International Journal of Approximate Reasoning 192 (2026) 109654

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Cauchy-Schwarz inequalities from the viewpoint of lower and upper previsions

Paolo Vicig *, Renato Pelessoni *

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

ARTICLE INFO

Keywords:
Cauchy-Schwarz inequality
Coherent and 2-coherent imprecise previsions
Centered convex previsions
Imprecise correlation

ABSTRACT

In recent papers we investigated how some classical probability inequalities still obtain, in a modified form, when bounded random numbers, also termed gambles, are evaluated by means of upper/lower (imprecise) previsions satisfying various consistency criteria. This kind of analysis is developed in this paper with reference to the Cauchy-Schwarz (CS) inequality. We show that two among the potential extensions of the CS inequality obtain without restrictions when the upper/lower previsions are coherent. We then consider weaker consistency requirements than coherence. When the imprecise previsions avoid sure loss or are centered convex, these two inequalities may fail to hold; nevertheless, we give sufficient conditions for them to apply. With 2-coherent upper/lower previsions, we provide several sufficient conditions for the two inequalities to hold, limiting the instances where they fail to rather peculiar cases. We also show that it is possible to obtain, in certain instances, information about the correlation of two gambles without requiring coherence.

1. Introduction

The problem of establishing if classical probability inequalities still obtain, in case in modified versions, within the theory of Imprecise Probabilities is at present only partially explored. It has instead been more extensively investigated expressing the inequalities by means of various non-additive integrals (cf., among others, [1–6]). These two approaches are largely distinct and with limited overlap, as discussed for the case of Jensen's inequality in Pelessoni and Vicig [7, Section 3.1]. In fact, the role of (non-additive) integrals is secondary within the theory of Imprecise Probabilities: *bounded* random numbers, termed *gambles*, are primarily evaluated through upper or lower previsions. Imprecise (upper and lower) previsions are uncertainty measures extending the notion of (precise) prevision, suggested by de Finetti [8]. Roughly speaking, a prevision for a gamble X, usually termed $P(X)$, is a real number summarising an agent's uncertain beliefs on X. Such a number may be difficult to assess, or unreliable, which leads to its replacement with a lower prevision (customarily identified with $\underline{P}(X)$) or an upper prevision ($\overline{P}(X)$). Of course, some rule or *consistency* notion is needed to prevent inconsistent evaluations, such as $\overline{P}(X) < \inf X$. Moreover, the rule applies to any evaluation on an *arbitrary* set of gambles. The prominent consistency concept for imprecise previsions is coherence [9], boiling down to de Finetti's coherence [8] (shortly: *dF-coherence*) with precise previsions (these and other concepts recalled in the Introduction will be properly defined in Section 2).

Whenever an expectation $E(X)$ is assessed for a gamble X, it is a dF-coherent prevision $P(X)$ for X [9, Sections 3.2.1, 3.2.2]. Some papers presenting results for inequalities involving gambles and coherent upper or lower previsions include [10–12].

* Corresponding authors.
E-mail addresses: paolo.vicig@deams.units.it (P. Vicig), renato.pelessoni@deams.units.it (R. Pelessoni).

https://doi.org/10.1016/j.ijar.2026.109654
Received 5 October 2025; Received in revised form 11 February 2026; Accepted 14 February 2026
Available online 18 February 2026
0888-613X/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

However, a distinguishing feature of most probability inequalities is that they ask for much weaker properties than those of (precise) probabilities. This is often interpreted in probability theory pointing out their *robustness*, i.e. the fact that they apply also with a partial knowledge of the underlying probability distributions. Yet, even coherence might be too strong a requirement for some inequalities. It is then interesting to ascertain whether, and to what extent, a given inequality still obtains (with coherence and) also with weaker consistency requirements.

In recent papers [7,13,14] we performed this analysis for Markov's, Jensen's, Cantelli's, Bhatia-Davis' and other inequalities. In this paper we focus on extending the classical Cauchy-Schwarz (shortly: CS) inequality

$$E(XY)^2 \leq E(X^2)E(Y^2), \tag{1}$$

where X, Y are gambles in our frame. To the best of our knowledge, the CS inequality has not yet been explored through the lens of imprecise previsions.

At a very preliminary stage, we observe that we should speak of CS inequalities and not of a single CS inequality in an imprecise setting, since there are potentially more candidates to extend (1). In fact, each expectation in (1) might be replaced by either a lower or an upper prevision. Thus, the general CS expression would be

$$\mu(XY)^2 \leq \mu'(X^2)\mu''(Y^2), \tag{2}$$

with each of μ, μ', μ'' being either $\underline{P}$ or $\overline{P}$, i.e. $\mu, \mu', \mu'' \in \{\underline{P}, \overline{P}\}$. We shall see later on, however, that not all alternatives for μ, μ', μ'' in (2) are viable.

The paper is organised as follows. Section 2 gathers basic definitions and properties of upper and lower previsions, focusing on those results that are employed in the paper. Section 3 is concerned with coherent previsions. It is natural to start our investigation with coherence because of its relevance, but also considering that it satisfies stronger properties than the other, weaker, consistency notions we consider. This means that the results for coherence give a useful indication as to what kinds of CS inequalities might possibly obtain when relaxing it. The main result in Section 3 is Proposition 3.1, proving the following CS inequality:

$$\overline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2). \tag{3}$$

While (3) can be proven in more alternative ways, some partly resorting to known techniques for proving (1), we give a standalone proof. It is based on coherence (Definition 2.1 (b)) and some of its properties, and lets us obtain, as a relevant by-product, the CS inequality for dF-coherent previsions:

$$P(XY)^2 \leq P(X^2)P(Y^2).$$

Next, we prove (Proposition 3.2) the CS inequality:

$$\underline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2), \tag{4}$$

while we show in Example 3.1 that the CS inequalities (2) do not generally hold when at least one of μ', μ'' is a lower prevision. In Section 4 we consider the role of the CS inequalities (3), (4) with the notions of avoiding sure loss (ASL), Definition 2.1 (d), and of centered convexity (C-convexity), Definition 2.1 (e). Both concepts are weaker than coherence, with C-convexity stronger than ASL. It turns out that (3), (4) obtain only with some additional conditions, more restrictive with ASL. In Section 5 we investigate the CS inequalities (3), (4) with a different relaxation of coherence, 2-coherence (Definition 2.1 (c)). With 2-coherence, the CS inequality (3) (and similarly (4)) applies in a number of cases, as shown by the many sufficient conditions we derive (Propositions 5.1,5.3,5.4,5.6,5.7), but not always. We hedge its non-applicability to a very special situation (Proposition 5.5). The role within our approach of the Choquet integral is analysed in Section 5.1.2. In Section 6 we suggest an application of CS inequalities, similar to the classical one giving the correlation coefficient. It enables us to evaluate the (imprecise) correlation between two gambles in certain cases, and may be especially useful when $\overline{P}$ or $\underline{P}$ are not coherent. Section 7 concludes the paper.

2. Preliminaries

The theory of lower and upper previsions has been mainly developed with reference to *gambles*, i.e. *bounded* random numbers [9,15–18]. In general, an upper (lower, precise) prevision $\overline{P}$ ($\underline{P}$, P) is a map $\overline{P} : \mathbb{D} \to \mathbb{R}$ ($\underline{P} : \mathbb{D} \to \mathbb{R}$, $P : \mathbb{D} \to \mathbb{R}$), where $\mathbb{D}$ is an *arbitrary* (non-empty) set of gambles. It is possible to refer to lower, or alternatively upper, previsions only by means of the customary *conjugacy* relationship

$$\overline{P}(X) = -\underline{P}(-X), \tag{5}$$

with X, $-X$ gambles. In the sequel, we shall be more concerned with upper previsions, but we shall employ lower previsions too, especially when they make certain inequalities more expressive. For instance, Eq. (4) requires assessing a lower prevision on XY and an upper prevision on X^2, Y^2, so that we have an assessment on $\{XY, X^2, Y^2\}$. By conjugacy, this is equivalent to assessing an upper prevision only, on $\{-XY, X^2, Y^2\}$, with $\overline{P}(-XY) = -\underline{P}(XY)$.

To be theoretically significant and operationally useful, an upper/lower prevision must obey some consistency requirement. More consistency notions have been proposed [9,19–21], coherence being the prominent one. It is derived from de Finetti's notion of coherence for (precise) previsions [8], termed here *dF-coherence*. Both dF-coherence and the other consistency concepts in this paper (coherence, 2-coherence, C-convexity, avoiding sure loss) may be defined in a similar way, as shown in the next Definition [8,9,19,20]:

Definition 2.1.

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

(b) $\overline{P} : \mathbb{D} \to \mathbb{R}$ is a coherent upper prevision on $\mathbb{D}$ iff, $\forall n \in \mathbb{N}, \forall s_0 \in \mathbb{R}$, $\forall s_1 \dots, s_n \geq 0$, $\forall X_0, \dots, X_n \in \mathbb{D}$, defining $\overline{G} = \sum_{i=1}^{n} s_i(\overline{P}(X_i) - X_i) - s_0(\overline{P}(X_0) - X_0)$, we have that $\sup \overline{G} \geq 0$.

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

(d) $\overline{P} : \mathbb{D} \to \mathbb{R}$ avoids sure loss (ASL) on $\mathbb{D}$ iff, $\forall n \in \mathbb{N}$, $\forall\ s_0, \dots, s_n$ real and non-negative, $\forall X_0, \dots, X_n \in \mathbb{D}$, defining $\overline{G} = \sum_{i=0}^{n} s_i(\overline{P}(X_i) - X_i)$, we have that $\sup \overline{G} \geq 0$.

(e) $\overline{P} : \mathbb{D} \to \mathbb{R}$ is a C-convex upper prevision on $\mathbb{D} \cup \{0\}$ iff $\overline{P}(0) = 0$ and $\forall n \in \mathbb{N}$, $\forall s_1 \dots, s_n \geq 0$ such that $\sum_{i=1}^{n} s_i = 1$, $\forall X_0, \dots, X_n \in \mathbb{D}$, defining $\overline{G} = \sum_{i=1}^{n} s_i(\overline{P}(X_i) - X_i) - (\overline{P}(X_0) - X_0)$, we have that $\sup \overline{G} \geq 0$.

We recall that similar definitions (cases (b), (c), (d) and (e)) can be given also for lower previsions, using (5). For instance, the definition of coherent lower prevision is provided below:

Definition 2.2. $\underline{P} : \mathbb{D} \to \mathbb{R}$ is a coherent lower prevision on $\mathbb{D}$ iff, $\forall n \in \mathbb{N}, \forall s_0 \in \mathbb{R}$, $\forall s_1 \dots, s_n \geq 0$, $\forall X_0, \dots, X_n \in \mathbb{D}$, defining $\underline{G} = \sum_{i=1}^{n} s_i(X_i - \underline{P}(X_i)) - s_0(X_0 - \underline{P}(X_0))$, we have that $\sup \underline{G} \geq 0$.

Further, $\underline{P}$ satisfies a certain consistency requirement on $\mathbb{D}$ if and only if its conjugate $\overline{P}$ does on $-\mathbb{D} = \{-X : X \in \mathbb{D}\}$.

Formally, the above definitions are all variants (due to different constraints on n and/or the sign of the coefficients s_i) of a unique pattern, having the well known interpretation of an idealised betting scheme [8,9]. This scheme requires that the gain (G, $\overline{G}$, $\overline{G}_2$, $\underline{G}$) from buying/selling gambles in $\mathbb{D}$ with certain rules and prices (P, $\overline{P}$, $\underline{P}$) must not be uniformly negative.

