

# Jensen's and Cantelli's inequalities with imprecise previsions

Renato Pelessoni \*, Paolo Vicig \*

DEAMS, University of Trieste, Piazzale Europa 1, 34127, Trieste, Italy

---

## Abstract

We investigate how basic probability inequalities can be extended to an imprecise framework, where (precise) probabilities and expectations are replaced by imprecise probabilities and lower/upper previsions. We focus on inequalities giving information on a single bounded random variable  $X$ , considering either convex/concave functions of  $X$  (Jensen's inequalities) or one-sided bounds such as  $(X \geq c)$  or  $(X \leq c)$  (Markov's and Cantelli's inequalities). As for the consistency of the relevant imprecise uncertainty measures, our analysis considers coherence as well as weaker requirements, notably 2-coherence, which proves to be often sufficient. Jensen-like inequalities are introduced, as well as a generalisation of a recent improvement to Jensen's inequality. Some of their applications are proposed: extensions of Lyapunov's inequality and inferential problems. After discussing upper and lower Markov's inequalities, Cantelli-like inequalities are proven with different degrees of consistency for the related lower/upper previsions. In the case of coherent imprecise previsions, the corresponding Cantelli's inequalities make use of Walley's lower and upper variances, generally ensuring better bounds.

© 2022 Elsevier B.V. All rights reserved.

*Keywords:* Lower previsions; Coherence; 2-coherence; Jensen's inequality; Cantelli's inequalities

---

## 1. Introduction

Probability theory is rich of a number of more or less famous inequalities, some of which, like Markov's inequality, usually presented already in introductory courses. Inequalities may serve several purposes, and a common one is offering bounds to uncertainty evaluations of a random number  $X$ . The bounds are often independent of the effective distribution of  $X$ , hence are especially useful when its knowledge is vague or imprecise. Thus, a natural question is: how are well known probability inequalities modified in the more general frame of Imprecise Probability theory? The theory deals explicitly with imprecise evaluations: for instance, lower/upper previsions replace and generalise expectations [21]. A specific feature of lower/upper previsions is that they may be defined on *arbitrary* sets of *gambles* (i.e. bounded random variables). This matches well with practical situations, where effective uncertainty knowledge may be limited to just a few gambles. Another relevant feature is that more consistency criteria for lower/upper previsions,

---

\* Corresponding authors.

E-mail addresses: [renato.pelessoni@deams.units.it](mailto:renato.pelessoni@deams.units.it) (R. Pelessoni), [paolo.vicig@deams.units.it](mailto:paolo.vicig@deams.units.it) (P. Vicig).

of different strengths, have been proposed. *Coherence* [21] is the most investigated and closest to the precise probability case. Weaker notions are *2-coherence* [15,16,21], the condition of avoiding sure loss [21, Section 2.4] and others. Hence, a second interesting question arises: if a probability inequality can be extended in some way to imprecise uncertainty measures, what degree of consistency is required for this? Should the measure be coherent, or does some weaker concept like 2-coherence already suffice?

In this paper, we investigate how some fundamental probability inequalities generalise in an imprecise framework. We shall see that:

- (a) More inequalities may correspond to a single probability inequality.
- (b) The generalised inequalities do not require exclusively coherence, even though coherence may ensure a larger applicability or tighter bounds.

These issues have been very little investigated in the literature on imprecise probabilities. Some papers [5,6] introduce inequalities for sums of random numbers while studying laws of large numbers for coherent lower previsions, and in passing a Markov inequality is derived.

A different and more investigated line of research (see e.g. [1,9,10,17,22,23]) extends probability inequalities to various integrals, such as the Choquet integral. This approach partly overlaps with the lower/upper previsions one, as discussed in the later Section 3.1.

The paper is organised as follows. Section 2 recalls preliminary notions from the theory of Imprecise Probabilities. In Section 3 we first derive versions of Jensen's inequality (Theorem 3.2) for 2-coherent previsions while extending a result in [1]. Then, we generalise a recently proposed [2] improvement of Jensen's inequality in Section 3.2. Our result (Theorem 3.3), while assuming at least 2-coherence, is more general than that in [2] even in the special precise probability environment. In Section 4 some applications of Jensen's imprecise inequalities are discussed. Lyapunov-like inequalities are obtained for 2-coherent previsions in Section 4.1, while Section 4.2 considers some basic inferences with Jensen's inequalities, including the even moment problem. We point out that if simple gambles only are involved and the imprecise previsions are coherent, a linear programming problem may be an attractive alternative to Jensen's inequality. In Section 5 we obtain two Markov's inequalities and comment on their extent according to whether 2-coherence or coherence is assumed.

In classical probability, given  $X \geq 0$ , Markov's inequality bounds the probability of event  $(X \geq c)$ , using only the expectation  $E(X)$ . Not requiring non-negativity for  $X$  and knowing additionally the variance  $\sigma^2(X)$ , events like  $(X \leq c)$  or  $(X \geq c)$  may be bounded by Cantelli's inequalities. These are known [8] to be the sharpest ones in the family of all distributions of  $X$  with given  $E(X)$ ,  $\sigma^2(X)$ . In Section 6 we investigate how Cantelli's inequalities can be extended to imprecise evaluations. Firstly, we derive them in Section 6.1 in a precise framework, but referring to linear or dF-coherent previsions rather than (less generally) expectations. Then we obtain more forms of Cantelli's inequalities, under weak consistency assumptions, in Section 6.2 (Proposition 6.2, Corollary 6.1). In Section 6.3 we require coherence, which lets us derive stricter Cantelli's inequalities (Proposition 6.3). Interestingly, they involve lower and upper variances, introduced in [21, Appendix G]. When both can be applied, we perform a comparison between Cantelli's and Markov's inequalities, showing that neither is uniformly preferable in an imprecise framework. Section 7 contains our conclusions and ideas for further developments.

## 2. Preliminaries

Throughout the paper, we shall be concerned with bounded random variables, termed *gambles*, defined on a common *partition*  $\mathcal{P}$  of the sure event  $\Omega$ . Partition  $\mathcal{P}$  (also called possibility space or universe of discourse) is made up of an arbitrary and possibly infinite number of exhaustive and pairwise disjoint (non-impossible) events. A gamble  $X$  is (identified by) a map from  $\mathcal{P}$  into  $\mathbb{R}$  and the *image set* of  $X$  is  $Im(X) = \{X(\omega) : \omega \in \mathcal{P}\} (\subset \mathbb{R})$ . If  $f$  is a real function whose domain includes  $Im(X)$ ,  $f(X)$  is a gamble with image set  $Im(f(X)) = \{f(X(\omega)) : \omega \in \mathcal{P}\}$ .

We term  $\mathcal{A}(\mathcal{P})$  the set of events logically dependent on  $\mathcal{P}$  (the powerset of  $\mathcal{P}$  in set-theoretical language); an event  $B$  belongs to  $\mathcal{A}(\mathcal{P})$  if and only if  $B$  is the logical sum of the events  $\omega \in \mathcal{P}$  implying  $B$  ( $B = \bigvee_{\omega \Rightarrow B} \omega$ ). For a given  $B \in \mathcal{A}(\mathcal{P})$ ,  $B \neq \emptyset$ , the conditional gamble  $X|B$  is identified by the restriction of  $X$  on  $B$ . Thus, the image set of  $X|B$  is  $\{X(\omega) : \omega \in \mathcal{P}, \omega \Rightarrow B\}$ , while  $X|B$  is undefined for  $\omega \not\Rightarrow B$ . Likewise, the image set of  $f(X|B)$  is given by  $Im(f(X|B)) = \{f(X(\omega)) : \omega \in \mathcal{P}, \omega \Rightarrow B\}$ .

## 2.1. Precise previsions

Lower and upper previsions [19,21] originate from the notion of (precise) prevision, a key concept in de Finetti's subjective probability approach [7]. Basically, given an arbitrary set of gambles  $\mathcal{D}$ , a (precise) *prevision*  $P : \mathcal{D} \rightarrow \mathbb{R}$  is a mapping that associates a real number  $P(X)$  that 'synthesises'  $X$  to each  $X \in \mathcal{D}$ . To avoid inconsistent syntheses, de Finetti developed the concept of coherent prevision, which we term here *dF-coherent prevision* to prevent confusion with the same term referred to lower or upper previsions:

**Definition 2.1.** A mapping  $P : \mathcal{D} \rightarrow \mathbb{R}$  is a *dF-coherent prevision* iff,  $\forall n \in \mathbb{N}, \forall s_0, \dots, s_n \in \mathbb{R}, \forall X_0, \dots, X_n \in \mathcal{D}$ , defining  $G = \sum_{i=0}^n s_i (X_i - P(X_i))$  we have that  $\sup G \geq 0$ .

While this definition has a well known betting interpretation [7,21], in the special case that  $\mathcal{D}$  is a linear space  $\mathcal{L}$  a functional  $P$  is a dF-coherent (or linear) prevision on  $\mathcal{L}$  iff it satisfies the following axioms [21, Section 2.3.6]:

- (L)  $P(X + Y) = P(X) + P(Y), \forall X, Y \in \mathcal{L}$  (linearity)
- (I)  $\inf X \leq P(X) \leq \sup X, \forall X \in \mathcal{L}$  (internality).

These axioms imply

- (H)  $P(\lambda X) = \lambda P(X), \forall \lambda \in \mathbb{R}, X \in \mathcal{L}$  (homogeneity)

and if  $\mathcal{D}$  is not a linear space, they are necessary (but not sufficient) conditions for dF-coherence of  $P$ .

Moreover, if a finitely additive probability  $P_0$  is given on  $\mathcal{A}(P)$ , its extension to a dF-coherent prevision for  $X$  is unique and coincides with the (linear) expectation of  $X$  [21, Sections 3.2.1, 3.2.2]. Yet, an agent might assess a dF-coherent prevision for any  $X$  without knowing or knowing only partially either  $P_0$  or even the distribution function of  $X$ , and in this sense the notion of dF-coherent prevision is more general than that of expectation.

## 2.2. Imprecise previsions

When assessing a precise prevision turns out to be difficult or unreliable, an agent may replace  $P(X)$  with a lower prevision  $\underline{P}(X)$  or an upper prevision  $\overline{P}(X)$  (or both). In practice, we shall assume, as customary, that lower and upper previsions are *conjugate*, meaning that

$$\overline{P}(X) = -\underline{P}(-X). \quad (1)$$

Thus, for instance, assessing  $\underline{P}(X), \overline{P}(X^2)$  is equivalent to assessing  $\underline{P}$  on  $\{X, -X^2\}$ . In general, conjugacy lets us refer to one type of imprecise prevision only.

