# Practical Representations of Incomplete Probabilistic Knowledge

C. Baudrit [a,*], D. Dubois [b]

[a] *Laboratoire Mathématiques et Applications, Physique Mathématique d'Orléans. Université d'Orléans, rue de Chartres BP 6759, Orléans 45067 cedex 2, France.*

[b] *Institut de Recherche en Informatique de Toulouse. Université Paul Sabatier, 118 route de Narbonne 31062 Toulouse Cedex 4, France.*

**Abstract**

The compact representation of incomplete probabilistic knowledge which can be encountered in risk evaluation problems, for instance in environmental studies is considered. Various kinds of knowledge are considered such as expert opinions about characteristics of distributions or poor statistical information. The approach is based on probability families encoded by possibility distributions and belief functions. In each case, a technique for representing the available imprecise probabilistic information faithfully is proposed, using different uncertainty frameworks, such as possibility theory, probability theory, and belief functions, etc. Moreover the use of probability-possibility transformations enables confidence intervals to be encompassed by cuts of possibility distributions, thus making the representation stronger. The respective appropriateness of pairs of cumulative distributions, continuous possibility distributions or discrete random sets for representing information about the mean value, the mode, the median and other fractiles of ill-known probability distributions is discussed in detail.

*Key words:* Imprecise Probabilities, Possibility, Belief functions, Probability-Boxes.

## 1 Introduction

In risk analysis, uncertainties are often captured within a purely probabilistic framework. It suggests that all uncertainties whether of a random or an epistemic nature should be represented in the same way. Under this assumption, the uncertainty associated with each parameter of a mathematical model of some phenomenon can

* Corresponding author.
*Email addresses:* baudrit@irit.fr (C. Baudrit), dubois@irit.fr (D. Dubois).

be described by a single probability distribution. According to the frequentist view, the occurrence of an event is a matter of chance. However, not all uncertainties are random nor can be objectively quantified, even if the choice of values for parameters is based as much as possible on on-site investigations. Due to time and financial constraints, information regarding model parameters is often incomplete. For example, it is quite common for a hydrogeologist to estimate the numerical values of acquifer parameters in the form of confidence intervals according to his/her experience and intuition (i.e expert judgment). We are then faced with a problem of processing incomplete knowledge.

Overall, uncertainty regarding model parameters may have essentially two origins. It may arise from randomness due to natural variability of observations resulting from heterogeneity (for instance, spatial heterogeneity) or the fluctuations of a quantity in time. Or it may be caused by imprecision due to a lack of information resulting, for example, from systematic measurement errors or expert opinions. As suggested by Ferson and Ginzburg [20] and more recently developed by Helton et al.[25], distinct representation methods are needed to adequately tell random variability (often referred to as "aleatory uncertainty") from imprecision (often referred to as "epistemic uncertainty").

A long philosophical tradition in probability theory dating back to Laplace demands that uniform distributions should be used by default in the absence of specific information about frequencies of possible values. For instance, when an expert gives his/her opinion on a parameter by claiming: "I only know that the value of $x$ lies in an interval $A$", the uniform probability with support $A$ is used. This view is also justified on the basis of the "maximum entropy" approach [24]. However, this point of view can be challenged. Adopting a uniform probability distribution to express ignorance is questionable. This choice introduces information that in fact is not available and may seriously bias the outcome of risk analysis in a non-conservative manner [20]. A more faithful representation of this knowledge on parameter $x$ is to use the characteristic function of the set $A$, say $\pi$ such that $\pi = 1$, $\forall\ x \in A$ and 0 otherwise. This is because $\pi$ is interpreted as a possibility distribution that encodes the family of all probability distribution functions with support in $A$ (Dubois and Prade [16]). Indeed, there exists an infinity of probability distributions with support in $A$, and the uniform distribution is just one among them.

In the context of risk evaluation, the knowledge really available on parameters is often vague or incomplete. This knowledge is not enough to isolate a single probability distribution in the domain of each parameter. When faced with this situation, representations of knowledge accepting such incompleteness look more in agreement with the available information. Of course, the Bayesian subjectivist approach maintains that only a standard probabilistic representation of uncertainty is rational, but this claim relies on a betting interpretation that enforces the use of a single probability distribution, in the scope of decision-making, not with a view to faithfully report the epistemic state of an agent (See Dubois, Prade and Smets [18] for more discussions on this topic). In practice, while information regarding variability is best conveyed using probability distributions, information regarding imprecision is more faithfully conveyed using families of probability distributions [32]. At the practical, level such families are most easily encoded either by probability boxes (upper and lower cumulative probability functions [21] [22]) or by possibility distributions (also called fuzzy intervals) [17,12] or yet by belief functions of Shafer [30].

This article proposes practical representation methods for incomplete probabilistic information, based on formal links existing between possibility theory, imprecise probability and belief functions. These results can be applied for modelling inputs to uncertainty propagation algorithms.

In Section 2, we recall basics of probability-boxes ( upper & lower cumulative distribution functions ), possibility distributions and belief functions. We also recall the links between these representations. All of them can encode families of probability functions. In section 3, the expressive power of probability boxes and possibility distributions is compared. In section 4, some results on the relation between prediction intervals and possibility theory are recalled. It allows a stronger form of encoding of a probability family by a possibility distribution, whereby the cuts of the latter enclose the prediction intervals of the probability functions. In Sections 5 and 6, we consider a non exhaustive list of knowledge types that one may meet after an information collection step in problems like environmental risk evaluation. We especially focus on incomplete non-parametric models, for which only some characteristic values are known, such as the mode, the mean or the median and other fractiles of the distribution. For each case we propose an adapted representation in terms of p-boxes, belief functions (Section 5), and especially possibility distributions (Section 6).

## 2 Formal Frameworks for Representing Imprecise Probability

Consider a measurable space $(\Omega, \mathcal{A})$ where $\mathcal{A}$ is an algebra of measurable subsets of $\Omega$. Let $\mathcal{P}$ be a set of probability measures on the referential $(\Omega, \mathcal{A})$. For all $A \subseteq \Omega$ measurable, we define:

$$\text{its upper probability} \quad \overline{P}(A) = \sup_{P \in \mathcal{P}} P(A)$$

and

$$\text{its lower probability} \quad \underline{P}(A) = \inf_{P \in \mathcal{P}} P(A).$$

Such a family may be natural to consider if a probabilistic parametric model is used but the parameters such as the mean value or the variance are ill-known (for instance they lie in an interval). It can be also obtained if the probabilistic model relies on imprecise (e.g. set-valued) statistics (Jaffray, [26]), or yet incomplete statistical information (only a set of conditional probabilities is available). In a subjectivist tradition, the lower probability $\underline{P}(A)$ can be interpreted as the maximal price one would be willing to pay for the gamble $A$, which pays 1 unit if event $A$ occurs (and nothing otherwise) [32]. Thus, $\underline{P}(A)$ is the maximal betting rate at which one would be disposed to bet on $A$. That means $\underline{P}(A)$ is a measure of evidence in favour of event $A$. The upper probability $\overline{P}(A)$ can be interpreted as the minimal selling price for the gamble $A$, or as one minus the maximal rate at which an agent would bet on $A$ [32]. That means $\overline{P}(A)$ measures the lack of evidence against $A$ since we have:

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

It is clear that representing and reasoning with a family of probabilities may be very complex. In the following we consider three frameworks for representing special sets of probability functions, which are more convenient for a practical handling. In the following we review three modes of representation of uncertainty that can be cast in the imprecise probability model.

### *2.1 Probability boxes*

Let $X$ be a random variable on $(\Omega, \mathcal{A})$. Recall that a cumulative distribution function is a non decreasing function $F : \mathbb{R} \to [0, 1]$ assigning to each $x \in \mathbb{R}$ the value $P(X \in (-\infty, x])$. This function encodes all the information pertaining to a probability measure, and is often very useful in practice.

A natural model of an ill-known probability measure is thus obtained by considering a pair $(\underline{F}, \overline{F})$ of non-intersecting cumulative distribution functions, generalising an interval. The interval $[\underline{F}, \overline{F}]$ is called a probability box (p-box) [21] [22]. A p-box encodes the class of probability measures whose cumulative distribution functions $F$ are restricted by the bounding pair of cumulative distribution functions $\underline{F}$ and $\overline{F}$ such that

$$\underline{F}(x) \leq F(x) \leq \overline{F}(x) \quad \forall x \in \mathbb{R}.$$

A p-box can be induced from a probability family $\mathcal{P}$ by:

$$\forall x \in \mathbb{R} \quad \underline{F}(x) = \underline{P}((-\infty, x])$$

and

$$\forall x \in \mathbb{R} \quad \overline{F}(x) = \overline{P}((-\infty, x]).$$

Let $\mathcal{P}(\underline{P} < \overline{P}) = \{P, \forall A \subseteq \Omega \text{ measurable}, \underline{P}(A) \leq P(A) \leq \overline{P}(A)\}$ be the probability family limited by upper $\overline{P}$ and lower $\underline{P}$ probabilities induced from $\mathcal{P}$. Clearly $\mathcal{P}$ is a proper subset of $\mathcal{P}(\underline{P} < \overline{P})$ generally. Let $\mathcal{P}(\underline{F} \leq \overline{F})$ be the probability family containing $\mathcal{P}$ and defined by:

$$\mathcal{P}(\underline{F} \leq \overline{F}) = \{P, \forall x, \underline{F}(x) \leq F(x) \leq \overline{F}(x)\}.$$

Generally, $\mathcal{P}(\underline{F} \leq \overline{F})$ strictly contains $\mathcal{P}(\underline{P} < \overline{P})$, hence also the set $\mathcal{P}$ it is built from. The probability box $[\underline{F}, \overline{F}]$ provides a bracketing of some ill-known cumulative distribution function and the gap between $\underline{F}$ and $\overline{F}$ reflects the incomplete nature of the knowledge, thus picturing the extent of what is ignored. However, as we shall see, this representation method can be very imprecise.

## *2.2 Basics of Numerical Possibility Theory*

Possibility theory [12] is relevant to represent consonant imprecise knowledge. The basic notion is the possibility distribution, denoted $\pi$, an upper semi-continuous mapping from the real line to the unit interval. A possibility distribution describes the more or less plausible values of some uncertain variable $X$. Possibility theory provides two evaluations of the likelihood of an event, for instance whether the value of a real variable $X$ does lie within a certain interval: the possibility $\Pi$ and the necessity $N$. The normalized measure of possibility $\Pi$ (respectively necessity $N$) is defined from the possibility distribution $\pi : \mathbb{R} \to [0,1]$ such that $\sup_{x\in\mathbb{R}} \pi(x) = 1$ as follows:

$$\Pi(A) = \sup_{x\in A} \pi(x) \tag{1}$$

and

$$N(A) = 1 - \Pi(\overline{A}) = \inf_{x\notin A} (1 - \pi(x)) . \tag{2}$$

- The possibility measure $\Pi$ verifies :

$$\forall A, B \subseteq \mathbb{R} \quad \Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \tag{3}$$

- The necessity measure $N$ verifies :

$$\forall A, B \subseteq \mathbb{R} \quad N(A \cap B) = \min(N(A), N(B)). \tag{4}$$

A possibility distribution $\pi_1$ is more specific than another one $\pi_2$ in the wide sense as soon as $\pi_1 \leq \pi_2$, i.e. $\pi_1$ is more informative than $\pi_2$.
A unimodal numerical possibility distribution may also be viewed as a nested set of confidence intervals, which are the $\alpha$-cuts $[\underline{x}_\alpha, \overline{x}_\alpha] = \{x, \pi(x) \geq \alpha\}$ of $\pi$. The degree of certainty that $[\underline{x}_\alpha, \overline{x}_\alpha]$ contains $X$ is $N([\underline{x}_\alpha, \overline{x}_\alpha])$ ($= 1 - \alpha$ if $\pi$ is continuous). Conversely, a nested set of intervals $A_i$ with degrees of certainty $\lambda_i$ that $A_i$ contains $X$ is equivalent to the possibility distribution

$$\pi(x) = \min_{i=1\dots n} \{1 - \lambda_i, x \in A_i\},$$

provided that $\lambda_i$ is interpreted as a lower bound on $N(A_i)$, and $\pi$ is chosen as the least specific possibility distribution satisfying these inequalities [16].
We can interpret any pair of dual functions necessity/possibility $[N, \Pi]$ as upper and lower probabilities induced from specific probability families.