For our purposes, in the following we will employ some characterisations of the notions of coherence, avoiding sure loss, and C-convexity in terms of sets of dF-coherent previsions:

Theorem 2.1 (Characterisation theorems [9,19])**.**

(a) (Envelope theorem) *An upper prevision $\overline{P}$ (a lower prevision $\underline{P}$) is* coherent *on an arbitrary set of gambles $\mathbb{D}$ iff there exists a non-empty set $\mathcal{M}^*_{\overline{P}}$ ($\mathcal{M}^*_{\underline{P}}$) of dF-coherent previsions on $\mathbb{D}$ such that*

$$\overline{P}(X) = \sup_{P \in \mathcal{M}^*_{\overline{P}}} P(X) \qquad (\underline{P}(X) = \inf_{P \in \mathcal{M}^*_{\underline{P}}} P(X)), \forall X \in \mathbb{D}.$$

The supremum is in particular achieved when $\mathcal{M}^*_{\overline{P}}$ coincides with the credal set *of $\overline{P}$, $\mathcal{M} = \{P : P \text{ is dF-coherent on } \mathbb{D}, P \leq \overline{P}\}$ (similarly with the infimum).*

(b) (Dominance theorem) *$\overline{P}$ ($\underline{P}$) avoids sure loss on $\mathbb{D}$ iff there exists a dF-coherent prevision P such that $\overline{P}(X) \geq P(X)$, $\forall X \in \mathbb{D}$ ($\underline{P}(X) \leq P(X)$, $\forall X \in \mathbb{D}$).*

(c) (Generalised envelope theorem[1]) *$\overline{P}$ is C-convex on $\mathbb{D} \cup \{0\}$ iff there exist a set $\mathcal{P}$ of previsions dF-coherent on $\mathbb{D}$ and a function $\alpha : \mathcal{P} \to \mathbb{R}$ such that*

$$\overline{P}(X) = \max_{P \in \mathcal{P}} \{P(X) - \alpha(P)\}, \forall X \in \mathbb{D}, \text{ and } \min_{P \in \mathcal{P}} \alpha(P) = 0. \tag{6}$$

It is almost immediate from Definitions 2.1, 2.2 and Theorem 2.1 to deduce the following. A dF-coherent prevision is also coherent as an upper prevision, and as a lower prevision too. Next, coherence implies C-convexity, the condition of avoiding sure loss (ASL) and 2-coherence. It is also known that C-convexity implies ASL [19, Proposition 3.5 (e)]. There is instead no implication between ASL and 2-coherence, see for this the beginning of Section 5. These relationships are visualized in Fig. 1.

In the sequel, we shall often employ the following properties, that are necessary conditions for 2-coherence of $\underline{P}$ (and its conjugate $\overline{P}$) [20]:

$$\begin{aligned} &\underline{P}(X + c) = \underline{P}(X) + c, \\ &\overline{P}(X + c) = \overline{P}(X) + c \end{aligned} \quad \forall X \in \mathbb{D}, \forall c \in \mathbb{R} \text{ (Translation invariance)} \tag{7}$$

$$X \leq Y \Rightarrow \underline{P}(X) \leq \underline{P}(Y), \overline{P}(X) \leq \overline{P}(Y), \forall X, Y \in \mathbb{D} \text{ (Monotonicity)} \tag{8}$$

$$\overline{P}(\lambda X) = \lambda \overline{P}(X), \forall X \in \mathbb{D}, \forall \lambda \geq 0 \text{ (Positive homogeneity)} \tag{9}$$

$$\overline{P}(\lambda X) \geq \lambda \overline{P}(X), \forall X \in \mathbb{D}, \forall \lambda \leq 0 \text{ (Negative homogeneity)} \tag{10}$$

$$\inf X \leq \underline{P}(X) \leq \overline{P}(X) \leq \sup X, \forall X \in \mathbb{D} \text{ (Internality)} \tag{11}$$

[1] Stated for upper previsions. The version for lower previsions is found in [19, Theorem 3.3].

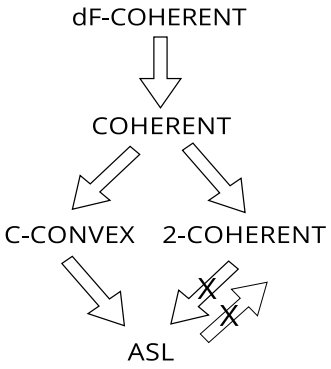

Fig. 1. Relationships between different notions of consistency.

Clearly, these axioms also apply to coherent and dF-coherent previsions, while coherence and dF-coherence satisfy additional or stronger properties. In particular, a dF-coherent prevision is additive and homogeneous:

$$P(X+Y) = P(X) + P(Y), P(\lambda X) = \lambda P(X), \forall \lambda \in \mathbb{R}. \tag{12}$$

Remark 1. DF-coherence, coherence, C-convexity and 2-coherence share an interesting feature: whenever a prevision satisfying one of these consistency properties is assessed on a set of gambles $\mathbb{D}$, it admits (at least) one extension with the same consistency requirement on any superset of gambles $\mathbb{D}' \supset \mathbb{D}$ [8,9,19,20]. For instance, a given upper prevision $\overline{P}$ 2-coherent on $\mathbb{D}$ may be extended as an upper prevision $\overline{P}'$, 2-coherent on $\mathbb{D}'$ and such that $\overline{P} = \overline{P}'$ on $\mathbb{D}$. It is also clear that the extension is often constrained: for instance, if $\overline{P}$ is (at least) 2-coherent on $\mathbb{D}$, $X \in \mathbb{D}$ while $X + c \notin \mathbb{D}$ ($c \in \mathbb{R}$), then necessarily $\overline{P}'(X + c) = \overline{P}(X) + c$ by translation invariance (7).

In the paper, we shall take advantage of this extension property for usually defining a lower/upper prevision on a minimal set $\mathbb{D}$, i.e. a set ensuring that just the previsions involved in the result at hand are defined. This is not restrictive, since we may always extend the given prevision if we wish or need to. In fact, the extension is performed (often implicitly) in several proofs, see e.g. Lemmas 3.1 and 3.2. On the other hand, this approach corresponds to many real-world situations, where uncertainty assessments are made with some difficulty and for sets of gambles that are arbitrary and typically not structured (i.e. they are generally not linear spaces, cones or other standardised sets).

Remark 2. A special extension will be employed in a couple of instances later on. In fact, given a lower prevision $\underline{P}$ that is coherent on $\mathbb{D}$, and a set of gambles $\mathbb{D}' \supset \mathbb{D}$, there always exists a coherent extension $\underline{E}$ of $\underline{P}$ on $\mathbb{D}'$ that is *least-committal*, meaning that if $\underline{P}'$ is any coherent extension of $\underline{P}$ on $\mathbb{D}'$, then $\underline{P}' \geq \underline{E}$ on $\mathbb{D}'$ (and $\underline{E} = \underline{P}$ on $\mathbb{D}$). $\underline{E}$ is the *natural extension* of $\underline{P}$, a concept investigated in Walley [9].

Replacing throughout 'coherent' with '2-coherent' above, we come to the analogous notion of 2-*coherent natural extension* $\underline{E}_{2C}$. Also $\underline{E}_{2C}$ always exists and is least-committal within the set of the 2-coherent extensions of $\underline{P}$ on $\mathbb{D}'$ [20,22]: for any $\underline{P}'$, 2-coherent extension of $\underline{P}$ on $\mathbb{D}'$, $\underline{P}' \geq \underline{E}_{2C}$ on $\mathbb{D}'$.

Finally, we recall that a prevision P (an upper prevision $\overline{P}$, a lower prevision $\underline{P}$) boils down to a probability (an upper probability, a lower probability) when applied to the indicator I_A of an event A. For instance, $P(I_A) = P(A)$ (the same for $\overline{P}$, $\underline{P}$).

More detailed information on the concepts recalled in this section may be found in [8,9,15–18,20,21,23].

3. Cauchy-Schwarz inequalities with coherent upper and lower previsions

The focus in this section is on establishing generalisations of the CS inequality (1) to coherent upper and lower (conjugate) previsions. However, a part of the results obtains also for 2-coherent previsions and will be stated in this form, in view of applying them for these measures in the later Section 5.

We first prove the CS inequality:

$$\overline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2).$$

This is done in Proposition 3.1. Prior to this, we prove the following preliminary lemmas.

Lemma 3.1. *Let $\overline{P}$ be a 2-coherent upper prevision on $\{XY, |XY|\}$, extended to $\{XY, |XY|, -|XY|\}$ preserving 2-coherence. Then*

$$|\overline{P}(XY)| \leq \overline{P}(|XY|). \tag{13}$$

Proof. Inequality (13) is equivalent to

$$-\overline{P}(|XY|) \leq \overline{P}(XY) \leq \overline{P}(|XY|).$$

The right-hand inequality is implied by monotonicity (8), the left-hand one by $\overline{P}(XY) \geq \overline{P}(-|XY|) \geq -\overline{P}(|XY|)$, obtained using (8) and then negative homogeneity (10). □

Lemma 3.2. *Let $\overline{P}$ be 2-coherent on $\{X^2, Y^2, XY\}$. If $\overline{P}(X^2) = 0$ then, extending $\overline{P}$ to $\{X^2, Y^2, XY, |XY|\}$ preserving 2-coherence, $\overline{P}(|XY|) = 0$ and $\overline{P}(XY) = 0$.*

Proof. We show first that, whatever is the 2-coherent extension of $\overline{P}$ to $|XY|$, $\overline{P}(|XY|) = 0$. For this, for any $\varepsilon > 0$, define the event $A_\varepsilon = (|XY| \geq \varepsilon)$. Applying Markov's inequality for 2-coherent upper previsions (see [7, Equation (36)]), monotonicity (8) and positive homogeneity (9) at, respectively, the second inequality, first and third inequality, fourth equality, we have that[2]

$$0 \leq \overline{P}(I_{A_\varepsilon}) = \overline{P}(A_\varepsilon) = \overline{P}(|XY| \geq \varepsilon) = \overline{P}(X^2Y^2 \geq \varepsilon^2) \leq \frac{\overline{P}(X^2Y^2)}{\varepsilon^2}$$
$$\leq \frac{\overline{P}(X^2 \sup(Y^2))}{\varepsilon^2} = \frac{\sup(Y^2)\overline{P}(X^2)}{\varepsilon^2} = 0,$$

and therefore

$$\overline{P}(A_\varepsilon) = 0, \forall \varepsilon > 0. \tag{14}$$

Since $|XY| = |XY|I_{|XY|\geq\varepsilon} + |XY|I_{|XY|<\varepsilon} \leq |XY|I_{|XY|\geq\varepsilon} + \varepsilon$, we get, using monotonicity (8), translation invariance (7), positive homogeneity (9) and lastly (14),

$$0 \leq \overline{P}(|XY|) \leq \overline{P}(|XY|I_{|XY|\geq\varepsilon} + \varepsilon) = \overline{P}(|XY|I_{|XY|\geq\varepsilon}) + \varepsilon$$
$$\leq \sup(|XY|)\overline{P}(I_{A_\varepsilon}) + \varepsilon = \varepsilon, \forall \varepsilon > 0.$$

Therefore, $\overline{P}(|XY|) = 0$.

That also $\overline{P}(XY) = 0$ follows then from this and Lemma 3.1 (cf. Equation (13)). □

Proposition 3.1. *Let $\overline{P}$ be a coherent upper prevision on $\{X^2, Y^2, XY\}$. Then, inequality (3) obtains:*

$$\overline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2).$$

Proof. If either $\overline{P}(X^2) = 0$ or $\overline{P}(Y^2) = 0$, (3) holds trivially, by Lemma 3.2.