The most widespread notion of consistency for lower/upper previsions is *coherence*. It is defined by introducing constraints on the sign of  $s_0, \dots, s_n$  in Definition 2.1 [21, Section 2.5], or by relaxing axioms (L), (I) and (H) if  $\mathcal{D}$  is a linear space  $\mathcal{L}$ . In this latter case,  $\underline{P}$  is a coherent lower prevision on  $\mathcal{L}$  iff it satisfies [21, Section 2.5.5]

- (S)  $\underline{P}(X + Y) \geq \underline{P}(X) + \underline{P}(Y), \forall X, Y \in \mathcal{L}$  (superlinearity).
- (LI)  $\inf X \leq \underline{P}(X), \forall X \in \mathcal{L}$  (lower internality).
- (PH)  $\underline{P}(\lambda X) = \lambda \underline{P}(X), \forall X \in \mathcal{L}, \forall \lambda \geq 0$  (positive homogeneity).

For what follows, it is useful to recall the characterisation of coherence given by the following

**Theorem 2.1 (Lower Envelope Theorem).**  $\underline{P} : \mathcal{D} \rightarrow \mathbb{R}$  is a coherent lower prevision on  $\mathcal{D}$  iff there exists a (non-empty) set  $\mathcal{M}^*$  of dF-coherent previsions on  $\mathcal{D}$  such that

$$\underline{P}(X) = \inf_{P \in \mathcal{M}^*} P(X), \forall X \in \mathcal{D}. \quad (2)$$

The infimum is attained in (2) when  $\mathcal{M}^*$  is the set  $\mathcal{M} = \{P : P \text{ is dF-coherent on } \mathcal{D}, P \geq \underline{P}\}$ , termed *credal set* of  $\underline{P}$ .

The Lower Envelope Theorem gives a robustness interpretation of coherent lower previsions:  $\mathcal{M}^*$  may be viewed as a set of potential previsions, among which the agent cannot easily establish which is the ‘true’ one. The credal set  $\mathcal{M}$  is the largest such set and is weak\*-compact and convex [21, Section 3.6.1].

A dF-coherent prevision  $\underline{P}$  is a special coherent lower and upper prevision such that  $\underline{P} = \overline{P} = P$ .

A lower prevision  $\underline{P}$  *avoids sure loss* (ASL) [21, Section 3.3.3 (a)] on  $\mathcal{D}$  iff the credal set  $\mathcal{M}$  is non-empty, i.e. iff there exists a dF-coherent prevision  $P \geq \underline{P}$  on  $\mathcal{D}$ . Thus, the notion of avoiding sure loss is weaker than coherence.

2-coherence is obtained introducing constraints in Definition 2.1. Next to the additional coherence constraints on  $s_0, \dots, s_n$ , it requires that  $n = 1$ , so that  $G$  is replaced by a summation of two terms. Hence, coherence implies 2-coherence, which is a weaker consistency requirement (unless  $D$  is made of one or two gambles only, in which case they are equivalent). We report the relevant definitions [15]:

**Definition 2.2.**

- (a)  $\underline{P} : \mathcal{D} \rightarrow \mathbb{R}$  is a *2-coherent lower prevision* on  $\mathcal{D}$  iff  $\forall X_0, X_1 \in \mathcal{D}, \forall s_1 \geq 0, \forall s_0 \in \mathbb{R}$ , defining  $\underline{G}_2 = s_1(X_1 - \underline{P}(X_1)) - s_0(X_0 - \underline{P}(X_0))$ , we have that  $\sup \underline{G}_2 \geq 0$ .
- (b)  $\overline{P} : \mathcal{D} \rightarrow \mathbb{R}$  is a *2-coherent upper prevision* on  $\mathcal{D}$  iff  $\forall X_0, X_1 \in \mathcal{D}, \forall s_1 \geq 0, \forall s_0 \in \mathbb{R}$ , defining  $\overline{G}_2 = s_1(\overline{P}(X_1) - X_1) - s_0(\overline{P}(X_0) - X_0)$ , we have that  $\sup \overline{G}_2 \geq 0$ .

2-coherence too can be defined by means of axioms on structured sets. Since a limited part of the sequel will involve conditional gambles, we shall state a result in the case that  $\mathcal{D}$  is now the linear space  $\mathcal{L}|B$  of all conditional gambles  $X|B$ , where  $X \in \mathcal{L}$ , a given linear space, and  $B \neq \emptyset, B \in \mathcal{A}(P)$ , is a fixed event.

Preliminarily, consider the following axioms, referred for later usage to a generic uncertainty measure  $\mu$ :

- (M) If  $X|B \geq Y|B$  then  $\mu(X|B) \geq \mu(Y|B)$  (monotonicity).
- (T)  $\mu(X + a|B) = \mu(X|B) + a, \forall a \in \mathbb{R}$  (translation invariance).
- (PH)  $\forall \lambda \geq 0, \mu(\lambda X|B) = \lambda \mu(X|B)$  (positive homogeneity).
- (NH)  $\forall \lambda < 0, \mu(\lambda X|B) \leq \lambda \mu(X|B)$  (negative homogeneity).

Then we have, as a special case of [15, Proposition 9] (cf. also [15, Remark 1])

**Theorem 2.2.**  $\underline{P} : \mathcal{L}|B \rightarrow \mathbb{R}$  is a *2-coherent lower prevision* on  $\mathcal{L}|B$  iff axioms (M), (T), (PH), (NH) above obtain (with  $\mu = \underline{P}$ , and for all  $X|B, Y|B \in \mathcal{L}|B$ ).

**Remark 2.1.** It is easy to verify that axioms (M), (T), (PH) imply further properties of  $\underline{P}$  and its conjugate  $\overline{P}$  (cf. also [15]). These include internality (I) (with  $X$  replaced by  $X|B$ ) and  $\underline{P}(c|B) = \overline{P}(c|B) = c, \forall c \in \mathbb{R}$ .

Importantly, it follows instead from (NH) and the conjugacy equality  $\overline{P}(X|B) = -\underline{P}(-X|B)$  that

$$\underline{P}(X|B) \leq \overline{P}(X|B). \tag{3}$$

This seemingly obvious property is nevertheless not always implied by weaker consistency notions than 2-coherence.

Axioms (M), (T), (PH), (NH) are necessary conditions for 2-coherence and therefore also for coherence on a generic domain  $\mathcal{D}$ . Theorem 2.2 applies in particular when  $B = \Omega$ , i.e. when  $\underline{P}$  is an unconditional measure on the linear space  $\mathcal{L}|\Omega = \mathcal{L}$ .

A fundamental result for 2-coherent, coherent and dF-coherent previsions defined on a set  $\mathcal{D}$  is that they allow an extension of the same type (2-coherent, coherent or dF-coherent, respectively) on any superset of gambles  $\mathcal{D}' \supset \mathcal{D}$  [7,16,21]. Therefore, assessing one such prevision on an arbitrary set of gambles is not penalising: we can always extend it preserving its degree of consistency.

The extension is usually not unique, but a lower prevision  $\underline{P}$  coherent on  $\mathcal{D}$  always has a least-committal coherent extension  $\underline{E}$  on  $\mathcal{D}'$ , termed *natural extension*. This means that  $\underline{E} \leq \underline{P}'$ , for any other coherent extension  $\underline{P}'$  of  $\underline{P}$  [19,21]. In the same manner, a 2-coherent lower prevision  $\underline{P}$  always admits a 2-coherent natural extension  $\underline{E}_2$  on any  $\mathcal{D}'$ . Again,  $\underline{E}_2$  is the least-committal extension of  $\underline{P}$  among its 2-coherent ones [15,16]. We shall encounter natural extensions and 2-coherent natural extensions in Section 4.2.

A special situation occurs when a gamble is the *indicator*  $I_A$  of an event  $A$ . The prevision of an indicator is the same as the probability of the corresponding event, and these terms will be used interchangeably in this paper. If  $\mathcal{D}$  is made of (indicators of) events only, then  $\underline{P}$  ( $\overline{P}$ ) is a lower (upper) probability on  $\mathcal{D}$ , whilst  $P$  is a dF-coherent probability. In this case, the conjugacy equality (1) is written  $\overline{P}(A) = 1 - \underline{P}(A^c)$ .

If the set of events  $\mathcal{D}$  is an algebra, a 2-coherent  $\underline{P}$  (or its conjugate  $\overline{P}$ ) is a (normalised) *capacity* or *fuzzy measure*, with the additional property (3). Coherent lower/upper previsions include a number of models as special cases, such as belief functions, possibility measures [19], coherent risk measures [14], several neighbourhood models [4,11,12], and many others. For further information on the material recalled in this section, see among others [15,16,19,21].

### 3. Jensen's and related inequalities

Various papers discuss extensions of Jensen's inequality beyond classical probabilities, among them [9,10,23]. Often, they make use of Choquet, Sugeno or other integrals. The work in [1] is the closest to our approach to Jensen's inequality with imprecise previsions, since it requires very general conditions. We start with a modified and conditional version of [1, Theorem 2]. For this, let  $f'_+(x)$ ,  $f'_-(x)$  be the right and left derivatives, respectively, of a real function  $f$  at  $x$ . It is well known that if  $f$  is convex on an interval then we have, at any two interior points  $x < y$  of the interval:

$$-\infty < f'_-(x) \leq f'_+(x) \leq f'_-(y) \leq f'_+(y) < +\infty; \quad (4)$$

the inequalities in (4) are reversed when  $f$  is concave (see e.g. [13, Theorem 1.4.2]).

**Theorem 3.1.** *Let  $X|B$  be a gamble,  $I \subset \mathbb{R}$  an interval that contains the image set of  $X|B$ ,  $\phi : I \rightarrow \mathbb{R}$  a convex function,  $\psi : I \rightarrow \mathbb{R}$  a concave function,  $X|B \in \mathcal{D}$ .<sup>1</sup> Further, let  $\mu : \mathcal{D} \rightarrow \mathbb{R}$  be an uncertainty measure satisfying axioms (M), (T), (PH) and  $\mu^c$  its conjugate (i.e.  $\mu^c(Y|B) = -\mu(-Y|B)$ ). Let also  $\mu(X|B)$  be an interior point of  $I$ .*

(a) Let  $J_\phi = [\phi'_-(\mu(X|B)), \phi'_+(\mu(X|B))]$ .

$$\text{If } \exists \lambda \in J_\phi, \lambda \geq 0 \text{ then } \mu(\phi(X|B)) \geq \phi(\mu(X|B)); \quad (5)$$

$$\text{If } \exists \lambda \in J_\phi, \lambda \leq 0 \text{ then } \mu^c(\phi(X|B)) \geq \phi(\mu(X|B)). \quad (6)$$

(b) Let  $J_\psi = [\psi'_+(\mu(X|B)), \psi'_-(\mu(X|B))]$ .

$$\text{If } \exists \lambda \in J_\psi, \lambda \geq 0 \text{ then } \mu(\psi(X|B)) \leq \psi(\mu(X|B)); \quad (7)$$

$$\text{If } \exists \lambda \in J_\psi, \lambda \leq 0 \text{ then } \mu^c(\psi(X|B)) \leq \psi(\mu(X|B)). \quad (8)$$

**Proof.** The proofs of (a) and (b) are similar. Let us prove (b). Recall for this that, since  $\psi$  is concave, it holds, for any  $x_0$  interior point of  $I$ , that

$$\psi(x) \leq \psi(x_0) + \lambda(x - x_0), \forall \lambda \in [\psi'_+(x_0), \psi'_-(x_0)]. \quad (9)$$

When  $x_0 = \mu(X|B)$ ,  $x = X|B$ , Equation (9) becomes

$$\psi(X|B) \leq \psi(\mu(X|B)) + \lambda(X|B - \mu(X|B)). \quad (10)$$

- If  $\lambda \geq 0$ , by axioms (M), (T), (PH), we get from (10)

$$\mu(\psi(X|B)) \leq \psi(\mu(X|B)) + \lambda(\mu(X|B) - \mu(X|B)) = \psi(\mu(X|B)).$$

- If  $\lambda \leq 0$ , rewrite (10) as

$$-\lambda(X|B - \mu(X|B)) \leq \psi(\mu(X|B)) - \psi(X|B).$$

<sup>1</sup> Here (and similarly in the next results) one may assume that  $\mathcal{D}$  contains all the gambles  $\mu$  is applied to in the proof. Alternatively,  $\mathcal{D} = \{X|B\}$  with the proviso that  $\mu$  is extended to any gamble in the proof according to axioms (M), (T), (PH).

