COMPUTING EXPECTATIONS WITH CONTINUOUS P-BOXES: UNIVARIATE CASE

LEV UTKIN AND SEBASTIEN DESTERCKE

ABSTRACT. Given an imprecise probabilistic model over a continuous space, computing lower/upper expectations is often computationally hard to achieve, even in simple cases. Because expectations are essential in decision making and risk analysis, tractable methods to compute them are crucial in many applications involving imprecise probabilistic models. We concentrate on p-boxes (a simple and popular model), and on the computation of lower expectations of non-monotone functions. This paper is devoted to the univariate case, that is where only one variable has uncertainty. We propose and compare two approaches : the first using general linear programming, and the second using the fact that p-boxes are special cases of random sets. We underline the complementarity of both approaches, as well as the differences.

1. INTRODUCTION

There are many situations where a unique probability distribution cannot be identified to describe our uncertainty about the value assumed by a variable on a state space. This can happen for example when data or expert judgments are not sufficient and/or are conflicting. In such cases, a solution is to model information by the means of imprecise probabilities, that is by considering either sets of probability distributions [17, 14] or bounds on expectations [18]. Note that, from a purely mathematical point of view, such representations encompass many other frameworks dealing with the representation of incomplete and conflicting information, such as random sets [7] and possibility theory [12].

When considering such models, the expectation of a real-valued bounded function over the state space is no longer precise and is lower- and upper-bounded by some value. In applications involving risk analysis or decision making, the decision process will be based on the values of these lower and upper expectations, using extensions of the classical expected utility criterion [25]. When the state space on which the variable assumes its value is finite, lower and upper expectations can be numerically computed by using, for instance, linear programming techniques [26]. The problem becomes quite more complicated when uncertainty models are defined over infinite state spaces (e.g., the real line, product spaces, . . .).

In this latter case, computing exactly and analytically the lower and upper expectations of a given function is impossible most of the time, and there are very few methods and algorithms around to compute approximations of these bounds [4, 21, 24]. In this paper, we study such analytical solutions for a specific case, that is the one where the uncertainty over a variable is described by a pair of upper and lower cumulative distributions (a so-called p-box [13]). In essence, such

a study comes down to search the extremal points of the p-box for which the expectation bounds are reached. The features of these solutions also allow us to suggest some ways to build more efficient numerical methods and algorithms, useful when analytical solutions cannot be computed. We also assume that the function over which lower and upper expectations have to be computed can be non-monotone but has a (partially) known behaviour. In this paper, we concentrate on the univariate case, i.e., where the value assumed by only one variable is tainted with uncertainty. The multivariate case as well as the case of mixed strategies (expectation bounds computed over mixture of functions) are left for forthcoming papers.

P-boxes are one of the simplest and most popular models of sets of probability distributions, directly extending cumulative distributions used in the precise case. P-boxes are often used in applications [16], as they can be easily derived from small samples [3] or from expert opinions expressed in terms of imprecise percentiles. consequently, our study is likely to be useful in many practical situations. P-box models can also be found in robust Bayesian analysis, where they are known as distribution band classes [2]. In other cases, the poor expressiveness of p-boxes compared to more general sets of probabilities is clearly a limitation [8]. However, as we shall see, their simplicity allows for more efficient computations, and they can provide quick first approximations. Eventually, if these first approximations already allow to take a decision, there is no need to consider more complex (and computationally demanding) models.

Methods developed in the paper are based on two different approaches, and we found it interesting to emphasize similarities and differences between these approaches, as well as how one approach can help the other: the first is based on the fact that the computation of bounding expectations can be viewed as a linear programming problem, while the second uses the fact that a p-box is a particular case of a random set [16, 8]. Approximating lower and upper expectations with these approaches mainly consists in discretizing the uncertainty models. In this sense, they are different from other approaches discretizing the state space [21, 24].

We first state the general problem in Section 2, how to solve it by using linear programming and random sets, and introduce the problem of conditioning by an observed event. We then study the computation of lower/upper expectations of a function over the p-box for different behaviours. Going from the simplest case to the most general one, we start with monotone functions in Section 3, pursue with functions having one extrema in Section 4, and finish by general (bounded) continuous functions in Section 5.

2. General problem statement

We assume that the information about a (real-valued) random variable X is (or can be) represented by a lower $\underline{F}$ and upper $\overline{F}$ cumulative probability distributions defining the p-box $[\underline{F}, \overline{F}]$ [13]. Lower $\underline{F}$ and upper $\overline{F}$ distributions thus define a set $\Phi(\underline{F}, \overline{F})$ of precise distributions such that

$$(1) \qquad \Phi(\underline{F}, \overline{F}) = \{F | \forall x \in \mathbb{R}, \ \underline{F}(x) \leq F(x) \leq \overline{F}(x)\}.$$

Given a function $h(X)$, lower ($\underline{\mathbb{E}}$) and upper ($\overline{\mathbb{E}}$) expectations over $[\underline{F}, \overline{F}]$ of $h(X)$ can be computed by means of a procedure sometimes called natural extension [30,

31], which corresponds to the following equations:

$$(2) \qquad \underline{\mathbb{E}}(h) = \inf_{F \in \Phi(\underline{F}, \overline{F})} \int_{\mathbb{R}} h(x) \mathrm{d}F, \overline{\mathbb{E}}(h) = \sup_{F \in \Phi(\underline{F}, \overline{F})} \int_{\mathbb{R}} h(x) \mathrm{d}F.$$

Computing the lower (resp. upper) expectation can be seen as finding the extremizing distribution F inside $\Phi(\underline{F}, \overline{F})$ reaching the infimum (resp. supremum) in Equations (2). If we consider the convex set of probabilities induced by $\Phi(\underline{F}, \overline{F})$, this is equivalent to find the extremum point (i.e., vertex) of this convex set where the bounds are reached, among all vertices (here infinitely many). Solving Equations (2) exactly is usually very difficult, although sometimes possible, even when analytical expressions of $h, \overline{F}, \underline{F}$ are known. In practice, numerical methods must often be used to solve the problem and estimate both the upper and lower expectations. Upper and lower expectations are dual [31, ch.2.], in the sense that $\underline{\mathbb{E}}(h) = -\overline{\mathbb{E}}(-h)$. This will allow us to concentrate only on the lower expectations for some cases studied in the sequel. We now detail the two generic approaches used throughout the paper to solve the above problem. Note that, through all the paper, we assume that we restrict ourselves either to σ-additive probabilities or to continuous functions h, as such assumptions are not, from a practical standpoint, very limiting.

We will denote by I_A the indicator function of the set A, that is the function such that $I_A(x) = 1$ if $x \in A$, zero otherwise. The lower (resp. upper) expectation of this function, $\underline{\mathbb{E}}(I_A)$ (resp. $\overline{\mathbb{E}}(I_A)$), have the same value as the lower (resp. upper) probability $\underline{P}(A)$ (resp. $\overline{P}(A)$ of the event A induced by the set $\Phi(\underline{F}, \overline{F})$.

2.1. **Linear programming view.** Although we assume that the readers have basic knowledge of linear programming (for an introduction to the topic, see for example Vanderbei [29]), we will recall basic results coming from this theory when they are used in the paper.

As sets of probabilities can be expressed through linear constraints over expectations, and as expectation is a linear functional, it is quite natural to translate Equations (2) into linear programs. The linear programs corresponding to lower expectation are summarized below.

Primal problem:	Dual problem:
Min. $\mathbf{v} = \int\limits_{-\infty}^{\infty} h(x)\rho(x)\,\mathrm{d}x$	Max. $\mathbf{w} = c_0 + \int\limits_{-\infty}^{\infty} \left(-c(t)\overline{F}(t) + d(t)\underline{F}(t)\right)\mathrm{d}t$
subject to	subject to
$\rho(x) \geq 0, \int\limits_{-\infty}^{\infty} \rho(x)\,\mathrm{d}x = 1,$	$c_0 + \int\limits_{x}^{\infty} (-c(t) + d(t))\,\mathrm{d}t \leq h(x),$
$-\int\limits_{-\infty}^{x} \rho(x)\,\mathrm{d}x \geq -\overline{F}(x),$	$c_0 \in \mathbb{R}, c(x) \geq 0, d(x) \geq 0.$
$\int\limits_{-\infty}^{x} \rho(x)\,\mathrm{d}x \geq \underline{F}(x).$	

Where $\mathbf{v}$ and $\mathbf{w}$ are the objective functions to respectively minimize and maximize for the primal and dual problems, and $\rho(x)$ is a probability density function having

a cumulative distribution inside $\Phi(\underline{F}, \overline{F})$. Since both the primal and dual problems are feasible (i.e. have solutions satisfying their constraints), then their optimal solutions coincide (due to strong duality [29, Ch.5]) and are equal to $\underline{\mathbb{E}}(h)$.

Numerically solving the above problem can be done by approximating the probability distribution function F by a set of N points $F(x_i)$, $i = 1, ..., N$, and by translating equations (2) into the corresponding linear programming problem with N optimization variables and where constraints correspond to equation (1). Those linear programming problems are of the form

$$(3) \qquad \underline{\mathbb{E}}^*(h) = \inf \sum_{k=1}^{N} h(x_k) z_k \;\; \text{or} \;\; \overline{\mathbb{E}}^*(h) = \sup \sum_{k=1}^{N} h(x_k) z_k$$

subject to

$$z_i \geq 0, \;\; i = 1, ..., N, \; \sum_{k=1}^{N} z_k = 1,$$

$$\sum_{k=1}^{i} z_k \leq \overline{F}(x_i), \; \sum_{k=1}^{i} z_k \geq \underline{F}(x_i), \;\; i = 1, ..., N.$$

where the z_k are the optimization variables, and objective function $\underline{\mathbb{E}}^*(h)$ (resp. $\overline{\mathbb{E}}^*(h)$) is an approximation of the lower (resp. upper) expectation. Note that the primal problem may not always be feasible (e.g., consider $N = 1$ and $\overline{F}(x_1) - \underline{F}(x_1) < 1$) if N is too small or values x_i are badly chosen. Also, the inequality $\underline{\mathbb{E}}(h) \leq \underline{\mathbb{E}}^*(h)$ (or its converse) does not always hold when solving the above discretized problem. The approximated solution $\underline{\mathbb{E}}^*$ is thus not a guaranteed inner or outer approximation. A solution to obtain a guaranteed inner approximation is to replace, for $i = 1, \ldots, N$, $\underline{F}(x_i)$ by $\underline{F}(x_{i+1})$ in constraints $\sum_{k=1}^{i} z_k \geq \underline{F}(x_i)$, with $\underline{F}(x_{N+1}) = 1$, since in this case, any solution to the linear program would be such that, for any $x \in [x_i, x_{i+1}]$,

$$\underline{F}(x) \leq \underline{F}(x_{i+1}) \leq \sum_{k=1}^{i} z_k \leq \overline{F}(x_i) \leq \overline{F}(x),$$

consequently the (discrete) cumulative distributions formed by the values z_k, $k = 1, \ldots, N$ is in $\Phi(\underline{F}, \overline{F})$. However, for this linear program to have a solution, we must be able to choose the x_i, $i = 1, \ldots, N$ on $\mathbb{R}$ such that $\overline{F}(x_i) \geq \underline{F}(x_{i+1})$. In addition to not be always possible, this puts necessary constraints over the chosen discretization of $\mathbb{R}$.

Let us write now the dual linear programming problem for computing $\underline{\mathbb{E}}^{**}(h)$, taking points y_i different from x_i,

$$(4) \qquad \underline{\mathbb{E}}^{**}(h) = \max \left(c_0 + \sum_{i=1}^{N} \left(d_i \underline{F}(y_i) - c_i \overline{F}(y_i) \right) \right)$$

subject to $c_0 \in \mathbb{R}$, $c_i \geq 0$, $d_i \geq 0$, and

$$c_0 + \sum_{k=i}^{N} (d_k - c_k) \leq h(y_i), \;\; i = 1, ..., N,$$

where c_0, c_i, d_i are the optimization variables, $y_i = (x_{i-1} + x_i)/2$.

When both problems are discretized, equality between their optimal solutions no longer holds, but converge towards the same value as N grows. To approximate the solution, one can let N grow iteratively until the difference $|\underline{\mathbb{E}}^*(h) - \underline{\mathbb{E}}^{**}(h)|$ is smaller than a given value $\varepsilon > 0$ characterizing the accuracy of the solutions. However, this way of determining the lower and upper expectations meets some computation difficulties if many iterations are needed and if the value of N is rather large. Indeed, the primal optimization problem have N variables and $3N+1$ constraints. On the other hand, solving the primal and dual approximated problems only once with a small value of N can lead to bad approximations of the exact value. Also important is the question of how to choose or sample the values x_i to improve numerical convergence? In other words, is there some regions that should be more sampled than others. A generic algorithm (for $\underline{\mathbb{E}}$) would look as follows:

(1) Fix a precision threshold ϵ and an initial value of N
(2) Sample N values x_i s.t. $\overline{F}(x_i) > 0$ and $\underline{F}(x_i) < 1$
(3) Compute $\underline{\mathbb{E}}^*(h)$ and $\underline{\mathbb{E}}^{**}(h)$
(4) If $|\underline{\mathbb{E}}^*(h) - \underline{\mathbb{E}}^{**}(h)| \le \epsilon$, stop, else increase N and return to step 2.