Suppose then that $\overline{P}(X^2) > 0$ and $\overline{P}(Y^2) > 0$. Let us extend coherently $\overline{P}$ on $|XY|$ and consider the following gamble $\overline{G}$, that is one of the admissible gains $\overline{G}$ in Definition 2.1 (*b*):

$$\overline{G} = \overline{P}(Y^2)(\overline{P}(X^2) - X^2) + \overline{P}(X^2)(\overline{P}(Y^2) - Y^2) - 2\sqrt{\overline{P}(Y^2)\overline{P}(X^2)}(\overline{P}(|XY|) - |XY|)$$
$$= -\left(|X|\sqrt{\overline{P}(Y^2)} - |Y|\sqrt{\overline{P}(X^2)}\right)^2 + 2\overline{P}(X^2)\overline{P}(Y^2) - 2\overline{P}(|XY|)\sqrt{\overline{P}(X^2)\overline{P}(Y^2)}.$$

Coherence of $\overline{P}$ implies $\sup \overline{G} \geq 0$ and therefore

$$2\overline{P}(X^2)\overline{P}(Y^2) - 2\overline{P}(|XY|)\sqrt{\overline{P}(X^2)\overline{P}(Y^2)} \geq \inf\left(|X|\sqrt{\overline{P}(Y^2)} - |Y|\sqrt{\overline{P}(X^2)}\right)^2 \geq 0.$$

It ensues that $\overline{P}(|XY|) \leq \sqrt{\overline{P}(X^2)\overline{P}(Y^2)}$.

From here, and as a consequence of Eq. (13) at the first inequality, we obtain

$$\overline{P}(XY)^2 \leq \overline{P}(|XY|)^2 \leq \overline{P}(X^2)\overline{P}(Y^2).$$

□

Remark 3. The proof of inequality (3) could also be phrased in alternative ways, closer to those of the classical CS inequality (1) in Probability Theory (and additionally exploiting Theorem 2.1 (*a*)). We stress that our proof relies instead on the definition of coherence and some of its properties. Furthermore, it also applies to dF-coherent previsions, that is, as an important by-product, it proves the inequality

$$P(XY)^2 \leq P(X^2)P(Y^2) \tag{15}$$

with P a dF-coherent prevision on $\{X^2, Y^2, XY\}$. This is easily seen either by replacing $\overline{P}$ with P throughout the proof of Proposition 3.1 or simply recalling that every dF-coherent P is also a coherent upper prevision.

The proof of Proposition 3.1 does not extend to 2-coherent previsions: its gamble $\overline{G}$ is obtained as a sum of three terms, which is not allowed by Definition 2.1 (*c*).

To investigate other versions of CS inequalities for coherent upper/lower previsions, we state the following result (note that it does not require coherence).

[2] In the next derivation, it is understood that it obtains for every 2-coherent extension of $\overline{P}$ to the additional gambles I_{A_ε}, X^2Y^2, $X^2\sup(Y^2)$, cf. Remark 1. Analogously, $\overline{P}$ is 2-coherently extended also to $|XY|I_{|XY|\geq\varepsilon} + \varepsilon$, $|XY|I_{|XY|\geq\varepsilon}$ in the sequel of the proof. We shall not always mention explicitly this kind of extension in later similar passages.

Table 1
Data for Example 3.1.

$\mathbb{P}$	$X(\omega)$	$Y(\omega)$	$XY(\omega)$	P_1	P_2
ω_1	-1	y_1	$-y_1$	0	$\frac{1}{3}$
ω_2	0	y_2	0	1	$\frac{1}{3}$
ω_3	1	y_3	y_3	0	$\frac{1}{3}$

Proposition 3.2. *Let $\overline{P}$ be defined on $\{X^2, Y^2, -XY\}$ and let $\underline{P}$ be its conjugate. If $\overline{P}$ satisfies inequality (3), then inequality (4) holds as well:*

$$\underline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2).$$

Proof. Using the conjugacy equality $\underline{P}(XY) = -\overline{P}(-XY)$ and applying (3), we obtain $\underline{P}(XY)^2 = \overline{P}(-XY)^2 \leq \overline{P}((-X)^2)\overline{P}(Y^2) = \overline{P}(X^2)\overline{P}(Y^2)$. □

As an immediate consequence of Propositions 3.1 and 3.2, it holds that

Corollary 3.1. *If $\overline{P}$ is defined and coherent on $\{X^2, Y^2, -XY\}$, then the CS inequality (4) obtains.*

A subsequent doubt is whether inequality (4) or also inequality (3) can be made tighter, replacing at least one between $\overline{P}(X^2)$, $\overline{P}(Y^2)$ with a lower prevision there. The answer is negative, in general, as shown by the next example (but see also the later Lemma 5.2 for an interesting exception).

Example 3.1.

(a) Assign X, Y and P_1, P_2 on the partition $\mathbb{P} = \{\omega_1, \omega_2, \omega_3\}$ as in Table 1. P_1, P_2 originate the dF-coherent previsions (expectations)

$$P_1(X^2) = 0, P_1(XY) = 0, P_2(X^2) = \frac{2}{3}, P_2(XY) = \frac{-y_1}{3} + \frac{y_3}{3}.$$

Now suppose that $y_1 > y_3$. Next, consider the lower prevision $\underline{P}$ defined on $\{X^2, XY\}$ as $\underline{P}(Z) = \min\{P_1(Z), P_2(Z)\}$, $Z = X^2$ *or* $Z = XY$: $\underline{P}(X^2) = 0$, $\underline{P}(XY) = \frac{-y_1}{3} + \frac{y_3}{3} < 0$.

$\underline{P}$ is coherent on $\{X^2, XY\}$ by Theorem 2.1 (*a*). Yet, no matter how $\underline{P}$ is extended to a coherent assignment on $\{X^2, Y^2, XY\}$, it holds that $\underline{P}(XY)^2 = (\frac{-y_1}{3} + \frac{y_3}{3})^2 > 0 = \underline{P}(X^2)\underline{P}(Y^2)$.

Therefore the CS inequality $\underline{P}(XY)^2 \leq \underline{P}(X^2)\underline{P}(Y^2)$ is false.

If, alternatively, we extend in whatever way $\underline{P}$ to $\{X^2, -Y^2, XY\}$ (which by conjugacy is equivalent to assessing arbitrarily $\overline{P}(Y^2)$), we come to an analogous conclusion:

the CS inequality $\underline{P}(XY)^2 \leq \underline{P}(X^2)\overline{P}(Y^2)$ is false too.

(b) Now suppose instead that $y_1 < y_3$ (without changing anything else in Table 1), and consider the upper prevision $\overline{P}'$ defined as $\overline{P}'(Z) = \max\{P_1(Z), P_2(Z)\}$, $Z \in \{-X^2, XY\}$, coherent by Theorem 2.1(a). By conjugacy (5), this is equivalent to consider the couple $\underline{P}'(X^2) = -\overline{P}'(-X^2)$, $\overline{P}'(XY)$. Easy computations show that $\overline{P}'(XY) = \frac{y_3 - y_1}{3} > 0$, $\underline{P}'(X^2) = 0$. Therefore, similarly to case (a), the CS inequalities

$$\overline{P}'(XY)^2 \leq \underline{P}'(X^2)\underline{P}'(Y^2),\ \overline{P}'(XY)^2 \leq \underline{P}'(X^2)\overline{P}'(Y^2) \tag{16}$$

are false too.

Remark 4. We point out that the conclusions of Example 3.1 (*a*) apply irrespective of which is the extension of $\underline{P}$ to either $\{X^2, Y^2, XY\}$ or $\{X^2, -Y^2, XY\}$ (i.e. whether any of such extensions is coherent or not). This suggests that the inequalities

$$\underline{P}(XY)^2 \leq \underline{P}(X^2)\underline{P}(Y^2),\ \underline{P}(XY)^2 \leq \underline{P}(X^2)\overline{P}(Y^2)$$

might be hard to obtain even with different consistency properties for $\underline{P}$.[3] The same conclusions hold for the CS inequalities (16) in Example 3.1 (*b*). Thus, summing up, *the CS inequality (2) does not unconditionally obtain when at least one of μ', μ'' is a lower prevision.*

4. Cauchy-Schwarz inequalities without asking coherence

Although coherence is the most natural and widespread requirement for consistency of a lower/upper prevision, other consistency notions play a role within Imprecise Probability theory.

[3] The alternative conjecture that the reverse inequality $\underline{P}(XY)^2 \geq \underline{P}(X^2)\underline{P}(Y^2)$ might always obtain is not well-grounded either, see for this the later Lemma 5.2.

4.1. The condition of avoiding sure loss (ASL)

The property of avoiding sure loss is investigated at length in Walley [9]. Being considerably weaker than coherence, and therefore departing even more from dF-coherence for precise probabilities and previsions, it has to be expected that ASL lower and upper previsions might not comply with the Cauchy-Schwarz inequalities (3) and (4), without additional novel constraints. Indeed, we have:

Proposition 4.1.

(a) Let $\overline{P}$ avoid sure loss on $\mathbb{D} = \{X^2, Y^2, XY\}$.

$$\text{If } \overline{P}(XY) \leq 0 \text{ then } \overline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2). \tag{17}$$

(b) Let $\underline{P}$ avoid sure loss on $\{-X^2, -Y^2, XY\}$ and $\overline{P}$ be its conjugate.

$$\text{If } \underline{P}(XY) \geq 0 \text{ then } \underline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2). \tag{18}$$

Proof.

(a) From Theorem 2.1 (*b*), there is a dF-coherent prevision P such that $P \leq \overline{P}$ on $\mathbb{D}$, hence $0 \geq \overline{P}(XY) \geq P(XY)$. From this and applying (15) to P, we obtain (17): $\overline{P}(XY)^2 \leq P(XY)^2 \leq P(X^2)P(Y^2) \leq \overline{P}(X^2)\overline{P}(Y^2)$.

(b) Again from Theorem 2.1 (*b*), there is a dF-coherent prevision P such that: $\underline{P}(XY) \leq P(XY)$; $\underline{P}(-X^2) = -\overline{P}(X^2) \leq P(-X^2) = -P(X^2)$, that is $P(X^2) \leq \overline{P}(X^2)$; similarly, $P(Y^2) \leq \overline{P}(Y^2)$. Applying then (15) to P gives (18): $\underline{P}(XY)^2 \leq P(XY)^2 \leq P(X^2)P(Y^2) \leq \overline{P}(X^2)\overline{P}(Y^2)$.

□

It is fairly easy to realize that (17), (18) cannot be improved without additional conditions on $\underline{P}$, $\overline{P}$, i.e., requiring only that they avoid sure loss. The intuition is developed for inequality (17) in the next example.

Example 4.1. Let $\overline{P}$ be an upper prevision that avoids sure loss on $\mathbb{D} = \{X^2, Y^2, XY\}$. Therefore there is a dF-coherent prevision P that is dominated by $\overline{P}$, i.e., $P \leq \overline{P}$ on $\mathbb{D}$. Now consider another upper prevision $\overline{P}^*$, such that $\overline{P}^*(X^2) = \overline{P}(X^2)$, $\overline{P}^*(Y^2) = \overline{P}(Y^2)$, while $\overline{P}^*(XY) > \overline{P}(XY)$. Clearly, $\overline{P}^* \geq P$ on D; hence $\overline{P}^*$ avoids sure loss on $\mathbb{D}$ too, by Theorem 2.1 (*b*), no matter how large $\overline{P}^*(XY)$ is. However, if $\overline{P}^*(XY)$ is large enough, $\overline{P}^*$ will not be coherent (if $\overline{P}^*(XY) > \sup(XY)$ it is not, because of internality (11)), nor will it satisfy the Cauchy-Schwarz inequality (3), when $\overline{P}^*(XY) > \sqrt{\overline{P}(X^2)\overline{P}(Y^2)}$.

As a simple special instance of additional condition that ensures the CS inequality (3), we point out the following lemma. It will be applied in the next Section 4.2.

Lemma 4.1. *Let $\overline{P}$ avoid sure loss on $\mathbb{D} = \{X^2, Y^2, XY\}$. If there exists a dF-coherent prevision P such that $P \leq \overline{P}$ on $\mathbb{D}$ and $P(XY) = \overline{P}(XY)$, then (3) obtains.*

Proof. From the assumptions and (15),

$$\overline{P}(XY)^2 = P(XY)^2 \leq P(X^2)P(Y^2) \leq \overline{P}(X^2)\overline{P}(Y^2).$$

□

4.2. The condition of C-convexity

Centered convex or *C-convex* upper/lower previsions, investigated in [19], are a notable type of imprecise previsions that avoid sure loss (cf. [19, Proposition 3.5 (*e*)]). They are close to coherent previsions in some sense (cf. Theorem 2.1), and are related to convex risk measures in risk measurement [19].