Now apply  $(M)$ ,  $(T)$ ,  $(PH)$ :

$$\mu(-\lambda(X|B - \mu(X|B))) = -\lambda(\mu(X|B) - \mu(X|B)) = 0 \leq \psi(\mu(X|B)) + \mu(-\psi(X|B)).$$

Hence,

$$-\mu(-\psi(X|B)) = \mu^c(\psi(X|B)) \leq \psi(\mu(X|B)). \quad \square$$

Now, a key point for the sequel is that Theorem 3.1 can be applied to 2-coherent lower and upper previsions, since axioms  $(M)$ ,  $(T)$ ,  $(PH)$  obtain for them, cf. Theorem 2.2. Recalling the same Theorem 2.2 and Remark 2.1, it appears clearly that 2-coherence is the weakest consistency notion for imprecise previsions that Theorem 3.1 can be reasonably applied to. In fact, 2-coherence requires, in addition to the axioms in Theorem 3.1, only axiom (NH), necessary to guarantee the very desirable property (3).

In classical probability theory, the commonest versions of Jensen's inequality refer to an unconditional environment. We too shall derive now their unconditional imprecise counterparts. For this, put  $B = \Omega$  in Theorem 3.1, so that  $X|B = X|\Omega = X$ .

**Theorem 3.2** (*Jensen's inequalities*). *In the assumptions of Theorem 3.1, let further  $B = \Omega$ ,  $\{\mu, \mu^c\} = \{\underline{P}, \overline{P}\}$ , with  $\underline{P}, \overline{P}$  2-coherent on their domain. Then,*

$$\underline{P}(\psi(X)) \leq \min\{\psi(\underline{P}(X)), \psi(\overline{P}(X))\}, \quad (11)$$

$$\overline{P}(\phi(X)) \geq \max\{\phi(\underline{P}(X)), \phi(\overline{P}(X))\}. \quad (12)$$

Besides,

$$\text{if } \phi'_+(\underline{P}(X)) \geq 0 \text{ then } \underline{P}(\phi(X)) \geq \phi(\underline{P}(X)), \quad (13)$$

$$\text{if } \phi'_-(\overline{P}(X)) \leq 0 \text{ then } \underline{P}(\phi(X)) \geq \phi(\overline{P}(X));$$

$$\text{if } \psi'_-(\overline{P}(X)) \geq 0 \text{ then } \overline{P}(\psi(X)) \leq \psi(\overline{P}(X)), \quad (14)$$

$$\text{if } \psi'_+(\underline{P}(X)) \leq 0 \text{ then } \overline{P}(\psi(X)) \leq \psi(\underline{P}(X)).$$

**Proof.** To prove (11), put  $\mu = \underline{P}$  in (7) and (8) getting, respectively

$$\underline{P}(\psi(X)) \leq \psi(\underline{P}(X)), \text{ if } \lambda \geq 0,$$

$$\underline{P}(\psi(X)) \leq \overline{P}(\psi(X)) \leq \psi(\underline{P}(X)), \text{ if } \lambda \leq 0.$$

Hence, no matter which is the sign of  $\lambda$ , it holds that

$$\underline{P}(\psi(X)) \leq \psi(\underline{P}(X)). \quad (15)$$

Putting instead  $\mu^c = \underline{P}$  in (7) and (8) we come to the inequalities, respectively

$$\underline{P}(\psi(X)) \leq \overline{P}(\psi(X)) \leq \psi(\overline{P}(X)), \text{ if } \lambda \geq 0,$$

$$\underline{P}(\psi(X)) \leq \psi(\overline{P}(X)), \text{ if } \lambda \leq 0,$$

and hence we conclude that

$$\underline{P}(\psi(X)) \leq \psi(\overline{P}(X)). \quad (16)$$

Grouping (15) and (16) we obtain (11).

We prove now (13). Put  $\mu = \underline{P}$  in (5) and note that condition  $\phi'_+(\underline{P}(X)) \geq 0$  is equivalent (cf. (4)) to requiring that there is  $\lambda \in J_\phi, \lambda \geq 0$ . Then, (5) implies  $\underline{P}(\phi(X)) \geq \phi(\underline{P}(X))$ .

Similarly, when  $\phi'_-(\overline{P}(X)) \leq 0$ , apply (6) with  $\mu = \overline{P}$  to obtain  $\underline{P}(\phi(X)) \geq \phi(\overline{P}(X))$ .

The proofs of (12) and (14) are analogous to those of (11) and (13), respectively.  $\square$

**Remark 3.1.** Theorem 3.2 improves over Theorem 3.1 in that its inequalities (11), (12) are *independent* of  $\lambda$ . With different consistency notions this kind of independence is no longer ensured.

It is important to observe that an additional *monotonicity hypothesis* of either  $\phi(x)$  or  $\psi(x)$  on  $[\underline{P}(X), \overline{P}(X)]$  simplifies and enhances Jensen's inequalities. Monotonicity of  $\phi$  solves easily the maximum in (12) and ensures that one of the equations (13) always applies. Similarly with  $\psi$ .

Another notable simplification occurs when  $\underline{P} = \overline{P} = P$ , and  $P$  is a dF-coherent prevision:

**Corollary 3.1.** *Given  $X, I, \phi, \psi$  as in Theorem 3.1,  $B = \Omega$ , let  $P : \mathcal{D} \rightarrow \mathbb{R}$  be a dF-coherent prevision. Then,*

$$P(\phi(X)) \geq \phi(P(X)), \quad P(\psi(X)) \leq \psi(P(X)). \quad (17)$$

**Proof.** Since  $P$  is also 2-coherent, Theorem 3.2 applies. Substituting  $\underline{P} = \overline{P} = P$  in (11) and (12) we get Equation (17).  $\square$

Recall that, for a gamble  $X$ , its expectation  $E(X)$  is a coherent prevision. Therefore, we may replace  $P$  with  $E$  in (17) to obtain the familiar Jensen's inequality for convex or concave functions of  $X$ :

$$E(\phi(X)) \geq \phi(E(X)), \quad E(\psi(X)) \leq \psi(E(X)). \quad (18)$$

In this paper, we mostly focus on inequalities in an unconditional setting. Yet, Theorem 3.1 can be exploited to derive a conditional version of Theorem 3.2 and Corollary 3.1, with simple modifications in their proofs. We restrict ourselves to stating the conditional Corollary 3.1, since it will be needed in the proof of the later Lemma 6.1.

**Corollary 3.2.** *Given  $X|B, I, \phi, \psi$  as in Theorem 3.1, let  $P : \mathcal{D} \rightarrow \mathbb{R}$  be a dF-coherent prevision. Then,*

$$P(\phi(X|B)) \geq \phi(P(X|B)), \quad P(\psi(X|B)) \leq \psi(P(X|B)). \quad (19)$$

### 3.1. Jensen's inequalities with and without integrals

In classical probability theory, the expectations in Jensen's inequalities (18) are written as (Lebesgue) integrals. Quite naturally then, various papers, including [9,10,17,22,23], investigated if and how Jensen's inequality can be extended to other types of integrals.

As a common feature, all these approaches require an uncertainty measure  $\mu$  to be preliminarily assessed on a structured subset of the powerset  $\mathcal{A}(\mathcal{P})$ , such as an algebra that may be  $\mathcal{A}(\mathcal{P})$  itself. The measure  $\mu$  is at least a capacity (monotone, and  $\mu(\emptyset) = 0$ ), usually normalised ( $\mu(\Omega) = 1$ ), but more restrictive conditions are sometimes assumed (for instance, lower and upper semicontinuity in [17]). Then,  $\mu$  is extended to a set of gambles  $\mathcal{D}$  by means of the chosen integral.

By contrast, much weaker assumptions are needed with the imprecise previsions framework of this paper: the measure  $\mu$  may be defined *only* on the set  $\mathcal{D}$  of gambles. Moreover,  $\mu$  and its conjugate  $\mu^c$  are 2-coherent lower and upper previsions, so that  $\{\mu, \mu^c\} = \{\underline{P}, \overline{P}\}$ . 2-coherence establishes a connection with the integral approach in the following sense: whenever  $\underline{P}$  or  $\overline{P}$  is (2-coherently) extended to  $\mathcal{D} \cup \mathcal{A}(\mathcal{P})$  (which is anyway not needed at all in our framework), the extension is a capacity on  $\mathcal{A}(\mathcal{P})$ . Our only extra assumption is the very reasonable condition  $\underline{P}(\cdot) \leq \overline{P}(\cdot)$ .

Studying Jensen's inequality by means of integrals brought to non-homogeneous results. It is shown in [9, Example 2.4] that recovering faithfully the structure of Equations (18) is not generally feasible with the Choquet, Sugeno or Shilkret integrals. With assumptions overlapping those in Theorem 3.1, but referring to the Choquet integral  $(C) \int X d\mu = \mu(X)$ , [9] obtains the inequality

$$\phi(\mu(X)) = \phi\left((C) \int X d\mu\right) \leq \max\left\{(C) \int \phi(X) d\mu, (C) \int \phi(X) d\mu^c\right\}. \quad (20)$$

In [23], a Jensen-type inequality for the Choquet integral in [22] is amended by requiring monotonicity of  $\phi(x)$ , cf. also [10] for analogous results under finiteness of  $\mathcal{P}$  and non-negativity of gambles.

The classical Choquet integral satisfies axioms (M), (T), (PH) (see e.g. [19, Appendix C]), hence these results are implied by Theorem 3.1. Equation (20) follows immediately from (5) and (6), Jensen's inequality in [23, Theorem 2.1] from (5).

