Introduction to imprecise probabilities

Inés Couso and Enrique Miranda
University of Oviedo
(couso,mirandaenrique)@uniovi.es

Overview

1. Introduction.
2. Models of non-additive measures.
3. Extension to expectation operators.
4. Preference modelling.
5. Independence.

What is the goal of probability?

Probability seeks to determine the plausibility of the different outcomes of an experiment when these cannot be predicted beforehand.

- What is the probability of guessing the 6 winning numbers in the lottery?
- What is the probability of arriving in 30' from the airport to the center of Oviedo by car?
- What is the probability of having a sunny day tomorrow?

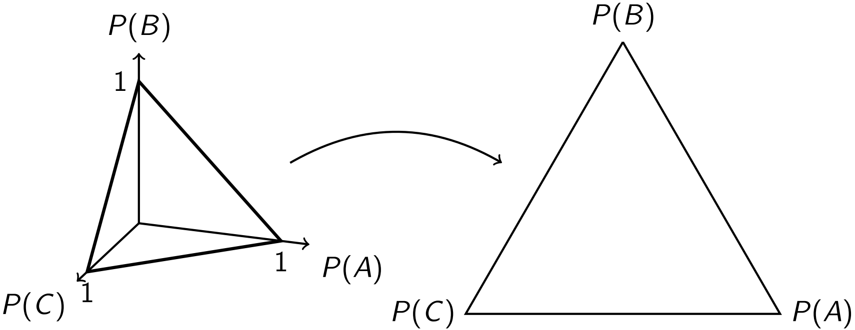
In this talk, we will consider **finite** spaces only.

Given such a space Ω, a probability is a functional P on $\wp(\Omega)$ satisfying:

- $P(\emptyset) = 0, P(\Omega) = 1.$
- $A \subseteq B \Rightarrow P(A) \leq P(B).$
- $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B).$

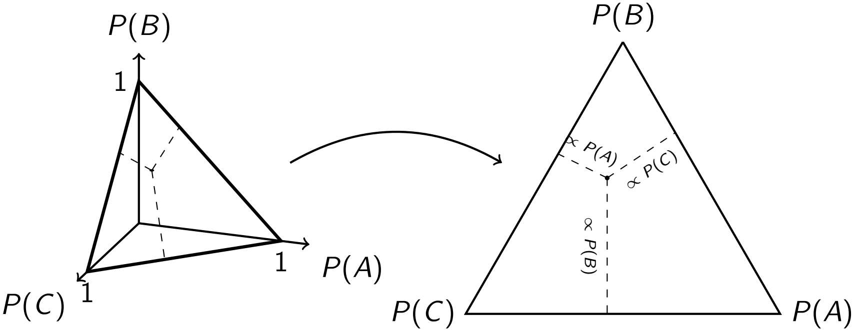
Graphical representations as points in $|\Omega|$ space

$$P(A) = 0.2,\ P(B) = 0.5,\ P(C) = 0.3$$



Graphical representations as points in $|\Omega|$ space

$$P(A) = 0.2,\ P(B) = 0.5,\ P(C) = 0.3$$



Aleatory vs. epistemic probabilities

In some cases, the probability of an event A is a property of the event, meaning that it does not depend on the subject making the assessment. We talk then of aleatory probabilities.

However, and specially in the framework of decision making, we may need to assess probabilities that represent *our* beliefs. Hence, these may vary depending on the subject or on the amount of information he possesses at the time. We talk then of subjective probabilities.

Example: frequentist probabilities

$$P(A) := \lim_{N \to \infty} \frac{\text{number of occurrences of A in N trials}}{N}$$

Examples

- game of chance (loteries, poker, roulette)
- physical quantities in
  - engineering (component failure, product defect)
  - biology (patient variability)
  - economics , . . .

But. . .

Frequentist probabilities: end of the story?

. . . some uncertain quantities are not repeatable/not statistical quantities:

- what's the age of the king of Sweden?
- has it rained in Oviedo yesterday?
- when will YOUR phone fail? has THIS altimeter failed? is THIS camera not operating?
- what are the chances that Spain wins the next World Cup of football? Or Eurovision? Or anything???

$\Rightarrow$ can we still use probability to model these uncertainties?

Credal sets

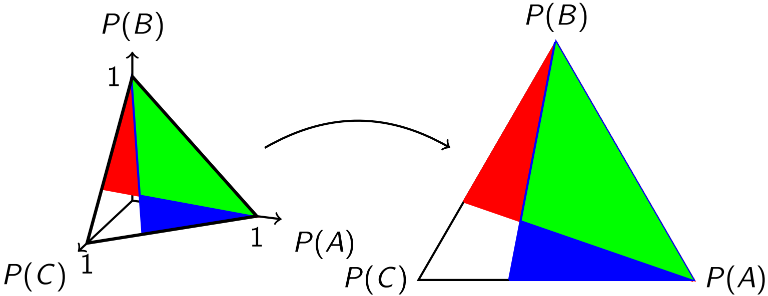
In a situation of imprecise information, we can then consider, instead of a probability measure, a set $\mathcal{M}$ of probability measures. Then for each event A we have a set of possible values $\{P(A) : P \in \mathcal{M}\}$. By taking lower and upper envelopes, we obtain the smallest and greatest values for $P(A)$ that are compatible with the available information:

$$\underline{P}(A) = \min_{P \in \mathcal{M}} P(A) \text{ and } \overline{P}(A) = \max_{P \in \mathcal{M}} P(A) \ \forall A \subseteq \Omega.$$

The two functions are conjugate: $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \subseteq \Omega$, so it suffices to work with $\underline{P}$.

Credal set example

$$X_1(A) = 0,\ X_1(B) = 20,\ X_1(C) = -10,\ \underline{P}(X_1) = 0$$
$$X_2(A) = 20,\ X_2(B) = -10,\ X_2(C) = -10,\ \underline{P}(X_2) = 0$$



Exercise

Before jumping off the wall, Humpty Dumpty tells Alice the following:

> *"I have a farm with pigs, cows and hens. There are at least as many pigs as cows and hens together, and at least as many hens as cows. How many pigs, cows and hens do I have?"*

- What are the probabilities compatible with this information?
- What is the lower probability of the set {hens, cows}?

Convexity: why (not)?