- Let $\pi$ be a possibility distribution inducing a pair of functions $[N, \Pi]$. We define the probability family $\mathcal{P}(\pi) = \{P, \forall A$ measurable, $N(A) \leq P(A)\} = \{P, \forall A$ measurable, $P(A) \leq \Pi(A)\}$. In this case, $\sup_{P\in\mathcal{P}(\pi)} P(A) = \Pi(A)$ and $\inf_{P\in\mathcal{P}(\pi)} P(A) =$

$N(A)$ (see [7,16]) hold. In other words, the family $\mathcal{P}(\pi)$ is entirely determined by the probability intervals it generates.

- Suppose pairs (interval $A_i$, necessity weight $\lambda_i$) supplied by an expert are interpreted as stating that the probability $P(A_i)$ is at least equal to $\lambda_i$ where $A_i$ is a measurable set. We define the probability family as follows: $\mathcal{P}(\pi) = \{P, \forall A_i, \lambda_i \leq P(A_i)\}$. We thus know that $\overline{P} = \Pi$ and $\underline{P} = N$ (see [16], and in the infinite case [7]).

### 2.3 *Imprecise probability induced by random intervals*

The theory of imprecise probabilities introduced by Dempster [8] (and elaborated further by Shafer [30] and Smets [31] in a different context) allows imprecision and variability to be treated separately within a single framework. Indeed, it provides mathematical tools to process information which is at the same time of random and imprecise nature. Contrary to probability theory, which in the finite case assigns probability weights to atoms (elements of the referential), in this approach we may assign such weights to any subset, called focal set, with the understanding that portions of this weight may move freely from one element to another in a focal set. We typically find this kind of knowledge when some measurement device is tainted with limited perception capabilities and a random error (variability) due to the variability of a phenomenon. We may obtain a sample of random intervals of the form $([m_i - \delta, m_i + \delta])_{i=1...K}$ supposedly containing the true value, where $\delta$ is a perception threshold, $m_i$ is a measured value and $K$ is the number of interval observations. Each interval is attached a probability $\nu_i$ of observing the measured value $m_i$. That is, we obtain a mass distribution $(\nu_i)_{i=1...K}$ on intervals, thus defining a random interval. The probability mass $\nu_i$ can be freely re-allocated to points within interval $[m_i - \delta, m_i + \delta]$. However, there is not enough information to do it.
Like possibility theory, this theory provides two indicators, called plausibility $Pl$ and belief $Bel$ by Shafer [30]. They qualify the validity of a proposition stating that the value of variable $X$ should lie within a set $A$ (a certain interval for example). Plausibility $Pl$ and belief $Bel$ measures are defined from the mass distribution assigning positive weights to a finite set $\mathcal{F}$ of measurable subsets of $\Omega$:

$$\nu : \mathcal{F} \rightarrow [0, 1] \quad \text{such that} \quad \sum_{E \in \mathcal{F}} \nu(E) = 1, \tag{5}$$

as follows:

$$Bel(A) = \sum_{E, E \subseteq A} \nu(E) \tag{6}$$

and

$$Pl(A) = \sum_{E, E \cap A \neq \emptyset} \nu(E) = 1 - Bel(\bar{A}), \tag{7}$$

where $E \in \mathcal{F}$ is called a focal element.$Bel(A)$ gathers the imprecise evidence that asserts $A$ ; $Pl(A)$ gathers the imprecise evidence that does not contradict $A$.

A mass distribution $\nu$ may encode the probability family $\mathcal{P}(\nu) = \{P, \forall A$ measurable, $Bel(A) \leq P(A)\} = \{P, \forall A$ measurable, $P(A) \leq Pl(A)\}$ [8]. In this case we have: $\overline{P} = Pl$ and $\underline{P} = Bel$, so that:

$$\forall P \in \mathcal{P}(\nu), Bel \leq P \leq Pl. \tag{8}$$

This view of belief functions is at odds with the theory of evidence of Shafer and the transferable belief model of Smets, who never refer to an imprecisely located probability distribution. Originally, Dempster [8] considered imprecise probabilities induced from a probability space via a set-valued mapping. In this scope, $Bel(A)$ is the minimal amount of probability that must be assigned to $A$ by sharing the probability weights defined by the mass function among single values in the focal sets. $Pl(A)$ is the maximal amount of probability that can be likewise assigned to $A$. We may define an upper $\overline{F}$ and a lower $\underline{F}$ cumulative distribution function (a particular p-box) such that $\forall x \in \mathbb{R} \quad \underline{F}(x) \leq F(x) \leq \overline{F}(x)$ with :

$$\overline{F}(x) = Pl(X \in (-\infty, x]) \tag{9}$$

and

$$\underline{F}(x) = Bel(X \in (-\infty, x]). \tag{10}$$

But this p-box contains many more probability functions than $\mathcal{P}(\nu)$.

The setting of belief and plausibility functions encompasses possibility and probability theories, at least in the finite case:

· When focal elements are nested, a belief measure $Bel$ is a necessity measure, that is $Bel = N$. A plausibility measure $Pl$ is a possibility measure, that is $Pl = \Pi$.
· When focal elements are some disjoint intervals, plausibility $Pl$ and belief $Bel$ measures are both probability measures, that is, we have $Bel = P = Pl$, for unions of such intervals.

Thus, all discrete probability distributions and possibility distributions may be interpreted by mass functions. However, continuous belief functions have not received much attention so far (except in the scope of random sets).

The above notions offer a common framework to treat the information of imprecise and random nature. However an obvious question is how to compare the expressivity of p-boxes, possibility distributions and belief functions. As we shall see, a p-box generally contains less information than a belief function and a possibility measure from which this p-box is derived. Possibility measures also offer the capability of approximating confidence intervals. A representation using belief functions is potentially more complex than the two other representation modes because a mass function must be specified for all subsets. However, using only a few focal subsets may be enough in practice. In the next section we focus on the respective expressive power of p-boxes and possibility measures.

## 3 Comparative expressivity of probability boxes and possibility distributions

Consider a unimodal continuous possibility distribution $\pi$ with core $\{a\}$ (i.e. $\Pi(\{a\}) = \pi(a) = 1$ and $\forall\, x \neq a, \pi(x) \neq 1$). We assume a unimodal $\pi$ for simplicity. Results in this Section readily adapt to the case when the core of $\pi$ is of the form $[a, b]$. The set of probability measures induced by $\pi$, that is, $\mathcal{P}(\pi)$, can be more conveniently described by a condition on the cumulative distribution functions of these probabilities (as first pointed out by Dubois and Prade [13]):

**Theorem 1** *Let $\pi$ be a unimodal continuous possibility distribution with core $\{a\}$. Then* $\mathcal{P}(\pi) = \{P, \forall x, y, x \leq a \leq y, F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y))\}$

***Proof.*** see appendix A

Note that we can choose $x$ and $y$ such that $\pi(x) = \pi(y)$ in the expression of $\mathcal{P}(\pi)$, i.e. suppose that $[x, y]$ is a cut of $\pi$. If $I_\alpha$ is the $\alpha$-cut of $\pi$, it holds that: $\mathcal{P}(\pi) = \{P, P(I_\alpha) \geq N(I_\alpha), \forall \alpha \in (0, 1]\}$. Thus by putting $\forall x \leq a$, $f(x) = \sup\{y, \pi(x) \geq \pi(y)\}$, we can prove that [11]:

$$\mathcal{P}(\pi) = \{P, \forall x \leq a, F(x) + 1 - F(f(x)) \leq \pi(x)\}.$$

Define a particular probability box $[\underline{F}, \overline{F}]$ such that:

$$\overline{F}(x) = \Pi(X \in (-\infty, x]) \tag{11}$$

and

$$\underline{F}(x) = N(X \in (-\infty, x]). \tag{12}$$

It is clear that $\overline{F}(x) = \pi(x)\ \forall x$ such that $\overline{F}(x) < 1$ and $\underline{F}(x) = 1 - \pi(x)\ \forall x$ such that $\underline{F}(x) > 0$. Define

$$\pi^+(x) = \begin{cases} \pi(x) & for\ x \leq a \\ 1 & for\ x \geq a \end{cases} \quad \text{and} \quad \pi^-(x) = \begin{cases} \pi(x) & for\ x \geq a \\ 1 & for\ x \leq a \end{cases}$$

the functions $\pi^+(x)$ and $1 - \pi^-(x)$ can be equated to the cumulative distribution functions $\overline{F}$ and $\underline{F}$. The probability box $(\overline{F}, \underline{F}) = (\pi^+, 1 - \pi^-)$ has an important specific feature: there exists a real value $a$ such that $\overline{F}(a) = 1$ and $\underline{F}(a) = 0$. It means that the p-box contains the deterministic value $a$, so that the two cumulative distributions are acting in disjoint areas of the real line separated by this value. We can retrieve a possibility distribution from such two cumulative distribution functions as $\pi = \min(\overline{F}, 1 - \underline{F})$ and thus retrieve the possibility distribution that generated the p-box. However it is clear that this process applied to any p-box does not yield a normalized possibility distribution, when the cumulative distributions are too close. A probability box can be a precise tool for approximating a probability distribution in the latter case, but it then forbids the case where the modelled unknown may be deterministic.

Moreover the two sets of probability functions $\mathcal{P}(\pi)$ and $\mathcal{P}(\underline{F} < \overline{F})$ differ. The following results indicate that the former is more precise than the latter (Dubois and Prade [13]):

**Theorem 2** *The probability family encoded by the unimodal continuous possibility distribution $\pi$ is included in the probability family encoded by the probability box $[\underline{F}, \overline{F}]$ induced from $\pi$:*

$$\mathcal{P}(\pi) \subset \mathcal{P}(\underline{F} < \overline{F}) \text{ for } \underline{F} = 1 - \pi^- \text{ and } \overline{F} = \pi^+.$$

***Proof.*** Let be $P \in \mathcal{P}(\pi)$. As $\lim_{y \to +\infty} F(x) + 1 - F(y) = F(x)$ and $\lim_{y \to +\infty} \max(\pi(x), \pi(y)) = \pi^+(x)$, we obtain according to Theorem 1: $F(x) \leq \pi^+(x)$. In the same way, $\lim_{x \to -\infty} F(x) + 1 - F(y) = 1 - F(y)$ and $\lim_{x \to -\infty} \max(\pi(x), \pi(y)) = \pi^-(y)$, thus $F(y) \geq 1 - \pi^-(y)$. Hence we have $P \in \mathcal{P}(\underline{F} < \overline{F})$.
On the other hand, the other inclusion is false, indeed take for example the triangular possibility distribution $\pi$ with support $[0, 2]$ and core $\{1\}$. Define $\mathcal{P}(\underline{F} < \overline{F}) = \{P, \forall x, 1 - \pi^-(x) \leq F(x) \leq \pi^+(x)\}$ and $P$ a probability measure such that $P(\{0.5\}) = 0.5$ and $P(\{1.5\}) = 0.5$. We do have $P \in \mathcal{P}(\underline{F} < \overline{F})$, however $P \notin \mathcal{P}(\pi)$ because, for $A = (-\infty, 0.5] \cup [1.5, +\infty)$, it holds $P(A) = 1 > \Pi(A) = 0.5$.□

We can systematize this counterexample and find probability families included in the probability box $[\underline{F}, \overline{F}]$ induced by $\pi$, which are not present in $\mathcal{P}(\pi)$. The following result improves a previous one due to Dubois and Prade [13]:

**Theorem 3** *Let $P$ be a probability measure in $\mathcal{P}(\underline{F} < \overline{F})$ such that:*

- *There exists $\gamma \geq a$ satisfying $P((-\infty, \gamma]) = \underline{F}(\gamma)$ (see Figure 1.a) (or $P((-\infty, \gamma]) = \overline{F}(\gamma)$, (see Figure 1.b)).*
- *There exists $\theta \in \{x \leq a / \overline{F}(x) \leq 1 - \underline{F}(\gamma)\}$ such that $P((-\infty, \theta]) \neq 0$ (see Figure 1.a) (or $\theta \in \{x \geq a / 1 - \underline{F}(x) \leq \overline{F}(\gamma)\}$ such that $P((-\infty, \theta]) \neq 1$), (see Figure 1.b)).*

*Then $P \notin \mathcal{P}(\pi)$.*

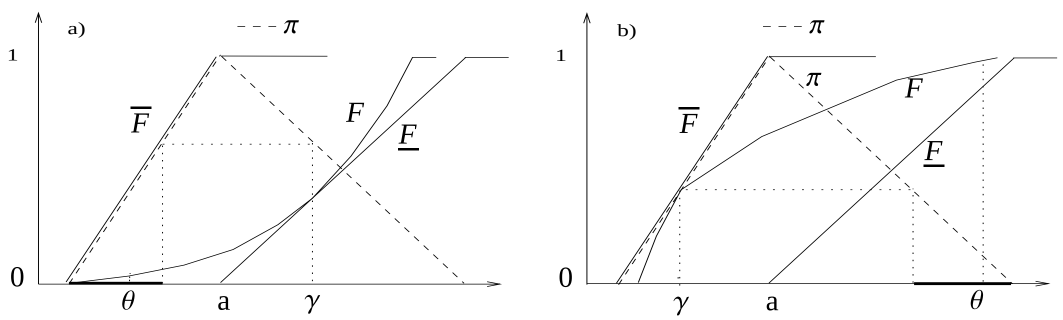


Fig. 1. Probabilities in $\mathcal{P}(\underline{F} < \overline{F})$ but not in $\mathcal{P}(\pi)$.