Other approaches, while starting from different hypotheses, are not directly comparable with ours. In particular, [17] obtains results with the Sugeno integral removing the convexity assumption on  $\phi(x)$ , while requiring its monotonicity and some further conditions.

### 3.2. Improving Jensen's inequalities

In a recent paper [2], an improvement to the classical Jensen's inequality (18) has been proposed in the case that  $X$  takes values in  $\mathbb{Z}$ . We investigate its extension to 2-coherent imprecise previsions under more general requirements for  $X$ . Roughly speaking, it suffices that the image set of  $X$  has some 'hole', and that  $\underline{P}(X)$  or  $\overline{P}(X)$  'falls into' one such hole. Let us formalise:

**Definition 3.1.** Given a gamble  $X$  with image set  $Im(X)$ , define for  $k \in [\inf X, \sup X]$

$$l(k) = \sup\{x \in Im(X) : x \leq k\}$$

$$u(k) = \inf\{x \in Im(X) : x \geq k\}.$$

**Theorem 3.3.** Given the gamble  $X \in \mathcal{D}$ , the interval  $I = [\inf X, \sup X]$ ,  $\phi : I \rightarrow \mathbb{R}$  convex function,  $\psi : I \rightarrow \mathbb{R}$  concave function, let  $\underline{P} : \mathcal{D} \rightarrow \mathbb{R}$  be 2-coherent, with  $\overline{P}$  its conjugate. Define

$$x_L = l(\underline{P}(X)), x_U = u(\underline{P}(X)), z_L = l(\overline{P}(X)), z_U = u(\overline{P}(X)).$$

If  $\inf X \leq x_L < x_U \leq \sup X$  (in items (a1), (b2)) and  $\inf X \leq z_L < z_U \leq \sup X$  (in items (a2), (b1)), then

$$(a1) \quad \begin{aligned} \underline{P}(\psi(X)) &\leq \psi(x_U) \frac{\underline{P}(X) - x_L}{x_U - x_L} + \psi(x_L) \left(1 - \frac{\underline{P}(X) - x_L}{x_U - x_L}\right) \\ &\leq \psi(\underline{P}(X)). \end{aligned} \quad (21)$$

(a2) If  $\psi(z_L) \leq \psi(z_U)$ ,

$$(22) \quad \begin{aligned} \overline{P}(\psi(X)) &\leq \psi(z_U) \frac{\overline{P}(X) - z_L}{z_U - z_L} + \psi(z_L) \left(1 - \frac{\overline{P}(X) - z_L}{z_U - z_L}\right) \\ &\leq \psi(\overline{P}(X)). \end{aligned}$$

(b1)

$$(23) \quad \begin{aligned} \overline{P}(\phi(X)) &\geq \phi(z_U) \frac{\overline{P}(X) - z_L}{z_U - z_L} + \phi(z_L) \left(1 - \frac{\overline{P}(X) - z_L}{z_U - z_L}\right) \\ &\geq \phi(\overline{P}(X)). \end{aligned}$$

(b2) If  $\phi(x_L) \leq \phi(x_U)$ ,

$$(24) \quad \begin{aligned} \underline{P}(\phi(X)) &\geq \phi(x_U) \frac{\underline{P}(X) - x_L}{x_U - x_L} + \phi(x_L) \left(1 - \frac{\underline{P}(X) - x_L}{x_U - x_L}\right) \\ &\geq \phi(\underline{P}(X)). \end{aligned}$$

**Proof.** We prove (a1). Let  $y = s(x)$  be the straight line joining  $(x_L, \psi(x_L))$ ,  $(x_U, \psi(x_U))$ ,

$$s(x) = \frac{\psi(x_U) - \psi(x_L)}{x_U - x_L} (x - x_L) + \psi(x_L). \quad (25)$$



Since  $\psi$  is concave, it holds that  $\psi(x) \leq s(x)$  for both  $x \leq x_L$  and  $x \geq x_U$ .<sup>2</sup> On the other hand,  $X$  takes no value in  $]x_L, x_U[$ , thus it is true that

$$\psi(X) \leq s(X). \quad (26)$$

From (26), we obtain (21) through the derivation that follows. Its first inequality is due to monotonicity of  $\underline{P}$ , the second to (25), (T) and to axiom (PH) if  $\psi(x_U) \geq \psi(x_L)$ , to (NH) otherwise,<sup>3</sup> the third to concavity of  $\psi$ :

$$\begin{aligned} \underline{P}(\psi(X)) &\leq \underline{P}(s(X)) \leq \frac{\psi(x_U) - \psi(x_L)}{x_U - x_L} (\underline{P}(X) - x_L) + \psi(x_L) \\ &= \psi(x_U) \frac{\underline{P}(X) - x_L}{x_U - x_L} + \psi(x_L) \left( 1 - \frac{\underline{P}(X) - x_L}{x_U - x_L} \right) \leq \psi(\underline{P}(X)). \end{aligned}$$

To prove (a2), follow the same course of reasoning as in (a1), replacing  $x_L, x_U, \underline{P}$  with  $z_L, z_U, \overline{P}$  and applying (PH) at the second inequality<sup>4</sup> of the final derivation.

The proof of (b1), (b2) is analogous to (a1), (a2) respectively.  $\square$

How can Theorem 3.3 be exploited, operationally? To exemplify, let us discuss (a1) and (a2). Note that the hypothesis  $\inf X \leq x_L < x_U \leq \sup X$

(a) is compatible with the assessments  $\underline{P}(X) = x_L > \inf X$  or  $\underline{P}(X) = x_U < \sup X$ .

Besides,

(b) the values  $\underline{P}(X) = x_L$  or  $\underline{P}(X) = x_U$  make (21) useless. In fact, in this case, (21) boils down to Equation (15),  $\underline{P}(\psi(X)) \leq \psi(\underline{P}(X))$ , and cannot improve any of Jensen's inequalities.

Condition  $\inf X \leq z_L < z_U \leq \sup X$  plays the same role with  $\overline{P}$  in (a2) as for (a), but not for (b) as far as Equations (14) are concerned, as will appear shortly.

Items (a1) and (a2) may tighten Jensen's inequalities (11) and (14). To see this, let us label the right-hand terms of the inequalities (21) and (22):

$$\begin{aligned} M_1 &= \psi(x_U) \frac{\underline{P}(X) - x_L}{x_U - x_L} + \psi(x_L) \left( 1 - \frac{\underline{P}(X) - x_L}{x_U - x_L} \right) \\ M_2 &= \psi(z_U) \frac{\overline{P}(X) - z_L}{z_U - z_L} + \psi(z_L) \left( 1 - \frac{\overline{P}(X) - z_L}{z_U - z_L} \right) \end{aligned}$$

Consider inequality (11). By (a1), we may replace it with

$$\underline{P}(\psi(X)) \leq \min\{M_1, \psi(\overline{P}(X))\}, \quad (27)$$

since  $M_1 \leq \psi(\underline{P}(X))$ . Note that if  $\psi$  is monotone non-decreasing on  $[\underline{P}(X), \overline{P}(X)]$ , or just when  $\psi(\underline{P}(X)) \leq \psi(\overline{P}(X))$ , (27) simplifies to  $\underline{P}(\psi(X)) \leq M_1$ .

When (a2) is applicable, we have that

$$\underline{P}(\psi(X)) \leq \overline{P}(\psi(X)) \leq M_2 \leq \psi(\overline{P}(X)). \quad (28)$$

This brings to the next tightening of (27) and hence of inequality (11):

$$\underline{P}(\psi(X)) \leq \min\{M_1, M_2\}.$$

<sup>2</sup> This standard result can be derived for instance using the *three chords inequality* for concave functions (cf. [13, Equation (1.16)]).

<sup>3</sup> Note that when  $\psi(x_U) \geq \psi(x_L)$ , the second inequality is an equality.

<sup>4</sup> This is the only asymmetry with the proof of (a1): when  $\psi(z_L) > \psi(z_U)$ , we cannot apply the upper prevision version of (NH)  $\overline{P}(\lambda X) \geq \lambda \overline{P}(X)$ , because the inequality direction is opposite to that of the majorisation we would need.

Now take inequalities (14). Here the extent of Theorem 3.3 is more limited: if (a2) can be applied, from (28), the first inequality (14) can be replaced by  $\overline{P}(\psi(X)) \leq M_2$ , i.e. by (22).

Note that (22) may apply, while the first inequality (14) can not, when  $\psi'_-(\overline{P}(X)) < 0$ . Since then  $\psi'_+(\overline{P}(X)) < 0$  (cf. (4)), using the second inequality (14) we obtain in this case that

$$\overline{P}(\psi(X)) \leq \min\{M_2, \psi(\underline{P}(X))\},$$

structurally similar to (27).

The role of (b1), (b2) in improving Jensen's inequalities (12), (13) is analogous.

Finally, when  $\underline{P} = \overline{P} = P$ ,  $P$  dF-coherent, (23) and (24) boil down to a single inequality, like (21) and (22). They can be applied to possibly improve Jensen's inequalities (17).

The special case of Theorem 3.3 established in [2] is obtained assuming additionally that an expectation  $E(X)$  is given, that  $Im(X) \subset \mathbb{Z}$ , and (whenever an improvement over Jensen's inequality can be achieved) that  $E(X)$  is non-integer. In particular, in [2]  $l(E(X))$  is set equal to the floor value  $\lfloor E(X) \rfloor$ ,  $u(E(X))$  to  $\lceil E(X) \rceil = \lfloor E(X) \rfloor + 1$ , so that  $x_U - x_L = z_U - z_L = 1$ .

## 4. Applying Jensen's inequalities

### 4.1. Lyapunov's inequalities

In classical probability theory, *Lyapunov's inequality* ensures that (see e.g. [18])

$$\text{for } 0 < s < t, [E(|X|^s)]^{\frac{1}{s}} \leq [E(|X|^t)]^{\frac{1}{t}},$$

provided that the expectations above are finite. The result may be obtained from Jensen's inequality [18]. Thus, a natural question is whether Lyapunov-like inequalities can be derived in Imprecise Probability theory from (some of) Jensen's inequalities (11)–(14). The next proposition gives an affirmative answer.

**Proposition 4.1.** *Let  $\underline{P}$  be 2-coherent wherever it is defined,  $\overline{P}$  being its conjugate. Letting  $X$  be a gamble and  $0 < s < t$ , we have that*

$$(a) [\underline{P}(|X|^s)]^{\frac{1}{s}} \leq [\overline{P}(|X|^s)]^{\frac{1}{s}} \leq [\overline{P}(|X|^t)]^{\frac{1}{t}}. \quad (29)$$

$$(b) \text{ If } X \geq 0, \text{ it holds that } [\underline{P}(X^s)]^{\frac{1}{s}} \leq [\underline{P}(X^t)]^{\frac{1}{t}}. \quad (30)$$

**Proof.** (a) The first inequality in (29) derives from (M) and the property (3) of 2-coherence, ensuring that  $0 \leq \underline{P}(|X|^s) \leq \overline{P}(|X|^s)$ , after raising its terms to  $\frac{1}{s}$ .

