# Probability boxes on totally preordered spaces for multivariate modelling

**Matthias Troffaes**
Durham University, Durham, UK
matthias.troffaes@gmail.com

**Sebastien Destercke**
CIRAD, Montpellier, France
sebastien.destercke@cirad.fr

## Abstract

Probability boxes (pairs of cumulative distribution functions) are among the most popular models used in imprecise probability theory. In this paper, we provide new efficient tools to construct multivariate p-boxes and develop algorithms to draw inferences from them. For this purpose, we formalise and extend the theory of p-boxes using lower previsions. We allow p-boxes to be defined on arbitrary totally preordered spaces, hence thereby also admitting multivariate p-boxes. We discuss the construction of multivariate p-boxes under various independence assumptions. An example demonstrates the practical feasibility of our results.

**Keywords.** p-box, natural extension, multivariate, elicitation, independence, Fréchet, lower prevision

## 1 Introduction

Imprecise probability [18] refers to uncertainty models applicable in situations where the available information does not allow us to single out a unique probability measure for the random variables involved. They require more complex mathematical tools, such as non-linear functionals. It is therefore of interest to consider models that yield simpler mathematical descriptions, at the expense of generality, but gaining ease of use, elicitation, and representation.

We consider one such model: pairs of lower and upper distribution functions, also called *probability boxes*, or p-boxes [9, 10]. They are often used in risk studies, where cumulative distributions are central. Many theoretical properties and practical aspects of p-boxes have already been studied in the literature. Previous work includes probabilistic arithmetic [20], which provides a very efficient numerical framework for particular inferences with p-boxes (and which we generalise in this paper). In [11], p-boxes are connected to info-gap theory [1]. The relation between p-boxes and random sets was investigated in [14]. Finally, an extension of p-boxes to arbitrary finite spaces [8] yields potential applications to much more general problems.

In this paper, we study p-boxes using lower previsions [19, 18]. From the point of view of lower previsions, p-boxes were studied briefly in [18, Section 4.6.6] and [17]. This has at least two advantages. Firstly, they can be defined on arbitrary spaces. Secondly, they come with a powerful inference tool, called *natural extension*. We will study the natural extension of a p-box, and we derive a number of useful expressions for it, whence providing new numerical tools for exact inferences on arbitrary random quantities and events.

As mentioned, [8] extended p-boxes to finite totally preordered spaces. In this paper, we extend p-boxes further to arbitrary totally preordered spaces, leading to many useful features that classical p-boxes do not have. Firstly, we encompass, in one sweep, p-boxes defined on finite spaces and on closed real intervals. Secondly, as we do not impose anti-symmetry on the ordering, we can also handle product spaces by considering an appropriate total preorder, and thus also admit multivariate non-finite p-boxes, which have not been considered before.[1] Whence, we can specify p-boxes directly on the product space. Contrast this with the usual multivariate approach to p-boxes, such as probabilistic arithmetic [20], that consider one marginal p-box per dimension and draw inferences from a joint model built around some information about variable dependencies. Finally, our approach is also useful in elicitation, as it allows uncertainty to be expressed as probability bounds over any collection of (possibly multivariate) nested sets, because we can always find a total preorder that is compatible with any collection of nested sets.

The paper is organised as follows: Section 2 provides a brief introduction to the theory of coherent lower

[1] We still require the preorder to be total. P-boxes for partially preordered spaces might be interesting, but are not considered in this paper.

previsions. Section 3 introduces and studies the p-box model from the point of view of lower previsions. Section 4 provides an expression for the natural extension of a p-box to all events, and Section 5 studies the natural extension to all gambles. Section 6 studies an important special case of p-boxes whose preorder is induced by a real-valued mapping, as this is a convenient way to specify a multivariate p-box. Section 7 discusses the construction of such multivariate p-boxes from marginal coherent lower previsions under arbitrary dependency models. Section 8 demonstrates the theory with an example.

## 2 Preliminaries

This section introduces lower previsions, see [2, 19, 18, 15] for details.

The possibility space is $\Omega$. A *gamble* on $\Omega$ is a bounded real-valued map on $\Omega$. The set of all gambles on $\Omega$ is $\mathcal{L}(\Omega)$, or $\mathcal{L}$ if $\Omega$ is evident. A subset of $\Omega$ is an *event*. The *indicator* of $A$ is the gamble that is 1 on $A$ and 0 elsewhere: write $I_A$, or $A$ if confusion fails.

A *lower prevision* $\underline{P}$ is a real-valued map on an arbitrary subset $\mathcal{K}$ of $\mathcal{L}$: for any $f$ in $\mathcal{K}$, $\underline{P}(f)$ represents a subject's supremum buying price for $f$ (see [18] for actual explanation). A lower prevision on a set of indicators of events is a *lower probability*.

$\overline{P}$ denotes the conjugate *upper prevision* of $\underline{P}$: for every $-f \in \mathcal{K}$, $\overline{P}(f) = -\underline{P}(-f)$; it represents a subject's infimum selling price for $f$.

A real-valued map $P$ on $\mathcal{L}$ satisfying $P(f) \geq \inf f$ and $P(f+g) = P(f) + P(g)$ for all $f$ and $g \in \mathcal{L}$ is a *linear prevision* on $\mathcal{L}$ [18, p. 88, Sec. 2.4.8]. The set of all linear previsions on $\mathcal{L}$ is denoted by $\mathcal{P}$. A linear prevision is essentially an expectation operator.

Of particular interest is the set

$$\mathcal{M}(\underline{P}) = \{Q \in \mathcal{P} \colon (\forall f \in \mathcal{K})(Q(f) \geq \underline{P}(f))\}.$$

If $\mathcal{M}(\underline{P}) \neq \emptyset$, then $\underline{P}$ is said to *avoid sure loss*, in which case the *natural extension* of $\underline{P}$ [18, Sec. 3.4.1]

$$\underline{E}(f) = \min_{Q \in \mathcal{M}(\underline{P})} Q(f) \text{ for all } f \in \mathcal{L}$$

extends $\underline{P}$ to $\mathcal{L}$. Finally, $\underline{P}$ is called *coherent* [19, p. 18] when it coincides with $\underline{E}$ on $\mathcal{K}$.

A lower prevision $\underline{P}$ defined on a lattice of gambles $\mathcal{K}$, i.e., a set of gambles closed under point-wise maximum and point-wise minimum, is called *n-monotone* if for all $p \in \mathbb{N}$, $p \leq n$, and all $f$, $f_1$, …, $f_p$ in $\mathcal{K}$ [5]:

$$\sum_{I \subseteq \{1,\dots,p\}} (-1)^{|I|} \underline{P}\left(f \wedge \bigwedge_{i \in I} f_i\right) \geq 0.$$

A lower prevision which is $n$-monotone for all $n \in \mathbb{N}$ is called *completely monotone*.

## 3 P-Boxes

Next, we introduce the formalism of p-boxes defined on totally preordered spaces. In contrast to [9], we do not restrict p-boxes to intervals on the real line.

