Kuznetsov Independence for Interval-valued Expectations and Sets of Probability Distributions: Properties and Algorithms

Fabio G. Cozman[a], Cassio Polpo de Campos[b]

[a] *Universidade de São Paulo – Av. Prof. Mello Moraes, 2231, São Paulo, SP – Brazil*
[b] *Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA) – Galleria 2, Manno, 6928 – Switzerland*

Abstract

Kuznetsov independence of variables X and Y means that, for any pair of bounded functions $f(X)$ and $g(Y)$, $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \boxtimes \mathbb{E}[g(Y)]$, where $\mathbb{E}[\cdot]$ denotes interval-valued expectation and $\boxtimes$ denotes interval multiplication. We present properties of Kuznetsov independence for several variables, and connect it with other concepts of independence in the literature; in particular we show that strong extensions are always included in sets of probability distributions whose lower and upper expectations satisfy Kuznetsov independence. We introduce an algorithm that computes lower expectations subject to judgments of Kuznetsov independence by mixing column generation techniques with nonlinear programming. Finally, we define a concept of conditional Kuznetsov independence, and study its graphoid properties.

Key words: Sets of probability distributions, lower expectations, probability and expectation intervals, independence concepts, graphoids.

1. Introduction

A considerable number of theories of inference and decision, in various fields, allow probability values and expectations to be imprecise or indeterminate. Statisticians have long used sets of probability distributions to represent both prior uncertainty [3, 39] or imprecise likelihoods [34, 38], or even lack of identifiability [50]. Economics has also been a prolific source of theories that deal with imprecision and indeterminacy in probabilities and expectations, often under the banner of *Knightian uncertainty* [25, 35, 41, 54]. Statisticians, economists, psychologists and philosophers have paid regular attention to axioms of "rational" behavior that accommodate partially ordered preferences through interval-valued expectations and sets of probability distributions; sometimes this is done to attend to descriptive concerns [7, 37, 60], while often the move to sets of probabilities is normative [26, 46, 55, 58, 63]. There are also several fields that, while perhaps not adopting sets of probability distributions as primitive concepts, do

Preprint submitted to Elsevier *August 30, 2013*

manipulate them explicitly — for example, information theory routinely deals with geometric properties of sets of probability distributions [10, 22].

Research on artificial intelligence has given attention to interval-valued expectations and sets of probability distributions in a variety of forms. Early representation schemes have explored probability intervals [29], multivalued mappings [20, 59], possibility measures [64], random sets [42, 53]. Sets of probability distributions and interval-valued expectations are central elements of most probability logics [31, 32, 36], including some recent logics geared towards ontology management [48, 49]. Sets of probability distributions have also been used to encode abstractions of complex statistical models [27, 30].

An important ingredient of standard probability theory is the concept of *independence*. In modeling languages such as Bayesian and Markov networks, one uses assumptions of stochastic independence to drastically reduce the number of parameters needed to specify a model [56]. Here stochastic independence of events A and B means that $P(A \cap B) = P(A)\,P(B)$. Stochastic independence of (random) variables $\{X_i\}_{i=1}^n$ means that $E[\prod_{i=1}^n f_i(X_i)] = \prod_{i=1}^n E[f_i(X_i)]$ for all bounded functions $f_i(X_i)$.

There is currently no unique concept of independence associated with sets of probability distributions and interval-valued expectations; several concepts have received attention in the literature [9, 16, 18]. A quite compelling proposal, due to V. P. Kuznetsov [43], is to say that two variables X and Y are independent if, for any two bounded functions $f(X)$ and $g(Y)$, we have

$$\mathbb{E}\left[f(X)g(Y)\right] = \mathbb{E}\left[f(X)\right] \boxtimes \mathbb{E}\left[g(Y)\right], \tag{1}$$

where $\mathbb{E}\left[\cdot\right]$ denotes interval-valued expectation, and the product $\boxtimes$ is understood as interval multiplication. Recall:

$$[a, b] \boxtimes [c, d] = [\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)]. \tag{2}$$

Unfortunately, relatively little is known about this concept of independence. The concept was introduced in a book only available in Russian [43], and discussed in a few of Kuznetsov's short publications [44, 45]. The death of Kuznetsov in 1998 stopped work in the concept for a while, and for some time there was discussion in the research community about the relationship between Kuznetsov's ideas and other concepts of independence in the literature. Some of these questions were solved by the first author of the present paper in 2001 [13]. In a subsequent paper the same author studied properties of a conditional version of Kuznetsov independence [14]. Only recently the study of Kuznetsov independence was picked up again by De Cooman, Miranda and Zaffalon [19]; these authors defined Kuznetsov independence for a finite set of variables, introduced several other related concepts of independence, and presented significant results for all of them.

In this paper we present new results for Kuznetsov independence and related concepts of independence. Our contributions are divided in three parts of somewhat different character, presented after some necessary background (Section 2).

Section 3 summarizes what is known about Kuznetsov independence and related concepts, and shows that a closed convex set of probability distributions whose lower and upper expectations satisfy Kuznetsov independence must contain the strong extension of its marginals (and containment can be strict). This result closes several questions left open by De Cooman et al. in their substantial work.

Section 4 examines the computation of lower expectations under judgments of Kuznetsov independence. We derive the optimization problem that must be solved, analyze its properties, and introduce an algorithm that solves it. We also report experiments with our implementation.

Finally, Section 5 proposes a conditional version of Kuznetsov independence, and examines its graphoid properties.

2. Background: Credal sets, lower expectations, extensions

In this paper all variables are assumed to have finitely many values, and all functions are real-valued; therefore, all functions are bounded (we remove the qualifier "bounded" whenever possible). A probability mass function for variable X is denoted by $p(X)$; a probability mass function simply assigns probability $p(x)$ to any value x of X, and it completely specifies an induced probability distribution for X. Given a function $f(X)$, $E_p[f(X)]$ denotes the expectation of function $f(X)$ with respect to $p(X)$. Stochastic independence of variables X and Y obtains when $p(X, Y) = p(X)\, p(Y)$.

A set of probability distributions is called a *credal set* [46]. In this paper we mostly focus on credal sets that are closed and convex; we take the topology induced by Euclidean distance throughout. A credal set defined by a collection of mass functions $p(X)$ is denoted by $K(X)$. We also use $K(X)$ to denote a set of probability mass functions $p(X)$. Given a credal set $K(X)$ and a function $f(X)$, the *lower* and *upper* expectations of $f(X)$ are defined respectively as $\underline{E}[f] = \inf_{p(X) \in K(X)} E_p[f]$ and $\overline{E}[f] = \sup_{p(X) \in K(X)} E_p[f]$; hence $\overline{E}[f] = -\underline{E}[-f]$. A closed credal set and its convex hull produce the same lower and upper expectations. A *lower expectation functional* maps every function to its lower expectation; an *upper expectation functional* maps every function to its upper expectation. The *lower probability* and the *upper probability* of event A are defined respectively as $\underline{P}(A) = \inf_{p(X) \in K(X)} P(A)$ and $\overline{P}(A) = \sup_{p(X) \in K(X)} P(A)$. For any function $f(X)$, a credal set induces an expectation interval $\mathbb{E}[f] = \left[\underline{E}[f], \overline{E}[f]\right]$. Likewise, a probability interval is induced for any event.

A closed convex credal set can be mapped to a unique lower expectation functional and vice-versa [63, Section 3.6.1]. Thus closed convex credal sets and interval-valued expectations have identical expressivity.

Assessments of lower/upper expectations can be viewed as constraints on probability values. An *extension* of a set of assessments is a credal set that satisfies all such constraints. In general we are interested in the largest extension of some given assessments. The largest extension is often referred to as the *natural extension* of the assessments (following terminology by Walley [63]).

A closed convex credal set $K(X)$ is *finitely generated* when it is a polytope in the space of probability distributions for X; that is, the intersection of finitely many closed halfspaces. A *closed halfspace* is a set $\{p \in \mathbb{R}^d : f \cdot p \geq \alpha\}$ for $f \neq 0$. A closed halfspace is defined by a *hyperplane*; that is, a set $\{p \in \mathbb{R}^d : f \cdot p = \alpha\}$ for $f \neq 0$ (f is the *normal vector* to the hyperplane). A function f can be viewed as a vector in $\mathbb{R}^d$; to simplify notation, we use the same letter (f, for instance) to denote a function, a vector, or a normal.

Given a probability mass $p(X)$, the conditional probability mass $p(X|A)$ is obtained by the usual Bayes rule whenever $P(A) > 0$. One may be interested in the set of conditional probability distributions such that $P(A) > 0$ [62], defined as

$$K^{>}(X|A) = \{P(\cdot|A) : P \in K(X) \text{ and } P(A) > 0\} \quad \text{whenever } \overline{P}(A) > 0.$$

If $K(X)$ is convex, then $K^{>}(X|A)$ is convex whenever $\overline{P}(A) > 0$ [46]; moreover, if $K(X)$ is finitely generated and $\overline{P}(A) > 0$, then $K^{>}(X|A)$ is finitely generated (hence closed). In general, $K^{>}(X|A)$ is not closed, as the next example demonstrates.

Example 1. Consider two binary variables X and Y and the closed convex credal set $K(X, Y)$ defined by constraints

$$p_{00} \leq p_{10}, \quad p_{00} \geq (p_{01} - 1/2)^2 + (p_{11} - 1/2)^2,$$

where $p_{xy} = P(\{X = x\} \cap \{Y = y\})$. We then have $P(X = 0|Y = 0) \in (0, 1/2]$ whenever $P(Y = 0) > 0$. The value $p_{00} = 0$ is only obtained by a single probability distribution for which $P(Y = 0) = 0$, hence $P(X = 0|Y = 0) = 0$ is not possible within $K^{>}(X|Y = 0)$. $\square$

We can define a functional as follows: for any $f(X)$,

$$\underline{E}^{>}[f|A] = \inf_{p(\cdot|A) \in K^{>}(X|A)} E_{p(\cdot|A)}[f|A] \quad \text{whenever } \overline{P}(A) > 0;$$

additionally, define $\overline{E}^{>}[f|A] = -\underline{E}^{>}[-f|A]$ whenever $\overline{P}(A) > 0$.

Whenever $\overline{P}(A) > 0$, we have [15, Lemma 1]:

$$\underline{E}^{>}[f(X)|A] = \sup\left(\lambda : \underline{E}[(f(X) - \lambda)I_A(X)] \geq 0\right), \tag{3}$$

where $I_A(X)$ is the indicator function of A (that is, $I_A(x) = 1$ if $x \in A$, and 0 otherwise).

The functional $\underline{E}^{>}$ is often called the *regular extension* of given assessments [63, Appendix J]. Such a functional can be understood as providing a definition of conditioning, even though a range of possible conditioning values can be defined when lower probabilities are equal to zero, as discussed by Miranda [52]. A popular alternative scheme is to consider te set all all conditional mass functions that are coherent with the credal set $K(X)$; in this case one obtain the *natural extension* of given assessments [63]. In this paper we do not adopt any specific definition of conditioning; we use $\underline{E}^{>}$ whenever needed to prove mathematical results.

