



# Identification of Imprecision in Data Using $\epsilon$ -Contamination Advanced Uncertainty Model

Keivan Shariatmadar<sup>1</sup>(✉), Hans Hallez<sup>1</sup>, and David Moens<sup>2</sup>

<sup>1</sup> M-Group, Campus Bruges, KU Leuven, Bruges, Belgium  
{keivan.shariatmadar,hans.hallez}@kuleuven.be  
<https://iiw.kuleuven.be/brugge/m-group>

<sup>2</sup> LMSD, Campus De Nayer Sint-Katelijne-Waver, KU Leuven, Leuven, Belgium  
david.moens@kuleuven.be  
<https://www.mech.kuleuven.be/en/mod>

**Abstract.** One of the importance of the contamination uncertainty model is to consider in-determinism in the uncertainty. We consider this advanced property and develop two methods. These methods identify if there is imprecision in a given model or data. In the first approach, we build two different—a probability distribution and an interval—models for a test function  $f$  via given data/model. Then, we identify the level of imprecision by assessing, so-called model trust,  $\epsilon \in (0, 1)$  in the contamination model whether the weight is higher for the probabilistic/interval model or not. In the second approach, we calculate the lowest and highest previsions for the test function and identify the imprecision interval out of them. We further discuss and show the idea via two simple production and clutch design problems to illustrate our novel results.

**Keywords:** Imprecision ·  $\epsilon$ -Contamination · Uncertainty · Indecision

## 1 Introduction

Dealing with uncertainty is one of the problems which is needed for the problems under uncertainty. The uncertainty is present because of lack of information or data. One of the uncertainty models is probabilistic (data-driven or analytical) model. These models' intentions are to represent e.g., agents' beliefs (agent like human, machines, or robots) about the domain/area they are operating in, which describe and even determine the actions they will take or decide in a diversity of situations or realisations [38]. Probability theory provides a normative system for *reasoning* and decision making in the face of uncertainty. Bayesian or precise probability models have the property that they are purely *decisive* i.e., a Bayesian agent always has an optimal choice when faced with several alternatives, whatever his state of information is, see e.g., [19, 38]. While many may view this as an advantage, it is not always realistic. There are two problems, Gilboa [17] offers historical surveys with (precise) probabilities as a

model to describe uncertainty: (i) the interpretation is not clear or at least, the consequences in the real world are not clear. Therefore, we want an operational and behavioural model (ii) the model is unique and static while the real model behaviour is dynamic. In any precise decision problem, there is always an optimal solution. You can—beholding some degenerate cases—decide between two actions. The idea whether there is a fair price or is not (either to accept/buy or reject/sell a gamble) is not vital, the possibility of *indecision* is rather important [19, 38]. Imprecise probability (data-driven, grey/white box) models deal with said issues by explicitly allowing for indecision while retaining the normative, coherent stance of the Bayesian approach, see for more details, [5, 19, 38, 42, 44].

In this paper, our main goal is to answer a question about the existence of the imprecision in a data or model i.e., how to know that there is imprecision in the uncertainty made via the given data or model? In this section, we describe the advanced uncertainty modelling in depth via some simple examples to understand the concepts and especially the generic theory of lower and upper previsions. In our recent works [36–40], we have focused on the novel approach to make decisions under different types of imprecise uncertainties in linear optimisation problems (as one of the applications). We proposed two different solutions under two decision criteria—Maximinity and Maximality i.e., the worst-case solutions (the least risky solutions) and less conservative solutions (more optimal solutions). With these approaches, we can always decide based on the applications and preferences (from the final decision maker) to choose whether the more optimal (more risky) solutions or less risky (less optimal) solutions<sup>1</sup>. In the next Sect. 1.1. first, we give an overview of the state-of-the-art and history about the uncertainty. Second, in Sect. 2.1. we explain the uncertainty briefly under Walley’s integration [42].

### 1.1 Literature Status and History

There is a long history about using imprecise probability models starting from the middle of the 19<sup>th</sup> century [38]. For instance, in *probabilistic logic*: it was already known to George Boole [4] that the result of probabilistic inferences may be a set of probabilities (an imprecise probability model), rather than a single probability. In 1920, Keynes [22] worked on an explicit interval estimate method to probabilities. Work on imprecise probability models proceeded in the 20<sup>th</sup> century, by A. Kolmogorov [23] in 1933, B. Koopman [24] in 1940, C. A. B. Smith [41] in 1961, I. J. Good [18] in 1965, A. Dempster [13] in 1967, H. Kyburg [21] in 1969, B. de Finetti [16] in 1975, G. Shafer [34] in 1976, P. M. Williams [48] in 1978, I. Levi [26] in 1980, P. Walley [42] in 1991, T. Seidenfeld [33], and G. de Cooman [5, 44] in 1999. In 1990, P. Walley’s published the reference book: *Statistical Reasoning with Imprecise Probabilities* [42] representing the theory of *imprecise probability*. He also interpreted the subjective probabilities as accepting/buying and rejecting/selling prices in gambling. In 1990 some

<sup>1</sup> The risk is the distance between the worst-case solution and the less conservative solutions e.g., in the linear optimisation problem, the risk is the distance between the objective function at maximin point and the maximal solutions.

important works published by Kuznetsov [25] and Weichselberger [45, 46] about the interval probabilities. Also Weichselberger generalizes the Kolmogorov's work [23] in 1933. In 2000, R. Fabrizio [30] presented the robust statistics. In 2004, T. Augustin [1] provided non-parametric statistics. In 2008, the important concept about Choquet integration is proposed by G. de Cooman [9]. This work together with the work of P. Huber [20] about two-monotone and totally monotone capacities have been the foundation of artificial intelligence. Moreover, in 2008, G. de Cooman and F. Hermans [8] proposed imprecise game theory (as the extension of the work of Safer and Vovk [35]). Dealing with missing or incomplete data, leading to so-called partial identification of probabilities, is proposed by G. de Cooman and C. F. Manski [10, 27]. Another application in network domain so-called *credal nets* were proposed by F. Cozman [6, 7] which are essentially Bayesian nets with imprecise conditional probabilities.