***Proof.*** Let $P \in \mathcal{P}(\underline{F} < \overline{F})$ with cumulative function $F$.
Consider the case where $\gamma \geq a$ exists satisfying $P((-\infty, \gamma]) = \underline{F}(\gamma)$ (see Figure 1.a). Using Theorem 1 and the features of $F$ on $(-\infty, \theta]$, we have $P((-\infty, \theta] \cup [\gamma, +\infty)) = F(\theta) + 1 - \underline{F}(\gamma) > 1 - \underline{F}(\gamma) = \pi(\gamma) = max(\pi(\theta), \pi(\gamma))$ Thus, $P \notin \mathcal{P}(\pi)$.
Similarly, if there exists $\gamma \leq a$ satisfying $P((-\infty, \gamma]) = \overline{F}(\gamma)$, (see Figure 1.b); we have $\overline{F}(\gamma) + 1 - F(\theta) > \overline{F}(\gamma) = \pi(\gamma) = max(\pi(\gamma), \pi(\theta))$. □

The probability box induced by $\pi$ can thus contain multimodal distributions (if $F(\theta) = F(\gamma)$ for instance), and some unimodal distributions with mode different from $a$ which are ruled out by the probability family encoded by the possibility distribution $\pi$.

Theorem 3 identifies a set of probability measures which are not in $\mathcal{P}(\pi)$. Consider $\pi$ and $F$ such that $F \leq \overline{F} = \pi^+$. If $F$ is known on $(-\infty, a]$, we can define a lower bound $F_*$ of $F$ on $[a, +\infty)$ such that the corresponding probability measure $P$ belongs to $\mathcal{P}(\pi)$ if and only if $F \geq F_*$. Consider the function $g : y \mapsto \min\{x \leq a | \pi(x) = \pi(y)\}$. From theorem 1, we deduce that $F \in \mathcal{P}(\pi)$ if and only if

$$F(y) \geq 1 - \pi(y) + F(g(y)), \ \forall y \geq a.$$

The function $1 - \pi(y) + F(g(y))$ is not necessarily increasing (see Figure 2), we can thus define

$$F_*(y) = \max(F(a), 1 - \pi(y) + F(g(y))) \text{ for } y \geq a \text{ (see Figure 2).}$$

Conversely if $F$ is known on $[a, +\infty)$ an upper bound $F^*$ can be found on $(-\infty, a]$ such that $P \in \mathcal{P}(\pi)$ if and only if

$$F(x) \leq F^*(x) = \min(F(a), \pi(a) - 1 + F(f(x))).$$

It is clear that sets $\{P | F(x) \leq \overline{F}(x) \ \forall x \leq a \text{ and } F(x) = 1 \ \forall x \geq a\}$ and $\{P | F(x) \geq \underline{F}(x) \ \forall x \geq a \text{ and } F(x) = 0 \ \forall x \leq a\}$ are included in $\mathcal{P}(\pi)$. They correspond to probability densities with support $[\min, a]$ or limited by $[a, \max]$ and their mode is not the mode of $\pi$.

Conversely, suppose $\underline{F} < \overline{F}$ is the available information, and there exists a real value $a$ such that $\overline{F}(a) = 1$ and $\underline{F}(a) = 0$. The above results show that the possibility distribution $\min(\overline{F}, 1 - \underline{F})$ cannot encompass all probability distributions restricted by the p-box. An obvious example is as follows: consider a probability distribution $P$, and the probability box $(\underline{F}, \overline{F})$ such that $\overline{F}(x) = F(x)$ for $x < a$ and 1 otherwise; $\underline{F}(x) = F(x)$ for $x > a$ and 0 otherwise. The corresponding possibility distribution is $\pi(x) = F(x)$ if $x < a$, $\pi(x) = 1 - F(x)$ if $x > a$ and $\pi(x) = 1$ if $x = a$. It can be checked that $P \notin \mathcal{P}(\pi)$ while $\underline{F} < F < \overline{F}$.

Can we find a non-trivial possibility distribution $\pi^\star$ such that $\mathcal{P}(\pi^\star)$ contains the probability box $[\underline{F}, \overline{F}]$ ? Note that $\overline{F}(x) \geq F(x) \geq \underline{F}(x)$ implies $\forall x \leq a \leq y, F(x) +$

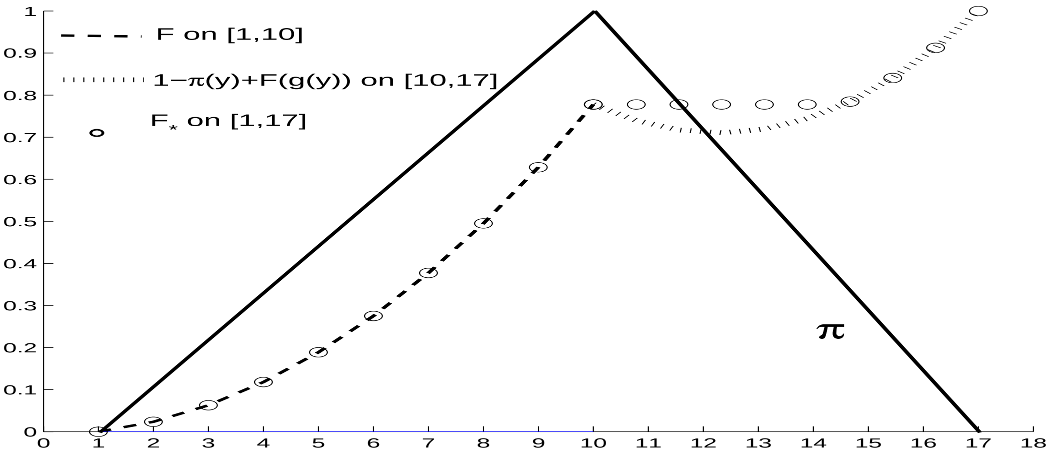


Fig. 2. Example of probability families included in $\mathcal{P}(\pi)$ induced by the triangular possibility distribution of core $\{10\}$ and support $[1, 17]$

$1 - F(y) \leq \overline{F}(x) + 1 - \underline{F}(y)$. Thus, with intervals $[x, g(x)]$, where $g(x)$ is such that $\overline{F}(x) = 1 - \underline{F}(g(x))$, $\forall x \leq a, F(x) + 1 - F(g(x)) \leq 2\overline{F}(x)$. Thus, by letting $\pi^{\star+}(x) = \min(1, 2\overline{F}(x))$ and $\pi^{\star-}(y) = \min(1, 2(1 - \underline{F}(y)))$ we do build a possibility distribution (often very imprecise) $\pi^\star$ such that $\mathcal{P}(\underline{F} < \overline{F}) \subset \mathcal{P}(\pi^\star)$.

We may conclude: To represent knowledge using a possibility distribution is more precise than using the upper & lower cumulative distribution functions $\underline{F}, \overline{F}$ it induces.

As a consequence, if we seek to estimate the probability $P(X \in [x, y])$ using the probability box $[\underline{F}, \overline{F}]$ induced by $\pi$, for some $x \neq y$, we may obtain a bracketing of this probability larger than that the one obtained from the possibility distribution. From the probability box, we can estimate a best envelope of probability $P(X \in [x, y])$ by:

$$\max(0, \underline{F}(y) - \overline{F}(x)) \leq P(X \in [x, y]) \leq \overline{F}(y) - \underline{F}(x).$$

With a similar reasoning as in the proof of theorem 1, we can show $\exists\, (x, y)$, $x \neq y$ such that:

$$\max(0, \underline{F}(y) - \overline{F}(x)) < N([x, y]).$$

Indeed, for $x < a < y$ such that $\pi^+(x) > 0$ and $\pi^-(y) > 0$, we have:

$$\max(0, \underline{F}(y) - \overline{F}(x)) = \max(0, 1 - \pi^-(y) - \pi^+(x)) = 0 \text{ if } \pi^+(x) + \pi^-(y) > 1$$

and

$$N([x, y]) = 1 - \max(\pi^+(x), \pi^-(y)) > 0.$$

It is clear that $\pi^+(x) + \pi^-(y) > \max(\pi^+(x), \pi^-(y))$ which implies $N([x, y]) > \underline{F}(y) - \overline{F}(x)$.

Note that the cumulative distributions describing any p-box can be generated by a belief function, contrary to the case of possibility distributions (see Ferson et al. [21]). So an open interesting problem is to characterize the set of belief functions whose upper and lower cumulative distributions coincide with a (discrete) p-box, and among them the least informative one (for instance maximizing the expected size of the focal elements). This is left for further research.

## 4 Approximating probability families by possibility distributions

Let $p$ be a unimodal probability density function. Denote by $M$ the mode of $p$. Let $P$ be the probability measure associated to $p$. So far we considered possibility distributions $\pi$ which verify the following condition (dubbed *Dominance condition*):

$$P(A) \leq \Pi(A), \text{ for all measurable events } A.$$

We say that the possibility measure dominates the probability measure and it means $P \in \mathcal{P}(\pi)$.

More generally, $\pi$ dominates a probability family $\mathcal{P}$ if and only if $\mathcal{P} \subseteq \mathcal{P}(\pi)$. It holds that $\mathcal{P}(\max(\pi_1, \pi_2))$ is the convex closure of $\mathcal{P}(\pi_1) \cup \mathcal{P}(\pi_2)$, since the possibility distribution $\max(\pi_1, \pi_2)$ generates the possibility measure $\max(\Pi_1, \Pi_2)$. On the contrary, $\mathcal{P}(\min(\pi_1, \pi_2)) \neq \mathcal{P}(\pi_1) \cap \mathcal{P}(\pi_2)$ in general, because $\min(\Pi_1, \Pi_2)$ is not a possibility measure. So, if a probability family $\mathcal{P}$ is dominated by two possibility distributions $\pi_1, \pi_2$, one cannot deduce that $\mathcal{P}$ is dominated by $\min(\pi_1, \pi_2)$, even if $\min(\pi_1, \pi_2)$ is normalized.

An approximate (covering) possibilistic representation of a given family $\mathcal{P}$ is any $\pi$ such that $\mathcal{P} \subseteq \mathcal{P}(\pi)$. Clearly it means that $\pi$ dominates all probability functions in $\mathcal{P}$. Ideally $\pi$ should be such that $\mathcal{P}(\pi)$ is as small as possible. However such optimal covering approximations of probability families generally do not exist (see Dubois and Prade [14]). Nevertheless, in the remainder of the paper we shall lay bare various informative approximate covering possibilistic representations of probability families induced by incomplete probabilistic data.

However as previously seen, a possibility measure also encodes a set of nested confidence intervals provided by an expert. A possibility measure $\pi$ such that $P \in \mathcal{P}(\pi)$ can be constructed as follows ([11], [15]): Let $J_\lambda = [x(\lambda), y(\lambda)]$, for $\lambda \in [0, 1]$ be a continuous nested interval family such that $J_\lambda \subseteq J_\beta$ if $\lambda \geq \beta$, $J_0 = \{x_*\} \subset supp(p)$ and $J_1 = supp(p)$ where $supp(p)$ is the support of a unimodal probability density $p$. Then, the possibility distribution $\pi$ given by :

$$\pi(x(\lambda)) = \pi(y(\lambda)) = 1 - P(J_\lambda) \quad \forall \lambda, \tag{13}$$

dominates $p$. That is $p \in \mathcal{P}(\pi)$ (or $P \leq \Pi$). If we choose $J_\lambda$ such that:

$$J_\lambda = \{x, p(x) \geq \lambda\}, \quad \forall \lambda \in [0, sup(p)]. \tag{14}$$

$J_\lambda$ is of the form $[x, f(x)]$ where $f(x) = max\{y, p(y) \geq p(x)\}$. Then, $J_\lambda$ is the narrowest prediction interval of probability $\alpha = P(J_\lambda)$, $x_*$ is the mode $M$ of $p$ and $J_\lambda$ is also the most probable interval of length $|y(\lambda) - x(\lambda)|$ [12].