There are several concepts of independence that can be applied to credal sets [9, 16, 18]. Two concepts that have an intuitive appeal and interesting properties are *epistemic independence* and *strong independence*.

Epistemic independence is based on the concept of *epistemic irrelevance*; this concept initially appeared in the work of Keynes [40] and was later applied to imprecise probabilities by Walley [63, Chapter 9]: Variable Y is *epistemically irrelevant* to X if $\underline{E}[f(X)|Y = y] = \underline{E}[f(X)]$ for any function $f(X)$ and any possible value y of Y (recall that Walley adopts conditioning even on events of zero probability). We then have: Variables X and Y are *epistemically independent* if X is irrelevant to Y and Y is irrelevant to X.

The *epistemic extension* of "marginal" credal sets $K(X)$ and $K(Y)$ is the largest joint credal set that satisfies epistemic independence with marginals $K(X)$ and $K(Y)$ (this has been called *independent natural extension* [19, 63]; we use "epistemic extension" here to emphasize that it adopts epistemic independence). De Cooman et al. have studied epistemic independence and epistemic extensions for sets of variables in great generality [19].

Strong independence focuses instead on factorization of probability distributions: Variables X and Y are *strongly independent* when $K(X, Y)$ is the convex hull of a set of distributions where each distribution satisfies $p(X, Y) = p(X)\, p(Y)$. The generalization for n variables should be clear: their credal set must be the convex hull of a set where each joint distribution factorizes according to stochastic independence.

The *strong extension* of marginal credal sets $K(X_1), \ldots, K(X_n)$ is the largest joint credal set that satisfies strong independence with marginals $K(X_i)$ [9, 12]. The strong extension is intuitively the "product" of the marginal credal sets: Every extreme point of $K(X_i)$ is combined with (multiplied by) every extreme point of $K(X_j)$ (for $i \neq j$) [19, Proposition 8(ii)].

3. Kuznetsov independence and Kuznetsov extensions

Following De Cooman et al. [19], say that $X_1, \ldots, X_n$ are Kuznetsov independent, and that credal set $K(X_1, \ldots, X_n)$ is Kuznetsov, when, for any functions $f_1(X_1), \ldots, f_n(X_n)$,

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i\right] = \boxtimes_{i=1}^{n} \mathbb{E}\left[f_i\right]. \tag{4}$$

If X and Y are Kuznetsov independent, then for any $f(X)$, $g(Y)$,

$$\underline{E}[fg] = \min\left(\underline{E}[f]\,\underline{E}[g]\,, \underline{E}[f]\,\overline{E}[g]\,, \overline{E}[f]\,\underline{E}[g]\,, \overline{E}[f]\,\overline{E}[g]\right); \tag{5}$$

a similar expression can be written for the upper expectation $\overline{E}[fg]$ using Expression (2).

Also, say that $K(X_1, \ldots, X_n)$ is *factorizing* when Expression (4) holds, but, for each set of functions $\{f_1(X_1), \ldots, f_n(X_n)\}$, *only one* function f_j can take negative values, and all other f_k for $k \neq j$ must be non-negative; consequently:

$$\underline{E}\left[\prod_{i=1}^{n} f_i\right] = \min\left(\underline{E}[f_j]\prod_{k\neq j}\underline{E}[f_k]\,, \underline{E}[f_j]\prod_{k\neq j}\overline{E}[f_k]\right),$$

$$\overline{E}\left[\prod_{i=1}^{n} f_i\right] = \max\left(\overline{E}[f_j]\prod_{k\neq j}\underline{E}[f_k]\,, \overline{E}[f_j]\prod_{k\neq j}\overline{E}[f_k]\right).$$

Suppose credal sets $K_1(X_1, \ldots, X_n)$ and $K_2(X_1, \ldots, X_n)$, both with identical marginal credal sets $K(X_i)$, are both Kuznetsov. Clearly, their union is also Kuznetsov. Moreover, any $K(X_1, \ldots, X_n)$ with the same marginal credal sets $K(X_i)$ and such that $K_1 \subseteq K \subseteq K_2$ is clearly Kuznetsov [19, Proposition 31]. The same statements are true if "Kuznetsov" is replaced by "factorizing".

We are often interested in the largest credal set that is Kuznetsov and that satisfies all given assessments; we call this credal set the *Kuznetsov extension* of the assessments. Consider *separately specified credal sets* $K(X_1), \ldots, K(X_n)$ (that is, there is no assessment that involves elements of two distinct marginal credal sets). Their strong extension satisfies Expression (4); hence, their strong extension is contained in their Kuznetsov extension [19, Proposition 8(iv)]. As a consequence, the Kuznetsov extension of separately specified $K(X_1), \ldots, K(X_n)$ must satisfy *external additivity*: for any functions $f_1(X_1), \ldots, f_n(X_n)$,

$$\mathbb{E}\left[\sum_{i=1}^{n} f_i\right] = \boxplus_{i=1}^{n}\mathbb{E}\left[f_i\right],$$

where $\boxplus$ denotes interval addition ($[a, b] \boxplus [c, d] = [a + c, b + d]$). External additivity holds because we always have $\underline{E}[\sum_{i=1}^{n} f_i] \geq \sum_{i=1}^{n}\underline{E}[f_i]$ and $\overline{E}[\sum_{i=1}^{n} f_i] \leq \sum_{i=1}^{n}\overline{E}[f_i]$, and the strong extension guarantees equalities. The name "external additivity" is due to De Cooman et al. [19], who proved external additivity of Kuznetsov extensions of separately specified marginal credal sets. De Cooman et al. also studied a related concept of *strong external additivity*.

We now set out to prove that *any* closed convex credal set $K(X_1, \ldots, X_n)$ that is factorizing must contain the strong extension of its marginal credal sets $K(X_1), \ldots, K(X_n)$. Hence *any* closed convex credal set $K(X_1, \ldots, X_n)$ that is Kuznetsov must contain the strong extension of its marginal credal sets. Consequently any closed convex credal set that is factorizing, and any closed convex credal set that is Kuznetsov, satisfy external additivity. These questions were left open by De Cooman et al. in their investigation [19, Section 9].

We start by examining relationships between factorizing credal sets and conditional expectations. Consider variables $X_1, \ldots, X_n$ and a factorizing credal set $K(X_1, \ldots, X_n)$. Take any X_j and any function $f_j(X_j)$, and any set of events $\{A_i\}_{i\neq j}$ where each A_i is a set of values of X_i. Using Expression (3) and the

	$p(0,0)$	$p(0,1)$	$p(1,0)$	$p(1,1)$
p_1	1/4	1/4	1/4	1/4
p_2	9/25	6/25	6/25	4/25
p_3	3/10	3/10	1/5	1/5
p_4	3/10	1/5	3/10	1/5
p_5	1/3	2/9	2/9	2/9
p_6	3/11	3/11	3/11	2/11

Table 1: The mass functions $p_1(X,Y)$ to $p_6(X,Y)$ for the epistemic extension of $P(X=1) \in [2/5, 1/2]$ and $P(Y=1) \in [2/5, 1/2]$.

fact that $K(X_1, \ldots, X_n)$ is factorizing: whenever $\overline{P}(\cap_{i \neq j} A_i) > 0$,

$$\underline{E}^{>}[f_j(X_j)| \cap_{i \neq j} A_i] = \sup \left(\lambda : \min \left(\begin{array}{c} \underline{E}[f_j - \lambda]\, \underline{P}(\cap_{i \neq j} A_i)\,, \\ \underline{E}[f_j - \lambda]\, \overline{P}(\cap_{i \neq j} A_i) \end{array} \right) \geq 0 \right).$$

If $\lambda \leq \underline{E}[f_j]$, then the condition in the supremum is satisfied; if $\lambda > \underline{E}[f_j]$, then the condition in the supremum is not satisfied. Thus $\lambda = \underline{E}[f_j]$ is the supremum; hence,

$$\underline{E}^{>}[f_j| \cap_{i \neq j} A_i] = \underline{E}[f_j] \qquad \text{whenever } \overline{P}(\cap_{i \neq j} A_i) > 0. \tag{6}$$

De Cooman et al. proved a result similar to Expression (6), showing that a factorizing credal set is a *many-to-one independent product* [19, Section 7].

Our discussion above shows that Kuznetsov independence of X and Y implies epistemic independence of X and Y whenever all lower probabilities are positive. The reverse is not true [13, 19]. For instance, take variables X and Y with values 0 and 1, and build the epistemic extension of assessments $P(X=1) \in [2/5, 1/2]$ and $P(Y=1) \in [2/5, 1/2]$; this epistemic extension has six extreme points listed in Table 1 [63, Section 9.3.4]. For the function $f(X)g(Y)$, where $f(0) = -f(1) = g(0) = -g(1) = 1$, we have $\underline{E}[fg] = -1/11$ with respect to the epistemic extension, but $\underline{E}[fg] = 0$ according to Expression (1).

We now have the tools we need to prove the main result of this section.

Theorem 1. *Suppose that closed convex credal set $K(X_1, \ldots, X_n)$ is factorizing. Then $K(X_1, \ldots, X_n)$ contains the strong extension of its marginal credal sets $K(X_1), \ldots, K(X_n)$.*

PROOF. The proof has three parts. First we study products of strictly positive functions. Then we reason by contradiction to establish that some extreme points of the strong extension must be in $K(X_1, \ldots, X_n)$. Finally, we extend the reasoning to all extreme points of the strong extension.

For each i, adopt $Y_i = \{X_j\}_{j \neq i}$. For a set of functions $\{f_1(X_1), \ldots, f_n(X_n)\}$, for each i, adopt $g_i(Y_i) = \prod_{j \neq i} f_j(X_j)$.