The paper is organised as follows. In the next Sect. 2 we explain the theory of imprecise probability and show the differences between precise and imprecise uncertainties via several simple examples. An advanced— $\epsilon$ -contamination—model as well as two novel methods to identify imprecision are discussed in Sect. 3. In Sect. 4 we propose a numerical production problem to illustrate the results. We conclude and discuss the future works in Sect. 5.

## 2 Uncertainty

Generally, uncertainty is the consequence of lack of data, information, or knowledge. Conventional methods of introducing uncertainty into a problem, ignore the following cases: (a) imprecision, (b) mixed or combined precise and imprecise models, or (c) choosing best imprecise models for the available amount of data. In this paper, we consider (a) and (b) to propose two methods to identify if there is imprecision in a given uncertainty model or not. In this section, we first explain the difference between precise and imprecise uncertainty. To understand this better, we illustrate these concepts via several simple examples. Second, we use define a prevision operator to measure the uncertainty. We interpret lower and upper prevision operators to quantify the imprecise uncertainty. Finally, we define an advanced mixed/combined model to identify imprecision in a given uncertainty model (analytical or data-driven model) in two ways.

### 2.1 Interpretation of Lower and Upper Previsions

Most of the above mentioned works on imprecise probability theory was introduced by Walley [42]. In this paper, we follow the terminology and school of thought of Walley [42, 43] who follows the tradition of Frank Ramsey [29], Bruno de Finetti [12] and Peter Williams [50] in trying to establish a rational model for a subject's beliefs and reasoning. In the subjective interpretation of Walley, the upper and lower previsions/expectations for gamblers are seen as prices. A gambler's highest desirable buying price and the lowest desirable selling price, respectively. In gambling, which is about exchanging of gambles, assume that a

gambler (decision maker) wants to make a profit whether (s)he wants/accepts to buy or sell a gamble. By knowing the highest desirable price to buy the gamble and the lowest desirable price to sell the gamble, (s)he can make any desirable decision to not to lose money.

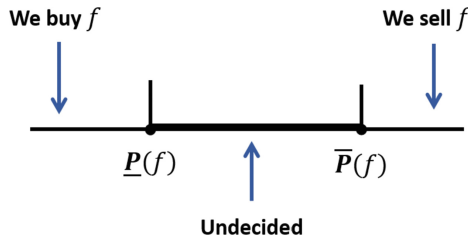
Generally, a decision maker's lower prevision/expectation  $\underline{P}(\cdot)$  is the highest acceptable price  $\alpha$  to buy a gamble/utility function  $f$ . In other words,  $\underline{P}(f)$  is supremum price to buy the gamble  $f$ . Mathematically  $\underline{P}(f)$  is defined as:

$$\underline{P}(f) := \sup_{\alpha \in \mathbb{R}} \{\alpha : f - \alpha \geq 0\}, \quad (1)$$

and the upper prevision/expectation  $\overline{P}(\cdot)$  is the lowest acceptable price  $\beta$  to sell the gamble  $f$ . In other words,  $\overline{P}(f)$  is infimum price to sell the gamble  $f$ . Mathematically  $\overline{P}(f)$  is defined as:

$$\overline{P}(f) := \inf_{\beta \in \mathbb{R}} \{\beta : \beta - f \geq 0\}. \quad (2)$$

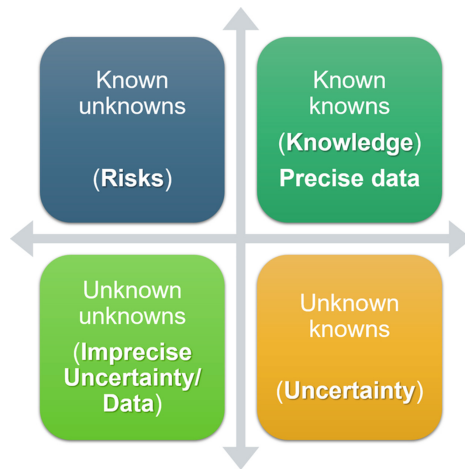
In classical probability theory, the upper and lower previsions are coincided:  $\underline{P}(f) = \overline{P}(f) := P(f)$ . Then  $P(f)$  is interpreted as the gambler's *fair price* for the gamble  $f$ . The price that the decision-maker accepts to buy  $f$  for any lower price and sell it for any higher price than  $P(f)$ . The gap between  $\underline{P}(f)$  and  $\overline{P}(f)$  is called *imprecision* or *indecision*. This is the main difference between precise and imprecise probability theories—as shown in Fig. 1, imprecise models allow for *indecision*/imprecision. Such gaps arise naturally e.g., in betting markets which happen to be financially illiquid due to asymmetric information, for more information see [21, 26]. As an interpretation, for instance in gambling (which is about exchanging of a gamble  $f$ ),  $\overline{P}(f)$  is the lowest desirable price to sell the gamble  $f$ . In other words, if a gambler knows the lowest acceptable price of a gamble then (s)he can accept any higher price than  $\overline{P}(f)$ . To explain the importance and deeper view about the in-deterministic uncertainty, in the next Sect. 2.2., different types of uncertainty, as well as some simple examples, will be talked to clarify the distinction between precise and imprecise uncertainty. In Sect. 3 a general overview about modelling uncertainty via one of the advanced models called  $\epsilon$ -contamination as well as two methods (imprecision identification methods) will be discussed. Next, a simple example will be given to illustrate the results in Sect. 4. Conclusions and further discussions will be in Sect. 5.



