

# Computing Lower and Upper Expectations under Epistemic Independence

Cassio Polpo de Campos <sup>a</sup>, Fabio Gagliardi Cozman <sup>b</sup>

<sup>a</sup>*Escola de Artes, Ciências e Humanidades, USP Leste  
Univ. de São Paulo, São Paulo, SP, Brazil*

<sup>b</sup>*Escola Politécnica, Univ. de São Paulo, São Paulo, SP, Brazil*

---

## Abstract

This paper investigates the computation of lower/upper expectations that must cohere with a collection of probabilistic assessments and a collection of judgements of epistemic independence. New algorithms, based on multilinear programming, are presented, both for independence among events and among random variables. Separation properties of graphical models are also investigated.

*Key words:* Sets of probability measures, concepts of independence, imprecise probabilities, epistemic independence, multilinear programming

---

## 1 Introduction

Among the concepts of independence that have been investigated in connection with sets of probability measures, the concept of *epistemic irrelevance* is probably the easiest to explain — intuitively,  $Y$  is epistemically irrelevant to  $X$  if assessments about  $X$  do not change when we observe  $Y$  [36]. *Epistemic independence* is the symmetric concept:  $X$  and  $Y$  are epistemically independent if each one is epistemically irrelevant to the other. Despite their intuitive content, epistemic irrelevance and epistemic independence are quite difficult to handle computationally. Given probabilistic assessments and judgements of epistemic irrelevance, how can one compute lower and upper expectations?

Our main contribution in this paper is to show that judgements of epistemic irrelevance can generally be recast as multilinear constraints. We show how to compute

---

*Email addresses:* cassio@ire.usp.br (Cassio Polpo de Campos),  
fgcozman@usp.br (Fabio Gagliardi Cozman).

lower/upper probabilities that take into account epistemic irrelevance among events through multilinear programming. We then extend the multilinear programming approach to independence among variables — in the presence of sets of probability measures, there are essential differences between independence among events and among variables. We apply our multilinear approach to multivariate models with graph-theoretical representations, often called credal networks. We consider credal networks under “epistemic irrelevance” and “epistemic independence” semantics, and investigate separation properties of these networks.

Section 2 presents a few relevant definitions and results. Section 3 introduces our multilinear approach to epistemic irrelevance among events (Appendix A compares our approach to Walley’s algorithm for epistemic irrelevance). Sections 4 and 5 look respectively into credal networks and separation properties of Markov chains. Section 6 concludes the paper.

## 2 Credal sets and concepts of independence

We deal with *categorical* random variables. To distinguish random variables from variables used in optimization problems, we refer to the latter as *optimization variables*.

A set of probability measures induced by distributions on random variable  $X$  is denoted by  $K(X)$  and called a *credal set*. A *joint* credal set  $K(\mathbf{X})$  contains joint probability measures for random variables  $\mathbf{X}$ . Given a credal set  $K(X)$ , the *lower expectation* and the *upper expectation* of a bounded function  $f(X)$  are defined respectively as  $\underline{E}[f(X)] = \inf_{P \in K(X)} E_P[f(X)]$  and  $\overline{E}[f(X)] = \sup_{P \in K(X)} E_P[f(X)]$ , where  $E_P[f(X)]$  is standard expectation. Lower/upper probabilities are defined similarly. Conditioning is performed by applying Bayes rule to each measure in a credal set; the posterior credal set is the union of all posterior probability measures [19]. A *conditional* credal set  $K(X|A)$  contains conditional measures on the event  $A$ . We defer to future work the very important case of conditioning on events with zero probability [5,10,34]; here we assume throughout that any conditioning event has lower probability strictly larger than zero.

Lower and upper expectations can be viewed as linear constraints on probabilities:  $\underline{E}[f(X)] \leq E_P[f(X)] \leq \overline{E}[f(X)]$ . Conditional lower and upper expectations also yield linear constraints, as  $\underline{E}[f(X)|A] = \alpha$  is equivalent to  $\underline{E}[A(X)(f(X) - \alpha)] = 0$ , where we use  $A(X)$  for the indicator function of event  $A$  defined by  $X$  (this equation is Walley’s *generalized Bayes rule* [36]; we note that this equation is valid only under the assumption of positive lower probabilities). If a collection of lower/upper expectations defines a convex set of probability measures, such that every constraint is tight, we say that the lower/upper expectations are *coherent*. We do not assume that every given set of assessments is coherent; we assume only that any such set

defines a non-empty set of measures and thus can be made coherent by adjusting some assessments. A set of constraints with this property is said to *avoid uniform loss* [36,37], or alternatively, to be *g-coherent* [2]. In general, we are interested in the *largest* set of probability measures that satisfies a set of constraints — constraints may not be coherent but must avoid uniform loss. We call this largest set the *natural extension* of the constraints, adapting Walley’s terminology [36].

Several concepts of independence can be used when one deals with credal sets [6,11,16,36]. We review here three non-equivalent concepts; relationships between them have received considerable attention in the literature [9,13,25].

The most commonly adopted concept is *strong independence*:<sup>1</sup> Events  $A$  and  $B$  are *strongly independent* when every vertex of the underlying credal set  $K$  satisfies standard stochastic independence of  $A$  and  $B$ . Likewise, random variables  $X$  and  $Y$  are strongly independent when every vertex of the underlying credal set satisfies standard stochastic independence of  $X$  and  $Y$ . Conditional strong independence (for events and for random variables) is obtained by demanding that vertices satisfy conditional stochastic independence. Strong independence usually produces a multilinear program, as the following example illustrates.<sup>2</sup>

**Example 1** Consider a generalized version of Boole’s challenge problem [21]. Take three Boolean random variables  $X_1, X_2$  and  $X_3$ ; random variable  $X_i$  takes values  $i$  and  $\hat{i}$ . We want to find tight bounds on  $P(X_3 = 3)$ . Whenever possible we indicate the events  $\{X_i = i\}$  and  $\{X_i = \hat{i}\}$  simply by  $i$  and  $\hat{i}$ , and we indicate conjunction of events  $A \wedge B$  simply by  $A, B$ . Suppose we have  $P(1) \in [l_1, u_1]$ ,  $P(2) \in [l_2, u_2]$ ,  $P(1, 3) \in [l_3, u_3]$ ,  $P(2, 3) \in [l_4, u_4]$ ,  $P(\hat{1}, \hat{2}, 3) = 0$ , with  $l_i$  strictly larger than 0 for all  $i$ . Suppose also that  $X_1$  and  $X_2$  are strongly independent; given that relevant probabilities are positive, strong independence implies  $P(1, 2) = P(1)P(2)$  for every vertex of  $K(X_1, X_2)$ . Defining  $p_1 = P(1, 2, 3)$ ,  $p_2 = P(1, 2, \hat{3})$ ,  $p_3 = P(1, \hat{2}, 3)$ ,  $p_4 = P(1, \hat{2}, \hat{3})$ ,  $p_5 = P(\hat{1}, 2, 3)$ ,  $p_6 = P(\hat{1}, 2, \hat{3})$ ,  $p_7 = P(\hat{1}, \hat{2}, 3)$ ,  $p_8 = P(\hat{1}, \hat{2}, \hat{3})$ , we obtain bounds for  $P(3)$  by computing:

$$\begin{aligned} & \max / \min \quad p_1 + p_3 + p_5 + p_7, \\ & \text{subject to} \quad p_1 + p_2 + p_3 + p_4 = \pi_1, \quad p_1 + p_2 + p_5 + p_6 = \pi_2, \\ & \quad p_1 + p_3 = \pi_3, \quad p_1 + p_5 = \pi_4, \quad p_7 = 0, \quad p_1 + p_2 = \pi_1 \pi_2, \\ & \quad p_1 + \dots + p_8 = 1, \quad l_i \leq \pi_i \leq u_i, \quad p_k \geq 0. \end{aligned}$$

Suppose  $l_1 = 0.1$ ,  $l_2 = 0.2$ ,  $l_3 = 0.1$ ,  $l_4 = 0.3$ ,  $u_1 = 0.5$ ,  $u_2 = 0.8$ ,  $u_3 = 0.3$ ,  $u_4 = 0.7$ . The solution of this multilinear program yields  $P(3) \in [0.3, 0.79]$ . If the independence judgement is dropped, a linear program produces  $P(3) \in [0.3, 1.0]$ .