Part 1) Consider strictly positive functions $f_1(X_1), \ldots, f_n(X_n)$. There must be an extreme point $p(X_1, \ldots, X_n)$ of $K(X_1, \ldots, X_n)$ such that $E_p[\prod_{i=1}^n f_i] = \underline{E}[\prod_{i=1}^n f_i]$. Because the credal set $K(X_1, \ldots, X_n)$ is factorizing,

$$E_p[f_i g_i] = E_p\left[\prod_{i=1}^n f_i\right] = \underline{E}\left[\prod_{i=1}^n f_i\right] = \underline{E}[f_i] \prod_{j \neq i} \underline{E}[f_j] = \underline{E}[f_i]\, \underline{E}[g_i] \,.$$

If for some value y_i we have $p(y_i) > 0$, we obviously know that $\overline{P}(Y_i = y_i) > 0$, and then $E_p[f_i|Y_i = y_i] \geq \underline{E}[f_i]$ by Expression (6). Write $E_p[f_i|Y_i] = \underline{E}[f_i] + \sigma(Y_i)$ where $\sigma(y_i) \geq 0$ when $p(y_i) > 0$. Then:

$$\begin{aligned}
\underline{E}[f_i]\, \underline{E}[g_i] = E_p[f_i g_i] &= \sum_{Y_i} \sum_{X_i} f(x_i) g(y_i) p(x_i, y_i) \\
&= \sum_{Y_i : p(y_i) > 0} \sum_{X_i} f(x_i) g(y_i) p(x_i|y_i)\, p(y_i) \\
&= \sum_{Y_i : p(y_i) > 0} E_p[f_i|Y_i = y_i]\, g(y_i) p(y_i) \\
&= \sum_{Y_i : p(y_i) > 0} \underline{E}[f_i]\, g(y_i) p(y_i) + \sum_{Y_i : p(y_i) > 0} \sigma(y_i) g(y_i) p(y_i) \\
&= \underline{E}[f_i]\, E_p[g_i] + \sum_{Y_i : p(y_i) > 0} \sigma(y_i) g(y_i) p(y_i) \\
&\geq \underline{E}[f_i]\, \underline{E}[g_i] + \sum_{Y_i : p(y_i) > 0} \sigma(y_i) g(y_i) p(y_i) \,.
\end{aligned}$$

Hence $\sum_{Y_i : p(y_i) > 0} \sigma(y_i) g(y_i) p(y_i) \leq 0$ and this is only possible if $\sigma(Y_i)$ is zero whenever $p(y_i) > 0$. Consequently, $E_p[f_i|Y_i = y_i] = \underline{E}[f_i]$ whenever $p(y_i) > 0$.

Part 2) Suppose that an extreme point $\prod_{i=1}^n q_i(X_i)$ of the strong extension does not belong to $K(X_1, \ldots, X_n)$. Moreover, assume that each q_i is an *exposed* point of its corresponding (closed and convex) marginal credal set $K(X_i)$. Thus for each $K(X_i)$ we find a function $f_i(X_i)$ such that $E_{q_i}[f_i] = \underline{E}[f_i]$ and $E_{q'_i}[f_i] > \underline{E}[f_i]$ for any other $q'_i(X_i) \in K(X_i)$. As we can add any positive quantity to f_i while maintaining these equalities and inequalities, we can assume each f_i to be strictly positive. For this selection of strictly positive functions $f_1(X_1), \ldots, f_n(X_n)$, take an extreme point of $K(X_1, \ldots, X_n)$, a probability mass function $p(X_1, \ldots, X_n)$ such that $E_p[\prod_{i=1}^n f_i] = \underline{E}[\prod_{i=1}^n f_i]$. Using the first part of the proof: $E_p[f_i|Y_i = y_i] = \underline{E}[f_i]$ whenever $p(y_i) > 0$. Now consider $p(X_i|Y_i = y_i)$ for y_i such that $p(y_i) > 0$. The closed convex credal set $K(X_i)$ is completely characterized by the constraints $E[f(X_i)] \geq \underline{E}[f(X_i)]$; using Expression (6), we have that $p(X_i|Y_i = y_i)$ must satisfy (at least) the same constraints. And within all probability mass functions that satisfy these constraints, only $q_i(X_i)$ is such that $E[f_i] = \underline{E}[f_i]$. So, we must have $p(X_i|Y_i = y_i) = q_i(X_i)$ whenever $p(y_i) > 0$. This implies that $p(X_1, \ldots, X_n) = \prod_{i=1}^n q_i(X_i)$, but this contradicts the fact that $\prod_{i=1}^n q_i(X_i)$ is not in $K(X_1, \ldots, X_n)$. Hence every extreme point of the strong extension that is the product of exposed points of marginal credal sets $K(X_i)$ is in $K(X_1, \dots, X_n)$.

Part 3) Suppose that an extreme point $\prod_{i=1}^n q_i(X_i)$ of the strong extension does not belong to $K(X_1, \dots, X_n)$, and for this extreme point we have that some of the q_i are not exposed points. (Note that an extreme point may fail to be an exposed point [57, Chapter 18].) Because $K(X_1, \dots, X_n)$ is closed, there must be a ball B_δ of radius $\delta > 0$, centered at $\prod_{i=1}^n q_i(X_i)$, lying outside of $K(X_1, \dots, X_n)$. We now construct a joint mass function $\prod_{i=1}^n q_i'(X_i)$ that belongs to B_δ and that is a product of exposed points of $K(X_i)$. To construct $\prod_{i=1}^n q_i'(X_i)$, we use the fact that the set of exposed points is dense in the set of extreme points [57, Theorem 18.6]: there must be an exposed point q_i' of $K(X_i)$ such that $||q_i' - q_i|| \leq \epsilon$ for any $\epsilon > 0$ (recall we are using Euclidean norm). Then $\max |q_i' - q_i| \leq \epsilon$ and, by taking $\epsilon < 1$,

$$
\begin{aligned}
\max \left| \prod_{i=1}^n q_i'(X_i) - \prod_{i=1}^n q_i(X_i) \right| &\leq \max \left| \prod_{i=1}^n (q_i(X_i) + \epsilon) - \prod_{i=1}^n q_i(X_i) \right| \\
&\leq (1+\epsilon)^n - 1 \\
&= \sum_{k=1}^n \binom{n}{k} \epsilon^k \\
&\leq \epsilon \sum_{k=1}^n \binom{n}{k} \\
&= \epsilon(2^n - 1).
\end{aligned}
$$

Hence $|| \prod_{i=1}^n q_i'(X_i) - \prod_{i=1}^n q_i(X_i)|| \leq \epsilon 2^n \sqrt{d}$, where d is the number of values of $(X_1, \dots, X_n)$. Now by taking $\epsilon = \delta/(2^n\sqrt{d})$, we have that $\prod_{i=1}^n q_i'(X_i)$ must belong to B_δ, and so it cannot belong to $K(X_1, \dots, X_n)$. Note that $\prod_{i=1}^n q_i'(X_i)$ must be an extreme point of the strong extension of $K(X_1), \dots, K(X_n)$. For suppose not; then $\prod_{i=1}^n q_i'(X_i) = \alpha \prod_{i=1}^n r_i(X_i) + (1-\alpha) \prod_{i=1}^n s_i(X_i)$ for some $\alpha \in (0,1)$ and mass functions r_i and s_i; now marginalize to obtain $q_i'(X_i) = \alpha r_i(X_i) + (1-\alpha) s_i(X_i)$ for any X_i, a contradiction because q_i' is an exposed point of $K(X_i)$. The fact that $\prod_{i=1}^n q_i'(X_i)$ is an extreme point of the strong extension contradicts the fact that it belongs to B_δ (using the second part of the proof), thus showing that $\prod_{i=1}^n q_i(X_i)$ must belong to $K(X_1, \dots, X_n)$. $\square$

As a digression, note that the proof of Theorem 1 requires only factorization on positive gambles (weaker than the factorizing condition).

We thus have, easily:

Corollary 1. *Suppose that closed convex credal set $K(X_1, \dots, X_n)$ is factorizing. Then $K(X_1, \dots, X_n)$ satisfies external additivity.*

Corollary 2. *Suppose that closed convex credal set $K(X_1, \dots, X_n)$ is Kuznetsov. Then $K(X_1, \dots, X_n)$ contains the strong extension of its marginal credal sets $K(X_1), \dots, K(X_n)$, and satisfies external additivity.*

A natural question is whether the Kuznetsov and the strong extensions of separately specified marginal credal sets are in fact identical. This can be answered positively when variables are binary:[1]

Proposition 1. *Take binary variables X and Y and separately specified closed convex credal sets $K(X)$ and $K(Y)$. The strong and the Kuznetsov extensions of $K(X)$ and $K(Y)$ are identical.*

Proof. Suppose X and Y have values 0 and 1. Suppose $K(X)$ has two distinct extreme points $p_1(X)$ and $p_2(X)$ such that $p_1(0) < p_2(0)$; hence $K(X)$ is specified by inequalities $\sum_X f_i(X)p(X) \geq 0$ for $i = \{1, 2\}$, where $f_1(0) = 1 - p_1(0)$ and $f_1(1) = -p_1(0)$, $f_2(0) = p_2(0) - 1$ and $f_2(1) = p_2(0)$. Likewise, suppose $K(Y)$ has two distinct extreme points $q_1(Y)$ and $q_2(Y)$ such that $q_1(0) < q_2(0)$; hence $K(Y)$ is specified by inequalities $\sum_Y g_i(Y)p(Y) \geq 0$ for $i = \{1, 2\}$. The four extreme points of the strong extension are $p_{11} = p_1q_1$, $p_{12} = p_1q_2$, $p_{21} = p_2q_1$, $p_{22} = p_2q_2$. Each one of the four hyperplanes that define the strong extension goes through three extreme points plus the origin: hyperplane H_1 contains p_{11}, p_{12}, p_{21}; hyperplane H_2 contains p_{11}, p_{12}, p_{22}; hyperplane H_3 contains p_{11}, p_{21}, p_{22}; hyperplane H_4 contains p_{12}, p_{21}, p_{22}. Now note that H_1 is defined by equality $\sum_{X,Y} h(x, y)p(x, y) = 0$ for some $h(X, Y)$; a simple verification shows that $\sum_{X,Y} f_1(x)g_1(y)p(x, y) = 0$ goes through p_{11}, p_{12}, p_{21} and therefore H_1 is specified by the decomposable function $f_1(x)g_1(y)$. Likewise, H_2 must be specified by f_1g_2; H_3 must be specified by f_2g_1; H_4 must be specified by f_2g_2. These decomposable functions define hyperplanes that must also support the Kuznetsov extension; hence the Kuznetsov extension cannot be larger than the strong extension. As the latter must be contained in the former, both are equal. Now suppose that $K(X)$ is actually a singleton containing only $p_1(X)$. Pick up a point $p_2(X)$ such that $p_1(0) < p_2(0)$, construct the hyperplanes $H_1, \ldots, H_4$ as before. Now take the hyperplane H_5 given by $\sum_{X,Y}(-f_1(x))p(x, y) \geq 0$; this hyperplane imposes $P(X = 0) \leq p_1(0)$. Take the intersection of the halfspaces defined by these five hyperplanes, each one of them defined by a decomposable function. The intersection has extreme points p_1q_1 and p_1q_2; hence it is exactly the original strong extension, and the previous reasoning applies. (The only difficulty here is if $p_1(0) = 1$; then take $p_2(X)$ such that $p_2(0) < 1$, and rename p_1 and p_2.) The same argument works for the case where $K(X)$ contains two distinct extreme points but $K(Y)$ is a singleton, simply by renaming extreme points. Finally, if both $K(X)$ and $K(Y)$ are singletons, both the strong and the Kuznetsov extensions are subject to the same constraint $p(X, Y) = p(X)\, p(Y)$. $\square$

The proof of Proposition 1 uses the fact that Kuznetsov independence only deals with decomposable functions. A geometric picture of the situation is that we must carve a region of the unitary simplex with a special chisel, one that can

[1] This result and Example 2 appeared in preliminary form in Ref. [13]; improved versions are presented in this paper.