Let $(\Omega, \preceq)$ be a total preorder: so $\preceq$ is transitive and reflexive and any two elements are comparable. We write $x \prec y$ for $x \preceq y$ and $x \not\succeq y$, $x \succ y$ for $y \prec x$, and $x \simeq y$ for $x \preceq y$ and $y \preceq x$. For any two $x$, $y \in \Omega$ exactly one of $x \prec y$, $x \simeq y$, or $x \succ y$ holds. We also use the following common notation for intervals in $\Omega$:

$$[x, y] = \{z \in \Omega \colon x \preceq z \preceq y\}$$
$$(x, y) = \{z \in \Omega \colon x \prec z \prec y\}$$

and similarly for $[x, y)$ and $(x, y]$.

For simplicity, we assume that $\Omega$ has a smallest element $0_\Omega$ and a largest element $1_\Omega$ (we can always add them to $\Omega$).

A *cumulative distribution function* is a mapping $F : \Omega \to [0, 1]$ which is non-decreasing and satisfies moreover $F(1_\Omega) = 1$. For each $x \in \Omega$, we interpret $F(x)$ as the probability of the interval $[0_\Omega, x]$. We do not impose $F(0_\Omega) = 0$, so we allow $\{0_\Omega\}$ to carry non-zero mass, which happens commonly if $\Omega$ is finite. No continuity assumptions are made.

By $\Omega/\simeq$ we denote the quotient set of $\Omega$ with respect to the equivalence relation $\simeq$ induced by $\preceq$, that is:

$$[x]_\simeq = \{y \in \Omega \colon y \simeq x\} \text{ for any } x \in \Omega$$
$$\Omega/\simeq = \{[x]_\simeq \colon x \in \Omega\}$$

Because $F$ is non-decreasing, $F$ is constant on elements $[x]_\simeq$ of $\Omega/\simeq$.

**Definition 1.** A *probability box*, or *p-box*, is a pair $(\underline{F}, \overline{F})$ of cumulative distribution functions from $\Omega$ to $[0, 1]$ satisfying $\underline{F} \leq \overline{F}$.

A p-box is interpreted as a lower and an upper cumulative distribution function. In Walley's framework, this means that a p-box is interpreted as a lower probability $\underline{P}_{\underline{F},\overline{F}}$ on the set of events

$$\mathcal{K} = \{[0_\Omega, x] \colon x \in \Omega\} \cup \{(y, 1_\Omega] \colon y \in \Omega\}$$

by

$$\underline{P}_{\underline{F},\overline{F}}([0_\Omega, x]) = \underline{F}(x) \text{ and } \underline{P}_{\underline{F},\overline{F}}((y, 1_\Omega]) = 1 - \overline{F}(y).$$

P-boxes on a totally preordered space $(\Omega, \preceq)$ are coherent (the proof is virtually identical to the one given in [17, p. 93, Thm. 3.59], which considered p-boxes on $[a, b] \subseteq \mathbb{R}$). We denote by $\underline{E}_{\underline{F},\overline{F}}$ the natural extension of $\underline{P}_{\underline{F},\overline{F}}$ to all gambles.

When $\underline{F} = \overline{F}$, we say that $(\underline{F}, \overline{F})$ is *precise*, and we denote the corresponding lower prevision on $\mathcal{K}$ by $\underline{P}_F$ and its natural extension to $\mathcal{L}$ by $\underline{E}_F$ (with $F := \underline{F} = \overline{F}$).

We end with a useful approximation theorem:

**Theorem 2.** *Let $\underline{P}$ be any coherent lower prevision defined on $\mathcal{L}$. The least conservative p-box $(\underline{F}, \overline{F})$ on $(\Omega, \preceq)$ whose natural extension is dominated by $\underline{P}$ is*

$$\underline{F}(x) = \underline{P}([0_\Omega, x])), \quad \overline{F}(x) = \overline{P}([0_\Omega, x]), \quad \forall x \in \Omega.$$

# 4 Natural Extension to All Events

The remainder of this paper is devoted to finding convenient expressions for the natural extension $\underline{E}_{\underline{F},\overline{F}}$ of $\underline{P}_{\underline{F},\overline{F}}$. We start by giving the form of the natural extension on the field of events generated by $\mathcal{K}$.

## 4.1 Extension to the Field Generated by the Domain

Let $\mathcal{H}$ be the field of events generated by the domain $\mathcal{K}$ of the p-box, i.e., events of the type

$$[0_\Omega, x_1] \cup (x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$$

for $x_1 \prec x_2 \prec x_3 \prec \cdots \prec x_{2n+1}$ in $\Omega$ (if $n$ is 0 we simply take this expression to be $[0_\Omega, x_1]$) and

$$(x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$$

for $x_2 \prec x_3 \prec \cdots \prec x_{2n+1}$ in $\Omega$. Clearly, these events form a field: the union and intersection of any two events in $\mathcal{H}$ is again in $\mathcal{H}$, and the complement of any event in $\mathcal{H}$ also is again in $\mathcal{H}$.

To simplify the description of this field, and the expression of natural extension, we introduce an element $0_\Omega-$ such that $0_\Omega- \prec x$ for all $x \in \Omega$ and:

$$F(0_\Omega-) = \underline{F}(0_\Omega-) = \overline{F}(0_\Omega-) = 0$$

So, $(0_\Omega-, x] = [0_\Omega, x]$. With $\Omega^* = \Omega \cup \{0_\Omega-\}$,

$$\mathcal{H} = \{(x_0, x_1] \cup (x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}] : x_0 \prec x_1 \prec \cdots \prec x_{2n+1} \text{ in } \Omega^*\}. \tag{1}$$

To calculate the natural extension of $\underline{P}_{\underline{F},\overline{F}}$ to all gambles, we first consider the extension from $\mathcal{K}$ to $\mathcal{H}$, then to all events, and finally to all gambles.

A precise p-box $\underline{P}_F$ has a unique extension to a finitely additive probability measure on $\mathcal{H}$:

**Proposition 3.** *$\underline{E}_F$ restricted to $\mathcal{H}$ is a finitely additive probability measure. Moreover, for any $A \in \mathcal{H}$, that is $A = (x_0, x_1] \cup (x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$ with $x_0 \prec x_1 \prec \cdots \prec x_{2n+1}$ in $\Omega^*$, it holds that*

$$\underline{E}_F(A) = \sum_{k=0}^{n} (F(x_{2k+1}) - F(x_{2k})) \tag{2}$$

Proposition 3 extends to p-boxes as follows:

**Proposition 4.** *For any $A \in \mathcal{H}$, that is $A = (x_0, x_1] \cup (x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$ with $x_0 \prec x_1 \prec \cdots \prec x_{2n+1}$ in $\Omega^*$, it holds that $\underline{E}_{\underline{F},\overline{F}}(A) = \underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}}(A)$, where*

$$\underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}}(A) = \sum_{k=0}^{n} \max\{0, \underline{F}(x_{2k+1}) - \overline{F}(x_{2k})\}. \tag{3}$$

For $\overline{E}_{\underline{F},\overline{F}}$, use $\overline{E}_{\underline{F},\overline{F}}(A) = 1 - \underline{E}_{\underline{F},\overline{F}}(A^c)$.