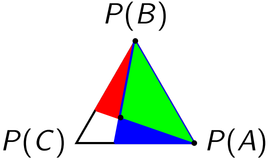
- The lower and upper envelopes of a set $\mathcal{M}$ of probability measures coincide with the lower and upper envelopes of its convex hull $CH(\mathcal{M})$.
- Closed convex sets are representable by their extreme points, and we have a number of mathematical tools at our disposal.
- However, the assumption of convexity is not always innocuous:
  - It may have implications when modelling independence.
  - It is also important in connection with preference modelling.

Extreme points

P is an extreme point of the credal set $\mathcal{M}$ when there are no $P_1 \neq P_2$ in $\mathcal{M}$ and $\alpha \in (0,1)$ such that $P = \alpha P_1 + (1-\alpha)P_2$.

For instance, the extreme points of the previous credal set on $\{A, B, C\}$ are:

- $p(A) = 1, p(B) = 0, p(C) = 0$
- $p(A) = 0, p(B) = 0, p(C) = 1$
- $p(A) = 0.25, p(B) = 0.5, p(C) = 0.25$



Credal sets and lower and upper probabilities

A set $\mathcal{M}$ of probability measures always determines a lower and an upper probability, but there may be different sets associated with the same $\underline{P}, \overline{P}$. The largest one is

$$\mathcal{M}(\underline{P}) := \{P : P(A) \geq \underline{P}(A) \ \forall A \subseteq \Omega\},$$

and we call it the credal set associated with $\underline{P}$. It holds that

$$\mathcal{M}(\overline{P}) := \{P : P(A) \leq \overline{P}(A) \ \forall A \subseteq \Omega\},$$

where $\overline{P}$ is the conjugate of $\underline{P}$.

Credal sets or lower probabilities?

In some cases, the easiest thing in practice is to determine the set $\mathcal{M}$ of probability measures compatible with the available information. This can be done with assessments such as comparative probabilities (A is more probable than B), linear constraints (the probability of A is at least 0.6), etc. Examples will appear in the lecture of Cassio de Campos.

Even if sets of probabilities are the primary model, it may be more efficient to work with the lower and upper probabilities they determine. These receive different names depending on the mathematical properties they satisfy.

Capacities

Let $\underline{P} : \wp(\Omega) \to [0,1]$. It is called a capacity or non-additive measure when it satisfies:

1. $\underline{P}(\emptyset) = 0, \underline{P}(\Omega) = 1$ (normalisation).
2. $A \subseteq B \Rightarrow \underline{P}(A) \leq \underline{P}(B)$ (monotonicity).

Capacities are also called fuzzy measures or Choquet capacities of the 1st order. When they are interpreted as lower (resp.,upper) bounds of a probability measure they are also called lower (resp., upper) probabilities.

Examples of properties of lower and upper probabilities

Among the properties that capacities may satisfy, we can consider some among 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

Consider non-additive measure $\underline{P}$ on $\wp(\Omega)$ and its conjugate $\overline{P}$:

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

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

Avoiding sure loss

A first assessment we can make on a lower probability $\underline{P}$ is that its associated credal set $\mathcal{M}(\underline{P})$ is non-empty. In that case, we say that $\underline{P}$ avoids sure loss.

For instance, if $\Omega = \{1, 2\}$ and we assess $\underline{P}(\{1\}) = \underline{P}(\{2\}) = 0.6$, the condition is not satisfied.

This is a minimal requirement if we want to interpret $\underline{P}$ as a summary of a credal set.

Some types of non-additive measures

- Coherent lower probabilities.
- 2-monotone capacities.
- Belief functions.
- Possibility/necessity measures.
- Probability boxes.

Coherent lower probabilities

We say that $\underline{P} : \mathcal{P}(\Omega) \to [0,1]$ is a coherent lower probability when

$$\underline{P}(A) = \min\{P(A) : P \in \mathcal{M}(\underline{P})\} \ \forall A \subseteq \Omega.$$

Equivalently, its conjugate function $\overline{P}$ satisfies

$$\overline{P}(A) = \max\{P(A) : P \in \mathcal{M}(\overline{P})\} \ \forall A \subseteq \Omega.$$

Example

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

$$\begin{array}{lll} \underline{P}(\{1\}) = 0.1 & \underline{P}(\{2\}) = 0.2 & \underline{P}(\{3\}) = 0.3 \\ \underline{P}(\{1,2\}) = 0.6 & \underline{P}(\{1,3\}) = 0.6 & \underline{P}(\{2,3\}) = 0.6. \end{array}$$

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

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 $\mathcal{M}(\underline{P})$.

Exercise

Mr. Play-it-safe is planning his upcoming holidays in the Canary Islands, and he is taking into account three possible disruptions: an unexpected illness (A), severe weather problems (B) and the unannounced visit of his mother in law (C).
He has assessed his lower and upper probabilities for these events:

	A	B	C	D
$\underline{P}$	0.05	0.05	0.2	0.5
$\overline{P}$	0.2	0.1	0.5	0.8

where D denotes the event 'Nothing bad happens'. He also assumes that no two disruptions can happen simultaneously.
Are these assessments coherent?

2-monotone capacities

Let $\underline{P}$ be a lower probability defined on $\mathcal{P}(\Omega)$. It is called a 2-monotone capacity when

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

for every pair of subsets A, B of Ω.

2-monotone capacities are also called submodular or convex.

They do not have an easy interpretation in the behavioural theory, but they possess interesting mathematical properties.

Properties (Denneberg, 1994)

- A 2-monotone capacity is a coherent lower probability (Walley, 1981).
- Let P be a probability measure and let $f : [0,1] \to [0,1]$ be a convex function with $f(0) = 0$. The lower probability given by $\underline{P}(\Omega) = 1$, $\underline{P}(A) = f(P(A))$ for every $A \neq \Omega$ is a 2-monotone capacity.

Example

A roulette has an unknown dependence between the red and black outcomes, in the sense that the first outcome is random but the second may depend on the first (with the same type of dependence in both cases). Let H_i="the i-th outcome is red", $i = 1, 2$.

