# Lower & upper covariance

Erik Quaeghebeur

The original publication by Springer is available at the DOI 10.1007/978-3-540-85027-4_39 in the book "Soft Methods for Handling Variability and Imprecision" in the series "Advances in Soft Computing". This book contains the proceedings (edited by D. Dubois et al.) of SMPS 2008, which was held in Toulouse on 8–10 September 2008.
Cite as: E. Quaeghebeur (2008). *Lower and Upper Covariance*, Advances in Soft Computing **48**, 323–330.

**Abstract** We give a definition for lower and upper covariance in Walley's theory of imprecise probabilities (or coherent lower previsions) that is direct, i.e., does not refer to credal sets. It generalizes Walley's definition for lower and upper variance. Just like Walley's definition of lower and upper variance, our definition for lower and upper covariance is compatible with the credal set approach; i.e., we also provide a covariance envelope theorem. Our approach mirrors the one taken by Walley: we first reformulate the calculation of a covariance as an optimization problem and then generalize this optimization problem to lower and upper previsions. We also briefly discuss the still unclear meaning of lower and upper (co)variances and mention some ideas about generalizations to other central moments.

**Key words:** variance, covariance, central moment, theory of imprecise probabilities, envelope theorem

## 1 Introduction

In the statistical and probabilistic literature, the variance and covariance of a random variable are important quantities. The generalization of these and other concepts to Walley's [5] theory of imprecise probabilities – which encompasses, in some sense, classical probability theory, Dempster–Shafer theory, and possibility theory – is usually done by taking lower and upper envelopes [see, e.g., 3, 4]. Walley [5, §G] gives a more direct, equivalent definition of lower and upper variance. In this paper, we similarly give a direct definition of the lower and upper covariance concepts.

In the rest of this introduction, we go over some necessary concepts from the theory of classical and imprecise probabilities, respectively. Then follows Section 2, treating Walley's definition of lower and upper variance. This forms a basis for

SYSTeMS Research Group, Ghent University, Technologiepark 914, 9052 Zwijnaarde, Belgium
e-mail: `Erik.Quaeghebeur@UGent.be`

Section 3, which culminates in our definition for lower and upper covariance. The concluding Section 4 briefly touches on further generalizations to other central moments and the meaning of the concepts we have been discussing.

Throughout the paper, we give a running example, recognizable by its small print.

**Concepts from classical probability theory** In the classical setting, a subject is uncertain about the value of some random variable, that, e.g., represents the outcome of an experiment. We here only consider real-valued, bounded random variables, which – because they can be seen as uncertain gains or losses – we call gambles.

Even though he may be uncertain, the subject can express what incomplete knowledge he has by giving an estimate for a number of gambles. In classical probability theory, these estimates are expected values, or fair prices for the gambles [2]. These expected values are collected in the subject's expectation functional or prevision $P$, which maps gambles to fair prices (real numbers). We assume that the subject has specified prices for all gambles defined on some possibility space.

If the subject wants his prices to be reasonable and consistent, such a prevision must satisfy some rationality criteria called coherence conditions [5, §2.8]. Let $f$ and $g$ be any two gambles and let $\lambda$ be a real number, then these criteria are:[1]

$$\text{Boundedness:} \quad \inf f \leq Pf \leq \sup f, \tag{1}$$
$$\text{Linearity:} \quad P(\lambda \cdot f + g) = \lambda \cdot Pf + Pg. \tag{2}$$

Assume that $P$ satisfies them. Call $\mathscr{P}$ the set of previsions satisfying these criteria.[2]

Let us introduce our running example: Take the unit interval $[0,1]$ as the possibility space and let $U$ be the prevision corresponding to the uniform distribution on this possibility space. We shall be using two gambles, the identity gamble id and $1-\mathrm{id}$. We then find $U\,\mathrm{id} = U(1-\mathrm{id}) = \frac{1}{2}$.

**Variance & covariance** When focusing on the uncertainty governing the outcome of some gambles, there are – apart from their expected value – also other so-called statistics that summarize what we know about the gambles. One class of these are the central moments corresponding to a given prevision $P$: let $\mathscr{H}$ be a finite multiset of gambles,[3] then the corresponding central moment $M_P\mathscr{H}$ is defined by

$$M_P\mathscr{H} := P\big(\textstyle\prod_{h\in\mathscr{H}}(h - Ph)\big). \tag{3}$$

In this paper, we are mainly interested in the case where $\mathscr{H}$ contains two gambles, so to the so-called second order central moments. The first of these is the variance of $f$, defined by taking $\mathscr{H} = \{f, f\}$ above: (even in $f$)

$$V_P f := P(f - Pf)^2 = Pf^2 - (Pf)^2. \tag{4}$$

[1] Unless noted otherwise, all introduced symbols retain their meaning in the rest of the paper. For clarity, we do not use parentheses for function application, but only for grouping.

