Special Cases

Sebastien Destercke, Didier Dubois

Special cases

S. Destercke, D. Dubois

Tuesday 3rd June, 2014

Contents

Chapter 1

Special Cases

by Sébastien Destercke, Didier Dubois

1.1 Introduction

As argued in Chapters **??** and **??**, lower previsions and sets of desirable gambles are very general models of uncertainty that have solid foundations and a clear behavioural interpretation. However, this generality goes along with a high computational complexity and a difficulty to easily explain such representations to users that are not experts in imprecise probability theories.

Therefore, in practical applications, simplified representations can greatly improve the applicability of imprecise probability theories. There are three main reasons for using simplified representations:

- to facilitate the elicitation or information collection process;
- to improve the computational tractability of mathematical models;
- to improve the interpretability of results when answering some particular question of interest.

The main objection to the use of such representations, or more precisely for restricting oneself to them, is that they may not be general enough to *exactly* model the available information. Moreover, even if some pieces of information can be exactly modelled by such simplified representations, further processing may result in information no longer exactly representable in a simple form (see Walley [54], for example). Indeed, if one works with processing tools specific to coherent lower previsions (and therefore closed for such representations), one can hardly expect more specific properties to be preserved (this remains true for other theories, see Chapter **??**). Nevertheless, summarising complex outputs in the form of simplified representations may make these outputs easier to interpret.

In this chapter, we will review several simplified representations that are special cases of coherent lower previsions (actually, most of them are coherent lower probabilities). To do so, we will adopt the languages of the previous chapters. For each representation, we emphasise its interests and limitations.

1.2 Capacities and n-monotonicity

Capacities are set functions defined over the events of a possibility space $\mathcal{X}$ that can be used to represent uncertainty about an experiment.

Definition 1.1. Given a finite possibility space[1] $\mathcal{X}$, a capacity on $\mathcal{X}$ is a function g, defined on $\wp(\mathcal{X})$, such that:

- $g(\emptyset) = 0$, $g(\mathcal{X}) = 1$, and
- for all $A, B \subseteq \mathcal{X}$, $A \subseteq B$ implies $g(A) \leq g(B)$ (monotonicity property).

The monotonicity property stresses the fact that if $x \in A$ implies $x \in B$ then one cannot have more confidence in A than in B. Capacities were first introduced by Choquet [7], and are also known as fuzzy measures [39].

A capacity is said to be *super-additive* if the property

$$A \cap B = \emptyset \implies g(A \cup B) \geq g(A) + g(B) \tag{1.1}$$

holds for all events $A, B \subseteq \mathcal{X}$. The dual notion, called *sub-additivity*, is obtained by reversing the inequality. In particular, coherent lower and upper probabilities are super-additive and sub-additive capacities, respectively. Among capacities, those satisfying n-monotonicity are particularly interesting.

Definition 1.2. Let $n \in \mathbb{N}_0$, $n \geq 2$. A capacity g is said to be *n-monotone* whenever for any collection $\mathcal{A}_n \subseteq \wp(\mathcal{X})$ of n events, it holds that

$$g\left(\bigcup_{A \in \mathcal{A}_n} A\right) \geq \sum_{\emptyset \neq \mathcal{A}' \subseteq \mathcal{A}_n} (-1)^{|\mathcal{A}'|+1} g\left(\bigcap_{A \in \mathcal{A}'} A\right)$$

A capacity g is said to be ∞-*monotone*, totally monotone or *completely monotone*, whenever it is n-monotone for every $n \in \mathbb{N}_0$, $n \geq 2$. [2]

If a capacity is n-monotone, then it is m-monotone for all $m \in \mathbb{N}_0$, $2 \leq m \leq n$. Given any capacity g defined on a finite possibility space $\mathcal{X}$, we can define its Möbius inverse:

Definition 1.3. The Möbius inverse $m_g \colon \wp(\mathcal{X}) \to \mathbb{R}$ of a capacity g is defined, for every event $E \subseteq \mathcal{X}$, as

$$m_g(E) := \sum_{A \subseteq E} (-1)^{|E \setminus A|} g(A). \tag{1.2}$$

[1] (Pre-)Capacities on infinite spaces are discussed in the context of game-theoretic probability, see page **??**.

[2] The conjugated capacity $\widetilde{g}$ with $\widetilde{g}(A) := 1 - g(A^c)$ for all $\subseteq \Omega$ is then *p-alternating*.

Conversely, from a Möbius inverse m_g, one can retrieve the value of $g(A)$ on any event A by computing

$$g(A) = \sum_{E \subseteq A} m_g(E). \tag{1.3}$$

Chateauneuf and Jaffray [6] have shown that the Möbius inverse m_g and the n-monotonicity of the capacity g are related in the following way: if a capacity g is n-monotone, then its Möbius inverse m_g is positive for every subset $A \subseteq \mathcal{X}$ such that $|A| \leq n$. Also note that a Möbius inverse m_g is always such that $\sum_{E \subseteq \mathcal{X}} m_g(E) = 1$. Two specific kinds of capacities are of particular importance for practical purposes: 2-monotone and ∞-monotone capacities.

1.3 2-monotone capacities

A 2-monotone capacity, that is a capacity that satisfies for every pair of events $A, B \subseteq \mathcal{X}$ the inequality $g(A \cup B) + g(A \cap B) \geq g(A) + g(B)$, is a coherent lower probability $\underline{P}$. This means that it induces a non-empty credal set $\mathcal{M}(\underline{P})$, of which it is the lower envelope on events. In contrast, satisfying only super-additivity (1.1) does not guarantee to have $\mathcal{M}(\underline{P}) \neq \emptyset$ (for a counter-example, see Papamarcou and Fine [42]). Due to their practical interest, they have received particular attention in the literature [6, 53, 36, 5, 8]. When $\mathcal{X}$ is finite, 2-monotone capacities have the following interesting properties (all detailed in [6]):

- The natural extension $\underline{E}(f)$ of a 2-monotone lower probability $\underline{P}$ for any bounded gamble f is given by the Choquet integral (see Section **??** on page **??** for details on the Choquet integral).
- Let $\mathcal{X} = \{x_1, \ldots, x_k\}$, and let Σ denote the set of all permutations of $\{1, \ldots, k\}$. For any permutation $\sigma \in \Sigma$, we can define a probability distribution P_σ as follows: for any $i \in \{1, \ldots, k\}$,

  $$P_\sigma(\{x_{\sigma(i)}\}) \coloneqq \underline{P}(A_{\sigma(i)}) - \underline{P}(A_{\sigma(i-1)}),$$

  with $A_0 = \emptyset$ and $A_{\sigma(i)} = \{x_{\sigma(1)}, \ldots, x_{\sigma(i)}\}$. Then, the set of extreme points of the convex set $\mathcal{M}(\underline{P})$ is given by $\mathrm{ext}(\mathcal{M}(\underline{P})) = \{P_\sigma : \sigma \in \Sigma\}$.