Since $P(\text{red}) = P(\text{black}) = 0.5$ we should consider

$$\underline{P}(H_1) = \overline{P}(H_1) = \underline{P}(H_2) = \overline{P}(H_2) = 0.5$$
$$\underline{P}(H_1 \cap H_2) = 0, \overline{P}(H_1 \cap H_2) = 0.5, \underline{P}(H_1 \cup H_2) = 0.5, \overline{P}(H_1 \cup H_2) = 1.$$

Then $\underline{P}(H_1 \cup H_2) + \underline{P}(H_1 \cap H_2) = 0.5 < \underline{P}(H_1) + \underline{P}(H_2) = 1$, and our beliefs would not be 2-monotone.

Exercises

1. Let $\underline{P}$ a 2-monotone capacity defined on a field of sets $\mathcal{A}$, and let us extend it to $\wp(\Omega)$ by

$$\underline{P}_*(A) = \sup\{\underline{P}(B) : B \subseteq A\}.$$

   Show that $\underline{P}_*$ is also 2-monotone.

2. Consider $\Omega = \{1,2,3,4\}$, and let $\underline{P}$ be the lower envelope of the probabilities P_1, P_2 given by

$$P_1(\{1\}) = P_1(\{2\}) = 0.5, P_1(\{3\}) = P_1(\{4\}) = 0$$
$$P_2(\{1\}) = P_2(\{2\}) = P_2(\{3\}) = P_2(\{4\}) = 0.25.$$

   Show that $\underline{P}$ is not 2-monotone.

Further reading on 2-monotonicity

- P. Walley, *Coherent lower (and upper) probabilities*, Statistics Research Report. University of Warwick, 1981.
- D. Denneberg, *Non-additive measure and integral*, Kluwer, 1994.
- G. Choquet, *Theory of capacities*. Annales de l'Institute Fourier, 1953.
- G. de Cooman, M. Troffaes, E. Miranda, J. of Math. Analysis and Applications, 347(1), 133-146, 2009.

Belief functions (Shafer, 1976)

A lower probability $\underline{P}$ is called ∞-monotone when for every natural number n and every family $\{A_1, \ldots, A_n\}$ of subsets of Ω, it holds that

$$\underline{P}(A_1 \cup \ldots A_n) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \underline{P}(\cap_{i \in I} A_i). \tag{1}$$

When Ω is finite ∞-monotone capacities are called belief functions.

The evidential interpretation

Belief functions were mostly developed by Shafer starting from some works by Dempster in the 1960s. The belief of a set A, $\underline{P}(A)$, represents the existing evidence that supports A.

We usually assume the existence of a true (and unknown) state in Ω for the problem we are interested in. However, this does not imply that $\underline{P}$ is defined only on singletons, nor that it is characterised by its restriction to them.

Example

A crime has been committed and the police has two suspects, Chucky and Demian. An unreliable witness claims to have seen Chucky in the crime scene. We consider two possibilities: either (a) he really saw Chucky or (b) he saw nothing. In the first case, the list of suspects reduces to Chucky, and in the second it remains unchanged.

If we assign $P((a)) = \alpha, P((b)) = 1 - \alpha$, we obtain the belief function $\underline{P}$ given by
$\underline{P}(Chucky) = \alpha, \underline{P}(Demian) = 0, \underline{P}(Chucky, Demian) = 1.$

Probability measures

A particular case of belief functions are the probability measures. They satisfy Eq. (1) with $=$ for every n.

This implies that all the non-additive models we have seen so far (coherent lower probabilities, 2- and n-monotone capacities, belief functions) include as a particular case probability measures.

Basic probability assignment

When Ω is finite, we can give another representation of belief functions, using the so-called basic probability assignment.

A function $m : \wp(\Omega) \to [0,1]$ is called a basic probability assignment when it satisfies the following conditions:

1. $m(\emptyset) = 0$.
2. $\sum_{A \subseteq \Omega} m(A) = 1$.

Relationship with belief functions

- Given a basic probability assignment m, the function $\underline{P} : \wp(\Omega) \to [0,1]$ given by

$$\underline{P}(A) = \sum_{B \subseteq A} m(B)$$

is a belief function.

- If $\underline{P}$ is a belief function, the map $m : \wp(\Omega) \to [0,1]$ given by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \underline{P}(B)$$

is a basic probability assignment.

Moreover, this correspondence is one-to-one.

The function m is also called Möbius inverse of the belief function $\underline{P}$. The concept can also be applied to 2-monotone capacities. The function $m : \wp(\Omega) \to \mathbb{R}$ given by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \underline{P}(B)$$

is the Möbius inverse of $\underline{P}$, and it holds that $\underline{P}(A) = \sum_{B \subseteq A} m(B)$. Note that m need not take positive values only; in fact,

$$\underline{P} \text{ belief function} \iff m \text{ non-negative}$$

Focal elements

Given a lower probability $\underline{P}$ with Möbius inverse m, a subset A of Ω is called a focal element of m when $m(A) \neq 0$. In particular, the focal elements of a belief function are those sets for which $m(A) > 0$.

The focal elements are useful when working with a lower probability. In this sense, in game theory we have the so-called k-additive measures, which are those whose focal elements have cardinality smaller or equal than k.

Relationship with upper probabilities

Given a belief function $Bel : \wp(\Omega) \to [0,1]$, its conjugate $Pl : \wp(\Omega) \to [0,1]$, given by

$$Pl(A) = 1 - Bel(A^c),$$

is called a plausibility function.

Pl are Bel related to the same basic probability assignment, in the case of Pl by the formula

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B).$$

Exercises

Consider $\Omega = \{1, 2, 3\}$.

1. Let m be the basic probability assignment given by $m(\{1,2\}) = 0.5, m(\{3\}) = 0.2, m(\{2,3\}) = 0.3$. Determine the belief function associated with m.

2. Consider the belief function $\underline{P}$ given by $\underline{P}(A) = \frac{|A|}{3}$ for every $A \subseteq \Omega$. Determine its basic probability assignment.

Further reading on belief functions

- G. Shafer, *A mathematical theory of evidence*. Princeton, 1976.
- A. Demspter. Ann. of Mathematical Statistics, 38:325-339, 1967.
- R. Yager and L. Liu (eds.), *Classic works in the Dempster-Shafer theory of belief functions*. Studies in Fuzziness and Soft Computing 219. Springer, 2008.

...and the talk by Sébastien Destercke on Saturday!

