Probability Inequalities with Imprecise Previsions

Renato Pelessoni[0000−0002−7820−8173] and Paolo Vicig[0000−0001−8842−442X]

University of Trieste, Piazzale Europa 1 I-34127 Trieste, Italy
{renato.pelessoni,paolo.vicig}@deams.units.it

Abstract. We investigate how various well known probability inequalities extend to lower and upper previsions. Our focus is especially on Markov's, Bhatia-Davis, Jensen's and Cantelli's inequalities. In all such cases, imprecise versions of these inequalities are available even requiring the weak consistency notion of 2-coherence, which implies that they obtain for a large number of uncertainty models. However, stronger results may be achieved with coherent lower and upper previsions. In particular, it is possible to bound lower and upper variances. Various bounds for lower and upper covariances are also presented; while being generally not tight, they require very limited amounts of information to obtain.

Keywords: Markov's inequality · Bhatia-Davis inequality · Coherent and 2-coherent imprecise previsions · Lower and upper covariance.

1 Introduction

While a great number of inequalities are well known in classical probability theory, the issue of investigating how they are modified in imprecise probability theory received considerably less attention. Some results are available within studies on laws of large numbers [4, 5] or other specific problems [16].

In [11, 12] we started a systematic study in this direction of some known inequalities, which we report and integrate in the present paper. The considered inequalities regard evaluations of gambles, and provide therefore bounds on lower or upper previsions for these gambles or transformations of theirs. Typical distinguishing features of imprecise probability inequalities, as will constantly appear in the sequel, are:

(a) More imprecise probability inequalities may be counterparts to the same probability inequality.
(b) Several imprecise probability inequalities apply requiring only relatively weak degrees of consistency.

We believe that (b) is particularly interesting, as it lets us shed light on the truly minimal properties that underlie an uncertainty inequality. This fact cannot be adequately investigated as long as precise probabilities only are considered, unlike imprecise probabilities that offer more notions of consistency. Thus, we have inequalities for coherent lower/upper previsions, but often also versions with imprecise previsions that are 2-coherent, which is a much weaker requirement.

The paper is organised as follows: Section 2 contains basic information on the consistency notions for lower/upper previsions useful in the sequel. Section 3 presents Markov's inequalities, requiring 2-coherence. Section 4 discusses a number of extensions of Bhatia-Davis inequality, originally upper bounding the variance $\sigma_X^2 = E[(X - E(X))^2]$ of a gamble X with $(\sup X - E(X))(E(X) - \inf X)$. Bounds are given when the expectations in σ_X^2 are replaced with upper ($\overline{P}$) or lower ($\underline{P}$) 2-coherent previsions. Further results requiring that $\underline{P}$, $\overline{P}$ are coherent are: bounds on the upper and lower variance of X (defined in [18]), and a bivariate generalisation of Bhatia-Davis inequality. The latter bounds upper and lower covariances. Although in general all these bounds require a limited uncertainty knowledge, when even less information is available some weaker bounds obtain, intermediate between Bhatia-Davis and Popoviciu's inequalities. Section 5 presents basic results on imprecise versions of Jensen's and Cantelli's inequalities. Section 6 concludes the paper.

2 Preliminaries

We shall be concerned with bounds on *gambles*, i.e. on bounded random numbers. If X is a gamble, define $M_X = \sup X$, $m_X = \inf X$.

Lower ($\underline{P}$) and upper ($\overline{P}$) previsions are imprecise evaluations of gambles. In principle, $\underline{P}$ or $\overline{P}$ is a mapping from an *arbitrary* set $\mathcal{D}$ of gambles into $\mathbb{R}$. Thus, $\mathcal{D}$ need not be a structured set, such as a linear space.

More consistency notions are available for lower and upper previsions. They are variants of de Finetti's definition of coherence [7] for a (precise) prevision P, termed here dF-coherence:

Definition 1. *A mapping $P : \mathcal{D} \to \mathbb{R}$ 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* $\sup G \geq 0$.

In Definition 1, the gamble G has the behavioural interpretation of a gain from $n+1$ transactions $s_i(X_i - P(X_i))$ on X_i with price $P(X_i), i = 0, \dots, n$.

Whenever an expectation $E(X)$ is assessed, it is a dF-coherent prevision for X [18, Sections 3.2.1, 3.2.2]. The variance σ_X^2 of X can be defined in terms of previsions as $V_P(X) = P[(X - P(X))^2]$. In what follows, the symbols $E(X)$, σ_X^2 will only be employed when quoting results from classical probability theory.

Replacing P with $\underline{P}$ in Definition 1, we obtain the following consistency notions:

(a) the definition of *coherent lower prevision* [18] if we restrict the coefficients $s_1, \dots, s_n$ (but not s_0) to be non-negative;
(b) the definition of 2-*coherent lower prevision* [10] if we require further that $n \leq 1$, so that we have (at most) two coefficients in the expression of the gain, $s_0 \in \mathbb{R}$, $s_1 \geq 0$.

As customary, we shall assume in the sequel that $\underline{P}$ and $\overline{P}$ are *conjugate*, meaning that

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

It is also understood that $\underline{P}$, $\overline{P}$ are defined on the same domain $\mathcal{D}$. (This is equivalent to defining just one between $\underline{P}$, $\overline{P}$ on a domain $\mathcal{D}'$ with the property $X \in \mathcal{D}' \Rightarrow -X \in \mathcal{D}'$.)