To prove that

$$[\overline{P}(|X|^s)]^{\frac{1}{s}} \leq [\overline{P}(|X|^t)]^{\frac{1}{t}} \quad (31)$$

let  $r = \frac{t}{s} > 1$ ,  $\phi(x) = |x|^r$ .

Noting that  $\phi$  is a convex function and making use of (12) at the second inequality, we obtain:

$$0 \leq |\overline{P}(Y)|^r = \phi(\overline{P}(Y)) \leq \overline{P}(\phi(Y)) = \overline{P}(|Y|^r). \quad (32)$$

Now, putting  $Y = |X|^s$  and replacing  $r$  in (32),

$$|\overline{P}(Y)|^r = (\overline{P}(|X|^s))^{\frac{t}{s}} \leq \overline{P}(|X|^{s \cdot \frac{t}{s}}) = \overline{P}(|X|^t),$$

implying (31).

(b) Define  $r$  as in (a), and the convex function  $\phi(x) = x^r$ . Clearly,  $\phi(x)$  is increasing for  $x \geq 0$ . Because of this, and since  $\overline{P}(Y) \geq \underline{P}(Y) \geq 0$  for any  $Y \geq 0$ , we have that

$$(\underline{P}(Y))^r = \phi(\underline{P}(Y)) \leq \underline{P}(\phi(Y)) = \underline{P}(Y^r),$$

where the inequality is due to the first of (13). The proof continues now like that of (a): take  $Y = X^s$  and replace  $r = \frac{t}{s}$  in the last derivation to obtain  $(\underline{P}(X^s))^{\frac{t}{s}} \leq \underline{P}(X^t)$ , and then (30).  $\square$

We observe that again a probability inequality is split into more imprecise probability inequalities and that the weak requirement of 2-coherence is enough to obtain them.

As a by-product of Lyapunov's inequalities we have:

**Corollary 4.1.** *If  $X \geq 0$ , then for 2-coherent and conjugate  $\underline{P}, \overline{P}$  we have:*

$$[\underline{P}(X)]^2 \leq \underline{P}(X^2), [\overline{P}(X)]^2 \leq \overline{P}(X^2). \quad (33)$$

**Proof.** Put  $s = 1, t = 2$  in Proposition 4.1. This specialises (30) into  $(0 \leq) \underline{P}(X) \leq [\underline{P}(X^2)]^{\frac{1}{2}}$ . Squaring gives the first inequality in (33). Taking instead the second inequality in (29) as a starting point we obtain the remaining inequality in (33).  $\square$

Note that inequalities (33) in the form  $\mu(X^2) - [\mu(X)]^2 \geq 0$  ( $\mu = \underline{P}$  or alternatively  $\mu = \overline{P}$ ) appear to be an imprecise probability generalisation of the variance property  $E(X^2) - E^2(X) = \sigma^2(X) \geq 0$ .

#### 4.2. Basic inferences with Jensen's inequalities

The most natural *basic inferential problem* a Jensen's inequality can be applied to is the following: a *coherent* lower (upper) prevision  $\underline{P}$  ( $\overline{P}$ ) is given on  $\{X\}$  and some bound is sought for its *coherent* extensions to a new gamble  $Y$ , which is a convex or concave function of  $X$ . We point out that an agent's uncertainty information may be very essential here and limited to just  $\underline{P}(X)$  or  $\overline{P}(X)$ . For instance, these values may have been communicated by an expert, without further details on how they were obtained.

This problem includes the imprecise version of the *even moment problem*, where  $Y = X^{2n}, n \in \mathbb{N}^+$ . In fact, since  $Y = \phi(X)$ , with  $\phi(x) = x^{2n}$  being a convex function, if  $\underline{P}(X)$  and  $\overline{P}(X)$  are both assessed Equation (12) returns a lower bound on  $\overline{P}(X^{2n})$ .

Some further result is possible under additional hypotheses. For instance, assessing only  $\underline{P}(X)$  may be enough if  $X \geq 0$ , since then  $\phi$  is increasing and the first inequality in (13) ensures that  $\underline{P}(Y) \geq (\underline{P}(X))^{2n}$ . Still while  $X \geq 0$ , the *odd moment problem* with  $Y = X^{2n+1}, n \in \mathbb{N}^+$ , can be tackled too applying (13), giving  $\underline{P}(Y) \geq (\underline{P}(X))^{2n+1}$ .

Returning to the general *basic inferential problem*, an interesting question is: how useful or convenient is answering it via Jensen's inequalities, even when monotonicity of  $\phi$  or  $\psi$  simplifies them? The answer depends on how difficult it is to find an extension of  $\underline{P}$  or  $\overline{P}$  to  $Y$  with alternative methods.

When  $X$  is a *simple* gamble, hence taking finitely many values  $x_1, x_2, \dots, x_n$ , we may detect the *natural extension*  $\underline{E}(Y)$ , i.e. the smallest coherent extension of  $\underline{P}$  to  $Y$ , via linear programming (LP). The LP problem to solve is<sup>5</sup>

$$\begin{aligned} \text{minimize} \quad & E_P(Y) = \sum_{i=1}^n p_i \phi(x_i) \\ \text{subject to} \quad & \sum_{i=1}^n p_i x_i = E_P(X) \geq \underline{P}(X), \\ & \sum_{i=1}^n p_i = 1, \quad p_i \geq 0 \quad (i = 1, \dots, n). \end{aligned} \quad (34)$$

Its unknowns are  $p_1, \dots, p_n$ , with  $p_i$  probability of  $(X = x_i)$ . The LP problem determines  $\underline{E}(Y)$  as the minimum expectation  $E_P(Y)$  for  $Y = \phi(X)$ , among all probabilities  $P$  that guarantee the dominance condition  $E_P(X) \geq \underline{P}(X)$ . The fact that the LP solution always exists and is the natural extension  $\underline{E}(Y)$  relies on well known properties of the natural extension (cf. [21, Sections 3.1.1, 3.3.3, 3.4.1]). We will not detail the explanation, since this procedure is tangential in this paper.

**Example 4.1.** Consider the partition  $\mathcal{P} = \{\omega_1, \omega_2, \omega_3\}$  and the gamble  $X$  defined by  $X(\omega_1) = x_1 = -1, X(\omega_2) = x_2 = 1, X(\omega_3) = x_3 = 2$ . Next to this, we only know that  $\underline{P}(X) = 0.75$ .

<sup>5</sup> To exemplify we consider function  $\phi$ . It could be replaced by  $\psi$ .

We may compute the natural extension  $\underline{E}(X^2)$  of  $\underline{P}$  to  $\phi(X) = X^2$  solving the LP problem (34), i.e.

$$\begin{aligned} & \text{minimize} && E_P(X^2) = p_1 + p_2 + 4p_3 \\ & \text{subject to} && -p_1 + p_2 + 2p_3 \geq 0.75, \\ & && p_1 + p_2 + p_3 = 1, \quad p_i \geq 0 \quad (i = 1, 2, 3). \end{aligned}$$

The set of optimal solutions is made of the triples  $(1 - p_2, p_2, 0)$ , with  $p_2 \in [\frac{7}{8}, 1]$ . Consequently,  $\underline{E}(X^2) = 1 - p_2 + p_2 + 0 = 1$ .

We may alternatively apply the first Jensen's inequality (13). This gives the bound  $\underline{E}(X^2) \geq (\underline{P}(X))^2 = 0.5625$ .

It is also possible to apply equation (24), since  $x_L = -1$ ,  $x_U = 1$  and  $\phi(x_L) = \phi(x_U) = 1$ . We obtain the (best) bound  $\underline{E}(X^2) \geq 1$ .

Note that, in the LP approach we are discussing, we need only assessing  $\underline{P}$  on  $X$  as an input, while Jensen's inequalities may require knowing also  $\overline{P}(X)$ . Moreover, after performing the LP problem we obtain a lower prevision on two gambles only,  $X$  and  $Y$ . Therefore coherence and 2-coherence are not distinguishable here, and the LP problem detects also the 2-coherent natural extension  $\underline{E}_2$ . In fact, in this case  $\underline{E}_2 = \underline{E}$ , as appears from their definitions in [16, Definition 4].

If, more generally,  $\underline{P}$  is given on a finite set  $\mathcal{D}$  of simple gambles (including  $X$ , but not  $Y$ ), we can still find out  $\underline{E}(Y)$  via linear programming. It is necessary that  $\underline{P}$  is coherent or at least avoids sure loss on  $\mathcal{D}$ . Given this, it suffices to apply, in the new LP problem, the dominance constraint (34) to any gamble in  $\mathcal{D}$ . However, we generally no longer obtain  $\underline{E}_2(Y)$  in this way, but only an upper bound. In fact, it is known that  $\underline{E}_2 \leq \underline{E}$  [16, Lemma 1].

Clearly, the LP alternative does not apply to non-simple gambles. If our beliefs are encoded by some specific models, (relatively) simple formulae for computing  $\underline{E}(Y)$  may be available. This requires anyway knowing more than just  $\underline{P}(X)$  or  $\overline{P}(X)$ . For instance, let a 2-monotone lower probability be assessed on the powerset  $\mathcal{A}(\mathcal{P})$  of partition  $\mathcal{P}$ . Then, the natural extension  $\underline{E}(Y)$  is obtained as a Choquet integral [19]. In general, however, it is precisely with non-simple gambles that Jensen's inequalities are operationally most useful. The improved inequalities of Theorem 3.3 are instead useless if, for instance, the image of  $X$  is an interval, but can be helpful in other situations (including the basic problem, with  $X$  simple).

## 5. Markov's inequalities

Markov's inequality, very well known in classical probability theory, operates under rather mild requirements: if  $X$  is a non-negative random number and its expectation  $E(X)$  exists, then

$$\forall a > 0, P(X \geq a) \leq \frac{E(X)}{a}.$$

The interesting fact we are going to point out is that Markov's inequality does not really depend on our employing precise probabilities, but is still valid with much weaker uncertainty measures such as 2-coherent imprecise previsions. This ensues from the next proposition.

**Proposition 5.1** (Markov's inequalities). *Let  $X$  be a non-negative gamble,  $X \geq 0$ .*

(a) *Let  $\underline{P}$  be a 2-coherent lower prevision on  $\{X, (X \geq a)\}$ . Then,*

$$\underline{P}(X \geq a) \leq \frac{\underline{P}(X)}{a}, \forall a > 0 \text{ (Lower Markov Inequality)}. \quad (35)$$

(b) *Let  $\overline{P}$  be a 2-coherent upper prevision on  $\{X, (X \geq a)\}$ . Then,*

$$\overline{P}(X \geq a) \leq \frac{\overline{P}(X)}{a}, \forall a > 0 \text{ (Upper Markov Inequality)}. \quad (36)$$

**Proof.** (a) Apply Definition 2.2 (a) with  $s_1 = a$ ,  $s_0 = 1$ ,  $X_1 = I_{(X \geq a)}$ ,  $X_0 = X$ . We obtain

$$\underline{G}_2 = a(I_{(X \geq a)} - \underline{P}(X \geq a)) - (X - \underline{P}(X)) = \underline{P}(X) - a\underline{P}(X \geq a) + aI_{(X \geq a)} - X.$$

Note that  $aI_{(X \geq a)} - X \leq 0$ . Therefore, it is necessary for the 2-coherence condition  $\sup G_2 \geq 0$  that  $\underline{P}(X) - a\underline{P}(X \geq a) \geq 0$ , which is the Lower Markov Inequality.