Hence, if we choose $J_\lambda$ as in (14) and $\pi$ as in (13), we obtain a possibility distribution $\pi_p^\star$ such that $\Pi \geq P$ (*dominance condition*) and the $\alpha$-cut of $\pi_p^\star$ is indeed the narrowest prediction interval of $p$ of confidence level $1 - \alpha$ (*prediction interval condition*). Such $\pi_p^\star$ is called the *optimal transform* of $p$,

$$\pi_p^\star(x) = 1 - P(\{y, p(y) \geq p(x)\}) = F(x) + 1 - F(f(x)),$$

and $\pi_p^\star(M) = 1$. This transformation is optimal in the sense that it provides the most specific possibility distribution among those that dominate $p$, and preserve the order induced by $p$ on the support interval.

It is clear that function $\pi_p^\star$ is a kind of cumulative distribution. More precisely, given any total ordering of values $\succ$ on the real line, and any value $x$, let $A_\succ(x) = \{y, x \geq y\}$, and assume $A_\succ(x)$ is measurable for all $x$. The function $F_\succ(x) = P(A_\succ(x)))$ is the cumulative function according to the order relation $\succ$. If $\succ=>$ the usual ordering on the real line, then $F = F_\succ$. Now, choosing the ordering induced by the density $p$, that is, $x \succ_p y$ if and only if $p(x) > p(y)$, then $F_{\succ_p} = \pi_p^\star$.

Computing $\pi_p^\star$ is not so obvious in general, but the case of symmetric densities has been considered in [11]; it is shown that $\pi_p^\star$ is convex on each side of the mode of $p$. This result no longer holds in the general case, but we can approximate any unimodal density $p$ by means of a piecewise continuous function. Then we can easily show he following result:

**Theorem 4** *Let $p$ be a unimodal continuous probability distribution function of mode $M$ and of bounded support $supp(p)$. If $p$ is (piecewise) linear, then its optimal transform is piecewise convex.*

***Proof.*** see appendix B

Using the idea of narrowest prediction intervals described above, it is also interesting to characterize approximate covering possibilistic representations of probability families $\mathcal{P}$ that account for such prediction intervals. A possibility distribution $\pi$ is said to strongly dominate a probability measure $P$ with density $p$ if $\{x; p(x) \geq \lambda\} = J_\lambda \subseteq \{x, \pi(x) \geq \alpha\}$ for $\alpha = 1 - P(J_\lambda)$ (dubbed *prediction interval condition*). Given a probability family $\mathcal{P}$ one may try to find the most specific possibility distribution $\pi_{\mathcal{P}}$ thatstrongly dominates all $P \in \mathcal{P}$ such a possibility distribution is $\pi_{\mathcal{P}} = sup_{p \in \mathcal{P}} \pi_p^\star$. $\pi_{\mathcal{P}}$ has the peculiarity that any of its $\alpha$-cuts contain the

$(1-\alpha)$-prediction interval of any $p \in \mathcal{P}$. Note that this approach is enabled by the property $\mathcal{P}(\pi_1) \cup \mathcal{P}(\pi_2)) \subseteq \mathcal{P}(\max(\pi_1, \pi_2))$.

## 5 Simple Models of Incomplete Probabilistic Knowledge using p-boxes and belief functions

The preceding results can be applied to the definition of faithful representations of poor probabilistic knowledge by means of probability boxes, belief functions and especially possibility distributions. The extreme case is when an expert provides an interval containing the unknown value. Generally there is a little more information than a simple interval: an expert may have an idea on typical values in the interval: the median , the mean, the mode. Additional information on a distribution may be the knowledge of appropriate fractiles and confidence intervals. These pieces of information define constraints restricting a probability family. The problem is whether such family can be simply described or approximated by means of the simple tools described in the previous sections. These representation techniques suggest that simple non-parametric representations of available uncertain knowledge, where incompleteness and variability receive specific treatments, are feasible in the scope of further uncertainty propagation steps in risk analysis problems. This section recalls representation methods proposed by Ferson using p-boxes, when the mean-value of a density is known, and for the modelling of a small set of precise observations by means of imprecise probabilities. Moreover, belief functions can be directly used for exploiting the knowledge of fractiles.

### *5.1 Representations by Probability Boxes*

As discussed earlier, probability boxes generalise the idea of interval from a pair of points to a pair of cumulative distribution functions. They are a very natural way of extending the notion of interval. They are especially informative when the two cumulative distributions are close to each other. They come up as a natural choice for imprecise parametric models with imprecise parameters. For instance, a Gaussian model where the mean-value and/or the variance is known to lie in a prescribed interval may naturally yield a narrow p-box (even if the latter contains non-Gaussian distributions). However, we do not deal with parametric models here. The p-box model has been especially investigated by Ferson et al [21]. We recall his proposals for representing distributions with fixed mean value as well as for using the Kolmogorov-Smirnov confidence limits in order to derive a p-box from small data samples.

*5.1.1 Probability distributions with known mean and support*

Suppose an expert supplies the mean $\mu$ and the support $I = [b, c]$. Let $\mathcal{P}_I^{mean}$ denote the set of probabilities with support $I$ and prescribed mean, equal to $\mu$. Ferson [21] proposes to represent this knowledge by a probability box $[\underline{F}, \overline{F}]$. To obtain it, he separately solves two problems for each value $x$ as follows: $\overline{F}(x) = \sup_{F:E(X)=\mu} F(x)$ and $\underline{F}(x) = \inf_{F:E(X)=\mu} F(x)$ (the unknown is $F$). Using the characteristic property of the mean:

$$\int_{\inf(I)}^{\mu} F(y)dy = \int_{\mu}^{\sup(I)} (1 - F(y))dy,$$

one obtains the following results

$$\underline{F}(x) = \begin{cases} \frac{x-\mu}{x-b} & \forall x \in [\mu, c] \\ 0 & \forall x \in [b, \mu] \end{cases} \quad \text{and} \quad \overline{F}(x) = \begin{cases} 1 & \forall x \in [\mu, c] \\ \frac{c-\mu}{c-x} & \forall x \in [b, \mu] \end{cases}$$

The probability box $[\underline{F}, \overline{F}]$ (see Figure 3 for an example) defines a probability

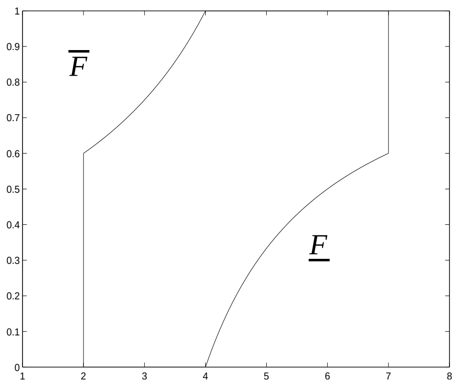


Fig. 3. Probability box built from $x \in [2, 7]$ and $E(X) = 4$.

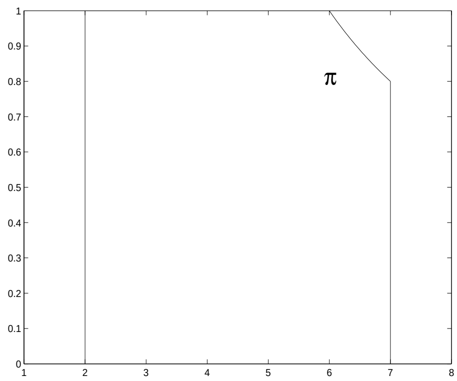


Fig. 4. Possibility distribution $\pi$ containing the probability box $[\underline{F}, \overline{F}]$.

family $\mathcal{P}(\underline{F}, \overline{F})$ which contains $\mathcal{P}_I^{mean}$. It could be tempting to use the probability family induced by possibility distribution $\pi$ such that $\pi(x) = (c - \mu)/(c - x)$ for $x \in [b, \mu]$ and $\pi(x) = 1 - (x - \mu)/(x - b)$ for $x \in [\mu, c]$. But, as expected from the previous sections, the inclusion $\mathcal{P}_I^{mean} \subset \mathcal{P}(\pi)$ does not hold. The probability $P$, defined by $P(X = 2) = 3/5$ and $P(X = 7) = 2/5$, is enough to show we do not have the inclusion. Indeed, we do have $E(X) = 4$ but $P(X = 2 \text{ or } X = 7) = 1$ and $\Pi(X = 2 \text{ or } X = 7) = 0.6$, which is contradictory with $P \leq \Pi$. As pointed out earlier, the probability family $\mathcal{P}(\pi)$ such that $\pi^+(x) = \min(1, 2\overline{F}(x))$ and $\pi^-(y) = \min(1, 2(1 - \underline{F}(y)))$ (see Figure 4), contains $\mathcal{P}_I^{mean}$ and $\mathcal{P}(\underline{F}, \overline{F})$. However, it is clear that this p-box is poorly informative, and that the covering possibility is even more so. In fact, the mean value does not seem to bring much information on the distribution, and the problem of finding a better, tighter representation of this kind of information remains open. Moreover, while the average value is very easy and often natural to compute from statistical data, it is not clear that this value is cognitively plausible, that is, one may doubt that a single representative value of an ill-known quantity provided by an expert refers to the mean value. For instance, while some quantities like average income can be easily figured out, the average human size sounds like a very artificial notion, and would not be directly perceived by individuals.

### 5.1.2 *Representing small data samples by a p-box*

When the available knowledge is just a small data sample $(x_1, ..., x_n)$ coming from the unknown cumulative distribution function $F$, Ferson et al. [21] define a probability box $[\underline{F}, \overline{F}]$ by using Kolmogorov-Smirnov confidence limits (noted K.S.) [19,28]. These confidence limits are distribution-free bounds about the sample empirical cumulative distribution function $F_n$ where $n$ is the size of the sample. We can define $F_n$ as follows:

$$F_n(x) = \begin{cases} 0 \text{ for } & x < x_{(1)} \\ \quad \vdots & \\ \frac{i}{n} \text{ for } & x_{(i)} \leq x < x_{(i+1)}, \\ \quad \vdots & \\ 1 \text{ for } & x \geq x_{(n)} \end{cases}$$

where $x_{(i)}$ are the order statistics of the sample.
$\underline{F}_n$ and $\overline{F}_n$ converge to the empirical cumulative distribution $F_n$ when the sample becomes very large, although convergence is rather slow. Kolmogorov-Smirnov limits require that the samples be independent and identically distributed. This is a very standard assumption, but it is sometimes hard to justify (if the values come from heterogeneous sources, for instance). To obtain these bounds, we use the maximum deviation $D_{KS}$ between $F_n$ and $F$ defined as follows:

$$D_{KS} = \max_{i=1,\dots,n} \left( |F(x_{(i)}) - \frac{i}{n}|, |F(x_{(i)}) - \frac{i-1}{n}| \right).$$

$D_{KS}$ is a random variable whose exact distribution is not known but Kolmogorov found that $\sqrt{n}D_{KS}$ has a limiting distribution given by:

$$\forall\, t \geq 0 \lim_{n\to\infty} P(\sqrt{n}D_{KS} \leq t) = 1 - 2\sum_{k=1}^{+\infty}(-1)^{k+1}e^{-2k^2t^2}.$$

This limit has been tabulated and allows for each confidence level $\alpha$ to find a value $D_n(\alpha)$ such that $P(D_{KS} \leq D_n(\alpha)) = 1 - \alpha$. To conclude, the K.S. bounds are computed with the expression $min(1, max(0, F_n(x) \pm D_n(\alpha))$ for a fixed confidence level $\alpha$. For instance, at 95% confidence level, for a sample size of 10, the value of $D_n(\alpha)$

is 0.40925 (see Figure 5). These limits are often used to express the reliability of results of a simulation or to test if the sampling from the simulation follows some probability laws. However, it is not common to use K.S. limits on input parameters to define a probability family respecting the available data. We must be aware that K.S. limits are not sure bounds but statistical ones. It means for instance that 95% of the time the true distribution will lie inside the bounds.

