Chapter 3
Introduction to the Theory of Imprecise Probability

Erik Quaeghebeur

Abstract The theory of imprecise probability is a generalization of classical 'precise' probability theory that allows modeling imprecision and indecision. This is a practical advantage in situations where a unique precise uncertainty model cannot be justified. This arises, for example, when there is a relatively small amount of data available to learn the uncertainty model or when the model's structure cannot be defined uniquely. The tools the theory provides make it possible to draw conclusions and make decisions that correctly reflect the limited information or knowledge available for the uncertainty modeling task. This extra expressivity however often implies a higher computational burden. The goal of this chapter is to primarily give you the necessary knowledge to be able to read literature that makes use of the theory of imprecise probability. A secondary goal is to provide the insight needed to use imprecise probabilities in your own research. To achieve the goals, we present the essential concepts and techniques from the theory, as well as give a less in-depth overview of the various specific uncertainty models used. Throughout, examples are used to make things concrete. We build on the assumed basic knowledge of classical probability theory.

3.1   Introduction

The theory of imprecise probability is a generalization of classical 'precise' probability theory that allows modeling imprecision and indecision. Why is such a theory necessary? Because in many practical applications a lack of information—e.g., about model parameters—and paucity of data—especially if we also consider conditional models—make it impossible to create a reliable model.

For example, consider a Bayesian context where a so-called prior probability distribution must be chosen as part of the modeling effort. The lack of information

---

(This chapter was written while at Delft University of Technology)

E. Quaeghebeur (✉)
Eindhoven University of Technology, Eindhoven, The Netherlands

may make it difficult to determine the type of the prior distribution, let alone its parameters. Then, even if we assume some prior has been chosen—e.g., a normal one—in a somewhat arbitrary way, a paucity of data will make the parameters of the posterior—updated—distribution depend to a large degree on the prior's somewhat arbitrary parameters. The consequence is that conclusions drawn from the posterior are unreliable and decisions based on it somewhat arbitrary.

The theory of imprecise probability provides us with a set of tools for dealing with the problem described above. For the example above, instead of choosing a single prior distribution, a whole set of priors is used, one that is large enough to sufficiently reduce or even eliminate the arbitrariness of this modeling step. The consequence is that conclusions drawn from an imprecise probabilistic model are more reliable by being less committal—more vague, if you wish; some would say 'more honest'—and that decisions based on it allow for indecision.

In this chapter, we will go over the basic concepts of the theory of imprecise probability theory. Therefore, we will consider 'small' problems, with finite possibility spaces. However, the theory can be applied to infinite—countable and uncountable—possibility spaces as well. Also, only the basics of more advanced topics such as conditioning will be touched upon. But, and this is the chapter's goal, after having understood the material we do treat, the imprecise probability literature should have become substantially more accessible. Good extensive general treatments are available [2, 16, 20] and the proceedings of the ISIPTA conferences provide an extensive selection of papers developing imprecise probability theory or applying it [1, 3, 4, 6–12].

Concretely, we start with a discussion of the fundamental concepts in Sect. 3.2. This is done in terms of the more basic notion of sets of acceptable gambles. Probabilities only appear thereafter, in Sect. 3.3, together with the related notion of prevision (expectation). The connection with sets of probabilities is made next, in Sect. 3.4. Then we touch upon conditioning, in Sect. 3.5, and before closing add some remarks about continuous possibility spaces, in Sect. 3.6. Throughout we will spend ample time on a running example to illustrate the theory that is introduced.

3.2 Fundamental Concepts

In this section, we introduce the fundamental concepts of the theory of imprecise probability [18] [20, §3.7]. First, in Sect. 3.2.1, we get started with some basic concepts. Then, in Sect. 3.2.2, we list and discuss the coherence criteria on which the whole theory is built.

3.2.1 Basic Concepts

Consider an *agent* reasoning about an *experiment* with an uncertain *outcome*. This experiment is modeled using a *possibility space*—a set—X of outcomes x. Now consider the linear space $\mathcal{L} = X \to \mathbb{R}$ of real-valued functions over the outcomes. We view these functions as *gambles* because they give a value, seen as a *payoff*, for each outcome and because the outcome is uncertain and therefore the payoff is as well. A special class of gambles are the outcome *indicators* 1_x or subset *indicators* 1_B, which take the value one on that outcome or subset and zero elsewhere.

The agent can then express her uncertainty by specifying a set of gambles, called an *assessment* $\mathcal{A}$, that she considers *acceptable*. Starting from such an assessment, she can reason about other gambles and decide whether she should also accept them or not. If she were to do this for all gambles, then the *natural extension* $\mathcal{E}$ of her assessment would be the set of all acceptable gambles. To reason in a principled way, she needs some guiding criteria; these are the next section's topic.