<sup>1</sup> We should note that terminology is not completely standardized on this topic [6,9].

<sup>2</sup> Multilinear programming has also been related to other concepts of independence, for example independence in comparative probabilities [4].

Thus strong independence requires us to solve multilinear programs. Existing solution methods produce sequences of sub-problems using either branch-and-bound or cutting-plane techniques [20,22,24,28,31]. The algorithms of Maranas and Floudas [24], and Gochet and Smeers [20] produce convex nonlinear sub-problems, while Sherali and Adams' algorithm produces linear sub-problems [28]. We employ Sherali and Adams' branch-and-bound algorithm in our calculations, as it is particularly appropriate for computing lower/upper expectations — because the sub-problems generated by Sherali and Adams' algorithm are linear programs, column generation and other valuable techniques can be employed [21].

A different definition of independence is Kuznetsov's [23]:  $X$  and  $Y$  are *Kuznetsov independent* when the interval of expected values  $\mathbb{E}[f(X)g(Y)]$  is equal to the interval-product of the intervals  $\mathbb{E}[f(X)]$  and  $\mathbb{E}[g(Y)]$ , for any bounded  $f(X)$  and  $g(Y)$ . Little is known about the computation of lower/upper expectations under judgements of Kuznetsov independence; the existing method works by explicitly constructing a joint credal set [11], a potentially complex operation that is not applicable to large multivariate settings in any obvious way.

A third concept of independence for credal sets is *epistemic independence* [35,36]. In many ways, this is the concept with the most appealing definition, because it can be given a direct behavioral interpretation. We now present the relevant definitions both for events and random variables:

**Definition 2** *Event  $A$  is epistemically irrelevant to event  $B$  given event  $C$  when  $\underline{P}(B|A, C) = \underline{P}(B|A^c, C) = \underline{P}(B|C)$  and  $\overline{P}(B|A, C) = \overline{P}(B|A^c, C) = \overline{P}(B|C)$ .*

We indicate that  $A$  is epistemically irrelevant to  $B$  given  $C$  by  $EIR(A, B|C)$ . Unlike standard stochastic independence, epistemic irrelevance is not symmetric. Walley defines epistemic independence as the “symmetrized” concept:

**Definition 3** *Events  $A$  and  $B$  are epistemically independent given event  $C$  when  $EIR(A, B|C)$  and  $EIR(B, A|C)$ .*

We indicate that  $A$  and  $B$  are epistemically independent given  $C$  by  $EIN(A, B|C)$ .

Consider now definitions for random variables.

**Definition 4** *Random variable  $X$  is epistemically irrelevant to random variable  $Y$  given event  $C$  when  $\underline{E}[f(Y)|X = x, C] = \underline{E}[f(Y)|C]$  for any bounded  $f(Y)$  and any  $x$ .*

We indicate that  $X$  is epistemically irrelevant to  $Y$  given  $C$  by  $EIR(X, Y|C)$ .

**Definition 5** *Random variables  $X$  and  $Y$  are epistemically independent given event  $C$  when  $EIR(X, Y|C)$  and  $EIR(Y, X|C)$ .*

We use  $EIN(X, Y|C)$  to indicate that  $X$  and  $Y$  are epistemically independent given  $C$ . We can also have irrelevance and independence conditional on a random variable  $Z$ ; as we restrict ourselves to categorical random variables, the judgement  $EIR(X, Y|Z)$  simply means that  $EIR(X, Y|Z = z)$  for every possible value  $z$  of  $Z$  (and likewise for epistemic independence).

Note that irrelevance for random variables is not a straightforward generalization of irrelevance for events (in fact, this leads us to consider different algorithms for events and random variables in the remainder of the paper). Equality of lower expectations for all bounded functions means equality of the *convex hull* of credal sets (assuming that sets are closed; as we assume that assessments do not include strict inequalities, we in fact deal with closed sets). Thus  $X$  is epistemically irrelevant to  $Y$  when the convex hull of  $K(Y|X = x)$  is equal to the convex hull of  $K(Y)$  for any  $x$ ; we indicate this by  $K(Y|X = x) \cong K(Y)$ .

### 3 Epistemic independence for events

In this section we propose a multilinear programming formulation for the computation of upper expectations under judgements of epistemic irrelevance of events. The computation of lower expectations can be tackled similarly. We focus on epistemic irrelevance as any judgement of epistemic independence can be expressed as two judgements of epistemic irrelevance.

Consider that  $s$  assessments are given as pairs  $(\underline{P}(F_i|G_i), \alpha_i)$ . We interpret such an assessment as  $P(F_i|G_i) \geq \alpha_i$  rather than  $\underline{P}(F_i|G_i) = \alpha_i$ ; if the assessments are not coherent, it may be impossible to enforce equality. Assessments can be encoded as

$$E[G_i(F_i - \alpha_i)] \geq 0, \tag{1}$$

given our assumption of strictly positive lower probabilities for conditioning events.

Suppose we want to compute  $\overline{P}(D)$  for an event  $D$ . We are then dealing with events  $D$ ,  $\{F_i\}_{i=1}^c$  and  $\{G_i\}_{i=1}^s$ . Now take the set  $\Omega$  containing the  $N$  atomic events  $\omega_k$ , where each atomic event  $\omega_k$  is a complete conjunction of events (or their negations). Note that  $N$  can be exponential on the number of assessments and judgements. Each event  $A$  can be viewed as a function  $A : \Omega \rightarrow \{0, 1\}$  (that is, the event can be identified with its indicator function). For example, if the  $i$ th assessment is unconditional, then  $G_i(\omega_k) = 1$  for every  $\omega_k$ .

Denote by  $p_k$  the probability of the  $k$ th atomic event  $\omega_k$ . Hence we can write  $\overline{P}(D)$  as  $\max \sum_k D(\omega_k)p_k$  subject to linear constraints (1), also expressed in terms of the  $p_k$  given the assumption of strictly positive lower probabilities. At this point we have encoded assessments into a linear program, as usually done in probabilistic logic [21].

