{P}(\mathcal{X}) \to [0,1]$.

Among the properties that capacities may satisfy, we can consider the following:

- $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)\ \forall A, B$ disjoint (super-additivity).
- $\underline{P}(A \cup B) \leq \underline{P}(A) + \underline{P}(B)\ \forall A, B$ disjoint (subadditivity).
- $\underline{P}(\cup_n A_n) = \sup_n \underline{P}(A_n)$ for every increasing sequence (lower continuity).
- $\underline{P}(\cap_n A_n) = \inf_n \underline{P}(A_n)$ for every decreasing sequence (upper continuity).

The choice between them depends on the interpretation of $\underline{P}$.

Conjugate functions

Given a capacity $\underline{P}$ on $\mathcal{P}(\mathcal{X})$, its conjugate $\overline{P}$ is given by

$$\overline{P}(A) = 1 - \underline{P}(A^c)\ \forall A \subseteq \mathcal{X}.$$

- $\underline{P}$ is superadditive $\Leftrightarrow$ $\overline{P}$ subadditive.
- $\underline{P}$ lower continuous $\Leftrightarrow$ $\overline{P}$ is upper continuous.

Back Forward

(Coherent) lower probabilities

We shall call $\underline{P}$ a lower probability when we interpret it as a lower bound of a probability measure.

In that case, its associated credal set is

$$\mathcal{M}(\underline{P}) = \{P \text{ probability} \mid P(A) \geq \underline{P}(A)\ \forall A \subseteq \mathcal{X}\}.$$

$\underline{P}$ avoids sure loss if $\mathcal{M}(\underline{P}) \neq \emptyset$, and is called coherent when

$$\underline{P}(A) = \min_{P \in \mathcal{M}(\underline{P})} P(A)\ \forall A \subseteq \mathcal{X}.$$

Its conjugate upper probability is defined by $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \subseteq \mathcal{X}$.

The epistemic interpretation

Assume the uncertainty about the experiment is modelled by some probability measure P_0 that is partially known, in that for every event of interest A we can only establish that

$$P_0(A) \in [\underline{P}(A), \overline{P}(A)].$$

The functionals $\underline{P}, \overline{P}$ thus defined are lower and upper probabilities.

The credal set $\{P : \underline{P} \leq P \leq \overline{P}\}$ would be the set of possible candidates for P_0.

$\underline{P}$ lower probability $\not\Rightarrow$ $\underline{P}$ avoids sure loss

Consider $\mathcal{X} = \{1, 2\}$ and the lower probability $\underline{P}$ given by

$$\underline{P}(\emptyset) = 0,\ \underline{P}(\{1\}) = 0.6,\ \underline{P}(\{2\}) = 0.6,\ \underline{P}(\{1, 2\}) = 1.$$

$\underline{P}$ lower probability $\not\Rightarrow$ $\underline{P}$ avoids sure loss

Consider $\mathcal{X} = \{1, 2\}$ and the lower probability $\underline{P}$ given by

$$\underline{P}(\emptyset) = 0,\ \underline{P}(\{1\}) = 0.6,\ \underline{P}(\{2\}) = 0.6,\ \underline{P}(\{1, 2\}) = 1.$$

$\underline{P}$ is monotone and normalised, but there is no $P \geq \underline{P}$: it would satisfy

$$P(\{1, 2\}) = P(\{1\}) + P(\{2\}) \geq 0.6 + 0.6 = 1.2 > 1!$$

This means that $\underline{P}$ incurs a sure loss.

$\underline{P}$ avoiding sure loss $\not\Rightarrow$ $\underline{P}$ coherent

Assume that $\mathcal{X} = \{1,2,3\}$ and that $\underline{P}$ is given by:

$$\begin{array}{llll} \underline{P}(\emptyset) = 0 & \underline{P}(\{1\}) = 0.1 & \underline{P}(\{2\}) = 0.2 & \underline{P}(\{3\}) = 0.3 \\ \underline{P}(\{1,2\}) = 0.6 & \underline{P}(\{1,3\}) = 0.6 & \underline{P}(\{2,3\}) = 0.6 & \underline{P}(\mathcal{X}) = 1 \end{array}$$

$\underline{P}$ avoiding sure loss $\not\Rightarrow$ $\underline{P}$ coherent

Assume that $\mathcal{X} = \{1, 2, 3\}$ and that $\underline{P}$ is given by:

$$\begin{array}{llll} \underline{P}(\emptyset) = 0 & \underline{P}(\{1\}) = 0.1 & \underline{P}(\{2\}) = 0.2 & \underline{P}(\{3\}) = 0.3 \\ \underline{P}(\{1,2\}) = 0.6 & \underline{P}(\{1,3\}) = 0.6 & \underline{P}(\{2,3\}) = 0.6 & \underline{P}(\mathcal{X}) = 1 \end{array}$$

$\underline{P}$ avoids sure loss: the probability with mass function $(0.2, 0.4, 0.4)$ dominates it.