A lower prevision $\underline{P}$ and its conjugate $\overline{P}$ are coherent on $\mathcal{D}$ iff $\underline{P}(X) = \min\{P(X) : P \in \mathcal{M}\}, \overline{P}(X) = \max\{P(X) : P \in \mathcal{M}\}$, $\forall X \in \mathcal{D}$, where $\mathcal{M} = \{P :$ P dF-coherent on $\mathcal{D}$, $\underline{P} \leq P \leq \overline{P}\}$ is the *credal set* of $\underline{P}$ and $\overline{P}$ [18].

A dF-coherent prevision P is both a lower and an upper coherent prevision, $P = \underline{P} = \overline{P}$. Lower/upper coherent previsions are clearly also 2-coherent. The properties of dF-coherent previsions are stronger than those of lower/upper coherent previsions, which are stronger than those of lower/upper 2-coherent previsions. For instance, a dF-coherent P is linear: $P(X+Y) = P(X) + P(Y)$, while only the property $\underline{P}(X+Y) \geq \underline{P}(X) + \underline{P}(Y)$ obtains in general if $\underline{P}$ is coherent, not even that if it is 2-coherent. 2-coherent previsions ensure however some minimal properties (cf. [10]): we recall, for $\mu \in \{\underline{P}, \overline{P}\}$:

- $\mu(X) \in [m_X, M_X]$ *(internality)*; implies $\mu(c) = c$, $\forall c \in \mathbb{R}$;
- $X \leq Y \Rightarrow \mu(X) \leq \mu(Y)$ *(monotonicity)*;
- $\mu(\lambda X) = \lambda\mu(X), \forall \lambda \geq 0$ *(positive homogeneity)*; applies with $\lambda \in \mathbb{R}$ for dF-coherent previsions;
- $\mu(X + c) = \mu(X) + c, \forall c \in \mathbb{R}$ *(translation invariance)*;
- $\underline{P}, \overline{P}$ conjugate implies $\underline{P}(X) \leq \overline{P}(X)$.

In the results in the paper, we shall point out which is the minimal set $\mathcal{D}$ on which P, or $\underline{P}$ and $\overline{P}$, have to be assessed. In their proofs, they may be extended to larger sets applying the properties listed above, thus guaranteeing 2-coherence or coherence of the extension. Recall also that there always exists a coherent (2-coherent) extension on any set of gambles $\mathcal{D}' \supset \mathcal{D}$ of a lower/upper prevision coherent (2-coherent) on $\mathcal{D}$ [10, 18].

When applied to (indicators of) events, all these kinds of previsions boil down to probabilities (precise or lower/upper), while the conjugacy relation becomes, for an event A, $\underline{P}(A) = 1 - \overline{P}(A^c)$.

While being a weak consistency requirement, 2-coherence includes a larger number of uncertainty models than coherence. For instance, if $\mathcal{D}$ is the powerset of a partition of the sure event Ω into atomic events, assigning conjugate $\underline{P}$, $\overline{P}$ 2-coherent on $\mathcal{D}$ corresponds to giving a couple of (normalised) capacities $(\underline{P}, \overline{P})$ with the additional condition $\underline{P}(A) \leq \overline{P}(A) \, \forall A \in \mathcal{D}$. 2-coherent previsions on linear spaces are instead a prominent case of niveloids [6]. Hence, inequalities that apply to 2-coherent previsions are valid for a great number of models.

3 Markov's Inequalities

To start our investigation of what corresponds to standard probability inequalities in an imprecise setting, in this section we consider the well known Markov's inequality $P(X \geq a) \leq \frac{E(X)}{a}, \forall a > 0$, with X non-negative. While requiring simple computations, the problem of generalising Markov's inequality to imprecise previsions already displays the core features of this kind of analysis, pointed out in the Introduction. In fact, *two* Markov's inequalities have been proven in [11, Proposition 5.1], requiring *only* 2-coherence of $\underline{P}$, $\overline{P}$:

Proposition 1 (Markov's inequalities). *Let X be a non-negative gamble.*

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

$$\underline{P}(X \geq a) \leq \frac{\underline{P}(X)}{a}, \forall a > 0 \tag{1}$$

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

$$\overline{P}(X \geq a) \leq \frac{\overline{P}(X)}{a}, \forall a > 0 \tag{2}$$

Next we introduce further Markov-like inequalities, still with the weak consistency requirement of 2-coherence.

Proposition 2 (Reverse Markov Inequalities). *Let $\underline{P}$, $\overline{P}$ be conjugate and 2-coherent on $\mathcal{D} \supseteq \{X, (X \leq a)\}$, $X \geq 0$. For any $0 < a < M_X$, it holds that*

$$1 - \frac{\overline{P}(X)}{a} \leq \underline{P}(X \leq a) \leq \frac{M_X - \overline{P}(X)}{M_X - a}, \tag{3}$$

$$1 - \frac{\underline{P}(X)}{a} \leq \overline{P}(X \leq a) \leq \frac{M_X - \underline{P}(X)}{M_X - a}. \tag{4}$$