In the sequel, we will see that knowing h and its behaviour can significantly improve both accuracy and efficiency of expectation bound computations. It also provides some insight as to how values x_i could be sampled.

2.2. **Random set view.** Now that we have given a global sketch of the linear programming approach, we can detail the one using random sets. Formally, a random set is a mapping Γ from a probability space to the power set $\wp(X)$ of another space X, also called a multi-valued mapping. This mapping induces lower and upper probabilities on X [7]. Here, we consider the unit interval $[0, 1]$ equipped with Lebesgue measure as the probability space, and $\wp(X)$ are the measurable subsets of the real line $\mathbb{R}$.

Given the p-box $[\underline{F}, \overline{F}]$, we will denote $A_\gamma = [a_{*\gamma}, a_\gamma^*]$ the set such that

$$a_{*\gamma} := \sup\{x \in \mathbb{R} : \overline{F}(x) < \gamma\} = \overline{F}^{-1}(\gamma),$$
$$a_\gamma^* := \inf\{x \in \mathbb{R} : \underline{F}(x) > \gamma\} = \underline{F}^{-1}(\gamma),$$

FIGURE 1. P-box as random set, illustration

By extending existing results [16, 13] to the continuous real line [9, 1], we can conclude that the p-box $[\underline{F}, \overline{F}]$ is equivalent to the continuous random set with a uniform mass density on $[0, 1]$ and a mapping (see figure 1) such that

$$\Gamma(\gamma) = A_\gamma = [a_{*\gamma}, a_\gamma^*], \ \gamma \in [0, 1].$$

Note that both $\overline{F}^{-1}(\gamma), \underline{F}^{-1}(\gamma)$ are non-decreasing functions of γ. The interest of this mapping Γ is that it allows us to rewrite equations (2) in the following form:

$$(5) \qquad \qquad \underline{\mathbb{E}}(h) = \int_0^1 \inf_{x \in A_\gamma} h(x) \, d\gamma,$$

$$(6) \qquad \qquad \overline{\mathbb{E}}(h) = \int_0^1 \sup_{x \in A_\gamma} h(x) \, d\gamma.$$

Again, finding analytical solutions of such integrals is not easy in the general case, but numerical approximations can be computed (with more or less difficulty) by discretizing the p-box on a finite number of levels γ_i, the main difficulty in the general case being to find the infimum or supremum of $h(X)$ for each discretized level. Note that, in the finite case, a random set can be represented by non-null weights, here denoted m, given to subsets of space X and summing up to one (i.e., $\sum_{E \subseteq X} m(E) = 1$). Let $\gamma_0 = 0 \leq \gamma_1 \leq \ldots \leq \gamma_M = 1$ and define the discrete random set $\overline{\Gamma}$ such that for $i = 1, \ldots, M$

$$\overline{\Gamma} := \left\{ \begin{array}{l} A_{\overline{\gamma_i}} = [a_{*\gamma_{i-1}}, a_{\gamma_i}^*], \\ m(A_{\overline{\gamma_i}}) = \gamma_i - \gamma_{i-1} \end{array} \right.$$

We denote by $\Phi(\underline{F}, \overline{F})_{\overline{\Gamma}}$ the set of precise distributions induced by $\overline{\Gamma}$. This discretization, which is an outer approximation of the p-box $[\underline{F}, \overline{F}]$ (i.e., $\Phi(\underline{F}, \overline{F}) \subset \Phi(\underline{F}, \overline{F})_{\overline{\Gamma}}$), is sometimes referred to as the ODM (Outer discretization Method) and has been studied by other authors [23]. Working with $\overline{\Gamma}$, Equations (5), (6) can be rewritten as

$$\underline{\mathbb{E}}^{\overline{\Gamma}}(h) = \sum_{i=1}^M m(A_{\overline{\gamma_i}}) \inf_{x \in A_{\overline{\gamma_i}}} h(x) \text{ and } \overline{\mathbb{E}}^{\overline{\Gamma}}(h) = \sum_{i=1}^M m(A_{\overline{\gamma_i}}) \sup_{x \in A_{\overline{\gamma_i}}} h(x).$$

Let us now define another discrete random set $\underline{\Gamma}$ such that for $i = 1, \ldots, M$

$$\underline{\Gamma} := \left\{ \begin{array}{c} A_{\underline{\gamma_i}} = [a_{*\gamma_i}, a_{\gamma_{i-1}}^*] \text{ if } a_{*\gamma_i} \leq a_{\gamma_{i-1}}^*, \quad \emptyset \text{ otherwise} \\ m(A_{\underline{\gamma_i}}) = \gamma_i - \gamma_{i-1} \end{array} \right.$$

We denote by $\Phi(\underline{F}, \overline{F})_{\underline{\Gamma}}$ the set of precise distributions induced by $\underline{\Gamma}$. $\underline{\Gamma}$ is an inner approximation of the p-box (i.e., $\Phi(\underline{F}, \overline{F})_{\underline{\Gamma}} \subset \Phi(\underline{F}, \overline{F})$), and Equations (5), (6) can again be rewritten

$$\underline{\mathbb{E}}^{\underline{\Gamma}}(h) = \sum_{i=1}^M m(A_{\underline{\gamma_i}}) \inf_{x \in A_{\underline{\gamma_i}}} h(x) \text{ and } \overline{\mathbb{E}}^{\underline{\Gamma}}(h) = \sum_{i=1}^M m(A_{\underline{\gamma_i}}) \sup_{x \in A_{\underline{\gamma_i}}} h(x).$$

Note that when there is an index i for which $A_{\underline{\gamma_i}} = \emptyset$, $\underline{\Gamma}$ does no longer describe a non-empty set of probabilities, and we will name such a random set inconsistent. This case can be compared to the case when the linear program giving guaranteed inner approximation has no feasible solutions.

We have that $\underline{\mathbb{E}}^{\overline{\Gamma}}(h) \leq \underline{\mathbb{E}}(h) \leq \underline{\mathbb{E}}^{\underline{\Gamma}}(h)$ (due to inclusions $\Phi(\underline{F}, \overline{F})_{\underline{\Gamma}} \subset \Phi(\underline{F}, \overline{F}) \subset \Phi(\underline{F}, \overline{F})_{\overline{\Gamma}}$). Thus, to approximate the solution we can again let M grow until

$|\overline{\mathbb{E}}^{\overline{\Gamma}}(h) - \underline{\mathbb{E}}^{\underline{\Gamma}}(h)|$ is smaller than a given accuracy $\varepsilon > 0$. As in the case of linear programming, choosing too few levels γ_i or using poor heuristics to find the infinimum/supremum over sets can lead to bad approximations, and if those infinimum/supremum are hard to find, computational difficulties can arise. A generic algorithm (for $\underline{\mathbb{E}}$) using random sets would be as follows

(1) Fix a precision threshold ϵ and an initial value of M
(2) Sample M values γ_i
(3) Compute $\underline{\mathbb{E}}^{\overline{\Gamma}}(h)$ and $\underline{\mathbb{E}}^{\underline{\Gamma}}(h)$
(4) If $|\underline{\mathbb{E}}^{\overline{\Gamma}}(h) - \underline{\mathbb{E}}^{\underline{\Gamma}}(h)| \leq \epsilon$, stop, else increase M and return to step 2.

Note that the distance between two consecutive γ_i, γ_{i+1} does not have to be constant. If $\underline{\Gamma}$ is inconsistent, an alternative is to use one of the two random sets Γ_1, Γ_2 such that for $i = 1, \ldots, M$

$$\Gamma_1 := \left\{ \begin{array}{l} A_{\gamma_{i,1}} = [a_{*\gamma_{i-1}}, a^*_{\gamma_{i-1}}], \\ m(A_{\gamma_{i,1}}) = \gamma_i - \gamma_{i-1}, \end{array} \right. \qquad \Gamma_2 := \left\{ \begin{array}{l} A_{\gamma_{i,2}} = [a_{*\gamma_i}, a^*_{\gamma_i}], \\ m(A_{\gamma_{i,2}}) = \gamma_i - \gamma_{i-1}. \end{array} \right.$$

The corresponding approximations read, for $j = 1, 2$,

$$\underline{\mathbb{E}}^{\Gamma_j}(h) = \sum_{i=1}^{M} m(A_{\gamma_{i,j}}) \inf_{x \in A_{\gamma_{i,j}}} h(x) \text{ and } \overline{\mathbb{E}}^{\Gamma_j}(h) = \sum_{i=1}^{M} m(A_{\gamma_{i,j}}) \sup_{x \in A_{\gamma_{i,j}}} h(x).$$

Compared to $\underline{\Gamma}$, Γ_1, Γ_2 have the advantage to always be consistent, but the obtained approximations can either outer- or inner-approximate the exact values, even if they converge towards it as M increases.

2.3. Conditional lower/upper expectations. Another quite common problem when dealing with imprecise probabilities is the procedure of conditioning and the computations of associated lower/upper conditional expectations. Suppose that we observe an event $B = [b_0, b_1]$. Then the lower and upper conditional expectations, given the p-box $[\underline{F}, \overline{F}]$ and under condition of B, can be determined as follows:

$$\underline{\mathbb{E}}(h|B) = \inf_{\underline{F} \leq F \leq \overline{F}} \frac{\int_{\mathbb{R}} h(x) I_B(x) \mathrm{d}F}{\int_{\mathbb{R}} I_B(x) \mathrm{d}F},$$

$$\overline{\mathbb{E}}(h|B) = \sup_{\underline{F} \leq F \leq \overline{F}} \frac{\int_{\mathbb{R}} h(x) I_B(x) \mathrm{d}F}{\int_{\mathbb{R}} I_B(x) \mathrm{d}F}.$$

The above formulas are equivalent to applying Bayes formula to every probability measure inside $\Phi(\underline{F}, \overline{F})$, and then retrieving the optimal bounds. Other generalisations of Bayes formula to imprecise probabilistic framework exist [11, 31], but we will restrict ourselves to the above solution, as it is by far the most used within frameworks using lower/upper expectation bounds. Also, we assume that B is large enough (or the two distributions $[\underline{F}, \overline{F}]$ close enough) so that $\underline{F}(b_1) > \overline{F}(b_0)$. This is equivalent to require $\underline{P}(B) > 0$, thus avoiding conditioning on an event of probability 0. Indeed, there are still some discussions about what should be done in presence of such events (see Miranda [18] for an introductory discussion and Cozman [5] for possible numerical solutions).

Similarly to unconditional expectations, the above problems can numerically be solved by approximating the probability distribution function F by a set of N

points $F(x_i)$, $i = 1, ..., N$, and by writing linear-fractional optimization problems[1] and then associated linear programming problems. Problems mentioned for the unconditional case can again occur. The next proposition indicates that previous results can be used to provide a more attractive formulation of $\underline{\mathbb{E}}(h|B), \overline{\mathbb{E}}(h|B)$.