Let us now introduce our running example:

Wiske and Yoko Tsuno want to bet on Belgium vs. Japan

Given a sports match between *Belgium* and *Japan*, there is uncertainty about which country's team will win. So we consider the possibility space $\{\text{BE}, \text{JP}\}$. There are to agents—gamblers—: *Wiske* and *Yoko Tsuno*, two comic book heroines. Each has an assessment consisting of a single gamble that they find acceptable:

- Wiske accepts losing 5 coins if Japan wins for the opportunity to win 1 coin if Belgium wins; so $\mathcal{A}_W = \{1_{\text{BE}} - 5 \cdot 1_{\text{JP}}\}$.
- Yoko Tsuno accepts losing 4 coins if belgium wins for the opportunity to win 1 coin if Japan wins; so $\mathcal{A}_Y = \{-4 \cdot 1_{\text{BE}} + 1_{\text{JP}}\}$.

The heroines are also discussing joining forces and forming a betting pool. The pools they consider are

- 'Simple', formed by combining their assessments; so

$$\mathcal{A}_{\text{SP}} = \{1_{\text{BE}} - 5 \cdot 1_{\text{JP}}, -4 \cdot 1_{\text{BE}} + 1_{\text{JP}}\}.$$

- 'Empty' in case of disagreement, without any acceptable gambles; so $\mathcal{A}_{\text{EP}} = \emptyset$.

3.2.2 Coherence

In the theory of imprecise probabilities, the classical rationality criteria used for reasoning about assessments are called *coherence* criteria. These are typically formulated as four rules that should apply to any gambles f and g. (There are different

variants in the literature, but the differences are not relevant in this introductory text.) We divide the criteria into two classes.

Constructive   State how to generate acceptable gambles from the assessment:
  Positive scaling   If f is acceptable and $\lambda > 0$, then $\lambda \cdot f$ is acceptable.
        Addition   If f and g are acceptable, then $f + g$ is acceptable.

Background   State which gambles are always or never acceptable:
      Accepting gain   If f is nonnegative for all outcomes, then f is acceptable.
  Avoiding sure loss   If g is negative for all outcomes, then g is not acceptable.

These criteria are quite broadly seen as reasonable, under the assumption that the payoffs are 'not too large'.

The last criterion, 'Avoiding sure loss', puts a constraint on what is considered coherent; if it is violated, we say that an assessment *incurs sure loss*. The first three rules can be used to create an explicit expression for the natural extension:

$$\mathcal{E} = \left\{ \sum_{f \in \mathcal{K}} \lambda_f \cdot f : \mathcal{K} \Subset \mathcal{A} \cup \{f \in \mathcal{L} : f \geq 0\} \text{ and } (\forall f \in \mathcal{K} : \lambda_f \geq 0) \right\},$$

where $\Subset$ denotes the finite subset relation. Then $\mathcal{E}$ is the smallest convex cone of gambles encompassing the assessment $\mathcal{A}$ and the nonnegative gambles—including the zero gamble.

Let us apply the natural extension to our running example:

The natural extensions of Wiske, Yoko Tsuno, and the betting pools

For our finite possibility space,

$$\{f \in \mathcal{L} : f \geq 0\} = \left\{ \sum_{x \in \mathcal{X}} \mu_x \cdot 1_x : (\forall x \in \mathcal{X} : \mu_x \geq 0) \right\}$$

So, with $\lambda_A, \mu_x \geq 0$ for all outcomes x and agent identifiers A, we get the following expressions that characterize the natural extensions:

$$\text{Wiske} \quad \begin{aligned} \lambda_W \cdot (1_{BE} - 5 \cdot 1_{JP}) &+ \mu_{BE} \cdot 1_{BE} + \mu_{JP} \cdot 1_{JP} \\ &= (\lambda_W + \mu_{BE}) \cdot 1_{BE} + (-5 \cdot \lambda_W + \mu_{JP}) \cdot 1_{JP}, \end{aligned}$$

$$\text{Yoko Tsuno} \quad \begin{aligned} \lambda_Y \cdot (-4 \cdot 1_{BE} + 1_{JP}) &+ \mu_{BE} \cdot 1_{BE} + \mu_{JP} \cdot 1_{JP} \\ &= (-4 \cdot \lambda_Y + \mu_{BE}) \cdot 1_{BE} + (\lambda_Y + \mu_{JP}) \cdot 1_{JP}, \end{aligned}$$

$$\text{Simple pool} \quad (\lambda_W - 4 \cdot \lambda_Y + \mu_{BE}) \cdot 1_{BE} + (-5 \cdot \lambda_W + \lambda_Y + \mu_{JP}) \cdot 1_{JP},$$

$$\text{Empty pool} \quad \mu_{BE} \cdot 1_{BE} + \mu_{JP} \cdot 1_{JP}.$$

To check whether the natural extension incurs sure loss, we must check whether the coefficients of 1_{BE} and 1_{JP} can become negative at the same time. Only the simple pool incurs sure loss; e.g., fill in $\lambda_W = \lambda_Y = 1$ and $\mu_{BE} = \mu_{JP} = 0$ to convince yourself. (Convince yourself as well that the others avoid sure loss indeed.)