[2] We choose the weak* topology (i.e., the topology of pointwise convergence) on the set $\mathscr{P}$ [5, §D3]. This allows us to later talk about the compactness of its subsets (e.g., credal sets) and about the continuity of functions defined on it (i.e., that have previsions as an argument).

[3] A multiset can contain the same object multiple times, but – like a set – is unordered.

The second, the covariance of $f$ and $g$, is defined by taking $\mathscr{H} = \{f,g\}$: (odd in $f$ and $g$, even in $(f,g)$)

$$C_P\{f,g\} := P\big((f-Pf)\cdot(g-Pg)\big) = P(f\cdot g) - Pf\cdot Pg. \tag{5}$$

For both Equation (4) and (5) the last expression follows from the boundedness (1) and linearity (2) of $P$. They show that both $V_P f$ and $C_P\{f,g\}$ are continuous functions of $P$, a fact that will be of use later on.

For our running example, we find that $V_U\,\mathrm{id} = V_U(1-\mathrm{id}) = \frac{1}{12}$ and $C_U\{\mathrm{id}, 1-\mathrm{id}\} = -\frac{1}{12}$.

**Concepts from the theory of imprecise probabilities** The theory of imprecise probabilities and other generalizations of classical probability theory were all conceived to model aspects of uncertainty that cannot be captured classically. The major difference is that now the subject's prevision – which specifies unique fair (buying *and* selling) prices for gambles – is replaced by a lower prevision $\underline{P}$ and an upper prevision $\overline{P}$, which respectively specify acceptable supremum buying prices and acceptable infimum selling prices (both real numbers) that need not coincide.

In the context of this paper, where lower and upper previsions are defined for all gambles, it is sufficient to give either one of them, because they are related by conjugacy: $\overline{P}f = -\underline{P}(-f)$. We shall here take a lower prevision to be the fundamental concept and an upper prevision the derived one.

The conditions (1) and (2) now get replaced by a weaker set of coherence conditions [5, §2.3], which we assume $\underline{P}$ and $\overline{P}$ to satisfy: (take $\lambda$ positive)

$$\text{Boundedness:} \quad \inf f \le \underline{P}f \le \overline{P}f \le \sup f, \tag{6}$$

(Mixed) super and sublinearity:

$$\lambda\cdot\underline{P}f + \underline{P}g \le \underline{P}(\lambda\cdot f + g) \le \lambda\cdot\underline{P}f + \overline{P}g \le \overline{P}(\lambda\cdot f+g) \le \lambda\cdot\overline{P}f + \overline{P}g. \tag{7}$$

For our example, the subject's lower (upper) prevision is a so-called linear-vacuous mixture: $\underline{U} = \delta\cdot U + (1-\delta)\cdot\inf$, $\overline{U} = \delta\cdot U + (1-\delta)\cdot\sup$, with $\delta\in[0,1]$. Then $\underline{U}\,\mathrm{id} = \underline{U}(1-\mathrm{id}) = \frac{\delta}{2}$ and $\overline{U}\,\mathrm{id} = \overline{U}(1-\mathrm{id}) = 1-\frac{\delta}{2}$. Further on, we also encounter the two gambles $\frac{1}{2}\cdot\big(\mathrm{id}+(1-\mathrm{id})\big) = \frac{1}{2}$ and $\frac{1}{2}\cdot\big(\mathrm{id}-(1-\mathrm{id})\big) = \mathrm{id}-\frac{1}{2}$; for these, we find $\underline{U}\frac{1}{2} = \overline{U}\frac{1}{2} = \frac{1}{2}$, $\underline{U}(\mathrm{id}-\frac{1}{2}) = -\frac{1-\delta}{2}$, and $\overline{U}(\mathrm{id}-\frac{1}{2}) = \frac{1-\delta}{2}$. Taking $\delta$ to be 0 or 1 gives us the vacuous[4] or uniform case, respectively.