$\underline{P}$ avoiding sure loss $\not\Rightarrow$ $\underline{P}$ coherent

Assume that $\mathcal{X} = \{1,2,3\}$ and that $\underline{P}$ is given by:

$$\begin{array}{llll}
\underline{P}(\emptyset) = 0 & \underline{P}(\{1\}) = 0.1 & \underline{P}(\{2\}) = 0.2 & \underline{P}(\{3\}) = 0.3 \\
\underline{P}(\{1,2\}) = 0.6 & \underline{P}(\{1,3\}) = 0.6 & \underline{P}(\{2,3\}) = 0.6 & \underline{P}(\mathcal{X}) = 1
\end{array}$$

$\underline{P}$ avoids sure loss: the probability with mass function $(0.2, 0.4, 0.4)$ dominates it.

However, $\underline{P}$ is NOT coherent: it is impossible to find a probability measure $P \geq \underline{P}$ such that $P(\{1\}) = 0.1$.

Becoming coherent: the natural extension

If $\underline{P}$ is not coherent but its associated credal set $\mathcal{M}(\underline{P})$ is not empty, we can make a minimal correction so as to obtain a coherent model: there is a smallest $\underline{P}^{'} \geq \underline{P}$ that is coherent. This is called the natural extension of $\underline{P}$.

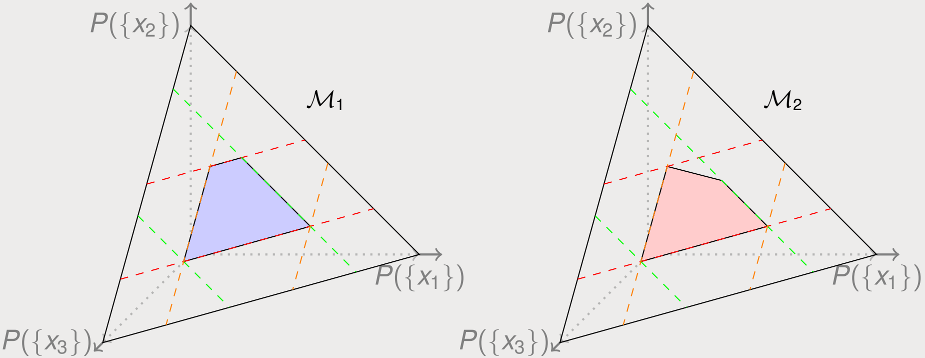
To obtain it, we simply have to take the lower envelope of the credal set $\mathcal{M}(\underline{P})$.

Why (not) use coherent lower probabilities?

- $\underline{P}$ is coherent $\iff \underline{P} = \min \mathcal{M}(\underline{P})$. ✓
- In addition to the epistemic interpretation, they also have a behavioural interpretation. ✓
- However, the structure of $\mathcal{M}(\underline{P})$ may be complex. ✗
- The extension to expectations (the so-called *coherent lower previsions*) is not unique. ✗

Lower previsions $\neq$ lower probabilities

Two different sets of probabilities may determine the same coherent lower probability.



2-monotone capacities

A lower probability is 2-monotone when

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B)\ \forall A, B.$$

2-monotone capacities are also called submodular or convex:

$$P_0 \text{ probability}, f : [0,1] \to [0,1] \text{ convex} \Rightarrow \underline{P} := f(P_0)\ 2-\text{monotone}.$$

A 2-monotone lower probability is always coherent.

$\underline{P}$ coherent $\not\Rightarrow$ $\underline{P}$ 2-monotone

Consider $\mathcal{X} = \{1, 2, 3, 4\}$ and let $\underline{P} := \min\{P_1, P_2\}$, where P_1, P_2 are determined by the mass functions $(0.5, 0.5, 0, 0), (0.25, 0.25, 0.25, 0.25)$.

$\underline{P}$ coherent $\not\Rightarrow$ $\underline{P}$ 2-monotone

Consider $\mathcal{X} = \{1, 2, 3, 4\}$ and let $\underline{P} := \min\{P_1, P_2\}$, where P_1, P_2 are determined by the mass functions $(0.5, 0.5, 0, 0), (0.25, 0.25, 0.25, 0.25)$.