**Fig. 1.** Highest buying and lowest selling prices for a gamble  $f$

## 2.2 Classification of Uncertainty

There are four levels of certainty (or uncertainty) about knowledge or data. In Fig. 2, four levels of these certainties or uncertainty are illustrated from known knowns (knowledge) towards unknown unknowns (imprecise uncertainty). Our main focus here is on the *Unknown unknowns* where the unknown data is not precise. In other words, the probability of an event or a phenomenon is vague. In real-life problems, the nature of the uncertainties are usually imprecise uncertainties and one of the sources of the imprecisions, in which we have researched about, is human, also weather, traffic, and so on, [39]. One of the interesting purposes in almost all of those real-life problems is to find the best choice under some conditions dealing with the uncertainties. In other words, one of the major problems is to make the best (optimal) decision based on the restrictions (uncertainty, constraints, and so on) within some criteria. Mathematically, the idea can be formulated as optimising a goal function under an uncertain domain given by constraints. But the important point is how to deal with the uncertainty? Even more importantly, how to know that the uncertainty is not deterministic? To understand deeper the idea of the existence of indeterminism in the uncertainty, let's point out three real-life examples.



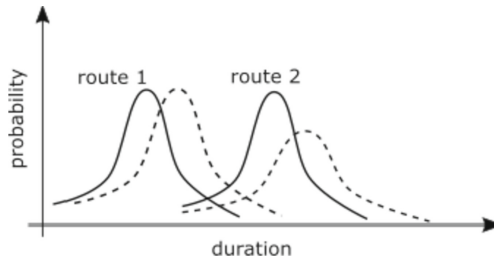
**Fig. 2.** Precision vs. imprecision

## 2.3 Probability Under Different Conditions—Travelling to Work

Assume the problem of driving a car each day from home to work and back over a (long) distance. Consider there are two possible routes. Typically, one would measure the duration of travel for both routes over some period, let's say a year and compute probability or cumulative distribution function (CDF) from the data. The goal is—using the computed CDF results in some tool—to

decide, based on the probability, which route is beneficial. As we have seen from a real database [36] sent by a factory here in Belgium<sup>2</sup>, the CDF functions differ and are not unique. Consequently, a single CDF function cannot capture the true distribution because of the indeterministic parameters influencing this duration, e.g., weather conditions, human (driver) mood, or traffic status that might change during travel which one path might be highly influential by these weather, driver, or traffic conditions in contrast to the other path. The variation on the CDF of one path might be much higher than for the other which is not possible to model via one single CDF. It is, therefore, better to capture this uncertainty by use of an advanced model (considering the indeterminism).

One of the best models to describe the imprecise uncertainty for this problem is sets of distributions functions which is called *probability box (p-box)* [15]. This model is developed and discussed briefly under an optimisation problem in [37], which is the most informative model. Another alternative is to use the contamination model, which is simpler and doesn't require lots of data, we will discuss it in the next section. In this case, every CDFs is collected in a set which is bounded from below and above, called *upper* and *lower* bounds, where can model *variations* in the probabilities (imprecision). The variation on the probability of the duration for both routes is illustrated in Fig. 3 where the full and dashed lines represent the probability under different conditions. In the case of the driving example, subset division can be made based on whether or traffic conditions (obtained from a weather or traffic database), as well.



**Fig. 3.** Probability density of the duration for both routes 1 and 2

This, in turn, allows for more robust decision making in the future (which path to take) based on the imprecision in the data, captured by the advanced models<sup>3</sup>. Again, the main question is how to measure this imprecision and find out there is imprecision in the data (or model), generically?

<sup>2</sup> Because of the confidentiality about the agreement, we can not make the names and details of the database public unless under an official confidentiality agreement.

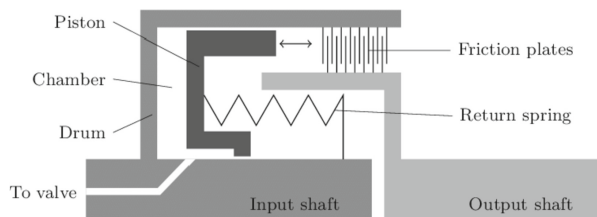
<sup>3</sup> Generally, p-box uncertainty model, described above, belong to *coherent upper* and *lower previsions* family, see e.g., for details and terminology [28, 42].

## 2.4 Probability Under Different Conditions—Diagnosis and Treatment

Logical decision making is a major part of all sciences, engineering and decision-based professions, where scientists, engineers or specialists apply their knowledge or beliefs in a given area to make optimal decisions. However, the decision under uncertainty is one of more advance topics compared to the deterministic decisions. Even more challenging where the uncertainty is not precise. For example, in the medical science area, decision making often involves a diagnosis and the selection (decision) of appropriate (optimal) treatment under a vague data—meaning, the data is not large enough or incomplete because of several restrictions such as expensive tests, test-case limitation, missing data, or unknown unmeasurable parameters, to gather enough data—where we call *life involved* (high-risk) problems. The nature of uncertainty is not unique. In other words, for instance, the uncertainty is not the same from one patient to another. In these kinds of areas, when the uncertainty is imprecise, we do not have a single (optimal) decision to make, however, it is very important to know at least the extreme cases e.g., the *worst/best cases*<sup>4</sup>.