Proof. For a unified proof of (3), (4), let $\mu \in \{\underline{P}, \overline{P}\}$ and μ^{c} be its conjugate. Using either (1) (when $\mu^{\mathsf{c}} = \underline{P}$) or (2) (when $\mu^{\mathsf{c}} = \overline{P}$) at the last inequality of the next derivation, we obtain the left-hand inequalities in (3), (4):

$$\mu(X \leq a) = 1 - \mu^{\mathsf{c}}(X > a) \geq 1 - \mu^{\mathsf{c}}(X \geq a) \geq 1 - \frac{\mu^{\mathsf{c}}(X)}{a}.$$

As for the right-hand inequalities, observing that $(X \leq a) = (M_X - X \geq M_X - a)$ and applying either (1) or (2) at the inequality, translation invariance and conjugacy at the second equality, we have that

$$\mu(X \leq a) = \mu(M_X - X \geq M_X - a) \leq \frac{\mu(M_X - X)}{M_X - a} = \frac{M_X - \mu^{\mathsf{c}}(X)}{M_X - a}.$$

□

When $\underline{P} = \overline{P} = P$, the right-hand inequalities (3), (4) boil down to $P(X \leq a) \leq \frac{M_X - P(X)}{M_X - a}$, known (for expectations) as Reverse Markov Inequality.

The next lemma is an easy follow-up of Proposition 2:

Lemma 1. *In the assumptions of Proposition 2, it holds that*

$$\underline{P}(X \geq a) \geq \frac{\underline{P}(X) - a}{M_X - a}, \overline{P}(X \geq a) \geq \frac{\overline{P}(X) - a}{M_X - a}.$$

Proof. Using again $\mu \in \{\underline{P}, \overline{P}\}$ and μ^c, and by either (3) or (4) at the second inequality,

$$\mu(X \geq a) \geq \mu(X > a) = 1 - \mu^c(X \leq a) \geq 1 - \frac{M_X - \mu(X)}{M_X - a} = \frac{\mu(X) - a}{M_X - a}.$$

□

By Lemma 1, a bilateral version of inequalities (1), (2) obtains:

$$\frac{\underline{P}(X) - a}{M_X - a} \leq \underline{P}(X \geq a) \leq \frac{\underline{P}(X)}{a}, \ \frac{\overline{P}(X) - a}{M_X - a} \leq \overline{P}(X \geq a) \leq \frac{\overline{P}(X)}{a}. \quad (5)$$

Note that only one inequality in (3) is non-trivial when $a \neq \overline{P}(X)$, and no one for $a = \overline{P}(X)$. For instance, if $a > \overline{P}(X)$ then $\frac{M_X - \overline{P}(X)}{M_X - a} > 1$, which makes the right-hand bound in (3) useless: by internality, $\underline{P}(X \leq a) \leq 1$. A similar reasoning applies to the left-hand bound in (3) and to the double inequalities in (4) and in (5).

Markov's inequalities (1), (2) originate Chebyshev-like inequalities. For example, for given $b > 0$, we have that $\underline{P}(|X - \underline{P}(X)| \geq b) \leq b^{-2}\underline{P}[(X - \underline{P}(X))^2]$. The right-hand term of this inequality is upper bounded by one of the Bhatia-Davis inequalities studied in the next section, see the following Proposition 3 (a).

4 Bhatia-Davis Inequalities

The classical Bhatia-Davis inequality [2] gives an upper bound to the variance σ_X^2 of a gamble X:

$$\sigma_X^2 = E[(X - E(X))^2] \leq (M_X - E(X))(E(X) - m_X). \quad (6)$$

When extending it to lower and upper previsions, we have more options to replace the inner and outer expectations in the left-hand term of (6) with either $\underline{P}$ or $\overline{P}$. Intuitively, it should be simpler to bound $\underline{P}[(X - \mu(X))^2]$ rather than $\overline{P}[(X - \mu(X))^2]$, given that, assuming 2-coherence, the former is less than or equal to the latter, for $\mu(X) = \underline{P}(X)$ or $\mu(X) = \overline{P}(X)$. In fact, the following proposition [12] obtains for $\underline{P}[(X - \mu(X))^2]$:

Proposition 3. *(a) Given $\underline{P}$ 2-coherent on $\mathcal{D} \supseteq \{X, (X - \underline{P}(X))^2\}$, it holds that*

$$\underline{P}[(X - \underline{P}(X))^2] \leq (M_X - \underline{P}(X))(\underline{P}(X) - m_X)$$

.

(b) Given $\underline{P}$, $\overline{P}$ 2-coherent on $\mathcal{D} \supseteq \{X, (X - \overline{P}(X))^2\}$, it holds that

$$\underline{P}[(X - \overline{P}(X))^2] \leq (M_X - \overline{P}(X))(\overline{P}(X) - m_X).$$

The next proposition establishes instead two more structured bounds for $\overline{P}[(X - \mu(X))^2]$, when $\mu(X) = \underline{P}(X)$ or $\mu(X) = \overline{P}(X)$, respectively [12].