Now consider that  $r$  judgements of epistemic irrelevance  $EIR(A_j, B_j|C_j)$  are given, in addition to the  $s$  constraints  $P(F_i|G_i) \geq \alpha_i$ . Each judgement  $EIR(A_j, B_j|C_j)$  introduces constraints such as  $\min P(B_j|A_j, C_j) = \min P(B_j|C_j)$ , where both minima are taken with respect to the underlying credal set. As we now show, it is possible to express irrelevance relations through multilinear constraints. To do so, introduce new optimization variables  $\nu_j$  and  $\mu_j$ , and generate the following inequalities (note that inequality symbols are numbered, as their order is used later):

$$\begin{aligned} \nu_j &\leq_1 P(B_j|A_j, C_j) \leq_4 \mu_j, \\ \nu_j &\leq_2 P(B_j|A_j^G, C_j) \leq_5 \mu_j, \\ \nu_j &\leq_3 P(B_j|C_j) \leq_6 \mu_j. \end{aligned} \tag{2}$$

By clearing the denominators, these inequalities become multilinear expressions on the  $p_k$ ,  $\nu_j$  and  $\mu_j$ . Note that we can clear the denominators given our assumption of positive probability for conditioning events.

Denote by  $\mathcal{C}_0$  the set of linear constraints  $E[G_i(F_i - \alpha_i)] \geq 0$ , plus the constraints  $p_k \geq 0$ ,  $\sum_k p_k = 1$  and the  $6r$  inequalities (2). Now construct  $6r$  additional sets of  $N$  optimization variables. Denote by  $\mathbf{q}_{j,l}$  each one of these  $6r$  sets of optimization variables — there is one set for each judgement of irrelevance (where  $j = 1, \dots, r$ ) and for each inequality in (2) (where  $l = 1, \dots, 6$  indicates which inequality is used, following the numbering in (2)).

The idea is simple. For the  $r$ th judgement of irrelevance and the  $l$ th inequality, there must be a measure on the underlying joint credal that satisfies the inequality with equality. As each inequality may be satisfied with equality by a different measure, we must create as many measures as there are inequalities. For example, optimization variables  $\mathbf{q}_{3,4}$  will have to satisfy  $P(B_3|A_3, C_3) = \mu_3$ , or rather

$$P(A_3, B_3, C_3) = \mu_3 P(A_3, C_3). \tag{3}$$

Thus we construct  $6r$  sets of constraints, a set of constraints for each set of variables  $\mathbf{q}_{j,l}$ . In fact, the set of constraints  $\mathcal{C}_{j,l}$  only refers to optimization variables in  $\mathbf{q}_{j,l}$ . The constraints are identical to the ones in  $\mathcal{C}_0$ , except that: (1) instead of optimization variable  $p_k$  we have  $q_{j,l,k}$ ; (2) the  $l$ th inequality of the  $j$ th judgement is replaced by equality. We obtain a set of  $6r + 1$  loosely coupled systems of multilinear constraints; the connection between these systems is given by the  $\nu_j$  and  $\mu_j$ . By construction, we have:

**Theorem 6** *The value of  $\overline{P}(D)$  is given by  $\max P(D)$  (as a linear expression of  $p_k$ ) subject to  $\mathcal{C}_0$  and  $\mathcal{C}_{j,l}$  for  $j = 1, \dots, r$  and  $l = 1, \dots, 6$ .*

To illustrate this result, we revisit Example 1:

**Example 7** *Consider the same assessments described in Example 1, but replace*

*the strong independence judgement with the epistemic independence judgement  $EIN(X_1, X_2)$  (that is, two judgements of epistemic irrelevance). As we deal only with binary variables, we can treat them as events. To compute  $P(3)$  we must deal with 13 groups of 8 optimization variables and approximately 300 constraints, many of which are multilinear. Our implementation of Sherali and Adams' method readily produces  $P(3) \in [0.3, 0.85]$ .*

The previous discussion can be adapted to produce conditional upper expectations of the form  $\bar{P}(D|E)$ . We start with a fractional multilinear program where the objective function is  $\max P(D, E)/P(E)$ . Now define  $t = P(E)$ ; the objective function then is  $\max t^{-1} \sum_k D(\omega_k)E(\omega_k)p_k$ . Given our assumption that  $t > 0$ , we can multiply by  $t^{-1}$  both sides of constraints (1), (2) or (3). If we distribute  $t^{-1}$  and replace every product  $t^{-1}p_k$  by a new optimization variable  $p'_k$ , and every product  $t^{-1}q_{j,l,k}$  by a new optimization variable  $q'_{j,l,k}$ , we obtain a multilinear program that is essentially identical to the original fractional multilinear program. There are a few differences; most notably, the objective function becomes  $\max \sum_k D(\omega_k)E(\omega_k)p'_k$ . Also, the definition  $t = P(E)$  leads to the new constraint  $\sum_k E(\omega_k)p'_k = 1$ . Finally, the unitary constraint  $\sum_k p_k = 1$  becomes  $\sum_k p'_k = t^{-1}$ ; likewise,  $\sum_k q_{j,l,k} = 1$  becomes  $\sum_k q'_{j,l,k} = t^{-1}$ . As the last constraint is simply  $\sum_k q'_{j,l,k} = \sum_k p'_k$ , the auxiliary optimization variable  $t$  can be ignored in the presence of the other constraints. Note that this technique is similar to the Charnes-Cooper transformation used in linear fractional programming [7].

The techniques outlined in this section remain essentially untouched for assessments of lower expectations; that is, assessments (not necessarily coherent, but avoiding uniform loss) involving functions of random variables such as  $(\underline{E}[f_i|G_i], \alpha_i)$ , interpreted as linear constraints  $E[G_i(f_i - \alpha_i)] \geq 0$ .

Section 6 and Appendix A briefly compare our multilinear programming approach with a proposal by Walley to deal with epistemic independence of events.

#### **4 Epistemic independence for random variables: credal networks**

While judgements of epistemic independence between events imply a fixed number of equalities among lower and upper probabilities, epistemic independence between random variables requires that credal sets have identical convex hulls, and these convex hulls can be rather complex objects. It does not seem that arbitrary judgements of epistemic irrelevance can be dealt with in any simple form — in Appendix A we discuss how Walley's algorithm for irrelevance among events can be generalized to deal with irrelevance among random variables, but the resulting method faces steep theoretical and computational difficulties. In this section we take a different route. Instead of dealing with arbitrary judgements of irrelevance among random variables, we focus on judgements that can be organized

using graphs: we explore compact representations of credal sets that are inspired by Bayesian networks and other graphical models [27].

We thus consider *credal networks* as our representation for judgements of epistemic irrelevance and independence [1,3,9,12,17]. A credal network consists of a directed acyclic graph associated with random variables and “local” credal sets. Each node of the graph is associated with a random variable  $X_i$ ; we refer to variables and nodes interchangeably. If there is an edge from  $X_j$  to  $X_i$ , then  $X_j$  is a *parent* of  $X_i$ . The parents of  $X_i$  are denoted by  $\mathbf{Z}_i$ . Each variable  $X_i$  is further associated with a local credal set  $K(X_i|\mathbf{Z}_i = \mathbf{z})$  for each value  $\mathbf{z}$  of  $\mathbf{Z}_i$ . We assume that local credal sets are *separately specified*, that is,  $K(X_i|\mathbf{Z}_i = \mathbf{z}')$  and  $K(X_i|\mathbf{Z}_i = \mathbf{z}'')$  impose no constraints on each other for  $\mathbf{z}' \neq \mathbf{z}''$ .