Possibility and necessity measures (Dubois and Prade, 1988)

A possibility measure on Ω is a function $\Pi : \wp(\Omega) \to [0,1]$ such that

$$\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\} \; \forall A, B \subseteq \Omega.$$

The conjugate function of a possibility measure, given by $Nec(A) = 1 - \Pi(A^c)$, is called a necessity measure, and satisfies

$$Nec(A \cap B) = \min\{Nec(A), Nec(B)\}$$

for every $A, B \subseteq \Omega$.

Properties

- A possibility measure is a plausibility function, and a necessity measure is a belief function. They correspond to the case where the focal elements are nested.

- A possibility measure is characterised by its possibility distribution $\pi : \Omega \to [0, 1]$, which is given by $\pi(\omega) = \Pi(\{\omega\})$. It holds

$$\Pi(A) = \max_{\omega \in A} \pi(\omega) \ \forall A \subseteq \Omega$$

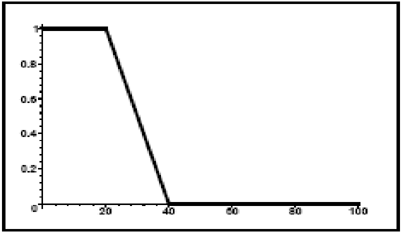
Exercise

Consider $\Omega = \{1, 2, 3, 4\}$.

1. Let Π be the possibility measure associated with the possibility distribution $\pi(1) = 0.3, \pi(2) = 0.5, \pi(3) = 1, \pi(4) = 0.7$. Determine its focal elements and its basic probability assignment.
2. Given the basic probability assignment $m(\{1\}) = 0.2, m(\{1,3\}) = 0.1, m(\{1,2,3\}) = 0.4, m(\{1,2,3,4\}) = 0.3$, determine the associated possibility measure and its possibility distribution.

Relationship with fuzzy sets

Let $\tilde{X} : \Omega \to [0,1]$ be a fuzzy set. We can interpret $\tilde{X}(\omega)$ as the degree of compatibility of ω with the concept described by $\tilde{X}$. On the other hand, given evidence of the type "Ω is $\tilde{X}$", $\tilde{X}(\omega)$ would be the possibility that Ω takes the value ω.



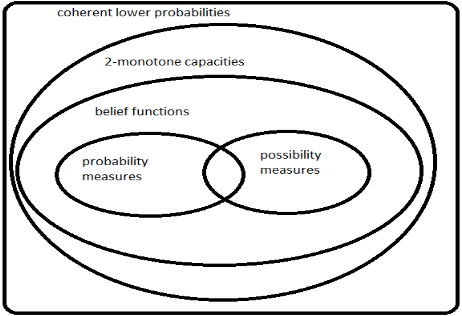
The possibility measure Π associated with the possibility distribution $\pi = \tilde{X}$ provides the possibility that Ω takes values in the set A.

Hence, in the previous figure $\Pi(A)$ would be the degree of possibility of the proposition "a young person's age belongs to the set A".

There are other interpretations of π in terms of likelihood functions, probability bounds, random sets, etc.

Relationships between the definitions

The relationships between the different types of lower and upper probabilities are summarised in the following figure:



Further reading on possibility theory

- D. Dubois and H. Prade, *Possibility theory*. Plenum, 1988.
- L. Zadeh, Fuzzy Sets and Systems, 1, 3-28, 1978.
- G. Shafer, *A mathematical theory of evidence*, Princeton, 1976.
- G. de Cooman, Int. J. of General Systems, 25, 291-371, 1997.

Distribution functions and p-boxes

We shall call a function $F : [0,1] \to [0,1]$ a distribution function when it satisfies the following two properties:

1. $\omega_1 \leq \omega_2 \Rightarrow F(\omega_1) \leq F(\omega_2)$ (monotonicity).
2. $F(1) = 1$ (normalisation).

A p-box is a pair of distribution functions, $(\underline{F}, \overline{F})$, satisfying $\underline{F}(\omega) \leq \overline{F}(\omega)$ for every $\omega \in [0,1]$.

The concept can be extended to arbitrary ordered spaces, producing then the so-called generalised p-boxes.

Particular cases

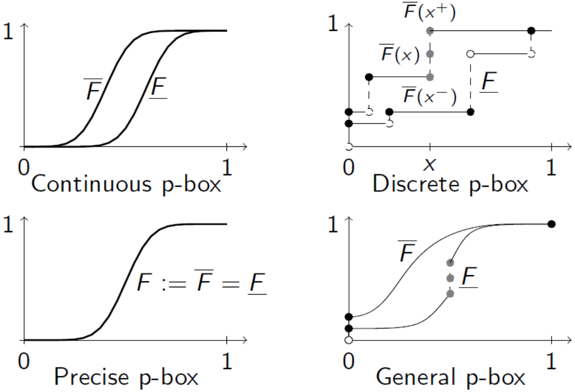
A p-box $(\underline{F}, \overline{F})$ is called:

- precise when $\underline{F} = \overline{F}$.
- continuous when both $\underline{F}, \overline{F}$ are continuous, meaning that

$$\underline{F}(\omega) = \sup_{\omega' < \omega} \underline{F}(\omega') = \inf_{\omega' > \omega} \underline{F}(\omega'), \overline{F}(\omega) = \sup_{\omega' < \omega} \overline{F}(\omega') = \inf_{\omega' > \omega} \overline{F}(\omega)$$

  for every $\omega \in [0, 1]$.
- discrete when both $\underline{F}, \overline{F}$ assume a countable number of different values.

Examples


1
$\overline{F}$
$\underline{F}$
0
1
Continuous p-box
1
$\overline{F}(x^+)$
$\overline{F}(x)$
$\overline{F}(x^-)$
$\underline{F}$
0
x
1
Discrete p-box
1
$F := \overline{F} = \underline{F}$
0
1
Precise p-box
1
$\overline{F}$
$\underline{F}$
0
1
General p-box

P-boxes as non-additive measures

A p-box can be represented as a lower probability $\underline{P}_{\underline{F},\overline{F}}$ on

$$\mathcal{K} = \{[0,\omega] : \omega \in [0,1]\} \cup \{(\omega,1] : \omega \in [0,1]\}$$

by

$$\underline{P}_{\underline{F},\overline{F}}([0,\omega]) := \underline{F}(\omega) \text{ and } \underline{P}_{\underline{F},\overline{F}}((\omega,1]) = 1 - \overline{F}(\omega).$$

- $\underline{P}_{\underline{F},\overline{F}}$ is a belief function.