The above properties allow performing various computational tasks (coherence checking, statistical testing[3], ...) in an easier way than with generic coherent lower previsions. However, 2-monotone lower probabilities still require, in the general case, to store $2^{|\mathcal{X}|}$ values. Representations used in practice will, however, be of reduced complexity.

1.4 Probability intervals on singletons

Probability intervals on singletons [11, 55] are popular representations that play an important role in graphical models (See Chapter **??**) and in classification procedures (See

[3]See the so-called Huber-Strassen theory on generalized Neyman-Pearson testing outlined in Section **??**.

Chapter **??**). They are defined as lower and upper probabilities specified on singletons of a finite possibility space $\mathcal{X}$. That is, they correspond to assessments $\underline{P}(\{x\})$ and $\overline{P}(\{x\}) = 1 - \underline{P}(\{x\}^c)$ for every $x \in \mathcal{X}$.

De Campos *et al.* [11] have shown that probability intervals $\underline{P}$ avoid sure loss whenever

$$\sum_{x\in\mathcal{X}} \underline{P}(\{x\}) \leq 1 \leq \sum_{x\in\mathcal{X}} \overline{P}(\{x\}) \tag{1.4}$$

and that $\underline{P}$ is coherent when, for each $x \in \mathcal{X}$, the two following inequalities

$$\overline{P}(\{x\}) + \sum_{y\in\mathcal{X}\setminus\{x\}} \underline{P}(\{y\}) \leq 1 \text{ and } \underline{P}(\{x\}) + \sum_{y\in\mathcal{X}\setminus\{x\}} \overline{P}(\{y\}) \geq 1 \tag{1.5}$$

hold. Compared to generic algorithms of Chapter **??**, these criteria are easy to check. Provided $\underline{P}$ avoids sure loss (satisfies Eq. (1.4)), De Campos *et al.* [11] also describe efficient methods to compute probability intervals that are coherent. From now on, we will only consider such coherent probability intervals.

The natural extension $\underline{E}$ of probability intervals $\underline{P}$ to any event $A \subseteq \mathcal{X}$ can easily be computed using the following formulas:

$$\underline{E}(A) = \max\left\{\sum_{x\in A} \underline{P}(\{x\}), 1 - \sum_{x\in A^c} \overline{P}(\{x\})\right\}, \tag{1.6a}$$

$$\overline{E}(A) = 1 - \underline{E}(A^c) = \min\left\{\sum_{x\in A} \overline{P}(\{x\}), 1 - \sum_{x\in A^c} \underline{P}(\{x\})\right\}. \tag{1.6b}$$

It can also be shown that probability intervals are 2-monotone, allowing one to use corresponding computational tools (De Campos *et al.* [11] also propose various algorithms optimized for probability intervals).

From any lower prevision $\underline{P}$, one can easily get outer-approximating probability intervals by computing the natural extension of $\underline{P}$ over events $\{x\}$ and $\{x\}^c$. Note that only $2|\mathcal{X}|$ values need to be stored to describe probability intervals, instead of the $2^{|\mathcal{X}|}$ needed values for generic 2-monotone capacities. As usual, this complexity reduction goes along with a limited expressive power (i.e., not all 2-monotone capacities can be expressed by probability intervals).

1.5 ∞-monotone capacities

∞-monotone capacities also play an important role as special cases of coherent lower probabilities. Indeed, given the relation between n-monotonicity and the Möbius inverse, any ∞-monotone lower probability $\underline{P}$ on a finite space $\mathcal{X}$ has a non-negative Möbius inverse $m_{\underline{P}} : \wp(\mathcal{X}) \to [0,1]$. This is a characteristic property, as any mapping $m : \wp(\mathcal{X})\setminus\emptyset \to [0,1]$ such that $\sum_{A\subseteq\mathcal{X}} m(A) = 1$ will induce a ∞-monotone measure by using Eq. (1.3).

1.5.1 Constructing ∞-monotone capacities

The non-negativeness of m means that it can be seen as a probability mass function defined over $\wp(\mathcal{X})$. In infinite spaces $\mathcal{X}$, random sets [14, 38] also induce ∞-monotone lower probabilities. Indeed, functions m often result from a situation where the available (statistical) pieces of information only partially determine the quantity of interest. This is typically the case when only a compatibility relation (instead of a mapping) between a probability space and the possibility space $\mathcal{X}$ of interest is available. Suppose there is a multimapping $\Gamma : \mathcal{Y} \to \wp(\mathcal{X})$ that defines for each value $y \in \mathcal{Y}$ of the quantity Y the set $\Gamma(y)$ of possible values of the ill-known quantity x in $\mathcal{X}$. If the subject knows $Y = y$, she only knows that $x \in \Gamma(y)$ and nothing else. From the knowledge of a probability function on $\mathcal{Y}$, only a mass assignment on $\mathcal{X}$ is derived, namely: $\forall E \subseteq \mathcal{X}, m(E) = P(\{y \in \mathcal{Y} \colon \Gamma(y) = E\})$ if $\exists y \in \mathcal{Y}, E = \Gamma(y)$, and 0 otherwise. The probability space $\mathcal{Y}$ can be considered as a sample space like in the framework of frequentist probabilities, but it is then assumed that observations are imprecise.[4]

Example 1.4. : Consider an opinion poll pertaining to a French presidential election. The set of candidates is $\mathcal{X} = \{a, b, c, d, e\}$, going from left-wing ($\{a, b\}$) to right-wing ($\{d, e\}$). There is a population $\mathcal{Y} = \{y_1, \ldots, y_n\}$ of n individuals that supply their preferences. But since the opinion poll takes place well before the election, individuals may not have made a final choice, even if they do have an opinion. The opinion of individual y_i is modeled by the subset $\Gamma(y_i) \subseteq \mathcal{X}$. For instance, a left-wing vote is modeled by $\Gamma(y_i) = \{a, b\}$; for an individual having no opinion, $\Gamma(y_i) = \mathcal{X}$, etc. In this framework, if individual responses of this form are collected, $m(E)$ is the proportion of opinions of the form $\Gamma(y_i) = E$. ♦

Another method for constructing Γ can be devised when the frame $\mathcal{X}$ is multidimensional, i.e., $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k$, and a probability distribution P is available on part of the frame, like $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_i$, $i < k$, and there is a set of constraints relating the various variables $X_1, X_2, \ldots, X_k$, thus forming a relation R on $\mathcal{X}$. R represents all admissible tuples in $\mathcal{X}$. Let $\mathcal{Y} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_i$. Then if $y = (x_1, x_2, \ldots, x_i)$ and if $[y]$ denotes its cylindrical extension to $\mathcal{X}$, $\Gamma(y) = R \cap [y]$. The watch Example 1.6 in the next section is of this kind.