- $\underline{P}$ is coherent, because all lower envelopes of probability measures are.

$\underline{P}$ coherent $\not\Rightarrow$ $\underline{P}$ 2-monotone

Consider $\mathcal{X} = \{1,2,3,4\}$ and let $\underline{P} := \min\{P_1, P_2\}$, where P_1, P_2 are determined by the mass functions $(0.5, 0.5, 0, 0), (0.25, 0.25, 0.25, 0.25)$.

- $\underline{P}$ is coherent, because all lower envelopes of probability measures are.
- However, given $A = \{1,3\}, B = \{2,3\}$, it holds that

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) = 0.75 + 0 = 0.75 < 1 = 0.5 + 0.5 = \underline{P}(A) + \underline{P}(B),$$

whence $\underline{P}$ is NOT 2-monotone.

Example: probability intervals

Given a finite space $\mathcal{X} = \{x_1, \ldots, x_n\}$, a probability interval consists in determining $[\underline{P}(\{x_i\}), \overline{P}(\{x_i\})]$ for $i = 1, \ldots, n$.

- This models avoids sure loss iff $\sum_{i=1}^{n} \underline{P}(\{x_i\}) \leq 1 \leq \sum_{i=1}^{n} \overline{P}(\{x_i\})$.
- Similarly, Moral et al. characterised in which cases the model is coherent.
- A coherent probability interval is always 2-monotone.
- Some distortion models, such as the contamination models, are included in this family.

Why (not) use 2-monotone capacities?

If $\underline{P}$ is 2-monotone:

- There is a unique way of defining the expectation operator (the *Choquet integral*). ✓
- The extreme points of $\mathcal{M}(\underline{P})$ can be computed easily. ✓
- Almost all models in the literature satisfy this condition. ✓
- The behavioural interpretation of 2-monotonicity is not too clear. ✗

k-monotonicity

Similarly, a lower probability is k-monotone when

$$\underline{P}\Big(\cup_{i=1}^{k} A_i\Big) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \underline{P}\Big(\cap_{i \in I} A_i\Big)$$

for every $A_1, \ldots, A_k$ in $\mathcal{P}(\mathcal{X})$. If it is k-monotone for every k, it is called ∞-monotone.

The conjugate upper probability is called k-alternating.

$\underline{P}$ 2-monotone $\not\Rightarrow$ $\underline{P}$ 3-monotone

Consider $\mathcal{X} = \{1, 2, 3\}$, and let $\underline{P}$ be the lower envelope of

$$\{(0.4, 0.4, 0.2), (0.4, 0.2, 0.4), (0.2, 0.4, 0.4)\}.$$

$\underline{P}$ 2-monotone $\not\Rightarrow$ $\underline{P}$ 3-monotone

Consider $\mathcal{X} = \{1,2,3\}$, and let $\underline{P}$ be the lower envelope of

$$\{(0.4,0.4,0.2),(0.4,0.2,0.4),(0.2,0.4,0.4)\}.$$

- $\underline{P}$ is coherent, and also 2-monotone because $|\mathcal{X}| = 3$ (Walley).

$\underline{P}$ 2-monotone $\not\Rightarrow$ $\underline{P}$ 3-monotone

Consider $\mathcal{X} = \{1, 2, 3\}$, and let $\underline{P}$ be the lower envelope of

$$\{(0.4, 0.4, 0.2), (0.4, 0.2, 0.4), (0.2, 0.4, 0.4)\}.$$

- $\underline{P}$ is coherent, and also 2-monotone because $|\mathcal{X}| = 3$ (Walley).
- Given $A_1 = \{1, 2\}, A_2 = \{1, 3\}, A_3 = \{2, 3\}$,

$$\underline{P}(A_1 \cup A_2 \cup A_3) = 1 < \sum_i \underline{P}(A_i) - \sum_{i \neq j} \underline{P}(A_i \cap A_j) + \underline{P}(A_1 \cap A_2 \cap A_3)$$
$$= 1.8 - 0.6 + 0 = 1.2.$$

Thus, $\underline{P}$ is NOT 3-monotone.