Further reading on p-boxes

- S. Ferson, V. Kreinovich, I. Ginzburg, D. Mayers, K. Sentz. Technical report SAND2002-4015. 2003.
- M. Troffaes, S. Destercke. Int. J. of Approximate Reasoning, 52(6), 767-791, 2011.
- R. Pelessoni, P. Vicig, I. Montes, E. Miranda. IJUFKS, 24(2), 229-263, 2016.

...and the talk by Scott Ferson tomorrow!

Gambles

A function $X : \Omega \to \mathbb{R}$ is called a gamble.
If we specify a probability measure P on $\wp(\Omega)$, it uniquely determines its expectation on any gamble $X : \Omega \to \mathbb{R}$:

$$E_p(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

Similary, if we have a set of probabilities $\mathcal{M}$, it determines lower and upper expectations:

$$\underline{E}(X) = \min_{P \in \mathcal{M}} E_p(X) \text{ and } \overline{E}(X) = \max_{P \in \mathcal{M}} E_p(X).$$

Coherent lower previsions

If we define $\mathcal{L}(\Omega) := \{X : \Omega \to \mathbb{R}\}$, a coherent lower prevision on $\mathcal{L}(\Omega)$ is a function $\underline{P}$ such that

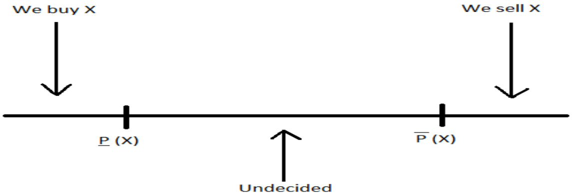
- $\underline{P}(X) \geq \min X$
- $\underline{P}(\lambda X) = \lambda \underline{P}(X)$
- $\underline{P}(X + Y) \geq \underline{P}(X) + \underline{P}(Y)$

for every $X, Y \in \mathcal{L}(\Omega)$ and every $\lambda > 0$.

They can be given a behavioural interpretation in terms of acceptable buying prices.

The behavioural interpretation

The lower prevision of X can be understood as the supremum acceptable buying price for X: $X - \mu$ is desirable for any $\mu < \underline{P}(X)$.
Similarly, the upper prevision of X would be the infimum acceptable selling price for X: $\mu - X$ is desirable for any $\mu > \overline{P}(X)$.



Exercise

Let $\underline{P}$ be the lower prevision on $\mathcal{L}(\{1,2,3\})$ given by

$$\underline{P}(X) = \frac{\min\{X(1), X(2), X(3)\}}{2} + \frac{\max\{X(1), X(2), X(3)\}}{2}.$$

Is it coherent?

Does it matter?

In general, coherent lower previsions are more expressive than coherent lower probabilities:

- Although the restriction to indicators of events of a coherent lower prevision is a coherent lower probability, a coherent lower probability may have more than one extension to gambles.
- There is a one-to-one correspondence between coherent lower previsions and convex sets of probability measures.
- However, the credal sets determined by a coherent lower probability are not as general: they always have a finite number of extreme points, for instance.

For these reasons, we may work with coherent lower previsions as the primary model.

Extension from events to gambles

Given a coherent lower probability $\underline{P} : \wp(\Omega) \to [0,1]$, its natural extension to $\mathcal{L}(\Omega)$ is

$$\underline{E}(X) := \min\{E_p(X) : P(A) \geq \underline{P}(A)\ \forall A \subseteq \Omega\}.$$

It is the smallest coherent lower prevision on $\mathcal{L}(\Omega)$ that agrees with $\underline{P}$ on $\wp(\Omega)$.

- When $\underline{P}$ is 2-monotone, the natural extension can be computed with the Choquet integral: we have

$$\underline{E}(X) = \sum_{i=1}^{n} (X(\omega_i) - X(\omega_{i+1}))\underline{P}(\{\omega_1, \dots, \omega_i\}),$$

where $X(\omega_1) \geq X(\omega_2) \geq \cdots \geq X(\omega_n)$, and with $X(\omega_{n+1}) = 0$.

Exercise

Let $\underline{P}_A$ be the vacuous lower probability relative to a set A, given by the assessment $\underline{P}_A(A) = 1$.

Prove that the natural extension $\underline{E}$ of $\underline{P}_A$ is equal to the vacuous lower prevision relative to A:

$$\underline{E}(X) = \min_{\omega \in A} X(\omega),$$

for any $X \in \mathcal{L}(\Omega)$.

Further reading on coherent lower previsions

- P. Walley, *Statistical reasoning with imprecise probabilities*. Chapman and Hall, 1991.
- T. Augustin, F. Coolen, G. de Cooman, M. Troffaes (eds.), *Introduction to imprecise probabilities*. Wiley, 2014.
- M. Troffaes, G. de Cooman, *Lower previsions*. Wiley, 2014.

Sets of desirable gambles

If model the available information with a set $\mathcal{M}$ of probability measures, we can consider the non-additive measure it induces (a coherent lower probability) or the expectation operator it determines (a coherent lower prevision).

Equivalently, we can assess which gambles we consider desirable or not.

In the precise case, we say that X is desirable when its expectation is positive.

How to convey this idea with imprecision?

Rationality axioms for sets of desirable gambles

If we consider a set of gambles that we find desirable, there are a number of rationality requirements we can consider:

- A gamble that makes us lose money, no matter the outcome, should not be desirable, and a gamble which never makes us lose money should be desirable.
- A change of utility scale should not affect our desirability.
- If two transactions are desirable, so should be their sum.

These ideas define the notion of coherence for sets of gambles.

Coherence of sets of desirable gambles

A set of desirable gambles is coherent if and only if

(D1) If $X \leq 0$, then $X \notin \mathcal{D}$.

(D2) If $X \gneq 0$, then $X \in \mathcal{D}$.

(D3) If $X, Y \in \mathcal{D}$, then $X + Y \in \mathcal{D}$.

(D4) If $X \in \mathcal{D}, \lambda > 0$, then $\lambda X \in \mathcal{D}$.

Exercise

Let $\Omega = \{1, 2, 3\}$, and consider the following sets of desirable gambles:

$$\mathcal{D}_1 := \{X : X(1) + X(2) + X(3) > 0\}.$$
$$\mathcal{D}_2 := \{X : \max\{X(1), X(2), X(3)\} > 0\}.$$

Is $\mathcal{D}_1$ coherent? And $\mathcal{D}_2$?