1.5.2 Simple support functions

A particular instance of ∞-monotone capacities that plays an important role in other interpretations (see Chapter **??**) are simple support functions.

Definition 1.5. A simple support function is a mass function where a mass $\alpha = m(A)$ is given to a set A and $1 - \alpha = m(\mathcal{X})$.

Such functions are often used to model the reliability of some source of information, as in the next example.

Example 1.6. Consider an unreliable watch. The failure probability ϵ is known. The set $\mathcal{Y}$ describes the possible states of the watch $\mathcal{Y} = \{KO, OK\}$. The subject cares for the time

[4] See also Section **??**.

it is. So, $\mathcal{X}$ is the set of possible time-points (discretised according to the watch precision). Suppose the watch indicates time x. Then the multimapping Γ is such that $\Gamma(OK) = \{x\}$ (if the watch is in order, it provides the right time), and $\Gamma(KO) = \mathcal{X}$ (if the watch does not work properly, the time it is is unknown). The induced mass assignment on $\mathcal{X}$ is thus $m(\{x\}) = 1 - \epsilon$ and $m(\mathcal{X}) = \epsilon$, which is the probability of not knowing the time it is. ♦

1.5.3 Further elements

Apart from offering practical ways to build coherent lower probabilities from imprecise observations, the fact that m (or Γ) can be interpreted as a probability mass function (or a random variable) over $\wp(\mathcal{X})$ means that usual sampling methods such as Monte-Carlo methods can be used to simulate them. A good review of such methods is given by Wilson[5] [58]. In finite spaces $\mathcal{X}$, ∞-monotone lower probabilities are mathematically equivalent to so-called belief functions [47, 14] (see Chapter **??** for further discussion on belief function interpretations). This means that practical results concerning such models can be transposed to ∞-monotone capacities. For example, the fast Möbius transform [31] can be used to quickly compute $m_{\underline{P}}$, or approximation methods [15] can be used to reduce the representation complexity. Indeed, such models still generally require $2^{|\mathcal{X}|}$ values to be entirely specified, however in practice many sets $A \subseteq \mathcal{X}$ will be such that $m(A) = 0$ (e.g., k-additive belief functions [28], that will not be discussed further in this chapter, give positive masses only to sets whose cardinality is at most k).

In infinite settings, the definition of ∞-monotone lower probabilities poses tricky mathematical problems (initially discussed by Shafer [48] in the setting of belief functions and Matheron [34] in the setting of random sets). Nevertheless, it is possible to define a belief function on the reals, based on a continuous mass density bearing on closed intervals [50] (see Smets [49] for more details and Alvarez [1] for simulation techniques).

Except in specific [55, p.40] [11, Sec. 6], there are no specific relations between ∞-monotone capacities and probability intervals studied in Section 1.4. However, several authors have proposed mappings from probability intervals to ∞-monotone capacities (see Denoeux [16], Hall and Lawry [29] and Quaeghebeur [44])

1.6 Possibility distributions, p-boxes, clouds and related models

In this section, we study practical representations linked by the fact that they correspond to assessments over some collections of nested sets. Like previous models, they have a limited expressiveness but possess interesting properties making them handy.

[5] The review concentrates on a different interpretation of m, but many tools can be used within a lower prevision approach.

1.6.1 Possibility distributions

A possibility distribution [23, 24, 22] is a function $\pi : \mathcal{X} \to [0,1]$ with $\pi(x) = 1$ for at least one $x \in \mathcal{X}$. From this distribution can be defined a possibility measure $\overline{P}$ such that

$$\overline{P}(A) = \sup_{x \in A} \pi(x). \tag{1.7}$$

This measure is supremum-preserving [12], in the sense that for any $\mathcal{A} \subseteq \wp(\mathcal{X})$, we have

$$\overline{P}\left(\bigcup_{A \in \mathcal{A}} A\right) = \sup_{A \in \mathcal{A}} \overline{P}(A).$$

$\overline{P}$ is a coherent upper probability, and the dual coherent lower probability (called necessity measure) is defined as $\underline{P}(A) = 1 - \overline{P}(A^c) = 1 - \sup_{x \in A^c} \pi(x)$. When $\mathcal{X}$ is finite, the supremum in Eq. (1.7) can be replaced by a maximum.

This supremum-preserving property makes possibility measures very easy to use (i.e., evaluations of $\overline{P}(A)$ or $\underline{P}(A)$ are straightforward). On the other hand, this property induces that the inequality $\underline{P}(A) \leq \overline{P}(A)$ holds in a strong form, as $\underline{P}(A) > 0$ implies $\overline{P}(A) = 1$. In particular, we cannot have $\underline{P}(A) = \overline{P}(A)$ for values other than zero and one, hence possibility measures are unable to model linear previsions except in the most trivial cases, this in contrast with previous simple representations. However, possibility distributions and measures have strong connections with the probabilistic setting, as we will see next.

From a possibility distribution π and for any value $\alpha \in [0,1]$, the strong and regular α-cuts are subsets respectively defined as $A_{\overline{\alpha}} = \{x \in \mathcal{X} \colon \pi(x) > \alpha\}$ and $A_\alpha = \{x \in \mathcal{X} \colon \pi(x) \geq \alpha\}$. These α-cuts are nested, since if $\alpha > \beta$, then $A_\alpha \subseteq A_\beta$. The linear previsions included in the corresponding credal sets $\mathcal{M}(\underline{P})$ can be characterised in a particularly interesting way [21], that links them to the concept of strong α-cut (see [10] for extensions on general spaces):

Proposition 1.7. *Given a possibility distribution π and the corresponding credal set $\mathcal{M}(\underline{P})$, we have for all α in $(0,1]$, $P \in \mathcal{M}(\underline{P})$ if and only if*

$$1 - \alpha \leq P(A_{\overline{\alpha}})$$

Conversely, given any indexed family of nested intervals $A_\alpha, \alpha \in [0,1]$, such that $A_\alpha \subseteq A_\beta$ whenever $\alpha \geq \beta$, the credal set $\mathcal{M} = \{P : P(A_\alpha) \geq 1 - \alpha\}$ defines a possibility measure, with distribution $\pi(x) = \inf_{x \notin A_\alpha} \alpha$ [21].