We investigate in this subsection whether they satisfy the Cauchy-Schwarz inequalities (3), (4) under more general conditions than generic previsions avoiding sure loss. We shall see that further sufficient conditions may be proven for them, even though (3), (4) do not always apply, unlike coherence.

We start by considering inequality (3). To prove a sufficient condition for (3), we shall make use of the next lemma. In itself, it establishes, so to say, an arithmetical property of the Cauchy-Schwarz inequality, unrelated to C-convexity. We therefore state it for a generic map μ.

Lemma 4.2. *Define the map $\mu : \{X^2, Y^2, XY\} \to \mathbb{R}$.*

(i) Let μ satisfy the CS inequality

$$\mu(XY)^2 \leq \mu(X^2)\mu(Y^2). \tag{19}$$

Then, given $\alpha > 0$, the map $\mu - \alpha$ satisfies the analogous CS inequality

$$(\mu(XY) - \alpha)^2 \leq (\mu(X^2) - \alpha)(\mu(Y^2) - \alpha) \tag{20}$$

iff

(a) either $\mu(X^2)+\mu(Y^2)-2\mu(XY)\leq 0$

(b) or $\mu(X^2)+\mu(Y^2)-2\mu(XY)>0$ *and* $\alpha\leq \frac{\mu(X^2)\mu(Y^2)-\mu(XY)^2}{\mu(X^2)+\mu(Y^2)-2\mu(XY)}$.

(ii) If μ *is a dF-coherent prevision* P*, the inequality in* (a) *boils down to the equality*

$$P[(X-Y)^2]=0, \tag{21}$$

while the denominator in (b) *simplifies to* $P[(X-Y)^2]$.

Proof.

(i) Easy computations show that inequality (20) is equivalent to the inequality

$$\alpha(\mu(X^2)+\mu(Y^2)-2\mu(XY))\leq \mu(X^2)\mu(Y^2)-\mu(XY)^2$$

whose right term is non-negative by (19). Therefore, this inequality applies iff either (a) or (b) does.

(ii) Let $\mu=P$. By the additivity and homogeneity properties of dF-coherent previsions (12), and internality (11), the denominator in (b) is

$$P(X^2)+P(Y^2)-2P(XY)=P[(X-Y)^2]\geq 0.$$

Therefore (a) may only obtain, when it does, with equality.

□

Lemma 4.2, combined with the generalised envelope theorem (Theorem 2.1(c)), lets us prove the following

Proposition 4.2. *Let* $\overline{P}$ *be a C-convex upper prevision on* $\mathbb{D}=\{X^2,Y^2,XY,0\}$ *and let* $\alpha\geq 0$ *and* $\mathcal{P}$ *be respectively the function and the set of dF-coherent previsions on* $\mathbb{D}$ *defined in Theorem 2.1* (c)*. Let also* $P\in\mathcal{P}$ *be such that* $\overline{P}(XY)=P(XY)-\alpha(P)$.

If

i) either $\alpha(P)=0$

ii) or $0<\alpha(P)\leq \max\{P(X^2),P(Y^2)\}$ *and one of the following conditions applies:*

(a) $P[(X-Y)^2]=0$

(b) $P[(X-Y)^2]>0$ *and* $\alpha(P)\leq \frac{P(X^2)P(Y^2)-P(XY)^2}{P[(X-Y)^2]}$

then the CS inequality $\overline{P}(XY)^2\leq \overline{P}(X^2)\overline{P}(Y^2)$ *obtains.*

Proof.

If $\alpha(P)=0$, then $P\leq\overline{P}$ on $\mathbb{D}$, while $\overline{P}$ avoids sure loss. Hence, the thesis follows from Lemma 4.1.

If $\alpha(P)>0$, use Lemma 4.2 with $\mu=P$ to ensure that $\mu-\alpha(P)=P-\alpha(P)$ obeys the CS inequality (20). Because of this, the first inequality in the next derivation obtains. Besides, condition $\alpha(P)\leq\max\{P(X^2),P(Y^2)\}$, insuring that both factors $P(X^2)-\alpha(P)$ and $P(Y^2)-\alpha(P)$ are non-negative, justifies the second inequality:

$$\overline{P}(XY)^2=(P(XY)-\alpha(P))^2\leq(P(X^2)-\alpha(P))(P(Y^2)-\alpha(P))\leq\overline{P}(X^2)\overline{P}(Y^2).$$

□

Proposition 4.2 exhibits two interesting features. Primarily, it gives a sufficient condition for (3) that is independent of the sign of $\overline{P}(XY)$, unlike Proposition 4.1.

Moreover, it sheds light on how a C-convex prevision may deviate from coherence while still satisfying the CS inequality (3). For this, observe that by Theorem 2.1 (a) a coherent upper prevision may be viewed as a way of grouping dF-coherent previsions, while Theorem 2.1 (c) lets us interpret a C-convex upper prevision $\overline{P}$ as a pooling of *corrected* (downwards) dF-coherent previsions. The correction term $\alpha(P)$ may in general prevent $\overline{P}$ from being also coherent, and $P-\alpha(P)$ from being dF-coherent. If each $P\in\mathcal{P}$ is an expert's opinion, $\alpha(P)$ is an agent's correction of that opinion. Then, Proposition 4.2 requires that the correction $\alpha(P^*)$ (of a prevision P^*, such that $\overline{P}(XY)=P^*(XY)-\alpha(P^*)$), is not "too large": it is upper bounded by both condition $\alpha(P^*)\leq\max\{P^*(X^2),P^*(Y^2)\}$ and the inequality in Proposition 4.2 (b).

Although empirical checks suggest that (3) is often verified with C-convexity, this is not always true, as the next example shows.

Example 4.2. Define the gambles X,Y and the probability measures P_1,P_2 on partition $\mathbb{P}=\{\omega_1,\omega_2,\omega_3\}$ as in Table 2.

We then obtain a C-convex upper prevision on $\mathbb{D}=\{X^2,Y^2,XY,0\}$ by extending P_1 and P_2 (as expectations/dF-coherent previsions) on $\mathbb{D}$, fixing $\alpha(P_1)=0$, $\alpha(P_2)=0.6$ and applying Theorem 2.1 (c) to $\overline{P}=\max\{P_1,P_2-0.6\}$. The results are

$$P_1(X^2)=0.82,\ P_1(Y^2)=0.565,\ P_1(XY)=0.68,\ P_1(0)=0;$$

$$P_2(X^2)-0.6=0.84,\ P_2(Y^2)-0.6=1.36,$$

$$P_2(XY)-0.6=1.08,\ P_2(0)-0.6=-0.6.$$

Table 2
Data for Example 4.2.

$\mathbb{P}$	X	Y	XY	P_1	P_2
ω_1	1.2	1.4	1.68	0	1
ω_2	1	0.8	0.8	0.5	0
ω_3	0.8	0.7	0.56	0.5	0

It ensues that $\overline{P}$ coincides with $P_2 - \alpha(P_2)$ on $\{X^2, Y^2, XY\}$ (while $\overline{P}(0) = P_1(0) = 0$). We cannot apply Proposition 4.1 (a) here, since $\overline{P}(XY) = 1.08 > 0$. Furthermore, neither (i) nor (ii) in Proposition 4.2 obtains, nor does (3). In fact,

$$\alpha(P_2) = 0.6 > \frac{P_2(X^2)P_2(Y^2) - P_2(XY)^2}{P_2(X^2) + P_2(Y^2) - 2P_2(XY)} = 0.$$
$$\overline{P}(XY)^2 = 1.1664 > 1.1424 = \overline{P}(X^2)\overline{P}(Y^2).$$

We conclude that improving the ASL condition with C-convexity is not sufficient to unconditionally guarantee (3).

Turning to the CS inequality (4), a sufficient condition similar to Proposition 4.2 may be proven.

Proposition 4.3. *Let $\overline{P}$ be C-convex on $\mathbb{D} = \{X^2, Y^2, -XY, 0\}$, with $\underline{P}$ its conjugate, and let $\alpha \geq 0$ and $\mathcal{P}$ be respectively the function and the set of dF-coherent previsions on $\mathbb{D}$ defined in Theorem 2.1 (c). Let also $P \in \mathcal{P}$ be such that $\overline{P}(-XY) = P(-XY) - \alpha(P)$. If*

i) either $\alpha(P) = 0$
ii) or $0 < \alpha(P) \leq \max\{P(X^2), P(Y^2)\}$ and one of the following conditions applies:
  (a) $P[(X+Y)^2] = 0$
  (b) $P[(X+Y)^2] > 0$ and $\alpha(P) \leq \frac{P(X^2)P(Y^2) - P(XY)^2}{P[(X+Y)^2]}$,

then

$$\underline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2). \tag{22}$$

Proof. The proof is immediate, after recognising that the statement of Proposition 4.3 is equivalent to that of Proposition 4.2, when replacing X with $-X$ there. In fact, the replacement modifies the domain to $\{(-X)^2 = X^2, Y^2, -XY\}$, while not affecting condition $i)$. As for condition $ii)$, its (a) in Proposition 4.3 follows at once from $P[(-X-Y)^2] = P[(X+Y)^2]$. Condition (b) ensues from

$$\frac{P((-X)^2)P(Y^2) - P(-XY)^2}{P[(-X-Y)^2]} = \frac{P(X^2)P(Y^2) - [-P(XY)]^2}{P[(X+Y)^2]}$$
$$= \frac{P(X^2)P(Y^2) - P(XY)^2}{P[(X+Y)^2]}.$$

Hence, by Proposition 4.2, we get $\overline{P}(-XY)^2 = (-\underline{P}(XY))^2 = \underline{P}(XY)^2 \leq \overline{P}((-X)^2)\overline{P}(Y^2) = \overline{P}(X^2)\overline{P}(Y^2)$. □

The interpretation of this result is similar to that of Proposition 4.2: the correction $\underline{P}(XY)$ of $P(XY)$ must not be 'too large' ($\underline{P}(XY) = -\overline{P}(-XY) = -(P(-XY) - \alpha(P)) = P(XY) + \alpha(P)$). This time however the correction is upwards.

5. Cauchy-Schwarz inequalities and 2-coherence

Comparing the consistency conditions of avoiding sure loss and of 2-coherence, neither implies the other one. In fact, to see that ASL does not imply 2-coherence, take the upper prevision $\overline{P}^*$ from Example 4.1: it still avoids sure loss on $\mathbb{D}$ when $\overline{P}^*(XY) > \sup(XY)$, but cannot be 2-coherent by internality (11). The example below shows that the converse implication may fail too.

Example 5.1. Define the lower probability $\underline{P}$ on $\mathcal{A}$, the set of events logically dependent on partition $\mathbb{P} = \{\omega_1, \omega_2, \omega_3\}$, as follows:

$$\underline{P}(\varnothing) = 0, \underline{P}(\omega_1) = 0.4, \underline{P}(\omega_2) = 0.3, \underline{P}(\omega_3) = 0.4,$$
$$\underline{P}(\omega_1 \vee \omega_2) = \underline{P}(\omega_1 \vee \omega_3) = \underline{P}(\omega_2 \vee \omega_3) = 0.6, \underline{P}(\Omega) = 1.$$

- $\underline{P}$ is 2-coherent on $\mathcal{A}$. In fact, it satisfies the conditions: $\underline{P}(\varnothing) = 0, \underline{P}(\Omega) = 1$; $\underline{P}(A) + \underline{P}(A^c) \leq 1, \forall A \in \mathcal{A}$; and $\forall A, B \in \mathcal{A}$ s.t. $A \Rightarrow B, \underline{P}(A) \leq \underline{P}(B)$, which are sufficient for $\underline{P}$ to be 2-coherent on $\mathcal{A}$ by [24, Proposition 4].
- $\underline{P}$ does not avoid sure loss on $\mathcal{A}$. For this, there should exist a dF-coherent probability $P \geq \underline{P}$ on $\mathcal{A}$, by Theorem 2.1 (b). Thus, any such P should ensure that $P(\omega_1) \geq 0.4, P(\omega_2) \geq 0.3, P(\omega_3) \geq 0.4$, hence $P(\omega_1 \vee \omega_2 \vee \omega_3) = P(\Omega) \geq 1.1$, which contradicts dF-coherence.