(b) The proof is analogous to (a), using Definition 2.2 (b) with  $s_1 = 1, s_0 = a, X_1 = X, X_0 = I_{(X \geq a)}$ .  $\square$

Obviously, Markov's inequalities apply to coherent upper/lower previsions too. Yet, deducing that working with coherence or 2-coherence is always indifferent as far as Markov's inequalities are concerned would be erroneous. In fact, coherent imprecise previsions satisfy a greater number of properties. This makes the use of Markov's inequalities possible for them in more general situations than with 2-coherence. See for this the next example.

**Example 5.1.** Let  $X, Y, Z$  be non-negative gambles such that  $\overline{P}(X) = 60, \overline{P}(Y) = 50, \overline{P}(Z) = 10$ . If  $\overline{P}$  is coherent, we can get a bound to  $\overline{P}(X + Y + Z \geq 150)$  with the following derivation, making use of the Upper Markov's inequality (36) firstly, and then of the sublinearity property of coherent upper previsions [21, Section 2.6.1 (m)]:

$$\overline{P}(X + Y + Z \geq 150) \leq \frac{\overline{P}(X + Y + Z)}{150} \leq \frac{\overline{P}(X) + \overline{P}(Y) + \overline{P}(Z)}{150} = \frac{4}{5}.$$

However, sublinearity is not guaranteed with 2-coherence, which does not therefore allow the above computation.

## 6. Cantelli's inequalities

Markov's inequality is probably the best known instance of one-sided inequality, bounding the probability of  $(X \geq a)$ . Further known inequalities of this kind assume knowledge of some higher order moments.

In a form investigated among others by Uspensky [20, p. 198], Cantelli's inequalities<sup>6</sup> apply to a random number  $X$  with expectation  $E(X) = 0$  and variance  $\sigma^2 = E(X^2)$ . Then,  $\forall \varepsilon > 0$ ,

$$P(X \leq -\varepsilon) \leq \frac{E(X^2)}{E(X^2) + \varepsilon^2} \tag{37}$$

$$P(X \geq \varepsilon) \leq \frac{E(X^2)}{E(X^2) + \varepsilon^2}. \tag{38}$$

Inequality (38) is similar in its structure to Markov's. While requiring knowledge of a second moment, it is not restricted to non-negative random numbers and is never vacuous, unlike Markov's inequality. Inequality (37) is a left-sided variant.

When  $E(X) \neq 0$ , both (37) and (38) obtain replacing  $X$  with  $X - E(X)$ :

$$P(X \leq E(X) - \varepsilon) \leq \frac{\sigma^2(X)}{\sigma^2(X) + \varepsilon^2}, P(X \geq E(X) + \varepsilon) \leq \frac{\sigma^2(X)}{\sigma^2(X) + \varepsilon^2}. \tag{39}$$

### 6.1. Cantelli's inequalities with dF-coherent previsions

Cantelli's inequalities (37), (38), (39) generalise to imprecise previsions. Before seeing this, we take an intermediate step and prove a version of (37), (38), where  $X$  is a gamble and  $E(X), E(X^2)$  are replaced by the dF-coherent previsions  $P(X), P(X^2)$ . A preliminary lemma is needed<sup>7</sup>:

**Lemma 6.1.** *Given a gamble  $X$ , event  $A = (X > 0)$  with indicator  $I_A$ , a dF-coherent prevision  $P$  on  $\{A, I_A X, I_A X^2\}$ , it holds that*

$$[P(I_A X)]^2 \leq P(A)P(I_A X^2). \tag{40}$$

<sup>6</sup> Also termed Chebyshev-Cantelli inequalities or one-sided Chebyshev inequalities, as their earliest clues appear in Chebyshev's work. Cantelli [3] investigated them in 1928 (possibly also in some now hardly available paper around 1910), Uspensky [20] in 1937.

<sup>7</sup> Our proof applies Jensen's inequality (19). Alternatively, the Cauchy-Schwarz inequality may be employed.

**Proof.** Suppose first  $P(A) = 0$ . From  $0 \leq I_A X \leq I_A \sup X$ , monotonicity and homogeneity of  $P$  imply that  $0 \leq P(I_A X) \leq P(A) \sup X = 0$ . Therefore,  $P(I_A X) = P(A) = 0$  and (40) holds.

Let now  $P(A) > 0$ . We have that

$$[P(I_A X)]^2 \leq P(A)P(I_A X^2) \text{ iff } \left[\frac{P(I_A X)}{P(A)}\right]^2 \leq \frac{P(I_A X^2)}{P(A)} \text{ iff } [P(X|A)]^2 \leq P(X^2|A),$$

which is a true inequality, as follows applying (19) with  $\phi(x) = x^2$ .  $\square$

**Proposition 6.1.** *Given a gamble  $X$ ,  $\varepsilon > 0$ , and a dF-coherent prevision  $P$  on  $\{X, X^2, (X \leq -\varepsilon)\}$  such that  $P(X) = 0$ ,*

$$P(X \leq -\varepsilon) \leq \frac{P(X^2)}{P(X^2) + \varepsilon^2}, \quad (41)$$

$$P(X \geq \varepsilon) \leq \frac{P(X^2)}{P(X^2) + \varepsilon^2}. \quad (42)$$

**Proof.** We prove first inequality (41). Define the event  $A = (X > -\varepsilon) = (X + \varepsilon > 0)$ . Since  $I_A(X + \varepsilon) \geq X + \varepsilon$ , applying monotonicity and translation invariance to (any dF-coherent extension to  $I_A(X + \varepsilon)$ ,  $X + \varepsilon$  of)  $P$  we obtain

$$P(I_A(X + \varepsilon)) \geq P(X + \varepsilon) = \varepsilon > 0. \quad (43)$$

By (43) at the first inequality, (40) at the second, and noting that  $I_A(X + \varepsilon)^2 \leq (X + \varepsilon)^2$  at the third, it turns out that

$$\begin{aligned} \varepsilon^2 &\leq [P(I_A(X + \varepsilon))]^2 \leq P(A)P(I_A(X + \varepsilon)^2) \leq P(A)P((X + \varepsilon)^2) \\ &= P(A)(P(X^2) + \varepsilon^2) = (1 - P(X \leq -\varepsilon))(P(X^2) + \varepsilon^2), \end{aligned}$$

using linearity and homogeneity of  $P$  at the equalities. Inequality (41) follows straightforwardly.

The proof of inequality (42) ensues from (41) applied to  $-X$  (noting that  $P(-X) = -P(X) = 0$  and  $(X \geq \varepsilon) = (-X \leq -\varepsilon)$ ).  $\square$

We point out that Cantelli's inequalities in Proposition 6.1 are more general than their classical counterparts (37) and (38). In fact, they require no probability assessment on (subsets of)  $\mathcal{A}(P)$ . Further, if such an assessment is available, it may be a dF-coherent probability, not necessarily countably additive.

## 6.2. Cantelli's inequalities with imprecise previsions

We focus at first on establishing an imprecise version of Proposition 6.1 while assuming very loose consistency requirements for the relevant imprecise previsions.

**Proposition 6.2** (Cantelli's inequalities). *Let  $X$  be a gamble, and take  $\varepsilon > 0$ .*

(a) *Let  $\underline{P}$  be defined on  $\mathcal{D}_0^{\leq}(X) = \{(X \leq -\varepsilon), -X^2\}$  and define  $\mathcal{M}_0^{\leq}(X) = \{P : P \text{ dF-coherent, } P \geq \underline{P} \text{ on } \mathcal{D}_0^{\leq}(X) \text{ and } P(X) = 0\}$ . If  $\mathcal{M}_0^{\leq}(X) \neq \emptyset$ , then*

$$\underline{P}(X \leq -\varepsilon) \leq \frac{\overline{P}(X^2)}{\overline{P}(X^2) + \varepsilon^2}.$$

(b) *Let  $\underline{P}$  be defined on  $\mathcal{D}_0^{\geq}(X) = \{(X \geq \varepsilon), -X^2\}$  and define  $\mathcal{M}_0^{\geq}(X) = \{P : P \text{ dF-coherent, } P \geq \underline{P} \text{ on } \mathcal{D}_0^{\geq}(X) \text{ and } P(X) = 0\}$ . If  $\mathcal{M}_0^{\geq}(X) \neq \emptyset$ , then*

$$\underline{P}(X \geq \varepsilon) \leq \frac{\overline{P}(X^2)}{\overline{P}(X^2) + \varepsilon^2}.$$

**Proof.** (a) Take  $P \in \mathcal{M}_0^{\leq}(X)$ . The proof follows from the derivation

$$\underline{P}(X \leq -\varepsilon) \leq P(X \leq -\varepsilon) \leq \frac{P(X^2)}{P(X^2) + \varepsilon^2} \leq \frac{-\underline{P}(-X^2)}{-\underline{P}(-X^2) + \varepsilon^2} = \frac{\overline{P}(X^2)}{\overline{P}(X^2) + \varepsilon^2},$$

using (41) at the second inequality. For the last inequality, observe first that  $P(X^2) \leq -\underline{P}(-X^2)$  iff  $-P(X^2) = P(-X^2) \geq \underline{P}(-X^2)$ , which obtains because  $P \in \mathcal{M}_0^{\leq}(X)$ . Then, apply the arithmetic inequality  $\frac{a}{a+b} \leq \frac{c}{c+b}$  iff  $a \leq c$ , when  $b(a+b)(c+b) > 0$  (with  $a = P(X^2)$ ,  $c = -\underline{P}(-X^2)$ ,  $b = \varepsilon^2$ ).

(b) The derivation is similar to (a), taking  $P \in \mathcal{M}_0^{\geq}(X)$  and using (42).  $\square$

We point out that  $\underline{P}$  is asked, in Proposition 6.2 (a), to *avoid sure loss* (ASL) on  $\mathcal{D}_0^{\leq}(X)$ , since  $\mathcal{M}_0^{\leq}(X) \neq \emptyset$  implies that the credal set of  $\underline{P}$  is non-empty (cf. Section 2.2). If  $\underline{P}$  is 2-coherent on  $\mathcal{D}_0^{\leq}(X)$ , it is also coherent and ASL on  $\mathcal{D}_0^{\leq}(X)$ , but 2-coherence is not really required in (a). Instead, there is the additional condition  $P(X) = 0$  for  $P \in \mathcal{M}_0^{\leq}(X)$ . The assumptions of Proposition 6.2 (b) are analogous.