One useful consequence of this result is that any lower prevision $\underline{P}$ can be outer approximated by a possibility distribution, simply specifying a set of nested sets $A_1 \subset \ldots \subset A_n$ and considering the natural extension $\underline{E}(A_i)$ for each A_i. This also means that any probabilistic inequality of the form $\{P([x^* - \alpha a, x^* + \alpha a]) \geq f(\alpha) : \alpha \geq 0\}$, where x^* is some landmark value, is captured by a possibility distribution. For instance, in the Chebychev inequality, x^* is the mean, a is the variance, $f(\alpha) = \frac{a^2}{\alpha^2}$ and $\pi(x) = \min(1, 1 - \frac{a^2}{(x-x^*)^2})$. On finite spaces, the set $\{\pi(x) \colon x \in \mathcal{X}\}$ is of the form

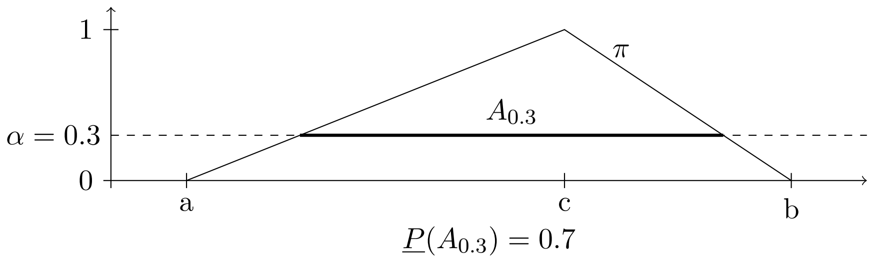

Figure 1.1: Illustration of a triangular fuzzy interval with α-cut of level 0.3

$\alpha_0 = 0 < \alpha_1 < \ldots < \alpha_M = 1$, and there are M distinct α-cuts in this case[6]. In practice, this means that the coherent lower probability induced by a possibility distribution can be expressed (in the finite case) by M lower bounds on the probability of nested events $A_{\overline{\alpha_i}}$, $i = 0, \ldots, M-1$.

Finally, we note that lower probabilities induced by possibility distributions are specific cases of ∞-monotone lower probabilities. The Möbius inverse m of the lower probability $\underline{P}$ induced by π can be easily computed: assuming that elements of $\mathcal{X}$ are ranked such that $\pi(x_1) \geq \ldots \geq \pi(x_{|\mathcal{X}|})$, m is such that

$$m(E_i) = \pi(x_i) - \pi(x_{i+1}), i = 1, \ldots, |\mathcal{X}|, \tag{1.8}$$

where $E_i = \{x_1, \ldots, x_i\}$ and letting $\pi(x_{|\mathcal{X}|+1}) = 0$. Note that events E_i are nested. Shafer [47] showed that there is a one-to-one correspondence between possibility measures and upper probabilities $\overline{P}$ whose Möbius inverse is positive on a collection of nested sets only.

1.6.2 Fuzzy intervals

When working with possibility distributions on the real line, fuzzy intervals constitute by far the most usual representation.

Definition 1.8. A *fuzzy interval* $\pi : \mathbb{R} \to [0, 1]$ is a possibility distribution such that for all x and $y \in \mathbb{R}$,

$$\pi(z) \geq \min\{\pi(x), \pi(y)\} \text{ for all } z \in [x, y].$$

A fuzzy interval π is said to be *normalised* when there is an $x \in \mathbb{R}$ such that $\pi(x) = 1$.

The α-cut A_α of a fuzzy interval π is an interval of the real-line. The piece of information A_α can then be processed by using classical interval analysis or optimisation. Figure 1.1 pictures a triangular fuzzy interval together with an α-cut A_α. Fuzzy intervals have been proposed as a natural representation in practical situations where only limited (probabilistic) information is available:

[6] Note that this is true as long as π only assumes a finite number of values, even on non-finite spaces.

- when the expert gives a set of confidence intervals around a "best guess" value together with confidence levels (the result is then a possibility distribution with a finite number of values) [46];
- when only the support $[a, b]$ and modal value c of a distribution is known, it can be shown that the credal set induced by the triangular fuzzy set of Figure 1.1 includes all distributions having this support and this mode [2][7];
- when only a handful of (not necessarily nested) sensor measurements are available [35];
- when considering probabilistic inequalities providing sets of confidence intervals around a central value (for instance the Chebychev inequality).

The features of fuzzy intervals make them handy tool for various applications, including uncertainty propagation for risk analysis [41], uncertainty analysis in scheduling[20], signal filtering [33], ...

1.6.3 Clouds

As mentioned earlier, possibility distributions are useful but poorly expressive representations, as they cannot capture linear previsions. Clouds [40] are representations that extend possibility distributions while still remaining simple.

Definition 1.9. A cloud $[\delta, \pi]$ on $\mathcal{X}$ is a pair of mappings $\delta : \mathcal{X} \to [0, 1]$, $\pi : \mathcal{X} \to [0, 1]$ such that δ is point-wise less than π (i.e., $\delta \leq \pi$). Moreover, $\pi(x) = 1$ for at least one element x in $\mathcal{X}$, and $\delta(y) = 0$ for at least one element y in $\mathcal{X}$. δ and π are called the lower and upper distributions of the cloud $[\delta, \pi]$, respectively.

From a cloud $[\delta, \pi]$, Neumaier [40] defines the following probabilistic constraints for every $\alpha \in [0, 1]$:

$$P(B_\alpha \coloneqq \{x \in \mathcal{X} \colon \delta(x) \geq \alpha\}) \leq 1 - \alpha \leq P(A_{\overline{\alpha}} = \{x \in \mathcal{X} \colon \pi(x) > \alpha\}). \tag{1.9}$$

These constraints are equivalent to the lower prevision such that $\underline{P}(A_{\overline{\alpha}}) = 1 - \alpha$ and $\underline{P}(B^c_\alpha) = \alpha$ for $\alpha \in [0, 1]$. If $\mathcal{X}$ is continuous, then $\mathcal{M}(\underline{P}) \neq \emptyset$ as soon as $[\delta, \pi]$ satisfies Definition 1.9. On finite spaces, $[\delta, \pi]$ must also satisfy the following condition [18]: $\forall A \subseteq \mathcal{X}$, $\max_{x \in A} \pi(x) \geq \min_{y \notin A} \delta(y)$. Clouds generalise possibility distributions in the sense that if $\delta = 0$, then $\underline{P}$ is the lower probability induced by the possibility distribution π. We also have that $\mathcal{M}(\underline{P}) = \mathcal{M}(\underline{P}_\pi) \cap \mathcal{M}(\underline{P}_{1-\delta})$, with $\underline{P}_\pi$ and $\underline{P}_{1-\delta}$ the lower probabilities induced by the possibility distributions π and $1 - \delta$, respectively.