Thus, 2-coherence does not imply the weakest of the other consistency notions encountered in this paper: there are 2-coherent previsions that incur sure loss.

In this section, the later Propositions 5.4 and 5.6 identify broad conditions that ensure inequality (3) with 2-coherence. Yet, (3) may not obtain in some cases: Proposition 5.5 hedges them, showing that they must satisfy very peculiar features, and supplies a weak CS inequality that obtains in such instances too.

The path to these results passes through a few propositions. The next is a sufficient condition for the CS inequality (3).

Proposition 5.1. *Let $\overline{P}$ be 2-coherent on $\mathbb{D} = \{X^2, Y^2, XY\}$. Suppose that there exist two previsions P_1, P_2 dF-coherent on $\mathbb{D}$ and such that*

$$P_1(XY) = \overline{P}(XY), \quad P_1(X^2) = \overline{P}(X^2) - \delta_1, \quad P_1(Y^2) = \overline{P}(Y^2) + \varepsilon_1 \tag{23}$$

$$P_2(XY) = \overline{P}(XY), \quad P_2(X^2) = \overline{P}(X^2) + \delta_2, \quad P_2(Y^2) = \overline{P}(Y^2) - \varepsilon_2 \tag{24}$$

with $\delta_1, \varepsilon_2 \geq 0$; $\varepsilon_1, \delta_2 > 0$. Then, the CS inequality (3) obtains if

$$\varepsilon_2 \delta_1 \geq \varepsilon_1 \delta_2 \tag{25}$$

Proof. Every convex combination P_λ of P_1, P_2, $P_\lambda = \lambda P_1 + (1-\lambda)P_2$, $\lambda \in [0,1]$, is a dF-coherent prevision too [9, Section 2.6.4], and moreover $P_\lambda(XY) = \overline{P}(XY)$.

Hence, (3) obtains if P_λ satisfies the inequality

$$P_\lambda(X^2)P_\lambda(Y^2) \leq \overline{P}(X^2)\overline{P}(Y^2). \tag{26}$$

In fact, then $\overline{P}(XY)^2 = P_\lambda(XY)^2 \leq P_\lambda(X^2)P_\lambda(Y^2) \leq \overline{P}(X^2)\overline{P}(Y^2)$, using (15).

Let us now show that such a P_λ exists when (25) holds. Taking account of (23), (24), we easily obtain

$$P_\lambda(X^2) = \overline{P}(X^2) - \lambda\delta_1 + (1-\lambda)\delta_2, \; P_\lambda(Y^2) = \overline{P}(Y^2) + \lambda\varepsilon_1 - (1-\lambda)\varepsilon_2.$$

Thus, inequality (26) applies if $P_\lambda(X^2) \leq \overline{P}(X^2)$ and $P_\lambda(Y^2) \leq \overline{P}(Y^2)$, i.e. if the following system is compatible:

$$\begin{cases} -\lambda\delta_1 + (1-\lambda)\delta_2 \leq 0 \\ \lambda\varepsilon_1 - (1-\lambda)\varepsilon_2 \leq 0 \end{cases}$$

or equivalently if

$$\frac{\delta_2}{\delta_1 + \delta_2} \leq \lambda \leq \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2}$$

There exists at least one such λ iff $\frac{\delta_2}{\delta_1+\delta_2} \leq \frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}$, which is equivalent to (25). □

Remark 5. The sufficient condition (25) in Proposition 5.1 may be interpreted in terms of how $\overline{P}$ may deviate from coherence while guaranteeing (3). For this, parameters ε_1, δ_2 measure how much P_1 and respectively P_2 exceed from values that would include them in the credal set $\mathcal{M}$ of $\overline{P}$, $\mathcal{M} = \{P : P \text{ is dF-coherent on } \mathbb{D}, P \leq \overline{P}\}$, while δ_1, ε_2 represent the 'margins' of P_1, P_2, i.e. the maximum values that may be added to P_1, P_2 without exiting $\mathcal{M}$. Thus, (25) requires that the product of the excess values is not larger than the product of the margin values.

Another key intermediate result is the next characterisation of 2-coherent upper previsions that are not coherent on a three-gamble domain.

Definition 5.1. *Given $D_0 = \{X_1, X_2, X_3\}$, $\overline{P} : D_0 \to \mathbb{R}$, let $\mathcal{M} = \{P : P \text{ dF-coherent on } D_0, P \leq \overline{P}\}$ be the* credal set *of $\overline{P}$. Define further the following* partial *credal sets, $\forall i, j \in \{1,2,3\}, i \neq j$:*

$$\mathcal{M}_{X_i,X_j} = \mathcal{M}_{X_j,X_i} = \{P : P \text{ is dF-coherent on } D_0, P \leq \overline{P} \text{ on } \{X_i, X_j\}\}$$

and, for each of the three partial credal sets, let

$$\mathcal{P}_{X_i}(\mathcal{M}_{X_i,X_j}) = \{P \in \mathcal{M}_{X_i,X_j} : P(X_i) = \overline{P}(X_i)\}$$

$$\mathcal{P}_{X_j}(\mathcal{M}_{X_i,X_j}) = \{P \in \mathcal{M}_{X_i,X_j} : P(X_j) = \overline{P}(X_j)\}.$$

Remark 6. When $\overline{P}$ is 2-coherent on D_0, this means that $\overline{P}$ is coherent on each $\{X_i, X_j\} \subset D_0, \forall i, j \in \{1,2,3\}, i \neq j$ (cf. Definitions 2.1 (*b*) and 2.1 (*c*)). Consequently, the envelope theorem (Theorem 2.1 (a)) applies piecewise, i.e., separately on every $\{X_i, X_j\}$. Hence, $\mathcal{M}_{X_i,X_j}$, $\mathcal{P}_{X_i}(\mathcal{M}_{X_i,X_j})$ and $\mathcal{P}_{X_j}(\mathcal{M}_{X_i,X_j})$ are all not empty.

Proposition 5.2. *Let $\overline{P}$ be 2-coherent on $D_0 = \{X_1, X_2, X_3\}$ and $\mathcal{M}$ be its credal set. Then, $\overline{P}$ is not coherent on D_0 iff there exists a partial credal set $\mathcal{M}_{X_i,X_j}$ $(i, j \in \{1,2,3\}, i \neq j)$ such that*

$$\mathcal{P}_{X_i}(\mathcal{M}_{X_i,X_j}) \cap \mathcal{M} = \emptyset. \tag{27}$$

Proof. If condition (27) applies, then for any P such that $P(X_i) = \overline{P}(X_i)$, $P(X_j) \leq \overline{P}(X_j)$, it holds that $P(X_k) > \overline{P}(X_k)$, $k \in \{1,2,3\} \setminus \{i,j\}$. As a consequence, $\overline{P}$ is not coherent on D_0, by Theorem 2.1(a): there is no dF-coherent prevision $P \in \mathcal{M}$ such that $P(X_i) = \overline{P}(X_i)$.

Conversely, when (27) does not obtain, $\overline{P}$ may be computed as the upper envelope of three dF-coherent previsions in $\mathcal{M}$, which implies coherence by Theorem 2.1(a). □

Remark 7. Since $\mathcal{M}_{X_i,X_j} = \mathcal{M}_{X_j,X_i}$, condition (27) requires that there is a partial credal set $\mathcal{M}_{X_i,X_j}$ such that (27) applies to either $\mathcal{P}_{X_i}(\mathcal{M}_{X_i,X_j})$ or $\mathcal{P}_{X_j}(\mathcal{M}_{X_j,X_i})$.

A simple but useful sufficient condition for the CS inequality (3) is given next.

Proposition 5.3. *Let $\overline{P}$ be 2-coherent on $\mathbb{D} = \{X^2, Y^2, XY\}$ and $\overline{P}(XY) \geq 0$. Then, inequality (3) obtains if there exists an upper prevision $\overline{P}^*$, coherent on $\mathbb{D}$, such that $\overline{P}^*(X^2) \leq \overline{P}(X^2)$, $\overline{P}^*(Y^2) \leq \overline{P}(Y^2)$, $\overline{P}^*(XY) \geq \overline{P}(XY)$.*

Proof. Recalling Proposition 3.1, we obtain

$$\overline{P}(XY)^2 \leq \overline{P}^*(XY)^2 \leq \overline{P}^*(X^2)\overline{P}^*(Y^2) \leq \overline{P}(X^2)\overline{P}(Y^2).$$

□

The previous results let us state that the CS inequality (3) holds in a number of cases with 2-coherence, as shown by Propositions 5.4 and 5.6. We consider only 2-coherent upper previsions that are not coherent, since otherwise (3) is already guaranteed by Proposition 3.1.

Proposition 5.4. *Let $\overline{P}$ be 2-coherent, but not coherent, on $\mathbb{D} = \{X^2, Y^2, XY\}$. Then, inequality (3) holds if $\overline{P}$ avoids sure loss on $\mathbb{D}$ and one of the following alternatives obtains:*

(a) $\overline{P}(XY) \leq 0$;
(b) $\overline{P}(XY) > 0$ *and Proposition 5.3 holds.*

Proof. When (a) or (b) obtain, the thesis follows, respectively, from Proposition 4.1 (a) or Proposition 5.3. □

Proposition 5.4 lets us identify a situation when $\overline{P}$ is 2-coherent and avoids sure loss but the CS inequality (3) may fail to be true. We describe this case in Proposition 5.5, from the perspective of Proposition 5.2 in its part (a), considering its practical implications in parts (b), (c). The next lemma will be useful in its proof.

Lemma 5.1. *Let $\tilde{P} : \{X, Y\} \to \mathbb{R}$ be both a coherent upper prevision and a lower prevision avoiding sure loss. Then $\tilde{P}$ is a dF-coherent prevision.*

Proof. By Definition 2.1 (a), it suffices to show that, $\forall s_1, s_2 \in \mathbb{R}, \sup G_2 \geq 0$, with

$$G_2 = s_1(X - \tilde{P}(X)) + s_2(Y - \tilde{P}(Y)).$$

When $s_1 \geq 0, s_2 \geq 0$, we get $\sup G_2 \geq 0$ viewing $\tilde{P}$ as a lower prevision that avoids sure loss and applying Definition 2.1 (d).

In the other instances (i.e. when at least one between s_1, s_2 is negative), rewriting G_2 as $G_2 = -s_1(\tilde{P}(X) - X) - s_2(\tilde{P}(Y) - Y)$, we conclude again that $\sup G_2 \geq 0$ viewing $\tilde{P}$ as a coherent upper prevision and applying Definition 2.1 (a). □

Proposition 5.5. *Let the upper prevision $\overline{P}$ be 2-coherent, but not coherent, on $\mathbb{D} = \{X^2, Y^2, XY\}$, with $\overline{P}(XY) > 0$, while the hypothesis of Proposition 5.3 is not satisfied. Then,*

(a)

$$\mathcal{P}_{XY}(\mathcal{M}_{X^2,XY}) \cap \mathcal{M} = \emptyset, \ \mathcal{P}_{XY}(\mathcal{M}_{Y^2,XY}) \cap \mathcal{M} = \emptyset \tag{28}$$

(b) Moreover, $\overline{P}$ is coherent as a lower prevision on $\mathbb{D}$, and even dF-coherent on $\{X^2, Y^2\}$, on $\{X^2, XY\}$, and on $\{Y^2, XY\}$ separately.
(c) $\overline{P}$ satisfies the weak CS inequality

$$\begin{aligned} \overline{P}(XY)^2 &\leq \overline{P}(X^2)\overline{P}(Y^2) \min\left\{\frac{\sup(X^2)}{\overline{P}(X^2)}, \frac{\sup(Y^2)}{\overline{P}(Y^2)}\right\} \\ &= \min\left\{\overline{P}(Y^2)\sup(X^2), \overline{P}(X^2)\sup(Y^2)\right\} \end{aligned} \tag{29}$$

Proof.