$\underline{P}$ 2-monotone $\not\Rightarrow$ $\underline{P}$ 3-monotone

Consider $\mathcal{X} = \{1, 2, 3\}$, and let $\underline{P}$ be the lower envelope of

$$\{(0.4, 0.4, 0.2), (0.4, 0.2, 0.4), (0.2, 0.4, 0.4)\}.$$

- $\underline{P}$ is coherent, and also 2-monotone because $|\mathcal{X}| = 3$ (Walley).
- Given $A_1 = \{1, 2\}, A_2 = \{1, 3\}, A_3 = \{2, 3\}$,

$$\underline{P}(A_1 \cup A_2 \cup A_3) = 1 < \sum_i \underline{P}(A_i) - \sum_{i \neq j} \underline{P}(A_i \cap A_j) + \underline{P}(A_1 \cap A_2 \cap A_3)$$
$$= 1.8 - 0.6 + 0 = 1.2.$$

Thus, $\underline{P}$ is NOT 3-monotone.

Other 2-monotone models that are not always 3-monotone are the total variation distortion models.

Why (not) use *k*-monotone capacities?

- All the nice mathematical properties of 2-monotone properties hold in this particular case. ✓
- In addition, Choquet established a characterisation in terms of differences. ✓
- However, some interesting models that are 2-monotone need do not satisfy 3-monotonicity. ✗

Belief functions

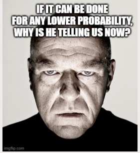
An ∞-monotone capacity is called belief function.

Like any other lower probability, when $\mathcal{X}$ is finite, it can be represented through their Möbius inverse:

$$m_{\underline{P}}(A) = \sum_{B\subseteq A} (-1)^{|A\setminus B|}\underline{P}(B) \quad \forall A \subseteq \mathcal{X},$$

and it holds that

$$\underline{P}(A) = \sum_{B\subseteq A} m_{\underline{P}}(B) \ \forall A \subseteq \mathcal{X}.$$



Back
Forward

Möbius inverse=mass function

While this can be done for any lower probability, in the case of belief functions (and only then) it holds that and moreover $m_{\underline{P}}(A) \geq 0$ for every A and $\sum_{A \subseteq \mathcal{X}} m_{\underline{P}}(A) = 1$: $m_{\underline{P}}$ is the mass function associated with the belief function!

The events A such that $m_{\underline{P}}(A) > 0$ are called focal elements of $\underline{P}$. Moreover, $m_{\underline{P}}$ also determines the conjugate plausibility function $\overline{P}$ by

$$\overline{P}(A) = \sum_{B \cap A \neq \emptyset} m_{\underline{P}}(B) \; \forall A \subseteq \mathcal{X}.$$

Example: *p*-boxes

Assume $\mathcal{X}$ is ordered, and let $\underline{F}, \overline{F} : \mathcal{X} \to [0,1]$ be distribution functions, i.e.,

$$x_i \leq x_j \Rightarrow \underline{F}(x_i) \leq \underline{F}(x_j) \text{ and } \overline{F}(x_i) \leq \overline{F}(x_j), \underline{F}(x_n) = \overline{F}(x_n) = 1.$$

The pair $(\underline{F}, \overline{F})$ is called a *p*-box, and the credal set

$$\mathcal{M}(\underline{F}, \overline{F}) = \{P : \underline{F} \leq F_P \leq \overline{F}\}$$

determines coherent lower and upper probabilities $\underline{P}, \overline{P}$ taking lower and upper envelopes.

$\underline{P}$ is a belief function and $\overline{P}$ is a plausibility function.

Why (not) use belief functions?

- Belief functions are the basis of evidence theory. ✓
- They are (basically) equivalent to random sets. ✓
- They are many tools established specifically for belief functions. ✓
- In the case of infinite spaces the model is a bit more complex. ✗
- Some interesting models do not satisfy ∞-monotonicity. ✗

Possibility and necessity measures

A possibility measure on $\mathcal{X}$ is a function $\Pi : \mathcal{P}(\mathcal{X}) \to [0, 1]$ that satisfies that for any family of subsets $(A_i)_{i \in I}$ of $\mathcal{X}$,

$$\Pi(\cup_{i \in I} A_i) = \sup_{i \in I} \Pi(A_i).$$

Its conjugate function, $Nec(A) = 1 - \Pi(A^c)$, is called necessity measure, and satisfies

$$Nec(\cap_{i \in I} A_i) = \inf_{i \in I} Nec(A_i)$$

for every $(A_i)_{i \in I}$.

Maxitive and minitive measures

In particular, a possibility measure is maxitive: $\overline{P}(A \cup B) = \max\{\overline{P}(A), \overline{P}(B)\}\ \forall A, B$ and a necessity measure is minitive: $\underline{P}(A \cap B) = \min\{\underline{P}(A), \underline{P}(B)\}\ \forall A, B$.