The information encoded in the lower prevision $\underline{P}$ can also be represented by the compact convex set $\mathscr{M}\underline{P}$ of previsions that dominate it [5, §3.3]. This so-called credal set is defined by $\mathscr{M}\underline{P} := \{P\in\mathscr{P} | P \ge \underline{P}\}$, where the inequality is a universally quantified pointwise one. The lower or upper prevision for any gamble can be calculated as a lower or upper envelope of the previsions for this gamble present in $\mathscr{M}\underline{P}$:

$$\underline{P}f = \min\nolimits_{P\in\mathscr{M}\underline{P}} Pf \quad\text{and}\quad \overline{P}f = \max\nolimits_{P\in\mathscr{M}\underline{P}} Pf. \tag{8}$$

The minimum and maximum are always attained in an extreme point of $\mathscr{M}\underline{P}$.

[4] With a vacuous lower (or upper) prevision, the subject expresses a total lack of knowledge.

## 2 Lower & upper variance

In this section, we state the results found in Appendix G of Walley's book [5].

**Variance as an optimization problem** What we need to do first, is take another look at the definition (4) of the variance of a gamble. We rewrite and reformulate it as follows: For all real numbers $\mu$, it holds that $V_P f = P\big((f-\mu)+(\mu-Pf)\big)^2 = P(f-\mu)^2-(Pf-\mu)^2$, so $V_P f+(Pf-\mu)^2 = P(f-\mu)^2$. The second left-hand side term is the expression for a parabola with a minimum 0 for $\mu$ in $Pf$; the right-hand side is therefore also an expression for a parabola, with the variance as a minimum, also attained in $Pf$. So an alternative definition for the variance is

$$V_P f = \min_{\mu\in\mathbb{R}} P(f-\mu)^2. \tag{9}$$

**Lower & upper variance** Walley [5, §G1] takes this last definition (9) as inspiration to define the lower and upper variance of a gamble: (even in $f$)

$$\underline{V}_{\underline{P}} f := \min_{\mu\in\mathbb{R}} \underline{P}(f-\mu)^2 \quad \text{and} \quad \overline{V}_{\underline{P}} f := \min_{\mu\in\mathbb{R}} \overline{P}(f-\mu)^2. \tag{10}$$

In this definition, minima and not infima are used, even though $\mathbb{R}$, the set of reals, is open. Let us justify this. First, let $\underline{\overline{P}}$ stand for both $\underline{P}$ and $\overline{P}$, so we can do a parallel derivation for both cases. Furthermore, take $\varepsilon = \mu-\underline{P}f$, then

$$\begin{aligned} \underline{\overline{P}}(f-\mu)^2 = \underline{\overline{P}}(f-\underline{P}f-\varepsilon)^2 &\geq \underline{\overline{P}}(f-\underline{P}f)^2+\varepsilon^2+\underline{P}\big(-2\cdot\varepsilon\cdot(f-\underline{P}f)\big) \\ &\geq \underline{\overline{P}}(f-\underline{P}f)^2+\varepsilon^2 \quad \text{if } \varepsilon\leq 0 \\ &> \underline{\overline{P}}(f-\underline{P}f)^2 \quad \text{if } \varepsilon<0, \text{ i.e., if } \mu<\underline{P}f. \end{aligned} \tag{11}$$

The first inequality follows from (mixed) superadditivity (7). An entirely similar derivation, now with $\varepsilon = \overline{P}f-\mu$, gives us

$$\underline{\overline{P}}(f-\mu)^2 > \underline{\overline{P}}(f-\overline{P}f)^2 \quad \text{if } \varepsilon<0, \text{ i.e., if } \mu>\overline{P}f. \tag{12}$$

Together with the fact that the interval $[\underline{P}f,\overline{P}f]$ is compact and that $\underline{\overline{P}}(f-\mu)^2$ is a continuous[5] function of $\mu$, inequalities (11) and (12) show that, for both the case of the lower and upper variance, a minimum is attained in a $\mu$ that belongs to $[\underline{P}f,\overline{P}f]$.