Proposition 1. *Given a p-box $[\underline{F}, \overline{F}]$, a function $h(x)$ and an event B, the upper and lower conditional expectations of $h(X)$ on $[\underline{F}, \overline{F}]$ after observing the event B can be written*

$$(7) \qquad \overline{\mathbb{E}}(h|B) = \sup_{\substack{\underline{F}(b_0) \leq \alpha \leq \overline{F}(b_0) \\ \underline{F}(b_1) \leq \beta \leq \overline{F}(b_1)}} \frac{1}{\beta - \alpha} \Psi(\alpha, \beta),$$

$$(8) \qquad \underline{\mathbb{E}}(h|B) = \inf_{\substack{\underline{F}(b_0) \leq \alpha \leq \overline{F}(b_0) \\ \underline{F}(b_1) \leq \beta \leq \overline{F}(b_1)}} \frac{1}{\beta - \alpha} \Phi(\alpha, \beta),$$

with

$$\Psi(\alpha, \beta) = \int_\alpha^\beta \sup_{x \in A_\gamma \cap B} h(x) \mathrm{d}\gamma.$$

$$\Phi(\alpha, \beta) = \int_\alpha^\beta \inf_{x \in A_\gamma \cap B} h(x) \mathrm{d}\gamma.$$

General proof. We consider only upper expectation. We do not know how the extremizing distribution function behaves outside the interval B. Therefore, we suppose that the value of the extremizing distribution function at point b_0 is $F(b_0) = \alpha \in [\underline{F}(b_0), \overline{F}(b_0)]$ and its value at point b_1 is $F(b_1) = \beta \in [\underline{F}(b_1), \overline{F}(b_1)]$ (see Fig. 4). Then there holds

$$\int_{\mathbb{R}} I_B(x) \mathrm{d}F(x) = \beta - \alpha.$$

Hence, we can write

$$\overline{\mathbb{E}}(h|B) = \sup_{\substack{\underline{F}(b_0) \leq \alpha \leq \overline{F}(b_0) \\ \underline{F}(b_1) \leq \beta \leq \overline{F}(b_1) \\ \underline{F} \leq F \leq \overline{F}}} \frac{1}{\beta - \alpha} \int_{\mathbb{R}} h(x) I_B(x) \mathrm{d}F(x)$$

$$= \sup_{\substack{\underline{F}(b_0) \leq \alpha \leq \overline{F}(b_0) \\ \underline{F}(b_1) \leq \beta \leq \overline{F}(b_1)}} \frac{1}{\beta - \alpha} \left(\sup_{\substack{\underline{F} \leq F \leq \overline{F} \\ F(b_0) = \alpha \\ F(b_1) = \beta}} \int_{\mathbb{R}} h(x) I_B(x) \mathrm{d}F(x) \right)$$

$$(9) \qquad = \sup_{\substack{\underline{F}(b_0) \leq \alpha \leq \overline{F}(b_0) \\ \underline{F}(b_1) \leq \beta \leq \overline{F}(b_1)}} \frac{1}{\beta - \alpha} \int_\alpha^\beta \sup_{x \in A_\gamma \cap B} h(x) \mathrm{d}\gamma.$$

By using the results obtained for the unconditional upper expectation, we can see that the integrand is equal to $\Psi(\alpha, \beta)$. The lower expectation is similarly proved.

$\square$

---

[1]Problems where the objective function is a fraction of two linear functions and constraints are linear.

As value $\beta - \alpha$ increases in Equations (7)-(8), so do the numerator and denominator, thus playing opposite role in the evolution of the objective function. Hence, in order to compute the upper (resp. lower) conditional expectation, one has to find the values β and α such that any increase (decrease) in the value $\beta - \alpha$ is greater (resp. lower) than the corresponding increase (resp. decrease) in $\Psi(\alpha, \beta)$ $(\Phi(\alpha, \beta))$.

A crude algorithm to approximate the solution would be to samples different values $\alpha \in [\underline{F}(b_0), \overline{F}(b_0)]$ and $\beta \in [\underline{F}(b_1), \overline{F}(b_1)]$, evaluating Equations (7)-(8) for all combination $[\alpha, \beta]$ and retaining the highest obtained value (note that we can have $\overline{F}(b_0) \geq \underline{F}(b_1)$, hence the need to make sure by adding constraint that $[\alpha, \beta]$ is not void).

Another interesting point to note is that the proof takes advantage of both views, since the idea to use levels α and β comes from fractional linear programming, while the final equation (9) can be elegantly formulated by using the random set view.

In any cases (lower/upper and conditional/unconditional expectations), it is obvious that the extremizing probability distribution F providing the minimum (resp. maximum) expectation of h depends on the form of the function h. If this form follows some typical cases, efficient solutions can be found to compute lower (resp. upper) expectations. The simplest examples (for which solutions are well known) of such typical cases are monotone functions.

3. The simple case of monotone functions

We first consider the case where h is a monotone function that is non-decreasing (resp. non-increasing) in $\mathbb{R}$. We will also introduce the running example used throughout the paper.

3.1. Unconditional expectations. In the case of a monotone non-decreasing (resp. non-increasing) function, existing results [31] tell us that we have:

$$(10) \qquad \underline{\mathbb{E}}(h) = \int_{\mathbb{R}} h(x) \mathrm{d}\overline{F} \ \left(\underline{\mathbb{E}}(h) = \int_{\mathbb{R}} h(x) \mathrm{d}\underline{F} \right),$$

$$(11) \qquad \overline{\mathbb{E}}(h) = \int_{\mathbb{R}} h(x) \mathrm{d}\underline{F} \ \left(\overline{\mathbb{E}}(h) = \int_{\mathbb{R}} h(x) \mathrm{d}\overline{F} \right),$$

and we see from (10)-(11) that lower and upper expectations are completely determined by bounding distributions $\underline{F}$ and $\overline{F}$. Using equations (5)-(6), we get the following formulas

$$(12) \qquad \underline{\mathbb{E}}(h) = \int_0^1 h(a_{*\gamma}) d\gamma \ \left(\underline{\mathbb{E}}(h) = \int_0^1 h(a_\gamma^*) d\gamma \right),$$

$$(13) \qquad \overline{\mathbb{E}}(h) = \int_0^1 h(a_\gamma^*) d\gamma \ \left(\overline{\mathbb{E}}(h) = \int_0^1 h(a_{*\gamma}) d\gamma \right),$$

which are the counterparts of equations (10)-(11). Here, expectations are totally determined by extreme values of the mappings. When h is non-monotone, equations (10)-(13) only provide inner approximations of $\underline{\mathbb{E}}(h), \overline{\mathbb{E}}(h)$. When using numerical procedures over monotone functions, there appears to be no specific sampling strategies of values that would allow for faster convergence.

We now introduce the example that will illustrate our results all along the paper.

Example 1. *Assume that we have to estimate the loss incurred by the failure of a unit of some industrial item. Suppose that this loss is the function of time $h(x) = 20 - x$, and it is known that the unit time to failure is governed by a distribution whose bounds are exponential distributions with a failure rate 0.2 and 0.5 (note that only the bounds are of exponential nature). h is decreasing and can, for example, model the fact that the later the unit fails, the less it costs to replace it. Let us compute the expected losses as the expectation of h. The lower and upper distribution functions of the unit time to failure are $1 - \exp(-0.2x)$ and $1 - \exp(-0.5x)$, respectively. Hence*

$$\overline{\mathbb{E}}(h) = \int_0^\infty (20 - x)\mathrm{d}(1 - \exp(-0.5x)) = \int_0^\infty (20 - x)0.5e^{-0.5x}\mathrm{d}x = 18,$$

$$\underline{\mathbb{E}}(h) = \int_0^\infty (20 - x)\mathrm{d}(1 - \exp(-0.2x)) = \int_0^\infty (20 - x)0.2e^{-0.2x}\mathrm{d}x = 15.$$

Finally, we obtain that the expected losses are in the interval $[15, 18]$.

Let us use the random set approach. Since $\overline{F}^{-1}(\gamma) = -2\ln(1 - \gamma) = a_\gamma^*$ and $\underline{F}^{-1}(\gamma) = -5\ln(1 - \gamma) = a_{*\gamma}$, then

$$\overline{\mathbb{E}}(h) = \int_0^1 (20 + 2\ln(1 - \gamma))\mathrm{d}\gamma = 18,$$

$$\underline{\mathbb{E}}(h) = \int_0^1 (20 + 5\ln(1 - \gamma))\mathrm{d}\gamma = 15.$$

We get the same values of the lower and upper expectations of h.

3.2. **Conditional expectations.** We now consider that we want to know the lower and upper expectations in the case where event $B = [b_0, b_1]$ occurs. That is, we want to compute Equations (7), (8) for a monotone h. Lower and upper expectations are then given by the following proposition.

Proposition 2. *Given a p-box $[\underline{F}, \overline{F}]$, a monotone function $h(x)$ and an event B, the upper and lower conditional expectation of $h(X)$ on $[\underline{F}, \overline{F}]$ after observing the event B can be written*

$$\overline{\mathbb{E}}(h|B) = \sup_{\substack{\underline{F}(b_0) \leq \alpha \leq \overline{F}(b_0) \\ \underline{F}(b_1) \leq \beta \leq \overline{F}(b_1)}} \frac{1}{\beta - \alpha} \int_\alpha^\beta \sup_{x \in A_\gamma \cap B} h(x)\mathrm{d}\gamma$$

$$= \frac{1}{\overline{F}(b_1) - \overline{F}(b_0)} \left(\int_{\underline{F}^{-1}(\overline{F}(b_0))}^{b_1} h(x)\mathrm{d}\underline{F}(x) + h(b_1)\left(\overline{F}(b_1) - \underline{F}(b_1)\right) \right),$$

$$\underline{\mathbb{E}}(h|B) = \inf_{\substack{\underline{F}(b_0) \leq \alpha \leq \overline{F}(b_0) \\ \underline{F}(b_1) \leq \beta \leq \overline{F}(b_1)}} \frac{1}{\beta - \alpha} \int_\alpha^\beta \inf_{x \in A_\gamma \cap B} h(x)\mathrm{d}\gamma$$

$$= \frac{1}{\underline{F}(b_1) - \underline{F}(b_0)} \left(h(b_0)\left(\overline{F}(b_0) - \underline{F}(b_0)\right) + \int_{b_0}^{\overline{F}^{-1}(\underline{F}(b_1))} h(x)\mathrm{d}\overline{F}(x) \right)$$

if h is non-decreasing and

$$\overline{\mathbb{E}}(h|B) = \frac{1}{\underline{F}(b_1) - \underline{F}(b_0)} \left(h(b_0) \left(\overline{F}(b_0) - \underline{F}(b_0) \right) + \int_{b_0}^{\overline{F}^{-1}(\underline{F}(b_1))} h(x) \mathrm{d}\overline{F}(x) \right),$$

$$\underline{\mathbb{E}}(h|B) = \frac{1}{\overline{F}(b_1) - \overline{F}(b_0)} \left(\int_{\underline{F}^{-1}(\overline{F}(b_0))}^{b_1} h(x) \mathrm{d}\underline{F}(x) + h(b_1) \left(\overline{F}(b_1) - \underline{F}(b_1) \right) \right),$$

if h is non-increasing.

Proof. We will only prove the upper expectation for non-decreasing function h. Lower expectation can be derived likewise, and the case of non-increasing functions is then obtained by using duality between lower and upper expectations.

When h is non-decreasing, we know that $\sup_{x \in A_\gamma \cap B} h(x)$ is a non-decreasing function of γ that coincides with $\underline{F}^{-1}$. Using the integral mean value theorem, we know that there exists some $z \in [b_0, b_1]$ such that $\overline{\mathbb{E}}(h|B) = h(z)$, whatever the choice of α, β. For maximizing $\overline{\mathbb{E}}(h|B)$, values α, β should be chosen so that the retained values z and $h(z)$ (coinciding with $\underline{F}^{-1}$) are as high as possible. As h is non-decreasing, this corresponds to values $\alpha = \overline{F}(b_0)$, $\beta = \overline{F}(b_1)$, which settles the denominator of the objective function. We then have

$$\int_\alpha^\beta \sup_{x \in A_\gamma \cap B} h(x) \mathrm{d}\gamma = \int_{\underline{F}^{-1}(\overline{F}(b_0))}^{b_1} h(x) \mathrm{d}\underline{F}(x) + h(b_1) \left(\overline{F}(b_1) - \underline{F}(b_1) \right),$$

because for values $\gamma \in [\overline{F}(b_0), \underline{F}(b_1)]$, supremum of $h(x)$ on $A_\gamma \cap B$ is obtained for $x = \underline{F}^{-1}(\gamma)$, while for $\gamma \in [\underline{F}(b_1), \overline{F}(b_1)]$, supremum of $h(x) = b_1$. $\qquad\square$

FIGURE 2. Conditional expectations with monotone non-increasing functions

Example 2. *We consider the same p-box $[\underline{F}, \overline{F}]$ and function h as in Example 1, but now we consider that we want to know the incurred loss in case $x \in B = [1, 8]$, that is the failure is supposed to happen between 1 and 8 units of time. We have*

$$\underline{F}(b_0) = 1 - \exp(-0.2 \cdot 1) = 0.18, \quad \overline{F}(b_0) = 1 - \exp(-0.5 \cdot 1) = 0.39,$$

$$\underline{F}(b_1) = 1 - \exp(-0.2 \cdot 8) = 0.8, \quad \overline{F}(b_1) = 1 - \exp(-0.5 \cdot 8) = 0.98,$$

and we get

$$\overline{\mathbb{E}}(h|B) = \frac{1}{0.8 - 0.18}\left((20 - 1)(0.39 - 0.18) + \int_1^{\overline{F}^{-1}(0.8)}(20 - x)0.5e^{-0.5x}\mathrm{d}x\right)$$

$$= 18.298,$$

$$\underline{\mathbb{E}}(h|B) = \frac{1}{0.98 - 0.39}\left((20 - 8)(0.98 - 0.8) + \int_{\underline{F}^{-1}(0.39)}^8(20 - x)0.2e^{-0.2x}\mathrm{d}x\right)$$

$$= 14.219.$$

Note that, if we compare above values with those of Example 1, we have $[\underline{\mathbb{E}}(h), \overline{\mathbb{E}}(h)] \subset [\underline{\mathbb{E}}(h|B), \overline{\mathbb{E}}(h|B)]$.