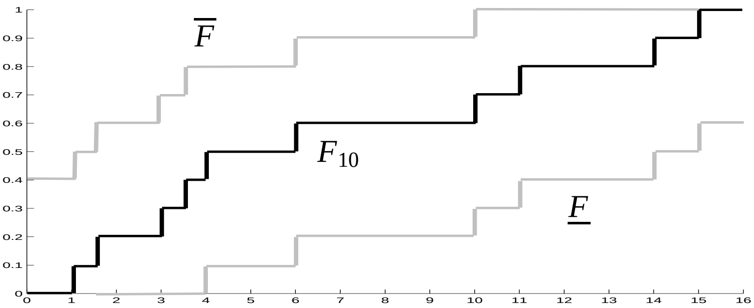


Fig. 5. Kolmogorov-Smirnov confidence limits (gray) about an empirical cumulative distributino function (black) assuming a sample size of 10.

The obtained p-box cannot be obtained from a possibility distribution, as it generally does not include the step-function corresponding to a deterministic value.

### *5.2 Discrete belief functions representations*

Very naturally the representation of a family of probabilities by means of a belief function (a discrete random set) is appropriate if the probability of prescribed events is known. This is typically the case when only the median $m$ of a distribution is known. The meaning of the median is: $P(X \leq m) = 0.5$. Let $\mathcal{P}_I^{med}$ be the set of probability functions with support $I = [b, c]$ and with median $m$. This knowledge can be exactly represented by a mass function $\nu_m$ such that $\nu_m([b, m]) = \nu_m(]m, c]) = 0.5$. The belief function $Bel_m$, deduced from $\nu_m$, encodes all probabilities with median $m$, i.e., $\mathcal{P}_I^{med} = \{P, \forall C, Bel_m(C) \leq P(C)\}$.

This representation naturally extends to the case when some fractiles and the support $I$ of the unknown probability distribution function are known. Suppose an expert supplies fractiles, say $x_1$, $x_2$ and $x_3$ at 5%, 50% and 95%. Denote $\mathcal{P}_I^{x_1,x_2,x_3}$ the set of probability distribution functions of support $I = [b, c]$ and of fractiles $x_1, x_2, x_3$. We can represent this knowledge in an exact way using a belief function by the following obvious mass function $\nu_{fract}$: $\nu_{fract}([b, x_1]) = 0.05$, $\nu_{fract}(]x_1, x_2]) = 0.45$, $\nu_{fract}(]x_2, x_3]) = 0.45$ and $\nu_{fract}(]x_3, c]) = 0.05$. The belief function $Bel_{frac}$, deduced from $\nu_{frac}$, is dominated by all probabilities with fractiles $x_1$, $x_2$ and $x_3$, i.e., $\mathcal{P}_I^{x_1,x_2,x_3} = \{P, \forall C, Bel_{frac}(C) \leq P(C)\}$.

Note that the mass function induced by fractiles bears on a partition of the support. On the contrary if an expert, provides a confidence interval, $x \in A \subseteq \mathbb{R}$ with a certainty degree $\lambda$, the most cautious interpretation corresponds to an inequality $P(A) \geq \lambda$. The corresponding mass function $\nu$ assigns $\lambda$ to $A$ and $1 - \lambda$ to the real line itself. This is called a simple support function by Shafer. Note that the two focal elements are nested. The knowledge of a confidence interval with confidence $\lambda$ is less precise than a fractile: if $A = [x_1, x_3]$ with confidence at least $\lambda$, we cannot deduce the probability degrees associated to intervals $(-\infty, x_1]$ and $[x_3, +\infty)$, except if we assume the symmetry of the underlying density.

A confidence interval can be represented by the possibility distribution [16]:

$$\forall x \in \mathbb{R} \quad \pi(x) = \begin{cases} 1 & if \quad x \in A \\ 1 - \lambda & if \quad x \notin A \end{cases}$$

where $\pi$ encodes the probability family $\mathcal{P}(\pi) = \{P, \lambda \leq P(A)\}$. When $A$ is large enough, but, in practice, bounded, the level of confidence is 1. This representation extends when several nested confidence intervals $\{A_1 \subset A_2 \subset \ldots A_k\}$ are obtained for several confidence levels $\{\lambda_1 < \lambda_2 < \ldots \lambda_k\}$ as suggested previously. The corresponding mass assignment is $\nu(A_i) = \lambda_{i+1} - \lambda_i$, assuming $\lambda_0 = 0$ . It yields a discrete possibility distribution. The next section considers the case of continuous possibility distributions.

## 6 Representations by continuous possibility distributions

The use of continuous possibility distributions for representing probability families heavily relies on probabilistic inequalities. Such inequalities provide probability bounds for intervals forming a continuous nested family around a typical value. This nestedness property leads to interpreting the corresponding family as being induced by a possibility measure. While these bounds are usually used for proving convergence properties, we propose here to use them for representing knowledge. This is the case of the Chebyshev inequality for instance. The classical Chebyshev inequality [27] defines a bracketing approximation on the confidence intervals around the known mean $\mu$ of a random variable $X$, knowing its standard deviation $\sigma$. The Chebyshev inequality can be written as follows:

$$P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2} \qquad \text{for } k \geq 1.$$

By referring to the Section 2.2, Chebyshev inequality allows to define a possibility distribution $\pi$ by considering intervals $[\mu - k\sigma, \mu + k\sigma]$ as $\alpha$-cuts of $\pi$ and letting $\pi(\mu - k\sigma) = \pi(\mu + k\sigma) = 1/k^2$ (see Figure 6). This possibility distribution (see [11]) defines a probability family $\mathcal{P}(\pi)$ such that $\mathcal{P}^{\mu,\sigma} \subseteq \mathcal{P}(\pi)$ containing all distributions with known mean and standard deviation, whether the unknown probability distribution function is symmetric or not, unimodal or not. If it is moreover assumed that the unknown probability distribution is unimodal and symmetric, we can improve the possibility distribution $\pi$ by means of Camp-Meidel inequality [27] (see Figure 6).

$$P(|X - \mu| \leq k\sigma) \geq 1 - \frac{4}{9k^2} \qquad \text{for } k \geq \frac{2}{3}.$$

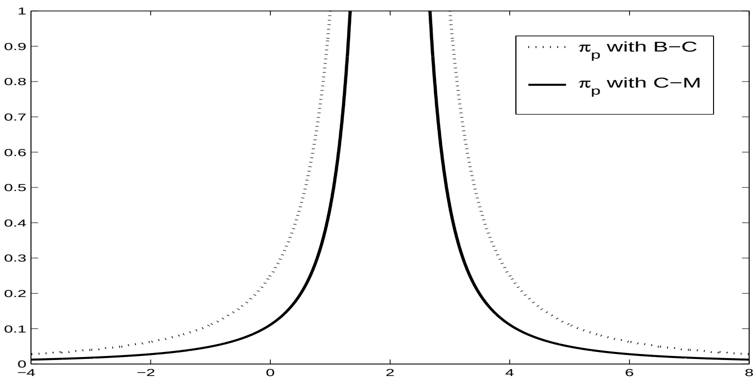


Fig. 6. Optimal possibility distribution $\pi_p$ knowing $\mu = 2$ and $\sigma = 1$ and using Bienaymé-Chebychev and Camp-Meidel inequality.

Very often, and as seen above, the nested intervals share the same midpoints, thus yielding symmetric possibility distributions. In the following we do not make this restriction. On the contrary, we shall also rely on the most narrow intervals of fixed confidence levels as introduced earlier in this paper. It leads to exploit information on the mode of distributions rather than the mean. Moreover we make the additional assumption that the distributions have a bounded support. Some assumptions can be made on the shape of the density (without going to the point of choosing a particular mathematical model like a Gaussian): symmetry, convexity or concavity can be assumed, for instance.

### *6.1 Distributions with known mode and support: simple dominance*

Suppose the mode $M$ and the support $I$ of the unknown probability distribution function $p$ is supplied by an expert. In this section unimodality of distributions is assumed. One might argue that the mode best corresponds to the notion of usual value, as being the most frequently observed value. Even if the mode is known to be difficult to extract from a sample of statistical data, one may consider that the most frequent value (or a most frequent small range of values) is the natural feature extracted from repeated observations by humans. So the problem of representing this kind of knowledge looks natural. We can take advantage of the fact that the cumulative distribution function $F$, associated to a unimodal (asymmetric) probability distribution function $p$ with mode $M$ and bounded support $I$, satisfies the following properties:

- $F$ is convex on $[\inf(I), M]$ since $p$ increases on $[\inf(I), M]$.
- $F$ is concave on $[M, \sup(I)]$ since $p$ decreases on $[M, \sup(I)]$.

Thus, the concavity of $F$ changes at $M$. Let $\mathcal{P}_I^M$ be the set of probabilities with support $I = [b, c]$ and with mode $M$. Ferson (in [22]) proposes to represent this knowledge by the probability box $[\underline{F_L}, \overline{F_L}]$ such that

- $\overline{F_L}(x) = (x - b)/(M - b)$ for $x \in [b, M]$ and 1 otherwise.
- $\underline{F_L}(x) = (x - M)/(c - M)$ for $x \in [M, c]$ and 0 otherwise.

Indeed it is obvious that any probability distribution with mode $M$ and support $I$ is such that $\overline{F_L} > F > \underline{F_L}$.

**Theorem 5** *The triangular possibility distribution $\pi_L = \min(\overline{F_L}, 1 - \underline{F_L})$, with support $[b, c]$ and core $\{M\}$ dominates all probabilities lying in $\mathcal{P}_I^M$ .*

***Proof.*** consider the nested family of intervals $[x, y]$ such that:

$$\frac{(x-b)}{M-b} = \frac{(c-y)}{c-M}.$$

They are cuts of the triangular possibility distribution $\pi_L$. Define the cumulative distribution $F_L$ as follows: $F_L(x) = (F(M)(x - b))/(M - b)$ for $x \leq M$, and $F_L(x) = F(M) + ((x - M)(1 - F(M)))/(c - M)$ for $x \geq M$. Due to the convexity of any $F$ before the mode, and its concavity after the mode, it is clear that $F_L(x) \geq F(x)$ for $x \leq M$, and $F_L(x) \leq F(x)$ for $x \geq M$. Using (13) in Section 4, it is clear that:

$$\begin{aligned}
\forall (P, x) \in \mathcal{P}_I^M \times [b, M],\ P([x, y]^c) &= F(x) + 1 - F(y) \\
&\leq F_L(x) + 1 - F_L(y) \\
&= \tfrac{F(M)(x-b)}{M-b} + 1 - (F(M) + \tfrac{(1-F(M))(y-M)}{c-M}) \\
&= \tfrac{x-b}{M-b} = \pi_L(x).
\end{aligned}$$

So it holds that $\Pi_L(A) \geq P(A)\ \forall A, \forall P \in \mathcal{P}_I^M$.□

Clearly this result corresponds to a Chebyshev-like probabilistic inequality built from the $\alpha$-cuts of $\pi_L$. The triangular possibility distribution $\pi_L$ of mode $M$ is thus a more precise representation than the p-box $[\overline{F_L}, \underline{F_L}]$. Namely, the probability family $\mathcal{P}(\pi_L)$ is a better approximation of $\mathcal{P}_I^M$ than the probability box $[\underline{F}, \overline{F}]$ proposed by Ferson. Note that the assumption of bounded support is crucial in getting this piecewise linear representation. Moreover it is noticeable that this distribution does not depend on the value $F(M)$.

Suppose now this value is known. Let $\mathcal{P}_I^{M,F(M)}$ be the set of probabilities with support $I = [b, c]$, with mode $M$ and value $F(M)$ at $M$. The latter information can be modelled by a belief function (see Section 2.3) but we may wish to preserve a possibilistic representation and alter its shape so as to account for this fractile, still ensuring the *Dominance condition*. Assume $F(M) \leq 0.5$. We choose nested intervals $J_x = [x, F^{-1}(1 - F(x))]$ around the median, and let $\overline{M} = F^{-1}(1 - F(M))$. We have seen $F$ is concave on $[M, c]$. So, $F > F_L$ on $[M, c]$. Hence

$$\overline{M} \leq \overline{M_L} = \frac{1}{1 - F(M)}\{(c - M)(1 - F(M)) - cF(M) + M\}.$$

Now we can use $\pi_p(x) = 1 - P(J_x)$ as a possibility distribution dominating $p$. Let

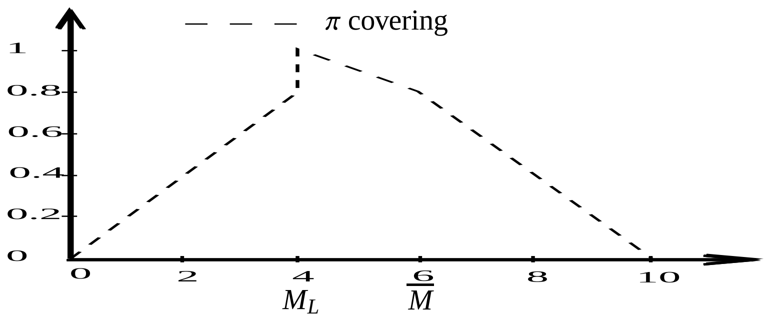


Fig. 7. $M$=4, $F(M)$ = 0.4, min = 0 and max = 10.