Connection with coherent lower previsions

- If $\mathcal{D}$ is a coherent set of gambles, then the lower prevision it induces by
$$\underline{P}(X) = \sup\{\mu : X - \mu \in \mathcal{D}\}$$
is coherent.
- Conversely, a coherent lower prevision $\underline{P}$ determines a coherent set of desirable gambles by $\mathcal{D} := \{X : \underline{P}(X) > 0\} \cup \{X \gneq 0\}$.

Hence, we have three equivalent representations of our beliefs:

1. Coherent lower previsions.
2. Closed and convex sets of probability measures.
3. Coherent sets of desirable gambles.

In fact, sets of desirable gambles have an extra layer of expressivity that helps dealing with the problem of conditioning on sets of probability zero.

Connection with 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 is one of the (many) possible optimality criteria when we want to establish our preferences with imprecise probabilities.

More of these will appear in the lecture of Matthias Troffaes on Thursday.

Further reading on sets of desirable gambles

- I. Couso, S. Moral. Int. J. of Appr. Reasoning, 52(7):1034-1055, 2011.
- E. Miranda, M. Zaffalon. Ann. Math. Artif. Intelligence, 60(3-4):251-309, 2010.
- E. Quaeghebeur. *Introduction to imprecise probabilities*, chapter 1. Wiley, 2014.

Notation

- **Joint probability:** $\mu : \wp(\Omega_1 \times \Omega_2) \to [0,1]$
- The **marginal probability** of μ **on** Ω_1 is $\mu_1 : \wp(\Omega_1) \to [0,1]$ defined as:

$$\mu_1(A) = \mu(A \times \Omega_2), \; \forall A \in \wp(\Omega_1).$$

- The **marginal probability** of μ **on** Ω_2 is $\mu_2 : \wp(\Omega_2) \to [0,1]$ defined as follows:

$$\mu_2(B) = \mu(\Omega_1 \times B), \; \forall B \in \wp(\Omega_2).$$

- For the sake of simplicity, Ω_1 and Ω_2 are assumed to be finite.

Stochastic independence in Probability Theory

- **Independent events:** A and B are independent if (three equivalent conditions):
  - $\mu(A \cap B) = \mu(A) \cdot \mu(B)$ or
  - $\mu(A|B) = \mu(A)$, if $\mu(B) > 0$ or
  - $\mu(B|A) = \mu(B)$, if $\mu(A) > 0$.
- **Independent variables:** X and Y are independent random variables if $X^{-1}(A)$ $Y^{-1}(B)$ are independent events for every $A \in \mathcal{X}$, $B \in \mathcal{Y}$.
- **Product probability:** μ is a "product probability" when

$$\mu(A \times B) = \mu_1(A) \cdot \mu_2(B), \ \forall A \in \wp(\Omega_1), B \in \wp(\Omega_2)$$

(Equivalently, when $A \times \Omega_2$ and $\Omega_1 \times B$ are independent events -wrt μ-, $\forall A \in \wp(\Omega_1), B \in \wp(\Omega_2)$.)

Notation

- Credal set $\mathcal{M}$.
- Marginal credal sets
  - $\mathcal{M}_i = \{\mu_i : \wp(\Omega_i) \to \mathbb{R} \,:\, \mu \in \mathcal{M}\}\ i = 1, 2.$
- Joint credal set associated to a marginal credal set
  - $\mathcal{M}_i^* = \{\mu : \wp(\Omega_1 \times \Omega_2) \to \mathbb{R} \,:\, \mu_i \in \mathcal{M}_i\}\ i = 1, 2.$
- $CH(\mathcal{P})$: convex hull of a (non-necessarily convex) set of probability measures.

Outline

- Independence conditions expressed in terms of $\mathcal{M}$, $\mathcal{M}_1$ and $\mathcal{M}_2$.
- Construction of the largest (joint) credal set satisfying certain independence condition from a pair of marginal credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$.

Independence concepts in Imprecise Probability

- Epistemic irrelevance
- Epistemic independence
- Independence in the selection

Epistemic irrelevance

Consider the (joint) credal set $\mathcal{M}$ on $\Omega_1 \times \Omega_2$. Consider an arbitrary $\mu \in \mathcal{M}$ and denote:

- $\mu_{2|\omega_1}$ the probability measure on Ω_2 defined as:

$$\mu_{2|\omega_1}(A) = \mu(\Omega_1 \times A | \{\omega_1\} \times \Omega_2), \ \forall A \subseteq \Omega_2.$$

- $\mathcal{M}_{2|\omega_1} = \{\mu_{2|\omega_1} : \mu \in \mathcal{M}\}, \forall \omega_1 \in \Omega_1.$

We say that the first experiment is *epistemically irrelevant* to the second one when $\mathcal{M}_{2|\omega_1} = \mathcal{M}_2, \ \forall \omega_1 \in \Omega_1$.

Irrelevant natural extension

Consider two credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ on Ω_1 and Ω_2 respectively. The largest credal set $\mathcal{M}$ under which the first experiment is epistemically irrelevant to the second, i.e, the set of joint distributions μ for which:

- $\mu_1 \in \mathcal{M}_1$
- $\mu_{2|\omega_1} \in \mathcal{M}_2, \ \forall \omega_1 \in \Omega_1$

is called the *irrelevant natural extension*.

Exercise

- We have three urns. Each of them has 10 balls which are coloured either red or white.
- Urn 1: 5 red, 2 white, 3 unknown; Urns 2 and 3: 3 red, 3 white, 4 unknown (not necessarily the same composition).
- A ball is randomly selected from the first urn.
- If the first ball is red then the second ball is selected randomly from the second urn, and if the first ball is white then the second ball is selected randomly from the third urn.

Exercise (cont.)

- Our uncertainty about the pair of colours is modelled by a set of joint probabilities of the form μ where
  - $\mu(\{(r, \omega_2)\}) = \mu_1(\{r\})\mu_{2|r}(\{\omega_2\}), \omega_2 \in \{r, w\}$
  - $\mu(\{(w, \omega_2)\}) = \mu_1(\{w\})\mu_{2|w}(\{\omega_2\}), \omega_2 \in \{r, w\}$,

  with
  - $0.5 \leq \mu_1(\{r\}) \leq 0.8$,
  - $0.3 \leq \mu_{2|r}(\{\omega_2\}) \leq 0.7$ and
  - $0.3 \leq \mu_{2|w}(\{\omega_2\}) \leq 0.7$.
- The above set has eight extreme points, each of them determined by a combination of the extremes of the marginal on Ω_1 and the two conditionals.