Extreme points of $K(X)$	Inward normals of facets of $K(X)$	Extreme points of $K(Y)$	Inward normals of facets of $K(Y)$
$[1/5, 3/10, 1/2]$	$[1/4, 3/2, -1]$	$[3/10, 3/10, 2/5]$	$[2/3, 2/3, -1]$
$[2/5, 1/5, 2/5]$	$[-1/6, 7/3, -1]$	$[1/5, 7/10, 1/10]$	$[4, -1, -1]$
$[7/10, 1/5, 1/10]$	$[-7/3, 23/3, 1]$	$[1/5, 2/5, 2/5]$	$[-2/3, 1/21, 1]$
$[3/5, 3/20, 1/4]$	$[7/17, -33/17, 1]$	$[2/5, 7/20, 1/4]$	$[-13/3, 17/3, -1]$

Table 2: Extreme points and inward normals of facets for Example 2.

only deal with decomposable hyperplanes. In fact, given separately specified credal sets $K(X)$ and $K(Y)$, we can make a mental picture of the Kuznetsov extension $K(X,Y)$: it is the smallest set that “wraps” the strong extension with decomposable supporting hyperplanes.

The following example shows that strong extensions can in fact be strictly contained in corresponding Kuznetsov extensions.

Example 2. Consider ternary variables X and Y, and credal sets $K(X)$ and $K(Y)$ with extreme points and facets in Table 2. Figure 1 shows these marginal sets in the same unitary simplex. The strong extension has 16 extreme points and 24 facets (the software **lrs** [1] was used to obtain facets); however, some of these facets cannot be specified using decomposable functions. For example, the hyperplane

$$\begin{aligned}[434, -301, -21, -2836, 1154, 1734, 1164, -96, -1116] &\quad \cdot \\ [p_{00}, p_{01}, p_{02}, p_{10}, p_{11}, p_{12}, p_{20}, p_{21}, p_{22}] &\quad = \quad 0,\end{aligned} \tag{7}$$

where $p_{ij} = p(x_i, y_j)$, supports the strong extension, but it cannot be written as $\sum_{X,Y} h(x,y)p(x,y) = 0$ for some $h(X,Y) = f(X)g(Y) + \alpha$, where α is a constant. (Note that if the function cannot be written as $f(X)g(Y)+\alpha$ for any α, then it cannot specify a hyperplane that supports the Kuznetsov extension.) In fact, the lower expectation of the function in Expression (7) is zero with respect to the strong extension, and -14.5 with respect to the Kuznetsov extension (value obtained with the algorithm in the next section). □

We now examine two other concepts introduced by De Cooman et al. [19]. Say that $K(X_1, \ldots, X_n)$ is *strongly Kuznetsov* when, for any functions $f(\mathbf{W})$ and $g(\mathbf{Z})$, where $\mathbf{W}$ and $\mathbf{Z}$ are disjoint subsets of $\{X_1, \ldots, X_n\}$, we have

$$\mathbb{E}[f(\mathbf{W})g(\mathbf{Z})] = \mathbb{E}[f(\mathbf{W})] \boxtimes \mathbb{E}[g(\mathbf{Z})].$$

Say that $K(X_1, \ldots, X_n)$ is *strongly factorizing* when the same definition holds, but $f(\mathbf{W})$ is restricted to be non-negative.

The strong extension of marginal credal sets $K(X_i)$ clearly satisfies these constraints and is therefore strongly Kuznetsov/factorizing [19, Proposition 8(iv)].

Clearly if $K(X_1, \ldots, X_n)$ is strongly Kuznetsov, it is Kuznetsov; and if $K(X_1, \ldots, X_n)$ is strongly factorizing, it is factorizing. Hence any closed convex

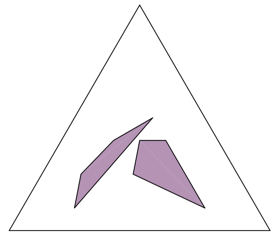

Figure 1: Credal sets in Example 2 in the unitary simplex, viewed from the point $[1,1,1]$.

$K(X_1,\ldots,X_n)$ that is strongly Kuznetsov/factorizing must contain the strong extension of its marginal credal sets $K(X_i)$, and must be externally additive. This closes a few questions left open by De Cooman et al. [19]; the only issue we do not settle here is whether the definitions of Kuznetsov and strongly Kuznetsov independence are equivalent or not.

It seems that any justification one might find for Kuznetsov independence should be a justification for strong Kuznetsov independence as well. However, their computational implications seem to be rather different when n grows, as discussed in the next section.

To conclude this section, we note that one might think that any credal set satisfying strong independence also satisfies Kuznetsov independence. This is not true; a simple example can be constructed by taking a credal set that is the convex hull of probability mass functions $p_1(X,Y)$ and $p_2(X,Y)$ in Table 1. This credal set is smaller than the strong extension of its marginals,[2] and it is not Kuznetsov: for functions $f(X)$ and $g(Y)$ such that $f(0)=3f(1)=3$ and $2g(0)=g(1)=2$, $\overline{E}[fg]=77/25<33/10=\overline{E}[f]\,\overline{E}[g]$.

4. Computing lower expectations

Suppose we have two variables X and Y, each with finitely many values. Suppose also we have separately specified closed convex credal sets $K(X)$ and $K(Y)$. We have a function $h(X,Y)$ and we wish to compute $\underline{E}[h(X,Y)]$ with respect to the Kuznetsov extension of $K(X)$ and $K(Y)$. How can we do it?

We must solve the following optimization problem:

$$\begin{aligned} \inf_{p(X,Y)} \quad & \sum_{X,Y} h(x,y)p(x,y)\,, && (8)\\ \text{subject to:} \quad & \forall x,y : p(x,y)\geq 0, \\ & \sum_{X,Y} p(x,y) = 1, \\ & \forall f(X), g(Y) : E_p[fg] \geq \underline{e}_{fg}, \end{aligned}$$

[2]The credal set satisfies the *repetition independence* concept of Couso et al. [9]: each extreme point of the credal set is the repeated product of a probability mass function.

where $\underline{e}_{fg}$ is the lower bound of the interval $\mathbb{E}[f] \boxtimes \mathbb{E}[g]$ (Expression (5)). As the strong extension of the marginal credal sets belongs to the feasible region, the last set of inequalities suffice to guarantee the equality constraints imposed by Expression (5). We thus have that the feasible region of Optimization problem (8) is the intersection of closed halfspaces, and therefore it is a closed set.

Constraint $\sum_{X,Y} p(x,y) = 1$ is redundant: just take both pairs $f(X) = g(Y) = 1$ and $f(X) = -g(Y) = 1$ to obtain it. Additionally, note that we can impose the following additional constraints on functions $f(X)$ and $g(Y)$, to better condition Optimization problem (8) without changing its result:

$$\forall f(X) : \forall x : f(x) \in [-1,1], \qquad \forall g(Y) : \forall y : g(y) \in [-1,1]. \tag{9}$$

Denote by $\mathcal{F}$ the set of all functions with values in $[-1,1]$. Condition (9) is simply: $f(X) \in \mathcal{F}$, $g(Y) \in \mathcal{F}$.

Hence we can rewrite our optimization problem as follows:

$$\begin{aligned} \min_{p(X,Y)} \quad & \sum_{X,Y} h(x,y)p(x,y)\,, & (10) \\ \text{subject to:} \quad & \forall x,y : p(x,y) \geq 0, \\ & \forall f(X) \in \mathcal{F}, \forall g(Y) \in \mathcal{F} : \sum_{X,Y} f(x)g(y)p(x,y) \geq \underline{e}_{fg}. \end{aligned}$$

This optimization problem is a semi-infinite linear program [28, 47], with infinitely many constraints indexed by f and g. We refer to this problem as the *primal problem*; it can be associated with the following (Haar) *dual problem*, where λ_{fg} denotes the optimization variable associated with the pair (f,g), and λ is the set of all such dual optimization variables:

$$\begin{aligned} \max_{\lambda} \quad & \sum_{f,g} \underline{e}_{fg}\lambda_{fg}, & (11) \\ \text{subject to:} \quad & \forall f \in \mathcal{F}, g \in \mathcal{F} : \lambda_{fg} \geq 0, \\ & \forall x,y : \sum_{f,g} f(x)g(y)\lambda_{fg} \leq h(x,y), \end{aligned}$$

with the constraint that only finitely many optimization variables can be positive.

We now generalize to variables $X_1, \ldots, X_n$. Suppose we have separately specified marginal credal sets $K(X_1), \ldots, K(X_n)$ and we take $K(X_1, \ldots, X_n)$ to be Kuznetsov. The strong extension of separately specified closed convex credal sets again satisfies all constraints. Therefore, to obtain $\underline{E}[h(X_1, \ldots, X_n)]$, we must solve:

$$\begin{aligned} \min_{p(X_1,\ldots,X_n)} \quad & \sum_{X_1,\ldots,X_n} h(x_1,\ldots,x_n)p(x_1,\ldots,x_n)\,, & (12) \\ \text{subject to:} \quad & \forall x_1,\ldots,x_n : p(x_1,\ldots,x_n) \geq 0 \\ & \forall f_i(X_i) \in \mathcal{F} : \sum_{X_1,\ldots,X_n} \left(\prod_{i=1}^{n} f_i(x_i)\right) p(x_1,\ldots,x_n) \geq \underline{e}_{f_1,\ldots,f_n}, \end{aligned}$$

where $\underline{e}_{f_1,\ldots,f_n}$ is the lower bound of the interval $\boxtimes_{i=1}^n \mathbb{E}[f_i]$. The dual problem is then:

$$\begin{aligned} \max_{\lambda} \quad & \sum_{f_1,\ldots,f_n} \underline{e}_{f_1,\ldots,f_n} \lambda_{f_1,\ldots,f_n}, \\ \text{subject to:} \quad & \forall f_i(X_i) \in \mathcal{F} : \lambda_{f_1,\ldots,f_n} \geq 0, \\ & \forall x_1,\ldots,x_n : \sum_{f_1,\ldots,f_n} \left(\prod_i f_i(x) \right) \lambda_{f_1,\ldots,f_n} \leq h(x_1,\ldots,x_n), \end{aligned} \tag{13}$$

with the constraint that only finitely many optimization variables can be positive.