(a) Since $\overline{P}$ is not coherent, by Proposition 5.2 Eq. (27) must obtain, for some $\mathcal{P}_{X_i}(\mathcal{M}_{X_i,X_j})$, $X_i, X_j \in \{X^2, Y^2, XY\}$.
On the other hand, since the hypothesis of Proposition 5.3 is not satisfied, it holds that

$$\forall P \in \mathcal{P}_{X^2}(\mathcal{M}_{X^2,Y^2}), P(XY) < \overline{P}(XY) \tag{30}$$

$$\forall P \in \mathcal{P}_{Y^2}(\mathcal{M}_{X^2,Y^2}), P(XY) < \overline{P}(XY) \tag{31}$$

Now, if the first equality in (28) is false (the argument for the second is identical), there exists $P^* \in \mathcal{P}_{XY}(\mathcal{M}_{X^2,XY}) \cap \mathcal{M}$. Then, $P^*(XY) = \overline{P}(XY)$, $P^*(X^2) \leq \overline{P}(X^2)$, $P^*(Y^2) \leq \overline{P}(Y^2)$. It follows that $\overline{P}$ is the upper envelope of P^* and of two dF-coherent previsions verifying, respectively, (30) and (31). Hence $\overline{P}$ is coherent on $\mathbb{D}$ by Theorem 2.1 (a), against the assumptions.

(b) Condition (28) ensures that

- any $P_1 \in \mathcal{P}_{XY}(\mathcal{M}_{X^2,XY})$ is such that

$$P_1(XY) = \overline{P}(XY), \ P_1(X^2) = \overline{P}(X^2) - \delta_1, \ P_1(Y^2) = \overline{P}(Y^2) + \varepsilon_1 \tag{32}$$

- any $P_2 \in \mathcal{P}_{XY}(\mathcal{M}_{Y^2,XY})$ is such that

$$P_2(XY) = \overline{P}(XY), \ P_2(X^2) = \overline{P}(X^2) + \delta_2, \ P_2(Y^2) = \overline{P}(Y^2) - \varepsilon_2 \tag{33}$$

with $\delta_1, \varepsilon_2 \geq 0, \varepsilon_1, \delta_2 > 0$. Then, for every P_1 of the type (32) and every P_2 of the type (33), it holds that

$$\varepsilon_2\delta_1 < \varepsilon_1\delta_2. \tag{34}$$

In fact, otherwise there are $P_1 \in \mathcal{P}_{XY}(\mathcal{M}_{X^2,XY})$ and $P_2 \in \mathcal{P}_{XY}(\mathcal{M}_{Y^2,XY})$ such that $\varepsilon_2\delta_1 \geq \varepsilon_1\delta_2$. But then Proposition 5.1 applies and the dF-coherent prevision P_λ in its proof belongs to both $\mathcal{P}_{XY}(\mathcal{M}_{X^2,XY}) \cap \mathcal{M}$ and $\mathcal{P}_{XY}(\mathcal{M}_{Y^2,XY}) \cap \mathcal{M}$, contrary to (28). Still recalling the proof of Proposition 5.1, we know that every $P^*_\lambda = \lambda P_1 + (1-\lambda)P_2$, $\lambda \in [0,1]$, is a dF-coherent prevision on $\mathbb{D}$, and $P^*_\lambda(XY) = \overline{P}(XY)$,

$$P^*_\lambda(X^2) = \overline{P}(X^2) - \lambda\delta_1 + (1-\lambda)\delta_2 \tag{35}$$

$$P^*_\lambda(Y^2) = \overline{P}(Y^2) + \lambda\varepsilon_1 - (1-\lambda)\varepsilon_2. \tag{36}$$

Now select, among all previsions P^*_λ, $P^{(1)}$ putting $\lambda = \frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}$ and $P^{(2)}$, obtained putting $\lambda = \frac{\delta_2}{\delta_1+\delta_2}$.

Recalling (35), (36) and, for the inequalities, (34), we easily obtain:

$$P^{(1)}(XY) = \overline{P}(XY), \quad P^{(1)}(Y^2) = \overline{P}(Y^2),$$

$$P^{(1)}(X^2) = \overline{P}(X^2) + \frac{\varepsilon_1\delta_2 - \varepsilon_2\delta_1}{\varepsilon_1+\varepsilon_2} > \overline{P}(X^2);$$

$$P^{(2)}(XY) = \overline{P}(XY), \quad P^{(2)}(X^2) = \overline{P}(X^2),$$

$$P^{(2)}(Y^2) = \overline{P}(Y^2) + \frac{\varepsilon_1\delta_2 - \varepsilon_2\delta_1}{\delta_1+\delta_2} > \overline{P}(Y^2).$$

Thus, $\overline{P} = \min\{P^{(1)}, P^{(2)}\}$ is a coherent lower probability by Theorem 2.1 (*a*).

There remains to prove that $\overline{P}$ is also dF-coherent on any 2-element subset of $\{X^2, Y^2, XY\}$.

For this, note that since $\overline{P}$ is a coherent lower prevision it avoids sure loss. Furthermore, being a 2-coherent upper prevision, it is coherent on every 2-element subset. Hence, to prove dF-coherence, apply Lemma 5.1 to each of these subsets.

(c) Take any P_1 satisfying Eq. (32). Using also Eq. (15) and internality (11), it holds that

$$\overline{P}(XY)^2 = P_1(XY)^2 \leq P_1(X^2)P_1(Y^2) \leq \overline{P}(X^2)\frac{P_1(Y^2)}{\overline{P}(Y^2)}\overline{P}(Y^2)$$

$$\leq \overline{P}(X^2)\overline{P}(Y^2)\frac{\sup(Y^2)}{\overline{P}(Y^2)}.$$

[4] Taking any P_2 satisfying (33), we obtain analogously the inequality

$$\overline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2)\frac{\sup(X^2)}{\overline{P}(X^2)},$$

and then (29).

□

In its part (*b*), Proposition 5.5 tells us that when we fail to apply (3) to $\overline{P}$ because it avoids sure loss but Proposition 5.4 does not obtain, then $\overline{P}$ satisfies very restrictive conditions. The doubt remains whether there exist 2-coherent upper previsions that may comply with such conditions, and if so whether (3) may fail. The following example gives an affirmative answer to both issues.

Example 5.2. Define the gambles X, Y on partition $\mathbb{P} = \{\omega_1, \omega_2, \omega_3\}$ as in the second and third column of Table 3. Next, define the upper prevision $\overline{P}$ on $\mathbb{D} = \{X^2, Y^2, XY\}$ as follows:

$$\overline{P}(X^2) = 5.4,\ \overline{P}(Y^2) = 55,\ \overline{P}(XY) = 18.$$

- The CS inequality (3) does not hold:

$$\overline{P}(XY)^2 = 324 > 297 = \overline{P}(X^2)\overline{P}(Y^2).$$

Therefore, $\overline{P}$ is not coherent on $\mathbb{D}$.

- $\overline{P}$ is 2-coherent and avoids sure loss on $\mathbb{D}$, is coherent as a lower prevision on $\mathbb{D}$, and is dF-coherent on each 2-gamble subset of $\mathbb{D}$.

One way to see this[5] begins defining the probabilities P, $P^{(1)}$, $P^{(2)}$ on $\mathbb{P}$ as in the three rightmost columns in Table 3.

The expectations of X^2, Y^2, XY computed from P, $P^{(1)}$, $P^{(2)}$ are dF-coherent previsions on $\{X^2, Y^2, XY\}$, still termed P, $P^{(1)}$, $P^{(2)}$ respectively:

[4] Note that the assumption $\overline{P}(XY) > 0$ implies that $\overline{P}(X^2) > 0$, $\overline{P}(Y^2) > 0$ by Lemma 3.2.

[5] The upper prevision $\overline{P}$ on $\mathbb{D}$ in this example has been derived as the conjugate of $\underline{E}_{2C}$, the 2-coherent natural extension (cf. Remark 2) on $\{-X^2, -Y^2, -XY\}$ of the lower probability $\underline{P}$ of Example 5.1. $\underline{E}_{2C}$ has been computed by means of the formulae in [22, Proposition 10]. Hence, $\overline{P}$ is 2-coherent on $\mathbb{D}$. Other properties of $\overline{P}$ may be obtained exploiting Definition 2.1.

Table 3
Values of X, Y and probabilities $P, P^{(1)}, P^{(2)}$ in Example 5.2.

$\mathbb{P}$	X	Y	P	$P^{(1)}$	$P^{(2)}$
ω_1	-3	-4	$\frac{5}{28}$	0	$\frac{5}{13}$
ω_2	-3	-10	$\frac{29}{70}$	$\frac{3}{5}$	$\frac{29}{65}$
ω_3	0	5	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{11}{65}$

$$P(X^2) = 5.4 = \overline{P}(X^2) \qquad P^{(1)}(X^2) = \overline{P}(X^2) \qquad P^{(2)}(X^2) = \frac{486}{65} > \overline{P}(X^2)$$
$$P(Y^2) = 55 = \overline{P}(Y^2) \qquad P^{(1)}(Y^2) = 60 > \overline{P}(Y^2) \qquad P^{(2)}(Y^2) = \overline{P}(Y^2)$$
$$P(XY) = \frac{207}{14} < \overline{P}(XY) = 18 \qquad P^{(1)}(XY) = \overline{P}(XY) \qquad P^{(2)}(XY) = \overline{P}(XY)$$

Then, $\overline{P}$ is dF-coherent on $\{X^2, Y^2\}$, on $\{X^2, XY\}$, on $\{Y^2, XY\}$ separately, being the restriction on these sets of, respectively, P, $P^{(1)}$, $P^{(2)}$. It is coherent on each of these three subsets of $\mathbb{D}$ (dF-coherence implies coherence), and therefore 2-coherent on $\mathbb{D}$. $\overline{P}$ avoids sure loss, by Theorem 2.1 (b): $P \leq \overline{P}$ on $\mathbb{D}$. Finally, $\overline{P}$ is coherent as a lower prevision on $\mathbb{D}$ by Theorem 2.1 (a): $\overline{P} = \min\{P^{(1)}, P^{(2)}\}$. Note that $P^{(1)}, P^{(2)}$ have precisely the role of previsions $P^{(1)}, P^{(2)}$ in the proof of Proposition 5.5 (b).

- *As for the weak CS inequality (29), it gives:*

$$\overline{P}(XY)^2 = 324 < 297 \cdot \min\left\{\frac{9}{5.4}, \frac{100}{55}\right\} = 495.$$

Turning to 2-coherent upper previsions that incur sure loss, we have the proposition that follows.

Proposition 5.6. *Let $\overline{P}$ be 2-coherent on $\mathbb{D} = \{X^2, Y^2, XY\}$. Then, inequality (3) holds if $\overline{P}$ incurs sure loss and one of the following conditions (a) or (b) holds:*

(a) $\overline{P}(XY) \geq 0$.
(b) There exists a 2-coherent extension of $\overline{P}$ on $\mathbb{D} \cup \{|XY|\}$, still termed $\overline{P}$, such that
  (b1) either $\overline{P}$ incurs sure loss on $\mathbb{D}^+ = \{X^2, Y^2, |XY|\}$
  (b2) or $\overline{P}$ avoids sure loss on $\mathbb{D}^+$ and Proposition 5.3 obtains (with XY, $\mathbb{D}$ replaced by $|XY|$, $\mathbb{D}^+$ there).

Proof.