Determine the collection of eight extreme points of the above set.

Exercise

Consider the last exercise where our uncertainty about the pair of colours is modelled by a set of joint probabilities of the form μ where

- $\mu(\{(r,\omega_2)\}) = \mu_1(\{r\})\mu_{2|r}(\{\omega_2\}), \omega_2 \in \{r,w\}$
- $\mu(\{(w,\omega_2)\}) = \mu_1(\{w\})\mu_{2|w}(\{\omega_2\}), \omega_2 \in \{r,w\}$,

with

- $0.5 \leq \mu_1(\{r\}) \leq 0.8$,
- $0.3 \leq \mu_{2|r}(\{\omega_2\}) \leq 0.7$ and
- $0.3 \leq \mu_{2|w}(\{\omega_2\}) \leq 0.7$.

Calculate the upper probability that the first ball is red, given the colour of the second ball. Does the collection of conditional probabilities $\mathcal{M}_{1|r} = \{\mu_{1|r} : \mu \in \mathcal{M}\}$ coincide with $\mathcal{M}_1$?

Epistemic independence and independent natural extension

Consider the (joint) credal set $\mathcal{M}$ on $\Omega_1 \times \Omega_2$. We say that the two experiments are *epistemically independent* when each one is epistemically irrelevant to the other. The independent natural extension $\mathcal{M}$ can be constructed as the intersection of two irrelevant natural extensions.

Independence in the selection and strong independence

We say that there is independence in the selection when every extreme point μ of $\mathcal{M}$ factorizes as $\mu = \mu_1 \otimes \mu_2$. $\mathcal{M}$ satisfies strong independence if it can be expressed as:

$$\mathcal{M} = CH(\{\mu_1 \otimes \mu_2 \ : \ \mu_1 \in \mathcal{M}_1, \mu_2 \in \mathcal{M}_2\}).$$

Exercise

- Assume that we have two urns with the following composition: Urn 1: 5 red, 2 white, 3 unknown; Urn 2: 3 red, 3 white, 4 unknown;
- the 7 balls in the two urns whose colours are unknown are all the same colour;
- the drawings from the two urns are stochastically independent.

Determine the convex hull of the set of probabilities that is compatible with the above information. Does it satisfy independence in the selection? Does it satisfy strong independence?

Exercise

- Assume that we have two urns with the following composition: Urn 1: 5 red, 2 white, 3 unknown; Urn 2: 3 red, 3 white, 4 unknown;
- the drawings from the two urns are stochastically independent.

Determine the convex hull of the set of probabilities that is compatible with the above information. Does it satisfy independence in the selection? Does it satisfy strong independence?

Independence of the marginal sets

$\mathcal{M}$ satisfies *independence of the marginal sets* if for any $\mu_1 \in \mathcal{M}_1$ and any $\mu_2 \in \mathcal{M}_2$ there exists $\mu \in \mathcal{P}$ whose marginals are μ_1 and μ_2.

Exercise

Consider the product possibility space $\Omega = \Omega_1 \times \Omega_2$ where $\Omega_1 = \Omega_2 = \{r, w\}$. Consider the credal set $\mathcal{M} = CH(\{\mu, \mu'\})$ where $\mu = (0.01, 0.09, 0.09, 0.81)$ and $\mu' = (0.81, 0.09, 0.09, 0.01)$. Is independence of the marginal sets satisfied?

Unknown interaction

Consider two credal sets $\mathcal{M}_1$ and any $\mathcal{M}_2$ on Ω_1 and Ω_2, respectively. Let $\mathcal{M}_1^*$ and $\mathcal{M}_2^*$ denote the (convex) collections of joint probability measures:

$$\mathcal{M}_1^* = \{\mu : \mu_1 \in \mathcal{M}_1\},\ \mathcal{M}_2^* = \{\mu : \mu_2 \in \mathcal{M}_2\}.$$

The largest credal set induced by $\mathcal{M}_1$ and $\mathcal{M}_2$ and satisfying independence of the marginal sets is $\mathcal{M}_1^* \cap \mathcal{M}_2^*$. If $\mathcal{M} = \mathcal{M}_1^* \cap \mathcal{M}_2^*$ we say that there is *unknown interaction*.

Exercise

- We have two urns. Each of them has 10 balls which are coloured either red or white.
- Urn 1: 5 red, 2 white, 3 unknown; Urn 2: 3 red, 3 white, 4 unknown.
- One ball is chosen at random from each urn. We have no information about the interaction between the two drawings.

Exercise (cont.)

(a) Determine the marginal credal set on $\Omega_1 = \{r, w\}$ characterizing our incomplete information about the first drawing. Denote it $\mathcal{M}_1$.

(b) Determine the marginal credal set on $\Omega_2 = \{r, w\}$ characterizing our incomplete information about the second drawing. Denote it $\mathcal{M}_2$.

(c) Consider the joint possibility space $\Omega = \{rr, rw, wr, ww\}$ and the joint probability $\mu = (0.2, 0.4, 0.3, 0.1)$ defined on it.

- Check that it belongs to the set $\mathcal{M} = \mathcal{M}_1^* \cap \mathcal{M}_2^*$.
- Design a random experiment compatible with the above incomplete information associated to this joint probability.

Exercise

Consider the last example of two urns. Characterize our uncertainty about the color of both balls by means of a credal set on $\Omega = \Omega_1 \times \Omega_2$.

Set-valued data

- Consider a two-dimensional random vector (X, Y) representing a pair of attributes.
- Suppose that we are provided with set-valued information about each outcome of X and Y.
- Let the random set Γ_X (resp. Γ_Y) represent our incomplete information about X (resp. about Y).
- Information about $X(\omega)$ (resp. about $Y(\omega)$) : $X(\omega) \in \Gamma_X(\omega)$ (resp. $Y(\omega) \in \Gamma_Y(\omega)$).
- Let m_1 and m_2 respectively denote the bma induced by Γ_X and Γ_Y.
- Let m denote the bma associated to $\Gamma = \Gamma_X \times \Gamma_Y$.