The following theorem collects facts about our primal and dual problems. Some terminology is needed [47]. First, the *duality gap* is the difference between the minimum of the primal problem and the maximum of the dual problem. Second, a *grid* T is a finite set of constraints of the primal problem. A problem is *weakly discretizable* if there is a sequence of grids T_k such that the optimal value subject to constraints in the grid T_k goes to the optimal value of the original problem as $k \to \infty$. A problem is *discretizable* if for every sequence of grids T_k such that the supremum of the distance between constraints goes to zero (precisely: $\sup_{f,g} \min_{f',g'} ||(f,g),(f',g')|| \to 0$), the optimal value subject to constraints in the grid T_k goes to the optimal value of the original problem as $k \to \infty$. Finally, a problem is *finitely reducible* if there is a grid T such that the optimal value subject to constraints in T is equal to the optimal value of the original problem. The fact that a problem is finitely reducible does not mean that its feasible set is finitely generated; it simply means that, given an objective function, one can build an approximate feasible set with finitely many constraints, such that the optimal value is obtained.

Theorem 2. *Optimization problem (12) has a nonempty, bounded, closed and convex feasible region; it is discretizable and finitely reducible; the dual Optimization problem (13) is solvable and the duality gap is zero.*

PROOF. The feasible region contains the (nonempty) strong extension and belongs to the unitary simplex (so it is not the entire space); it is closed and convex because it is the intersection of closed halfspaces. The dual is always solvable because we can build a feasible λ as follows. For each non-zero $h(x_1,\ldots,x_n)$, consider a function $(h(x_1,\ldots,x_n)/|h(x_1,\ldots,x_n)|) \prod_{i=1}^n I_{x_i}(X_i)$, associated with an element of λ equal to $|h(x_1,\ldots,x_n)|$; set all other values of λ to zero, thus producing a feasible λ. Now consider the set $M = \text{cone}\{\prod_i f_i, \forall f_1,\ldots,f_n\}$, called the *first-moment* cone of the primal problem. We have that M is equal to the whole space; hence the relative interior of M is the whole space, and consequently the duality gap is zero [47, Theorem 4(v)]. And then the primal is finitely reducible and weakly discretizable [47, Theorem 7(a)]. Now note that all constraints that depend on f_i are continuous functions of f_i as they are products and minima over summations involving products of f_i and elements of $K(X_i)$ with each other. Other than the finitely many constraints $p(x_1,\ldots,x_n) \geq 0$, the constraints are indexed by a vector $[f_1, \ldots, f_n]$ whose values belong to a closed and convex subset of an Euclidean space (hence a compact set, and consequently a compact Hausdorff topological space). Consequently the primal is discretizable [47, Corollary 1]. $\Box$

Note that if we do not constrain f_i to be in $\mathcal{F}$, all results in the theorem hold except that the primal problem is weakly discretizable instead of discretizable (using the same proof, except for the fact that the index set of constraints is not a compact set).

The literature on semi-infinite linear programming offers a number of schemes to tackle our primal and dual problems — as we have a discretizable primal, we might use several discretization or exchange methods [47]. The gist of these methods is to solve the dual and then to verify whether the primal is satisfied by the obtained solution; if yes, stop, if not, then select a primal constraint so as to add a column to the dual problem. The focus on dual methods is partially motivated by the empirical observation that problems with many columns tend to be more efficiently solved than problems with many constraints [6, Chapter 4]. More importantly, the dual problem leads to column generation methods [4, 21] that come with a host of techniques for speeding-up and bounding solutions. In this paper we emphasize the dual problem and column generation.

We can write down the constraints of the dual problem as $\mathbf{A}\lambda \leq \mathbf{h}$, where $\mathbf{A}$ is a matrix with infinitely many columns and $\mathbf{h}$ is a vector with D elements encoding the function $h(X_1, \ldots, X_n)$. To solve this problem, we must find finitely many tuples $(f_1, \ldots, f_n)$ that construct a suitable sub-matrix of $\mathbf{A}$ (each tuple corresponds to a column of $\mathbf{A}$). In fact, we do not ever need more than D columns at once, where D is the number of tuples $(x_1, \ldots, x_n)$. To generate the columns that matter at the solution of the optimization problem, one must start up with D columns generated arbitrarily, and must iterate by selecting new columns to be added to the pool of columns. When a column is added, the corresponding dual variable $\lambda_{f_1,\ldots,f_n}$ for some other column goes to zero (we can use any linear programming scheme for column removal, as implemented in a linear solver of choice); for improved performance, that column could be removed from the pool of columns.

A column that enters into the pool of columns in a particular iteration must be a column whose associated *reduced cost* is positive (our problem is a maximization); if there is no such column, the maximum of the dual has been found. The reduced cost of a column is given by

$$\underline{e}_{f_1,\ldots,f_n} - \sum_{X_1,\ldots,X_n} \left(\prod_{i=1}^{n} f_i(x_i) \right) p(x_1, \ldots, x_n) \,, \tag{14}$$

where $p(X_1, \ldots, X_n)$ denotes the current primal solution, fixed during each iteration as the algorithm looks for a positive reduced cost. Note that usually a linear programming solver can return the current primal feasible point when solving the dual problem.

Now note that we do not need to explicitly write down the expression of $\underline{e}_{fg}$ while we look for a positive reduced cost. Rather, we can create an optimization problem, starting from Expression (14), as follows:

$$\begin{aligned} \max_{\rho, f_1, \ldots, f_n} \qquad & \rho - \sum_{X_1, \ldots, X_n} \left(\prod_{i=1}^{n} f_i(x_i) \right) p(x_1, \ldots, x_n)\,, \\ \text{subject to:} \qquad & \forall q_i(X_i) \in K(X_i) : \rho \leq \sum_{X_1, \ldots, X_n} \left(\prod_{i=1}^{n} f_i(x_i) q_i(x_i) \right). \end{aligned} \tag{15}$$

The key here is that the maximum of Optimization problem (15) is indeed attained when ρ is equal to $\underline{e}_{f_1,\ldots,f_n}$. So, we obtain another semi-infinite optimization program, one where the objective function and the constraints are multilinear functions of optimization variables $\{f_i(X_i)\}$, and constraints are indexed by $\{q_i(X_i)\}$. Direct analysis of this optimization problem does not seem trivial; we now introduce an assumption that substantially simplifies the optimization. Suppose $K(X_i)$ are finitely generated, so that we have their extreme points; our optimization problem is now finite:

$$\begin{aligned} \max_{\rho, f_1, \ldots, f_n} \qquad & \rho - \sum_{X_1, \ldots, X_n} \left(\prod_{i=1}^{n} f_i(x_i) \right) p(x_1, \ldots, x_n)\,, \\ \text{subject to:} \qquad & \forall q_i(X_i) \in \mathrm{ext} K(X_i) : \rho \leq \sum_{X_1, \ldots, X_n} \left(\prod_{i=1}^{n} f_i(x_i) q_i(x_i) \right), \end{aligned}$$

where $\mathrm{ext}S$ is the set of extreme points of a set S. Clearly, the computational cost comes from the exponential blow up in the number of combinations of extreme points to be analyzed in these auxiliary problems.

In short, the column generation method proceeds by solving auxiliary multilinear optimization problems. At each iteration we must find a column by maximizing reduced cost, exchange columns in and out of the pool of columns, and run a linear optimizer to obtain new dual variables. At any step, if there is no positive reduced cost, we stop as $\underline{E}[h]$ has been found. At any step, if the reduced cost is positive, we have a lower bound for $\underline{E}[h]$.

Due to numeric error (recall that constraints may be infinitely close), we may face difficulties if we wish to run this process until we have the exact minimum. Using column generation techniques we can find heuristic arguments concerning early stopping. We now describe some heuristics we have used in our implementation.

Note first that our primal problem is equivalent to:

$$\begin{aligned}
\min_{p(X_1,\ldots,X_n)} \quad & \sum_{X_1,\ldots,X_n} h(x_1,\ldots,x_n)p(x_1,\ldots,x_n)\,, \\
\text{subject to:} \quad & \forall x_1,\ldots,x_n : p(x_1,\ldots,x_n) \geq 0 \\
& \sum_{X_1,\ldots,X_n} p(x_1,\ldots,x_n) \leq 1, \\
& \forall f_i(X_i) \in \mathcal{F} : \\
& \qquad \sum_{X_1,\ldots,X_n} \left(2 + \prod_{i=1}^{n} f_i(x_i)\right) p(x_1,\ldots,x_n) \geq 2 + \underline{e}_{f_1,\ldots,f_n}.
\end{aligned}$$

This is true because by selecting $f_i(X_i) = 0$ for some X_i, we obtain the constraint $\sum_{X_1,\ldots,X_n} 2p(x_1,\ldots,x_n) \geq 2$, as needed. Now consider the following modified problem, where γ is a large positive constant:

$$\begin{aligned}
\min_{q,p(X_1,\ldots,X_n)} \quad & \gamma q + \sum_{X_1,\ldots,X_n} h(X_1,\ldots,X_n)p(X_1,\ldots,X_n)\,, \qquad (16) \\
\text{subject to:} \quad & q \geq 0,\ \ \forall x_1,\ldots,x_n : p(x_1,\ldots,x_n) \geq 0 \\
& \sum_{X_1,\ldots,X_n} p(X_1,\ldots,X_n) \leq 1 + q, \\
& \forall f_i(X_i) \in \mathcal{F} : \\
& \qquad \sum_{X_1,\ldots,X_n} \left(2 + \prod_{i=1}^{n} f_i(x_i)\right) p(x_1,\ldots,x_n) \geq 2 + \underline{e}_{f_1,\ldots,f_n}.
\end{aligned}$$

The large penalty introduced by γ forces the unitary constraint to hold, at least approximately. Optimization problem (16) has a larger feasible region than the original primal problem; hence the feasible region is nonempty (yet different from the entire space). The relative interior of the first-moment cone M used in the proof of Theorem 2 is again the whole space; hence the duality gap is zero [47, Theorem 4(v)]. This leads us to the following dual problem:

$$\begin{aligned}
\max_{w,\lambda} \quad & -w + \sum_{f_1,\ldots,f_n} (2 + \underline{e}_{f_1,\ldots,f_n})\lambda_{f_1,\ldots,f_n}, \qquad (17) \\
\text{subject to:} \quad & w \geq 0,\ \ \forall f_i(X_i) \in \mathcal{F} : \lambda_{f_1,\ldots,f_n} \geq 0, \\
& w \leq \gamma, \\
& \forall x_1,\ldots,x_n : \\
& \qquad \sum_{f_1,\ldots,f_n} \left(2 + \prod_{i} f_i(x_i)\right) \lambda_{f_1,\ldots,f_n} \leq w + h(x_1,\ldots,x_n),
\end{aligned}$$

with the constraint that only finitely many optimization variables can be positive.