3.3 Previsions and Probabilities

In this section, we move from modeling uncertainty using sets of acceptable gambles to the more familiar language of expectation—or, synonymously, *prevision*—and probability [17]. We first transition from acceptable gambles to previsions in Sect. 3.3.1 [18, §1.6.3] [17, §2.2] and in a second step, in Sect. 3.3.2, give the connection to probabilities [20, §2.6]. Next, in Sect. 3.3.3, we consider assessments in terms of previsions and what the other fundamental concepts of Sect. 3.2 then look like [17, §2.2.1, §2.2.4] [20, §2.4–5, §3.1]. Finally, in Sect. 3.3.4, we consider the important special case of assessments in terms of previsions defined on a linear space of gambles [17, §2.2.1] [20, §2.3.2–6].

3.3.1 Previsions as Prices for Gambles

Before we start: 'prevision' is in much of the imprecise probability literature used as a synonym for 'expectation'; we here follow that tradition.

Now, how do we get an agent's *previsions* for a gamble—equivalently: expectation of a random variable—given that we know the agent's assessment as a set of acceptable gambles $\mathcal{A}$? We first define a *price* to be a constant gamble and identify this constant gamble with its constant payoff value. Then we define the agent's previsions as specific types of *acceptable* prices:

- The *lower prevision* $\underline{P}(f)$ is the supremum acceptable buying price of f:

$$\underline{P}(f) = \sup \{\nu \in \mathbb{R} : f - \nu \in \mathcal{E}\}.$$

- The *upper prevision* $\overline{P}(f)$ is the infimum acceptable selling price of f:

$$\overline{P}(f) = \inf \{\kappa \in \mathbb{R} : \kappa - f \in \mathcal{E}\}.$$

If $\mathcal{E}$ is coherent, then $\underline{P}$ and $\overline{P}$ are also called coherent. There is a *conjugacy* relation between coherent lower and upper previsions: $\overline{P}(f) = -\underline{P}(-f)$. It allows us to work in terms of either type of prevision; we will mainly use the lower one.

In case $\underline{P}(f) = \overline{P}(f)$, then $P(f) = \underline{P}(f)$ is the called the (precise) *prevision* of the gamble f.

3.3.2 Probabilities as Previsions of Indicator Gambles

Now that we have definitions for lower and upper previsions, we can derive probabilities from those. For classical probability, we have that the probability of an event—a

subset B of the possibility space $\mathcal{X}$—is the prevision of the indicator for that event. For lower and upper previsions, we get:

- The *lower probability*: $\underline{P}(B) = \underline{P}(1_B)$.
- The *upper probability*: $\overline{P}(B) = \overline{P}(1_B)$.

Notice that we reuse the same symbol for the prevision and probability functions, as is common in the literature. As long as the nature of the argument—gamble or event—is clear, this does not cause ambiguity. If $\underline{P}$ and $\overline{P}$ are coherent as previsions, then so are they as probabilities. Also the conjugacy relationship can be translated to coherent lower and upper probabilities; let $B^c = \mathcal{X} \setminus B$, then

$$\overline{P}(B) = \overline{P}(1_B) = \overline{P}(1 - 1_{B^c}) = -\underline{P}(-1 + 1_{B^c}) = 1 - \underline{P}(1_{B^c}) = 1 - \underline{P}(B^c).$$

In case $\underline{P}(B) = \overline{P}(B)$, then $P(B) = \underline{P}(B)$ is called the (precise) *probability* of B.