## 2.5 Probability Under Different Conditions—Clutch Design

Another example, in the mechanical engineering area, is decision making (usually) about a design of a component under some conditions such as selections of right parameters for a design—for instance, a clutch design—e.g., diameters, friction disks, friction coefficient (uncertain parameter), torque capacity, speed, gear parameters, cooling system parameters (uncertain parameter), and so on. To design a safe clutch pack, in one hand, the engineer needs to make a safe decision i.e., tries to find the worst-case solution to avoid the risks, on the other hand, concerning the total cost of ownership, he/she needs to decide to have minimum cost i.e., less conservative solutions. In both examples, the nature of uncertainty is not unique. In other words, for instance, the uncertainty is not the same from one clutch to another e.g., the friction coefficient is not known and is changing in different temperature ranges (coming from energy loss by friction or oil condition) and different geometry see Fig. 4.



**Fig. 4.** Wet-plate clutch of an automatic transmission

<sup>4</sup> It is also interesting to know what are possible *less conservative* cases/decisions.

Modelling those unknowns is not possible via classical uncertainty because of the imprecision like the uncertainties. Knowing in advance, the presence of the imprecision might help to choose the right uncertainty model and have more robust and stable (optimal) decisions. Then the important question is that is there an imprecision in a problem or data? How to find out that there are random fluctuations in a problem/data? We discuss this in detail in the next section.

### 3 $\epsilon$ –Contamination Model

To a Bayesian<sup>5</sup> analyst, the distinction between fixed, random, and mixed models boils down to a specification of the number of stages in a given hierarchical model. One of these mixed models is called  $\epsilon$ -contamination model. This model is more advanced than the interval model [40] i.e., it links the precise model to the imprecise model. This model is recommended to be used to also analyse if there is imprecision in a given uncertainty model or not<sup>6</sup>. Furthermore, the  $\epsilon$ -contamination model is easier to build as well as implement compared to the p-box (or other imprecise models e.g., possibility distribution model [40]). In literature [2], several classes of prior distribution have been proposed but the most commonly used one is the contamination class e.g., works of Good [18], Huber [20], Dempster [14], Rubin [31], and Berger [3] to mention a few. In particular, it is concerned with what they call the posterior robustness<sup>7</sup>. The idea is to acknowledge the prior uncertainty by specifying a class/set  $M$  of possible prior distributions and then investigating the robustness of the posterior distribution as the prior varies over  $M$ . It had been mentioned by Berger [3] and Huber [20], to work with the contamination class of priors when investigating posterior robustness. They proposed the contamination class of combining elicited prior—termed the base prior—with a contamination class of arbitrary priors. These approaches are popular with Bayesian sensitivity analysis—first, to elicit an additive probability measure  $P$ , and then consider possible inaccuracies in the assessment<sup>8</sup> of  $P$ , [42]. Those contamination models, achieve statistical efficiency and robustness simultaneously, however, not much attention has been paid to this framework in non-deterministic advanced uncertainty cases (pure non-probabilistic such as intervals or high-dimensional cases like  $\epsilon$ -contamination or probability box). In the next section, we explain the  $\epsilon$ -contamination model for a given probability measure  $E$  and an imprecise interval model  $\underline{E}$ .

#### 3.1 Definition

$\epsilon$ -contamination model  $\underline{P}(\cdot)$  is described as a convex combination of two uncertainty models: (i) linear prevision model—Probabilistic model, e.g., Normal

<sup>5</sup> In this paper, without any intention, we call a researcher who works on the deterministic uncertainty framework, a Bayesian analyst.

<sup>6</sup> We can decide if a pure precise model could be suitable or not.

<sup>7</sup> Which was different from the robustness defined by White [47].

<sup>8</sup> The second step is called constructing a neighbourhood set of  $P$ .



distributed model  $E$ , and (ii) lower prevision (imprecise) model e.g., interval vacuous model  $\underline{E}$ , which is described as follows:

$$\underline{P}(f) = (1 - \epsilon)E(f) + \epsilon\underline{E}(f) \quad (3)$$

where  $E$  is the set of dominating linear previsions by  $\underline{E}$  i.e.,  $E \in \mathcal{M}\{\underline{E}\} = \{E : \forall f \in \mathcal{L}(\mathcal{Y}), E(f) \geq \underline{E}(f)\}$ , for a given interval  $[a, b] \subset \mathbb{R}$ , the  $\underline{E}(f) := \min_{y \in [a, b]} f(y)$ , and  $0 < \epsilon < 1$  is called (here) level of *model-trust/importance*<sup>9</sup>. One question is, how to build or get the  $\epsilon$ -contamination model? Let's consider a simple example. We need to build two models (out of given data or model), one probabilistic and one imprecise model. Assume, there is a %60 chance (precise model) of having heavy traffic in a road  $A$  around time  $t$ , where  $t$  varies between 1:00 and 2:00 o'clock, i.e.,  $t \in [1, 2]$  hours (imprecise model). We are not sure about the time  $t$ : sometimes  $t = 13 : 00$  and sometimes  $t = 14 : 00$ . Suppose we have an equal belief to the precise and the imprecise models, i.e.,  $\epsilon = 0.5$ . Therefore, the uncertainty model for a given test-function  $f$  in this problem becomes the average of both models,

$$\underline{P}(f) = 0.5E(f) + 0.5\underline{E}(f) = 0.5(E(f) + \underline{E}(f)).$$

### 3.2 Rationale

One of the important properties for this model is that this model considers both probabilistic (a probability measure  $E$ ) and non-probabilistic (an interval  $[a, b]$ ) models where we can tune it by choosing the right trust value ( $\epsilon$ ). This needs some expert knowledge or historical information about selecting the right level. However, since the problem is convex, we can always generate all possible outcomes for all  $\epsilon \in (0, 1)$ , mathematically<sup>10</sup>. In many real-life problems e.g., said traffic problem, we have both, a variation, and a guess or chance in the real-life problems. The variation can be found via a robustness test or experiment. By the time, with enough information, via e.g., sensitivity analysis, or reliability tests/experiments we can also obtain the percentage of beliefs about the unknown parameter, event, realisation, or phenomenon. Mathematically, to find an interval model we need the lower and upper values of the realisation which is varying between them i.e., the two boundary values are enough to build the interval. For the probabilistic model, normally, we need more data to get those percentages and guesses. But to consider both models, current classical (precise) uncertainty models are not able to handle and deal with both models, simultaneously. We believe that to start moving towards advanced uncertainty (after interval case) the  $\epsilon$ -contamination model is one of the best models to use in many real problems and applications [40].