Here we are interested in semantics for credal networks that are based on epistemic irrelevance; we thus consider two possible interpretations for a credal network [8]:

- The *extension based on epistemic irrelevance* is the largest joint credal set such that nondescendants nonparents of a random variable  $X_i$  are epistemically irrelevant to  $X_i$  given the parents of  $X_i$ .
- The *extension based on epistemic independence*, or simply *epistemic extension*, is the largest joint credal set such that nondescendants nonparents of a random variable  $X_i$  are epistemically independent of  $X_i$  given the parents of  $X_i$ .

These extensions are clearly based on different Markov conditions. Given the asymmetric character of epistemic irrelevance, it might seem that an extension based on epistemic irrelevance should be the most natural interpretation of a *directed* acyclic graph. However, such extensions are quite weak in what they represent, as shown by the next example.

**Example 8** Consider three binary variables  $X_1$ ,  $X_2$ , and  $X_3$  where variable  $X_i$  has values  $i$  and  $\hat{i}$ , following the conventions in Example 1. Consider a “chain” as in Figure 2.a. Suppose this graph is associated with assessments  $P(1) = \alpha_1$ ,  $P(2|1) = \alpha_2$ ,  $P(2|\hat{1}) = \alpha_3$ ,  $P(3|2) \in [\beta_1, \beta_2]$ ,  $P(3|\hat{2}) = \beta_3$ , where all  $\alpha_i$  and  $\beta_j$  are different. The extension based on epistemic irrelevance requires that  $EIR(X_1, X_3|X_2)$ ; it seems in fact reasonable to assume that  $X_2$  “separates”  $X_1$  and  $X_3$ . Consider then a credal set with three vertices (each vertex is a joint distribution  $P(X_1, X_2, X_3)$ ). The first distribution has  $P(3|2, 1) = P(3|2, \hat{1}) = \beta_1$ ; the second distribution has  $P(3|2, 1) = P(3|2, \hat{1}) = \beta_2$ ; and the third distribution has  $P(3|2, 1) = \beta_1$ ,  $P(3|2, \hat{1}) = \beta_2$ . We have  $P(1|2, 3) = P(1|2, \hat{3})$  for the first two distributions, while  $P(1|2, 3) \neq P(1|2, \hat{3})$  for the third one. Consequently,  $X_3$  is not irrelevant to  $X_1$  conditional on  $X_2$  for the credal set, so  $X_2$  does not really “separate”  $X_1$  and  $X_3$ .

Despite their weaknesses, extensions based on epistemic irrelevance can often be manipulated through straightforward linear programming. Suppose a credal network is given and we must compute the upper probability  $\bar{P}(Q|E)$ , where  $Q$  and  $E$  denote events defined by (possibly several)  $X_i$ . For the epistemic extension based



on irrelevance, this computation can be reduced to a linear program [8]. To understand this reduction, consider the judgement:

$$K(X_i|\mathbf{Z}_i, \mathbf{Y}_i) \cong K(X_i|\mathbf{Z}_i), \quad (4)$$

where  $\mathbf{Z}_i$  denotes the parents of  $X_i$ , and  $\mathbf{Y}_i$  denotes the nondescendants nonparents of  $X_i$  (we use this notation in the remainder of this section). Again we emphasize that  $\cong$  denotes equality of convex hulls. The right hand side of expression (4) is known, as it is part of the network definition. So we can express constraints in the epistemic extension based on irrelevance by taking the constraints over  $K(X_i|\mathbf{Z}_i = \mathbf{z}_i)$  and replicating them for all sets  $K(X_i|\mathbf{Z}_i = \mathbf{z}_i, \mathbf{Y}_i = \mathbf{y}_i)$ , for every  $(\mathbf{y}_i, \mathbf{z}_i)$ . Constraints must be expressed over  $p_k$ , the probabilities of atomic events; as the number of atomic events is exponential on the number of random variables  $X_i$ , we obtain a potentially large linear program (these constraints can be satisfied at least by the *strong extension* of the network [12], so they in fact characterize epistemic irrelevance; the proof of Lemma 9 deals with a similar issue).

Handling epistemic extensions based on independence raises more difficulties. Such extensions must satisfy constraints (4) and the “backward” judgements

$$K(\mathbf{Y}_i|\mathbf{Z}_i, X_i) \cong K(\mathbf{Y}_i|\mathbf{Z}_i). \quad (5)$$

Neither side of these constraints is directly specified by the network. This difficulty is circumvented in a “brute-force” manner by the only existing algorithm for epistemic extensions [8], which we call the E3 algorithm (for *Extensive Epistemic Extension* algorithm). This algorithm explicitly builds each set appearing on the right hand side of expression (5). This construction is exponential on the number of variables; even worse, the number of constraints grows extremely fast as it requires exponentially many projections of polyhedra (each one of which with worst-case exponential complexity). Such complexity level has prevented networks with more than four variables to be dealt with in practice. Alas, the E3 algorithm offers no clear path to approximation schemes — a frustrating situation as it seems that approximation algorithms are a necessary route to follow.

In the remainder of this section we offer a multilinear programming formulation for epistemic extensions, summarized in Figure 1. The algorithm we derive is significantly simpler to implement than the E3 algorithm, as it does not require an explicit construction of the epistemic extension.

Given a credal network, we start by creating optimization variables  $p_k$  as in Section 2: these optimization variables now represent atomic probabilities over conjunctions of values of random variables. We now ask, what are the constraints over  $p_k$  such that these optimization variables do represent a measure in the epistemic extension? Clearly we must have  $p_k \geq 0$  for all  $p_k$ , the unitary constraint  $\sum_k p_k = 1$ , and the “forward” judgements of irrelevance in Expression (4), generated by the replication technique already discussed. Steps 1 and 2 in Figure 1 generate these

MULTILINEAREXTENSION( $\mathbf{X}, \{p_k\}$ )

$\mathbf{X}$ : set of random variables  $X_i$  that constitute a network,

$\{p_k\}$ : set of optimization variables representing atomic probabilities over  $\mathbf{X}$ .

1. Generate constraints  $p_k \geq 0$  for all  $k$  and the constraint  $\sum_k p_k = 1$ .
2. For every “forward” irrelevance judgement (4), generate constraints that enforce  $P(X_i|\mathbf{Z}_i, \mathbf{Y}_i) \in K(X_i|\mathbf{Z}_i)$  for every value of  $(\mathbf{Y}_i, \mathbf{Z}_i)$ , using the constraints for  $K(X_i|\mathbf{Z}_i)$  specified in the network description.
3. For every random variable  $X_i$ , and for every possible  $x$ :
  - 3.1. Introduce optimization variables  $q_{i,x}^{\mathbf{y},\mathbf{z}}$ , and generate constraints (6), one per value of  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$ .
  - 3.2. If the top sub-network represented by  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$  has more than one node and contains irrelevance relations, then recursively call

$$\text{MULTILINEAREXTENSION}(\{\mathbf{Y}_i, \mathbf{Z}_i\}, \{q_{i,x}^{\mathbf{y},\mathbf{z}}\});$$

otherwise just impose the (linear) constraints on this sub-network over  $q_{i,x}^{\mathbf{y},\mathbf{z}}$ .

Fig. 1. The procedure MULTILINEAREXTENSION.

constraints.