Suppose we fix γ, and solve the dual problem using column generation. If we reach the optimal value *and* the primal problem satisfies $q = 0$, we have clearly reached the optimum of the original problem. If we are still exchanging columns, we can quantify the error incurred by Optimization problem (17) at that iteration, by exploring properties of linear programs [21, Section 2.1]. The reduced cost is the change in the objective function per unit increase in the lower bound of the variable. Hence if we have the value ζ for the current (dual) feasible point, and the maximum reduced cost $\eta > 0$, and an upper bound $\overline{S}$ on the summation $w + \sum_{f_1,\ldots,f_n} \lambda_{f_1,\ldots,f_n}$, we know that $\zeta^* \leq \zeta + \eta \times \overline{S}$, where ζ^* is the maximum in Optimization problem (17) (the dual objective function cannot increase more than $\eta \times \overline{S}$ from the current feasible point). To apply this result, use the fact that, for each $(x_1, \ldots, x_n)$,

$$\begin{aligned} w + \sum_{f_1,\ldots,f_n} \lambda_{f_1,\ldots,f_n} &\leq w + \sum_{f_1,\ldots,f_n} \left(2 + \prod_i f_i(x_i)\right) \lambda_{f_1,\ldots,f_n} \\ &\leq w + w + h(x_1, \ldots, x_n). \end{aligned}$$

Consequently,

$$w + \sum_{f_1,\ldots,f_n} \lambda_{f_1,\ldots,f_n} \leq 2\gamma + \min_{x_1,\ldots,x_n} h(x_1, \ldots, x_n),$$

and, for Optimization problem (17),

$$\zeta \leq \zeta^* \leq \zeta + \eta \times \left(2\gamma + \min_{x_1,\ldots,x_n} h(x_1, \ldots, x_n)\right).$$

If, at an iteration of column generation, the constraint $w \leq \gamma$ is satisfied with equality, we simply increase γ (in doing so we stress the unitary constraint, so moving closer to the original problem). Indeed, there must be a large value of γ that forces q to be zero, given that the duality gap is zero.

In our tests, we normalized $h(X_1, \ldots, X_n)$ such that its values belong to the interval $[0, 1]$, and we always started by selecting $\gamma = 1$; we *always* observed exact satisfaction of the primal constraint $\sum_{X_1,\ldots,X_n} p(x_1, \ldots, x_n) = 1$ without ever increasing γ from its initial value.

As the maximum reduced cost η comes from a nonlinear program, it is important to use a global solver that gives guaranteed upper bounds for the optimal η and guaranteed lower bounds for $\underline{E}[h]$. If the number of variables and states is relatively small, global optimizers such as Couenne [2] are quite effective. In our implementation we have coded the algorithm above with AMPL, using the CPLEX program as the linear optimizer and Couenne [2] as the nonlinear optimizer. We now report some experiments with this implementation. Additional comments on theoretical complexity of our problem can be found at the end of the next section.

First, consider again Example 2. Denote by $h(X, Y)$ the function in Expression (7). We used the algorithm above to obtain $\underline{E}[h]$ as reported in Example 2. Suppose we vary the value of $h(x, y)$ for $x = 0$ and $y = 0$; at $h(0, 0) = 434$ we have the results reported in Example 2. Figure 2 shows the lower expectation of h for varying values of $h(0, 0)$. The result for the strong extension is

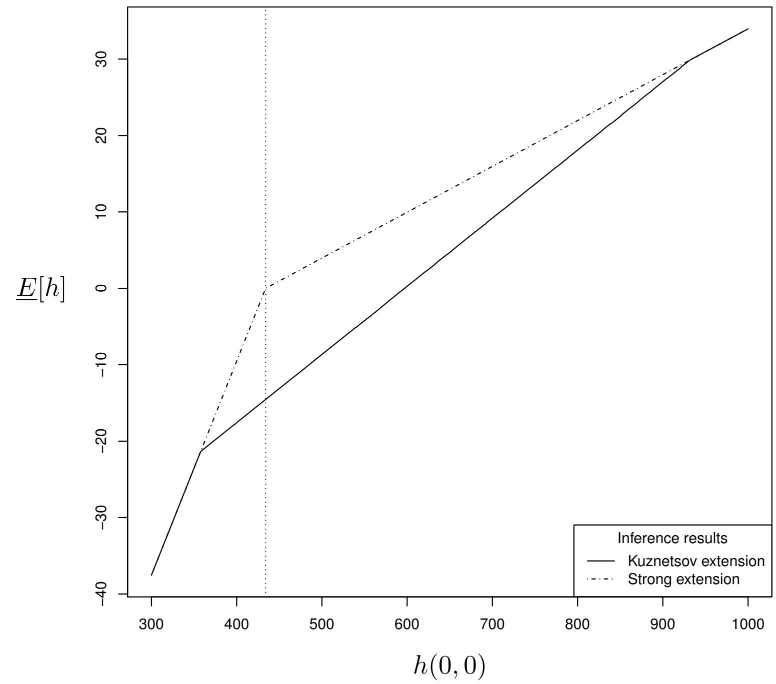

Figure 2: Differences between the lower expectations with respect to strong extension and with respect to Kuznetsov extension using h from Example 2, but varying $h(0,0)$ from 300 to 1000. The vertical line indicates the value of $h(0,0)$ used in Example 2.

shown exactly with dotted-dashed lines. We note that there is nothing special with respect to our choice of varying the value $h(0,0)$; similar figures would be obtained if we allowed other values of h to vary. We also note that only finitely many extreme points of the extensions seem to matter; however we have not been able to theoretically determine whether or not the Kuznetsov extensions of finitely generated marginal credal sets are themselves finitely generated. In all experiments, we stopped iterations when bounds for the dual problem were smaller than 10^{-4}, always at points where $\sum_{X_1,\ldots,X_n} p(x_1,\ldots,x_n) = 1$.

Second, Table 3 depicts the computational effort for randomly generated credal sets $K(X)$ and $K(Y)$ (always separately specified and finitely generated). The number of values for variables X and Y are respectively d_X and d_Y; the number of extreme points of $K(X)$ and $K(Y)$ are respectively v_X and v_Y. Each row of the table presents the mean of time spent (in seconds) and number of generated columns over 20 randomly generated problems. The vast majority of resources are spent by the (linear and nonlinear) solvers themselves.

We have also computed lower expectations with Kuznetsov independence using more than two variables for illustrative purposes. The increase in computational effort is substantial as we move from two to three variables. Suppose we take the same credal sets of Example 2 for the variables X and Y, plus an addi-

d_X	v_X	d_Y	v_Y	Time (sec)	# of Generated Columns
3	3	3	3	0.1 [0,0.2]	14.9 [10,22]
4	4	4	4	0.6 [0.2,1.3]	33.3 [17,58]
5	5	5	5	2.6 [1.3,4.8]	64.1 [40,91]
6	6	6	6	9.3 [5.3,15.4]	140.5 [78,244]
7	7	7	7	26.2 [16.1,51.8]	225.7 [138,520]
8	8	8	8	129.3 [72.5,445.8]	365.1 [192,721]
9	9	9	9	415.6 [215.7,880.2]	553.5 [274,1138]
10	10	10	10	1166.0 [683.4,2537.5]	812.5 [397,2012]
4	5	4	5	0.8 [0.3,2.9]	42.5 [25,75]
4	10	4	10	8.9 [2.9,43.8]	46.0 [25,90]
4	15	4	15	215.8 [36.2,961.4]	120.8 [36,277]
4	20	4	20	1501.4 [131.3,9119.7]	237.9 [57,753]
4	25	4	25	5505.3 [483.7,17574.1]	628.4 [83,1947]

Table 3: Time (mean time [minimum time, maximum time]) and number of generated columns (mean number [minimum number, maximum number]) to solve the optimization problem in random experiments with 20 runs for each scenario.

tional ternary variable Z associated with a credal set $K(Z)$ that has the same extreme points as $K(Y)$. Suppose Z has values 0, 1 and 2, and take $h(X, Y, Z)$ such that $h(X, Y, 0) = h(X, Y)$ of Example 2, and $h(X, Y, 1) = h(X, Y, 2) = 1$. We obtained $\underline{E}[h(X, Y, Z)] = -5.21$ after about an hour of processing, stopping when the bounds indicated less than 1% error. If instead we take $K(Z)$ to be a singleton containing only the uniform distribution and take $h(X, Y, 0) = h(X, Y)$, $h(X, Y, 1) = h(X, Y, 2) = 0$, we obtain -4.84 after about one our of processing, again with bounds indicating less than 1% error (note that the value is exactly one third of the value obtained in Example 2, as expected).

To conclude this section, we now turn to strong Kuznetsov independence. If the credal set $K(X_1, \dots, X_n)$ is strongly Kuznetsov, then we must solve:

$$
\begin{aligned}
\min_{p(X_1,\dots,X_n)} \quad & \sum_{X_1,\dots,X_n} h(x_1, \dots, x_n) p(x_1, \dots, x_n)\,, \qquad (18)\\
\text{subject to:} \quad & \forall x_1, \dots, x_n : p(x_1, \dots, x_n) \geq 0; \text{ and,}\\
& \text{for every partition } \{\{X_{i_1} \dots, X_{i_m}\}, \{X_{i_{m+1}}, \dots, X_{i_n}\}\}\\
& \text{of } \{X_1, \dots, X_n\},\\
& \forall f(X_{i_1} \dots, X_{i_m}) \in \mathcal{F}, g(X_{i_{m+1}}, \dots, X_{i_n}) \in \mathcal{F} :\\
& \sum_{X_1,\dots,X_n} f(x_{i_1} \dots, x_{i_m}) g(x_{i_{m+1}}, \dots, x_{i_n}) p(x_1, \dots, x_n) \geq \underline{e}_{fg}.
\end{aligned}
$$

The dual problem is:

$$
\begin{aligned}
\max_{\lambda} \qquad & \sum_{f,g} \underline{e}_{f,g} \lambda_{f,g}, \\
\text{subject to:} \qquad & \forall f, g \in \mathcal{F} : \lambda_{f,g} \geq 0, \\
& \forall x, \ldots, x_n : \sum_{f,g} fg\lambda_{f,g} \leq h(x_1, \ldots, x_n),
\end{aligned}
\tag{19}
$$

with the constraint that only finitely many optimization variables can be positive, and with the understanding that f and g are functions that operate on disjoint subsets of $\{X_1, \ldots, X_n\}$.

The strong extension of separately specified credal sets satisfies the constraints in Optimization problem (18). Therefore, using the same arguments in the proof of Theorem 2, we obtain:

Theorem 3. *Optimization problem (18) has a nonempty, bounded, closed and convex feasible region; it is discretizable and finitely reducible; the dual Optimization problem (19) is solvable and the duality gap is zero.*

The computational effort demanded by judgments of strong Kuznetsov independence is highly nontrivial. To apply column generation to the dual problem, one must face the maximization of reduced cost under constraints of Kuznetsov independence (to obtain the value of $\underline{e}_{fg}$, one must in general handle Kuznetsov independence). We leave as an open problem the derivation of an actual algorithm that handles strong Kuznetsov independence.