In general, $\underline{P}$ is not 2-monotone [18], hence harming the practical interest of clouds. However, comonotonic clouds are specific clouds for which $\underline{P}$ is ∞-monotone.

[7] Note that it also includes multi-modal distributions, and that the full language of lower previsions is needed to exactly represent this information.

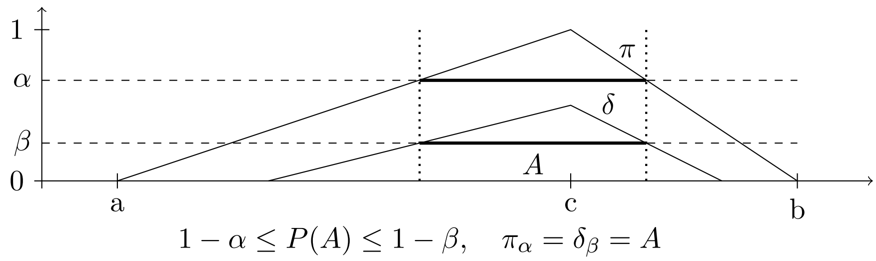

Figure 1.2: Illustration of a comonotonic cloud and of the bounds over event A.

Definition 1.10. A comonotonic cloud $[\delta, \pi]$ is defined as a cloud such that δ and π are comonotone, i.e., for any $x, y \in \mathcal{X}$, $\pi(x) > \pi(y) \implies \delta(x) \geq \delta(y)$

In the finite case, comonotonicity means that there exists a common permutation σ of $\mathcal{X} = \{x_1, \ldots, x_{|\mathcal{X}|}\}$ such that

$$\pi(x_{\sigma(1)}) \geq \pi(x_{\sigma(2)}) \geq \cdots \geq \pi(x_{\sigma(|\mathcal{X}|)})$$

and

$$\delta(x_{\sigma(1)}) \geq \delta(x_{\sigma(2)}) \geq \cdots \geq \delta(x_{\sigma(|\mathcal{X}|)})$$

The simplest comonotonic cloud, that is the cloud summarised by a constraint $\alpha \leq \underline{P}(A) \leq 1 - \beta$ with $\alpha \leq 1 - \beta$, provides an interesting example as it directly extends simple support functions of Section 1.5.2. Indeed, its Möbius inverse m is such that $m(A) = \alpha, m(A^c) = \beta, m(\mathcal{X}) = 1 - \alpha - \beta$, and simple support functions are retrieved when $\alpha = 0$ or $\beta = 0$.

A cloud is comonotonic if and only if the sets $\{A_\alpha, B_\alpha \colon \alpha \in [0, 1]\}$ form a nested sequence (i.e., they are completely ordered w.r.t. inclusion). In the light of constraints in Eq. (1.9), this means that comonotonic clouds correspond to sets of nested intervals to which are associated upper and lower probabilistic bounds (thus adding an upper bound to the lower bound of possibility distributions). Figure 1.2 illustrates the notion of comonotonic cloud on the real line.

As for possibility distributions, a lower probability induced by a cloud can be described (in the finite case) by M constraints corresponding to the M distinct values assumed by the cloud distributions. Conversely, any lower prevision $\underline{P}$ can be outer-approximated by a comonotonic cloud by simply specifying a set of nested sets $A_1 \subset \ldots \subset A_n$ and considering the lower and upper natural extensions $\underline{E}(A_i)$ and $\overline{E}(A_i)$ for each A_i. Recently, clouds have been proposed as practical representations to deal with robust design problems [27] and signal filtering [19].

1.6.4 P-boxes

A p-box [26] (short for probability box) was originally defined as a pair $[\underline{F}, \overline{F}]$ of cumulative distributions, an upper one $\overline{F} : \mathbb{R} \to [0, 1]$ and a lower one $\underline{F} : \mathbb{R} \to [0, 1]$ such that both

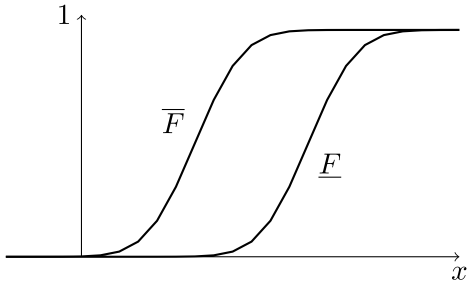

Figure 1.3: Illustration of p-box.

$\overline{F}, \underline{F}$ are non-decreasing and $\exists r$ such that $\overline{F}(r) = \underline{F}(r) = 1$. This model corresponds to specifying lower and upper probabilities on events $(-\infty, x]$ such that $\underline{P}([-\infty, x]) = \underline{F}(x)$ and $\overline{P}([-\infty, x]) = 1 - \underline{P}((x, \infty]) = \overline{F}(x)$. The resulting lower probability $\underline{P}$ is coherent and ∞-monotone.

The concept of p-box has been extended to preordered spaces [17, 51]. This allows to define them on nested confidence regions, or to consider orders on $\mathbb{R}$ that differ from the natural order of numbers. In finite spaces $\mathcal{X}$, it has been shown that such p-boxes generalise possibility distributions and are actually equivalent to comonotonic clouds [18].

P-boxes have been proposed as a practical representation in situations where only limited (probabilistic) information is available [26, 2]:

- when the mean value μ and the support I of a random variable X are known. However, as underlined in [2], it results in a fairly imprecise p-box. Actually, the language of lower previsions is needed to represent exactly this kind of information;

- when only a small number of samples is available. In this case, the use of Kolmogorov-Smirnov confidence limits [25] allow to build bounds over events of the type $[-\infty, x]$;

- when experts provide imprecise assessments about percentiles (already in Walley [53, Sec. 4.6]).

To outer approximate a given lower prevision $\underline{P}$ by a p-box, it is sufficient to define a pre-order $\preceq$ on the possibility space $\mathcal{X}$ and to consider its lower and upper natural extension over events $[0_{\mathcal{X}}, x] := \{y \in \mathcal{X} \colon 0_{\mathcal{X}} \preceq y \preceq x\}$, where $0_{\mathcal{X}}$ denotes a smallest element of $\mathcal{X}$ with respect to $\preceq$.

P-boxes are most often used as initial or final representations in reliability and safety studies, as well as in (industrial) risk analysis (see, for example, Chapter **??**). The reason for that is that cumulative distributions answer the question: “Does quantity X exceeds a given (safety) threshold x or not?”