## 4.2 Inner Measure

The inner measure $\underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}\,*}$ of the coherent lower probability $\underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}}$ defined in Eq. (3) coincides with $\underline{E}_{\underline{F},\overline{F}}$ on all events [18, Cor. 3.1.9, p. 127]:

$$\underline{E}_{\underline{F},\overline{F}}(A) = \underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}\,*}(A) = \sup_{C \in \mathcal{H}, C \subseteq A} \underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}}(C). \tag{4}$$

For ease of notation, from now onwards, we denote $\underline{E}_{\underline{F},\overline{F}}$ by $\underline{E}$ when no confusion about the functions $\underline{F}$ and $\overline{F}$ determining the p-box can arise.

In principle, the problem of natural extension to all events is solved: simply calculate the inner measure as in Eq. (4), using Eq. (3) to calculate $\underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}}(C)$ for elements $C$ in $\mathcal{H}$. However, the inner measure still involves calculating a supremum. What we show next is that Eq. (3) can be extended to arbitrary events, by first taking the topological interior with respect to a very simple topology, followed by a (possibly infinite) sum over the so-called full components of this interior.

## 4.3 The Partition Topology

Consider the *partition topology* on $\Omega$ generated by $\tau := \{[x]_\simeq : x \in \Omega\}$. The open sets in this topology are all unions of equivalence classes (or, subsets of $\Omega/\simeq$, if you like). Hence, every open set is also closed. In particular, every interval in $(\Omega, \preceq)$ is clopen.

The topological interior of a set $A$ is given by the union of all equivalence classes contained in $A$:

$$\operatorname{int}(A) = \bigcup\{[x]_\simeq : [x]_\simeq \subseteq A\} \tag{5}$$

and the topological closure is given by the union of all equivalence classes which intersect with $A$:

$$\operatorname{cl}(A) = \bigcup\{[x]_\simeq : [x]_\simeq \cap A \neq \emptyset\}. \tag{6}$$

**Lemma 5.** *For any subset $A$ of $\Omega$, $\underline{E}(A) = \underline{E}(\mathrm{int}(A))$ and $\overline{E}(A) = \overline{E}(\mathrm{cl}(A))$.*

### 4.4 Additivity on Full Components

Next, we determine a constructive expression of the natural extension $\underline{E}$ on the clopen subsets of $\Omega$.

**Definition 6.** [16, §4.4] A set $S \subseteq \Omega$ is called *full* if $[a, b] \subseteq S$ for any $a \preceq b$ in $S$.

What do these full sets look like?

**Lemma 7.** *Every full set is clopen.*

Under an additional completeness assumption, the full sets are precisely the intervals.

**Lemma 8.** *If $\Omega/\simeq$ is order complete, that is, if every subset of $\Omega/\simeq$ has a supremum (minimal upper bound) and infimum (maximal lower bound), then every full set is an interval, that is, it can be written as $[x, y]$, $[x, y)$, $(x, y]$, or $(x, y)$, for some $x$, $y$ in $\Omega$.*

Note that $\Omega/\simeq$ can be made order complete via the Dedekind completion [16, §4.34].

**Definition 9.** [16, §4.4] Given a clopen set $A \subseteq \Omega$ and an element $x$ of $A$, the *full component* $C(x, A)$ of $x$ in $A$ is the largest full set $S$ which satisfies $x \in S \subseteq A$.

**Lemma 10.** *The full components of any clopen set $A$ form a partition of $A$.*

We can prove that the natural extension $\underline{E}$ is additive on full components. Recall that the sum of a family $(x_\lambda)_{\lambda\in\Lambda}$ of non-negative real numbers is defined as

$$\sum_{\lambda\in\Lambda} x_\lambda = \sup_{\substack{L\subseteq\Lambda \\ L \text{ finite}}} \sum_{\lambda\in L} x_\lambda$$

If the above sum is a finite number, at most countably many of the $x_\lambda$'s are non-zero [16, 10.40].

**Theorem 11.** *Let $B$ be a clopen subset of $\Omega$. Let $(B_\lambda)_{\lambda\in\Lambda}$ be the full components of $B$, and let $(C_\lambda)_{\lambda\in\Lambda'}$ be the full components of $B^c$. Then*

$$\underline{E}(B) = \sum_{\lambda\in\Lambda} \underline{E}(B_\lambda) \text{ and } \overline{E}(B) \quad = 1 - \sum_{\lambda\in\Lambda'} \underline{E}(C_\lambda)$$

In other words, the natural extension $\underline{E}$ of a p-box is *arbitrarily additive on full components* (but obviously not additive on arbitrary events). Interestingly, additivity on full components is not sufficient for a lower probability to be equivalent to a p-box.

### 4.5 Practical computations over events

Let us explain how Proposition 4 can be generalized to all events (at least when $\Omega/\simeq$ is order complete). Consider an arbitrary event $A$. By Lemma 5, it suffices to find the natural extension of $\mathrm{int}(A)$ or $\mathrm{cl}(A)$. Calculating the interior or closure with respect to the partition topology will usually be trivial (see examples further on). Because the topological interior or closure of a set is always clopen, we only need to know the natural extension of clopen sets.

Now, by Theorem 11, we only need to calculate the natural extension of the (clopen) full components $(B_\lambda)_{\lambda\in\Lambda}$ of $\mathrm{int}(A)$ or the (clopen) full components $(C_\lambda)_{\lambda\in\Lambda}$ of $\mathrm{cl}(A)^c = \mathrm{int}(A^c)$. Finding the full components will often be a trivial operation. By Lemma 8, if $\Omega/\simeq$ is order complete, then each full component is an interval. And for intervals, we immediately infer from Proposition 4 and Eq. (4) that (i.p. standing for immediate predecessor):

$$\underline{E}((x, y]) = \max\{0, \underline{F}(y) - \overline{F}(x)\} \tag{7a}$$

$$\underline{E}((x, y)) = \max\{0, \underline{F}(y-) - \overline{F}(x)\} \tag{7b}$$

$$\underline{E}([x, y]) = \begin{cases} \max\{0, \underline{F}(y) - \overline{F}(x)\} & \text{if } x \text{ has no i.p.} \\ \max\{0, \underline{F}(y) - \overline{F}(x-)\} & \text{if } x \text{ has an i.p.} \end{cases} \tag{7c}$$

$$\underline{E}([x, y)) = \begin{cases} \max\{0, \underline{F}(y-) - \overline{F}(x)\} & \text{if } x \text{ has no i.p.} \\ \max\{0, \underline{F}(y-) - \overline{F}(x-)\} & \text{if } x \text{ has an i.p.} \end{cases} \tag{7d}$$

for any $x \prec y$ in $\Omega$,[2] where $\underline{F}(y-)$ denotes $\sup_{z\prec y} \underline{F}(z)$ and similarly for $\overline{F}(x-)$. The equalities hold because, if $x \prec y$ in $\Omega$, and $x-$ is an immediate predecessor of $x$, then $[x, y] = (x-, y]$ and $[x, y) = (x-, y)$. Recall also that $\underline{F}(0_\Omega-) = \overline{F}(0_\Omega-) = 0$ by convention. If $\Omega/\simeq$ is finite, then one can think of $z-$ as the immediate predecessor of $z$ in $\Omega/\simeq$.