<sup>9</sup>  $\epsilon$  is also called *tuning* parameter or weight factor.

<sup>10</sup> For instance, we can easily calculate the outcome of the convex combination of two points which is a line between the two points.

### 3.3 Imprecision Identification–Method I

Another importance of the  $\epsilon$ -contamination model is to distinguish between imprecision and precision i.e., the question is how to identify the imprecision in a given problem or data? How to find out that there are random fluctuations in a problem/data? The answer is given via this  $\epsilon$ -contamination model as follows. From a given model or available data (database), we first assume a known outcome  $\underline{P}(f)$  for a given real-valued test function  $f$ <sup>11</sup>. Then we build a probability distribution as well as the variation interval for the test function  $f$  via given data. By calculation the expected values for  $f$  in both cases–Probability and Interval—we know  $E(f)$  and  $\underline{E}(f)$ . Finally, we solve the Eq. (3) to find the  $\epsilon$ . If  $0 \leq \epsilon \leq 0.5$  then there is less chance (less than %50) of having imprecision in the data or the model otherwise, there is imprecision with the probability of higher than %50.

### 3.4 Imprecision Identification–Method II

Assume an interval  $[a, b] \in \mathbb{R}$  and a probability distribution are given via a data (a database) or model. Another method to identify whether there is imprecision in the given data (via a database) or model<sup>12</sup>, is to calculate the lower as well as the upper previsions for a chosen test function  $f$  as follows via (3):

$$\text{the lower prevision } \underline{P}(f) \text{ is : } (1 - \underline{\epsilon})E(f) + \underline{\epsilon}\underline{E}(f), \text{ and} \quad (4)$$

$$\text{the upper prevision } \overline{P}(f) \text{ is : } (1 - \overline{\epsilon})E(f) + \overline{\epsilon}\overline{E}(f) \quad (5)$$

where  $\underline{\epsilon}, \overline{\epsilon} \in (0, 1)$ , the upper prevision in the given interval  $[a, b]$  is  $\overline{E}(f) := \max_{y \in [a, b]} f(y)$ . If  $\exists \epsilon^* = \max\{\underline{\epsilon} \in \underline{\mathcal{E}}, \overline{\epsilon} \in \overline{\mathcal{E}}\}$  where  $\underline{\epsilon}$  satisfies in (4),

$$\underline{\epsilon} \in \underline{\mathcal{E}} := \{\underline{\epsilon}_i : \underline{\epsilon}_i \in (0, 1)\} \text{ and } \overline{\epsilon} \text{ satisfies in (5)}$$

$$\overline{\epsilon} \in \overline{\mathcal{E}} := \{\overline{\epsilon}_i : \overline{\epsilon}_i \in (0, 1)\}$$

such that  $\underline{P}(f) < \overline{P}(f)$  then there is imprecision in the uncertainty model with probability of  $\epsilon^*$ , and the imprecision interval is  $[\underline{P}(f), \overline{P}(f)]$ .

## 4 Numerical Example

### 4.1 Chocolate Production Problem

Consider a chocolate manufacturer which produces two types of chocolate  $\mathfrak{A}$  and  $\mathfrak{B}$ . Both chocolates require Milk and Cacao only (for simplicity). Each unit of

<sup>11</sup> This can be done by an expert or historical data, to have the simplest case that we know the outcome of the realisation which is given via both uncertainty models.

<sup>12</sup> These models as discussed, could be estimated from the existing data or the available model under uncertainty. These estimations are not the aim of this paper but for instance the interval can be estimated via a sensitivity analysis and the probability distribution can be calculated via a normal distribution fitted to the data/model.

$\mathfrak{A}$  requires 1 unit of Milk and  $Y_1$  units of Cacao. Each unit of  $\mathfrak{B}$  requires 1 unit of Milk and  $Y_2$  units of Cacao. The company capacity has a total of 12 units of Cacao (no limit for milk). On each sale, the company makes profit of €1 per unit  $\mathfrak{A}$  and €1 per unit  $\mathfrak{B}$ . The goal is to maximise profit (how many units of  $\mathfrak{A}$  and  $\mathfrak{B}$  should be produced respectively). Mathematically, the problem is modelled as a linear programming problem:

$$\begin{aligned} \max \quad & x_1 + x_2 \\ \text{s.t} \quad & \begin{cases} Y_1 x_1 + Y_2 x_2 \leq 12 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned} \quad (6)$$

Assume that there are two sources of uncertainties: (i) a priori probabilistic information (about dealing with *experiments* that numerically describes the number of desired outcomes)—obtained by a historical data, expert knowledge, or sensitivity analysis—and (ii) a set of realisations obtained via e.g., reliability analysis (about robustness/variation). Suppose (i) the probabilistic models are given with distribution functions  $N_1 := N(\mu_1 = 7.5, \sigma_1 = 1)$  and  $N_2 := N(\mu_2 = 9.5, \sigma_2 = 1)$  about how likelihood we need the amount of cacao for both chocolates  $\mathfrak{A}$  and  $\mathfrak{B}$  in one year and (ii) also we know that the amounts of cacao for both chocolates  $\mathfrak{A} \in [7, 8]$  and  $\mathfrak{B} \in [9, 10]$  are varying. In other words, the problem is to maximise profit under the  $\epsilon$ -contamination uncertain constraint for Cacao, which has likelihood amounts given by the normal probability distributions and varying in the assumed lower and upper values, shown in Table 1.