$J_x = [x, y]$. The following possibility distribution $\pi_{L,F(M)}$ can then be used in place of $\pi_L$:

- For $x \in [b, M]$, $\pi_p(x) = 2 \cdot F(x)$, and $\pi_p$ is convex. So $\pi_p(x) \leq 2 \cdot F_L(x)$. So we let $\pi_{L,F(M)}(x) = 2 \cdot F_L(x)$.
- For $y \in [M, \overline{M}_L]$, $\pi_p(y) = F(x) + 1 - F(y) \leq F(M) + 1 - F(y)$, and the latter is convex. We let $\pi_{L,F(M)}(y) = F(M) + 1 - F_L(y)$.
- For $y \in [\overline{M}_L, c]$, $\pi_p(y) = 2 \cdot (1 - F(y))$ is convex. $\pi_p(y) \leq 2 \cdot (1 - F_L(y))$. So we let $\pi_{L,F(M)}(y) = 2 \cdot (1 - F_L(y))$.

Thus, we have $\mathcal{P}_I^{M,F(M)} \subseteq \mathcal{P}(\pi_{L,F(M)})$. The obtained shape is more realistic than the triangular fuzzy interval especially when $M$ is the center of $I$, because the lack of balance of the probability mass is reflected on the possibility distribution (see Figure 7). In the case where $F(M) = 0.5$, we obtain $M = \overline{M}$ and we thus retrieve the triangular $\pi_L$ with a support $[b, c]$ and a core $\{M\}$. Note that it cannot be refined by exploiting the fact that both $\pi_L$ and the above derived $\pi_{L,F(M)}$ dominate $\mathcal{P}_I^{M,F(M)}$ to refine the result, considering $min(\pi_{L,F(M)}, \pi_L)$ as a tighter approximant. Indeed, as pointed out earlier $min(\pi_{L,F(M)}, \pi_L)$ will not dominate $\mathcal{P}_I^{M,F(M)}$ generally, as $\mathcal{P}(min(\pi_{L,F(M)}, \pi_L))$ differs from $\mathcal{P}(\pi_{L,F(M)}) \cap \mathcal{P}(\pi_L)$.

Suppose the expert provides the mode $M$ and the median $m$ of the probability distribution. Let $\mathcal{P}_I^{M,m}$ be the set of such unimodal probability functions bounded by $I = [b, c]$ and assume $m < M$. Then we can refine the possibilistic approximation $\pi_L$ by accounting for the additional information on the median, namely that $F(m) = 0.5$. It means that $F$ goes through the point of coordinates $(m, 0.5)$. So, instead of $F_L$, we can consider the piecewise cumulative distribution $F_L^m$ made of the segments $[(b, 0), (m, 0.5)]$, $[(m, 0.5), (M, F(M))]$, $[(M, F(M)), (c, 1)]$. Clearly, $F \leq F_L^m < F_L$ on $[b, M]$. Hence by choosing again the intervals $[x, y]$ such that $(x - b)/(M - b) = (c - y)/(c - M)$, we obtain a more specific piecewise linear possibility distribution $\pi_L^m \leq \pi_L$ which dominates all probability distributions with mode $M$ and median $m$. That is, $\mathcal{P}_I^{M,m} \subset \mathcal{P}(\pi_L^m)$. In particular :

$$\pi_L^m(m) = \pi_L^m(\overline{m}) = 0.5 + (1 - F(M))\frac{m - b}{M - b},$$

where $(m - b)/(M - b) = (c - \overline{m})/(c - M)$.

Note that this possibility distribution $\pi_L^m$ depends on $F(M)$, and that if $M > m$, the inequality $F(M) \geq (M - b)/(2(m - b))$ holds, since $\pi_L^m \leq \pi_L$. When $F(M) = (M - b)/(2(m - b))$, the triangular possibility distribution $\pi_L$ is retrieved, for instance when the mode and the median coincide ($F(M) = 0.5$). If $F(M) = 1$ (the most asymmetric case) then $\pi_L^m(m) = 0.5$. Exploiting this representation needs an estimation of $F(M)$. But this quantity is a good measure of the asymmetry of the distribution.

This result is easily extended to any other fractiles, or any set of fractiles if they are known a priori. In particular, consider the case where an expert gives fractiles, say $x_1$, $x_2$ and $x_3$, at 5%, 50% and 95%, on top of the mode $M$. By definition $x_2$ is the median, and suppose that it coincides with the mode. Let $\mathcal{P}_I^{x_1,x_2,x_3}$ be the probability

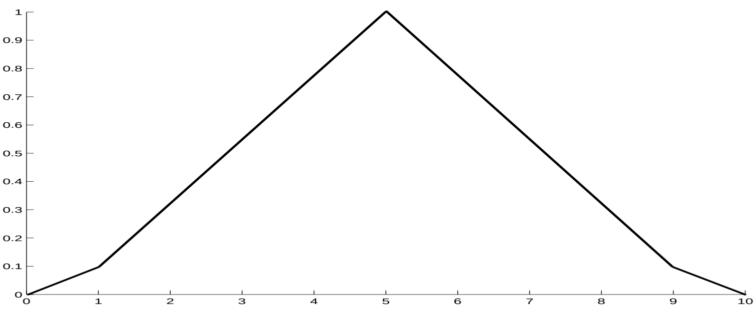


Fig. 8. Expert gives fractiles on 5%, 50% and 95% equal to 1, 5 and 9 on [0,10].

family having these fractiles defined in the Section 5.2. With the same reasoning as above, we can represent this knowledge by the following (symmetric) possibility distribution : $\pi(x_1) = \pi(x_3) = F(x_1) + 1 - F(x_3) = 0.1$, $\pi(x_2) = 1$ and linear interpolations on $[b, x_1]$, $[x_1, x_2]$, $[x_2, x_3]$ and $[x_3, c]$ for other values of $\pi(x)$ (see for instance Figure 8). Clearly $\mathcal{P}_I^{x_1,x_2,x_3} \cup \mathcal{P}^{M,m} \subset \mathcal{P}(\pi)$ (respecting the *Dominance condition* defined in Section 4).

### *6.3 Distributions with known mode and support: bracketing prediction intervals*

Suppose that $I = [b, c]$ contains the support of the unknown probability distribution function $p$ and the symmetry of $p$ is assumed. Let $\mathcal{P}_I^S$ be the set of such probabilities. Their mode is $(b + c)/2$ due to symmetry (but it includes the uniform probability on $I$). If $p$ is symmetric, the optimal transform $\pi_p^\star$ around the mode is convex on each side of the mode [11]. The symmetric triangular possibility distribution $\pi_S$ with support $I$ and core $(b+c)/2$ is thus such that $\pi_S \geq \pi_p^\star, \forall p$, and is really equal to $sup_{p\in\mathcal{P}_I^S}\pi_p^\star$ [11]. So not only do $\mathcal{P}(\pi_S)$ contain $\mathcal{P}_I^S$ but also, the $\alpha$-cuts of $\pi_S$ braket the narrowest prediction intervals of these probabilities. Nevertheless, $\mathcal{P}(\pi_S)$ also contains probability densities that are not symmetric and whose mode differ from $(b + c)/2$ (but they do not bracket their prediction intervals, necessarily). One may argue that the p-box $[\underline{F}_\star, \overline{F}^\star]$ defined by $\overline{F}^\star(x) = (x - b)/(c - b)$ if $x \leq (b + c)/2$, and 1 otherwise, $\underline{F}_\star(x) = (x - b)/(c - b)$ if $x \geq (b + c)/2$, and 0 otherwise, is a more informative representation of symmetric densities with support in $I$. But note that in this case, the possibility distribution $\pi(x) = \min(\overline{F}^\star(x), 1 - \underline{F}_\star(x))$ if $x \neq (b+c)/2$ is also dominating all such symmetric distributions, and is even more precise than the p-box. But of course, it cannot bracket their prediction intervals. The specific merit of $\pi_S$ is precisely to bracked the prediction intervals in $\mathcal{P}_I^S$, Interestingly, note that $\pi_S = 2 \cdot \min(\overline{F}^\star, 1 - \underline{F}_\star)$ for $x \neq (b + c)/2$. If we know some fractiles, we can refine the representation as explained in the previous section. Such refinements would respect the *prediction interval condition* (see Figure 8) due to the symmetry assumption.

When $p$ is asymmetric, the optimal transform $\pi_p^\star$, associated to $p$ may fail to be convex on each side of the mode $M$. So the $\alpha$-cuts of the triangular possibility distribution $\pi_L$ with core $\{M\}$ do not always contain the optimal $(1 - \alpha)$-prediction intervals of the probability measures of mode $M$, as clear from theorem 4 on the optimal transforms of piecewise linear densities. For instance consider the example on Figure 9 suggested in [11], where:

$$p(x) = 0.6x + 1.2 \text{ on } [-2, -1.5],$$

$$p(x) = (0.2/3)x + 0.4 \text{ on } [-1.5, 0]$$

and

$$p(x) = -0.2x + 0.4 \text{ on } [0, 2].$$

The interval $[-1.4, 1.4]$ corresponding to the $\alpha$-cut equal to 0.3 of the triangular possibility distribution does not contain the optimal 0.7-prediction interval of probability measure of mode 0, which is $[-1.5, 1.5]$ : the optimal transform of $p$ (in Section 4) is indeed not convex everywhere.

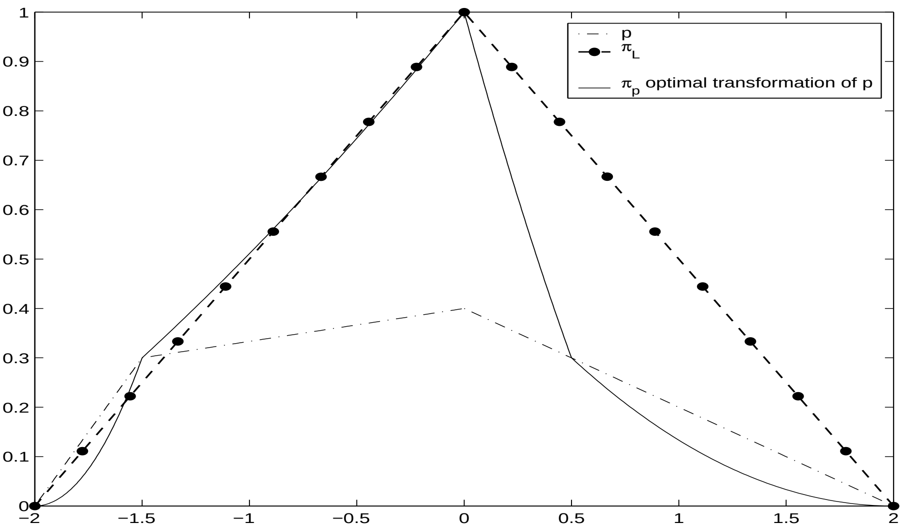


Fig. 9. Optimal transformation of $p$ around the mode.

We can nevertheless find an upper bound of $\pi_p^\star$ for a unimodal asymmetric continuous density $p$. Then, using the concavity of $F$ and considering nested intervals $J_x = [x, max\{y, p(y) \geq p(x)\} = f(x)]$ we have:

- For $x \leq M, \pi_p^\star(x) \leq F(x)+1-F(f(x)) \leq F_L(x)+1-F(M) = \frac{F(M)(x-b)}{M-b}+1-F(M)$.
- For $x \geq M, \pi_p^\star(x) \leq F(f^{-1}(x))+1-F(x) \leq F(M)+1-F_L(x) = 1-\frac{1-F(M)}{c-M}(x-M)$.

Knowing the value $F(M)$ is necessary to be able to define this approximation (see Figure 10 for instance). In general, it will be difficult to come up with a more informative possibility distribution which accounts for the prediction intervals of all probability measures on an interval $I$ with fixed mode, due to the wide range of such distributions. Some additional assumptions must be made, for instance on the convexity-concavity of the unknown probability density function $p$.