In other words, we have a simple constructive means of calculating the natural extension of any event.

### 4.6 Special Cases

The above equations hold for any $(\Omega, \preceq)$ with order complete quotient space. In most cases in practice, either $\Omega/\simeq$ is finite, or $\Omega/\simeq$ is connected, meaning that for any two elements $x \prec y$ in $\Omega$ there is a $z$ in $\Omega$ such that $x \prec z \prec y$,[3] (this is the case for instance when $\Omega$ is a closed interval in $\mathbb{R}$ and $\preceq$ is the usual ordering of reals). Moreover, if $\Omega/\simeq$ is connected, then, in practice, $\underline{F}$ will satisfy $\underline{F}(y-) = \underline{F}(y)$ for all $y$ in $\Omega$. For example, in case $\Omega$ is a closed interval in $\mathbb{R}$, this happens precisely when $\underline{F}(0) = 0$ and $\underline{F}$ is left-continuous in the usual sense.

[2] In case $x = 0_\Omega$, evidently, $0_\Omega-$ is the i.p.

[3] This terminology stems from the fact that, in this case, $\Omega/\simeq$ is connected with respect to the order topology [16, §15.46(6)].

If $\Omega/\simeq$ is finite, then every element of $\Omega$ has an immediate predecessor (remember, we take the immediate predecessor of $0_\Omega$ to be $0_\Omega-$), and if $\Omega/\simeq$ is connected, then no element except $0_\Omega$ has an immediate predecessor. So:

**Corollary 12.** *If $\Omega/\simeq$ is finite, then every full set $B \subseteq \Omega$ is of the form $[a, b]$ and for every event $A \subseteq \Omega$,*

$$\underline{E}(A) = \sum_{\lambda\in\Lambda} \max\{0, \underline{F}(b_\lambda) - \overline{F}(a_\lambda-)\}$$
$$\overline{E}(A) = 1 - \sum_{\lambda\in\Lambda'} \max\{0, \underline{F}(b'_\lambda) - \overline{F}(a'_\lambda-)\}$$

*where $([a_\lambda, b_\lambda])_{\lambda\in\Lambda}$ are the full components of* $\mathrm{int}(A)$, *and $([a'_\lambda, b'_\lambda])_{\lambda\in\Lambda'}$ are the full components of* $\mathrm{int}(A^c) = \mathrm{cl}(A)^c$.

**Corollary 13.** *If $\Omega/\simeq$ is order complete and connected, and $\underline{F}(y-) = \underline{F}(y)$ for all $y$ in $\Omega$, then*

$$\underline{E}(A) = \sum_{\lambda\in\Lambda} \max\{0, \underline{F}(\sup B_\lambda) - \overline{F}(\inf B_\lambda)\}$$
$$\overline{E}(A) = 1 - \sum_{\lambda\in\Lambda'} \max\{0, \underline{F}(\sup C_\lambda) - \overline{F}(\inf C_\lambda)\}$$

*where $(B_\lambda)_{\lambda\in\Lambda}$ are the full components of* $\mathrm{int}(A)$ *and $(C_\lambda)_{\lambda\in\Lambda'}$ are the full components of* $\mathrm{int}(A^c) = \mathrm{cl}(A)^c$.

Beware of $\underline{F}(0_\Omega) = \underline{F}(0_\Omega-) = 0$ in the last corollary.

### 4.7 Example

Let's investigate a particular type of p-boxes on the unit square $[0, 1]^2$. First, we must specify a preorder on $\Omega$. A natural yet naive way of doing so is, for instance, saying that $(x_1, y_1) \preceq (x_2, y_2)$ whenever $x_1 + y_1 \le x_2 + y_2$. Consider a p-box $(\underline{F}, \overline{F})$ on $([0, 1]^2, \preceq)$. Since $\underline{F}$ is required to be non-decreasing with respect to $\preceq$, it follows that $\underline{F}(x, y)$ is constant on elements of $[0, 1]^2/\simeq$, which means that $F(x_1, y_1) = F(x_2, y_2)$ whenever $x_1 + y_1 = x_2 + y_2$. Thus, we may think of $\underline{F}(x, y)$ as a function of a single variable $z = x + y$, and we write $\underline{F}(z)$. Similarly, we write $\overline{F}(z)$.

So, our p-box specifies bounds on the probability of right-angled triangles (restricted to $[0, 1]^2$) whose hypothenuses are orthogonal to the diagonal:

$$\underline{F}(z) \le p(\{(x, y) \in [0, 1]^2 : x + y \le z\}) \le \overline{F}(z) \quad (8)$$

Observe that the p-box is given directly on the two-dimensional product space, without the need to define marginal p-boxes for each dimension. The base $\tau$ for our partition topology is given by

$$\tau = \{\{(x, y) \in [0, 1]^2 : x + y = z\} : z \in [0, 2]\}$$

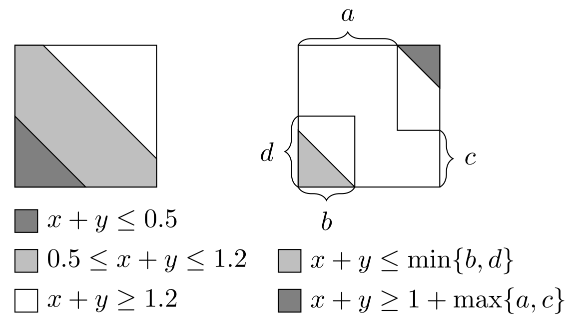


Figure 1: Shape of intervals induced by $\preceq$, and calculation of the topological interior.

For example, the topological interior of a rectangle $A = [a, b] \times [c, d]$ is empty, unless $a = c = 0$ or $b = d = 1$, because in all other cases, no element of $\tau$ is a subset of $A$. In the cases where $a = c = 0$ and $\min\{b, d\} < 1$, or $\max\{a, c\} > 0$ and $b = d = 1$ (if $a = c = 0$ and $b = d = 1$ then the interior is $\Omega$), respectively, we have:

$$\mathrm{int}([0, b] \times [0, d]) = \{(x, y) \in [0,1]^2 : x+y \le \min\{b, d\}\}$$
$$\mathrm{int}([a, 1] \times [c, 1]) = \{(x, y) \in [0,1]^2 : x+y \ge 1+\max\{a, c\}\}$$

Consequently, $\underline{E}(A) = 0$ for all rectangles $A$, except

$$\underline{E}([0, b] \times [0, d]) = \underline{F}(\min\{b, d\})$$
$$\underline{E}([a, 1] \times [c, 1]) = 1 - \overline{F}(1 + \max\{a, c\})$$

Fig. 1 illustrates the situation. So, for the purpose of making inferences about the lower probability of events that are rectangles, the ordering $\preceq$ was obviously poorly chosen. In general, *one should choose $\preceq$ in a way that $\Omega/\simeq$ contains good approximations for all events of interest.*

For example, a strategy would be to start from a reference point (e.g., an elicited modal value) and then to choose the ordering $\preceq$ such that intervals correspond to concentric regions of interests around the reference point. Again, all of this is possible because our theory concerns p-boxes on arbitrary totally preordered spaces, and is not limited to the real line with its natural ordering. More realistic examples in which such concentric regions are used are given in Section 8.