For our running example, we can use this optimization domain restriction to write $\underline{V}_{\underline{U}}\,\mathrm{id} = \min\big\{\underline{U}(\mathrm{id}-\mu)^2 \mid \mu\in[\frac{\delta}{2},1-\frac{\delta}{2}]\big\}$ and $\overline{V}_{\underline{U}}\,\mathrm{id} = \min\big\{\overline{U}(\mathrm{id}-\mu)^2 \mid \mu\in[\frac{\delta}{2},1-\frac{\delta}{2}]\big\}$. Working this out, we find that for both the minimum is attained for $\mu=\frac{1}{2}$; this gives $\underline{V}_{\underline{U}}\,\mathrm{id} = \frac{\delta}{12}$ and $\overline{V}_{\underline{U}}\,\mathrm{id} = \frac{1}{4}-\frac{\delta}{6}$. Similarly, $\underline{V}_{\underline{U}}\frac{1}{2} = \overline{V}_{\underline{U}}\frac{1}{2} = 0$, $\underline{V}_{\underline{U}}(\mathrm{id}-\frac{1}{2}) = \frac{\delta}{12}$, and $\overline{V}_{\underline{U}}(\mathrm{id}-\frac{1}{2}) = \frac{1}{4}-\frac{\delta}{6}$.

**The variance envelope theorem** Walley [5, §G2] also proves the variance envelope theorem; it allows calculation of lower and upper variances via the credal set:

$$\underline{V}_{\underline{P}} f = \min_{P\in\mathscr{M}\underline{P}} V_P f \quad \text{and} \quad \overline{V}_{\underline{P}} f = \max_{P\in\mathscr{M}\underline{P}} V_P f. \tag{13}$$

[5] A proof consists of the $\varepsilon$–$\delta$-technique together with (7).

In Walley's proof of this theorem, the minimax theorem is used, whose application requires the continuity of $V_P f$ as a function of $P$, which we mentioned below (5). Note that the maximum above is not necessarily attained in an extreme point of $\mathscr{M}\underline{P}$ [5, §G3], in contrast to the situation for upper previsions (8).

## 3 Lower & upper covariance

We are now ready to attack this paper's central topic: to find a *direct* definition for the lower and upper covariance of a pair of gambles, i.e., one that does not involve credal sets. The approach we are going to take is analogous to the one Walley has taken for the lower and upper variance of a gamble: (i) we first define the covariance of a gamble as an optimization problem and (ii) then use this to find expressions for the lower and upper covariance of a pair of gambles. However, for reasons that will become clear at that point, we are going to posit a covariance envelope theorem that mirrors the variance envelope theorem (13) right in between these two steps.

**Covariance as an optimization problem** As we did to obtain (9), we first reformulate the definition (5) of covariance as an optimization problem: For all real $\mu$ and $\nu$, it holds that $C_P\{f,g\} = P\big((f-\mu+\mu-Pf)\cdot(g-\nu+\nu-Pg)\big) = P\big((f-\mu)\cdot(g-\nu)\big) - (Pf-\mu)\cdot(Pg-\nu)$, so then $C_P\{f,g\} + (Pf-\mu)\cdot(Pg-\nu) = P\big((f-\mu)\cdot(g-\nu)\big)$. The second left-hand side term is the expression in $(\mu,\nu)$ for a hyperbolic paraboloid (or saddle surface); the same therefore again holds for the right-hand side. Its saddle point $(\underline{P}f,\underline{P}g)$ can be reached using a minimax (or maximin) operator. This is clearer after a substitution and some rewriting: let $\alpha$ and $\beta$ be real numbers such that $\mu=\alpha+\beta$ and $\nu=\alpha-\beta$, then it holds for all real $\alpha$ and $\beta$ that $P\big((\frac{f+g}{2}-\alpha)^2-(\frac{f-g}{2}-\beta)^2\big) = C_P\{f,g\} + (P\frac{f+g}{2}-\alpha)^2 - (P\frac{f-g}{2}-\beta)^2$, which gives rise to the following defining expression:

$$C_P\{f,g\} = V_P\frac{f+g}{2} - V_P\frac{f-g}{2} = \operatorname{opt}_{\alpha,\beta\in\mathbb{R}} P\big((\tfrac{f+g}{2}-\alpha)^2-(\tfrac{f-g}{2}-\beta)^2\big), \qquad (14)$$

where $\operatorname{opt}_{\alpha,\beta\in\mathbb{R}}$ can be either $\min_{\alpha\in\mathbb{R}}\max_{\beta\in\mathbb{R}}$ or $\max_{\beta\in\mathbb{R}}\min_{\alpha\in\mathbb{R}}$. Because $P$ is linear and its argument is the sum of terms in *either* $\alpha$ *or* $\beta$, it is the same whether we use a maximin or minimax operator.