We can further generalise Cantelli's inequalities in the following way, derived straightforwardly from Proposition 6.2.

**Corollary 6.1.** *Given a gamble  $X$  and  $c \in \mathbb{R}$ ,  $\varepsilon > 0$ ,*

(a) *Let  $\underline{P}$  be defined on  $\mathcal{D}_{\leq}(X) = \{(X \leq c - \varepsilon), -(X - c)^2\}$ , and let  $\mathcal{M}_{\leq}(X) = \{P : P \text{ dF-coherent}, P \geq \underline{P} \text{ on } \mathcal{D}_{\leq}(X) \text{ and } P(X) = c\}$ . If  $\mathcal{M}_{\leq}(X) \neq \emptyset$ , then*

$$\underline{P}(X \leq c - \varepsilon) \leq \frac{\overline{P}((X - c)^2)}{\overline{P}((X - c)^2) + \varepsilon^2}. \quad (44)$$

(b) *Let  $\underline{P}$  be defined on  $\mathcal{D}_{\geq}(X) = \{(X \geq c + \varepsilon), -(X - c)^2\}$ , and let  $\mathcal{M}_{\geq}(X) = \{P : P \text{ dF-coherent}, P \geq \underline{P} \text{ on } \mathcal{D}_{\geq}(X) \text{ and } P(X) = c\}$ . If  $\mathcal{M}_{\geq}(X) \neq \emptyset$ , then*

$$\underline{P}(X \geq c + \varepsilon) \leq \frac{\overline{P}((X - c)^2)}{\overline{P}((X - c)^2) + \varepsilon^2}. \quad (45)$$

**Proof.** Apply Proposition 6.2 to  $Y = X - c$  (noting, for (a), that  $\mathcal{D}_{\leq}(X) = \mathcal{D}_0^{\leq}(Y) = \{(Y \leq -\varepsilon), -Y^2\}$  and  $\mathcal{M}_{\leq}(X) = \mathcal{M}_0^{\leq}(Y)$ ; similarly for (b)).  $\square$

A natural choice for  $c$  is  $c = \underline{P}(X)$ . With this, (44) reads

$$\underline{P}(X \leq \underline{P}(X) - \varepsilon) \leq \frac{\overline{P}((X - \underline{P}(X))^2)}{\overline{P}((X - \underline{P}(X))^2) + \varepsilon^2}, \quad (46)$$

giving a non-trivial upper bound to the lower probability that  $X$  is 'too far away' (from below) from  $\underline{P}(X)$ . For instance, with  $\varepsilon = 3\sqrt{\overline{P}((X - \underline{P}(X))^2)}$  we obtain the bound

$$\underline{P}\left(X \leq \underline{P}(X) - 3\sqrt{\overline{P}((X - \underline{P}(X))^2)}\right) \leq \frac{1}{10}.$$

Appealing alternative choices for  $c$  are  $c = \overline{P}(X)$ ,  $c = \frac{\underline{P}(X) + \overline{P}(X)}{2}$ . Note that (a) in Corollary 6.1 does not apply when  $c \notin [\inf X, \sup X]$ , since then  $\mathcal{M}_{\leq}(X) = \emptyset$  ( $P(X) = c$  being not dF-coherent).

Analogous considerations apply to inequality (45). Thus, Cantelli's inequalities may be viewed as concentration inequalities, holding for lower provisions not necessarily coherent.

### 6.3. Cantelli's inequalities and coherence

The results of the previous section can be strengthened assuming coherence.

As a first step, take  $c = \underline{P}(X)$  and suppose that  $\underline{P}$  is coherent on  $\{(X \leq \underline{P}(X) - \varepsilon), -(X - \underline{P}(X))^2, X\}$ . Then,  $\mathcal{M}_{\leq}(X)$  in Corollary 6.1 is non-empty, by the Lower Envelope Theorem (Theorem 2.1), and inequality (46) obtains. Quite analogously, coherence of  $\underline{P}$  on  $\{(X \geq \underline{P}(X) + \varepsilon), -(X - \underline{P}(X))^2, X\}$  is sufficient to establish the corresponding inequality

$$\underline{P}(X \geq \underline{P}(X) + \varepsilon) \leq \frac{P((X - \underline{P}(X))^2)}{\overline{P}((X - \underline{P}(X))^2) + \varepsilon^2}. \quad (47)$$

If in particular  $\underline{P} = \overline{P} = P$ , with  $P$  dF-coherent prevision, (46) and (47) specialise into

$$P(X \leq P(X) - \varepsilon) \leq \frac{V_P(X)}{V_P(X) + \varepsilon^2} \quad (48)$$

$$P(X \geq P(X) + \varepsilon) \leq \frac{V_P(X)}{V_P(X) + \varepsilon^2} \quad (49)$$

where

$$V_P(X) = P((X - P(X))^2)$$

is the variance of  $X$  whenever  $P$  is an expectation, in which case (48), (49) boil down to (39).

However, stricter Cantelli's bounds than (46) and (47) apply with coherent lower previsions. Prior to seeing this, let us recall from [21, Appendix G] the definitions of lower and upper variance.

**Definition 6.1.** For a given gamble  $X$  and a coherent lower prevision  $\underline{P}$ , let

$$\underline{V}(X) = \min_{c \in \mathbb{R}} \{ \underline{P}((X - c)^2) \} \text{ (lower variance)} \quad (50)$$

$$\overline{V}(X) = \min_{c \in \mathbb{R}} \{ \overline{P}((X - c)^2) \} \text{ (upper variance)} \quad (51)$$

It is also proven in [21, Appendix G] that the minima in (50), (51) are attained by the previsions in the credal set  $\mathcal{M}$  of  $\underline{P}$ , termed  $P_1$  and  $P_2$  respectively, such that<sup>8</sup>

$$\underline{V}(X) = V_{P_1}(X) = \min_{P \in \mathcal{M}} \{ V_P \} \quad (52)$$

$$\overline{V}(X) = V_{P_2}(X) = \max_{P \in \mathcal{M}} \{ V_P \}. \quad (53)$$

Among the properties of  $\underline{V}(X)$ ,  $\overline{V}(X)$ , we recall that [21, Appendix G]

$$0 \leq \underline{V}(X) \leq \overline{V}(X), \quad \underline{V}(X) = \overline{V}(X) \Rightarrow \underline{P}(X) = \overline{P}(X). \quad (54)$$

Given this, we have that

**Proposition 6.3.** Let  $X$  be a gamble and  $\varepsilon > 0$ .

(a) Let  $\underline{P}$  be coherent on  $\{X, -X, (X \leq \underline{P}(X) - \varepsilon), (X \leq \overline{P}(X) - \varepsilon)\}$ . Then,

$$\underline{P}(X \leq \overline{P}(X) - \varepsilon) \leq \frac{\overline{V}(X)}{\overline{V}(X) + \varepsilon^2} \quad (55)$$

$$\underline{P}(X \leq \underline{P}(X) - \varepsilon) \leq \frac{\underline{V}(X)}{\underline{V}(X) + \varepsilon^2}. \quad (56)$$

(b) Let  $\underline{P}$  be coherent on  $\{X, -X, (X \geq \underline{P}(X) + \varepsilon), (X \geq \overline{P}(X) + \varepsilon)\}$ . Then,

$$\underline{P}(X \geq \underline{P}(X) + \varepsilon) \leq \frac{\overline{V}(X)}{\overline{V}(X) + \varepsilon^2} \quad (57)$$

$$\underline{P}(X \geq \overline{P}(X) + \varepsilon) \leq \frac{\underline{V}(X)}{\underline{V}(X) + \varepsilon^2}. \quad (58)$$

<sup>8</sup> In this result,  $\underline{P}$  is defined on the linear space  $\mathcal{L}$  of all gambles. In the sequel we shall operate with restrictions of  $\underline{P}$  onto the subset(s) of  $\mathcal{L}$  of interest. To see that these two approaches are equivalent, recall from Section 2.2 that given a coherent  $\underline{P}$  on any such subset  $\mathcal{D}$ ,  $\underline{P}$  always has a coherent extension on any superset of  $\mathcal{D}$ , in particular on  $\mathcal{L}$ .



**Proof.** We prove only part (a), the proof of (b) being analogous. We prove first (55). Because  $\underline{P}$  is coherent, there is  $P^* \in \mathcal{M}$  such that  $P^* \geq \underline{P}$  and  $P^*(-X) = \underline{P}(-X) = -\overline{P}(X)$ , implying  $P^*(X) = \overline{P}(X)$ . Hence,

$$\underline{P}(X \leq \overline{P}(X) - \varepsilon) \leq P^*(X \leq P^*(X) - \varepsilon) \leq \frac{V_{P^*}(X)}{V_{P^*}(X) + \varepsilon^2} \leq \frac{\overline{V}(X)}{\overline{V}(X) + \varepsilon^2},$$

where the second inequality is due to (48), the third to (53).

To prove (56), let  $P_1$  be the prevision in  $\mathcal{M}$  satisfying (52). Then we have, by (48) at the second inequality and by the properties of  $P_1$  elsewhere

$$\underline{P}(X \leq \underline{P}(X) - \varepsilon) \leq P_1(X \leq P_1(X) - \varepsilon) \leq \frac{V_{P_1}(X)}{V_{P_1}(X) + \varepsilon^2} = \frac{\underline{V}(X)}{\underline{V}(X) + \varepsilon^2}. \quad \square$$

Let us discuss the role of the new Cantelli's inequalities in Proposition 6.3. Considering first (55) and (56), both give a bound to the lower probability  $\underline{P}(X \leq c)$ . To apply (55), it is necessary that  $c < \overline{P}(X)$ . On the other hand, if  $c < \underline{P}(X)$ , inequality (56) obtains too and is *definitely preferable*. In fact, recalling (54), we have that

$$\frac{\underline{V}(X)}{\underline{V}(X) + \varepsilon^2} \leq \frac{\overline{V}(X)}{\overline{V}(X) + \varepsilon^2}, \quad (59)$$

so that the upper bound (55) is looser than (56).

Inequality (56) is also weakly stricter than inequality (46): this is easily seen from (59) and (51). However, (56) is *typically stricter* than (46), meaning that  $\underline{P}(X) = \overline{P}(X)$ , i.e.  $X$  is given a precise evaluation, when both inequalities return the same bound. In fact, in this case  $\underline{V}(X) = \overline{P}((X - \underline{P}(X))^2)$ , implying  $\underline{V}(X) = \overline{V}(X)$  (recalling that  $\underline{V}(X) \leq \overline{V}(X) \leq \overline{P}((X - \underline{P}(X))^2)$  by (51) and (54)). Therefore,  $\underline{P}(X) = \overline{P}(X)$  by (54). Thus, (56) is the strictest left-sided Cantelli's inequality we derived. In a comparison with inequality (46), it has to be said that (56) requires the additional task of determining  $\underline{V}(X)$ . Ways to compute  $\underline{V}(X)$  and  $\overline{V}(X)$  are discussed in [21, Appendix G]; it is reported there that the computation of  $\underline{V}(X)$  is simpler than that of  $\overline{V}(X)$ .