- Minitive measures are in particular belief functions.
- If $\mathcal{X}$ is finite, $\overline{P}$ possibility $\iff$ $\overline{P}$ maxitive.
- In general, a maxitive measure is a possibility iff it satisfies the property of condensability of Shafer.

$\underline{P}$ belief function $\not\Rightarrow$ $\underline{P}$ minitive measure

Consider $\mathcal{X} = \{1, 2, 3\}$ and let $\underline{P}$ be the lower envelope of

$$\{(0.5, 0.25, 0.25), (0.25, 0.5, 0.25), (0.25, 0.25, 0.5)\}.$$

$\underline{P}$ belief function $\not\Rightarrow$ $\underline{P}$ minitive measure

Consider $\mathcal{X} = \{1, 2, 3\}$ and let $\underline{P}$ be the lower envelope of

$$\{(0.5, 0.25, 0.25), (0.25, 0.5, 0.25), (0.25, 0.25, 0.5)\}.$$

- It can be checked that $\underline{P}$ is a belief function.

$\underline{P}$ belief function $\not\Rightarrow$ $\underline{P}$ minitive measure

Consider $\mathcal{X} = \{1, 2, 3\}$ and let $\underline{P}$ be the lower envelope of

$$\{(0.5, 0.25, 0.25), (0.25, 0.5, 0.25), (0.25, 0.25, 0.5)\}.$$

- It can be checked that $\underline{P}$ is a belief function.
- However, it is NOT minitive: given $A = \{1, 2\}, B = \{1, 3\}$,

$$\underline{P}(A \cap B) = \underline{P}(\{1\}) = 0.25 < \min\{\underline{P}(A), \underline{P}(B)\} = 0.5.$$

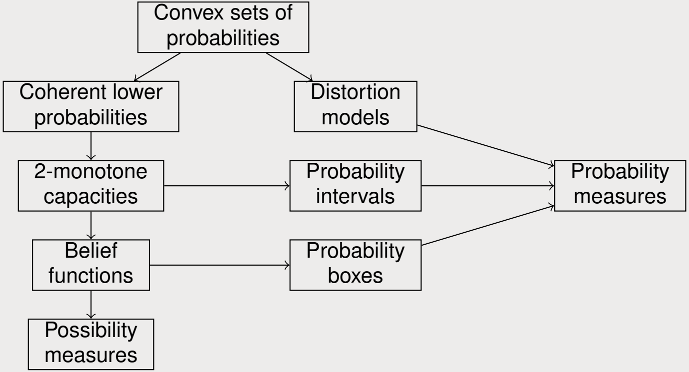
Why (not) use possibility measures?

- A possibility measure is ∞-alternating, so all the good properties of belief functions are satisfied. ✓
- They are simple from the computational point of view, since they are determined by their restriction to events. ✓
- Like belief functions, there are many tools established specifically for them. ✓
- They are related to fuzzy sets. ✓
- They are a very narrow model that does not behave well under transformations. ✗

Are these really imprecise *probabilities*?

- By construction, a probability measure is in particular a coherent lower probability. ✓
- It is also a belief function, and as a consequence also k-monotone for every k. ✓
- They can also be regarded as cases of probability intervals and p-boxes. ✓
- However, it will not be a possibility measure except for trivial cases. ✗

Relationship between the models



Convex sets of probabilities

Coherent lower probabilities

Distortion models

2-monotone capacities

Probability intervals

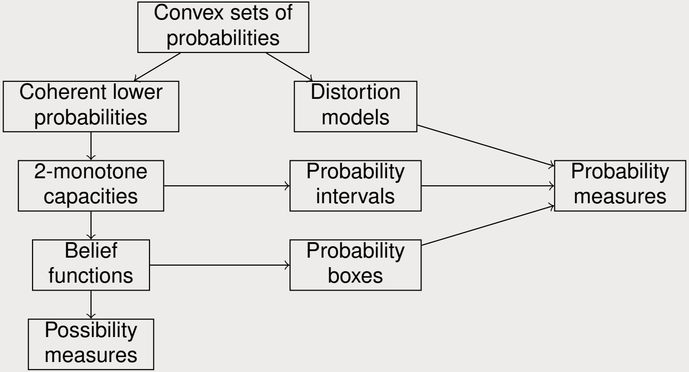
Probability measures

Belief functions

Probability boxes

Possibility measures

Relationship between the models



Convex sets of probabilities

Coherent lower probabilities

Distortion models

2-monotone capacities

Probability intervals

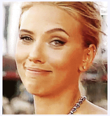
Probability measures

Belief functions

Probability boxes

Possibility measures

So many models...I wish there was a place to discuss them all!