Proposing a definition for lower and upper covariance would ideally have consisted of just replacing the linear prevision $P$ with the lower prevision $\underline{P}$ and its conjugate upper prevision $\overline{P}$. However, the fact that we have two operators to choose from leaves us with a dilemma: as neither $\underline{P}$ or $\overline{P}$ is linear, does it matter which operator to use for the definition of lower and upper covariance? Perhaps working with credal sets can shed some light on this issue and clarify which of the two choices should be taken (if they do not turn out to be equivalent). This is the next topic.

However, there is one thing we can already say: Independently of the operator, using the same reasoning that led to (11) and (12), it follows that the minimizing $\alpha$ belongs to $[\underline{P}\frac{f+g}{2},\overline{P}\frac{f+g}{2}]$ and – invoking conjugacy – that the maximizing $\beta$ belongs to $[\underline{P}\frac{f-g}{2},\overline{P}\frac{f-g}{2}]$.

For our running example, where we let $f = \mathrm{id}$ and $g = 1 - \mathrm{id}$, these intervals respectively become the singleton $\{\frac{1}{2}\}$ and $[-\frac{1-\delta}{2}, \frac{1-\delta}{2}]$.

**The covariance envelope theorem** Due to the confusing double options we have – with expression (14) – for generalizing the definition of covariance to imprecise probabilities, we need something that can help us choose between them (or show they are both good enough). This something is the covariance envelope theorem, which now rather functions as a definition, and not a theorem to be proven:

$$\underline{C}_{\underline{P}}\{f,g\} := \min_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} \quad \text{and} \quad \overline{C}_{\underline{P}}\{f,g\} := \max_{P\in\mathscr{M}\underline{P}} C_P\{f,g\}. \quad (15)$$

It states that the lower and upper covariance corresponding to $\underline{P}$ can be seen as lower and upper envelopes over the credal set $\mathscr{M}\underline{P}$ of classical covariances. Due to the compactness of $\mathscr{M}\underline{P}$ and the continuity of the covariance as a function of $P$, the minimum and maximum are attained, so this theorem is sensible as an indirect definition. Note that the minimum and maximum above are not necessarily attained in an extreme point of $\mathscr{M}\underline{P}$, in contrast to the situation for lower and upper previsions (8). This theorem is an immediate analog of the variance envelope theorem (13) and it expresses a desirable if not conceptually necessary property.

Looking back at the two equivalent definitions of covariance given by (14), it becomes clear we must investigate whether the maximin and minimax operator encountered there can be interchanged with the maximum or minimum over $P$ encountered in (15). Let us write this out more explicitly. First define the convex functions $u$ and $v$ by $u\alpha = (\frac{f+g}{2} - \alpha)^2$ and $v\beta = (\frac{f-g}{2} - \beta)^2$ then the question is: Which, if any, of the following statements can we ascertain to be true:

$$\begin{aligned} \min_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} &\stackrel{?}{=} \mathrm{opt}_{\alpha,\beta\in\mathbb{R}} \min_{P\in\mathscr{M}\underline{P}} P(u\alpha - v\beta), \\ \max_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} &\stackrel{?}{=} \mathrm{opt}_{\alpha,\beta\in\mathbb{R}} \max_{P\in\mathscr{M}\underline{P}} P(u\alpha - v\beta), \end{aligned} \quad (16)$$

where, as before, $\mathrm{opt}_{\alpha,\beta\in\mathbb{R}}$ can be either $\min_{\alpha\in\mathbb{R}}\max_{\beta\in\mathbb{R}}$ or $\max_{\beta\in\mathbb{R}}\min_{\alpha\in\mathbb{R}}$.

First of all, note that consecutive minimum operators or consecutive maximum operators can always be interchanged.