**Theorem 6** *If the density function $p$ is convex increasing on $]b, M[$ and concave strictly decreasing on $]M, c[$, then $\pi_p^\star$ is also convex on $]b, M[$.*

***Proof.*** see appendix C.

The assumption $F(M) < 0.5$ is consistent with the convexity of $p$ on $]b, M[$ and its concavity on $]M, c[$. In this case, a possibility distribution linearly increasing from 0 to 1 on $[b, M]$ covers all optimal transforms of such densities on this side. On the other side of the mode, using a linear shape is possible with $\pi(c) = 1 -$

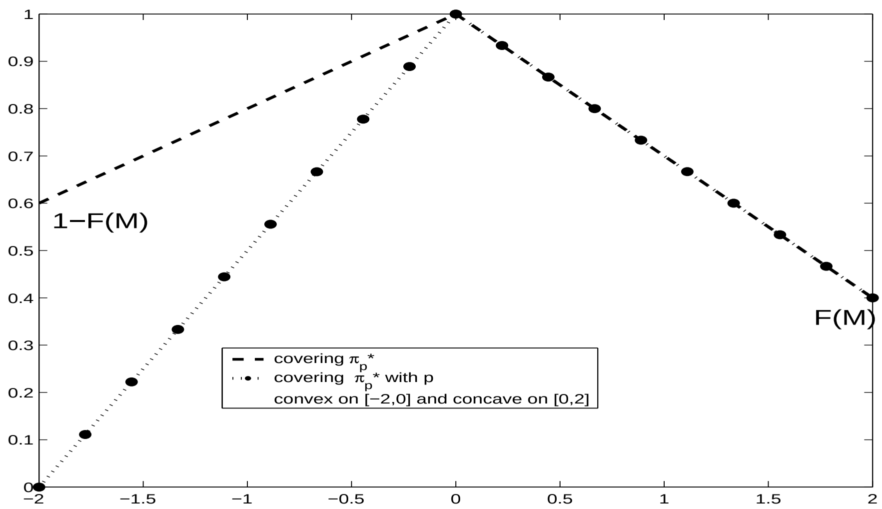


Fig. 10. the upper bound of $\pi_p^\star$ and its improvement when the concavity-convexity of $p$ is know for $M$=0 and $F(M) = 0.4$.

$F(M)$ (see Figure 10). In summary, assuming $F(M)$ is known and the assumption of theorem 5 on convexity & monotony of $p$ holds, then a more informative possibility distribution whose cuts contain the confidence intervals of distributions of mode $M$ having such characteristics can be computed.

## 7 Conclusion

The notion of imprecise probability offers a natural formal framework for representing imprecise knowledge on numerical quantities. Several types of information can be approximated by means of possibility distributions, others are directly and exactly representable by belief functions, yet other more naturally fit the probability box framework. In several cases, possibility distributions provide a concise approximate representation of a set of probability measures, sometimes interpretable in terms of confidence intervals of probabilities in such families. In fact each mode of representation seems to be adapted to the knowledge of specific characteristics of distributions. Only p-boxes seem to capture information about mean values in a reasonable way. Belief functions directly model fractile information, while possibility measures are particularly well-suited for representing families of distributions whose mode is known, and can integrate additional information on symmetry and concavity of densities, as well as known fractiles. The recent works of Neumaier [29] focus on probability families $\mathcal{P}$ of the form $\mathcal{P} = \mathcal{P}(\pi) \cap \mathcal{P}(1-\rho)$ where $\pi$ is a possibility distribution and $\rho$ is a function $I \to [0,1]$ acting as a lower bound of $\pi$, i.e. $\rho \leq \pi$. The probability family $\mathcal{P}(\pi) = \mathcal{P}$ is recovered when $\rho = 0$. The probability family $\mathcal{P}$ is more precise than $\mathcal{P}(\pi)$ and assessing its potential demands more future investigations.

Our representation tools using possibility theory are currently applied to risk management problems [5,6]. In such problems, straightforward Monte-Carlo methods involve too rich assumptions of complete probabilistic knowledge and stochastic independence between parameters. Moreover, uncertainty due to variability and uncertainty due to incomplete knowledge are mixed up in the resulting distribution. In contrast, Bardossy et al. [1], Bárdossy and Fodor [2], Dou et al. [10,9] present applications of possibility theory to propagate imprecise information in environmental models. However a proper handling of real cases requires the propagation of heterogeneous uncertain information where imprecision and variability of parameters are separately accounted for and propagated through numerical models. Guyonnet et al. [23] (see also Bárdossy and Fodor [2]) propose a method for the joint propagation of fuzzy intervals and probabilistic numbers. This method is further elaborated in [3,5]. Various joint possibility-probability propagation techniques are compared in [6], some involving independence assumptions, other ones, more conservative, avoiding such assumptions. Comparison with p-box propagation is also made. For a stimulating discussion of various uncertainty propagation techniques, using random intervals imprecise probability and possibility theory, see [25].

The unified representation framework proposed here makes it easy to represent poor data of various types in a faithful and yet simple way. It facilitates the definition of a uniform mode of propagation in risk management, in spite of the heterogeneous character of the data collected and the computation of conservative estimates, something that is not allowed by traditional probabilistic methods.


## Acknowledgements

This work is supported by the French Institutes B.R.G.M, I.R.S.N and I.N.E.R.I.S.


## A Proof of theorem 1

$\subseteq$: Let be $P \in \mathcal{P}(\pi)$ and an interval $A$ such that $A = [x, y]$ containing $a$.
By definition, $N(A) \leq P(A)$ is equivalent to $F(y) - F(x) \geq 1 - \sup_{z \notin [x,y]} \pi(z)$, i.e. $F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y))$. We thus have $\mathcal{P}(\pi) \subseteq \{P, \forall x, y, x \leq a \leq y, F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y))\}$.
$\supseteq$: Let be $P \in \{P, \forall x \leq a \leq y, F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y))\}$. Considering any measurable A.

(a) For $A = (-\infty, x]$ with $x \leq a$, $F(x) + 1 - F(+\infty) \leq \max(\pi(x), \pi(+\infty)) \Leftrightarrow F(x) \leq \pi^+(x) \Rightarrow P(A) \leq \Pi(A)$.
(b) For $A = [y, +\infty)$ with $y \geq a$, $F(-\infty) + 1 - F(y) \leq \max(\pi(y), \pi(-\infty)) \Leftrightarrow 1 - F(y) \leq \pi^-(y) \Rightarrow P(A) \leq \Pi(A)$.

(c) For $A = [x, y]$ with $y \leq a$, knowing that $F$ is increasing and according to case (a), we have $F(y) - F(x) \leq F(y) \leq \pi^+(y)$. Hence $P(A) \leq \Pi(A)$.

(d) For $A = [x, y]$ with $x \geq a$, knowing that $F$ is limited by 1 and according to case (b), we have $F(y) - F(x) \leq 1 - F(x) \leq \pi^-(x)$. Hence $P(A) \leq \Pi(A)$.

(e) For $A$, union of intervals such that $\Pi(A) < 1$. Suppose $\Pi(A)$ is obtained for some $y$ which lies on the right side of $a$. We may consider a set $A' = (-\infty, x] \cup [y, +\infty)$ such that $\pi(x) = \pi(y)$. Necessarily, $A'$ contains $A$, and we have $\Pi(A) = \Pi(A') = \pi(x)$ and $P(A) \leq P(A')$. We have $x \leq a \leq y$ ,thus $P(A) \leq P(A') = F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y)) = \Pi(A') = \Pi(A)$. We then have $P(A) \leq \Pi(A)$.

(f) For $A$, union of intervals such that $\Pi(A) = 1$, choose $y$ on the boundary of $A$ such that $\pi(y)$ is maximal. Suppose that $y$ is on the right of $a$, we can consider a set as $A' = [x, y] \subset A$ such that $\pi(x) = \pi(y)$. We have $\Pi(A) = \Pi(A') = 1$ and $N(A) = N(A')$, moreover $x \leq a \leq y$ thus $F(x) + 1 - F(y) \leq \max(\pi(x), \pi(y)) \Leftrightarrow F(y) - F(x) \geq 1 - \pi^-(y)$. we then have, $N(A) = N(A') \leq P(A') \leq P(A)$, thus $P(A) \geq N(A)$.□

## B Proof of theorem 4

First, we show that the optimal transform of a triangular density function $p$, is convex on each side of the mode $M$ (see Figure B.1). Let $[b, c]$ be the support of $p$.We have:

$$p_-(x) = \tfrac{p(M)}{M-b}(x - b) \text{ and } p_-^{-1}(\lambda) = \tfrac{M-b}{p(M)}\lambda + b$$

and

$$p_+(x) = \tfrac{p(M)}{c-M}(c - x) \text{ and } p_+^{-1}(\lambda) = c - \tfrac{c-M}{p(M)}\lambda.$$

For $\lambda \in [0, p(M)]$, we obtain $\pi_p^\star(p_-^{-1}(\lambda)) = \pi_p^\star(p_+^{-1}(\lambda)) = \frac{\lambda^2}{2p(M)}(c - b)$. Then:

- For $x \leq M$, by putting $\lambda = p_-(x)$, we have:

$$\pi_p^\star(x) = \frac{p(M)(c - b)}{2(M - b)^2}(x - b)^2,$$

  whose second derivative is positive, hence $\pi_p^\star$ is convex on $[b, M]$.
- Similarly, for $x \geq M$, by putting $\lambda = p_+(x)$, we have:

$$\pi_p^\star(x) = \frac{p(M)(c - b)}{2(c - M)^2}(x - c)^2.$$

Hence $\pi_p^\star$ is convex on $[M, c]$.

Now let $p$ be piecewise linear and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the ordinates of the points where the slope changes. In particular $p(M) = \lambda_n$ and $p(b) = p(c) = \lambda_1 = \lambda_2 = 0$.

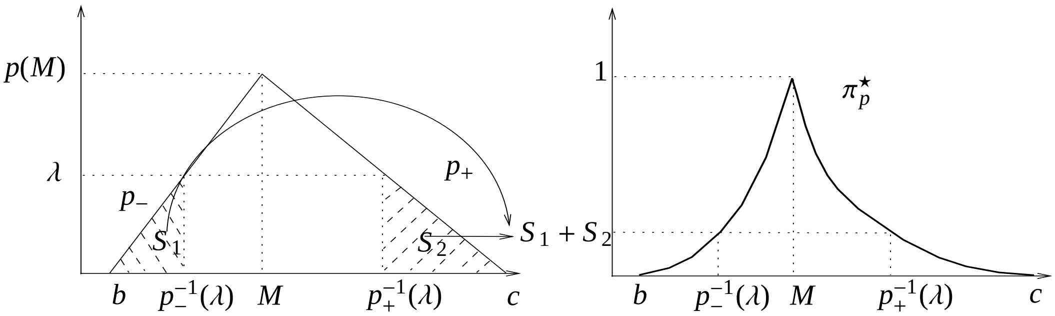


Fig. B.1. Triangular probability density $p$ on the left and the shape of its $\pi^\star_p$ optimal transformation on the right.

For illustration, we picture the case where the density $p$ is linear on 4 intervals $[b = \min(supp(p)), a_2]$, $[a_2, M]$, $[M, a_4]$ ,$[a_4, c = \max(supp(p))]$ (see Figure B.2).

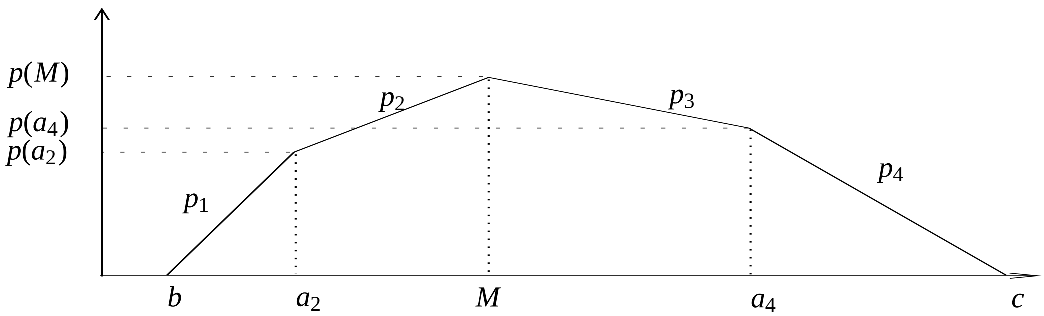


Fig. B.2. Linear unimodal continuous probability density.