## 5 Natural Extension to All Gambles

Next, we establish that the natural extension of p-boxes to all gambles can be expressed as a Choquet integral. We further simplify the calculation of this Choquet integral via the lower and upper oscillation of gambles with respect to the partition topology introduced earlier.

### 5.1 Choquet Integral Representation

Extending previous results [8] where the relation between p-boxes and complete monotonicity was established for finite spaces, we can show that the natural extension of p-boxes on totally pre-ordered spaces are completely monotone. Let $\underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}}$ denote the restriction of $\underline{E}_{\underline{F},\overline{F}}$ to $\mathcal{H}$, given by Proposition 4:

**Theorem 14.** *$\underline{P}^{\mathcal{H}}_{\underline{F},\overline{F}}$ is completely monotone.*

This allows us to characterise the natural extension on all gambles:

**Theorem 15.** *The natural extension $\underline{E}$ of $\underline{P}_{\underline{F},\overline{F}}$ is given by the Choquet integral*

$$\underline{E}(f) = \inf f + \int_{\inf f}^{\sup f} \underline{E}(\{f \geq t\})\,\mathrm{d}t$$

*for every gamble $f$. Moreover, $\underline{E}$ is completely monotone on all gambles. Similarly,*

$$\overline{E}(f) = \inf f + \int_{\inf f}^{\sup f} \overline{E}(\{f \geq t\})\,\mathrm{d}t.$$

### 5.2 Lower and Upper Oscillation

By Lemma 5, to turn Theorem 15 in an effective algorithm, we must calculate $\mathrm{int}(\{f \geq t\})$ for every $t$. Fortunately, there is a very simple way to do this.

For any gamble $f$ on $\Omega$ and any topological base $\tau$, define its *lower oscillation* as the gamble

$$\underline{osc}(f)(x) = \sup_{C \in \tau\colon x \in C} \inf_{y \in C} f(y)$$

For the partition topology which we introduced earlier, this simplifies to

$$\underline{osc}(f)(x) = \inf_{y \in [x]_{\simeq}} f(y) \tag{9}$$

The upper oscillation is:

$$\overline{osc}(f)(x) = -\underline{osc}(-f)(x) = \sup_{y \in [x]_{\simeq}} f(y) \tag{10}$$

For a subset $A$ of $\Omega$, the lower oscillation of $I_A$ is $I_{\mathrm{int}(A)}$, so the lower oscillation is the natural generalisation of the topological interior to gambles. Similarly, the upper oscillation of $I_A$ is $I_{\mathrm{cl}(A)}$.

**Proposition 16.** *For any gamble $f$ on $\Omega$,*

$$\mathrm{int}(\{f \geq t\}) = \{\underline{osc}(f) \geq t\}$$
$$\mathrm{cl}(\{f \geq t\}) = \{\overline{osc}(f) \geq t\}$$

*so, in particular,*

$$\underline{E}(f) = \inf \underline{osc}(f) + \int_{\inf \underline{osc}(f)}^{\sup \underline{osc}(f)} \underline{E}(\{\underline{osc}(f) \geq t\})\,\mathrm{d}t$$
$$\overline{E}(f) = \inf \overline{osc}(f) + \int_{\inf \overline{osc}(f)}^{\sup \overline{osc}(f)} \overline{E}(\{\overline{osc}(f) \geq t\})\,\mathrm{d}t$$

Concluding, to calculate the natural extension of any gamble, in practice, we must simply determine the full components of the cut sets of its lower or upper oscillation, and calculate a simple Riemann integral of a monotonic function.

Examples will be given in Section 8.

## 6 P-Boxes Whose Preorders are Induced by a Real-Valued Function

In practice, a convenient way to specify a preorder $\preceq$ on $\Omega$ such that $\Omega/\simeq$ is order complete and connected is by means of a bounded real-valued function $Z\colon \Omega \to \mathbb{R}$. For instance, in the example in Section 4.7, we used $Z(x,y) = x + y$. Also see [1, 12]. Let us assume from now onwards that $Z$ is a surjective mapping from $\Omega$ to $[0,1]$.

For any $x$ and $y$ in $\Omega$, define $x \preceq y$ whenever $Z(x) \leq Z(y)$. Because $Z$ is surjective, $\Omega/\simeq$ is order complete and connected. In particular, $\Omega$ has a smallest and largest element, for which $Z(0_\Omega) = 0$ and $Z(1_\Omega) = 1$. Moreover, we can think of any cumulative distribution function on $(\Omega, \preceq)$ as a function over a single variable $z \in [0,1]$. Consequently, we can think of any p-box on $(\Omega, \preceq)$ as a p-box on $([0,1], \leq)$. In particular, for any subset $I$ of $[0,1]$ we write $\underline{E}(I)$ for $\underline{E}(Z^{-1}(I))$. For example, for $a$, $b$ in $[0,1]$, and $A = Z^{-1}((a,b]) \subseteq \Omega$, we have that

$$\underline{E}(A) = \underline{E}((a,b]) = \max\{0, \underline{F}(a) - \overline{F}(b)\}$$

by Proposition 4. Similar expressions for other types of intervals follow from Eq. (7).

The topological interior and closure can be related to the so-called *lower and upper inverse* of $Z^{-1}$. Indeed, consider the multi-valued mapping $\Gamma := Z^{-1}\colon [0,1] \to \wp(\Omega)$. Because for every $x$ in $\Omega$, it holds that $[x]_{\simeq} = \Gamma(Z(x))$, it follows that, for any subset $A$ of $\Omega$, $\mathrm{int}(A) = \Gamma(\Gamma_*(A))$, and $\mathrm{cl}(A) = \Gamma(\Gamma^*(A))$, where $\Gamma_*$ and $\Gamma^*$ denote the lower and upper inverse of $\Gamma$ respectively, that is [7]

$$\Gamma_*(A) = \{z \in [0,1]\colon \Gamma(z) \subseteq A\}, \text{ and}$$
$$\Gamma^*(A) = \{z \in [0,1]\colon \Gamma(z) \cap A \neq \emptyset\}.$$

**Theorem 17.** *Let $A$ be any subset of $\Omega$. Then*

$$\underline{E}(A) = \sum_{\lambda \in \Lambda} \underline{E}(I_\lambda)$$
$$\overline{E}(A) = 1 - \sum_{\lambda \in \Lambda'} \underline{E}(J_\lambda)$$

*where $(I_\lambda)_{\lambda \in \Lambda}$ are the full components of $Z(\mathrm{int}(A)) = \Gamma_*(A)$ and $(J_\lambda)_{\lambda \in \Lambda'}$ are the full components of $Z(\mathrm{int}(A^c)) = Z(\mathrm{cl}(A)^c) = \Gamma_*(A^c) = (\Gamma^*(A))^c$.*

*If, in addition,* $\underline{F}$ *is left-continuous as a function of* $z \in [0,1]$ *and* $\underline{F}(0) = 0$*, then*

$$\underline{E}(A) = \sum_{\lambda\in\Lambda} \max\{0, \underline{F}(\sup I_\lambda) - \overline{F}(\inf I_\lambda)\}$$
$$\overline{E}(A) = 1 - \sum_{\lambda\in\Lambda'} \max\{0, \underline{F}(\sup J_\lambda) - \overline{F}(\inf J_\lambda)\}$$

For gambles, the lower oscillation is constant on equivalence classes. So, we may also consider $\underline{osc}(f)$ and $\overline{osc}(f)$ in Proposition 16 as functions of $z \in [0,1]$.