1.7 Neighbourhood models

Starting from an initial probability measure P_0, neighbourhood models consist in weakening this initial information by means of a parameter $\epsilon \in [0,1]$ that we will call here *discounting parameter*. The three main neighbourhood models are the so-called *pari-mutuel model*(named after horse races, for it was used to determine betting rates), the *linear-vacuous* model and the *odds-ratio* model.

1.7.1 Pari-mutuel

The pari-mutuel model is described by the following constraint on the upper probability of every event A: $\overline{P}_{pm}(A) = \min\{(1+\epsilon)P_0(A), 1\}$ and, by duality (conjugacy), $\underline{P}_{pm}(A) = \max\{(1+\epsilon)P_0(A) - \epsilon, 0\}$. It can be shown (see [53, Sec. 3.2.5]) that this lower probability is 2-monotone, hence previous results concerning this particular representation can be applied to it.

The *pari-mutuel model* inherits its name from the betting world. Consider a gambler betting on a event A and considering $P_0(A)$ as a fair price for a bet that returns 1 if A happens. In this case, the gamblers gain is $I_A - P_0(A)$, while the house (a bookmaker, an insurance, ...) gain is $P_0(A) - I_A$. In real-world situations, the house asks for a positive gain expectation. A way to insure such a positive expectation is to increase the price of the bet by $1+\epsilon$, transforming house gain into $(1+\epsilon)P_0(A) - I_A$. The coefficient ϵ is then interpreted as some kind of taxation or commission coming from the house. The use of this pari-mutuel model of the associated upper probability in insurance risk measurements is discussed by Pelessoni *et al.* [43].

1.7.2 Odds-ratio

The odds-ratio model is described by the following inequalities for each pair of events A, B:

$$\frac{P(A)}{P(B)} \leq (1-\epsilon)\frac{P_0(A)}{P_0(B)}.$$

In the finite case, this model corresponds to a finite set of quantitative comparisons between probabilities of events. Except for degenerate cases, it cannot be represented by means of lower probabilities alone (lower previsions are needed).

We will denote by $\underline{P}_{or}$ the lower prevision generated by this neighbourhood model. The odds-ratio neighbourhood model and the more general family of models called density ratio class in the robust Bayesian literature [4, Sec. 4] are particularly interesting for statistical inference purposes (see Chapter **??**). Indeed, the posterior generated from the combination of an odds-ratio prior with a likelihood function is also an odds-ratio model, which is not the case for other neighbourhood models.

1.7.3 Linear-vacuous

The linear-vacuous model, also called ϵ-contamination model in the robust Bayesian literature, corresponds to the convex mixture of the linear prevision P_0 (weighted by

$(1-\epsilon)$) with the vacuous lower prevision $\inf_{\mathcal{X}}(f) := \inf_{x \in \mathcal{X}} f(x)$ (which is equivalent to the vacuous necessity measure, $N(A) = 0$ for all $A \subset \mathcal{X}$, or to the vacuous mass assignment $m(\mathcal{X}) = 1$). It can be described by the following constraint on the lower probability of every event A: $\underline{P}_{lv}(A) = (1-\epsilon)P_0(A)$. The natural extension $\underline{E}(f)$ of a linear-vacuous model to a gamble $f \in \mathcal{L}(\mathcal{X})$ is given by

$$\underline{E}(f) = (1-\epsilon)E_{P_0}(f) + \epsilon \inf_{\mathcal{X}}(f), \tag{1.10}$$

with E_{P_0} the expectation of f given P_0. Since both linear and vacuous lower previsions are ∞-monotone measures, so is the linear-vacuous model. In this model, $1-\epsilon$ can be seen as the probability that the information P_0 is reliable, hence ϵ is the probability that the source is unreliable, i.e., the probability that we know nothing about a particular variable value (explaining why we combine the initial assessment with the vacuous lower prevision). A generalized version of this scheme is Shafer's discounting technique for belief functions [47]. The watch Example 1.6 is of this kind. Thanks to its simplicity, the linear-vacuous model has been used in many applications [59, 3].

This linear-vacuous model can be extended to the more general case where an information source provides an initial assessment in the form of a lower prevision $\underline{P}$ (or any other model interpretable as such). A lower-vacuous mixture $\underline{P}_\epsilon$ can then be defined, for any gamble f, as $\underline{P}_\epsilon(f) = (1-\epsilon)\underline{E}(f) + \epsilon \inf_{\mathcal{X}}(f)$, where $\underline{E}$ is the natural extension of $\underline{P}$. For the particular case of events, this gives $\underline{P}_\epsilon(A) = (1-\epsilon)\underline{P}(A)$. When $\underline{P}$ is ∞-monotone or when $\overline{P}$ is a possibility measure, this lower-vacuous mixture then coincides with the classical discounting operation of the corresponding theory. Also, simple support functions correspond to the case where $\underline{P}$ is itself vacuous w.r.t. some events A.

1.7.4 Relations between neighbourhood models

For given ϵ and P_0, the credal set induced by the odds-ratio model is included in the credal sets induced by the linear-vacuous and the pari-mutuel model, while the two latter generally just overlap, that is, $\max\{\underline{P}_{lv}, \underline{P}_{pm}\} \leq \underline{P}_{or}$

Example 1.11 (From [53], Sec. 4.6.). Let us consider a probability P_0 on a 3 element space $\mathcal{X} = \{x_1, x_2, x_3\}$ with $P_0(\{x_1\}) = 0.5$, $P_0(\{x_2\}) = 0.3$, $P_0(\{x_3\}) = 0.2$, with a reliability $1-\epsilon = 0.8$. The different credal sets induced by each neighbourhood model are illustrated in the simplex of Figure 1.4 (each point of the simplex is a probability distribution). Note that the odds-ratio neighbourhood is the only model whose edges are not parallel to one of the simplex edges, showing that it cannot be exactly represented by lower probabilities alone. ⧫

1.8 Summary

In this section, we have reviewed the main practical representations that have emerged as instrumental tools for representing imprecise probabilities and computing with them.