(a) Since 2-coherence is equivalent to coherence on a two-gamble set, $\overline{P}$ is coherent on $\{X^2, Y^2\}$. Let then $\overline{P}^*$ be a coherent extension of $\overline{P}$ from $\{X^2, Y^2\}$ to $\mathbb{D}$. Necessarily, $\overline{P}^*(XY) > \overline{P}(XY)$. In fact, since $\overline{P}^*$ avoids sure loss on $\mathbb{D}$, if $\overline{P}^*(XY) \leq \overline{P}(XY)$ then there exists by Theorem 2.1 (b) a dF-coherent prevision P such that $P \leq \overline{P}^* \leq \overline{P}$ on $\mathbb{D}$, a contradiction. Then, $\overline{P}^*(XY) > \overline{P}(XY)$ and Proposition 5.3 applies, giving (3).

(b) The proof of ($b1$) is firstly identical to that of (a). In fact, replacing XY and $\mathbb{D}$ with $|XY|$ and $\mathbb{D}^+$ there, we apply again Proposition 5.3 to obtain the inequality $\overline{P}(|XY|)^2 \leq \overline{P}(X^2)\overline{P}(Y^2)$. It follows from this and Eq. (13) that

$$\overline{P}(XY)^2 = |\overline{P}(XY)|^2 \leq \overline{P}(|XY|)^2 \leq \overline{P}(X^2)\overline{P}(Y^2).$$

As for ($b2$), it is a consequence of Proposition 5.3.

□

We observe that the case when the CS inequality (3) might not apply when $\overline{P}$ (is 2-coherent and) incurs sure loss is brought back by Proposition 5.6 to $\overline{P}$ satisfying the hypothesis of Proposition 5.5 (with XY, $\mathbb{D}$ replaced by $|XY|$, $\mathbb{D}^+$). In this sense, it suffices to explore the case of $\overline{P}$ avoiding sure loss.

Assuming this, it is at present not clear whether the conditions of Proposition 5.5 (b) are also sufficient to prevent that (3) obtains, with a 2-coherent upper prevision. Certainly, they identify a rather extreme situation, easy to detect: it suffices to check whether $\overline{P}$ is coherent as a lower prevision on $\{X^2, Y^2, XY\}$. Some numerical computations empirically suggest that the incidence of the case is very limited. At any rate, in this occurrence inequality (29) applies, even though it is generally loose. It performs well when at least one between $\overline{P}(X^2)$ and $\overline{P}(Y^2)$ tends to be vague, i.e. close to $\sup(X^2)$ or $\sup(Y^2)$ respectively.

It is possible to derive a further sufficient condition for applying inequality (3). It relies on the following result, which is a special case of [7, Theorem 3.2].

Theorem 5.1 (Jensen's inequality). *Let Z be a gamble, I an interval containing the image set of Z, $\psi : I \to \mathbb{R}$ a concave function, $\overline{P}$ an upper prevision 2-coherent on $\{Z, \psi(Z)\}$. If $\psi'(\overline{P}(Z)) \geq 0$, then*

$$\overline{P}(\psi(Z)) \leq \psi(\overline{P}(Z)). \tag{37}$$

The sufficient condition is given in the following proposition:

Proposition 5.7. *Let $\overline{P}$ be 2-coherent on $\{X^2, Y^2, XY, X^2Y^2\}$ and such that*

$$0 < \overline{P}(X^2Y^2) \leq \overline{P}(X^2)\overline{P}(Y^2). \tag{38}$$

Then inequality (3) obtains.

Proof. We may suppose that $\overline{P}(XY) \neq 0$ (otherwise (3) obtains trivially). This implies by Lemma 3.1 that $\overline{P}(|XY|) > 0$, whatever is the 2-coherent extension of $\overline{P}$ to $|XY|$. Next, define $Z = X^2Y^2$. The function $\psi(x) = \sqrt{x}$ is clearly concave, and $\psi'(x) > 0, \forall x > 0$, in particular when $x = \overline{P}(X^2Y^2)$. Applying Theorem 5.1 and then (38), it ensues that

$$\overline{P}(|XY|) = \overline{P}(\sqrt{X^2Y^2}) = \overline{P}(\psi(Z)) \leq \psi(\overline{P}(Z)) = \sqrt{\overline{P}(X^2Y^2)} \leq \sqrt{\overline{P}(X^2)\overline{P}(Y^2)}.$$

From this and (13) we get $\overline{P}(XY)^2 \leq \overline{P}(|XY|)^2 \leq \overline{P}(X^2)\overline{P}(Y^2)$. □

As for the lower prevision version (4) of the CS inequality, it is strictly related to inequality (3), as appears from the next result.

Proposition 5.8. *Let $\overline{P}$ be 2-coherent on $\{X^2, Y^2, -XY\}$ and $\underline{P}$ be its conjugate. Then it holds that*

$$\underline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2) \tag{39}$$

whenever the inequality $\overline{P}(-XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2)$ obtains.

Proof. Apply Proposition 3.2. □

5.1. Two special cases

We discuss here two special situations where the CS inequalities exhibit some peculiarities.

5.1.1. Assessing a capacity

In the first special case a couple of conjugate *capacities* $\underline{P}, \overline{P}$ (i.e. $\underline{P}$ and $\overline{P}$ are monotone and normalised) with $\underline{P} \leq \overline{P}$, is defined on $\mathcal{A} = \mathcal{A}(\mathbb{P})$, the powerset of a partition $\mathbb{P}$ of the sure event Ω. Then, it is well known that $\underline{P}$, $\overline{P}$ are 2-coherent [24]. In the CS inequalities, X, Y are now indicators of events, say I_A, I_B, or also events A, B when referring to probabilities instead of previsions. Thus, since $I_A = I_A^2, \forall A \in \mathcal{A}$, I_A^2, I_B^2 are replaced by I_A, I_B in any CS inequality. In this framework, (3), (4) always obtain, as well as other inequalities of the type (2) not holding in general (cf. Example 3.1).

Lemma 5.2. *Let $\underline{P}$ be a capacity on $\mathcal{A}$. Then, $\forall A, B \in \mathcal{A}$, it holds that*

$$\underline{P}(A \wedge B)^2 \leq \underline{P}(A)\underline{P}(B) \leq \underline{P}(A)\overline{P}(B) \leq \overline{P}(A)\overline{P}(B)$$

and

$$\overline{P}(A \wedge B)^2 \leq \overline{P}(A)\overline{P}(B).$$

Proof. Recalling that $I_A \cdot I_B = I_{A \wedge B} \leq \min\{I_A, I_B\}$, and applying monotonicity (8) and internality (11) of $\underline{P}$, we have that $\underline{P}(I_A \cdot I_B)^2 = \underline{P}(I_{A\wedge B})\underline{P}(I_{A\wedge B}) \leq \underline{P}(I_A)\underline{P}(I_B) \leq \underline{P}(I_A)\overline{P}(I_B) \leq \overline{P}(I_A)\overline{P}(I_B)$. Given that $\overline{P}(I_E) = \overline{P}(E), \underline{P}(I_E) = \underline{P}(E) \forall E \in \mathcal{A}$, we obtain the chain of inequalities in the thesis.

The derivation of the last inequality is analogous. □

5.1.2. Choquet integral extensions

Let us now suppose that a lower probability $\underline{P}$ is assessed on $\mathcal{A}$, and define $\mathcal{L}(\mathbb{P})$, the set of all gambles $X : \mathbb{P} \to \mathbb{R}$. One way to extend $\underline{P}$, as a lower prevision, to any gamble $X \in \mathcal{L}(\mathbb{P})$ is the *Choquet integral extension*, which obtains the extension $\underline{P}_C(X)$ as a Choquet integral (see e.g. [21, Appendix C] for basic properties of this integral):

$$\underline{P}_C(X) = (C) \int X d\underline{P}.$$

In the special case that the lower probability $\underline{P}$ is coherent and 2-*monotone* (i.e. $\underline{P}(A \vee B) + \underline{P}(A \wedge B) \geq \underline{P}(A) + \underline{P}(B), \forall A, B \in \mathcal{A}$), it is well-known [21, Section 6.2.2] that $\underline{P}_C$ is a coherent lower prevision on any $\mathcal{K} \subseteq \mathcal{L}(\mathbb{P})$. More precisely, $\underline{P}_C$ is the *natural extension* of $\underline{P}$, that is $\underline{P}_C = \underline{E}$ (cf. Remark 2). Coherence of $\underline{P}_C$ implies that the CS inequality (3) has a Choquet integral representation in this case: using conjugacy we easily obtain:

$$\begin{aligned} \overline{P}_C(XY)^2 &= \left((C)\int -XY d\underline{P}\right)^2 \leq \left((C)\int -X^2 d\underline{P}\right)\left((C)\int -Y^2 d\underline{P}\right) \\ &= \overline{P}_C(X^2)\overline{P}_C(Y^2). \end{aligned} \tag{40}$$

When $\underline{P}$ is coherent, but not 2-monotone, on $\mathcal{A}$, the Choquet integral extension produces an incoherent extension on $\mathcal{L}(\mathbb{P})$ [17,21]. Yet, in this case, or more generally when $\underline{P}$ is 2-*coherent* on $\mathcal{A}$, it has been proven in [22, Section 5.4] that $\underline{P}_C$ is again 2-coherent on $\mathcal{L}(\mathbb{P})$. Moreover, it does generally not coincide with the 2-coherent natural extension $\underline{E}_{2C}$ of $\underline{P}$ (cf. Remark 2), so that $\underline{P}_C \geq \underline{E}_{2C}$ on

$\mathcal{L}(\mathbb{P})$. The Choquet integral extension satisfies the CS inequality (40) whenever this is guaranteed by the sufficient conditions for the CS inequality (3) with 2-coherence.

By making use of the Choquet integral, the results in this subsection fall within the topic of seeking generalisations of the classical CS inequality by means of non-additive integrals. To the best of our knowledge, our approach and results differ from pre-existing ones in this area. For instance, Caballero and Sadarangani [4] is concerned with the Sugeno integral, replaces the product between $E(X^2)$ and $E(Y^2)$ with the 'sup' and 'inf' operators, X and Y are non-negative and either comonotonic or take values in $[0,1]$. [1] drops the comonotonicity hypothesis and extends the type of integrals involved, while still keeping non-negativity of X, Y, together with some additional assumptions.

6. Cauchy-Schwarz inequalities and (imprecise) correlation

Let us now consider again a generic 2-coherent upper prevision $\overline{P}$ and its conjugate $\underline{P}$. An immediate follow-up of the classical CS inequality (15) in Probability Theory is the definition of the correlation index $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ as a normalised and adimensional measure of the correlation between X and Y. Something of the kind may be introduced in an imprecise frame through inequalities (3), (4). Suppose that a centrality index c_X (c_Y) is available for X (for Y): c_X could be for instance $P(X)$, if known or estimated, or $\frac{\underline{P}(X)+\overline{P}(X)}{2}$. Consider the evaluations

$$c^U = \overline{P}[(X - c_X)(Y - c_Y)],\ c_L = \underline{P}[(X - c_X)(Y - c_Y)].$$

Knowledge of c^U, c_L may be informative about the dependence between X and Y in the following cases (a) and (b):

(a) $\overline{P}[(X - c_X)(Y - c_Y)] < 0$.
Then, also $\underline{P}[(X - c_X)(Y - c_Y)] < 0$, suggesting negative *near-correlation* between X and Y. Here we use the term 'near-correlation' to stress that, while not computing a classical covariance, the conclusions are analogous: values of X larger than c_X tend to happen with values of Y smaller than c_Y and vice versa.

(b) $\underline{P}[(X - c_X)(Y - c_Y)] > 0$.
This supports, similarly to (a), a positive near-correlation between X and Y.

It has to be observed that cases (a) and (b) are not exhaustive. In fact, they correspond to condition $c_L \cdot c^U > 0$. Alternatively, we have two cases:

(c1) $c_L \cdot c^U = 0$. When $c_L = c^U = 0$, we may assume that X and Y are uncorrelated. If $c_L = 0$ and $c^U > 0$, uncertainty remains between uncorrelation and positive near-correlation of X and Y. Similarly, condition $c_L < 0$ and $c^U = 0$ only rules out positive near-correlation of X and Y.