Proposition 4. *Let $\underline{P}$, $\overline{P}$ be 2-coherent on $\mathcal{D} \supseteq \{X, (X-\overline{P}(X))^2, (X-\underline{P}(X))^2\}$. Then,*

$$\overline{P}[(X - \overline{P}(X))^2] \leq \max\{(M_X - \overline{P}(X))(\overline{P}(X) - m_X), (\overline{P}(X) - m_X)^2\}, \quad (7)$$

$$\overline{P}[(X - \underline{P}(X))^2] \leq \max\{(M_X - \underline{P}(X))(\underline{P}(X) - m_X), (M_X - \underline{P}(X))^2\}. \quad (8)$$

The maximum in (7) *(in* (8)*) is equal to* $(M_X - \overline{P}(X))(\overline{P}(X) - m_X)$ *(to* $(M_X - \underline{P}(X))(\underline{P}(X) - m_X)$*) iff* $\overline{P}(X) \leq \frac{M_X+m_X}{2}$ *(iff* $\underline{P}(X) \geq \frac{M_X+m_X}{2}$*).*

Thus, the upper bound $(M_X - \overline{P}(X))(\overline{P}(X) - m_X)$ for $\underline{P}[(X - \overline{P}(X))^2]$ from Proposition 3 (b) bounds also $\overline{P}[(X - \overline{P}(X))^2]$, but only if $\overline{P}(X)$ is 'sufficiently low', and precisely smaller or equal to $\frac{M_X+m_X}{2}$; similarly for the bound $(M_X - \underline{P}(X))(\underline{P}(X) - m_X)$ in Proposition 3 (a). It can be shown that all these bounds are sharp, in the sense that they may obtain with equality for non-trivial gambles. (Examples are given in [12].)

A natural application of $\underline{P}[(X - \mu(X))^2]$, $\overline{P}[(X - \mu(X))^2]$ ($\mu \in \{\underline{P}, \overline{P}\}$) is to measure the dispersion of X. Regarding this, a distinction arises between 2-coherence and coherence. In fact, *when $\underline{P}$ and $\overline{P}$ are coherent* we may also resort to the lower and upper variance of X,

$$\underline{V}(X) = \min_{P \in \mathcal{M}}\{V_P(X)\}, \ \overline{V}(X) = \max_{P \in \mathcal{M}}\{V_P(X)\},$$

with $V_P(X) = P[(X - P(X))^2], \forall P \in \mathcal{M}$ [18, Appendix G, Section G2].

If we think that there is a 'true' but unknown prevision $P(X)$ in the credal set $\mathcal{M}$, preferring $\underline{V}(X)$ and $\overline{V}(X)$ as measures of dispersion is quite natural. However, we must consider that *(a)* the computation of $\underline{V}(X)$ and especially $\overline{V}(X)$ is in general not simple (special cases are tackled already in [18, Appendix G], see also [15] for related work), and *(b)* with 2-coherence only, $\mathcal{M}$ may be empty, $\underline{V}(X)$, $\overline{V}(X)$ being then undefined. In case *(b)*, we are necessarily left with the previsions $\underline{P}[X - \mu(X)]^2$, $\overline{P}[X - \mu(X)]^2$, $\mu(X) \in \{\underline{P}(X), \overline{P}(X)\}$. In case *(a)*, the same previsions (and consequently their Bhatia-Davis bounds) majorise $\underline{V}(X)$ and $\overline{V}(X)$, respectively, because of the more general inequalities $\underline{V}(X) \leq \underline{P}[(X - c)^2]$, $\overline{V}(X) \leq \overline{P}[(X - c)^2]$, $\forall c \in \mathbb{R}$ [18, Appendix G].

In the sequel of this section, *we shall assume coherence* of $\underline{P}$, $\overline{P}$. Coherence lets us bound the difference $|\mu[(X - \underline{P}(X))^2] - \mu[(X - \overline{P}(X))^2]|$, with $\mu = \underline{P}$ or alternatively $\mu = \overline{P}$:

Proposition 5. *If $\underline{P}$, $\overline{P}$ are coherent on $\mathcal{D} \supseteq \{X, (X - \overline{P}(X))^2, (X - \underline{P}(X))^2\}$, then*

$$|\underline{P}[(X - \overline{P}(X))^2] - \underline{P}[(X - \underline{P}(X))^2]| \leq (\overline{P}(X) - \underline{P}(X))^2 \quad (9)$$

$$|\overline{P}[(X - \underline{P}(X))^2] - \overline{P}[(X - \overline{P}(X))^2]| \leq (\overline{P}(X) - \underline{P}(X))^2. \quad (10)$$

Proof. By positive homogeneity, translation invariance, and the property of coherent lower previsions $\underline{P}(X + Y) \geq \underline{P}(X) + \underline{P}(Y)$, we have that

$$\begin{aligned}
&\underline{P}[(X - \underline{P}(X))^2] = \underline{P}[((X - \overline{P}(X)) + (\overline{P}(X) - \underline{P}(X)))^2] \\
&\geq \underline{P}[(X - \overline{P}(X))^2] + (\overline{P}(X) - \underline{P}(X))^2 + 2(\overline{P}(X) - \underline{P}(X))\underline{P}(X - \overline{P}(X)) \\
&= \underline{P}[(X - \overline{P}(X))^2] - (\overline{P}(X) - \underline{P}(X))^2,
\end{aligned}$$

which gives $\underline{P}[(X-\overline{P}(X))^2]-\underline{P}[(X-\underline{P}(X))^2]\leq(\overline{P}(X)-\underline{P}(X))^2$.