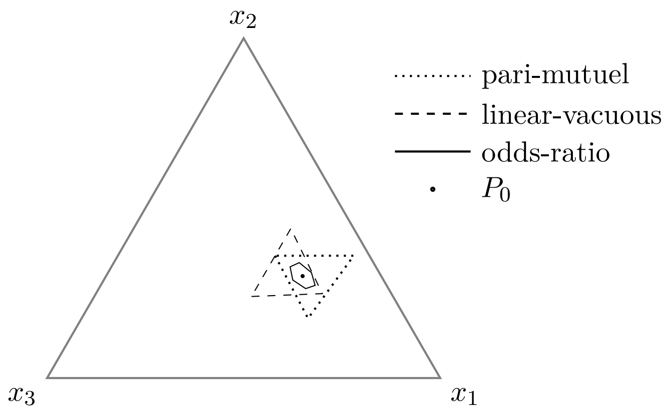

Figure 1.4: Neighbourhood models of Example 1.11

However, while they may be helpful in many situations, such as uncertainty propagation in risk analysis or uncertainty assessment procedures, there are still a number of situations where such simplified representations will not be sufficient. For instance, practical representations presented here cannot handle the case where the information consists of quantitative comparative assessments of probabilities [45], nor can lower and upper probabilities. Other typical situations where the use of generic lower previsions may be needed is in the extension of classical statistical notions such as:

- statistical inference (see Chapter **??**);
- combination of marginal assessments through independence concepts (see Chapter **??**);
- combination of conditional assessments (see Chapter **??**);
- extension of notions expressed by the means of expectation operators, such as characteristic or generating functions.

Indeed, for the above problems, even if one starts from simple representations, the representation resulting from information processing will usually not have the same properties.

Figure 1.5 summarises the relations between the different representations detailed in this chapter when $\mathcal{X}$ is finite. Note that for clarity purposes, some relations are not present, for instance probability intervals on singletons include linear vacuous mixtures, which themselves include linear and vacuous previsions as special cases. Similarly simple support functions, that are special cases of possibility distributions, are not in the picture. Note that such relations may not hold in infinite spaces, as in this case the different models may not coincide in some specific situations.

For example, when $\mathcal{X}$ is infinite, finitely maxitive upper probabilities may fail to possess an underlying possibility distribution, even in the Boolean case. For instance

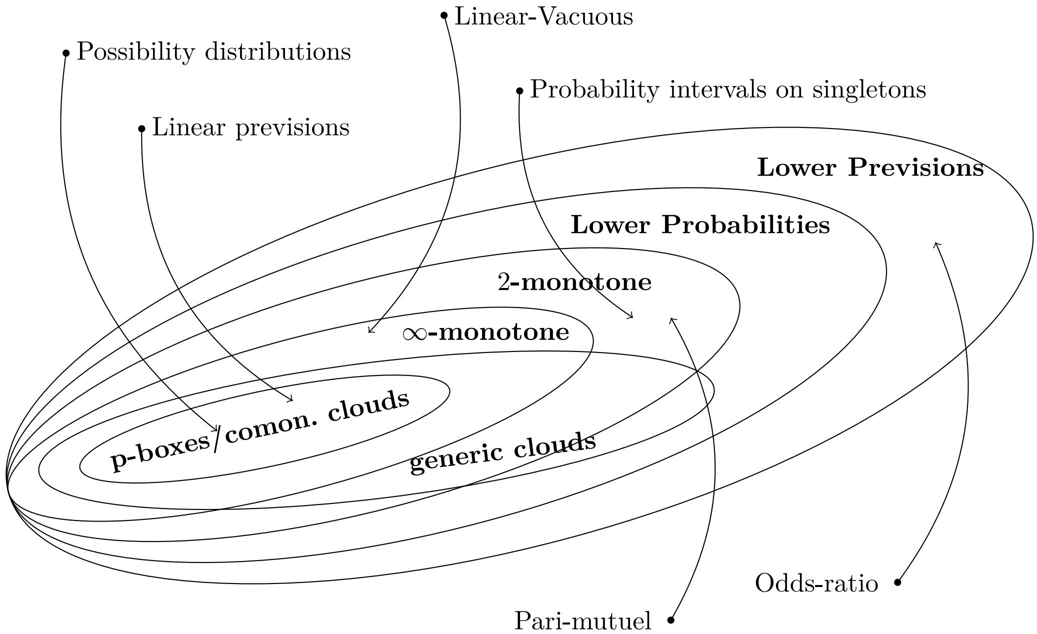

Figure 1.5: Summary of relations between practical representations

suppose $\mathcal{X}$ is the set of natural integers, and $\overline{P}(A) = 1$ if A is infinite, 0 if finite, then $\overline{P}(\{x\}) = 0, \forall x \in \mathcal{X}$. Similarly, extending Proposition 1.7 is not trivial [10]. Relations between models of Figure 1.5 for infinite spaces largely remain to be explored, but we can mention some results: Miranda *et al.* [37] explore the relation between consonant random sets and possibility measures in the general case, while other works [52] explore the links between p-boxes defined on totally pre-ordered spaces and possibility distributions.

Bibliography

[1] D. Alvarez. A Monte Carlo-based method for the estimation of lower and upper probabilities of events using infinite random sets of indexable type. *Fuzzy Sets and Systems*, 160:384–401, 2009.

[2] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51:86–108, 2006.

[3] A. Benavoli, M. Zaffalon, and E. Miranda. Reliable hidden Markov model filtering through coherent lower previsions. In *FUSION '09: Proceedings of the 12th International Conference on Information Fusion*, pages 1743–1750, Seattle, USA, 2009. IEEE.

[4] J. Berger, E. Moreno, L. Pericchi, M. Bayarri, J. Bernardo, J. Cano, J. De la Horra, J. Martín, D. Ríos-Insúa, B. Betrò, A. Dasgupta, P. Gustafson, L. Wasserman, J. Kadane, C. Srinivasan, M. Lavine, A. O'Hagan, W. Polasek, C. Robert, C. Goutis, F. Ruggeri, G. Salinetti, and S. Sivaganesan. An overview of robust Bayesian analysis. *TEST*, 3:5–124, 1994.

[5] A. Bronevich and T. Augustin. Approximation of coherent lower probabilities by 2-monotone measures. In T. Augustin, F. Coolen, M. Troffaes, and S. Moral, editors, *ISIPTA'09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 61–70, Durham, UK, 2009. SIPTA.

[6] A. Chateauneuf and J. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[7] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1954.

[8] L. Chrisman. Propagation of 2-monotone lower probabilities on an undirected graph. In E. Horvitz and F. Jensen, editors, *UAI '96: Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence*, pages 178–185, San Francisco, CA, 1996. Morgan Kaufmann.

[9] P. Coolen-Schrijner, F. Coolen, M. Troffaes, T. Augustin, and S. Gupta. *Imprecision in Statistical Theory and Practice.* Grace Scientific Publishing LLC, Greensboro, NC, USA, 2009.

[10] I. Couso, S. Montes, and P. Gil. The necessity of the strong alpha-cuts of a fuzzy set. *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:249–262, 2001.