**Table 1.** Uncertain Chocolate production problem— $\epsilon$ -contamination uncertainty

	Milk	Cacao	Profit per unit
$\mathfrak{A} := x_1$	1	$N(7.5, 1), [7, 8]$	1
$\mathfrak{B} := x_2$	1	$N(9.5, 1), [9, 10]$	1
Capacity	No limit	12	

Since, in this problem we do not have the lower and upper expected values then we use method II to identify if there is imprecision in this example or not. The lower and upper previsions are defined as follows,

$$(1 - \underline{\epsilon})((x_1 + x_2)|_{7.5x_1 + 9.5x_2 \leq 12}) + \underline{\epsilon}((x_1 + x_2)|_{8x_1 + 10x_2 \leq 12}), \text{ and} \quad (7)$$

$$(1 - \bar{\epsilon})((x_1 + x_2)|_{7.5x_1 + 9.5x_2 \leq 12}) + \bar{\epsilon}((x_1 + x_2)|_{7x_1 + 9x_2 \leq 12}) \quad (8)$$

For instance, to maximise the profit  $(x_1 + x_2)$ , from (7) we have:  $(1 - \underline{\epsilon})\frac{5}{3} + \underline{\epsilon}\frac{3}{2}$  and from (8) we have:  $(1 - \bar{\epsilon})\frac{5}{3} + \bar{\epsilon}\frac{12}{7}$ , where for all  $\bar{\epsilon}, \underline{\epsilon} \in (0, 1)$  the profit for upper prevision (8) is higher than (7) and the  $\epsilon^* = \max(0, 1) \approx 0.9999$ , meaning with the high probability 99.99 percent there is imprecision in the given model (6). There are many conditions such as traffic, weather, or human behaviour/mood

could affect e.g., transportations delays and consequently the exact amount of stock (Milk) or the exact availability of warehouse capacity. Furthermore, this inexact amount of stock or warehouse capacity is dynamically changing from day-to-day. Therefore, using only probabilistic (truncated) distributions for this problem will result in the suboptimal solution and the  $\epsilon$ -contamination model is the suitable model for (6).

## 4.2 Clutch Design Problem

Back to the Clutch Design example discussed in Sect. 2.5, one of the main ideas is to have a maximum torque transfer from one side of the clutch to the other side. We simplify the problem as follows. Assume we want to design the clutch to have a maximum friction torque  $\tau_f$  defined as:

$$\tau_f := \mu R N A P_k \Delta \omega \quad (9)$$

where  $R, A, N$  are the radius, area, and the number of friction disks, respectively.  $P_k$  is the internal oil clutch pressure (pushing the friction plates towards each other to close by increasing the pressure and open by decreasing it),  $\Delta \omega$  is the slip speed, and  $\mu$  is the uncertain friction coefficient. If the spring force  $f_s$  is higher than the friction force  $P_k A$ , then the clutch is open otherwise it is closed. This is controlled via the pressure  $P_k$  to create a smooth closing (opening) with less torque loss. As defined in (9), this pressure depends on the friction coefficient  $\mu$ . Currently, the friction coefficient is estimating, and it is a fixed value however there are many disturbances e.g., oil temperature, the air in the oil, centrifugal force, oil leakage, and so on, changing the friction coefficient. We use the data provided by the work of Schneider [32]. In the given test data (durability tests), we have seen that  $\mu$  is varying between 0.09 and 0.18 (interval model). Also a normal distribution function for  $\mu$  can be estimated as  $N(\mu = 0.11, \sigma = 1)$ . We calculate the lower and upper prevision for the following linear optimisation problem for a closing clutch.

$$\begin{array}{ll} \max_{\mu} & \mu R N A P_k \Delta \omega \\ \text{such that} & f_s < P_k A \\ & \mu, R, N, A, P_k, \Delta \omega \geq 0 \end{array} \quad \rightarrow \quad \begin{array}{ll} \max_t & t \\ \text{such that} & f_s < P_k A, \\ & \mu R N A P_k \Delta \omega \geq t \end{array} \quad (10)$$

The lower and upper previsions are defined as follows:

$$\begin{aligned} & (1 - \bar{\epsilon}) t_{|0.11 R N A P_k \Delta \omega \geq t} + \bar{\epsilon} t_{|0.18 R N A P_k \Delta \omega \geq t} \\ & (1 - \underline{\epsilon}) t_{|0.11 R N A P_k \Delta \omega \geq t} + \underline{\epsilon} t_{|0.09 R N A P_k \Delta \omega \geq t} \end{aligned}$$

Assume,  $R = 0.1$  m,  $N = 2$ ,  $A = 0.001132$  m<sup>2</sup>,  $P_k = 5$  bar,  $\Delta \omega = 7.5$  m/s, then

$$(1 - \bar{\epsilon}) t_{|93.39 \geq t} + \bar{\epsilon} t_{|152.82 \geq t} \quad (11)$$

$$(1 - \underline{\epsilon}) t_{|93.39 \geq t} + \underline{\epsilon} t_{|76.41 \geq t} \quad (12)$$

For instance, to maximise the objective  $t$ , from (11) we have:  $(1-\bar{\epsilon})93.39+\bar{\epsilon}152.82$  and from (12) we have:  $(1-\underline{\epsilon})93.39+\underline{\epsilon}76.41$ , where for all  $\bar{\epsilon}, \underline{\epsilon} \in (0, 1)$  the profit for upper prevision (11) is higher than (12) and the  $\epsilon^* = \max(0, 1) \approx 0.9999$ , meaning with the high probability 99.99 percent there is imprecision in the given model (10). So, we need to consider an imprecise model for the friction coefficient  $\mu$  rather than a fixed estimated value.