Substituting X with $-X$ in the derivation above, we obtain

$$\underline{P}[(-X-\underline{P}(-X))^2]\geq\underline{P}[(-X-\overline{P}(-X))^2]-(\overline{P}(-X)-\underline{P}(-X))^2. \quad (11)$$

By conjugacy, it is $(-X-\underline{P}(-X))^2=(-X+\overline{P}(X))^2=(X-\overline{P}(X))^2$ and $(-X-\overline{P}(-X))^2=(X-\underline{P}(X))^2$, while $(\overline{P}(-X)-\underline{P}(-X))^2=(\overline{P}(X)-\underline{P}(X))^2$. Taking account of these equalities in (11), it holds that $\underline{P}[(X-\overline{P}(X))^2]-\underline{P}[(X-\underline{P}(X))^2]\geq-(\overline{P}(X)-\underline{P}(X))^2$. Therefore, (9) obtains.

The proof of (10) is similar, applying instead $\overline{P}(X+Y)\leq\overline{P}(X)+\overline{P}(Y)$ to $\overline{P}[(X-\underline{P}(X))^2]=\overline{P}[((X-\overline{P}(X))+(\overline{P}(X)-\underline{P}(X)))^2]$. □

Another follow-up of Bhatia-Davis (imprecise) bounds regards upper and lower ($\overline{C}(X,Y)$ and $\underline{C}(X,Y)$, respectively) covariances, as defined in [13]:

$$\underline{C}(X,Y)=\min_{P\in\mathcal{M}}\{C_P(X,Y)\},\ \overline{C}(X,Y)=\max_{P\in\mathcal{M}}\{C_P(X,Y)\},$$

with $C_P(X,Y)=P[(X-P(X))(Y-P(Y))]=P(XY)-P(X)P(Y)$, $\forall P\in\mathcal{M}$.

The starting point is a bivariate generalisation of Bhatia-Davis inequality (6). In a form involving expectations, it was proven in [8], while a shorter proof is given in [12]. Here we give an alternative proof of this result, tailored for dF-coherent previsions.

Theorem 1 (Bivariate Bhatia-Davis inequalities). *Given a dF-coherent prevision P on $\mathcal{D}\supseteq\{X,Y,XY\}$, it holds that*

$$C_P(X,Y)\leq\min[(P(X)-m_X)(M_Y-P(Y)),(M_X-P(X))(P(Y)-m_Y)] \quad (12)$$

$$C_P(X,Y)\geq-\min[(P(X)-m_X)(P(Y)-m_Y),(M_X-P(X))(M_Y-P(Y))]. \quad (13)$$

Proof. For $X,Y,XY\in\mathcal{D}$, consider the gain

$$G=(XY-P(XY))-m_Y(X-P(X))-M_X(Y-P(Y))$$

which, by Definition 1, is such that $\sup G\geq 0$.

We rewrite G as

$$G=-P(XY)+m_YP(X)+M_XP(Y)-M_Xm_Y+Z,$$

where $Z=XY-m_YX-M_XY+M_Xm_Y=(X-M_X)(Y-m_Y)\leq 0$.

It follows that

$$\begin{aligned}0\leq\sup G&=-P(XY)+m_YP(X)+M_XP(Y)-M_Xm_Y+\sup Z\\&\leq-P(XY)+m_YP(X)+M_XP(Y)-M_Xm_Y\end{aligned}$$

and hence

$$\begin{aligned}C_P(X,Y)&=P(XY)-P(X)P(Y)\\&\leq m_YP(X)+M_XP(Y)-M_Xm_Y-P(X)P(Y) \quad (14)\\&=(M_X-P(X))(P(Y)-m_Y).\end{aligned}$$

Since $C_P(X,Y) = C_P(Y,X)$, we obtain from (14): $C_P(X,Y) = C_P(Y,X) \le (M_Y - P(Y))(P(X) - m_X)$, which completes the proof of (12).

As for (13), recalling that $m_{-X} = -M_X$, $M_{-X} = -m_X$, $C_P(-X,Y) = -C_P(X,Y)$ and using (12), we obtain

$$\begin{aligned} -C_P(X,Y) = C_P(-X,Y) &\le \min[\,(P(-X) - m_{-X})(M_Y - P(Y)),\\ &\qquad (M_{-X} - P(-X))(P(Y) - m_Y)]\\ &= \min[\,(M_X - P(X))(M_Y - P(Y)),\\ &\qquad (P(X) - m_X)(P(Y) - m_Y)], \end{aligned}$$

from which (13) follows immediately. □

The following generalisation of Theorem 1 to lower and upper covariances is proven in [12]:

Theorem 2. *Let $\underline{P}$ and its conjugate $\overline{P}$ be coherent on $\mathcal{D} \supseteq \{X,Y\}$. Then,*

$$\begin{aligned} \underline{C}(X,Y) &\le \min\{\,(\underline{P}(X) - m_X)(M_Y - \underline{P}(Y)), (M_Y - \overline{P}(Y))(\overline{P}(X) - m_X),\\ &\qquad (M_X - \overline{P}(X))(\overline{P}(Y) - m_Y), (M_X - \underline{P}(X))(\underline{P}(Y) - m_Y)\}\\ \underline{C}(X,Y) &\ge -\min\{\,(\overline{P}(X) - m_X)(\overline{P}(Y) - m_Y), (M_X - \underline{P}(X))(M_Y - \underline{P}(Y))\}\\ \overline{C}(X,Y) &\le \min\{\,(M_X - \underline{P}(X))(\overline{P}(Y) - m_Y), (\overline{P}(X) - m_X)(M_Y - \underline{P}(Y))\}\\ \overline{C}(X,Y) &\ge -\min\{\,(M_X - \overline{P}(X))(M_Y - \underline{P}(Y)), (M_X - \underline{P}(X))(M_Y - \overline{P}(Y)),\\ &\qquad (\underline{P}(X) - m_X)(\overline{P}(Y) - m_Y), (\overline{P}(X) - m_X)(\underline{P}(Y) - m_Y)\}. \end{aligned}$$