Consider now a “backward” constraint (5). To satisfy it, we must guarantee that there is a measure  $P(\mathbf{Y}_i|\mathbf{Z}_i, X_i = x)$  in the joint credal set that satisfies the constraints for  $K(\mathbf{Y}_i|\mathbf{Z}_i)$ , for each possible  $x$ . We now introduce a set of optimization variables  $\{q_{i,x}^{\mathbf{y},\mathbf{z}}\}$  that represent a “fresh” measure over  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$ , for *each possible*  $x$ . The “backward” constraint (5) requires exactly that, for each measure  $P(\mathbf{Y}_i|\mathbf{Z}_i, X_i = x)$  in  $K(\mathbf{Y}_i|\mathbf{Z}_i, X_i = x)$ , we have a “marginal” measure over  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$  that is identical to  $P(\mathbf{Y}_i|\mathbf{Z}_i, X_i = x)$ . Thus we introduce a multilinear constraint for each value  $\{\mathbf{y}, \mathbf{z}\}$  of  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$ :

$$q_{i,x}^{\mathbf{y},\mathbf{z}} = P(\mathbf{Y}_i = \mathbf{y}|\mathbf{Z}_i = \mathbf{z}, X_i = x) \times \sum_{\mathbf{y}' \in \mathcal{Y}_i} q_{i,x}^{\mathbf{y}',\mathbf{z}},$$

where  $\mathcal{Y}_i$  denotes the set of values of  $\mathbf{Y}_i$ . This constraint can be written as (given positivity assumptions):

$$P(X_i = x, \mathbf{Z}_i = \mathbf{z}) \times q_{i,x}^{\mathbf{y},\mathbf{z}} = P(X_i = x, \mathbf{Y}_i = \mathbf{y}, \mathbf{Z}_i = \mathbf{z}) \times \sum_{\mathbf{y}' \in \mathcal{Y}_i} q_{i,x}^{\mathbf{y}',\mathbf{z}}. \quad (6)$$

Note that  $P(X_i = x, \mathbf{Y}_i, \mathbf{Z}_i)$  and  $P(X_i = x, \mathbf{Z}_i)$  are linear functions of the optimization variables  $p_k$ , so we have obtained a multilinear constraint on  $p_k, q_{i,x}^{\mathbf{y},\mathbf{z}}$ .

The remaining problem is to constrain the optimization variables  $q_{i,x}^{\mathbf{y},\mathbf{z}}$  so that they do represent a valid marginal measure over  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$  — that is, a measure obtained by marginalization from a joint measure  $P(\mathbf{X})$  that satisfies all assessments and judgements of irrelevance. Note that we are now dealing with two joint measures simultaneously: we have a joint measure represented by the optimization variables

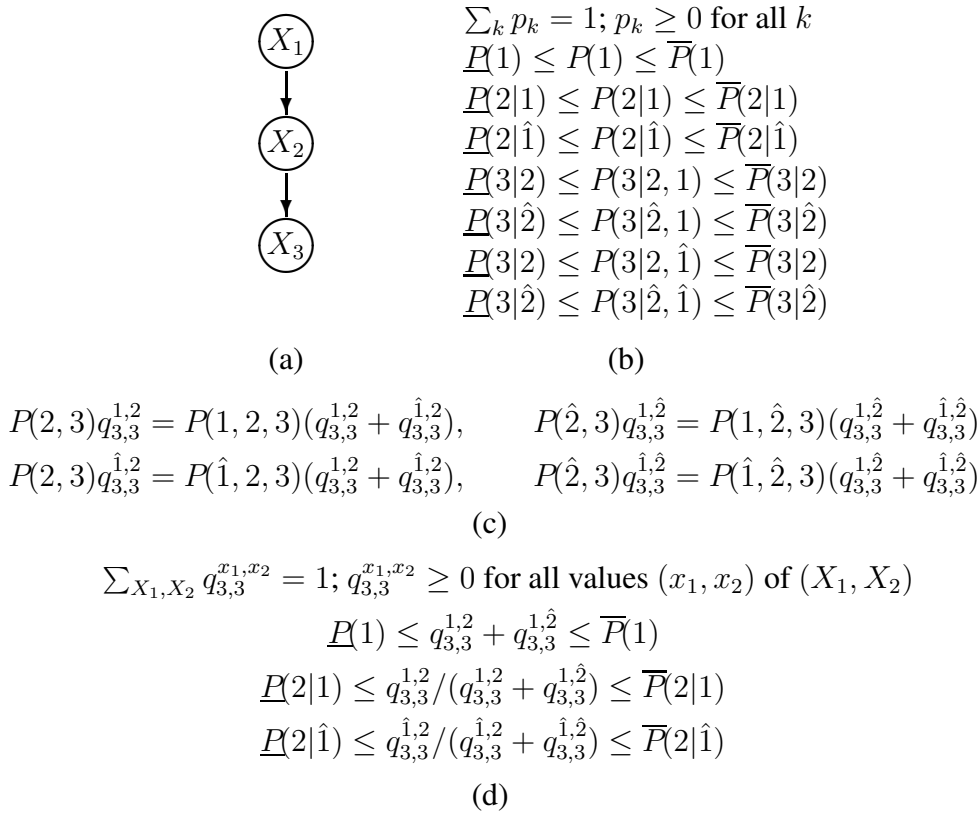


Fig. 2. Network and constraints for Example 10; values of  $P$  are combinations of the  $p_k$  (for instance,  $P(2|1) = (p_1 + p_2)/(p_1 + p_2 + p_3 + p_4)$  and  $P(2, 3) = p_1 + p_5$ ).

$p_k$ , and a joint measure that yields  $q_{i,x}^{y,z}$  by marginalization. As indicated by Step 3.2 of Figure 1, the constraints on  $q_{i,x}^{y,z}$  are obtained by recursively applying the same algorithm, now on the sub-network formed by random variables  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$ .

To show correctness of this recursive procedure, we must note that  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$  forms a *top sub-network* — that is, a sub-network such that if  $W_i$  is in the sub-network then all ascendants of  $W_i$  are in the sub-network. We now use the following result [8]: the natural extension of a top sub-network, taking into account independence relations in the top sub-network, is always equal to the marginal credal set obtained by marginalizing the complete epistemic extension. That is, if we “cut” a top sub-network out of a credal network, and compute the epistemic extension for this sub-network, we obtain the same credal set we would obtain if we started with the whole network and then marginalized the whole epistemic extension. Consequently, we can constrain the set of  $q_{i,x}^{y,z}$  to define a valid marginal measure by recursively calling the algorithm on the top sub-network formed by  $\{\mathbf{Y}_i, \mathbf{Z}_i\}$ . Clearly no recursive call is needed when a network with a single node is processed (or a network with no independence relation). As each recursive call is applied to a smaller network, the procedure must terminate.

The final step in showing correctness of the procedure is given by the following

result.