Random set independence vs independence in the selection

There is *random set independence* when:

$$m(A \times B) = m_1(A) \cdot m_2(B), \ \forall \ A \in \wp(\Omega_1), B \in \wp(\Omega_2)$$

(Or, equivalently, when the two random sets Γ_X and Γ_Y are stochastically independent, i.e:

$$P(\Gamma_X = A, \Gamma_Y = B) = P(\Gamma_X = A) \cdot P(\Gamma_Y = B),$$

$\forall \ A \in \wp(\Omega_1), B \in \wp(\Omega_2)$).

Example

- We have two urns. Each of them has 10 balls which are coloured either red or white.
- Urn 1: 5 red, 2 white, 3 unpainted; Urn 2: 3 red, 3 white, 4 unpainted.
- One ball is chosen at random from each urn.
- (If they have no colour, there may be arbitrary correlation between the colours they are finally assigned).

Example (cont.). Random sets

- Information about the color of the 1st ball:
  $P(\Gamma_X = \{r\}) = 0.5$, $P(\Gamma_X = \{w\}) = 0.2$,
  $P(\Gamma_X = \{r, w\}) = 0.2$.
- Information about the color of the 2nd ball:
  $P(\Gamma_Y = \{r\}) = 0.3$, $P(\Gamma_Y = \{w\}) = 0.3$,
  $P(\Gamma_Y = \{r, w\}) = 0.4$.
- Information about the pair of colors:
  $P(\Gamma_X = A_1, \Gamma_Y = A_2) = P(\Gamma_X = A_1) \cdot P(\Gamma_Y = A_2)$.

Example (cont.). Marginal and joint mass functions

m_1, m_2 and m respectively represent the mass functions of Γ_X, Γ_Y and $\Gamma = \Gamma_X \times \Gamma_Y$.

- Urn 1: $m_1(\{r\}) = 0.5, m_1(\{w\}) = 0.2, m_1(\{w, r\}) = 0.3$.
- Urn 2: $m_2(\{r\}) = 0.3, m_2(\{w\}) = 0.3, m_2(\{w, r\}) = 0.4$.
- Joint mass function $m(A_1 \times A_2) = m_1(A_1) \cdot m_2(A_2)$.

Exercise. Upper and lower bounds for the conditional probabilities

- Check that the focal sets of the joint mass function m are the following nine sets: $\{rr\}$, $\{rw\}$, $\{rr, rw\}$, $\{wr\}$, $\{ww\}$, $\{wr, ww\}$, $\{rr, wr\}$, $\{rw, ww\}$, $\{rr, rw, wr, ww\}$.
- Determine the mass values associated to those 9 focal sets.
- Consider the credal set $\mathcal{M}$ associated to m and calculate the minimum possible value for the conditional probability $\mu(\{r, w\} \times \{r\}|\{r\} \times \{r, w\})$:
$\min\{\mu(\{r, w\} \times \{r\}|\{r\} \times \{r, w\}) \mid \mu \in \mathcal{M}\} =$
$\min\left\{\frac{\mu(\{r,r\})}{\mu(\{rr,rw\})} \mid \mu \in \mathcal{M}\right\}$.
- Does the above minimum coincide with 0.3?

Exercise: Random set independence vs independence in the selection (I. Couso, D. Dubois and L. Sánchez, 2014)

- A light sensor displays numbers between 0 and 255.
- 10 measurements per second.
- If the brightness is higher than a threshold (255), the sensor displays 255 during 3/10s.

Complete the following table, about six consecutive measurements, where the actual values of brightness are independent from each other:

actual values	215	150	200	300	210	280
displayed quantities	215	150	200	255	__	__
set-valued information	{215}	{150}	__	__	__	__

Exercise (cont.)

Consider the information provided about the six measurements of a light sensor:

actual values	215	150	200	300	210	280
displayed quantities	215	150	200	255	__	__
set-valued information	$\{215\}$	$\{150\}$	__	__	__	__

Let Γ_i denote the random set that represents the (set-valued) information provided by the sensor in the i-th measurement. What is the value of the following conditional probability?:

$$P(\Gamma_i \supseteq [255,\infty) | \Gamma_{i-1} \supseteq [255,\infty), \Gamma_{i-2} \not\supseteq [255,\infty)).$$

Exercise: Random set independence vs independence in the selection (I. Couso, D. Dubois and L. Sánchez, 2014)

The random variables X_0 and Y_0 respectively represent the temperature (in $^{\mathrm{o}}$C) of an ill person taken at random in a hospital just before taking an antipyretic (X_0) and 3 hours later (Y_0). The random set Γ_1 represents the information about X_0 using a very crude measure (it reports always the same interval $[37, 39.5]$). The random set Γ_2 represents the information about Y_0 provided by a thermometer with $+/-0.5$ $^{\mathrm{o}}$C of precision.

(a) Are X_0 and Y_0 stochastically independent?

(b) Are Γ_1 and Γ_2 stochastically independent?

Alternative nomenclature

- Strict independence.- Cozman (2008) says that there is "strict independence" when every joint probability in the set can be factorized as the product of its marginals. This condition violates convexity. It has not been explicitly considered here.
- Independence in the selection.- Cozman (2008) calls it "strong independence". Campos and Moral (1995) call it "type 2 independence".
- Strong independence. Cozman (2008) calls it "strong extension". Walley (1991) calls it "type 1 extension". Campos and Moral (1995) call it "type 3 independence".
- Epistemic irrelevance.- Smith (1961) calls it "independence".

Further reading

All the notions reviewed here can be found in:

- I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. Risk, Decision and Policy, 5:165-181, 2000.
- I. Couso, D. Dubois, L. Sánchez, Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables: An Introduction for Ph.D. Students and Practitioners, Springer, 2014.

Further reading

- I. Couso, S. Moral, Independence concepts in evidence theory, International Journal of Approximate Reasoning 51: 748-758, 2010.
- F. G. Cozman, Sets of Probability Distributions and Independence, Technical Report presented at the 3rd edition of the SIPTA School (2008).
- L.M. de Campos and S. Moral. Independence concepts for convex sets of probabilities. In Conf. on Uncertainty in Artificial Intelligence, pages 108-115, San Francisco, California, 1995.
- V. P. Kuznetsov. Interval Statistical Methods. Radio i Svyaz Publ., (in Russian), 1991.
- P. Walley. Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London, 1991.