## 5 Conclusion

In this paper we consider two methods to identify if there is imprecision in a given problem under uncertainty with some degrees. The problem either is given via a database (black-box) or analytically (white-box) where there is uncertainty in either case, e.g., an unknown parameter where we know about distribution or variation in the parameter (in the model or the measured data). We use one of the advanced uncertainty— $\epsilon$ -contamination—models to identify the imprecision in the given data or model under uncertainty via two methods. If the lowest and the highest expected values on the problem are given (e.g., by a decision-maker) then we use method (I) proposed in Sect. 3.3 to search for  $\epsilon \in (0, 1)$ . Otherwise, if the expected values are not available, then we proposed method (II) discussed in Sect. 3.4 to search for the  $\epsilon^* \notin \emptyset$ . In both methods, the chance (degree) of having the imprecision is determined by the  $\epsilon$ . That is up to the final decision maker to decide whether using the imprecise uncertainty model is more optimal when the chance is low e.g., lower than %50, or not. The approach here to analyse and identify the existence of imprecision is a fundamental decision before modelling the uncertainty. By knowing that, we can decide to choose the best uncertainty model for the problem under uncertainty. This will avoid having further issues such as instability, inaccuracy, or wrong results from the model with wrong uncertainty model, and will help to have a more stable and accurate model for any decision (or design) problem. In both methods I and II, the problem is linear and convex i.e., the proposed methods are not NP-hard.

## References

1. Augustin, T., Coolen, F.: Nonparametric predictive inference and interval probability. *J. Stat. Plan. Infer.* **124**(2), 251–272 (2004). <https://doi.org/10.1016/j.jspi.2003.07.003>
2. Baltagi, B.H., Bresson, G., Chaturvedi, A., Lacroix, G.: Robust linear static panel data models using  $\epsilon$ -contamination. *Center Policy Res.* **239** (2017). <https://surface.syr.edu/cpr/239>
3. Berger, J.O.: *Statistical decision theory and Bayesian analysis*, 2nd edn. Springer, New York (1985)
4. Boole, G.: *The Laws of Thought*. Dover Publications, New York (1847, reprint 1961)
5. de Cooman, G., Aeyels, D.: Supremum preserving upper probabilities. *Inf. Sci.* **118**(1–4), 173–212 (1999)

6. Cozman, F.G.: Credal networks. *Artif. Intell.* **120**, 199–233 (2000). [https://doi.org/10.1016/S0004-3702\(00\)00029-1](https://doi.org/10.1016/S0004-3702(00)00029-1)
7. Cozman, F.G.: Graphical models for imprecise probabilities. *Int. J. Approx. Reason.* **39**(2–3), 167–184 (2005). <https://doi.org/10.1016/j.ijar.2004.10.003>
8. de Cooman, G., Hermans, F.: Imprecise probability trees: bridging two theories of imprecise probability. *Artif. Intell.* **172**, 1400–1427 (2008). <https://doi.org/10.1016/j.artint.2008.03.001>
9. de Cooman, G., Troffaes, M.C., Miranda, E.: n-Monotone exact functionals. *J. Math. Anal. Appl.* **347**(1), 143–156 (2008). <https://doi.org/10.1016/j.jmaa.2008.05.071>
10. de Cooman, G., Zaffalon, M.: Updating beliefs with incomplete observations. *Artif. Intell.* **159**(1–2), 75–125 (2004). <https://doi.org/10.1016/j.artint.2004.05.006>
11. de Finetti, B.: *Teoria delle Probabilità*. Einaudi, Turin (1970)
12. de Finetti, B.: *Theory of Probability: A Critical Introductory Treatment*. Wiley, Chichester (1974). English translation of [11], two volumes
13. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* **38**(2), 325–339 (1967). <https://doi.org/10.1214/aoms/1177698950>
14. Dempster, A.: Examples relevant to the robustness of applied inferences. In: Gupta, S.S., Moore, D.S. (eds.) *Statistical Decision Theory and Related Topics*, pp. 121–138. Academic Press (1977). <https://doi.org/10.1016/B978-0-12-307560-4.50010-7>. Supported in part by National Science Foundation Grant MCS75-01493
15. Destercke, S., Dubois, D., Chojnacki, E.: Unifying practical uncertainty representations: I. Generalized p-boxes. *Int. J. Approx. Reason.* **49**, 649–663 (2008)
16. de Finetti, B.: *Theory of Probability*, vol. 2. Wiley, Hoboken (1975)
17. Gilboa, I., Postlewaite, A.W., Schmeidler, D.: Probability and uncertainty in economic modeling. *J. Econ. Perspect.* **22**(3), 173–88 (2008). <https://doi.org/10.1257/jep.22.3.173>
18. Good, I.J.: *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. M.I.T. Press Research Monographs. M.I.T. Press, Cambridge (1965). <https://books.google.be/books?id=wxLvAAAAAAAJ>
19. Hermans, F.: *An operational approach to graphical uncertainty modelling*. Ph.D. thesis, Ghent University (2012)
20. Huber, P.J., Strassen, V.: Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Stat.* **1**(2), 251–263 (1973). <https://doi.org/10.1214/aos/1176342363>
21. Kyburg, Jr., H.E.: *Probability Theory*. Prentice-Hall, Englewood Cliffs (1969)
22. Keynes, J.M.: *A Treatise on Probability*. Macmillan, London (1921)
23. Kolmogorov, A.N.: *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Heidelberg (1933). English translation: *Foundations of the Theory of Probability*. Chelsea, New York (1950)
24. Koopman, B.: The bases of probability. *Bull. Am. Math. Soc.* **46**, 763–774 (1940)
25. Kuznetsov, V.P.: *Interval Statistical Models*. Radio i Svyaz Publication, Moscow (1991). (in Russian)
26. Levi, I.: *The Enterprise of Knowledge. An Essay on Knowledge, Credal Probability, and Chance*. MIT Press, Cambridge (1980). <https://doi.org/10.2307/2184951>
27. Manski, C.F.: *Partial Identification of Probability Distributions*. Springer, New York (2003)
28. Miranda, E.: A survey of the theory of coherent lower previsions. *Int. J. Approx. Reason.* **48**, 628–658 (2008). <https://doi.org/10.1016/j.ijar.2007.12.001>