To make the definitions for lower and upper previsions and probabilities concrete, let us apply them to our running example:

Lower and upper probabilities for all events and agents

We work out the calculation of Wiske's lower probability that Belgium will win.

$$
\begin{aligned}
\underline{P}_W(\text{BE}) =&\, \underline{P}_W(1_{\text{BE}}) \quad \text{(def. lower probability)}\\
=&\, \sup\left\{v \in \mathbb{R} : 1_{\text{BE}} - v \in \mathcal{E}_W\right\} \quad \text{(def. lower prevision)}\\
=&\, \sup\left\{v \in \mathbb{R} : \begin{bmatrix}1 - v\\0 - v\end{bmatrix} = \begin{bmatrix}\lambda_W + \mu_{\text{BE}}\\-5 \cdot \lambda_W + \mu_{\text{JP}}\end{bmatrix}, \lambda_W \geq 0, \mu_{\text{BE}} \geq 0, \mu_{\text{JP}} \geq 0\right\}\\
&\qquad\qquad\qquad \text{(write out natural extension } \mathcal{E}_W \text{ of } \mathcal{A}_W)\\
=&\, \sup\left\{5 \cdot \lambda_W - \mu_{\text{JP}} : 1 - 5 \cdot \lambda_W + \mu_{\text{JP}} = \lambda_W + \mu_{\text{BE}}, \lambda_W \geq 0, \mu_{\text{BE}} \geq 0, \mu_{\text{JP}} \geq 0\right\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(eliminate } v)\\
=&\, \sup\left\{5 \cdot \lambda_W - \mu_{\text{JP}} : \lambda_W = \tfrac{1}{6}(1 + \mu_{\text{JP}} - \mu_{\text{BE}}), \lambda_W \geq 0, \mu_{\text{BE}} \geq 0, \mu_{\text{JP}} \geq 0\right\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(solve constraint for } \lambda_W)\\
=&\, \sup\left\{\tfrac{5}{6} - \tfrac{1}{6}\mu_{\text{JP}} - \tfrac{5}{6}\mu_{\text{BE}} : 1 + \mu_{\text{JP}} \geq \mu_{\text{BE}}, \mu_{\text{BE}} \geq 0, \mu_{\text{JP}} \geq 0\right\} \quad \text{(eliminate } \lambda_W)\\
=&\, \frac{5}{6} \quad \text{(feasible solution } \mu_{\text{BE}} = 0, \mu_{\text{JP}} = 0 \text{ maximizes expression)}
\end{aligned}
$$

Do the calculations also for the other agents and Japan. Then apply conjugacy to find the following table of lower and upper probabilities:

Agents	$\underline{P}(\text{BE})$, $\overline{P}(\text{BE})$	$\underline{P}(\text{JP})$, $\overline{P}(\text{JP})$	Note
Wiske	5/6 , 1	0 , 1/6	Will not bet against Belgium
Yoko Tsuno	0 , 1/5	4/5 , 1	Will not bet against Japan
Simple pool	$+\infty$, $-\infty$	$+\infty$, $-\infty$	Sure loss, so absurd bounds
Empty pool	0 , 1	0 , 1	So-called *vacuous* model

While in the above example the calculation of the lower prevision can be done by hand, in general it realistically requires a linear program solver.

3.3.3 Assessments of Lower Previsions

Up until now, we assumed a set of acceptable gambles $\mathcal{A}$—an agent's assessment—to be given. But often the agent will directly specify lower and upper probabilities or previsions, e.g., as bounds on precise probabilities and previsions. However, the coherence criteria and expression for the natural extension are based on having a set of acceptable gambles. In this section we will provide expressions based on an assessment specified as lower prevision values for gambles in a given set $\mathcal{K}$.

The approach is to derive an assessment as a set of acceptable gambles $\mathcal{A}$ from these lower prevision. Irrespective of what its natural extension $\mathcal{E}$ actually looks like, it follows from the definition of the lower prevision as a supremum acceptable buying price that

$$0 \leq \sup \left\{ v - \underline{P}(f) : v \in \mathbb{R} \wedge f - v \in \mathcal{E} \supseteq \mathcal{A} \right\}$$
$$= \sup \left\{ \kappa \in \mathbb{R} : f - (\kappa + \underline{P}(f)) \in \mathcal{E} \right\} = \sup \left\{ \kappa \in \mathbb{R} : (f - \underline{P}(f)) - \kappa \in \mathcal{E} \right\}.$$

This implies that $f - \underline{P}(f) + \varepsilon \in \mathcal{E}$ for any $\varepsilon > 0$, because of coherence. We cannot take $\varepsilon = 0$, because the corresponding so-called *marginal gamble* $f - \underline{P}(f)$ is not included in $\mathcal{E}$ in general, as the supremum value $\kappa = 0$ is not necessarily attained inside the set. We therefore take $\mathcal{A} = \bigcup_{f \in \mathcal{K}} \left\{ f - \underline{P}(f) + \varepsilon : \varepsilon > 0 \right\}$.