Whether the interchange of a minimum and a maximum operator is allowed, can only be checked after a more thorough study of the functions involved: As function application is a linear operation, $P(u\alpha - v\beta)$ is linear in $P$ (and thus both convex and concave); as $P$ is linear, $P(u\alpha - v\beta)$ is convex in $\alpha$ and concave in $\beta$. Furthermore, $P(u\alpha - v\beta)$ is continuous in $\alpha$, $\beta$, and $P$. Together with the fact that the maximum or minimum is always attained in some convex compact set, this is enough to do a first operator interchange, i.e., we can apply the minimax theorem [5, §E6] if needed. We find the following modified statements:

$$\min_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} \begin{cases} \stackrel{?}{=} \min_{\alpha\in\mathbb{R}} \min_{P\in\mathscr{M}\underline{P}} \max_{\beta\in\mathbb{R}} P(u\alpha - v\beta), \\ \stackrel{?}{=} \max_{\beta\in\mathbb{R}} \min_{P\in\mathscr{M}\underline{P}} \min_{\alpha\in\mathbb{R}} P(u\alpha - v\beta), \end{cases}$$

$$\max_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} \begin{cases} \stackrel{?}{=} \min_{\alpha\in\mathbb{R}} \max_{P\in\mathscr{M}\underline{P}} \max_{\beta\in\mathbb{R}} P(u\alpha - v\beta), \\ \stackrel{?}{=} \max_{\beta\in\mathbb{R}} \max_{P\in\mathscr{M}\underline{P}} \min_{\alpha\in\mathbb{R}} P(u\alpha - v\beta). \end{cases}$$

As maximizing is a convex operation and minimizing is a concave operation, $\max_{\beta\in\mathbb{R}} P(u\alpha - v\beta)$ is convex as a function of $P$ and $\alpha$ and $\min_{\alpha\in\mathbb{R}} P(u\alpha - v\beta)$ is concave as a function of $P$ and $\beta$. This means that a second application of the maximin theorem is not possible, and a second interchange is not generally possible for all cases. We can therefore only be sure about two of the four initial statements (16):

$$\min_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} = \min_{\alpha\in\mathbb{R}} \max_{\beta\in\mathbb{R}} \underline{P}(u\alpha - v\beta),$$
$$\max_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} = \max_{\beta\in\mathbb{R}} \min_{\alpha\in\mathbb{R}} \overline{P}(u\alpha - v\beta).$$

Thus, the covariance envelope theorem implies a direct definition of lower and upper covariance, the starting point of what follows just below.

**Lower & upper covariance** By combining the covariance envelope theorem (15) with the last expressions we encountered, we find definitions for the lower and upper covariance of a pair of gambles that use a lower or upper prevision, but not the corresponding credal set: (odd in $f$ and $g$, even in $(f,g)$)

$$\begin{aligned} \underline{C}_{\underline{P}}\{f,g\} &= \min_{\alpha\in\mathbb{R}} \max_{\beta\in\mathbb{R}} \underline{P}\big((\tfrac{f+g}{2} - \alpha)^2 - (\tfrac{f-g}{2} - \beta)^2\big), \\ \overline{C}_{\underline{P}}\{f,g\} &= \max_{\beta\in\mathbb{R}} \min_{\alpha\in\mathbb{R}} \overline{P}\big((\tfrac{f+g}{2} - \alpha)^2 - (\tfrac{f-g}{2} - \beta)^2\big). \end{aligned} \tag{17}$$

For our running example, we know we must take $\alpha = \frac{1}{2}$, so then

$$\underline{C}_{\underline{U}}\{\mathrm{id}, 1-\mathrm{id}\} = \max_{\beta\in\mathbb{R}} \underline{U}\big(-(\mathrm{id} - \tfrac{1}{2} - \beta)^2\big) = -\min_{\mu\in\mathbb{R}} \overline{U}(\mathrm{id}-\mu)^2 = -\overline{V}_{\underline{U}}\,\mathrm{id} = -\tfrac{1}{4} + \tfrac{\delta}{6},$$
$$\overline{C}_{\underline{U}}\{\mathrm{id}, 1-\mathrm{id}\} = \max_{\beta\in\mathbb{R}} \overline{U}\big(-(\mathrm{id} - \tfrac{1}{2} - \beta)^2\big) = -\min_{\mu\in\mathbb{R}} \underline{U}(\mathrm{id}-\mu)^2 = -\underline{V}_{\underline{U}}\,\mathrm{id} = -\tfrac{\delta}{12}.$$