5. Conditional Kuznetsov independence and its graphoid properties

Say that two variables X and Y are conditionally Kuznetsov independent given event A if, for functions $f(X)$ and $g(Y)$,

$$
\mathbb{E}[fg|A] = \mathbb{E}[f|A] \boxtimes \mathbb{E}[g|A]. \tag{20}
$$

The interval-valued expectations given A may be defined in several ways [52]. For instance, we might take $\mathbb{E}[f|A]$ to mean the interval $\left[\underline{E}^{>}[f|A], \overline{E}^{>}[f|A]\right]$ whenever $\overline{P}(A) > 0$. Alternatives would be to define conditioning even on events of zero probability, for instance by resorting to full conditional measures [8, 23]; or to use Bayes rule whenever $\underline{P}(A) > 0$ and take the vacuous model otherwise [63]. Results in this section apply as long as there is agreement on conditioning when $\underline{P}(A) > 0$ (because examples 3, 4 and 5 deal with positive lower probabilities), with the obvious assumption that, whatever conditioning is adopted, $\underline{E}[\cdot|A]$ is the lower expectation of some set of probability mass functions.

Say that two variables X and Y are conditionally Kuznetsov independent given variable Z if, for functions $f(X)$ and $g(Y)$ and for any value of Z,

$$
\mathbb{E}[fg|Z = z] = \mathbb{E}[f|Z = z] \boxtimes \mathbb{E}[g|Z = z]. \tag{21}
$$

To what extent is this concept of conditional Kuznetsov independence a sensible idea? One way to study a concept of independence is to check which *graphoid* properties are satisfied by the concept. Indeed, graphoid properties have been studied in a variety of contexts and provide an abstract framework to study independence [17, 24, 56, 61]. A relation $(X \perp\!\!\!\perp Y \,|\, Z)$ is called a *graphoid* when it satisfies the following axioms [24]:

Symmetry: $(X \perp\!\!\!\perp Y \,|\, Z) \Rightarrow (Y \perp\!\!\!\perp X \,|\, Z)$.

Decomposition: $(X \perp\!\!\!\perp (W,Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,|\, Z)$.

Weak union: $(X \perp\!\!\!\perp (W,Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,|\, (W,Z))$.

Contraction: $(X \perp\!\!\!\perp Y \,|\, Z)$ & $(X \perp\!\!\!\perp W \,|\, (Y,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,|\, Z)$.

The following additional property is often considered:

Redundancy: $(X \perp\!\!\!\perp Y \,|\, X)$.

Finally, the following property is sometimes discussed in connection with positive probability distributions [17, 56]:

Intersection: $(X \perp\!\!\!\perp W \,|\, (Y,Z))$ & $(X \perp\!\!\!\perp Y \,|\, (W,Z)) \Rightarrow (X \perp\!\!\!\perp (W,Y) \,|\, Z)$.

Conditional Kuznetsov independence clearly satisfies Symmetry. Redundancy follows from

$$\begin{aligned} \mathbb{E}\left[f(X)g(Y)|X=x\right] &= f(x) \boxtimes \mathbb{E}\left[g(Y)|X=x\right] \\ &= \mathbb{E}\left[f(X)|X=x\right] \boxtimes \mathbb{E}\left[g(Y)|X=x\right] \end{aligned}$$

for any $f(X)$, $g(Y)$, and any x, whenever expectations are defined (note that the first equality holds both if $f(x) \geq 0$ and if $f(x) < 0$). Decomposition follows from the fact that any function of Y is also a function of Y and W, so we have

$$\mathbb{E}\left[f(X)g(Y)|Z=z\right] = \mathbb{E}\left[f(X)|Z=z\right] \boxtimes \mathbb{E}\left[g(Y)|Z=z\right]$$

when X and (W,Y) are conditionally Kuznetsov independent given Z, whenever expectations are defined.

As for the other properties, we have negative results concerning Contraction and Intersection, even when all events have positive lower probability. It is still an open question whether or not conditional Kuznetsov independence satisfies Weak Union.[3] Consider first failure of Contraction:

Example 3. Take binary variables W, X, and Y, and a credal set $K(W,X,Y)$ such that each extreme point decomposes as $p(W|Y)\,p(X)\,p(Y)$. That is, each

[3]Ref. [14] listed conditions that must be satisfied by Kuznetsov extensions (the conditions are claimed to be sufficient, but they are not), and from there argued that Weak Union holds. That argument is not correct and the status of Weak Union is open.

Extreme point P_i	$P_i(W=0\|Y=0)$	$P_i(W=0\|Y=1)$	$P_i(X=0)$	$P_i(Y=0)$	$E_{P_i}[f]$	$E_{P_i}[h]$	$E_{P_i}[fh]$
P_1	0.7	0.4	0.2	0.2	0.8	0.34	0.272
P_2	0.7	0.4	0.3	0.3	0.7	0.21	0.147
P_3	0.8	0.5	0.2	0.3	0.8	0.11	0.088
P_4	0.8	0.5	0.3	0.2	0.7	0.24	0.168

Table 4: Extreme points of credal set and expectations in Example 3.

extreme point satisfies stochastic independence of X and Y and stochastic independence of X and W conditional on Y. Suppose the credal set has four extreme points; values of $P(W = 0|Y = 0)$, $P(W = 0|Y = 1)$, $P(X = 0)$ and $P(Y = 0)$ are given in Table 4. It can be verified that $K(X, Y)$ contains every product of extreme points for $P(X = 0)$ and $P(Y = 0)$, so $K(X, Y)$ is the Kuznetsov extension for X and Y (using Proposition 1). Likewise, $K(W, X|Y = 0)$ is the Kuznetsov extension of W and X given $\{Y = 0\}$, and $K(W, X|Y = 1)$ is the Kuznetsov extension of W and X given $\{Y = 1\}$. Thus the credal set $K(W, X, Y)$ satisfies Kuznetsov independence of X and Y, and conditional Kuznetsov independence of X and W given Y; but it is not true that X and (W, Y) are Kuznetsov independent. Take the function $f(X)$ such that $f(0) = 0$ and $f(1) = 1$, and the function $h(W, Y)$ such that $h(0, 0) = -h(1, 1) = -1$ and $h(0, 1) = h(1, 0) = 0$; Kuznetsov's condition demands $\underline{E}[fh] = \underline{E}[f]\,\underline{E}[h] = 0.7 \times 0.11 = 0.077$, but $\underline{E}[fh] = 0.088$ for $K(W, X, Y)$. $\Box$

Despite the failure of Contraction for generic credal sets, some special cases may be interesting. For instance, suppose $K(X)$ contains a single probability mass function $p(X)$. Then if X and Y are Kuznetsov independent, X and W are Kuznetsov independent given Y, and moreover if all lower probabilities are positive, then X and (W, Y) are Kuznetsov independent. This is true because, using Expression (6), we have $K(X|W, Y) = \{p(X)\}$, so

$$\underline{E}[f(X)g(W, Y)] = \min_{p'} E_{p'}[gE_{p'}[f|W, Y]] = \min_{p'} E[f]\, E_{p'}[g]\,;$$

thus $\underline{E}[fg] = E[f]\,\underline{E}[g]$ if $E[f] \geq 0$, and $\underline{E}[fg] = E[f]\,\overline{E}[g]$ if $E[f] < 0$, as required by Kuznetsov independence.

Now consider the Intersection property. This property fails for conditional Kuznetsov independence even when all events have positive lower probability:

Example 4. Take binary variables W, X, and Y, and a credal set $K(W, X, Y)$ such that each extreme point decomposes as $p(W)\,p(X)\,p(Y)$. Suppose the credal set has four extreme points; values of $P(W = 0)$, $P(X = 0)$, and $P(Y = 0)$ are given in Table 5. It can be verified that for every w the set $K(X, Y|W = w)$ contains every product of extreme points of $K(X)$ and $K(Y)$; likewise, for every y the set $K(W, X|Y = y)$ contains every product of extreme points of $K(W)$ and $K(X)$. Thus X and W are conditionally Kuznetsov independent given Y, and X and Y are conditionally Kuznetsov independent given W (using Proposition

Extreme point P_i	$P_i(W=0)$	$P_i(X=0)$	$P_i(Y=0)$	$E_{P_i}[f]$	$E_{P_i}[h]$	$E_{P_i}[fh]$
P_1	0.7	0.4	0.2	0.4	0.06	0.024
P_2	0.7	0.5	0.3	0.5	0.09	0.045
P_3	0.8	0.4	0.3	0.4	0.06	0.024
P_4	0.8	0.5	0.2	0.5	0.04	0.020

Table 5: Extreme points of credal set and expectations in Example 4.

1), but it is not true that X and (W, Y) are Kuznetsov independent. Take the function $f(X)$ such that $f(0) = 1$ and $f(1) = 0$, and the function $h(W, Y)$ such that $h(1, 0) = 1$ and $h(W, Y) = 0$ otherwise. Kuznetsov's condition demands $\underline{E}[f]\,\underline{E}[h] = 0.4 \times 0.04 = 0.016$, but $\underline{E}[fh] = 0.020$ for $K(W, X, Y)$. $\square$

Obviously, the failure of some graphoid properties should not prevent us from considering assessments of conditional Kuznetsov independence. While we do not have an algorithm for general sets of assessments, the following example illustrates the matter. This example is interesting because we can exploit Proposition 1 to obtain answers exactly.

Example 5. A *credal network* is a directed acyclic graph where each node is a variable and where a Markov condition applies: every variable is independent of its nondescendants nonparents given its parents [11]. Consider the following graph:

$$X \quad \longrightarrow \quad Y \quad \longrightarrow \quad Z,$$

where X, Y and Z are binary variables, and adopt a version of the Markov condition where "independent" means Kuznetsov independent. This Markov condition implies that X and Z are Kuznetsov independent given Y. Because X and Z are binary variables, the Kuznetsov extension of $K(X|Y = y)$ and $K(Z|Y = y)$ is identical to the strong extension as these sets are separately specified. Thus X and Z are strongly independent given Y and the Kuznetsov extension is the largest credal set satisfying strong independence of X and Z given Y; that is, the Kuznetsov extension of all assessments is exactly their strong extension. We can thus construct an exact polytope that is the Kuznetsov extension in this case. For instance, consider the assessments:

$$\begin{aligned} &P(X=0) \in [1/10, 1/5], & \\ &P(Y=0|X=0) \in [3/5, 7/10], \qquad &P(Y=0|X=1) \in [3/10, 2/5], \\ &P(Z=0|Y=0) \in [2/5, 1/2], \qquad &P(Z=0|Y=1) \in [1/2, 3/5]. \end{aligned}$$

Suppose we are interested in calculating $\underline{P}(Z = 0)$; by enumerating the 32 extreme points of the Kuznetsov extension, we obtain $\underline{P}(Z = 0) = 0.454$. $\square$

In general, inference in credal networks under Kuznetsov independence is most likely a hard task. Such hardness comes from recent results on complexity of credal networks where imprecision in probability values is restricted to vacuous root nodes [51]. In that case, unconditional marginal inferences in the credal network are identical when one adopts either strong or epistemic independence. This fact implies NP-hardness of inferences even in very simple networks [51]. Because Kuznetsov independence leads to extensions that lie between extensions induced by these two other concepts of independence, at least when upper probabilities are positive, the same NP-hardness result should hold for Kuznetsov independence.