The above results indicate that, when h is monotone, computing lower/upper expectations exactly remains easy. Also, when using numerical methods, they provide insight as to how values should be sampled. For example, when computing upper conditional expectation by linear programming, values only need to be sampled in $[b_0, \overline{F}^{-1}(b_1)]$, and b_0 should be among the sampled values, since an important probability mass is concentrated at this value (see Fig. 2). When using random set approach and discretizing the unit interval $[0, 1]$, one should take $\gamma_1 = \underline{F}b_0$ and $\gamma_2 = \overline{F}(b_0)$, and not consider finer discretization of this interval, as this would not increase the precision. As we shall see, similar results can be derived for more complex cases.

4. FUNCTION WITH ONE MAXIMUM

In this section, we study the case where the function h has one maximum at point a, i.e. h is increasing (resp. decreasing) in $(-\infty, a]$ (resp. $[a, \infty)$). The case of h having one minimum follows by considering the function $-h$ and the duality between lower and upper expectations.

4.1. Unconditional expectations. As for monotone h, we first study the case of unconditional expectations. Before giving the main result, we show the next lemma that will be useful in subsequent proofs.

Lemma 1. *Given a p-box* $[\underline{F}, \overline{F}]$ *and a continuous function* $h(x)$ *with one maximum at* $x = a$, *there is always a solution* $\gamma \in [\underline{F}(a), \overline{F}(a)]$ *to the following equation*

$$(14) \qquad\qquad h\left(\overline{F}^{-1}(\gamma)\right) = h\left(\underline{F}^{-1}(\gamma)\right).$$

Proof. let us consider the function

$$\varphi(\alpha) = h\left(\overline{F}^{-1}(\alpha)\right) - h\left(\underline{F}^{-1}(\alpha)\right),$$

which, being a substraction of two continuous functions (by supposition), is continuous. Since the function h has its maximum at point $x = a$, then, by taking $\alpha = \underline{F}(a)$, we get the inequality

$$\varphi(\gamma) = h\left(\overline{F}^{-1}(\underline{F}(a))\right) - h(a) \le 0$$

and, by taking $\gamma = \overline{F}(a)$, we get the inequality

$$\varphi(\gamma) = h(a) - h\left(\underline{F}^{-1}(\overline{F}(a))\right) \ge 0.$$

Optimal F for $\overline{\mathbb{E}}(h)$ Optimal F for $\underline{\mathbb{E}}(h)$

FIGURE 3. Optimal distributions F with unimodal h

Consequently, there exists γ in the interval $\left(\underline{F}(a), \overline{F}(a)\right)$ such that $\varphi(\gamma) = 0$ (since φ is continuous). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The next proposition shows that, as for monotone h, the fact of knowing that h has one maximum in $x = a$ allows us to derive closed-form expressions of lower and upper expectations. The results of the proposition are illustrated in Fig. 3.

Proposition 3. *If the function h has one maximum at point $a \in \mathbb{R}$, then the upper and lower expectations of $h(X)$ on $[\underline{F}, \overline{F}]$ are*

$$(15) \qquad \overline{\mathbb{E}}(h) = \int_{-\infty}^{a} h(x)\mathrm{d}\underline{F} + h(a)\left[\overline{F}(a) - \underline{F}(a)\right] + \int_{a}^{\infty} h(x)\mathrm{d}\overline{F},$$

$$(16) \qquad \underline{\mathbb{E}}(h) = \left[\int_{-\infty}^{\overline{F}^{-1}(\alpha)} h(x)\mathrm{d}\overline{F} + \int_{\underline{F}^{-1}(\alpha)}^{\infty} h(x)\mathrm{d}\underline{F}\right],$$

or, equivalently

$$(17) \qquad \overline{\mathbb{E}}(h) = \int_{0}^{\underline{F}(a)} h(a_\gamma^*)d\gamma + [\overline{F}(a) - \underline{F}(a)]h(a) + \int_{\overline{F}(a)}^{1} h(a_{*\gamma})d\gamma$$

$$(18) \qquad \underline{\mathbb{E}}(h) = \int_{0}^{\alpha} h(a_{*\gamma})d\gamma + \int_{\alpha}^{1} h(a_\gamma^*)d\gamma,$$

where α is the solution of equation

$$(19) \qquad h\left(\overline{F}^{-1}(\alpha)\right) = h\left(\underline{F}^{-1}(\alpha)\right).$$

such that $\alpha \in [\underline{F}(a), \overline{F}(a)]$.

Proof using linear programming. We assume that the function $h(x)$ is differentiable in $\mathbb{R}$ and has a finite value as $x \to \infty$. The lower and upper cumulative probability functions $\underline{F}$ and $\overline{F}$ are also assumed to be differentiable. We also consider the primal and dual problems considered in Section 2.1 and recalled below.

Primal problem:	**Dual problem:**

Min. $\mathbf{v} = \int\limits_{-\infty}^{\infty} h\left(x\right)\rho\left(x\right)\mathrm{d}x$ Max. $\mathbf{w} = c_0 + \int\limits_{-\infty}^{\infty} \left(-c\left(t\right)\overline{F}\left(t\right) + d\left(t\right)\underline{F}\left(t\right)\right)\mathrm{d}t$

subject to   subject to

$$\rho\left(x\right) \geq 0, \int\limits_{-\infty}^{\infty}\rho\left(x\right)\mathrm{d}x = 1, \qquad\qquad c_0 + \int\limits_{x}^{\infty}\left(-c\left(t\right)+d\left(t\right)\right)\mathrm{d}t \leq h\left(x\right),$$

$$-\int\limits_{-\infty}^{x}\rho\left(x\right)\mathrm{d}x \geq -\overline{F}\left(x\right), \qquad\qquad c_0 \in \mathbb{R}, c\left(x\right) \geq 0, d\left(x\right) \geq 0.$$

$$\int\limits_{-\infty}^{x}\rho\left(x\right)\mathrm{d}x \geq \underline{F}\left(x\right).$$

The proof of Equations (15)-(16) and (19) can be separated in three main steps:

(1) We propose a feasible solution of the primal problem.
(2) We then consider the feasible solution of the dual problem corresponding to the one proposed for the primal problem.
(3) We show that the two solutions coincide and, therefore, according to the basic duality theorem of linear programming, these solutions are optimal ones.

First, we consider the primal problem. Let a' and a'' be real values. The function

$$\rho\left(x\right) = \begin{cases} \mathrm{d}\overline{F}\left(x\right)/\mathrm{d}x, & x < a' \\ 0, & a' \leq x \leq a'' \\ \mathrm{d}\underline{F}\left(x\right)/\mathrm{d}x, & a'' < x \end{cases}$$

is a feasible solution to the primal problem if the following conditions are respected:

$$\int_{-\infty}^{\infty}\rho\left(x\right)\mathrm{d}x = 1,$$

which, given the above solution, can be rewritten

$$\int_{-\infty}^{a'}\mathrm{d}\overline{F} + \int_{a''}^{\infty}\mathrm{d}\underline{F} = 1,$$

which is equivalent to the equality

(20) $$\overline{F}\left(a'\right) = \underline{F}\left(a''\right).$$

We now interest ourselves in the dual problem. Let us first consider the sole constraint

(21) $$c_0 + \int_{x}^{\infty}\left(-c\left(t\right)+d\left(t\right)\right)\mathrm{d}t \leq h\left(x\right),$$

which is the equivalent of the primal constraint $\rho\left(x\right) \geq 0$. We then consider the following feasible solution to the dual problem as $c_0 = h\left(\infty\right)$,

$$c\left(x\right) = \begin{cases} h'\left(x\right), & x < a' \\ 0, & x \geq a' \end{cases} \quad d\left(x\right) = \begin{cases} 0, & x < a'' \\ -h'\left(x\right), & x \geq a'' \end{cases}.$$

The inequalities $c\left(x\right) \geq 0$ and $d\left(x\right) \geq 0$ are valid provided we have the inequalities

$a' \leq a \leq a''$ (i.e. interval $[a', a'']$ encompasses maximum of h). By integrating $c(x)$ and $d(x)$, we get the increasing function

$$C(x) = -\int_x^\infty c(t)\,dt = \left\{ \begin{array}{ll} h(x) - h(a'), & x < a' \\ 0, & x \geq a' \end{array} \right.$$

and the decreasing function

$$D(x) = \int_x^\infty d(t)\,dt = \left\{ \begin{array}{ll} h(a'') - h(\infty), & x < a'' \\ h(x) - h(\infty), & x \geq a'' \end{array} \right. .$$

Let us rewrite condition (21) as follows:

$$(22) \qquad c_0 + C(x) + D(x) \leq h(x).$$

If $x < a'$, equation (22) becomes

$$c_0 + h(x) - h(a') + h(a'') - h(\infty) \leq h(x).$$

And, replacing the inequality by an equality (simply taking the upper bound of the constraint), we obtain

$$(23) \qquad h(a'') = h(a').$$

If $a' < x < a''$, we have $c_0 + h(a'') - h(\infty) \leq h(x)$ which means that for all $x \in (a', a'')$ we have $h(a'')\,(= h(a')) \leq h(x)$ (i.e. $h(a'')$ and a' are the minimal values of the function $h(x)$ in interval $x \in (a', a'')$.) If $x \geq a''$, then we get the trivial equality $c_0 + h(x) - h(\infty) = h(x)$. The two proposed solutions are valid iff there exist solutions to Eq. (20) and Eq. (23), respectively for the primal and dual problem. That such solutions exist can be seen by considering Lemma1 and taking $a' = \overline{F}^{-1}(\gamma)$ and $a'' = \underline{F}^{-1}(\gamma)$, with γ the solution of Eq. (19). We then find the admissible values of the objective functions

$$v_{\min} = \int_0^{a'} h(x)\,d\overline{F} + \int_{a''}^\infty h(x)\,d\underline{F},$$

$$w_{\max} = c_0 + \int_0^\infty \left(-c(t)\,\overline{F}(t) + d(t)\,\underline{F}(t) \right) dt.$$

By using integration by parts together with equations (20)-(23), we can show that equality $w_{\max} = v_{\min}$ holds, with γ the particular solution of equation (19) for which optimum is reached, as was to be proved. $\qquad \square$

Proof using random sets. Let us now consider equations (6)-(5). Looking first at equation (6), we see that before $\gamma = \underline{F}(a)$, the supremum of h on A_γ is $h(a_\gamma^*)$, since h is increasing between $[\infty, a]$. Between $\gamma = \underline{F}(a)$ and $\gamma = \overline{F}(a)$, the supremum of h on A_γ is $f(a)$. After $\gamma = \overline{F}(a)$, we can make the same reasoning as for the increasing part of h (except that it is now decreasing). Finally, this gives us the following formula:

$$(24) \qquad \overline{\mathbb{E}}(h) = \int_0^{\underline{F}(a)} h(a_\gamma^*)\,d\gamma + \int_{\underline{F}(a)}^{\overline{F}(a)} h(a)\,d\gamma + \int_{\overline{F}(a)}^1 h(a_{*\gamma})\,d\gamma$$

which is equivalent to (17). Let us now turn to the lower expectation. Before $\gamma = \underline{F}(a)$ and after $\gamma = \overline{F}(a)$, finding the infinimum is again not a problem (it is respectively $h(a_{*\gamma})$ and $h(a_\gamma^*)$). Between $\gamma = \underline{F}(a)$ and $\gamma = \overline{F}(a)$, since we know

that h is increasing before $x = a$ and decreasing after, infinimum is either $h(a_{*\gamma})$ or $h(a_\gamma^*)$. This gives us equation

$$(25) \qquad \underline{\mathbb{E}}h = \int_0^{\underline{F}(a)} h(a_{*\gamma})d\gamma + \int_{\underline{F}(a)}^{\overline{F}(a)} \min(h(a_{*\gamma}), h(a_\gamma^*))d\gamma + \int_{\overline{F}(a)}^1 h(a_\gamma^*)d\gamma$$

and if we use equations (20),(23) as in the first proof (reasoning used in the first proof to show that they have a solution is general, and thus applicable here), we know that there is a level α s.t. $h(\overline{F}^{-1}(\alpha)) = h(\underline{F}^{-1}(\alpha))$, and for which the above equation simplify in equation (19). $\qquad\square$

Figure 3 shows that the extremizing distribution corresponding to upper expectation consists in concentrating as much probability mass as possible on the maximum, as could have been expected, while the cumulative distribution reaching the lower expectation consists of an horizontal jump avoiding higher values. As we shall see, finding the level α satisfying Equation (20) and at which this jump occurs is sometimes feasible, and in this case exact lower and upper expectations can be found. In other cases, when computing the upper expectation by numerical methods and linear programming, results indicate that it is important to include the value a corresponding to the maximum of h in the sampled value, as well as values close to it when computing the upper expectation. When using the random set approach, they show that there are no need to consider values γ inside the interval $[\underline{F}(a), \overline{F}(a)]$, the bounds being sufficient. For the lower expectation, results indicate that when using linear programming, it is preferable to sample outside the interval $[\overline{F}^{-1}(\alpha), \underline{F}^{-1}(\alpha)]$.