We can then apply the theory described above to this assessment $\mathcal{A}$. This leads to the following nontrivial results for a lower prevision $\underline{P}$ defined on a set of gambles $\mathcal{K}$:

- It *avoids sure loss* if and only if for all $n \geq 0$ and $f_k \in \mathcal{K}$ it holds that

$$\sup_{x \in \mathcal{X}} \sum_{k=1}^{n} \left(f_k(x) - \underline{P}(f_k) \right) \geq 0.$$

- It is *coherent* if and only if for all $n, m \geq 0$ and $f_k \in \mathcal{K}$ it holds that

$$\sup_{x \in \mathcal{X}} \left(\sum_{k=1}^{n} \left(f_k(x) - \underline{P}(f_k) \right) - m \cdot (f_0 - \underline{P}(f_0)) \right) \geq 0.$$

- Its *natural extension* to any gamble f in $\mathcal{L}$ is

$$\underline{E}(f) = \sup \left\{ \inf_{x \in \mathcal{X}} \left\{ f(x) - \sum_{k=1}^{n} \lambda_k \cdot \left(f_k(x) - \underline{P}(f_k) \right) \right\} : n \geq 0, f_k \in \mathcal{K}, \lambda_k \geq 0 \right\}.$$

3.3.4   Working on Linear Spaces of Gambles

The coherence criterion for lower previsions on an arbitrary space $\mathcal{K}$ of gambles we gave in the preceding section is quite involved. However, in case $\mathcal{K}$ is a *linear space* of gambles, this criterion becomes considerably simpler. Namely, a lower prevision $\underline{P}$ must then satisfy the following criteria for all gambles f and g in $\mathcal{K}$ and $\lambda > 0$:

$$
\begin{aligned}
\textit{Accepting sure gains} \quad & \underline{P}(f) \geq \inf f, \\
\textit{Super-linearity} \quad & \underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g), \\
\textit{Positive homogeneity} \quad & \underline{P}(\lambda f) = \lambda \cdot \underline{P}(f).
\end{aligned}
$$

Expressed for upper previsions $\overline{P}$, these coherence criteria are very similar:

$$
\begin{aligned}
\textit{Accepting sure gains} \quad & \overline{P}(f) \leq \sup f, \\
\textit{Sub-linearity} \quad & \overline{P}(f+g) \leq \overline{P}(f) + \overline{P}(g), \\
\textit{Positive homogeneity} \quad & \overline{P}(\lambda f) = \lambda \cdot \overline{P}(f).
\end{aligned}
$$

From the coherence criteria, many useful properties can be derived for a coherent lower prevision $\underline{P}$ and its conjugate upper prevision $\overline{P}$. We provide a number of key ones, which hold for all gambles f and g in $\mathcal{K}$ and $\mu \in \mathbb{R}$; $\overline{\underline{P}}$ denotes either $\underline{P}$ or $\overline{P}$:

$$
\begin{aligned}
\textit{Upper dominates lower} \quad & \overline{P}(f) \geq \underline{P}(f), \\
\textit{Constants} \quad & \overline{\underline{P}}(\mu) = \mu, \\
\textit{Constant additivity} \quad & \overline{\underline{P}}(f+\mu) = \overline{\underline{P}}(f) + \mu, \\
\textit{Gamble dominance} \quad & \text{if } f \geq g + \mu \text{ then } \overline{\underline{P}}(f) \geq \overline{\underline{P}}(g) + \mu, \\
\textit{Mixed sub/super-additivity} \quad & \underline{P}(f+g) \leq \underline{P}(f) + \overline{P}(g) \leq \overline{P}(f+g).
\end{aligned}
$$

3.4   Sets of Probabilities

In Sect. 3.2 we modeled uncertainty using a set of acceptable gambles. In Sect. 3.3 we showed how this can also be done in terms of lower or upper previsions (or probabilities). In this section, we add a third representation, one using *credal sets*—sets of precise previsions [17, §2.2.2], [18, §1.6.2]. In Sect. 3.4.1 we show how to derive the credal set corresponding to a given lower prevision. In Sect. 3.4.2 we go the other direction and show how to go from a credal set to lower prevision values [20, §3.3].

3.4.1   From Lower Previsions to Credal Sets

A credal set is a subset of the set of all precise previsions $\mathcal{P}$. (For possibility spaces B different from $\mathcal{X}$, we write $\mathcal{P}_B$.) This set is convex, meaning that any convex mixture

of precise previsions is again a precise prevision. Because of this, a gamble's prevision is a linear function over this space. A *lower*—and *upper*—prevision can be seen as providing a bound on the value of the precise prevision for that gamble and thereby represent a *linear constraint* on the precise previsions. So the credal set $\mathcal{M}$ corresponding to a lower prevision $\underline{P}$ defined on a set of gambles $\mathcal{K}$ is the subset of $\mathcal{P}$ satisfying this constraint for all gambles in $\mathcal{K}$:

$$\mathcal{M} = \bigcap_{f \in \mathcal{K}} \left\{ P \in \mathcal{P} : P(f) \geq \underline{P}(f) \right\}.$$

Being defined as such an intersection, *such* credal sets are closed and convex.