The bounds of Theorem 2 are very general, in the sense that they require knowledge of $\underline{P}(X), \underline{P}(Y), \overline{P}(X), \overline{P}(Y)$ only to be applied. Of course, the counterpart for this is that they cannot be tight, in general, and might be rather loose in presence of additional information. Just think, for instance, of the case that the expert is sure that $\underline{C}(X,Y) > 0$. Then, the second bound in Theorem 2 is trivially true, as well as the fourth (since $\overline{C}(X,Y) \ge \underline{C}(X,Y)$).

In the case that not even $\underline{P}(\cdot)$, $\overline{P}(\cdot)$ are known with a reasonable accurateness or confidence, some weaker bounds than Bhatia-Davis are available. In particular (see [12] for details), the following bivariate extension of *Popoviciu's inequality*:

Lemma 2. *Given $\mathcal{D} \supseteq \{X,Y\}$, domain of a coherent lower prevision $\underline{P}$ and its conjugate $\overline{P}$, it holds that*

$$-\frac{(M_X - m_X)(M_Y - m_Y)}{4} \le \underline{C}(X,Y) \le \overline{C}(X,Y) \le \frac{(M_X - m_X)(M_Y - m_Y)}{4}. \tag{15}$$

When $X = Y$, (15) boils down to $\underline{V}(X) \le \overline{V}(X) \le \frac{(M_X - m_X)^2}{4}$. When $\underline{V}(X) = \overline{V}(X) = \sigma_X^2$, we reobtain Popoviciu's inequality [2].

From the perspective of the degree of information required, inequalities (15) are extreme: no uncertainty evaluation of X, Y is needed, nor a complete knowledge of their image sets, but just their infima and suprema. Interestingly, it is possible to derive some, so to say, *hybrid inequalities*, that require an intermediate uncertainty knowledge of X, Y, between the minimal one of (15) and Bhatia-Davis inequalities in Theorem 2. Precisely, a *qualitative* knowledge of which is larger between either $\underline{P}(X)$ or $\overline{P}(X)$ and either $\pm\underline{P}(Y)$ or $\pm\overline{P}(Y)$ is sufficient to apply the next result:

Proposition 6. *Let $\underline{P}$, $\overline{P}$ be coherent on $\mathcal{D} \supseteq \{X, Y\}$. It holds that*

$$\underline{C}(X,Y) \leq \begin{cases} \min\left\{\frac{(M_Y - m_X)^2}{4}, M_{P_0}\right\} \text{ if } \underline{P}(X) \leq \underline{P}(Y) \text{ or } \overline{P}(X) \leq \overline{P}(Y) \\ \min\left\{\frac{(M_X - m_Y)^2}{4}, M_{P_0}\right\} \text{ if } \underline{P}(X) \geq \underline{P}(Y) \text{ or } \overline{P}(X) \geq \overline{P}(Y) \end{cases} \tag{16}$$

$$\overline{C}(X,Y) \leq \begin{cases} \min\left\{\frac{(M_Y - m_X)^2}{4}, M_{P_0}\right\} \text{ if } \overline{P}(X) \leq \underline{P}(Y) \\ \min\left\{\frac{(M_X - m_Y)^2}{4}, M_{P_0}\right\} \text{ if } \underline{P}(X) \geq \overline{P}(Y) \end{cases} \tag{17}$$

$$\overline{C}(X,Y) \geq \begin{cases} -\min\left\{\frac{(M_X + M_Y)^2}{4}, M_{P_0}\right\} \text{ if } \underline{P}(X) \geq -\underline{P}(Y) \\ -\min\left\{\frac{(m_X + m_Y)^2}{4}, M_{P_0}\right\} \text{ if } \overline{P}(X) \leq -\overline{P}(Y) \end{cases} \tag{18}$$

$$\underline{C}(X,Y) \geq \begin{cases} -\min\left\{\frac{(M_X + M_Y)^2}{4}, M_{P_0}\right\} \text{ if } \begin{cases} \overline{P}(X) \geq -\underline{P}(Y) \text{ or} \\ \underline{P}(X) \geq -\overline{P}(Y) \end{cases} \\ -\min\left\{\frac{(m_X + m_Y)^2}{4}, M_{P_0}\right\} \text{ if } \begin{cases} \overline{P}(X) \leq -\underline{P}(Y) \text{ or} \\ \underline{P}(X) \leq -\overline{P}(Y) \end{cases} \end{cases} \tag{19}$$

where $M_{P_0} = \frac{(M_X - m_X)(M_Y - m_Y)}{4}$ is the upper bound in (15).

Proof. Inequality (16) is proven in [12].

Proof of (17). If $\overline{P}(X) \leq \underline{P}(Y)$, from the third inequality in Theorem 2 and applying then the Average Mean-Geometric Mean inequality $ab \leq (\frac{a+b}{2})^2$, we obtain $\overline{C}(X,Y) \leq (\overline{P}(X) - m_X)(M_Y - \underline{P}(Y)) \leq (\underline{P}(Y) - m_X)(M_Y - \underline{P}(Y)) \leq \frac{(M_Y - m_X)^2}{4}$.