However, it can happens that the exact value of α cannot be computed, but that the integrals in Eq. (15)-(16) can still be solved. In this case, lower and upper expectations have to be approximated, for example by scanning a more or less wide range of possible values for α (see [28] for an example).

Example 3. *We still consider the same p-box as in Example 1, but we now suppose that the loss is modelled by the function $h(x) = 60 - (x - 5)^2$. This loss function can express the idea that it is preferable for the unit to fail when it begins to work or when it has worked for a long time, rather than when it works at full capacity, as the cost of slowing a whole production line would then be quite higher. h has one maximum at $a = 5$, and we get*

$$\overline{\mathbb{E}}h = h(5) \left[\overline{F}(5) - \underline{F}(5) \right] + \int_0^5 h(x)\mathrm{d}\underline{F}(x) + \int_5^\infty h(x)\mathrm{d}\overline{F}(x)$$
$$= 60 \cdot (\exp(-0.2 \cdot 5) - \exp(-0.5 \cdot 5)) + 31.321 + 4.268$$
$$= 52.736.$$

Since $\overline{F}^{-1}(\alpha) = -2\ln(1 - \alpha)$ and $\underline{F}^{-1}(\alpha) = -5\ln(1 - \alpha)$, then α can be found by solving the following equality

$$60 - (-2\ln(1 - \alpha) - 5)^2 = 60 - (-5\ln(1 - \alpha) - 5)^2.$$

Hence, we have two solutions $\alpha = 1 - \exp(-10/7)$ and $\alpha = 0$. Since $\overline{F}^{-1}(0) = \underline{F}^{-1}(0)$, then the second solution has to be removed. Therefore, we get $\alpha = 1 -$

$\exp(-10/7) = 0.76$. *Hence, we obtain*

$$\mathbb{E}h = \int_{-\infty}^{-2\ln(1-0.76)} h(x)\mathrm{d}\overline{F}(x) + \int_{-5\ln(1-0.76)}^{\infty} h(x)\mathrm{d}\underline{F}(x)$$

$$= \int_{-\infty}^{2.85} \left(60 - (x-5)^2\right) 0.5e^{-0.5x}\mathrm{d}x + \int_{7.14}^{\infty} \left(60 - (x-5)^2\right) 0.2e^{-0.2x}\mathrm{d}x$$

$$= 29.745.$$

Finally, we obtain the interval of expected losses $[29.745, 52.736]$. *Using the random set approach, we get*

$$\overline{\mathbb{E}}(h) = \int_{0}^{1-\exp(-0.5\cdot5)} \left(60 - (-5\ln(1-\gamma)-5)^2\right)\mathrm{d}\gamma + h(5)\left[\overline{F}(5) - \underline{F}(5)\right]$$

$$+ \int_{1-\exp(-0.2\cdot5)}^{1} \left(60 - (-2\ln(1-\gamma)-5)^2\right)\mathrm{d}\gamma$$

$$= 52.736.$$

$$\underline{\mathbb{E}}(h) = \int_{0}^{0.76} \left(60 - (-5\ln(1-\gamma)-5)^2\right)\mathrm{d}\gamma + \int_{0.76}^{1} \left(60 - (-2\ln(1-\gamma)-5)^2\right)\mathrm{d}\gamma$$

$$= 29.745.$$

If the function h is symmetric about a, i.e., the equality $h(a - x) = h(a + x)$ is valid for all $x \in \mathbb{R}$, then the value of α in (19) does not depend on h and is determined as

$$a - \overline{F}^{-1}(\alpha) = \underline{F}^{-1}(\alpha) - a.$$

Note that expressions (10),(11) can be obtained from (15),(16) by taking $a \to \infty$.

4.2. **Conditional expectations.** We now consider conditioning by an event $B = [b_0, b_1]$, while h is still assumed to have one maximum. The following proposition indicates how lower and upper conditional expectations can be computed in this case.

Proposition 4. *If the function h has one maximum at point $a \in \mathbb{R}$, then the upper and lower conditional expectations of $h(X)$ on $[\underline{F}, \overline{F}]$ after observing the event B are*

$$\overline{\mathbb{E}}(h|B) = \sup_{\substack{\underline{F}(b_0)\leq\alpha\leq\overline{F}(b_0) \\ \underline{F}(b_1)\leq\beta\leq\overline{F}(b_1)}} \frac{1}{\beta-\alpha}\Psi(\alpha,\beta),$$

$$\underline{\mathbb{E}}(h|B) = \inf_{\substack{\underline{F}(b_0)\leq\alpha\leq\overline{F}(b_0) \\ \underline{F}(b_1)\leq\beta\leq\overline{F}(b_1)}} \frac{1}{\beta-\alpha}\Phi(\alpha,\beta),$$

with

$$\Psi(\alpha,\beta) = \; I_{(\alpha<\underline{F}^{-1}(a))} \int_{\underline{F}^{-1}(\alpha)}^{a} h(x)\mathrm{d}\underline{F} + I_{(\beta>\overline{F}^{-1}(a))} \int_{a}^{\overline{F}^{-1}(\beta)} h(x)\mathrm{d}\overline{F}$$

$$+ \, h(a)\left(\min(\overline{F}(a),\beta) - \max(\underline{F}(a),\alpha)\right)$$

$$\Phi(\alpha,\beta) = \; h(b_0)\left(\overline{F}(b_0)-\alpha\right) + \int_{b_0}^{\overline{F}^{-1}(\varepsilon)} h(x)\mathrm{d}\overline{F}$$

$$+ \, h(b_1)\left(\beta - \underline{F}(b_1)\right) + \int_{\underline{F}^{-1}(\varepsilon)}^{b_1} h(x)\mathrm{d}\underline{F}$$

Here $I_{(a<b)}$ is the indicator function taking 1 if $a < b$ and 0 if $a \geq b$; ε is one of the roots of the following equation:

(26) $$h\left(\overline{F}^{-1}(\varepsilon)\right) = h\left(\underline{F}^{-1}(\varepsilon)\right).$$

Proof. The proof follows from Proposition 1 where $\Psi(\alpha,\beta), \Phi(\alpha,\beta)$ are respectively replaced by formulas given in Proposition 3. □

Example 4. *We consider the same h as in Example 3, the same p-box $[\underline{F},\overline{F}]$ as in the other examples, and the conditioning event $B = [1,8]$. From Example 3, the solutions of Eq. (26) are $\varepsilon = 1-\exp(-10/7) = 0.76$, $\underline{F}^{-1}(\varepsilon) = 7.14$, $\overline{F}^{-1}(\varepsilon) = 2.85$. We also have $a = 5$, $\underline{F}(a) = 1 - \exp(-0.2 \cdot 5) = 0.63$, $\overline{F}(a) = 1 - \exp(-0.5 \cdot 5) = 0.92$. Let us first concentrate on*

$$\overline{\mathbb{E}}(h|B) = \sup_{\substack{0.18\leq\alpha\leq0.39 \\ 0.8\leq\beta\leq0.98}} \frac{1}{\beta-\alpha}\Psi(\alpha,\beta),$$

where

$$\Psi(\alpha,\beta) = \; I_{(\alpha<0.63)} \int_{-5\ln(1-\alpha)}^{5} \left(60 - (x-5)^2\right) 0.2e^{-0.2x}\mathrm{d}x$$

$$+ \; I_{(\beta>0.92)} \int_{5}^{-2\ln(1-\beta)} \left(60 - (x-5)^2\right) 0.5e^{-0.5x}\mathrm{d}x$$

$$+ \; 60\left(\min(1-e^{-0.5\cdot5},\beta) - \max(1-e^{-0.2\cdot5},\alpha)\right)$$

$$= \; \left(25\alpha\ln^2(1-\alpha) - 25\ln^2(1-\alpha) - 35\alpha + 31.32\right) + 60\left(\min(0.92,\beta) - 0.63\right)$$

$$+ \, I_{(\beta>0.92)}\left(4(1-\beta)\ln^2(1-\beta) + 12(1-\beta)\ln(1-\beta) + 47\beta - 42.73\right)$$

since $0.18 \leq \alpha \leq 0.39$, we have $I_{(\alpha<0.63)} = 1$. Let us then consider the two sets of value $[0.8, 0.92]$ and $(0.92, 0.98]$ for which $I_{(\beta>0.92)}$ takes different values, and the respective functions $\Psi_1(\alpha,\beta), \Psi_2(\alpha,\beta)$ associated to them:

$$\Psi_1(\alpha,\beta) = 25\alpha\ln^2(1-\alpha) - 25\ln^2(1-\alpha) - 35\alpha + 31.32 + 60(\beta - 0.63)$$

$$\Psi_2(\alpha,\beta) = 25\alpha\ln^2(1-\alpha) - 25\ln^2(1-\alpha) - 35\alpha + 31.32$$

$$+ 4(1-\beta)\ln^2(1-\beta) + 12(1-\beta)\ln(1-\beta) + 47\beta - 42.73 + 17.4$$

It can be checked that the derivative $\mathrm{d}\Psi_1(\alpha,\beta)/(\beta-\alpha)/\mathrm{d}\beta$ is positive for $0.18 \leq \alpha \leq 0.39$, hence the maximum of $\Psi_1(\alpha,\beta)/(\beta-\alpha)$ is achieved at $\beta = 0.98$. Also, since

FIGURE 4. Optimal distribution (thick) for computing upper conditional expectation on $B = [1, 8]$

$\Psi_1(\alpha, 0.98)/(0.98 - \alpha)$ *decreases as α increases, we have*

$$\sup \frac{1}{\beta - \alpha} \Psi_1(\alpha, \beta) = \frac{1}{0.98 - 0.18} \Psi_1(0.18, 0.98) = 56.52.$$

A similar analysis for $\Psi_2(\alpha,\beta)/(\beta-\alpha)$ *shows that maximum is achieved for $\alpha = 0.39$, $\beta = 0.8$. Hence*

$$\sup \frac{1}{\beta - \alpha} \Psi_2(\alpha, \beta) = \frac{1}{0.8 - 0.39} \Psi_2(0.39, 0.8) = 59.57.$$

and, finally, we have $\overline{\mathbb{E}}(h|B) = \max(56.52, 59.57) = 59.57$. Figure 4 gives an illustration of the extremizing cumulative distribution for which this upper conditional expectation is reached.

   Let us now detail the computations for

$$\underline{\mathbb{E}}(h|B) = \inf_{\substack{0.18 \leq \alpha \leq 0.39 \\ 0.8 \leq \beta \leq 0.98}} \frac{1}{\beta - \alpha} \Phi(\alpha, \beta),$$

where

$$\Phi(\alpha, \beta) = \left(60 - (1 - 5)^2\right)(0.39 - \alpha) + \int_1^{2.85} \left(60 - (x - 5)^2\right) 0.5 e^{-0.5x} \mathrm{d}x$$

$$+ \left(60 - (8 - 5)^2\right)(\beta - 0.8) + \int_{7.14}^8 \left(60 - (x - 5)^2\right) 0.2 e^{-0.2x} \mathrm{d}x$$

$$= 51\beta - 44\alpha - 3.54.$$

The function $\frac{1}{\beta-\alpha}\Phi(\alpha, \beta)$ increases as α increases by arbitrary $0.8 \leq \beta \leq 0.98$ and increases as β increases. This implies that $\underline{\mathbb{E}}(h|B) = 1/(0.8-0.18)\,(51 \cdot 0.8 - 44 \cdot 0.18 - 3.54) = 47.32$.

   Note that, in the general case, four functions Ψ_i (corresponding to all combinations of values of $I_{(\alpha < \underline{F}^{-1}(a))}$, $I_{(\beta > \overline{F}^{-1}(a))}$ inside $\{0, 1\}^2$) would have to be considered in the computation of $\overline{\mathbb{E}}(h|B)$. Example 4 well illustrates the fact that when h is non-monotone, analytical solutions can still be found in some cases, but that they tend to become tedious to compute. This will be confirmed in the next section.

5. Functions with local maxima/minima

Now we consider a general form of the function h, i.e., the function $h(x)$ has alternate local maxima at point a_i, $i = 1, 2, ...$ and minima at point b_i, $i = 0, 1, 2, ...$, such that

$$(27) \qquad b_0 < a_1 < b_1 \ldots < b_i < a_i < b_{i+1} < \ldots$$

Note that, in this case, studying the shape of the extremizing cumulative distribution reaching lower expectation is sufficient, thanks to the duality between lower and upper expectation.