## 7 Constructing Multivariate P-Boxes from Marginals

Next, we construct a multivariate p-box from marginal lower previsions under arbitrary rules of combination. We then focus on two special joint models: the first without any assumptions about dependence between variables (using the Fréchet-Hoeffding bounds [13]), and the second assuming epistemic independence between all variables (using the factorization property [3]). Finally, we derive Williamson and Downs's [20] probabilistic arithmetic as a special case of our framework.

Specifically, consider $n$ variables $X_1, \dots, X_n$ assuming values in $\mathcal{X}_1, \dots, \mathcal{X}_n$, and marginal lower previsions $\underline{P}_1, \dots, \underline{P}_n$ for each variable. Each $\underline{P}_i$ is a coherent lower prevision on $\mathcal{L}(\mathcal{X}_i)$.

### 7.1 Multivariate P-Boxes

First, we must define a mapping $Z$ to induce a preorder $\preceq$ on $\Omega = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. The following choice works perfectly for our purpose:

$$Z(x_1, \dots, x_n) = \max_{i=1}^{n} Z_i(x_i)$$

where each $Z_i$ is a surjective mapping from $\mathcal{X}_i$ to $[0,1]$ and hence, also induces a marginal preorder $\preceq_i$ on $\mathcal{X}_i$. Each $\underline{P}_i$ can be approximated by a p-box $(\underline{F}_i, \overline{F}_i)$ on $(\mathcal{X}_i, \preceq_i)$, defined by

$$\underline{F}_i(z) = \underline{P}_i(Z_i^{-1}([0,z])) \quad \overline{F}_i(z) = \overline{P}_i(Z_i^{-1}([0,z]))$$

This approximation is the best possible one, by Theorem 2.

Beware that even though different choices of $Z_i$ may induce the same total preorder $\preceq_i$, they might lead to a different total preorder $\preceq$ induced by $Z$. Roughly speaking, the $Z_i$ specify how the marginals scale relative to one another. This means that our choice of $Z_i$ affects the precision of our inferences: a good choice will ensure that any event of interest can be well approximated by elements of $\Omega/\simeq$. Of course, nothing prevents us, at least in theory, to consider the set of all $Z_i$ which induce some given marginal total preorders $\preceq_i$, and whence to work with a set of p-boxes. In Section 7.4, we will see an example where this approach is feasible.

Anyway, with this choice of $Z$, we can easily find the p-box which represents the joint as accurately as possible, under any rule of combination of coherent lower previsions:

**Theorem 18.** *Consider any rule of combination* $\odot$ *of coherent lower and upper previsions, mapping the marginals* $\underline{P}_1, \dots, \underline{P}_n$ *to a joint coherent lower prevision* $\bigodot_{i=1}^n \underline{P}_i$ *on all gambles. Suppose there are functions* $\ell$ *and* $u$ *for which:*

$$\bigodot_{i=1}^{n} \underline{P}_i \left(\prod_{i=1}^{n} A_i\right) = \ell(\underline{P}_1(A_1), \dots, \underline{P}_n(A_n)) \text{ and}$$
$$\bigodot_{i=1}^{n} \overline{P}_i \left(\prod_{i=1}^{n} A_i\right) = u(\overline{P}_1(A_1), \dots, \overline{P}_n(A_n)),$$

*for all* $A_1 \subseteq \mathcal{X}_1, \dots, A_n \subseteq \mathcal{X}_n$*. Then, the couple* $(\underline{F}, \overline{F})$ *defined by*

$$\underline{F}(z) = \ell(\underline{F}_1(z), \dots, \underline{F}_n(z)); \overline{F}(z) = u(\overline{F}_1(z), \dots, \overline{F}_n(z))$$

*is the least conservative p-box on* $(\Omega, \preceq)$ *whose natural extension* $\underline{E}_{\underline{F},\overline{F}}$ *is dominated by the combination* $\bigodot_{i=1}^n \underline{P}_i$ *of* $\underline{P}_1, \dots, \underline{P}_n$*.*

### 7.2 Natural Extension: The Fréchet Case

The *natural extension* $\boxtimes_{i=1}^n \underline{P}_i$ of $\underline{P}_1, \dots, \underline{P}_n$ is the lower envelope of all joint distributions whose marginal distributions are compatible with the given marginal lower previsions. So, the model is completely vacuous about the dependence structure. We refer to for instance [4, p. 120, §3.1] for a rigorous definition. In this paper, we only need to use the Fréchet bounds (see [21, p. 131]), in which case the functions $\ell$ and $u$ of Theorem 18 are respectively the Lukasiewicz and the minimum t-norms.

**Theorem 19.** *The p-box* $(\underline{F}, \overline{F})$ *defined by*

$$\underline{F}(z) = \max\left\{0, 1 - n + \sum_{i=1}^{n} \underline{F}_i(z)\right\} \quad \overline{F}(z) = \min_{i=1}^{n} \overline{F}_i(z)$$

*is the least conservative p-box on* $(\Omega, \preceq)$ *whose natural extension* $\underline{E}_{\underline{F},\overline{F}}$ *is dominated by the natural extension* $\boxtimes_{i=1}^n \underline{P}_i$ *of* $\underline{P}_1, \dots, \underline{P}_n$*.*

It is easily seen that the joint lower prevision $\boxtimes_{i=1}^n \underline{P}_i$ is in general not completely monotone, hence the joint p-box of Theorem 19 is in general only an outer approximation.

### 7.3 Independent Natural Extension

In contrast, the *independent natural extension* $\otimes_{i=1}^n \underline{P}_i$ of $\underline{P}_1, \ldots, \underline{P}_n$ models epistemic independence between $X_1, \ldots, X_n$. We refer to [3] for a rigorous definition and properties. In this paper we only need the factorization property, which implies that the functions $\ell$ and $u$ of Theorem 18 are the product rule.