[11] L. de Campos, J. Huete, and S. Moral. Probability intervals: A tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.

[12] G. de Cooman and D. Aeyels. Supremum preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.

[13] B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

[14] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[15] T. Denoeux. Inner and outer approximation of belief structures using a hierarchical clustering approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:437–460, 2001.

[16] T. Denoeux. Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, 42:228–252, 2006.

[17] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: I. Generalized p-boxes. *International Journal of Approximate Reasoning*, 49:649–663, 2008.

[18] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: II Clouds. *International Journal of Approximate Reasoning*, 49:649–663, 2008.

[19] S. Destercke and O. Strauss. Using cloudy kernels for imprecise linear filtering. In E. Hl̈lermeier, R. Kruse, and F. Hoffmann, editors, *IPMU '10: Proceedings of the 13th International Conference on Information Processing and Management of Uncertainty*, pages 198–207. Springer, 2010.

[20] D. Dubois, J. Fortin, and P. Zielinski. Interval PERT and its fuzzy extension. In C. Kahraman and M. Yavuz, editors, *Production Engineering and Management Under Fuzziness*, volume 252, pages 171–199. Springer, Berlin, 2010.

[21] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10:273–297, 2004.

[22] D. Dubois, H. Nguyen, and H. Prade. Possibility theory, probability and fuzzy sets: Misunderstandings, bridges and gaps. In *Fundamentals of Fuzzy Sets*, pages 343–438. Kluwer Academic Publishers, 2000.

[23] D. Dubois and H. Prade. Properties of measures of information in evidence and possibility theory. *Fuzzy Sets and Systems*, 24:161–182, 1987.

[24] D. Dubois and H. Prade. Possibility theory: Qualitative and quantitative aspects. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 169–226. Kluwer Academic Publishers, Dordrecht, 1998.

[25] W. Feller. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Annals of Mathematical Statistics*, 19:177–189, 1948.

[26] S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002-4015 (available from http://ebookbrowse.com/sand2002-4015-pdf-d22644198), Sandia National Laboratories, 2003.

[27] M. Fuchs and A. Neumaier. Potential based clouds in robust design optimization. *Journal of Statistical Theory and Practice*, 3:225–238, 2008. Reprinted in [9], pages 285-298.

[28] M. Grabisch. K-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.

[29] J. Hall and J. Lawry. Generation, combination and extension of random set approximations to coherent lower and upper probabilities. *Reliability Engineering & System Safety*, 85:89–101, 2004.

[30] J. Kadane, M. Schervish, and T. Seidenfeld. *Rethinking the Foundations of Statistics*. Cambridge University Press, 1999.

[31] R. Kennes. Computational aspects of the Möbius transformation of graphs. *IEEE Transactions on Systems, Man, and Cybernetics*, 22:201–223, 1992.

[32] H. Kyburg Jr. and H. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. Second edition (with new material) 1980.

[33] K. Loquin and O. Strauss. Noise quantization via possibilistic filtering. In T. Augustin, F. Coolen, S. Moral, and M. Troffaes, editors, *ISIPTA 09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 297–306, Durham, UK, 2009. SIPTA.

[34] G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, 1975.

[35] G. Mauris. Inferring a possibility distribution from very few measurements. In D. Dubois, M. Lubiano, H. Prade, M. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods for Handling Variability and Imprecision*, volume 48, pages 92–99, Berlin, 2008. Springer.

[36] E. Miranda, I. Couso, and P. Gil. Extreme points of credal sets generated by 2-alternating capacities. *International Journal of Approximate Reasoning*, 33:95–115, 2003.

[37] E. Miranda, I. Couso, and P. Gil. A random set characterization of possibility measures. *Information Science*, 168:51–75, 2004.

[38] E. Miranda, I. Couso, and P. Gil. Random sets as imprecise random variables. *Journal of Mathematical Analysis and Applications*, 307:32–47, 2005.

[39] T. Murofushi and M. Sugeno. An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure. *Fuzzy sets and Systems*, 29:201–227, 1989.

[40] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.

[41] M. Oberguggenberger, J. King, and B. Schmelzer. Classical and imprecise probability methods for sensitivity analysis in engineering: A case study. *International Journal of Approximate Reasoning*, 50:680–693, 2009.

[42] A. Papamarcou and T. Fine. A note on undominated lower probabilities. *The Annals of Probability*, 14:710–723, 1986.

[43] R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the Pari-Mutuel model. *International Journal of Approximate Reasoning*, 51:1145–1158, 2010.

[44] E. Quaeghebeur. Completely monotone outer approximations of lower probabilities on finite possibility spaces. In S. Li, X. Wang, Y. Okazaki, J. Kawabe, T. Murofushi, and L. Guan, editors, *Nonlinear Mathematics for Uncertainty and its Applications*, volume 100, pages 169–178. Springer, Berlin/Heidelberg, 2011.

[45] G. Regoli. Comparative probability orderings, 1999. http://www.sipta.org/documentation/comparative_prob/regoli.pdf.

[46] S. Sandri, D. Dubois, and H. Kalfsbeek. Elicitation, assessment and pooling of expert judgments using possibility theory. *IEEE Transactions on Fuzzy Systems*, 3:313–335, 1995.

[47] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[48] G. Shafer. Allocations of probability. *Annals of Probability*, 7:827–839, 1979.

[49] P. Smets. Belief functions on real numbers. *International Journal of Approximate Reasoning*, 40:181–223, 2005.

[50] T. Strat. Continuous belief functions for evidential reasoning. In R. Brachman, editor, *Proceedings of the Fourth National Conference on Artificial Intelligence (AAAI)*, pages 308–313, 1984.

[51] M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approxate Reasoning*, 52:767–791, 2011.

[52] M. Troffaes, E. Miranda, and S. Destercke. On the connection between probability boxes and possibility measures. *Information Sciences*, 224:88–108, 2013.

[53] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman & Hall, London, 1991.

[54] P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[55] K. Weichselberger and S. Pöhlmann. *A methodology for uncertainty in knowledge-based systems*, volume 419. Springer Heidelberg, 1990.

[56] P. Williams. Notes on conditional previsions. Technical report, University of Sussex, February 1975. Published as [57] ******************** unify? ************************.

[57] P. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383, 2007. Original: [56] ******************** unify? ************************.

[58] N. Wilson. Algorithms for Dempster-Shafer theory. In D. Gabbay, P. Smets, J. Kohlas, and S. Moral, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 5, pages 421–475. Kluwer, 2000.

[59] M. Zaffalon. The naive credal classifier. *Journal of Statistical Planning and Inference*, 105:5–21, 2002.