Proposition 5. *If local maxima (a_i) and minima (b_i) of the function h satisfy condition (27), then the extremizing distribution F for computing the lower unconditional expectation $\underline{\mathbb{E}}(h)$ has discontinuities (vertical jumps) at points b_i, $i = 1, ...$ of the size*

$$\min\left(\overline{F}(b_i), \alpha_{i+1}\right) - \max\left(\underline{F}(b_i), \alpha_i\right).$$

Between points b_{i-1} and b_i, that is between discontinuities numbered $i-1$ and i, the extremizing cumulative probability distribution function F is of the form:

$$F(x) = \begin{cases} \overline{F}(x), & x < a' \\ \alpha, & a' \le x \le a'' \\ \underline{F}(x), & a'' < x \end{cases},$$

where α is the root of the equation

$$h\left(\max\left(\overline{F}^{-1}(\alpha), b_{i-1}\right)\right) = h\left(\min\left(\underline{F}^{-1}(\alpha), b_i\right)\right)$$

in interval $\left[\underline{F}(a_i), \overline{F}(a_i)\right]$, and a',a'' are such that

$$a' = \max\left(\overline{F}^{-1}(\alpha), b_{i-1}\right), \; a'' = \min\left(\underline{F}^{-1}(\alpha), b_i\right).$$

The upper expectation $\overline{\mathbb{E}}(h)$ can be found from the condition $\overline{\mathbb{E}}(h) = -\underline{\mathbb{E}}(-h)$.

Proof using linear programming. This proof is based on the investigation of the following local primal and dual optimization problems for computing the lower expectation of h in finite interval $[b_0, b_1)$ where h has one maximum at point a_1:

Primal problem:	**Dual problem:**
Min. $v = \int_{b_0}^{b_1} h(x) f(x)\mathrm{d}x$	**Max.** $w = -c_0\overline{F}(b_0) + d_0\underline{F}(b_0) - c_1\overline{F}(b_1)$
subject to	$+d_1\underline{F}(b_1) + \int_{b_0}^{b_1}\left(-\overline{F}(x)c(x) + \underline{F}(x)d(x)\right)\mathrm{d}x$
$f(x) \ge 0$, $F_0 \ge 0$, $F_1 \ge 0$,	subject to
$-\int_{b_0}^{x} f(t)\,\mathrm{d}t - F_0 \ge -\overline{F}(x)$,	$e + \int_x^{b_1}\left(-c(t) + d(t)\right)\mathrm{d}t \le h(x)$,
$\int_{b_0}^{x} f(t)\,\mathrm{d}t + F_0 \ge \underline{F}(x)$,	$e - c_0 + d_0 + \int_{b_0}^{b_1}\left(-c(t) + d(t)\right)\mathrm{d}t \le 0$,
$-F_0 \ge -\overline{F}(b_0)$,$F_0 \ge \underline{F}(b_0)$,	$-e - c_1 + d_1 \le 0$,
$-F_1 \ge -\overline{F}(b_1)$,$F_1 \ge \underline{F}(b_1)$,	$c(x) \ge 0$,$c_0 \ge 0$,$c_1 \ge 0$,
$\int_{b_0}^{b_1} f(t)\,\mathrm{d}t + F_0 - F_1 = 0$.	$d(x) \ge 0$,$d_0 \ge 0$,$d_1 \ge 0$,$e \in \mathbb{R}$

The optimal solutions of the above problems correspond to the extremizing distribution for values $x \in [b_0, b_1)$. $F_0 := F(b_0)$ and $F_1 := F(b_1)$ respectively stand for the values of the extremizing F in b_0 and b_1. The proof then follows in two main steps:

FIGURE 5. Four cases of piece-wise extremizing F

(1) Find optimal solution (that is, propose a feasible solution which coincide for both the primal and dual problem) for the above primal and dual problems, and consequently the values of the extremizing F between any two local minima $[b_i, b_{i+1}]$
(2) Show that the combination of these piece-wise extremizing F correspond to a cumulative distribution.

Step (1) of the proof To find optimal solution between $x \in [b_0, b_1]$, we will consider every possible cases. First, we can differentiate between two main cases, depending on the inequality relation between $\overline{F}(b_0)$ and $\underline{F}(b_1)$.

Case 1. $\overline{F}(b_0) > \underline{F}(b_1)$. The optimal solution in this case is of the form: it corresponds to the solution $f(x) = 0$, $F(x) = F_0 = F_1 = \alpha$, where α is an arbitrary number satisfying the condition $\underline{F}(b_1) < \alpha < \overline{F}(b_0)$ for the primal problem and to the solution $c(x) = d(x) = 0$, $c_0 = d_0 = c_1 = d_1 = e = 0$ for the dual problem. See Fig. 5 for an illustration

Case 2. $\overline{F}(b_0) \leq \underline{F}(b_1)$. This case is similar to the one considered in Section 4, since between $[b_0, b_1]$, h has a maximum for $x = a_1$ and is increasing (resp. decreasing) in $[b_0, a_1]$ (resp. $[a_1, b_1)$). We will therefore proceed in the same way as in the proof of Proposition 3 to find the optimal solution. First recall (Lemma 1) that there is a value α which is a root of the function

$$\varphi(\alpha) = h\left(\max\left(\overline{F}^{-1}(\alpha), b_0\right)\right) - h\left(\min\left(\underline{F}^{-1}(\alpha), b_1\right)\right)$$

with $\alpha \in \left[\underline{F}\left(a_1\right), \overline{F}\left(a_1\right)\right]$. Three subcases can now occur, depending whether α is inside $[\overline{F}\left(b_0\right), \underline{F}\left(b_1\right)]$ or is higher/lower than any value in this interval. We now give details about each of these subcases, the reasoning being similar to the one in the proof of Proposition 3. All subcases and associated extremizing distribution are illustrated in Fig. 5

Subcase 2.1. $\overline{F}\left(b_0\right) \leq \alpha \leq \underline{F}\left(b_1\right)$ ($\alpha \in [\overline{F}\left(b_0\right), \underline{F}\left(b_1\right)]$). Let us denote $a' = \overline{F}^{-1}\left(\alpha\right)$, $a'' = \underline{F}^{-1}\left(\alpha\right)$. Then the optimal solution is of the form:

$$f\left(x\right) = \left\{ \begin{array}{cc} \mathrm{d}\overline{F}(x)/\mathrm{d}x, & b_0 < x < a' \\ 0, & a' \leqslant x \leqslant a'' \\ \mathrm{d}\underline{F}\left(x\right)/\mathrm{d}x, & a'' < x < b_1 \end{array} \right.,$$

$$F_0 = \overline{F}\left(b_0\right), \; F_1 = \underline{F}\left(b_1\right).$$

This implies that

$$F\left(x\right) = \int_{b_0}^{x} f\left(t\right)\mathrm{d}t + F_0 = \left\{ \begin{array}{cc} \overline{F}\left(x\right), & b_0 < x < a' \\ \alpha, & a' \leqslant x \leqslant a'' \\ \underline{F}\left(x\right), & a'' < x < b_1 \end{array} \right..$$

Let us now give the corresponding solution to the dual problem, and show that they are equal. According to relations between primal/dual problem, we have that if $a' < x < b_1$, then $c\left(x\right) = 0$, and if $b_0 < x < a''$, then $d\left(x\right) = 0$. It is obvious that $d_0 = c_1 = 0$. Consider the constraint

$$e + \int_{x}^{b_1} \left(-c\left(t\right) + d\left(t\right)\right)\mathrm{d}t \leq h\left(x\right)$$

for different intervals of x.

Let $a'' < x < b_1$. Then there holds

$$e + \int_{x}^{b_1} d\left(t\right)\mathrm{d}t = h\left(x\right).$$

Hence $d\left(x\right) = -h'\left(x\right)$ and $e = h\left(b_1\right)$.

Let $a' \leq x \leq a''$. Then the following inequality

$$e + \int_{a''}^{b_1} d\left(t\right)\mathrm{d}t \leq h\left(x\right)$$

or $h\left(a''\right) \leq h\left(x\right)$ has to be valid. Indeed, the inequality is valid due to the condition $h\left(a'\right) = h\left(a''\right)$.

Let $b_0 < x < a'$. Then

$$e - \int_{x}^{a'} c\left(t\right)\mathrm{d}t + \int_{a''}^{b_1} d\left(t\right)\mathrm{d}t = h\left(x\right)$$

or

$$-\int_{x}^{a'} c\left(t\right)\mathrm{d}t + h\left(a''\right) = h\left(x\right).$$

Hence $c\left(x\right) = h'\left(x\right)$. The equality

$$e - c_0 + d_0 + \int_{b_0}^{b_1} \left(-c\left(t\right) + d\left(t\right)\right)\mathrm{d}t = 0$$

shows that

$$h\left(b_1\right) - c_0 - h\left(a'\right) + h\left(b_0\right) - h\left(b_1\right) + h\left(a''\right) = 0$$

and $c_0 = h(b_0)$. It follows from the equality $-e - c_1 + d_1 = 0$ that there holds $d_1 = e = h(b_1)$. In sum, we have

$$c(x) = \begin{cases} h'(x), & b_0 < x < a' \\ 0, & a' \leqslant x \leqslant b_1 \end{cases},$$

$$d(x) = \begin{cases} 0, & b_0 < x < a'' \\ -h'(x), & a'' \leqslant x \leqslant b_1 \end{cases},$$

$$c_0 = h(b_0), \ d_0 = c_1 = 0, \ d_1 = e = h(b_1).$$

Let us now show that the two obtained solution coincide:

$$z_{\min} = \int_{b_0}^{a'} h(x) \, \mathrm{d}\overline{F}(x) + \int_{a''}^{b_1} h(x) \, \mathrm{d}\underline{F}(x)$$

$$w_{\max} = -\overline{F}(b_0) h(b_0) + \underline{F}(b_1) h(b_1) - \int_{b_0}^{a'} \overline{F}(x) h'(x) \mathrm{d}x - \int_{a''}^{b_1} \underline{F}(x) h'(x) \mathrm{d}x$$

or

$$w_{\max} = -\overline{F}(b_0) h(b_0) + \underline{F}(b_1) h(b_1)$$
$$+ \int_{b_0}^{a'} h(x) \mathrm{d}\overline{F}(x) - \overline{F}(a') h(a') + \overline{F}(b_0) h(b_0)$$
$$+ \int_{a''}^{b_1} h(x) \, \mathrm{d}\underline{F}(x) - \underline{F}(b_1) h(b_1) + \underline{F}(a'') h(a'')$$
$$= z_{\min}.$$

Hence the proposed solution is the optimal one.

Subcase 2.2. $\alpha > \underline{F}(b_1) \ ([\overline{F}(b_0), \underline{F}(b_1)] \leq \alpha)$. Denote $a' = \overline{F}^{-1}(\alpha)$. Then the optimal solution to the initial problem is:

$$f(x) = \begin{cases} \mathrm{d}\overline{F}(x)/\mathrm{d}x, & b_0 < x < a' \\ 0, & a' \leqslant x \leqslant b_1 \end{cases}, \ F_0 = \overline{F}(b_0), \ F_1 = \alpha,$$

$$F(x) = \int_{b_0}^{x} f(t) \, \mathrm{d}t + F_0 = \begin{cases} \overline{F}(x), & b_0 < x < a' \\ \alpha, & a' \leqslant x \leqslant b_1 \end{cases}.$$

The corresponding solution for the dual problem is such that if $a' < x < b_1$, then $c(x) = 0$, and if $b_0 < x < b_1$, then $d(x) = 0$, hence we have $d_0 = c_1 = 0$. Again, consider the constraint

$$e + \int_{x}^{b_1} (-c(t) + d(t)) \, \mathrm{d}t \leq h(x)$$

for different intervals. Let $a' < x < b_1$. Then the condition $e \leq h(x)$ must be valid. Let $b_0 < x < a'$. Then there holds

$$e - \int_{x}^{a'} c(t) \, \mathrm{d}t = h(x).$$

Consequently, there hold the equalities $c(x) = h'(x)$ and $e = h(a')$. Hence the inequality $e = h(a') \leq h(x)$ is valid for the interval $a' < x < b_1$. The equality

$$e - c_0 + d_0 + \int_{b_0}^{b_1} (-c(t) + d(t)) \, \mathrm{d}t = 0$$

shows that $h(a') - c_0 - h(a') + h(b_0) = 0$, and, therefore, $c_0 = h(b_0)$. It follows from the equality $-e - c_1 + d_1 = 0$ that there holds $d_1 = e = h(a')$. In sum, we get

$$c(x) = \begin{cases} h'(x), & b_0 < x < a' \\ 0, & a' \leqslant x \leqslant b_1 \end{cases},$$

$$d(x) = 0, c_0 = h(b_0), \ d_0 = c_1 = 0, \ d_1 = e = h(a').$$

The obtained solutions for the primal and dual problems are such that:

$$z_{\min} = \int_{b_0}^{a'} h(x) \, \mathrm{d}\overline{F}(x),$$

$$w_{\max} = -\overline{F}(b_0) h(b_0) + \overline{F}(a')h(a') - \int_{b_0}^{a'} \overline{F}(x) h'(x) \mathrm{d}x$$

or

$$w_{\max} = -\overline{F}(b_0) h(b_0) + \overline{F}(a')h(a')$$

$$+ \int_{b_0}^{a'} h(x)\mathrm{d}\overline{F}(x) - \overline{F}(a') h(a') + \overline{F}(b_0) h(b_0)$$

$$= z_{\min}.$$

Consequently, this is the optimal solution.