**Lemma 9** *Constraints generated by MULTILINEAREXTENSION are necessary and sufficient to define the epistemic extension.*

**PROOF.** The constraints generated by the algorithm are necessary. Consider for example the requirement that  $K(X_i|\mathbf{Z}_i, \mathbf{Y}_i)$  satisfies all constraints for  $K(X_i|\mathbf{Z}_i)$  — this guarantees that the former set cannot be larger than the latter; but it could be strictly smaller (which would violate some irrelevance relations). However, consider the strong extension of the credal network [12]; that is, the set of all distributions that factorize as  $\prod_i P(X_i|\mathbf{Z}_i)$ . In this extension we do have that  $K(X_i|\mathbf{Z}_i, \mathbf{Y}_i)$  is equal to  $K(X_i|\mathbf{Z}_i)$ . As the epistemic extension certainly contains the strong extension, we have that constraints generated by replication are in fact sufficient to guarantee (4). Now consider the “backward” constraints (5). We reason by induction on the iterations of the algorithm. For a single node, the constraints generated by the algorithm are sufficient for (5) (in fact, there is nothing to enforce). Now suppose that a top sub-network containing  $X_1$  to  $X_{i-1}$  has been created (all constraints for this sub-network are available), and a new node  $X_i$  is to be added. Consider an auxiliary extension formed by multiplying every distribution in  $K(X_i|\mathbf{Z}_i)$  by every distribution in  $K(X_1, \dots, X_{i-1})$ . This auxiliary extension does satisfy (5). And this auxiliary measure satisfies the Markov condition (for epistemic independence) on the top sub-network of  $X_1, \dots, X_i$ ; thus the epistemic extension contains this auxiliary extension, and the constraints generated by the algorithm are in fact sufficient.  $\square$

**Example 10** *Consider a Markov chain with three binary random variables  $X_1$ ,  $X_2$  and  $X_3$  (Figure 2.a). As in Example 1, random variable  $X_i$  takes values  $i$  and  $\hat{i}$ . Suppose we have separately specified sets  $K(X_1)$  (specified by  $\underline{P}(1)$  and  $\overline{P}(1)$ ),  $K(X_2|X_1)$  (specified by  $\underline{P}(2|1)$ ,  $\overline{P}(2|1)$ ,  $\underline{P}(2|\hat{1})$ ,  $\overline{P}(2|\hat{1})$ ), and  $K(X_3|X_2)$  (specified by  $\underline{P}(3|2)$ ,  $\overline{P}(3|2)$ ,  $\underline{P}(3|\hat{2})$ ,  $\overline{P}(3|\hat{2})$ ). The epistemic extension must satisfy the judgement  $EIN(X_1, X_3|X_2)$ . We have 8 variables  $p_k$  defined as in Example 1. Figure 2.b shows constraints on  $p_k$  implied directly by the local credal sets and the “forward” irrelevance judgement  $EIR(X_1, X_3|X_2)$ . To satisfy the judgement  $EIR(X_3, X_1|X_2)$ , introduce variables  $q_{3,3}^{X_1, X_2}$  and  $q_{3, \hat{3}}^{X_1, X_2}$ . Variables  $q_{3,3}^{X_1, X_2}$  are related to  $p_k$  by the multilinear constraints in Figure 2.c, and are subject to the constraints in Figure 2.d. Variables  $q_{3, \hat{3}}^{X_1, X_2}$  are subject to similar constraints (just replacing 3 by  $\hat{3}$ ).*

The previous example can be easily extended to Markov chains with  $n$  binary variables.<sup>3</sup> The number of multilinear constraints generated by the procedure at ran-

<sup>3</sup> A Markov chain with  $n$  variables/nodes has root node  $X_1$  and terminal node  $X_n$ , such that every node  $X_i$  between them has a single parent  $X_{i-1}$  and a single child  $X_{i+1}$ ;  $X_1$  has a single child  $X_2$  and  $X_n$  has a single parent  $X_{n-1}$ .

dom variable  $X_i$ , which we denote by  $T(i)$ , is recursively expressed as  $T(i) = \mathcal{O}(2^i) + 2T(i - 1)$ , thus we have  $T(i) = \mathcal{O}(i2^i)$ . The total number of multilinear constraints is of order  $\sum_{i=1}^n \mathcal{O}(i2^i)$ , and thus of order  $\mathcal{O}(n2^n)$ . The number of linear constraints follows the same pattern. Given the inherent complexity of epistemic independence, this exponential growth is not surprising in exact calculations.

Even if the MULTILINEAREXTENSION algorithm cannot deal with large networks, it does allow us move beyond the E3 algorithm. Consider a Markov chain with 5 nodes,  $X_1$  to  $X_5$ . The algorithm leads to 152 multilinear constraints, a number that can be handled by existing multilinear programming algorithms [15]. On the other hand the E3 algorithm cannot go beyond a Markov chain with 4 nodes — because the algorithm requires explicit manipulation of epistemic extensions, and the extension of a Markov chain with 4 binary nodes typically contains millions of vertices. Our experience indicates that multilinear programs with a few thousand variables can be solved with existing hardware, thus indicating that a (not too dense) network containing about 10 to 12 nodes can be processed in reasonable time. The limits of the algorithm depend on the network topology (the density of connections in the network) but also on the number of values of variables and the complexity of the local credal sets. Even though the viable networks are still small, they can serve as testing ground for approximate algorithms to be developed in the future.

More importantly, the MULTILINEAREXTENSION algorithm generates a program with a rather modular structure that is “incrementally” built in blocks; this structure can be explored by approximation techniques. For instance, consider the question: given a joint probability  $P(X_1, \dots, X_n)$ , does this measure belong to the epistemic extension of a network or not? With the E3 algorithm, the only way to answer this question is to construct the whole extension and then test for inclusion. The multilinear formulation offers a better route, as we can test whether a sequence of multilinear programs are satisfied or not. That is, we test whether the conditional and marginal distributions obtained from  $P(X_1, \dots, X_n)$  do in fact satisfy the multilinear programs that are built by the MULTILINEAREXTENSION algorithm. The existence of such an “inclusion test” may lead to algorithms that generate distributions, detect possible problems and modify them gradually — we leave this path for the future. In general, standard approximations from multilinear programming can be used [22,31], or approximations that are specific to epistemic extensions can be investigated. The E3 algorithm offers no such path.

Depending on the independence relations expressed in a network, several simplifications may be possible — as illustrated by the next example.

**Example 11** *Consider the network in Figure 3, taken from [8]. To process  $X_1$ , we must enforce the judgement  $EIN(X_1, (X_2, X_3, X_4))$ : we need 16 constraints and we must then enforce  $EIN(X_2, X_3)$  — however this second judgement can be directly enforced without any multilinear constraint, because here the “backward” constraint  $K(X_2|X_1) \cong K(X_1)$  deals with a credal set  $K(X_1)$  that is already*

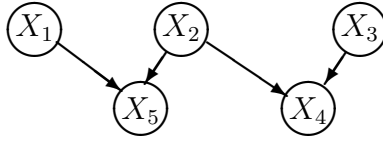


Fig. 3. Network for Example 11.

specified in the network. We must also enforce (among other things) the judgements  $EIN(X_3, (X_1, X_2, X_5))$  and  $EIN(X_3, X_5|X_1, X_2)$  — however the latter judgement is redundant as it is implied by the former (by the weak union property [13]).