If $\underline{P}(X) \geq \overline{P}(Y)$, apply this result to $\overline{C}(Y,X)$ to prove that $\overline{C}(X,Y) = \overline{C}(Y,X) \leq \frac{(M_X - m_Y)^2}{4}$. This, recalling also Lemma 2, proves (17).

Proof of (18). Using the property $\overline{C}(-X,Y) = -\overline{C}(X,Y)$, which is easy to prove, apply (17) to $\overline{C}(-X,Y)$. Simple computations, exploiting conjugacy and the properties $m_{-X} = -M_X$, $M_{-X} = -m_X$, give (18).

The proof of (19) is similar to that of (18), using $\underline{C}(-X,Y) = -\underline{C}(X,Y)$ and (16). □

Note that inequalities (17) and (18) are more restrictive than (16) and (19), respectively: when (17) (or (18)) obtains, (16) (or (19)) obtains too.

5 Further Inequalities

As already mentioned in the Introduction, little attention has been so far paid in the literature to probability inequalities with coherent lower/upper previsions, and even less with 2-coherent ones. Some inequalities that may be expressed by means of Lebesgue integrals in classical probability theory have instead been extended employing other kinds of integrals. It is the case of *Jensen's inequality*, that has been studied by means of Choquet and other integrals (see e.g. [9, 14, 19]). Jensen's inequality has been studied for imprecise previsions too in [11, Sections 3, 4], showing that 2-coherence is again sufficient to obtain versions of the inequality. We report the main result:

Theorem 3 (Jensen's inequalities). *Let $\underline{P}$, $\overline{P}$ be 2-coherent on $\mathcal{D}$, $X \in \mathcal{D}$, $I \subset \mathbb{R}$ an interval that contains the image set of X, $\phi : I \to \mathbb{R}$ a convex function, $\psi : I \to \mathbb{R}$ a concave function, with right (left) derivatives at x, respectively, $\phi'_+(x)$, $\psi'_+(x)$ ($\phi'_-(x)$, $\psi'_-(x)$). Let $\underline{P}(X)$, $\overline{P}(X)$ be interior points of I and $\phi(X), \psi(X) \in \mathcal{D}$. Then,*

$$\underline{P}(\psi(X)) \leq \min\{\psi(\underline{P}(X)), \psi(\overline{P}(X))\},\ \overline{P}(\phi(X)) \geq \max\{\phi(\underline{P}(X)), \phi(\overline{P}(X))\}.$$

Besides,

$$\begin{aligned}
&\textit{if } \phi'_+(\underline{P}(X)) \geq 0 \textit{ then } \underline{P}(\phi(X)) \geq \phi(\underline{P}(X)),\\
&\textit{if } \phi'_-(\overline{P}(X)) \leq 0 \textit{ then } \underline{P}(\phi(X)) \geq \phi(\overline{P}(X));\\
&\textit{if } \psi'_-(\overline{P}(X)) \geq 0 \textit{ then } \overline{P}(\psi(X)) \leq \psi(\overline{P}(X)),\\
&\textit{if } \psi'_+(\underline{P}(X)) \leq 0 \textit{ then } \overline{P}(\psi(X)) \leq \psi(\underline{P}(X)).
\end{aligned}$$

It is also possible to generalise an improvement to Jensen's inequality given in [1], cf. [11, Section 3.2]. While the original improvement applies in the precise case and when X takes values in $\mathbb{Z}$, the version in [11] requires 2-coherence and much less restrictive conditions on the image set of X.

Additional follow-ups of Jensen's inequalities include versions of Liapunov's inequality and, requiring coherence, applications to the moment problem [11, Section 4].

When seeking tail bounds to a gamble X, *Cantelli's inequalities* [3, 17] compete with Markov's. They state that, $\forall \varepsilon > 0$,

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

Thus and unlike Markov's inequalities, they do not require non-negativity of X, but involve its variance. Extensions of Cantelli's inequalities to imprecise previsions are introduced in [11, Section 6]. It is interesting to remark that:

i) Unlike Markov's inequalities, the most general versions of Cantelli's inequalities do not necessarily require 2-coherence, but some partly different conditions, see [11, Proposition 6.2].
ii) When $\underline{P}$ or $\overline{P}$ are coherent, Cantelli's inequalities involve the upper and lower variances $\overline{V}(X)$, $\underline{V}(X)$, as appears from the next proposition [11, Proposition 6.3].

Proposition 7. *Let X be a gamble, $\varepsilon > 0$, $\underline{P}$, $\overline{P}$ conjugate and coherent on $\mathcal{D} \supseteq \{X, (X \leq \underline{P}(X) - \varepsilon), (X \leq \overline{P}(X) - \varepsilon), (X \geq \underline{P}(X) + \varepsilon), (X \geq \overline{P}(X) + \varepsilon)\}$. Then,*

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

From the upper line formulae of Proposition 7, it appears that Cantelli's inequalities may bound from above the left tail of the lower distribution function of X, $\underline{F}_X(x) = \underline{P}(X \leq x)$. Applying conjugacy to the lower line inequalities gives formulae bounding from below the right tail of the upper distribution function $\overline{F}_X(x) = \overline{P}(X \leq x)$. Thus, these inequalities may play a role when the p-box $(\underline{F}_X, \overline{F}_X)$ is imprecisely known. When inadequately known, $\underline{V}(X)$, $\overline{V}(X)$ can be replaced by, respectively, $\underline{P}[(X - \mu(X))^2]$ and $\overline{P}[(X - \mu(X))^2]$ in the formulae of Proposition 7 ($\mu \in \{\underline{P}, \overline{P}\}$), because of the arithmetic inequality $\frac{a}{a+b} \leq \frac{c}{c+b}$ for $a, b, c > 0, a < c$. By the same inequality, $\underline{P}[(X - \mu(X))^2]$ and $\overline{P}[(X - \mu(X))^2]$ can be replaced, if unknown, by their Bhatia-Davis bounds (Propositions 4, 5).