Subcase 2.3. $\alpha < \overline{F}(b_0)$ $(\alpha \leq [\overline{F}(b_0), \underline{F}(b_1)])$. Denote $a'' = \underline{F}^{-1}(\overline{F}(b_0))$. Then the optimal solution to the primal problem is

$$f(x) = \begin{cases} 0, & b_0 \leqslant x \leqslant a'' \\ \mathrm{d}\underline{F}(x)/\mathrm{d}x, & a'' < x < b_1 \end{cases}, \ F_0 = \alpha, \ F_1 = \underline{F}(b_1).$$

$$F(x) = \begin{cases} \alpha, & b_0 \leqslant x \leqslant a'' \\ \underline{F}(x), & a'' < x < b_1 \end{cases}.$$

and the proof is similar to the one of above cases. Optimal shape of F for any interval $[b_i, b_{i+1}]$ can be obtained by replacing b_0 and b_1 by respectively b_i and b_{i+1} in the above proofs, as they are general (as pictured on Fig. 5). All is left to prove is that the concatenated F obtained by the piece-wise extremizing solutions is increasing (i.e., that F_i for $[b_{i-1}, b_i]$ is lower or equal than F_i for $[b_i, b_{i+1}]$).

Step (2) of the proof Now we show that the joint extremizing distribution function is increasing. Without loss of generality we consider only two intervals $[b_0, b_1]$ and $[b_1, b_2]$. The maximal value of the function $F(x)$ in the interval $[b_0, b_1]$ is $\max(\overline{F}(b_0), \underline{F}(b_1))$ for all the cases. The minimal value of the function $F(x)$ in the interval $[b_1, b_2]$ is $\min(\overline{F}(b_1), \underline{F}(b_2))$ for all the cases.

If $\underline{F}(b_2) \geq \overline{F}(b_0)$, then

$$\min(\overline{F}(b_1), \underline{F}(b_2)) \geq \max(\overline{F}(b_0), \underline{F}(b_1)).$$

This means that the function is increasing.

If $\underline{F}(b_2) < \overline{F}(b_0)$, then $\underline{F}(b_1) < \overline{F}(b_0)$ and we can take $F(x) = \underline{F}(b_1)$ for the left interval. On the other hand, $\underline{F}(b_2) < \overline{F}(b_1)$ and we can take $F(x) = \overline{F}(b_1)$ for the left interval. It follows from the condition $\underline{F}(b_1) < \overline{F}(b_1)$ that the function $F(x)$ is increasing in two neighbour intervals.

Figure 6 gives an example of a general extremizing distribution. $\qquad\qquad \square$

Proof using random sets. For convenience, we will consider that h begins with a local minimum and ends with a local maximum a_n. Formulas when h begins (resp. ends) with a local maximum (resp. minimum) are similar. Lower/upper expectations can be computed as follows:

$$\underline{\mathbb{E}}(h) = \int_0^{\underline{F}(b_n)} \min_{b_i \in A_\gamma} (h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) d\gamma + \int_{\underline{F}(b_n)}^1 h(a_{*\gamma}) d\gamma,$$

$$\overline{\mathbb{E}}(h) = \int_0^{\underline{F}(a_1)} h(a_\gamma^*) d\gamma + \int_{\underline{F}(a_1)}^{\overline{F}(a_n)} \max_{a_i \in A_\gamma} (h(a_{*\gamma}), h(a_i), h(a_\gamma^*)) d\gamma.$$

We concentrate on the formula giving the lower expectation (details for upper one are similar). The most interesting part is the first integral. We consider a particular level γ. Let $B = \{b_i, \ldots, b_j\}$ ($i \le j$) be the set of local minima included in the set A_γ (B can be empty). b_{i-1} and b_{j+1} are the closest local minima outside A_γ. We then consider the minimal $\Delta\gamma := \gamma + \delta\gamma$ such that $\min_{b_i \in A_\gamma}(h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) \ne \min_{b_i \in A_{\Delta\gamma}}(h(a_{*,\Delta\gamma}), h(b_i), h(a_{\Delta\gamma}^*))$ with $\min_{x \in A_{\Delta\gamma}} h(x) \ne h(a_{*,\Delta\gamma})$ if $\min_{x \in A_\gamma} h(x) = h(a_{*,\gamma})$ and $\min_{x \in A_{\Delta\gamma}} h(x) \ne h(a_{\Delta\gamma}^*)$ if $\min_{x \in A_\gamma} h(x) = h(a_\gamma^*)$. As in LP proof, four different cases can occur:

Case A: we have

$$\min_{b_i \in A_\gamma} (h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) = h(b_k)$$

and

$$\min_{b_i \in A_{\Delta\gamma}} (h(a_{*,\Delta\gamma}), h(b_i), h(a_{\Delta\gamma}^*)) = h(b_{k'}),$$

with $k \ne k'$ and where $h(b_k)$ and $h(b_{k'})$ are respectively the lowest local minima of $h(x)$ for $x \in A_\gamma$ and $x \in A_{\Delta\gamma}$. That is, probability mass is concentrated on b_k from γ to $\Delta\gamma$, and concentrates on $b_{k'}$ for values $\gamma' \ge \Delta\gamma$. This correspond to Case 1. of Fig. 5 and of the previous proof. In Fig. 6, it corresponds to the extremizing distribution between b_2 and b_3.

Case B: we have

$$\min_{b_i \in A_\gamma} (h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) = h(a_{*\gamma})$$

and

$$\min_{b_i \in A_{\Delta\gamma}} (h(a_{*,\Delta\gamma}), h(b_i), h(a_{\Delta\gamma}^*)) = h(a_{\Delta\gamma}^*).$$

This can happen when any local minimum inside $A_\gamma, A_{\Delta\gamma}$ is higher than local minima just outside it. In this case, it can happen that minimal values stand at the bounds of intervals $A_{\gamma'}$ for any $\gamma \le \gamma' \le \Delta\gamma$. This corresponds to Case 2.1. of Fig. 5 and of the previous proof. In Fig. 6, it corresponds to the extremizing distribution between b_4 and b_5.

Case C: we have

$$\min_{b_i \in A_\gamma} (h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) = h(b_k)$$

and

$$\min_{b_i \in A_{\Delta\gamma}} (h(a_{*,\Delta\gamma}), h(b_i), h(a_{\Delta\gamma}^*)) = h(a_{\Delta\gamma}^*).$$

FIGURE 6. Example of Optimal F with general h

With $h(b_k)$ the lowest local minima for $b_k \in A_\gamma$. The minimum shift from the left bound of A_γ (coinciding with $\overline{F}$) to b_k. This corresponds to Case 2.2. of Fig. 5 and of the previous proof. In Fig. 6, it corresponds to the extremizing distribution between b_1 and b_2.

Case D: we have

$$\min_{b_i \in A_\gamma} (h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) = ha_{*\gamma})$$

and

$$\min_{b_i \in A_{\Delta\gamma}} (h(a_{*,\Delta\gamma}), h(b_i), h(a_{\Delta\gamma}^*)) = h(b_{k'}).$$

With $h(b_{k'})$ the lowest local minima for $b_{k'} \in A_{\Delta\gamma}$. Situation is similar to the previous case, and corresponds to Case 2.3. of Fig. 5 and of the previous proof. In Fig. 6, it corresponds to the extremizing distribution between b_3 and b_4.

When $\min_{b_i \in A_\gamma}(h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) = \min_{b_i \in A_{\Delta\gamma}}(h(a_{*\gamma}), h(b_i), h(a_\gamma^*)) = h(b_k)$ with $b_k \in A_\gamma \cap A_{\Delta\gamma}$, probability mass stay concentrated on b_k, and this corresponds to a discontinuity mentioned in Proposition 5. By letting γ evolve from 0 to 1, we get the extremizing cumulative distribution of Proposition 5.   □

Looking at the extremizing distribution F pictured in Figure 6, we can see that computing the lower expectation consists in concentrating probability masses over local minima, while giving the less possible amount of probability mass to higher values of $h(x)$, as in the case of a function having one maximum. Thus, our results confirm what could have intuitively be guessed at first sight. They also give analytical and computational tools to compute lower and upper expectations. They are illustrated in the next example.

Example 5. *We consider the same p-box $[\underline{F}, \overline{F}]$ as in the previous examples (see Example 1). However, we assume that the loss function is of the type $h(x) = (0.6x)\cos(x)$. It could, for instance, model the return of a game based on the movement of a pendulum. It could also model the loss incurred by a unit failure whose functioning alternate between low and full capacity (failure during low capacity periods costing less). As a loss after failure has to be positive, one can consider $h(x)+\mu$, with μ a positive constant[2]. $h(x)$ is oscillating between local maxima and minima.*

_____

[2]This does not change further calculations, as $\underline{\mathbb{E}}(h + \mu) = \underline{\mathbb{E}}(h) + \mu$.

These extrema are solutions of $\cos(x) = x \sin(x)$:

$$a_1 = 0.860, \ b_1 = 3.426, \ a_2 = 6.437, \ b_2 = 9.529, \ a_3 = 12.645,$$

$$b_3 = 15.771, \ a_4 = 18.902, \ b_4 = 22.036, \ a_5 = 25.172, \ b_5 = 28.31.$$

We will compute the extremizing distribution for each intervals $[b_i, b_{i+1})$ *for* $i = 1, \ldots, 5$, *with* $b_0 = 0$. *Let us analyze the first interval* $[0, b_1)$. *The value* $\alpha \in (0, 1)$ *in this interval can be found as a root of the equation*

$$(\max(-2\ln(1-\alpha), 0)) \cdot \cos(\max(-2\ln(1-\alpha), 0))$$
$$= (\min(-5\ln(1-\alpha), 3.426)) \cdot \cos(\min(-5\ln(1-\alpha), 3.426)).$$

However, many different values of $\alpha \in (0, 1)$ *are solutions to the above equations. Relying on the proof of Proposition 5 and on the various subcases exposed therein (see Fig. 5), we should, for a given interval* $[b_i, b_{i+1})$, *take only root(s) which provides the interval* $[a', a'']$ *such that* $a_i \in [a', a'']$. *For* $[0, b_1)$, *this corresponds to* $\alpha = 0.215$, *for which values* a', a'' *are*

$$a' = \max(-2\ln(1-\alpha), b_{i-1}) = \max(-2\ln(1-0.215), 0) = 0.483,$$

$$a'' = \min(-5\ln(1-\alpha), b_i) = \min(-5\ln(1-0.215), 3.426) = 1.209.$$

It can be seen from the above that $a_1 = 0.860 \in [0.483, 1.209]$. *We can now determine the extremizing distribution function in* $[0, b_1)$, *which is as follows:*

$$F(x) = \begin{cases} 1 - \exp(-0.5 \cdot x), & x < 0.483 \\ 0.215, & 0.483 \le x \le 1.209 \\ 1 - \exp(-0.2 \cdot x), & 1.209 < x < 3.426 \end{cases}.$$

This corresponds to the case 2.1. of Figure 5. the "jump" (i.e., probability mass) at point b_1 *is of the size*

$$\min(1 - \exp(-0.5 \cdot 3.426), 0.808) - \max(1 - \exp(-0.2 \cdot 3.426), 0.215) = 0.312.$$

Since $\overline{F}(3.426) - \underline{F}(3.426) = 0.33 > 0.312$, *this means that the extremizing distribution in* $[b_1, b_2)$ *starts with a constant value* $F(b_1) = \underline{F}(3.426) + 0.312 = 0.808$ *and with an horizontal line. Moreover, we can check that 0.808 is the right starting point since it is a root of the equation*

$$\max(-2\ln(1-\alpha), 3.426) \cdot \cos(\max(-2\ln(1-\alpha), 3.426))$$
$$= \min(-5\ln(1-\alpha), 9.529) \cdot \cos(\min(-5\ln(1-\alpha), 9.529)).$$

And we have $a' = 3.426$ *and* $a'' = 8.263$ *for* $\alpha = 0.808$. *By taking into account the analysis of the first interval, we can write*

$$F(x) = \begin{cases} 0.808, & 3.426 \le x \le 8.263 \\ 1 - \exp(-0.2 \cdot x), & 8.263 < x < 9.529 \end{cases}.$$

This correspond to case 2.3. of Figure 5. the jump at b_2 *has value* 9.77×10^{-2}, *and we have again* $\overline{F}(9.529) - \underline{F}(9.529) = 0.14 > 9.77 \times 10^{-2}$. *Analysis for other intervals are similar (they all belong to case 2.3.). For the third interval* $[b_2, b_3)$, $\alpha = 0.948$, $a' = 9.529$, $a'' = 14.831$ *and we have*

$$F(x) = \begin{cases} 0.949, & 9.529 \le x \le 14.831 \\ 1 - \exp(-0.2 \cdot x), & 14.831 < x < 15.771 \end{cases}.$$

The jump at b_3 is of value 2.867×10^{-2}, and for $[b_3, b_4)$, we have $\alpha = 0.986$, $a' = 15.771$, $a'' = 21.255$ and

$$F\left(x\right) = \begin{cases} 0.986, & 15.771 \leq x \leq 21.255 \\ 1 - \exp(-0.2 \cdot x), & 21.255 < x < 22.036 \end{cases} .$$

The jump at b_4 is of value 8.189×10^{-3}, and for $[b_4, b_5)$, we have $\alpha = 0.996$, $a' = 22.036$, $a'' = 27.62$ and

$$F\left(x\right) = \begin{cases} 0.996, & 22.036 \leq x \leq 27.62 \\ 1 - \exp(-0.2 \cdot x), & 27.62 < x < 28.31 \end{cases} .$$

The jump at point b_5 is of the size 3.076×10^{-3}.