Turning to inequality (58), it is the strictest Cantelli's bound to  $\underline{P}(X \geq c)$ . It works when  $c > \overline{P}(X)$  and is preferable to the looser inequalities (57) and (47). On the other hand, (57) and (47) are applicable when  $c \in ]\underline{P}(X), \overline{P}(X)[$  and an advantage of (47) is that it does not require knowing  $\underline{V}(X)$ , nor  $\overline{V}(X)$  (and not even coherence).

Cantelli's inequalities assume an only limited knowledge of  $X$ . As such, they cannot be expected to produce very strict bounds, in general. To get some insight on this, a partial comparison between Cantelli's and Markov's inequalities is possible. Precisely, let us compare the Lower Markov's inequality (35) with Cantelli's inequality (58). Both inequalities apply if  $\underline{P}$  is coherent,  $X \geq 0$ ,  $\varepsilon > 0$  and  $a = \overline{P}(X) + \varepsilon$ , so that Markov's inequality is

$$\underline{P}(X \geq \overline{P}(X) + \varepsilon) \leq \frac{\underline{P}(X)}{\overline{P}(X) + \varepsilon}. \quad (60)$$

By (58),  $\underline{P}(X \geq \overline{P}(X) + \varepsilon)$  is majorised by  $\frac{\underline{V}(X)}{\underline{V}(X) + \varepsilon^2}$ , thus Cantelli's bound is better than Markov's iff

$$\frac{\underline{V}(X)}{\underline{V}(X) + \varepsilon^2} \leq \frac{\underline{P}(X)}{\overline{P}(X) + \varepsilon}. \quad (61)$$

Discarding the extreme case  $\underline{P}(X) = 0$  (where Markov's inequality (60) cannot be improved, implying  $\underline{P}(X \geq \overline{P}(X) + \varepsilon) = 0$ ), condition (61) is equivalent to

$$\underline{P}(X)\varepsilon^2 - \underline{V}(X)\varepsilon + \underline{V}(X)[\underline{P}(X) - \overline{P}(X)] \geq 0,$$

a second order inequality with

$$\Delta = \underline{V}(X)[\underline{V}(X) + 4\underline{P}(X)(\overline{P}(X) - \underline{P}(X))] \geq [\underline{V}(X)]^2 \geq 0.$$

It ensues that

$$\varepsilon_1 = \frac{\underline{V}(X) - \sqrt{\Delta}}{2\underline{P}(X)} \leq 0, \text{ while}$$

$$\varepsilon_2 = \frac{\underline{V}(X) + \sqrt{\Delta}}{2\underline{P}(X)} \geq \frac{\underline{V}(X)}{\underline{P}(X)} \geq 0.$$

Since  $\varepsilon > 0$ , we may conclude that Cantelli's inequality is stricter than Markov's iff  $\varepsilon > \varepsilon_2$ .

Note also that a simple sufficient condition for Markov's inequality to be preferable is that  $\varepsilon \underline{P}(X) < \underline{V}(X)$ .

We only mention that Markov's inequality (35), taking  $a = \underline{P}(X) + \varepsilon$ , can be confronted also with inequality (47). The procedure is similar, and again the result depends on the value of  $\varepsilon$ .

Finally, we point out that further Cantelli-like inequalities are easily derived from the previous ones via conjugacy. To exemplify, the bound in (56) applies by monotonicity also to  $\underline{P}(X < \underline{P}(X) - \varepsilon)$ , which implies that

$$\overline{P}(X \geq \underline{P}(X) - \varepsilon) = 1 - \underline{P}(X < \underline{P}(X) - \varepsilon) \geq 1 - \frac{\underline{V}(X)}{\underline{V}(X) + \varepsilon^2} = \frac{\varepsilon^2}{\underline{V}(X) + \varepsilon^2}.$$

## 7. Conclusions

In this paper we explored how some basic probability inequalities are modified when our uncertainty evaluations are imprecise. It turns out that the inequalities have a certain *robustness* with respect to the quality of our information. While in classical probability theory this means that a very limited (if any) distributional knowledge of a gamble  $X$  is required to apply the inequalities, here in addition we may have different degrees of consistency for our evaluations. As we have seen, coherence is anyway preferable. It may ensure inferences in a larger number of situations (Section 4.2, Example 5.1) or better bounds (Section 6.3). It remains to be explored whether some of these bounds can be further improved, in case referring to specific coherent models. It is on the other hand remarkable that versions of the investigated inequalities may be employed with much weaker consistency conditions, such as 2-coherence.

We expect that similar conclusions may be drawn with generalisations of other inequalities not explored here. This seems to be a promising area for future work. A question one may run into quite soon is how should concentration (or dispersion, equivalently) of lower or upper previsions be measured. While being an issue to be deepened, we guess that there might be no unique answer. Take for instance Cantelli's bounds. The variance  $\sigma^2(X)$  appears in the probability bounds (39). It is replaced by Walley's lower variance  $\underline{V}(X)$  in the bound (56) (or by  $\overline{V}(X)$  in (55)), but this requires coherence. Otherwise, the role of  $\sigma^2(X)$  is taken by  $\overline{P}((X - \underline{P}(X))^2)$  in inequality (46). Note also that Chebyshev-like inequalities are easily obtained from Markov's inequalities (35), (36). These involve different variance-like quantities. For instance, we have that, given  $b > 0$ ,

$$\overline{P}(|X - \underline{P}(X)| \geq b) \leq \frac{\overline{P}((X - \underline{P}(X))^2)}{b^2}, \text{ but also}$$

$$\overline{P}(|X - \overline{P}(X)| \geq b) \leq \frac{\overline{P}((X - \overline{P}(X))^2)}{b^2}.$$

In general,  $\underline{V}(X)$ ,  $\overline{V}(X)$  are coherence-dependent by construction, since they refer to the credal set  $\mathcal{M}$  of  $\underline{P}$ . By the Lower Envelope Theorem (Theorem 2.1), the connection between a coherent  $\underline{P}$  and its credal set is strict, but becomes looser with previsions that avoid sure loss. Would  $\underline{V}(X)$  still be a reliable variance measure in such a case? Further, with 2-coherence  $\mathcal{M}$  may even be empty. Thus, what precisely extends variance with imprecise judgements is another question waiting for more exhaustive answers.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] H. Agahi, Choquet functional and its applications in information theory, *Inf. Sci.* 433–434 (2018) 306–312, <https://doi.org/10.1016/j.ins.2017.12.035>.
- [2] D. Bethmann, An improvement to Jensen's inequality and its application to mating market clearing when paternity is uncertain, *Math. Soc. Sci.* 91 (2018) 71–74, <https://doi.org/10.1016/j.mathsocsci.2017.08.004>.
- [3] F.P. Cantelli, Sui confini della probabilità, in: *Atti del Congresso Internazionale dei Matematici*, vol. 6, Bologna, Italy, Zanichelli, 1929, pp. 47–59.

- [4] C. Corsato, R. Pelessoni, P. Vicig, Nearly-linear uncertainty measures, *Int. J. Approx. Reason.* 114 (2019) 1–28, <https://doi.org/10.1016/j.ijar.2019.08.001>.
- [5] F.G. Cozman, Concentration inequalities and laws of large numbers under epistemic and regular irrelevance, *Int. J. Approx. Reason.* 51 (9) (2010) 1069–1084, <https://doi.org/10.1016/j.ijar.2010.08.009>.
- [6] G. De Cooman, E. Miranda, Weak and strong laws of large numbers for coherent lower previsions, *J. Stat. Plan. Inference* 138 (8) (2008) 2409–2432, <https://doi.org/10.1016/j.jspi.2007.10.020>.
- [7] B. de Finetti, *Theory of Probability: A Critical Introductory Treatment*, John Wiley, New York, 1970.
- [8] B.K. Ghosh, Probability inequalities related to Markov’s theorem, *Am. Stat.* 56 (3) (Aug. 2002) 186–190, <https://doi.org/10.1198/000313002119>.
- [9] B. Giroto, S. Holzer, Chebyshev and Jensen inequalities for Choquet integral, *Math. Pannon.* 23 (2) (2012) 267–275.
- [10] R. Mesiar, J. Li, E. Pap, The Choquet integral as Lebesgue integral and related inequalities, *Kybernetika* 46 (6) (2010) 1098–1107.
- [11] I. Montes, E. Miranda, S. Destercke, Unifying neighbourhood and distortion models: part I — new results on old models, *Int. J. Gen. Syst.* 49 (6) (2020) 602–635.
- [12] I. Montes, E. Miranda, S. Destercke, Unifying neighbourhood and distortion models: part II — new models and synthesis, *Int. J. Gen. Syst.* 49 (6) (2020) 636–674.
- [13] C.P. Niculescu, L.E. Persson, *Convex Functions and Their Applications*, Springer Berlin Heidelberg, New York, NY, 2018.
- [14] R. Pelessoni, P. Vicig, Imprecise previsions for risk measurement, *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 11 (04) (2003) 393–412, <https://doi.org/10.1142/S0218488503002156>.
- [15] R. Pelessoni, P. Vicig, 2-coherent and 2-convex conditional lower previsions, *Int. J. Approx. Reason.* 77 (2016) 66–86, <https://doi.org/10.1016/j.ijar.2016.06.003>.
- [16] R. Pelessoni, P. Vicig, Weakly consistent extensions of lower previsions, *Fuzzy Sets Syst.* 328 (2017) 83–106, <https://doi.org/10.1016/j.fss.2017.07.021>.
- [17] H. Román-Flores, A. Flores-Franulič, Y. Chalco-Cano, A Jensen type inequality for fuzzy integrals, *Inf. Sci.* 177 (15) (2007) 3192–3201, <https://doi.org/10.1016/j.ins.2007.02.006>.
- [18] A.N. Shiryaev, *Probability*, vol. 95, Springer, New York, NY, 1996.
- [19] M.C.M. Troffaes, G. de Cooman, *Lower Previsions*, Wiley, Chichester, West Sussex, United Kingdom, 2014.
- [20] J.V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.
- [21] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London; New York, 1991.
- [22] R.-S. Wang, Some inequalities and convergence theorems for Choquet integrals, *J. Appl. Math. Comput.* 35 (1–2) (2011) 305–321, <https://doi.org/10.1007/s12190-009-0358-y>.
- [23] D. Zhang, R. Mesiar, E. Pap, Jensen’s inequality for Choquet integral revisited and a note on Jensen’s inequality for generalized Choquet integral, *Fuzzy Sets Syst.* 430 (2022) 79–87, <https://doi.org/10.1016/j.fss.2021.09.004>.