6. Conclusion

Results in this paper, together with results by De Cooman, Miranda and Zaffalon [19], should provide the basic machinery for further investigation of Kuznetsov independence. In this paper we have examined the connections between Kuznetsov independence and other concepts of independence (Section 3); in particular we have proved that any credal set that is factorizing must contain the strong extension of its marginal credal sets. Several results are derived from this fact, some of them closing open questions in the literature. Also, we have studied the optimization problem that must be solved when computing lower expectations under judgments of Kuznetsov independence. We have introduced an algorithm to calculate such lower expectations (Section 4), and presented a summary of experiments with our implementation. Finally, we have examined the graphoid properties of a conditional version of Kuznetsov independence (Section 5).

There are challenges left for future work. First, it is important to develop more efficient, perhaps approximate, algorithms for calculation of lower expectations, in particular when several variables interact. Second, it would be useful to know whether or not conditional Kuznetsov independence satisfies the Weak Union property, and to find ways to handle failure of graphoid properties. Third, it would be interesting to know whether strong Kuznetsov independence and Kuznetsov independence are equivalent or not. Finally, future work should evaluate the merits of Kuznetsov independence during elicitation in practical decision making problems. Applied experience would be important in deciding whether to adopt strong, Kuznetsov or epistemic independence in any specific decision problem.

Acknowledgements

Thanks to Peter Walley for bringing our attention to Kuznetsov's work and for showing that many functions of two binary variables can be written as decomposable functions, thus suggesting Proposition 1. Thanks to Serafín Moral for ideas concerning Kuznetsov independence and extensions; to Lev Utkin for generously translating pieces of Kuznetsov's book to English; to Igor Kozine for translating additional material from Kuznetsov's book to English; and to Enrique Miranda for clarifying results concerning many-to-one independent products. Thanks to David Avis for freely distributing the **lrs** package (used in the examples).

We thank the reviewers for detailed reviews that led to substantial improvements to this paper.

The first author has been partially supported by CNPq, and this work has been supported by FAPESP through grant 04/09568-0. The second author has been partially supported by the Hasler Foundation grant n. 10030.

References

[1] D. Avis. lrs: A revised implementation of the reverse search vertex enumeration algorithm. In G. Kalai and G. Ziegler, editors, *Polytopes - Combinatorics and Computation*, pages 177–198. Birkhauser-Verlag, 2000.

[2] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wachter. Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods and Software*, 24(4-5):597–634, 2009.

[3] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

[4] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Belmont, Massachusetts, 1997.

[5] L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 58(1):61–79, January 1991.

[6] S. P. Bradley, A. C. Hax, and T. L. Magnanti. *Applied Mathematical Programming*. Addison-Wesley, 1977.

[7] D. V. Budescu and T. M. Karelitz. Inter-personal communication of precise and imprecise subjective probabilities. In J.-M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *Third International Symposium on Imprecise Probabilities and Their Applications*, pages 91–105. Carleton Scientific, 2003.

[8] G. Coletti and R. Scozzafava. *Probabilistic Logic in a Coherent Setting*. Trends in logic, 15. Kluwer, Dordrecht, 2002.

[9] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5:165–181, 2000.

[10] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[11] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[12] F. G. Cozman. Separation properties of sets of probabilities. In C. Boutilier and M. Goldszmidt, editors, *Conference on Uncertainty in Artificial Intelligence*, pages 107–115, San Francisco, July 2000. Morgan Kaufmann.

[13] F. G. Cozman. Constructing sets of probability measures through Kuznetsov's independence condition. In *Second International Symposium on Imprecise Probabilities and Their Applications*, pages 104–111, Ithaca, New York, 2001.

[14] F. G. Cozman. Computing lower expectations with Kuznetsov's independence condition. In *Third International Symposium on Imprecise Probabilities and Their Applications*, pages 177–187, Lugano, Switzerland, 2003. Carleton Scientific.

[15] F. G. Cozman. Concentration inequalities and laws of large numbers under epistemic and regular irrelevance. *International Journal of Approximate Reasoning*, 51:1069–1084, 2010.

[16] F. G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

[17] A. P. Dawid. Conditional independence. In S. Kotz, C. B. Read, and D. L. Banks, editors, *Encyclopedia of Statistical Sciences, Update Volume 2*, pages 146–153. Wiley, New York, 1999.

[18] L. de Campos and S. Moral. Independence concepts for convex sets of probabilities. In P. Besnard and S. Hanks, editors, *Conference on Uncertainty in Artificial Intelligence*, pages 108–115, San Francisco, California, United States, 1995. Morgan Kaufmann.

[19] G. de Cooman, E. Miranda, and M. Zaffalon. Independent natural extension. *Artificial Intelligence*, 175:1911–1950, 2011.

[20] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[21] J. Desrosiers and M. E. Lübbecke. A primer in column generation. In G. Desaulniers, J. Desrosiers, and M. M. Solomon, editors, *Column Generation*, pages 1–32. Springer, Berlin, 2005.

[22] L. Devroye, L. Gyorfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer Verlag, New York, 1996.

[23] Dubins, L.E.: Finitely additive conditional probability, conglomerability and disintegrations. *Annals of Statistics* 3(1):89–99 (1975)

[24] D. Geiger, T. Verma, and J. Pearl. Identifying independence in Bayesian networks. *Networks*, 20:507–534, 1990.

[25] I. Gilboa and D. Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18(2):141–154, 1989.

[26] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 17–38. University Press, Valencia, Spain, 1980.

[27] R. Givan, S. M. Leach, and T. Dean. Bounded-parameter Markov decision processes. *Artificial Intelligence*, 122(1-2):71–109, 2000.

[28] M. A. Goberna and M. A. López. A comprehensive survey of linear semi-infinite optimization theory. In R. Reemtsen and J.-J. Rückmann, editors, *Semi-Infinite Programming*, pages 3–27. Kluwer Academic Publishers, The Netherlands, 1998.

[29] H. E. Kyburg Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[30] V. Ha and P. Haddawy. Theoretical foundations for abstraction-based probabilistic planning. In E. Horvitz and F. Jensen, editors, *Conference on Uncertainty in Artificial Intelligence*, pages 291–298, San Francisco, California, United States, 1996. Morgan Kaufmann.

[31] T. Hailperin. *Sentential Probability Logic.* Lehigh University Press, Bethlehem, United States, 1996.

[32] J. Y. Halpern. *Reasoning about Uncertainty.* MIT Press, Cambridge, Massachusetts, 2003.

[33] Peter J. Hammond. Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys, editor, *Patrick Suppes: Scientific Philosopher; Volume 1*, pages 25–59. Kluwer, Dordrecht, The Netherlands, 1994.

[34] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York, 1986.

[35] L. P. Hansen and T. J. Sargent. *Robustness.* Princeton University Press, 2007.

[36] P. Hansen and B. Jaumard. Probabilistic satisfiability. Technical Report G-96-31, Les Cahiers du GERAD, École Polytechique de Montréal, 1996.

[37] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel, and C. F. Camerer. Neural systems responding to degrees of uncertainty in human decision-making. *Science*, pages 1680–1683, December 2005.

[38] P. J. Huber. *Robust Statistics.* Wiley, New York, 1980.

[39] J. B. Kadane, editor. *Robustness of Bayesian Analyses*, volume 4 of *Studies in Bayesian Econometrics.* Elsevier Science Pub. Co., New York, 1984.

[40] J. M. Keynes. *A Treatise on Probability.* Macmillan and Co., London, 1921.

[41] F. H. Knight. *Risk, Uncertainty, and Profit.* Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, 1921.

[42] V. Kreinovich. Random sets unify, explain, and aid known uncertainty methods in expert systems. In J. Goutsias, R. P. S. Mahler, and H. T. Nguyen, editors, *Random Sets: Theory and Applications*, pages 321–345. Springer-Verlag, New York, 1997.

[43] V. P. Kuznetsov. *Interval Statistical Methods.* Radio i Svyaz Publ., (in Russian), 1991.

[44] V. P. Kuznetsov. Auxiliary problems of statistical data processing: Interval approach. In *Extended Abstracts of APIC95: International Workshop on Applications of Interval Computations*, pages 123–129, 1995.

[45] V. P. Kuznetsov. Interval methods for processing statistical characteristics. In *Extended Abstracts of APIC95: International Workshop on Applications of Interval Computations*, pages 116–122, 1995.

[46] I. Levi. *The Enterprise of Knowledge.* MIT Press, Cambridge, Massachusetts, 1980.

[47] M. López and G. Still. Semi-infinite programming. *European Journal of Operational Research*, 180(2):491–518, 2007.

[48] T. Lukasiewicz. Expressive probabilistic description logics. *Artificial Intelligence*, 172(6-7):852–883, April 2008.

[49] C. Lutz and L. Schröder. Probabilistic description logics for subjective uncertainty. In F. Lin, U. Sattler, and M. Truszczynski, editors, *Principles of Knowledge Representation and Reasoning*, pages 393–403. AAAI Press, 2010.

[50] C. F. Manski. *Partial Identification of Probability Distributions.* Springer-Verlag, 2003.

[51] D. D. Mauá, C. Polpo de Campos, A. Benavoli, and A. Antonucci. On the complexity of strong and epistemic credal networks. In Ann Nicholson and Padhraic Smyth, editors, *Conference on Uncertainty in Artificial Intelligence*, 2013.

[52] E. Miranda. Updating coherent previsions on finite spaces. *Fuzzy Sets and Systems*, 160(9):1286–1307, 2009.

[53] I. Molchanov. *Theory of Random Sets.* Springer, 2005.

[54] S. Mukerji. A survey of some applications of the idea of ambiguity aversion in economics. *International Journal of Approximate Reasoning*, 24:221–234, 2000.

[55] R. F. Nau. Indeterminate probabilities on finite sets. *Annals of Statistics*, 20(4):1737–1767, 1992.

[56] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.

[57] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[58] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. A representation of partially ordered preferences. *Annals of Statistics*, 23(6):2168–2217, 1995.

[59] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[60] M. Smithson, T. Bartos, and K. Takemura. Human judgement under sample space ignorance. In G. de Cooman, F. G. Cozman, S. Moral, and P. Walley, editors, *First International Symposium on Imprecise Probabilities and Their Applications*, pages 324–332, Ghent, Belgium, 1999. Universiteit Ghent.

[61] M. Studeny. Semigraphoids and structures of probabilistic conditional independence. *Annals of Mathematics and Artificial Intelligence*, 21(1):71–98, 1997.

[62] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3):149–170, 2000.

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

[64] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.