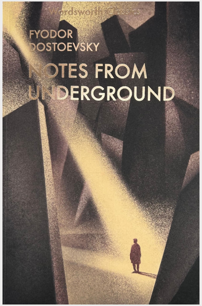


12th SIPTA School on Imprecise Probabilities
Munich, July 27-31, 2026

SIPTA

https://www.sipta.org/

And now for something completely different



"...there is one case, one only, when man may consciously, purposely, desire what is injurious to himself, what is stupid, very stupid–simply in order to have the right to desire for himself even what is very stupid and not to be bound by an obligation to desire only what is sensible."

3. The behavioural interpretation

The third approach is based on the ideas of de Finetti about coherent sets of desirable gambles $f : \mathcal{X} \to \mathbb{R}$.

He characterised finitely additive probabilities (=coherent linear previsions) as *fair* prices for a gamble, in the sense that a combination of desirable transactions does not make us lose money in an uniform manner.

His ideas were extended by Peter Williams and Peter Walley to allow for indecision.

Example

2026
FIFA WORLD CUP
UNITED STATES|CANADA|MÉXICO

Who will win the World football cup?

Consider the options a=*Brazil*, b=*France*, c=*Spain*, d=*Other*.

Take gambles (=games with uncertain prize) on $\mathcal{X} = \{a, b, c, d\}$, such as

$$f(a) = -3, f(b) = 2, f(c) = 5, f(d) = 10.$$

Depending on how probable victory is, we shall accept the gamble or not.

Previsions

If we knew the probability of each result of the experiment, then the fair price for a gamble f should be the expected win $P(f)$:

- for any smaller μ, the gamble $f - \mu$ produces a long-term win (so we should be disposed to buy f at this price);
- for any greater μ, the gamble $f - \mu$ produces a long-term loss (so we should accept its opposite $\mu - f$, that is, selling f at price μ).

Existence of indecision

If we do not have enough information, it may be difficult (and unreasonable) to establish a fair price $P(f)$: there will be values μ for which we are not disposed to buy nor sell f.

In terms of desirable gambles, this means we are *undecided*.

It is preferrable to give different values $\underline{P}(f) < \overline{P}(f)$ instead of giving a precise value (and possibly wrong) for the fair price.

We define then a lower prevision $\underline{P}$ and a upper prevision $\overline{P}$ from the set $\mathcal{L}(\mathcal{X})$ of all gambles from $\mathcal{X}$ to $\mathbb{R}$.

Rationale of accepting a gamble

The theory is based in the following considerations:

- If a gamble always makes us lose money we should not accept it, and if it never makes us lose money we should accept it.
- Changing the utility scale does not affect the desirability of a gamble.
- If two gambles are acceptable separately, we should be disposed to accept them together.

Avoiding sure loss: behavioural interpretation

A lower prevision $\underline{P} : \mathcal{L}(\mathcal{X}) \to \mathbb{R}$ avoids sure loss if and onfly if for every $k \in \mathbb{N}, f_1, \ldots, f_k \in \mathcal{L}(\mathcal{X}), \lambda_1, \ldots, \lambda_k \geq 0$,

$$\sup_{\omega \in \mathcal{X}} \left[\sum_{i=1}^{k} \lambda_i (f_i - \underline{P}(f_i)) \right] (\omega) \geq 0.$$

If this does not hold, a combination of acceptable transactions results in a sure loss $\hookrightarrow$ Dutch book argument.

Example

Assume that we consider the gambles f_1, f_2 in the following table:

	a	b	c	d
f_1	1	2	-3	0
f_2	0	-2	2.5	1.8

and that we establish $\underline{P}(f_1) = 1$ and $\underline{P}(f_2) = 1.8$.

Example

Assume that we consider the gambles f_1, f_2 in the following table:

	a	b	c	d
f_1	1	2	-3	0
f_2	0	-2	2.5	1.8

and that we establish $\underline{P}(f_1) = 1$ and $\underline{P}(f_2) = 1.8$. Then we should accept $g_1 = f_1 - 0.8$ and $g_2 = f_2 - 1.5$.

Example

Assume that we consider the gambles f_1, f_2 in the following table:

	a	b	c	d
f_1	1	2	-3	0
f_2	0	-2	2.5	1.8

and that we establish $\underline{P}(f_1) = 1$ and $\underline{P}(f_2) = 1.8$. Then we should accept $g_1 = f_1 - 0.8$ and $g_2 = f_2 - 1.5$. But if we buy g_1 once and g_2 twice we get

	a	b	c	d
$g_1 + 2g_2$	-2.8	-5.8	-1.8	-0.2

$\hookrightarrow$ We end up losing money, no matter the outcome!