   The rationality criteria for a lower prevision $\underline{P}$ we encountered before can also be expressed using its corresponding credal set $\mathcal{M}$:

- $\underline{P}$ incurs sure loss if and only if $\mathcal{M}$ is equal to the empty set.
- $\underline{P}$ is coherent if and only if all constraints are 'tight', i.e., if there exists a P in $\mathcal{M}$ such that $P(f) = \underline{P}(f)$ for all f in $\mathcal{K}$.

   Let us make the concept of a credal set concrete using our running example:

Yoko Tsuno's credal set

For a finite possibility space such as the one of our running example, a precise prevision P can be defined completely by the corresponding *probability mass function p* defined by $p_x = P(\{x\})$ for x in $X = \{\text{BE}, \text{JP}\}$. The set of all precise previsions can therefore be represented by the *probability simplex*—the set of all probability mass functions—on X. This set and the example probability mass function $(\frac{1}{2}, \frac{1}{2})$ is shown below left. Below right, we illustrate how Yoko Tsuno's lower prevision $\underline{P}_Y(\text{JP}) = \frac{4}{5}$ generates the credal set $\mathcal{M}_Y$: The gamble 1_{JP} as a linear function over the simplex is shown as an inclined line. This linear relationship between p—equivalently, the corresponding prevision P_p—and $P_p(1_{\text{JP}}) = P_p(\text{JP})$ transforms the bounds $\frac{4}{5} \leq P_p(\text{JP}) \leq 1$ into $\mathcal{M}_Y$.
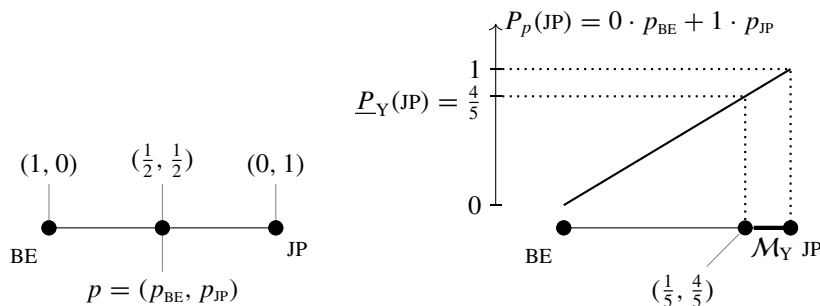
The set of extreme points $\mathcal{M}_Y^*$ of $\mathcal{M}_Y$ as probability mass functions is $\left\{ (\frac{1}{5}, \frac{4}{5}), (0, 1) \right\}$.

3.4.2   From Credal Sets to Lower Previsions

Now we assume that the agent's credal set $\mathcal{M}$ is given. Most generally this can be any set of precise previsions, i.e., any subset of $\mathcal{P}$. Often, to ensure equivalence between *coherent* lower previsions and *non-empty* credal sets, they are required to be closed and convex. In that case, a credal set is determined completely by its set of *extreme points* $\mathcal{M}^*$ in the sense that all other elements are convex mixtures of these.

   To determine the lower prevision corresponding to any credal set, we determine its value for each gamble f of interest using the *lower envelope theorem*:

$$\underline{P}(f) = \min\left\{P_p(f) : p \in \mathcal{M}\right\} = \min\left\{P_p(f) : p \in \mathcal{M}^*\right\}.$$

Let us again use the running example to provide a feeling for what this all means:

A credal set for the empty pool facing penalties

 Consider the empty pool. Because its assessment is empty, its credal set $\mathcal{M}_{EP}$ is the trivial one corresponding to all probability mass functions on $\mathcal{X} = \{BE, JP\}$. Now we add an extra element to the possibility space, 'Penalties'. Below left we show $\mathcal{M}_{EP}$ embedded in the corresponding larger probability simplex. Wiske and Yoko Tsuno decide to add the uniform probability mass function to it. Below right, you see the convex hull $\mathcal{M}_{EUP}$ of this extra probability mass function and the original credal set.
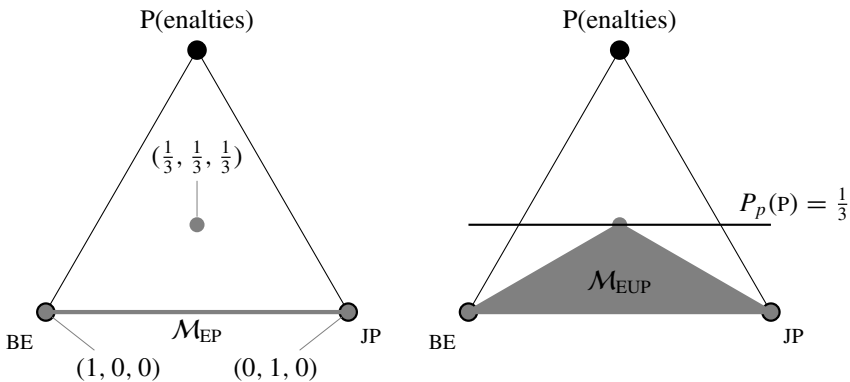
If we want to calculate lower and upper prevision values, we can here use the extreme point version of the lower and—similar—upper envelope theorem. For example, for the pool's upper probability for Penalties:

$$\overline{P}_{EP}(P) = \overline{P}_{EP}(1_P) = \max\left\{p^\top(0, 0, 1) : p \in \left\{(1, 0, 0), (0, 1, 0), (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3})\right\}\right\} = \frac{1}{3}.$$

 To make it explicit where this maximum is achieved, we above right show the line of probability mass functions p such that $P_p(P) = \frac{1}{3}$.

3.5 Basics of Conditioning

Conditioning an uncertainty model is the act of restricting attention to a subset B of the possibility space. It is often used to *update* an uncertainty model after having observed the event B [20, §6.1].

In the theory of imprecise probability, conditioning is a specific case of natural extension [17, §2.3.3], [20, §6.4.1]. In terms of acceptable gambles, conditioning on B corresponds to restricting the space of gambles to those that are zero outside B [18, §1.3.3]. For lower previsions, this translates to the following conditioning rule for all gambles f in $\mathcal{L}$:

$$\underline{E}(f \mid B) = \begin{cases} \inf_{x \in B} f(x) & \text{if } \underline{P}(B) = 0, \\ \max \left\{ \mu \in \mathbb{R} : \underline{P}(1_B(f - \mu) = 0) \right\} & \text{if } \underline{P}(B) > 0. \end{cases}$$

Conditioning a credal set $\mathcal{M}$ corresponds to taking the credal set $\mathcal{M}|B$ formed by conditioning each of the precise previsions in $\mathcal{M}$:

$$\mathcal{M}|B = \begin{cases} \mathcal{P}_B & \text{if } \exists P \in \mathcal{M} : P(B) = 0, \\ \{P(\cdot \mid B) : P \in \mathcal{M}\} & \text{if } \forall P \in \mathcal{M} : P(B) > 0. \end{cases}$$

These rules based on natural extension give vacuous conditionals whenever the lower probability of the conditioning event is zero. *Regular extension* is a less imprecise updating rule [17, §2.3.4], [18, §1.6.6], [20, App. J]: In credal set terms, it removes those precise previsions P such that $P(B) = 0$ from $\mathcal{M}$.

Let us apply the conditioning rules discussed here to our running example:

Conditioning the empty-uniform pool's credal set

We condition the empty-uniform pool's credal set on {JP, P}, i.e., Belgium not winning in regular time. Further down on the left, we show what happens if we apply natural extension: the conditional model is vacuous because $P_{(1,0,0)}(\{JP, P\}) = P_{(1,0,0)}(1_{\{JP,P\}}) = (1, 0, 0)^\top (0, 1, 1) = 0$. Further down on the right, we apply natural extension and therefore remove $P_{(1,0,0)}$ from $\mathcal{M}_{EUP}$; this results in a non-vacuous conditional credal set.
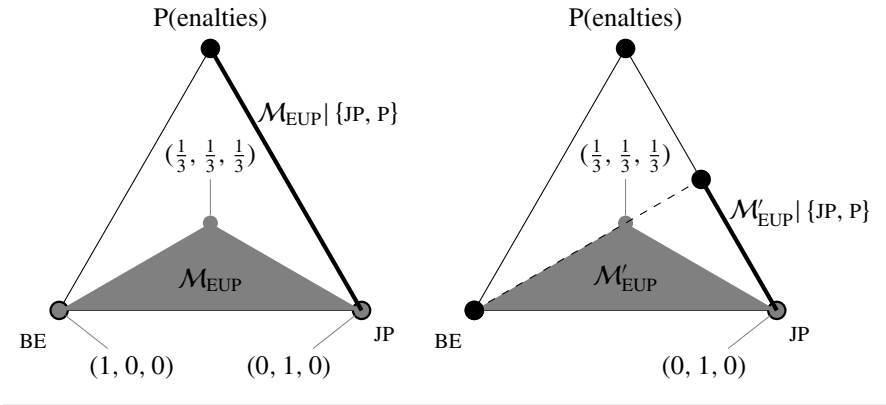
3.6 Remarks About Infinite Possibility Spaces