6 Conclusions

The results in this paper confirm that more versions of probability inequalities may be available with imprecise probabilities. Their expression and potential applications depend also on the degree of consistency required. Often, 2-coherence alone suffices. Anyway, this should not be taken as a rule: the inequalities we considered are based on evaluations on very few gambles, one or two each time. This is in line with the definition of 2-coherence, which considers at most two gambles simultaneously. Future work includes other types of inequalities, or more complex situations when a higher number of gambles is involved, or specific assumptions (like some forms of stochastic independence) are made.

Acknowledgments. We are grateful to the reviewers for their constructive comments. The authors are affiliated with the GNAMPA-INdAM research group.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bethmann, D.: An improvement to Jensen's inequality and its application to mating market clearing when paternity is uncertain. Mathematical Social Sciences **91**, 71–74 (2018). https://doi.org/10.1016/j.mathsocsci.2017.08.004
2. Bhatia, R., Davis, C.: A Better Bound on the Variance. The American Mathematical Monthly **107**(4), 353–357 (2000). https://doi.org/10.1080/00029890.2000.12005203

3. Cantelli, F.P.: Sui confini della probabilità. In: Atti del Congresso Internazionale dei Matematici. vol. 6, pp. 47–59. Zanichelli, Bologna, Italy (1929)
4. Cozman, F.G.: Concentration inequalities and laws of large numbers under epistemic and regular irrelevance. International Journal of Approximate Reasoning **51**(9), 1069–1084 (2010). https://doi.org/10.1016/j.ijar.2010.08.009
5. de Cooman, G., Miranda, E.: Weak and strong laws of large numbers for coherent lower previsions. Journal of Statistical Planning and Inference **138**(8), 2409–2432 (2008). https://doi.org/https://doi.org/10.1016/j.jspi.2007.10.020
6. Dolecki, S., Greco, G.H.: Niveloids. Topological Methods in Nonlinear Analysis **5**(1), 1–22 (1995). https://doi.org/10.12775/TMNA.1995.001
7. de Finetti, B.: Theory of Probability: A Critical Introductory Treatment. John Wiley, New York (1974)
8. Hössjer, O., Sjölander, A.: Sharp lower and upper bounds for the covariance of bounded random variables. Statistics & Probability Letters **182**, 109323 (2022). https://doi.org/https://doi.org/10.1016/j.spl.2021.109323
9. Mesiar, R., Li, J., Pap, E.: The Choquet integral as Lebesgue integral and related inequalities. Kybernetika **46**(6), 1098–1107 (2010)
10. Pelessoni, R., Vicig, P.: 2-coherent and 2-convex conditional lower previsions. International Journal of Approximate Reasoning **77**, 66–86 (2016). https://doi.org/10.1016/j.ijar.2016.06.003
11. Pelessoni, R., Vicig, P.: Jensen's and Cantelli's inequalities with imprecise previsions. Fuzzy Sets and Systems **458**, 50–68 (2023). https://doi.org/10.1016/j.fss.2022.06.021
12. Pelessoni, R., Vicig, P.: Bhatia-Davis inequalities for lower and upper previsions and covariances (submitted)
13. Quaeghebeur, E.: Lower and upper covariance. In: Dubois, D., et al. (eds.) Soft Methods for Handling Variability and Imprecision. pp. 323–330. Springer (2008). https://doi.org/10.1007/978-3-540-85027-4_39
14. Román-Flores, H., Flores-Franulič, A., Chalco-Cano, Y.: A Jensen type inequality for fuzzy integrals. Information Sciences **177**(15), 3192–3201 (2007). https://doi.org/10.1016/j.ins.2007.02.006
15. Salamanca, J.J., Couso, I.: The minimum variance of a random set on a Euclidean space. Fuzzy Sets and Systems **443**, 106–126 (2022). https://doi.org/10.1016/j.fss.2021.11.014
16. Troffaes, M., Basu, T.: A Cantelli-type inequality for constructing non-parametric p-boxes based on exchangeability. In: De Bock, J., et al. (eds.) Proc. of the 11th International Symposium on Imprecise Probabilities: Theories and Applications. Proc. of Machine Learning Research, vol. 103, pp. 386–393. PMLR (3–6 July 2019)
17. Uspensky, J.V.: Introduction to mathematical probability. McGraw-Hill (1937)
18. Walley, P.: Statistical reasoning with imprecise probabilities. Chapman and Hall, London; New York (1991)
19. Zhang, D., Mesiar, R., Pap, E.: Jensen's inequality for Choquet integral revisited and a note on Jensen's inequality for generalized Choquet integral. Fuzzy Sets and Systems **430**, 79–87 (2022). https://doi.org/10.1016/j.fss.2021.09.004