## 5 Separation properties

In a Bayesian network, the computation of a conditional probability  $P(Q|E)$  typically does not require manipulation of all nodes in the network [18]. Call *evidence* the set of random variables  $X_i$  that have their values fixed by the event  $E$ . There are two kinds of nodes that can be discarded given  $Q$  and  $E$ : *barren nodes* and nodes that are separated from  $Q$  by the evidence in the *moral graph* [29].<sup>4</sup> In a Bayesian network, the value of  $P(Q|E)$  can be obtained in the sub-network without barren and separated nodes. These separation properties have been elegantly condensed into the criterion of *d-separation*, an algorithmically simple (polynomial) test that detects independence in Bayesian networks [27]. However, the proof of soundness of d-separation depends on the *semi-graphoid* properties of stochastic independence [14,18,27,30]. The problem here is that one of the semi-graphoid properties, the *contraction* property, fails for epistemic independence [13], so the proof of d-separation does not extend to epistemic irrelevance/independence.

Can separation properties of Bayesian networks be extended to epistemic extensions based on irrelevance/independence? Some results are known: barren nodes can be removed from a credal network to compute epistemic extensions based on irrelevance/independence [8]. In the next theorem we focus on separation in Markov chains — the theorem shows that evidence in a node  $X_j$  makes “upstream” nodes independent of “downstream” nodes.

**Theorem 12** *Consider a Markov chain with  $n$  nodes, with separately specified local credal sets  $K(X_1)$  and  $K(X_i|X_{i-1})$  for  $i > 1$ , such that no conditioning event has zero lower probability. For  $i < j < k$ ,  $EIN(X_i, X_k|X_j)$  in the extension based on epistemic independence.*

<sup>4</sup> A node  $X_i$  is a *barren node* if it is not used to define events  $Q$  and  $E$ , and either it has no descendants, or its descendants are also barren nodes. The *moral graph* of a Bayesian network is obtained by connecting the parents of each node and then removing the direction of all edges.

**PROOF.** Consider first  $EIR(X_k, X_i|X_j)$  and the following inductive argument. If  $k = j + 1$ , the irrelevance is trivial: the Markov condition implies the irrelevance  $EIR(X_k, (X_1, \dots, X_{j-1})|X_j)$ , and the direct decomposition property (a graphoid property [13]) can be used to remove  $X_1$  to  $X_{j-1}$ , except  $X_i$ . Now consider  $j + l$  for  $l > 1$ . Assume  $EIR(X_{j+l}, X_i|X_j)$ . The Markov condition and direct decomposition imply  $EIR(X_{j+l+1}, (X_i, X_j)|X_{j+l})$ ; then  $EIR(X_{j+l+1}, X_i|(X_{j+l}, X_j))$  by direct weak union. By reverse contraction, both judgements  $EIR(X_{j+l}, X_i|X_j)$  and  $EIR(X_{j+l+1}, X_i|(X_{j+l}, X_j))$  imply  $EIR((X_{j+l}, X_{j+l+1}), X_i|X_j)$ , and  $X_{j+l}$  can be removed by reverse decomposition. The result is obtained when  $j + l + 1 = k$ .

Now consider  $EIR(X_i, X_k|X_j)$ . Again the result is trivial for  $k = j + 1$ . We use an inductive argument: we assume  $EIR(X_i, X_{j+l}|X_j)$  and we want to show that  $EIR(X_i, (X_{j+l}, X_{j+l+1})|X_j)$ . However we cannot use contraction here [13], so we must take a different route. Take an arbitrary function  $f(X_{j+l}, X_{j+l+1})$ ; to simplify notation, we use  $r$  for  $j + l$ . We must have  $\underline{E}[f(X_r, X_{r+1})|X_i, X_j] \geq \underline{E}[\underline{E}[f(X_r, X_{r+1})|X_r]|X_j]$ ; to show that these expressions are in fact identical, we exhibit a credal set that must belong to the epistemic extension and where equality is attained. Take the following distribution that is clearly independent of  $X_i$  and  $X_j$ :

$$P(X_{r+1}|X_1, \dots, X_r) = \arg \min_{P \in K(X_{r+1}|X_r)} E_P[f(X_{r+1}, X_r)|X_r]. \quad (7)$$

Consider an auxiliary extension generated by multiplying every distribution in the epistemic extension  $K(X_1, \dots, X_r)$  by the distribution in Expression (7). The resulting extension does satisfy the Markov condition for  $X_1, \dots, X_r$  and also for  $X_{r+1}$  (because Expression (7) defines the conditional probability of  $X_{r+1}$  given  $X_1, \dots, X_r$ , and this distribution is independent of  $X_1, \dots, X_{r-1}$ ). Thus the auxiliary extension belongs to the epistemic extension, and it contains an appropriate minimizing probability distribution. In the auxiliary extension we have:

$$\underline{E}[f(X_r, X_{r+1})|X_i, X_j] = \min E[E[f(X_r, X_{r+1})|X_i, X_j, X_r]|X_i, X_j],$$

and the last expression is equal to  $\min E[\underline{E}[f(X_r, X_{r+1})|X_r]|X_i, X_j]$  by construction (7). The last expression is equal to  $\underline{E}[\underline{E}[f(X_r, X_{r+1})|X_r]|X_i, X_j]$ . By assumption  $EIR(X_i, X_r|X_j)$ , so we have that  $\underline{E}[\underline{E}[f(X_r, X_{r+1})|X_r]|X_i, X_j]$  is equal to  $\underline{E}[\underline{E}[f(X_r, X_{r+1})|X_r]|X_j]$ , and consequently

$$\underline{E}[f(X_r, X_{r+1})|X_i, X_j] = \underline{E}[\underline{E}[f(X_r, X_{r+1})|X_r]|X_j]. \quad (8)$$

Likewise,  $\underline{E}[f(X_r, X_{r+1})|X_j] = \min E[\underline{E}[f(X_r, X_{r+1})|X_r]|X_j]$ ; the last expression is equal to  $\underline{E}[\underline{E}[f(X_r, X_{r+1})|X_r]|X_j]$  and by Expression (8) we obtain

$$\underline{E}[f(X_r, X_{r+1})|X_j] = \underline{E}[f(X_r, X_{r+1})|X_i, X_j].$$

As  $f(X_r, X_{r+1})$  is arbitrary, we obtain  $EIR(X_i, (X_{j+l+1}, X_{j+l})|X_j)$  and by direct decomposition  $EIR(X_i, X_{j+l+1}|X_j)$ .  $\square$

It is possible to adapt the proof of Theorem 12 to a number of more general situations. For example, we might consider two sets of variables  $\mathbf{X}_i$  and  $\mathbf{X}_j$ , where  $\mathbf{X}_i$  “precedes”  $\mathbf{X}_j$  and both precede the variable  $X_k$  in the Markov chain — we obtain that  $\mathbf{X}_i$  and  $X_k$  are epistemically independent conditional on  $\mathbf{X}_j$ .<sup>5</sup> However it seems that a substantially new approach would be needed to prove full d-separation in more general settings. When d-separation does not obtain, some simpler (possibly asymmetric) separation property may be valid [26,32–34].<sup>6</sup>

## 6 Conclusion

Epistemic irrelevance and independence offer a “behavioral” notion of independence for credal sets. However, these concepts are difficult to manipulate computationally. On the one hand, judgements of epistemic irrelevance and independence lead to very complex joint credal sets; on the other hand, little is known about their separation properties and other simplifications that are routinely applied with stochastic independence. In this paper we have tried to increase the current understanding about epistemic irrelevance and independence.