The theory we presented is also applicable to denumerable and continuous possibility spaces with some technical amendments to the coherence criteria and by considering only bounded gambles. However, the running example was based on finite possibility spaces, so no feeling was created for applications with infinite possibility spaces. Therefore we here give some remarks about imprecise probabilistic uncertainty models on continuous possibility spaces:

- They are mostly defined using credal sets whose extreme points are parametric distributions where the parameters vary in a set. A prime example are the imprecise Dirichlet model [21] and its generalizations [19].
- They are also commonly defined using probability mass assignments to subsets of the possibility space. This is in some way a reduction to the finite case. Examples are belief functions [13, §5.2.1.1], some P-boxes [14, §4.6.4], and NPI models [5, §7.6].
- Furthermore, models which bound some specific description of a precise prevision, such as cumulative distribution functions and probability density functions, are also popular in some domains. The extreme points of their credal set are, however, not known. General P-boxes [15] and lower and upper density functions [20, §4.6.3] are examples of this class.
- Calculating lower and upper previsions—i.e., performing natural extension—can easily become difficult optimization problems, so this should be a key consideration when choosing a specific type of model.

3.7 Conclusion

This introduction to the theory of imprecise probability has prepared you for accessing the broader literature on this topic and its applications. For those that wish to apply imprecise probabilistic techniques, this text only provides the first step: You should dive into the literature and contact experts to obtain the necessary knowledge and feedback. The references of this chapter and their authors or editors provide a starting point for that.

References

1. Alessandro Antonucci, Giorgio Corani, Inés Couso, and Sébastien Destercke, editors. *ISIPTA '17: Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, 2017.
2. Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics. Wiley, 2014.
3. Thomas Augustin, Frank P. A. Coolen, Serafín Moral, and Matthias C. M. Troffaes, editors. *ISIPTA '09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, 2009.
4. Thomas Augustin, Serena Doria, Enrique Miranda, and Erik Quaeghebeur, editors. *ISIPTA '15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, 2015.
5. Thomas Augustin, Gero Walter, and Frank P. A. Coolen. Statistical inference. In Augustin et al. [2], chapter 7, pages 135–189.
6. Jean-Marc Bernard, Teddy Seidenfeld, and Marco Zaffalon, editors. *ISIPTA '03: Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, volume 18 of *Proceedings in Informatics*, Waterloo, Ontario, Canada, 2003. Carleton Scientific.
7. Frank P. A. Coolen, Gert de Cooman, Thomas Fetz, and Michael Oberguggenberger, editors. *ISIPTA '11: Proceedings of the Seventh International Symposium on Imprecise Probability: Theories and Applications*, 2011.
8. Fabio Cozman, Thierry Denœux, Sébastien Destercke, and Teddy Seidenfeld, editors. *ISIPTA '13: Proceedings of the Eight International Symposium on Imprecise Probability: Theories and Applications*, 2013.
9. Fabio Gagliardi Cozman, Robert Nau, and Teddy Seidenfeld, editors. *ISIPTA '05: Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, 2005.
10. Gert de Cooman, Fabio Gagliardi Cozman, Serafín Moral, and Peter Walley, editors. *ISIPTA '99: Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, 1999.
11. Gert de Cooman, Terrence L. Fine, and Teddy Seidenfeld, editors. *ISIPTA '01: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, Maastricht, the Netherlands, 2001. Shaker Publishing.

12. Gert de Cooman, Jiřina Vejnarová, and Marco Zaffalon, editors. *Proceedings of the Fifth International Symposium on Imprecise Probabilities: Theories and Applications*. SIPTA, Action M Agency for SIPTA, 2007.
13. Sébastien Destercke and Didier Dubois. Other uncertainty theories based on capacities. In Augustin et al. [2], chapter 5, pages 93–113.
14. Sébastien Destercke and Didier Dubois. Special cases. In Augustin et al. [2], chapter 4, pages 79–92.
15. Scott Ferson, Vladik Kreinovich, Lev Ginzburg, Davis S. Myers, and Kari Sentz. Constructing probability boxes and dempster-shafer structures. Technical Report SAND2002-4015, Sandia National Laboratories, 2002.
16. Enrique Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.
17. Enrique Miranda and Gert de Cooman. Lower previsions. In Augustin et al. [2], chapter 2, pages 28–55.
18. Erik Quaeghebeur. Desirability. In Augustin et al. [2], chapter 1, pages 1–27.
19. Erik Quaeghebeur and Gert de Cooman. Imprecise probability models for inference in exponential families. In Cozman et al. [9], pages 287–296.
20. Peter Walley. *Statistical reasoning with imprecise probabilities*, volume 42 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1991.
21. Peter Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society B: Methodological*, 58(1):3–57, 1996.