29. Ramsey, F.P.: Truth and probability (1926). In: Braithwaite, R.B. (ed.) *The Foundations of Mathematics and other Logical Essays*, chap. VII, pp. 156–198. Taylor & Francis Group, London (1931)
30. Ríos, D., Ruggeri, F.: *Robust Bayesian Analysis*. Springer, Heidelberg (2000). <https://doi.org/10.1007/978-1-4612-1306-2>
31. Rubin, H.: Robust Bayesian estimation. In: Gupta, S.S., Moore, D.S. (eds.) *Statistical Decision Theory and Related Topics*, pp. 351–356. Academic Press, Cambridge (1977). <https://doi.org/10.1016/B978-0-12-307560-4.50023-5>
32. Schneider, T., Voelkel, K., Pflaum, H., Stahl, K.: Friction behavior of pre-damaged wet-running multi-plate clutches in an endurance test. *Lubricants* **8**(7) (2020). <https://doi.org/10.3390/lubricants8070068>
33. Seidenfeld, T.I.: The fiducial argument. Ph.D. thesis, Columbia University (1976)
34. Shafer, G.: *A Mathematical Theory of Evidence*. Princeton University Press, Princeton (1976)
35. Shafer, G., Vovk, V.: *Probability and Finance: It's Only a Game!*. Wiley, New York (2001)
36. Shariatmadar, K., Debrouwere, F., Misra, A., Versteyhe, M.: Improved uncertainty modelling of process variations in a smart industry manufacturing facility by use of probability box uncertainty models. In: *International Conference on Stochastic Processes and Algebraic Structures – From Theory Towards Applications*. Mälardalen University, Vasteras, Sweden (2019)
37. Shariatmadar, K., Versteyhe, M.: Linear programming under p-box uncertainty model. In: *Machines*, pp. 84–89. MDPI AG, TU Delft, Netherlands, August 2019. <https://doi.org/10.1109/ICCMA46720.2019.8988632>
38. Shariatmadar, K., Arrigo, A., Vallée, F., Hallez, H., Vandeveld, L., Moens, D.: Day-ahead energy and reserve dispatch problem under non-probabilistic uncertainty. *Energies* **14**(4) (2021). <https://doi.org/10.3390/en14041016>
39. Shariatmadar, K., De Ryck, M., Driesen, K., Debrouwere, F., Versteyhe, M.: Linear programming under  $\epsilon$ -contamination uncertainty. *Comput. Math. Methods* **2**(2) (2020). <https://doi.org/10.1002/cmm4.1077>
40. Shariatmadar, K., Versteyhe, M.: Numerical linear programming under non-probabilistic uncertainty models—interval and fuzzy sets. *Int. J. Uncertainty Fuzziness Knowl. Based Syst.* **28**(03), 469–495 (2020). <https://doi.org/10.1142/S0218488520500191>
41. Smith, C.A.B.: Consistency in statistical inference and decision. *J. Roy. Stat. Soc. Ser. B Methodol.* **23**(1), 1–37 (1961). <http://www.jstor.org/stable/2983842>
42. Walley, P.: *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, Monographs on Statistics and Applied Probability, vol. 42. Taylor & Francis (1991). <https://books.google.be/books?id=Nk9Qons1kHsC>
43. Walley, P.: Measures of uncertainty in expert systems. *Artif. Intell.* **83**(1), 1–58 (1996). [https://doi.org/10.1016/0004-3702\(95\)00009-7](https://doi.org/10.1016/0004-3702(95)00009-7)
44. Walley, P., de Cooman, G.: Coherence of rules for defining conditional possibility. *Int. J. Approx. Reason.* **21**, 63–107 (1999). [https://doi.org/10.1016/S0888-613X\(99\)00007-9](https://doi.org/10.1016/S0888-613X(99)00007-9)
45. Weichselberger, K.: The theory of interval-probability as a unifying model for uncertainty. *Int. J. Approx. Reason.* **24**, 149–170 (2000). [https://doi.org/10.1016/S0888-613X\(00\)00032-3](https://doi.org/10.1016/S0888-613X(00)00032-3)
46. Weichselberger, K.: Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept. *Physica*, Heidelberg (2001). [https://doi.org/10.1007/978-3-642-57583-9\\_3](https://doi.org/10.1007/978-3-642-57583-9_3). In cooperation with Augustin, T. and Wallner, A

47. White, H.: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–838 (1980). <https://doi.org/10.2307/1912934>
48. Williams, P.M.: On a new theory of epistemic probability. *Br. J. Philos. Sci.* **29**(4), 375–387 (1978). <https://doi.org/10.1093/bjps/29.4.375>
49. Williams, P.M.: Notes on conditional previsions. Technical report, School of Math and Physics Science, University of Sussex, February 1975. Published as [50]
50. Williams, P.M.: Notes on conditional previsions. *Int. J. Approx. Reason.* **44**, 366–383 (2007). <https://doi.org/10.1016/j.ijar.2006.07.019>. Original: [49]

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