**Theorem 20.** *The p-box* $(\underline{F}, \overline{F})$ *defined by*

$$\underline{F}(z) = \prod_{i=1}^{n} \underline{F}_i(z) \qquad \overline{F}(z) = \prod_{i=1}^{n} \overline{F}_i(z)$$

*is the least conservative p-box on* $(\Omega, \preceq)$ *whose natural extension* $\underline{E}_{\underline{F},\overline{F}}$ *is dominated by the indepedent natural extension* $\otimes_{i=1}^n \underline{P}_i$ *of* $\underline{P}_1, \ldots, \underline{P}_n$.

Again, the joint p-box will only be an outer approximation of the actual joint lower prevision.

### 7.4 Special Case: Probabilistic Arithmetic

Let $Y = X_1 + X_2$ with $X_1$ and $X_2$ real-valued random variables. Probabilistic arithmetic [21] estimates $\underline{P}_Y([-\infty, y]) = \underline{F}_Y(y)$ and $\overline{P}_Y([-\infty, y]) = \overline{F}_Y(y)$ for any $y \in \mathbb{R}$ under the assumptions that the uncertainty on $X_1$ and $X_2$ is given by p-boxes $(\underline{F}_1, \overline{F}_1)$ and $(\underline{F}_2, \overline{F}_2)$, with $\preceq_1$ and $\preceq_2$ the natural ordering of real numbers, and the dependence structure is completely unknown. Williamson and Downs [20] provide explicit formulae for common arithmetic operations, making inferences from marginal p-boxes very easy.

Let us show, for the particular case of addition, that their results are captured by our joint p-box proposed in Theorem 19. Cases of other arithmetic operators, not treated here to save space, follow from almost identical reasoning. The lower cumulative distribution function $\underline{F}_{X_1+X_2}(y)$ resulting from probabilistic arithmetic is, for any $y \in \mathbb{R}$,

$$\sup_{x_1, x_2 \colon x_1 + x_2 = y} \max\{0, \underline{F}_1(x_1) + \underline{F}_2(x_2) - 1\}. \quad (11)$$

Without much loss of generality, assume that both $X_1$ and $X_2$ lie in a bounded interval $[a, b]$.

Let $Z_1$ and $Z_2$ be any surjective maps $[a, b] \to [0, 1]$ which induce the usual ordering on $[0, 1]$ (so both must be continuous and strictly increasing).

To apply Theorem 19, consider the total preorder $\preceq$ on $\Omega = [a, b]^2$ induced by $Z(x_1, x_2) = \max\{Z_1(x_1), Z_2(x_2)\}$. Figure 2 illustrates the event[4] $\{X_1 + X_2 \leq y\}$, with $y \in [2a, 2b]$, as well as the largest interval $Z^{-1}([0, z])$ included in it. For $z$ such that

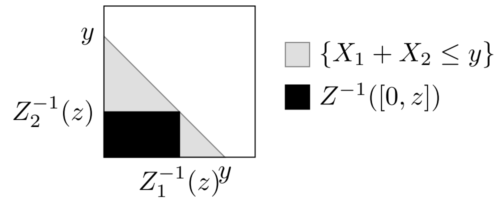


Figure 2: The event $\{X_1 + X_2 \leq y\}$, and the largest interval $Z^{-1}([0, z])$ included in it.

$Z_1^{-1}(z) + Z_2^{-1}(z) = y$, we achieve the largest interval $Z^{-1}([0, z])$ which is still included in $\{X_1 + X_2 \leq y\}$. There is always a unique such $z$ because also $Z_1^{-1} + Z_2^{-1}$ is continuous and strictly increasing.

Using Theorems 19 and 17, we find that

$$\begin{aligned}\underline{E}_{\underline{F},\overline{F}}(\{X_1 + X_2 \leq y\}) &= \underline{F}(Z^{-1}(z)) \\ &= \max\{0, \underline{F}_1(Z_1^{-1}(z)) + \underline{F}_2(Z_2^{-1}(z)) - 1\}\end{aligned}$$

But, this holds for every valid choice of $Z_1$ and $Z_2$, whence $\underline{P}_1 \boxtimes \underline{P}_2(\{X_1 + X_2 \leq y\})$ dominates Eq. (11).

## 8 Example

Next, we investigate an example in which p-boxes are used to model uncertainty around some parameters.

We aim to estimate the minimal required dike height $h$ along a stretch of river, using a model proposed in [6]. Although this model is quite simple, it provides a realistic industrial application. Skipping technical details, the model results in the following relationship:

$$h(q, k, u, d) = \begin{cases} \left(\dfrac{q}{k\sqrt{\frac{u-d}{\ell}}b}\right)^{\frac{3}{5}} & \text{if } q \geq 0 \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

with $b$ and $\ell$ the river width and length, $q$ the river flow rate, $k$ the Strickler coefficient and $u$, $d$ respectively the upriver and downriver water levels.

For this case study, the river width is $b = 300m$ and the length is $\ell = 6400m$. The remaining parameters are uncertain. Expert assessment leads to the following distributions.

The river flow rate $q$ has a Gumbel distribution with location and scale parameters $\mu = 1335m^3s^{-1}$ and $\beta = 716m^3s^{-1}$. To simplify calculations, we introduce a variable $r$ satisfying $q = \mu - \beta\ln(-\ln(r))$. If $r$ is uniform over $[0, 1]$, then $q$ is Gumbel with parameters $\mu$ and $\beta$. So, after transformation,

$$h(r, k, u, d) = \begin{cases} \left(\dfrac{\mu - \beta\ln(-\ln(r))}{k\sqrt{\frac{u-d}{\ell}}b}\right)^{\frac{3}{5}} & \text{if } q \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

[4] $\{X_1 + X_2 \leq y\}$ is $\{(x_1, x_2) \in [0, 1]^2 \colon x_1 + x_2 \leq y\}$.

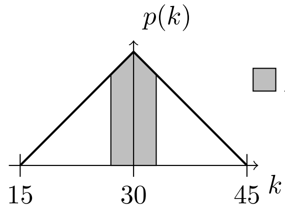


Figure 3: Derivation of the p-box for a triangular distribution.

The Strickler coefficient $k$ has a symmetric triangular distribution over the interval $[15m^{1/3}s^{-1}, 45m^{1/3}s^{-1}]$.

Upper and downriver water levels $u$ and $d$ are uncertain due to sedimentary conditions. Measured values are $u^* = 55m$ and $d^* = 50m$, with measurement error definitely less than $1m$. These are also modelled by symmetric triangular distributions, on $[54m, 56m]$ and $[49m, 51m]$ respectively.