Coherence: behavioural interpretation

A lower prevision $\underline{P} : \mathcal{L}(\mathcal{X}) \to \mathbb{R}$ is coherent if and onfly if for every $k \in \mathbb{N}, f_0, f_1, \dots, f_k \in \mathcal{L}(\mathcal{X}), \lambda_0, \lambda_1, \dots, \lambda_k \geq 0$,

$$\sup_{\omega \in \mathcal{X}} \left[\sum_{i=1}^{k} \lambda_i (f_i - \underline{P}(f_i)) - \lambda_0 (f_0 - \underline{P}(f_0)) \right] (\omega) \geq 0.$$

When this does not hold, it is possible to make a combination of acceptable transactions means that allows us to increase the supremum acceptable buying price for f_0.

Example

Assume now that for the gambles f_1, f_2, f_3 in the following table:

	a	b	c	d
f_1	1	2	-3	0
f_2	0	-2	2.5	1.8
f_3	2.5	0.9	0.4	3

we establish $\underline{P}(f_1) = 0$, $\underline{P}(f_2) = 1.8$ and $\underline{P}(f_3) = 2$.

Example

Assume now that for the gambles f_1, f_2, f_3 in the following table:

	a	b	c	d
f_1	1	2	-3	0
f_2	0	-2	2.5	1.8
f_3	2.5	0.9	0.4	3

we establish $\underline{P}(f_1) = 0$, $\underline{P}(f_2) = 1.8$ and $\underline{P}(f_3) = 2$. Then we should accept $g_1 = f_1 + 0.1$ and $g_2 = f_2 - 1.7$.

Example

Assume now that for the gambles f_1, f_2, f_3 in the following table:

	a	b	c	d
f_1	1	2	-3	0
f_2	0	-2	2.5	1.8
f_3	2.5	0.9	0.4	3

we establish $\underline{P}(f_1) = 0$, $\underline{P}(f_2) = 1.8$ and $\underline{P}(f_3) = 2$. Then we should accept $g_1 = f_1 + 0.1$ and $g_2 = f_2 - 1.7$. But if we buy g_1 and g_2 we get

	a	b	c	d
$g_1 + g_2$	-0.6	-1.6	-2.1	0.2

Example

Assume now that for the gambles f_1, f_2, f_3 in the following table:

	a	b	c	d
f_1	1	2	-3	0
f_2	0	-2	2.5	1.8
f_3	2.5	0.9	0.4	3

we establish $\underline{P}(f_1) = 0$, $\underline{P}(f_2) = 1.8$ and $\underline{P}(f_3) = 2$. Then we should accept $g_1 = f_1 + 0.1$ and $g_2 = f_2 - 1.7$. But if we buy g_1 and g_2 we get

	a	b	c	d
$g_1 + g_2$	-0.6	-1.6	-2.1	0.2

This is dominated by the gamble $f_3 - 2.5$:

	a	b	c	d
$f_3 - 2.5$	0	-1.6	-2	0.5

$\hookrightarrow$ We would conclude that 2.5 is an acceptable buying price for f_3, so 2 is NOT the supremum acceptable buying price.

Back Forward

What does this have to do with precise probability?

Assume $\underline{P}$ coincides with its conjugate upper prevision: $\underline{P}(f) = \overline{P}(f) = -\underline{P}(-f)$ for every f, and denote it $P := \underline{P} = \overline{P}$. Then the following are equivalent:

1. P avoids sure loss.
2. P is coherent.
3. The restriction to events of P is a finitely additive probability, and P is its expectation operator.

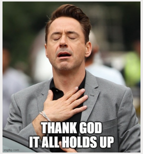
In any of those cases, P is called a coherent prevision.

Is there a connection with the previous notions?

Let $\underline{P} : \mathcal{L}(\mathcal{X}) \to \mathbb{R}$ be a lower prevision, and denote

$$\mathcal{M}(\underline{P}) = \{P \text{ coherent prevision} : P(f) \geq \underline{P}(f)\ \forall f\}.$$

- $\underline{P}$ avoids sure loss $\Leftrightarrow \mathcal{M}(\underline{P}) \neq \emptyset$.
- $\underline{P}$ coherent $\Leftrightarrow \underline{P}(f) = \min_{P \in \mathcal{M}(\underline{P})} P(f)\ \forall f$.