(c2) $c_L \cdot c^U < 0$. In this occurence, uncertainty on the direction of the dependence between X and Y remains largely unsolved if nothing else is evaluated, since then $c^U > 0$ and $c_L < 0$.

In the sequel, we focus on cases (a) or (b), since they ensure clear-cut conclusions. That is, we assume that

$$c_L \cdot c^U > 0. \tag{41}$$

At this point, let us make a step forward towards getting imprecise correlation coefficients. Here the CS inequalities (3), (4) play an essential role. In fact, let us suppose that (3), (4) hold (now with X, Y replaced by $X - c_X$, $Y - c_Y$, respectively). Then we have that $\overline{P}[(X - c_X)^2] \cdot \overline{P}[(Y - c_Y)^2] > 0$: this ensues from (41), ensuring that $c^U \neq 0$, and from (3), giving $0 < (c^U)^2 \leq \overline{P}[(X - c_X)^2] \cdot \overline{P}[(Y - c_Y)^2]$.

Then, the following quantities are well-defined:

$$\rho^U = \frac{c^U}{\sqrt{\overline{P}[(X - c_X)^2]\overline{P}[(Y - c_Y)^2]}},\ \rho_L = \frac{c_L}{\sqrt{\overline{P}[(X - c_X)^2]\overline{P}[(Y - c_Y)^2]}}.$$

Unlike c^U, c_L, but similarly to the classical correlation coefficient, ρ^U, ρ_L are adimensional and (because of (3), (4)) normalised.

Concerning the role of ρ^U, ρ_L, and of c^U, c_L, a distinction arises between evaluations that are coherent or (only) 2-coherent.

When $\overline{P}$, $\underline{P}$ are *coherent*, (3) and (4) always hold, and ρ^U, ρ_L have the meaning of imprecise correlation coefficients. As for c^U, c_L, they exhibit some similarity with, respectively, the upper and lower covariance $\overline{C}$, $\underline{C}$ of X, Y, defined following [25] as[6]

$$\overline{C}(X,Y) = \max_{P \in \mathcal{M}} \{P[(X - P(X))(Y - P(Y))]\},$$

$$\underline{C}(X,Y) = \min_{P \in \mathcal{M}} \{P[(X - P(X))(Y - P(Y))]\}.$$

Take for instance c^U. Under coherence, we have that

$$c^U = \max_{P \in \mathcal{M}} \{P[(X - c_X)(Y - c_Y)]\}.$$

[6] When $X = Y$, $\overline{C}$ and $\underline{C}$ boil down to respectively the upper and lower variance introduced in [9, Appendix G].

Then, c^U may be an upper covariance only if, $\forall P \in \mathcal{M}$, $P(X) = c_X$, $P(Y) = c_Y$ in the expression of $\overline{C}(X, Y)$. In other words, when employing an upper covariance, an agent is typically uncertain about what could work as c_X, c_Y, unlike the agent employing c^U.

When $\overline{P}$, $\underline{P}$ are 2-*coherent* (but not coherent) the just noted relationship with upper and lower covariances breaks down: the reported definition of $\overline{C}$, $\underline{C}$ from [25] requires coherence (or else the credal set $\mathcal{M}$ can be empty or too large). However, it is still useful to employ ρ^U, ρ_L whenever the CS inequalities (3), (4) hold. Then, these coefficients preserve their meaning of upper and lower correlation coefficients.

7. Conclusions

In this paper we investigated how the CS inequality may be extended to imprecise uncertainty evaluations represented by lower and upper previsions. This analysis falls within the more general problem of extending classical probability inequalities within the theory of imprecise probabilities. A basic feature of this topic is that more imprecise inequalities may correspond to a single probability inequality. Moreover, the number and expression of the corresponding inequalities may vary with the degree of consistency of $\underline{P}$, $\overline{P}$. The situation is relatively simple with Markov's inequality: we obtain two imprecise Markov's inequalities [7, Proposition 5.1] and two couples of Reverse Markov's inequalities [13, Proposition 2], both requiring only 2-coherence. Other inequalities allow more structured (or stratified by consistency) generalisations. See for this the various kinds of Jensen's inequalities in [7, Section 3] and the even more varied extensions of Bhatia-Davis inequality in Pelessoni and Vicig [14].

In this picture, the CS inequality returns instead a relatively simple frame. In fact, while Eq. (2) suggests a potentially high number of extensions, Example 3.1 and Remark 4 demonstrate that most of them do not apply, generally. They do with special additional conditions, such as the case of capacities discussed in Section 5.1.1. In general instead, the developments in this paper show that we are left with two CS inequalities, (3) and (4). But basically we can take (3), $\overline{P}(XY)^2 \leq \overline{P}(X^2)\overline{P}(Y^2)$, as the fundamental inequality. In fact, by Proposition 3.2, (4) can be viewed as a variant of (3), applied to $\{-XY, (-X)^2, Y^2\}$ instead of $\{XY, X^2, Y^2\}$.

Regarding the role of consistency properties of the upper/lower previsions within CS inequalities, we have proven that with coherent imprecise previsions inequalities (3) and (4) obtain in all cases. The condition of avoiding sure loss and C-convexity do not guarantee (3) or (4) without further restrictions. This is true with 2-coherence too, but the instances of non-validity of (3) are circumscribed by Propositions 5.5 and 5.6. It appears that they involve, at most, rather specific situations. Thus, when $\overline{P}$ is (2-coherent and) avoids sure loss on $\mathbb{D} = \{XY, X^2, Y^2\}$, we know that (3) obtains if $\overline{P}$ is not coherent as a lower prevision on $\mathbb{D}$. When $\overline{P}$ incurs sure loss, the same argument applies to $\mathbb{D}^+ = \{|XY|, X^2, Y^2\}$. Moreover, Proposition 5.7 gives a different, but relatively common, sufficient condition for (3). Section 5.1.2 applies the Choquet integral extension to (3), bridging our approach with that of extending the classical CS inequality via non-additive integrals. In general, the contact points between these approaches are relatively few, the initial assumptions being different.

Yet, the occurrence of (3) with 2-coherence is not fully characterised, so far. Another avenue for future investigations concerns applications of the imprecise CS inequalities. We outlined a direct follow-up of theirs in Section 6, regarding evaluations corresponding to correlations in classical probability. Other ones might be of interest, for instance exploring some imprecise version of Cramér-Rao lower bound. At a more general level, much work remains to do in exploring imprecise extensions of other probability inequalities.

CRediT authorship contribution statement

Paolo Vicig: Writing – original draft, Investigation, Formal analysis, Conceptualization; **Renato Pelessoni:** Investigation, Formal analysis, Conceptualization.

Data availability

No data was used for the research described in the article.

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.

Acknowledgements

The authors are affiliated with the GNAMPA-INdAM research group. We wish to thank the reviewers for their helpful comments. At an earlier stage our work benefited from fruitful discussions at the ISIPTA 2025 Symposium.

References

[1] H. Agahi, R. Mesiar, On Cauchy-Schwarz's inequality for Choquet-like integrals without the comonotonicity condition, Soft comput. 19 (6) (2015) 1627–1634. https://doi.org/10.1007/s00500-014-1578-0

[2] H. Agahi, R. Mesiar, Y. Ouyang, Chebyshev type inequalities for pseudo-integrals, Nonlinear Anal. Theory Methods Appl. 72 (6) (2010) 2737–2743. https://doi.org//10.1016/j.na.2009.11.017

[3] H. Agahi, A. Mohammadpour, R. Mesiar, Generalizations of the Chebyshev-type inequality for Choquet-like expectation, Inf. Sci. 236 (2013) 168–173. https://doi.org/10.1016/j.ins.2013.02.019
[4] J. Caballero, K. Sadarangani, A Cauchy-Schwarz type inequality for fuzzy integrals, Nonlinear Analysis Theory Methods Appl. 73 (10) (2010) 3329–3335. https://doi.org/10.1016/j.na.2010.07.013
[5] A. Flores-Franuliç, H. Román-Flores, Y. Chalco-Cano, Markov type inequalities for fuzzy integrals, Appl. Math. Comput. 207 (1) (2009) 242–247. https://doi.org/10.1016/j.amc.2008.10.029
[6] H. Román-Flores, A. Flores-Franuliç, 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
[7] R. Pelessoni, P. Vicig, Jensen's and Cantelli's inequalities with imprecise previsions, Fuzzy Sets Syst. 458 (2023) 50–68. https://doi.org/10.1016/j.fss.2022.06.021
[8] B. de Finetti, Theory of Probability: A Critical Introductory Treatment, John Wiley, New York, 1974.
[9] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London; New York, 1991.
[10] F.G. Cozman, Concentration inequalities and laws of large numbers under epistemic and regular irrelevance, Int. J. Approximate Reasoning 51 (9) (2010) 1069-1084. https://doi.org/10.1016/j.ijar.2010.08.009
[11] 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
[12] M. Troffaes, T. Basu, A Cantelli-type inequality for constructing non-parametric P-boxes based on exchangeability, in: J. De Bock, C.P. de Campos, G. de Cooman, E. Quaeghebeur, G. Wheeler (Eds.), Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications, 103 of *Proceedings of Machine Learning Research*, PMLR, 2019, pp. 386–393.
[13] R. Pelessoni, P. Vicig, Probability inequalities with imprecise previsions, in: M.-J. Lesot, S. Vieira, M.Z. Reformat, J.P. Carvalho, F. Batista, B. Bouchon-Meunier, R.R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer Nature Switzerland, 2025, pp. 65–77.
[14] R. Pelessoni, P. Vicig, Bhatia-Davis inequalities for lower and upper previsions and covariances, Fuzzy Sets Syst. 498 (2025) 109145. https://doi.org/10.1016/j.fss.2024.109145
[15] T. Augustin, F.P.A. Coolen, G. de Cooman, M.C.M. Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley Series in Probability and Statistics, Wiley, 2014. https://doi.org/10.1002/9781118763117
[16] E. Miranda, A survey of the theory of coherent lower previsions, Int. J. Approximate Reasoning 48 (2) (2008) 628–658. https://doi.org/10.1016/j.ijar.2007.12.001
[17] P. Walley, Coherent Lower (and upper) Probabilities, Technical Report 22, Dept. of Statistics, University of Warwick, Coventry, 1981.
[18] P. Walley, Towards a unified theory of imprecise probability, Int. J. Approximate Reasoning 24 (2-3) (2000) 125–148. https://doi.org/10.1016/S0888-613X(00)00031-1
[19] R. Pelessoni, P. Vicig, Convex imprecise previsions, Reliab. Comput. 9 (6) (2003) 465–485. https://doi.org/10.1023/A:1025870204905
[20] R. Pelessoni, P. Vicig, 2-Coherent and 2-convex conditional lower previsions, Int. J. Approximate Reasoning 77 (2016) 66-86. https://doi.org/10.1016/j.ijar.2016.06.003
[21] M.C.M. Troffaes, G. de Cooman, Lower previsions, Wiley, Chichester, West Sussex, United Kingdom, 2014.
[22] 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
[23] R. Pelessoni, P. Vicig, Imprecise previsions for risk measurement, Int. J. Uncertainty Fuzziness Knowledge Based Syst. 11 (04) (2003) 393–412. https://doi.org/10.1142/S0218488503002156
[24] R. Pelessoni, P. Vicig, I. Montes, E. Miranda, Bivariate *p*-boxes, Int. J. Uncertainty Fuzziness Knowledge Based Syst. 24 (02) (2016) 229–263. https://doi.org/10.1142/S0218488516500124
[25] E. Quaeghebeur, Lower and upper covariance, in: D. Dubois, M.A. Lubiano, H. Prade, M.Á. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Soft Methods for Handling Variability and Imprecision, Springer Berlin Heidelberg, 2008, pp. 323–330.