A natural choice for $Z$ is the distance between the expected values ($r^* = 1/2, k^* = 30, u^* = 55, d^* = 50$) and the actual values $(r, k, u, d)$:

$$Z(r, k, u, d) = \max\{2|r - \tfrac{1}{2}|, \tfrac{|k-30|}{15}, |u - 55|, |d - 50|\}.$$

The scale of the distances has been chosen such that $Z(r, k, u, d) \leq 1$ for all points of interest. Equivalence classes $[(r, k, u, d)]_\simeq$ are borders of 4-dimentional boxes with vertices (with $z = Z(r, k, u, d)$)

$$((1 \pm z)/2, 30 \pm 15z, 55 \pm z, 50 \pm z).$$

The marginal p-boxes are, for $r$:

$$\underline{F}_1(z) = \overline{F}_1(z) = p(2|r - 1/2| \leq z) = z$$

because $r$ is uniformly distributed over $[0, 1]$. For $k$:

$$\underline{F}_2(z) = \overline{F}_2(z) = p(|k - 30|/15 \leq z) = 1 - (1 - z)^2$$

(see Fig. 3). Similarly, for $u$ and $d$, it is easily verified that $\underline{F}_3(z) = \overline{F}_3(z) = \underline{F}_4(z) = \overline{F}_4(z) = 1 - (1 - z)^2$.

Next, $\underline{osc}(h)$ and $\overline{osc}(h)$ are:

$$\underline{osc}(h)(z) = \inf_{(r,k,u,d)\colon Z(r,k,u,d)=z} h(r, k, u, d) = o(-z)$$
$$\overline{osc}(h)(z) = \sup_{(r,k,u,d)\colon Z(r,k,u,d)=z} h(r, k, u, d) = o(z)$$

with

$$o(z) = \begin{cases} \left( \frac{\mu - \beta \ln(-\ln((1+z)/2))}{(30-15z)\sqrt{\frac{5-2z}{\ell}}b} \right)^{\frac{3}{5}} & \text{if } \cdots \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The function $o(z)$ is increasing, with $o(-1) = 0$, $o(0) = 3.032$, and $o(1) = +\infty$.

Hence, $\underline{osc}(h)(z)$ and $\overline{osc}(h)(z)$ are decreasing and increasing in $z$, respectively. So, the full components of the events

$$L_t = \{z \in [0, 1] \colon \underline{osc}(h)(z) \geq t\} = \{z \in [0, 1] \colon o(-z) \geq t\}$$
$$U_t = \{z \in [0, 1] \colon \overline{osc}(h)(z) \geq t\} = \{z \in [0, 1] \colon o(z) \geq t\}$$

are of the form $L_t = [0, \ell_t]$ and $U_t = [u_t, 1]$, with

$$\ell_t = -o^{-1}(t) \text{ for } t \leq o(0) \quad u_t = o^{-1}(t) \text{ for } t \geq o(0)$$

With unknown dependence, using Theorem 19,

$$\underline{F}(z) = \max\{0, -3 + z + 3(1 - (1 - z)^2)\}$$

and whence

$$\underline{E}(h) = \int_0^{o(0)} \underline{F}(-o^{-1}(t))\, \mathrm{d}t = 1.515$$
$$\overline{E}(h) = o(0) + \int_{o(0)}^{+\infty} \left(1 - \underline{F}(o^{-1}(t))\right) \mathrm{d}t = 6.423$$

Therefore, we should consider average overflowing heights of at least $6.5m$. For comparison, using traditional methods and assuming independence between all variables, $h$ has expectation $3.2m$, which lies between our lower and upper expectation, as expected. Note that the imprecision has two sources: we have reduced a multivariate problem to a univariate one and we have not made any assumption of independence.

Calculations were relatively simple due to the monotonicity of the target function with respect to the uncertain variables. This may not be the case in general.

## 9 Conclusions

We studied inferences (lower and upper expectations) from p-boxes on arbitrary totally preordered spaces. For this purpose, we represented p-boxes as coherent lower previsions, and studied their natural extension. Defining p-boxes on totally pre-ordered spaces allowed us to unify p-boxes on finite spaces and on real intervals, and to extend the theory to the multivariate case.

One interesting result is a practical means of calculating the natural extension of a p-box in this general setting: we proved that it suffices to calculate the full components of the cut sets of the lower oscillation, followed by a simple Riemann integral (Proposition 16).

As examples of how this model can be used in practice, we have detailed the cases of p-boxes whose preorders are induced by a real-valued mapping, and of joint p-boxes built from marginals under various combination rules. We demonstrated our methodology on inference about a river dike assessment, showing that calculations are generally straightforward.

Of course, many open problems regarding p-boxes remain. For instance, can the dependency model inform the choice of preorder, to arrive at tighter bounds? Our choice led to simple expressions, but other choices giving more precise inference could be investigated. Also, the connection of p-boxes with other uncertainty models, such as possibility measures and clouds, deserves further investigation.

## Acknowledgements

We are particularly grateful to Enrique Miranda for the many very fruitful discussions, extremely useful suggestions, and various contributions to this paper. We also thank Gert de Cooman and Didier Dubois for their help with a very early draft of this paper. Finally, we thank both reviewers for their valuable comments and suggestions.

## References

[1] Y. Ben-Haim. *Info-gap decision theory: decisions under severe uncertainty.* Academic Press, London, 2006.

[2] G. Boole. *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities.* Walton and Maberly, London, 1854.

[3] G. de Cooman, E. Miranda, and M. Zaffalon. Independent natural extension. In E. Hüllermeier, R. Kruse, and F. Hoffmann, editors, *Computational Intelligence for Knowledge-Based Systems Design*, Lecture Notes in Computer Science, pages 737–746. Springer, 2010.

[4] G. de Cooman and M. C. M. Troffaes. Coherent lower previsions in systems modelling: products and aggregation rules. *Reliability Engineering and System Safety*, 85:113–134, 2004.

[5] G. de Cooman, M. C. M. Troffaes, and E. Miranda. $n$-Monotone exact functionals. *Journal of Mathematical Analysis and Applications*, 347(1):143–156, 2008.

[6] E. de Rocquigny, N. Devictor, and S. Tarantola. *Uncertainty in Industrial Practice: A guide to Quantitative Uncertainty Management*, chapter 10. Partial Safety Factors to Deal with Uncertainties in Slope Stability of River Dykes. Wiley, 2008.

[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. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008.

[9] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories, January 2003.

[10] S. Ferson and W. Tucker. Sensitivity analysis using probability bounding. *Reliability engineering and system safety*, 91(10-11):1435–1442, 2006.

[11] S. Ferson and W. Tucker. Probability boxes as info-gap models. In *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society*, New York (USA), 2008.

[12] M. Fuchs and A. Neumaier. Potential based clouds in robust design optimization. *Journal of Statistical Theory and Practice, Special Issue on Imprecision*, 3(1):225–238, 2008.

[13] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the Americal Statistical Association*, 58:13–30, 1963.

[14] E. Kriegler and H. Held. Utilizing belief functions for the estimation of future climate change. *International Journal of Approximate Reasoning*, 39:185–209, 2005.

[15] E. Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.

[16] E. Schechter. *Handbook of Analysis and Its Foundations.* Academic Press, San Diego, CA, 1997.

[17] M. C. M. Troffaes. *Optimality, Uncertainty, and Dynamic Programming with Lower Previsions.* PhD thesis, Ghent University, Ghent, Belgium, March 2005.

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

[19] P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975.

[20] R. C. Williamson and T. Downs. Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4:89–158, 1990.

[21] R.C. Williamson. *Probabilistic Arithmetic.* PhD thesis, University of Queensland, Australia, 1989.