Coherent sets of gambles

A subset $\mathcal{D} \subseteq \mathcal{L}(\mathcal{X})$ is called coherent when:

- $f \gneq 0 \Rightarrow f \in \mathcal{D}$.
- $f \leq 0 \Rightarrow f \notin \mathcal{D}$.
- $f \in \mathcal{D}, \lambda > 0 \Rightarrow \lambda f \in \mathcal{D}$.
- $f, g \in \mathcal{D} \Rightarrow f + g \in \mathcal{D}$.

When we tell which transactions we accept, we are implicitly determining the lower and upper previsions.

Connection with coherent previsions and credal sets

A coherent set of desirable gambles $\mathcal{D}$ determines a coherent lower prevision by

$$\underline{P}(f) := \sup\{\mu : f - \mu \in \mathcal{D}\},$$

and a credal set

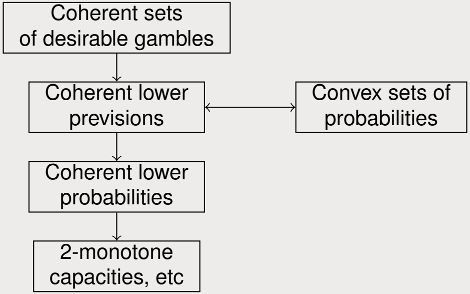
$$\mathcal{M} := \{P : P(f) \geq 0 \ \forall f \in \mathcal{D}\}.$$

In general, a coherent lower prevision or a credal set can be determined by several different coherent sets of desirable gambles. The smallest is

$$\mathcal{D}_{\underline{P}} := \{f : \underline{P}(f) > 0\} \cup \{f \gneq 0\}.$$

↪ Thus, coherent sets of desirable gambles are a *more informative* model.

Relationship between the models (II)



Coherent sets of desirable gambles

Coherent lower previsions

Convex sets of probabilities

Coherent lower probabilities

2-monotone capacities, etc

Preference relations and decision theory

If we have a coherent set of desirable gambles $\mathcal{D}$, we can define a preference relation $\succ$ by

$$X \succ Y \iff X - Y \in \mathcal{D}.$$

This notion is called maximality. It is one of the (several) possible optimality criteria when we want to establish our preferences with imprecise probabilities.

Others are E-admissibility, interval dominance or maximin, and are based on the associated credal sets or lower/upper previsions.

Why (not) use sets of desirable gambles?

- Coherent sets of desirable gambles are more informative than all the previous models. ✓
- They solve some issues related to conditioning on sets of zero lower probability. ✓
- The extension to unbounded gambles is not immediate. ✗
- There is an underlying hypothesis of linear utility that is not always applicable. ✗

For more information...

General introduction:

- T. Augustin, F. Coolen, G. de Cooman, M. Troffaes (editors). *Introduction to imprecise probabilities*. Wiley, 2014.
- Society for Imprecise Probabilities: Theories and Applications (SIPTA). `https://sipta.org/`.

Non-additive measures:

- D. Denneberg. Non-additive measure and integral. Kluwer, 1994.
- M. Grabisch. Set functions, games and capacities for decision making. Springer, 2016.

Coherent lower previsions, sets of desirable gambles:

- P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.

For more information...

Distortion models:

- I. Montes, E. Miranda, S. Destercke. *Univfying neighbourhood and distortion models. Parts I and II*. Int. J. of General Systems, 49(6), 602-674, 2020.

2- and k-monotone capacities:

- G. Choquet. *Theory of capacities*. Annales de l'Institut Fourier, 1953-1954.
- G. de Cooman, M. Troffaes, E. Miranda. *n*-monotone exact functionals. J. of Mathematical Analysis and Applications, 347, 143-156, 2008.

Probability intervals:

- L.M. de Campos, J. Huete, S. Moral. *Probability intervals: a tool for uncertain reasoning*. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, 2, 167-196, 1994.

For more information...

Belief functions:

- G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.

Probability boxes:

- S. Ferson, V. Kreinovich, L. Ginzburg, K. Sentz and D.S. Myers. *Constructing probability boxes and Dempster-Shafer structures*. Sandia Technical Report, 2003.
- M. Troffaes, S. Destercke. *probability boxes on totally preordered spaces for multivariate modelling*. Int. J. of Approximate Reasoning, 52(6), 767-791, 2011.

Possibility measures:

- D. Dubois, H. Prade. *Possibility theory*. Plenum Press, 1988.

Thank you for the attention...

...and for the questions!