Note that jump sizes decrease as index i increase. This is not true in general, and is here due to the particular shape of $h(x)$. By computing the extremizing distribution for every interval $[b_{i-1}, b_i)$, we can reach the lower expectation. That is, if we note $\underline{\mathbb{E}}_i(h)$ the lower expectation of h computed with the extremizing distribution obtained for i intervals $[b_{j-1}, b_j), j = 1, \ldots, i$, and if h have a finite number of local maxima and minima, say r, then $\underline{\mathbb{E}}(h) = \underline{\mathbb{E}}_r(h)$. However, in this example, $r = \infty$ and $\underline{\mathbb{E}}(h) = \lim_{r \to \infty} \underline{\mathbb{E}}_r(h)$. Therefore, only an approximate solution can be found[3]. We can therefore let r increase until $\left| \underline{\mathbb{E}}_r(h) - \underline{\mathbb{E}}_{r-1}(h) \right| \leq \varepsilon$, with $\varepsilon > 0$ a prescribed precision. For instance, we have

$$\begin{aligned} \underline{\mathbb{E}}_1(h) &= \int_0^{0.483} 0.6x \cos(x) \cdot 0.5 e^{-0.5x} \mathrm{d}x \\ &+ \int_{1.209}^{3.426} 0.6x \cos(x) \cdot 0.2 e^{-0.2x} \mathrm{d}x \\ &+ 0.6 \cdot 3.426 \cos(3.426) \cdot 0.312 \\ &= -0.82. \end{aligned}$$

Pursuing the computations, we have

$$\underline{\mathbb{E}}_2(h) = -1.558, \quad \underline{\mathbb{E}}_3(h) = -1.9, \quad \underline{\mathbb{E}}_4(h) = -2.033, \quad \underline{\mathbb{E}}_5(h) = -2.093.$$

If we take $\varepsilon = 0.1$, then $|\underline{\mathbb{E}}_5(h) - \underline{\mathbb{E}}_4(h)| = 0.06 < 0.1$, and we consider $\underline{\mathbb{E}}_5(h) = -2.093$ as a sufficient approximation of the true (but unknown) lower approximation. Upper expectation of h can be obtained by considering the function $-h(x)$ and by computing $\underline{\mathbb{E}}(-h)$. Hence $\overline{\mathbb{E}}(h) = -\underline{\mathbb{E}}(-h) = 1.94$ (approximation with $\varepsilon = 0.1$).

This example is useful in two respects: first, it illustrates why it is useful to have results concerning the piece-wise extremizing distribution; second, it shows that even when analytical calculations are possible, it is not always possible to compute an exact value, hence the interest of the generic methods proposed in Section 2. This is particularly true when h has an infinity of local extrema and when $\overline{F}, \underline{F}$ have infinite support. It also addresses the question of the choice of levels α when many solutions are possible.

Coming back to numerical approximations using linear programming, our results indicates that some regions should be sampled in priority. For example, when computing lower expectations, one should primarily consider values b_i (local minima) and sample in neighbourhoods of these values, as it is where probability masses are concentrated. The converse (sampling around local maxima) holds when computing upper expectations.

---

[3]We assume here that the expectation $\underline{\mathbb{E}}(h)$ exists.

If we now consider random set, we can formulate the problem of computing lower expectations as follows: let m be the number of local minima, and let $\gamma_{j_*}, \gamma_{j^*}$ be the two values bounding the probability mass concentrated on local minima b_j, for $j = 1, \ldots, m$ (for example, for the local minima b_2 in Figure 6, we would have $\gamma_{2_*} = \alpha_1, \gamma_{2^*} = \alpha_2$), then

$$(28) \qquad \underline{\mathbb{E}}(h) = \sum_{j=1}^{m} \left(\int_{\gamma_{(j-1)^*}}^{\gamma_{j_*}} \min(h(a_{*\gamma}), h(a_{\gamma}^*)) d\gamma + (\gamma_{(j)^*} - \gamma_{j_*}) h(b_j) \right).$$

This comes down to sum all the probability masses concentrated on local minima, and to calculate integrals when the extremizing distribution coincide either with $\overline{F}$ or $\underline{F}$. Note that, as in Example 5, m could be equal to ∞. This formulation clearly shows that, when using numerical methods with the random set approach, there is no need to discretize in finer intervals the intervals $[\gamma_{j_*}, \gamma_{(j)^*}]$, as it won't improve the precision of the result.

The case of conditional expectation with general function will not be treated here, as it would require long development that wouldn't bring many new ideas.

6. Conclusions

We have considered the problem of computing lower and upper expectations on p-boxes and particular functions under two different approaches: by using linear programming and by using the fact that p-boxes are special cases of random sets. Although the two approaches try to solve equivalent problems, their differences suggest different ways to approximate the solutions of those problems. As we have seen, knowing the behaviour of the function over which lower and upper expectations are to be estimated can greatly increase the computational efficiency (and even permit analytical computation).

However, more important than their differences is the complementarity of both approaches. Indeed, one approach can shed light on some problems obscured by the other approach (e.g., the level α of proposition 3). Another advantage of combining both approaches is the ease with which some problems are solved and the elegant formulation resulting from this combination (e.g., the conditional case). Let us nevertheless note that the constraint programming approach can be applied to imprecise probabilities in general, while the random set approach is indeed limited to random sets.

In this paper, we have concentrated on the case where uncertainty bears on one variable. The case where multiple variables are tainted with uncertainty described by p-boxes will be studied in a forthcoming paper. Concerning future work related to this topic, three lines of research seem interesting to us:

- study of other simple representations : it is desirable to achieve similar studies for other simple uncertainty representations involving sets of probabilities. This includes probability intervals [6], possibility distributions [10], clouds [20].
- Discretization schemes : when exact solutions cannot be computed, what is the best choice of points $x_1, \ldots, x_N$ or of levels $\gamma_1, \ldots, \gamma_M$, respectively to approximate the solution by using LP or RS (already mentioned by other authors [23]). We have mentioned how our results can possibly help in this

task, but proposing generic algorithms and empirically testing them largely remains to be done.

- Convex mixture of functions : in some applications, one can choose a strategy that is a convex mixture between a finite set of options having utility $h_1, \ldots, h_N$. For such cases, one often has to find the weights $\lambda_1, \ldots, \lambda_N$ such that $\sum_{i=1,N} \lambda_i h_i$ have the maximal lower expectation. It would be interesting to study whether similar results as the ones exposed in this paper also exists for this problem when using simple uncertainty representations (e.g., p-boxes).

We would like to end this paper with two final remarks:

- it is clear from our results that extreme distributions over which the upper and lower expectations will be reached will be, in general, discontinuous. Since any discontinuous functions can be approximated as close as one wants by continuous ones, we do not see it as a big flaw. However, in some cases, it could be desirable to add constraints about which cumulative distributions inside $[\underline{F}, \overline{F}]$ are admissible. This kind of questions is adressed, for example, by Kozine and Krymsky [15].
- We mention at the beginning of the paper that our study is restricted to the case where either cumulative distributions were assumed to be σ-additive or where h was continuous. Again, this is not a big limitation when dealing with practical applications, and this avoids many mathematical subtleties arising with the consideration of finitely additive probabilities [19].

References

[1] Diego A. Alvarez. On the calculation of the bounds of probability of events using infinite random sets. *I. J. of Approximate Reasoning*, 43:241–267, 2006.

[2] S. Basu and A. DasGupta. Robust bayesian analysis with distribution bands. *Statistics and Decision*, 13:333–349, 1995.

[3] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.

[4] F.G. Cozman. Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. *Journal of Computational and Graphical Statistics*, 8:824–838, 1999.

[5] F.G. Cozman. Algorithms for conditioning on events of zero lower probability. In *Proceedings of the Fifteenth International FLAIRS Conference*, pages 248–252, 2002.

[6] L.M. de Campos, J.F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.

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

[8] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: I. generalized p-boxes. *Int. J. of Approximate Reasoning*, (In press).

[9] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: Ii. clouds. *Int. J. of Approximate Reasoning*, (In press).

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

[11] D. Dubois and H. Prade. Evidence, knowledge, and belief functions. *Int. J. of Approximate Reasoning*, 6:295–319, 1992.

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

[13] S. Ferson, L. Ginzburg, V. Kreinovich, D.M. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.

[14] P.J. Huber. *Robust statistics*. Wiley, New York, 1981.

[15] I. Kozine and V. Krymsky. Enhancement of natural extension. In *Proc. 5th Int. Symp. on Imprecise Probabilities: Theories and Applications*, 2007.

[16] E. Kriegler and H. Held. Utilizing random sets for the estimation of future climate change. *I. J. of Approximate Reasoning*, 39:185–209, 2005.

[17] I. Levi. *The Enterprise of Knowledge*. MIT Press, London, 1980.

[18] E. Miranda. A survey of the theory of coherent lower previsions. *Int. J. of Approximate Reasoning*, In press, 2008.

[19] E. Miranda, G. de Cooman, and E. Quaeghebeur. Finitely additive extensions of distribution functions and moment sequences: the coherent lower prevision approach. *International Journal of Approximate Reasoning*, 2007. In press.

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

[21] M. Obermeier and T. Augustin. Lucenos discretization methods and its application in decision making under ambiguity. In *Proc. 5th Int. Symp. on Imprecise Probabilities: Theories and Applications*, 2007.

[22] T. Seidenfeld. Dilation for sets of probabilities. *The Annals of Statistics*, 21:1139–1154, 1993.

[23] F. Tonon. Some properties of a random set approximation to upper and lower distribution functions. *Int. J. of Approximate Reasoning*, 48:174–184, 2008.

[24] M. C. M. Troffaes. Finite approximations to coherent choice. *Int. J. of Approximate Reasoning*, In press.

[25] M.C.M. Troffaes. Decision making under uncertainty using imprecise probabilities. *Int. J. of Approximate Reasoning*, 45:17–29, 2007.

[26] L. Utkin and T. Augustin. Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In *Proc. of the fourth Int. Symp. on Imprecise Probabilities and Their Applications*, 2005.

[27] L. Utkin and S. Destercke. Computing expectations with p-boxes: two views of the same problem. In *Proc. of the fifth Int. Symp. on Imprecise Probabilities and Their Applications*, 2007.

[28] L.V. Utkin. Risk analysis under partial prior information and non-monotone utility functions. *Int. J. of Information Technology and Decision Making*, 6:625–647, 2007.

[29] R.J. Vanderbei. *Linear Programming*. Springer Verlag, 2007.

[30] P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York, 1991.

[31] P. Walley. Measures of uncertainty in expert systems. *Artifical Intelligence*, 83:1–58, 1996.

St. Petersburgh Forest Technical Academy, Dept. of Computer Science, Institutski per. 5, 194021, St. Petersburgh, Russia
E-mail address: `lev.utkin@list.ru`

Institut de Radioprotection et de Sureté Nucléaire (IRSN), Cadarache, France
E-mail address: `desterck@irit.fr`