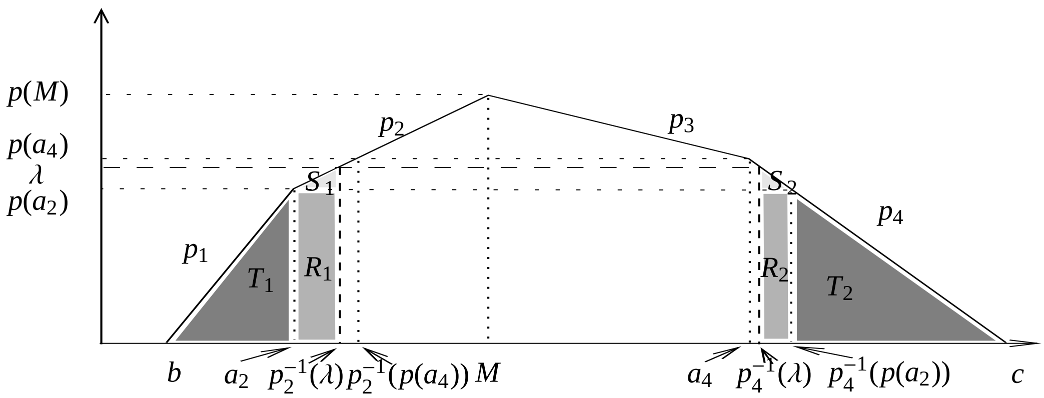


Fig. B.3. Step of the optimal transformation of linear unimodal continuous probability density when $\lambda \in [p(a_2), p(a_3)]$.

Consider index $i$ such that $\lambda_i < \lambda_{i+1}$, $\lambda \in [\lambda_i, \lambda_{i+1}]$. Denote $[\underline{b}_i, \overline{b}_i]$ and $[\underline{b}_{i+1}, \overline{b}_{i+1}]$ the intervals whose end-points have ordinates $\lambda_i$ and $\lambda_{i+1}$, and $[x, y]$ such that $p(x) = p(y) = \lambda$ (see Figure B.3 where $\lambda_i = p(a_2)$, $\lambda_{i+1} = p(a_4)$, $[\underline{b}_i, \overline{b}_i] = [a_2, p_4^{-1}(p(a_2))]$, $[\underline{b}_{i+1}, \overline{b}_{i+1}] = [p_2^{-1}(p(a_4)), a_4]$) and $[x, y] = [p_2^{-1}(\lambda), p_4^{-1}(\lambda)]$. The integral computing $\pi^\star_p(x) = \pi^\star_p(y)$ contains a constant part corresponding to the areas under $p$ outside the interval $[\underline{b}_i, \overline{b}_i]$ ($T_1$ and $T_2$ in Figure B.3), plus a part linear in $\lambda$ corresponding to rectangles ($R_1$ and $R_2$ in Figure B.3) of areas $\lambda_i \cdot (x - \underline{b}_i)$ and $\lambda_i \cdot (\overline{b}_i - y)$ inside the

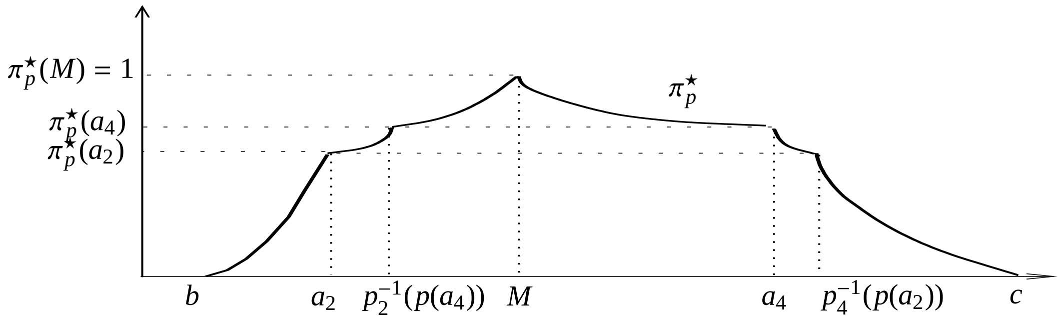


Fig. B.4. Shape of $\pi_p^\star$ optimal transformation of linear unimodal continuous probability density.

intervals $[\underline{b}_i, x]$ and $[y, \overline{b}_i]$, plus a quadratic part in $\lambda$ corresponding to the area of the remaining triangles ($S_1$ and $S_2$ in Figure B.3) located inside the intervals $[\underline{b}_i, x]$ and $[y, \overline{b}_i]$ and bounded by $p$ and the horizontal line of ordinate $\lambda_i$ and the verical lines of abcissae $x$ and $y$ respectively. The second derivative of $\pi_p^\star$ is this equal to zero except for the quadratic part of $\pi_p^\star$ for which it is constant. Hence, we find the same expression as in the optimal transformation triangular density $p$ (see Figure B.1). Then, $\pi_p^\star$ is piecewise convex and Figure B.4 shows the shape of $\pi_p^\star$ in the case where the density $p$ is linear on 5 points.□

## C Proof. of theorem 6

We must show that the second derivative of $\pi_p^\star$ is positive on $]b, M[$. Consider $p_1$ (the left part of $p$) and $p_2$ (the right part of $p$) defined as follows:

- $\forall x \in [b, M]$, $p_1(x) = p(x)$ and 0 otherwise.
- $\forall x \in [M, c]$, $p_2(x) = p(x)$ and 0 otherwise.

For $x \in [b, M]$, $\pi_p^\star(x) = F(x) + 1 - F(f(x))$ where $f(x) = max\{y, p(y) \geq p(x)\}$.
If we differentiate $\pi_p^\star$ on $]b, M[$, we obtain:

$$\pi_p^{\star'}(x) = F^{'}(x) - f^{'}(x)F^{'}(f(x)) = p_1(x) - f^{'}(x)p_2(f(x)).$$

However $p_1(x) = p_2(f(x))$, thus:

$$\pi_p^{\star'}(x) = p_1(x)\left(1 - f^{'}(x)\right).$$

Hence differentiating again:

$$\pi_p^{\star''}(x) = p_1^{'}(x)\left(1 - f^{'}(x)\right) - p_1(x)f^{''}(x).$$

We know that $p_1(x) = p_2(f(x))$; if we differentiate this equality, we obtain:

$$f^{'}(x) = \frac{p_1^{'}(x)}{p_2^{'}(f(x))}.$$

The function $p_1$ increases on $]b, M[$, then $p_1^{'} \geq 0$. The function $p_2$ strictly decreases on $]M, c[$, then $p_2^{'} < 0$. We thus deduce that $f^{'} \leq 0 \leq 1$. We conclude that:

$$p_1^{'}(x)(1 - f^{'}(x)) \geq 0, \ \forall x \in ]b, M[.$$

By differentiating again $f^{'}$, we obtain

$$f^{''}(x) = \frac{p_1^{''}(x) - (f^{'}(x))^2 p_2^{''}(f(x))}{p_2^{'}(x)}.$$

We know that p is convex on $]b, M[$ (resp. concave on $]M, c[$), we have $p_1^{''}(x) \geq 0$ for all $x \in ]b, M[$ (resp. $p_2^{''}(x) \leq 0$ for all $x \in ]M, c[$). Hence, $p_1^{''}(x) - (f^{'}(x))^2 p_2^{''}(f(x)) \geq 0$ for all $x \in ]b, M[$ and thus $f^{''}(x) \leq 0$ for all $x \in ]b, M[$. We thus conclude that $p_1(x) f^{''}(x) \leq 0, \ \forall x \in ]b, M[$.

To summarize, we have $p_1^{'}(x)(1 - f^{'}(x)) \geq 0$ and $p_1(x) f^{''}(x) \leq 0, \ \forall x \in ]b, M[$ We thus have proved that $\pi_p^{\star''}$ is positive on $]b, M[$, and hence the convexity of $\pi_p^{\star}$ on $]b, M[$.□

## References


[1] Bardossy, A., Bronstert, A., and Merz, B. 1-, 2- and 3-dimensional modeling of groundwater movement in the unsaturated soil matrix using a fuzzy approach. *Advances in Water Resources*, 18(4), 237-251, 1995.

[2] Bárdossy, G., Fodor, J. Evaluation of Uncertainties and Risks in Geology: New Mathematical Approaches for their Handling. *Springer, Berlin*, ISBN 3-540-20622-1, 2004.

[3] Baudrit, C., Dubois, D., Guyonnet, G. Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment Models. *Accepted to IEEE Transaction on Fuzzy Systems.*

[4] Baudrit, C., D. Dubois, H. Fargier, H.. Practical representation of incomplete probabilistic information. *Advances in Soft Computing:Soft Methods of Probability and Statistics conference, Oviedo*, 149-156, 2004.

[5] Baudrit, C., Dubois, D., Guyonnet, D., Fargier, H. Joint Treatment of imprecision and Randomness in Uncertainty Propagation. *Proc. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Perugia*, 873-880, 2004.

[6] Baudrit, C., Dubois, D. Comparing Methods for Joint Objective and Subjective Uncertainty Propagation with an example in a risk assessment. *Fourth International Symposium on Imprecise Probabilities and Their Application (ISIPTA'05), Pittsburg (USA, Pennsylvanie)*, p 31-40, 2005.

[7] De Cooman, G., Aeyels, D. Supremum-preserving upper probabilities. *Information Sciences*, 118, 173-212, 1999.

[8] Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals of Mathematical Statistics*, 38, 325-339, 1967.

[9] Dou, C., Woldt, W., Bogardi, I., and Dahab, M. Numerical solute transport simulation using fuzzy sets approach. *Journal of Contaminant Hydrology.*, 27, 107-126, 1997.

[10] Dou, C., Woldt, W., Bogardi, I., and Dahab, M. Steady-state groundwater fbw simulation with imprecise parameters. *Water Resource Res.*, 31(11), 2709-2719, 1995.

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

[12] Dubois, D., Nguyen, H.T., Prade, H. Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. *Fundamentals of Fuzzy Sets*, Dubois,D. Prade,H., Eds: Kluwer , Boston, Mass, 343-438 , 2000.

[13] Dubois, D., Prade, H. The mean value of a fuzzy number. *Fuzzy Sets & Systems*, 24, 279-300, 1987.

[14] Dubois, D., Prade, H. Consonant approximations of belief functions. *Int. J. Approximate Reasoning, Special Issue: Belief Functions and Belief Maintenance in Artificial Intelligence*, 4(5/6), 1990, 419-449.

[15] Dubois, D., Prade, H., Sandri, S.A. On possibility/probability transformations.*Fuzzy Logic: State of the Art* (R. Lowen, M. Lowen, eds.), Kluwer Academic Publ., 103-112, 1993.

[16] Dubois, D., Prade, H. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49, 65-74, 1992.

[17] Dubois, D., Prade, H. Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.

[18] Dubois, D., Prade, H., Smets, P. Representing partial ignorance. IEEE Trans. on Systems, Man and Cybernetics, 26(3), 1996, 361-377.

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

[20] Ferson, S., Ginzburg, L.R. Different methods are needed to propagate ignorance and variability. *Reliability Engineering and Systems Safety*, 54, 133-144, 1996.

[21] Ferson, S., Ginzburg, L., Kreinovich, V., Myers, D.M., Sentz, K. Construction of Probability Boxes and Dempster-Shafer structures. *Sandia National Laboratories, Technical report SANDD2002-4015*, 2003.
Available at www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf.

[22] Ferson, S., Ginzburg, L., Akcakaya, R. Whereof one cannot speak: when input distributions are unknown. To appear in *Risk Analysis*.

[23] Guyonnet D., Bourgine B., Dubois D., Fargier H., Côme B., Chilès J;-P. Hybrid Approach for Addressing Uncertainty in Risk Assessments. Journal of Environmental Engineering V. 129 N. 1 , p. 68-78 , 2003.

[24] Gzyl, H. The Method of Maximum Entropy. *Series on Advances in Mathematics for Applied Sciences*, 29,1995.

[25] Helton, J.C., J.D. Johnson, and W.L. Oberkampf. 2004. An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions,Ó Reliability Engineering and System Safety. Vol. 85, no. 1-3, pp. 39-71.

[26] Jaffray J.Y. (1992). Bayesian updating and belief functions, IEEE Trans. on Systems, Man and Cybernetics, 22, 1144-1152.

[27] Kendall, M. and Stuart, A. The Advanced Theory of Statistics. *ED. Griffin and Co.*, 1977.

[28] Miller, L.H. Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association*, 51, 111-121, 1956.

[29] Neumaier, A. Clouds, Fuzzy Sets and Probability Intervals. *Reliable Computing, Kluwer Academic Publishers*, V.10, N.4, p. 249-272, 2004.

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

[31] Smets P. and Kennes R. (1994). The transferable belief model, Artificial Intelligence, 66, 191-234.

[32] Walley, P. Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991

[33] Williamson, R.C., Downs, T. Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4, 8-158, 1990.