An interesting property of classical covariance is that it can be written as a difference of two variances (see equation (14)). For our generalized definition, this identity becomes a string of inequalities:

$$\begin{aligned} &\underline{V}_{\underline{P}}(\tfrac{f+g}{2}) - \overline{V}_{\underline{P}}(\tfrac{f-g}{2}) \le \underline{C}_{\underline{P}}\{f,g\} \\ &\qquad \le \min\big\{\underline{V}_{\underline{P}}(\tfrac{f+g}{2}) - \underline{V}_{\underline{P}}(\tfrac{f-g}{2}), \overline{V}_{\underline{P}}(\tfrac{f+g}{2}) - \overline{V}_{\underline{P}}(\tfrac{f-g}{2})\big\} \\ &\qquad \le \max\big\{\underline{V}_{\underline{P}}(\tfrac{f+g}{2}) - \underline{V}_{\underline{P}}(\tfrac{f-g}{2}), \overline{V}_{\underline{P}}(\tfrac{f+g}{2}) - \overline{V}_{\underline{P}}(\tfrac{f-g}{2})\big\} \\ &\qquad\qquad \le \overline{C}_{\underline{P}}\{f,g\} \le \overline{V}_{\underline{P}}(\tfrac{f+g}{2}) - \underline{V}_{\underline{P}}(\tfrac{f-g}{2}). \end{aligned} \tag{18}$$

These inequalities are obtained starting from the definitions of lower and upper covariance (17). They are related to lower and upper variance (10) by using (mixed) super and sublinearity (7).

For our running example, all but the third of the inequalities in (18) become equalities.

## 4 Conclusions

**Musing on other lower & upper central moments** We started with the definition (3) of a central moment and then restricted ourselves to the second order ones. An obvious (still open) question would now be: Can we generalize the ideas of this paper to higher order central moments?

Independently of whether it is possible or not to give definitions for arbitrary lower and upper central moments using a lower or upper prevision only, it seems desirable that these definitions satisfy a central moment envelope theorem:

$$\underline{M}_{\underline{P}}\mathscr{H} := \min_{P \in \mathscr{M}\underline{P}} M_P \mathscr{H} \quad \text{and} \quad \overline{M}_{\underline{P}}\mathscr{H} := \max_{P \in \mathscr{M}\underline{P}} M_P \mathscr{H}. \tag{19}$$

As before, we could try to write the definition (3) of a central moment as an optimization problem by replacing each $h - Ph$ by $(h - \mu_h) + (\mu_h - Ph)$ (where $h \in \mathscr{H}$ and the $\mu_h$ are real numbers) and separating the term $\underline{P}\big(\prod_{h \in \mathscr{H}}(h - \mu_h)\big)$, as we did to obtain (9) and (14). This could lead to central moments defined as optimization problems; the difficulty with this is that we are going to be dealing with much more complex multilinear expressions (in the $\mu_h$) than parabola or saddle surfaces.

**But what does it mean?** Another thing that is still an open question – to me personally – is: What is the meaning of a lower and upper variance and covariance?

An intuitive interpretation is the one typically given to their precise counterparts:

(i) variance is a statistic describing how much a gamble is believed to vary,
(ii) covariance is a statistic describing how and how much a pair of gambles is believed to vary together.

I have found no satisfactory behavioral interpretation; they could be seen as prices – as we do for previsions –, but trying to say for what leads to all too convoluted explanations, I think. Perhaps variance and covariance should just be seen as useful for the description of probability density or mass functions, and any 'generalization' as mathematically interesting at most.

On the other hand, the fact that no appealing, or intuitively simple behavioral interpretation is known to me, does not mean that it could not be found, e.g., in the economic literature or in other non-classical theories for uncertainty and indeterminacy. Couso et al [1], for example, give a nice overview of definitions for the variance of a *fuzzy* random variable and their interpretation.

**Acknowledgements** I wish to thank Gert de Cooman, Filip Hermans, and two anonymous referees for their useful comments and suggestions. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author.

## References

[1] Couso I, Dubois D, Montes S, Sánchez L (2007) On various definitions of the variance of a fuzzy random variable. In: De Cooman G, Vejnarová J, Zaffalon M (eds) ISIPTA'07: Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, Prague, pp 135–144
[2] de Finetti B (1974–1975) Theory of Probability. Wiley, Chichester
[3] Hall JW (2006) Uncertainty-based sensitivity indices for imprecise probability distributions. Reliability Engineering & System Safety 91(10-11):1443–1451, DOI 10.1016/j.ress.2005.11.042
[4] Kreinovich V, Xiang G, Ferson S (2006) Computing mean and variance under Dempster-Shafer uncertainty: Towards faster algorithms. International Journal of Approximate Reasoning 42(3):212–227, DOI 10.1016/j.ijar.2005.12.001
[5] Walley P (1991) Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London