First, we have presented multilinear programming methods that handle general judgements about events, and judgements about random variables expressed through credal networks. These techniques are more efficient than existing methods, particularly in connection with random variables, because they do not require explicit construction of extensions. The algorithms inherit convergence guarantees from multilinear programming (it is an open question whether such guarantees can be given for Walley’s algorithm and its generalizations), and they allow judgements of epistemic and strong irrelevance to be mixed in the same framework. However, it is clear that our algorithms are still not able to produce fast inferences for large credal networks. It may be that the main advantage of the multilinear programming approach is that it allows approximation methods from multilinear programming to be applied to epistemic irrelevance, something that cannot be easily done with existing methods. We leave for the future the exploration of approximation methods. In fact, we leave several avenues open for future exploration; for example, a precise characterization of computational complexity for epistemic irrelevance and independence is still open.

We have also shown in this paper that usual separation properties employed in Bayesian networks hold for Markov chains. Many important properties of stochastic independence have no known analogues for epistemic irrelevance and independence; an interesting avenue for further is exactly to find such analogues.

---

<sup>5</sup> We thank Heloisa Hanania for noticing this fact and working out the proofs.

<sup>6</sup> We thank a reviewer for indicating this possibility and the relevant references.



1. Start with  $\nu_j = 0$  and  $\mu_j = 1$  for all  $j$  ranging over the  $r$  judgements of irrelevance (optimization variables defined by Expression (2)).
2. Repeat:
  - 2.1. Form the constraints  $\mathcal{C}_0$  as in Section 3. (Note that now  $\mu_j$  and  $\nu_j$  are not optimization variables; their values change from iteration to iteration.)
  - 2.2. With the constraints in the previous step, compute for all  $j$  (using fractional linear programming):
$$\begin{aligned} \nu'_j &= \underline{P}(B_j|A_j, C_j), & \nu''_j &= \underline{P}(B_j|A_j^G, C_j), & \nu^*_j &= \underline{P}(B_j|C_j), \\ \mu'_j &= \overline{P}(B_j|A_j, C_j), & \mu''_j &= \overline{P}(B_j|A_j^G, C_j), & \mu^*_j &= \overline{P}(B_j|C_j). \end{aligned}$$
  - 2.3. If  $(\nu'_j = \nu''_j = \nu^*_j)$  and  $(\mu'_j = \mu''_j = \mu^*_j)$  for all  $j$ , stop; otherwise, take  $\nu_j = \max(\nu'_j, \nu''_j, \nu^*_j)$ ,  $\mu_j = \min(\mu'_j, \mu''_j, \mu^*_j)$  for all  $j$ , and return to step 2.1.
3. Using the constraints reached when step 2.3 breaks the loop, compute  $\max P(D|E)$  using linear fractional programming.

Fig. A.1. Walley's method for inferences with epistemic irrelevance among events.

## Acknowledgements

We thank Peter Walley for sharing with us the algorithm in Figure A.1. This work has received generous support from HP Brazil R&D. The work has also been supported by CNPq (through grant 3000183/98-4) and FAPESP (through grant 04/09568-0).

## A Walley's iterative algorithm

The iterative procedure described in Figure A.1 produces inferences for an event  $D$  conditional on another event  $E$ , under judgements of epistemic irrelevance. The method has been conceived by Walley (personal communication) and seems not to be published at the moment. We present here a *very* brief summary of Walley's algorithm, so as to compare it to our multilinear programming approach. The idea of Walley's algorithm is to start with the weakest possible bounds ( $\nu_j = 0$  and  $\mu_j = 1$ ) and then to check, at each iteration, whether irrelevance assessments are satisfied by the resulting constraints; if not, then the smallest change to assessments that can lead to satisfaction of irrelevance judgements is computed and the current constraints are modified accordingly. Each iteration modifies at least one of current assessments (or stops). The algorithm gradually converges to a set of constraints that represents the whole natural extension. Obvious changes to Walley's algorithm can account for assessments such as  $E[f_i|G_i] \geq \alpha$ .

Walley's algorithm deals only with events. So as to facilitate comparison with

our methods in Section 4, we outline here a possible strategy to deal with random variables. Consider judgements of the form  $EIR(X_j, Y_j|C)$ . We start with all assessments other than judgements of independence, as in step 2.1 of Walley’s algorithm. For each judgement  $EIR(X_j, Y_j|C)$ , we generate an explicit description of  $K(Y_j|C)$  and of  $K(Y_j|X_j = x, C)$ ; this has the same purpose of step 2.2 in Walley’s algorithm. Note that to generate an explicit description of  $K(Y_j|C)$  or  $K(Y_j|X_j = x, C)$ , we must resort either to Fourier-Motzkin elimination or to a vertex enumeration procedure [21]. If the credal sets  $K(Y_j|C)$  and  $K(Y_j|X_j = x, C)$  have the same convex hull for every value of  $X_j$ , for every  $j$ , then we stop (as in the “first half” of step 2.3). Suppose that, for a given  $j$ ,  $K(Y_j|C)$  and  $K(Y_j|X_j = x, C)$  have different convex hulls. Now we simply enforce that each one of these sets must satisfy all constraints in their current intersection (take the union of constraints defining these sets) — this is similar to the “second half” of step 2.3 in Walley’s algorithm. This procedure gradually constructs the whole natural extension. However its computational feasibility is not clear at the moment as there are several difficulties to face. First, the explicit description of sets  $K(Y_j|C)$  and  $K(Y_j|X_j = x, C)$  may lead to an exponential growth in the number of constraints; second, it is not easy to detect when sets have identical convex hulls; finally, it is not clear that this extended algorithm is always convergent, let alone finitely convergent.

## References

- [1] K. A. Andersen and J. N. Hooker. Bayesian logic. *Decision Support Systems*, 11:191–210, 1994.
- [2] V. Biazzo, A. Gilio, T. Lukasiewicz, and G. Sanfilippo. Probabilistic logic under coherence, model-theoretic probabilistic logic, and default reasoning in system P. *Journal of Applied Non-Classical Logics*, 12(2), 189-213, 2002.
- [3] J. Cano, M. Delgado, and S. Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8:253–280, 1993.
- [4] A. Capotorti, G. Coletti, and R. Scozzafava. Algorithms for processing partial probabilistic assessments. In *Third Congress Europeenne des Systemes*, pages 506–511. Kappa, 1996.
- [5] G. Coletti and R. Scozzafava. Stochastic independence in a coherent setting. *Annals of Mathematics and Artificial Intelligence*, 35:151–176, 2002.
- [6] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5:165–181, 2000.
- [7] F. G. Cozman. Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. *Journal of Computational and Graphical Statistics*, 8(4):824–838, 1999.

- [8] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.
- [9] F. G. Cozman. Separation properties of sets of probabilities. In C. Boutilier and M. Goldszmidt, editors, *XVI Conference on Uncertainty in Artificial Intelligence*, pages 107–115, San Francisco, July 2000. Morgan Kaufmann.
- [10] F. G. Cozman. Algorithms for conditioning on events of zero lower probability. In *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*, pages 248–252, Pensacola, Florida, 2002.
- [11] F. G. Cozman. Computing lower expectations with Kuznetsov’s independence condition. In J.-M. Bernard, T. Seidenfeld, M. Zaffalon, editors, *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, pages 177–187, Lugano, Switzerland, 2003. Carleton Scientific.
- [12] F. G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39(2-3):167–184, 2005.
- [13] F. G. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence*, 45:173–195, 2005.
- [14] 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.
- [15] C. P. de Campos and F. G. Cozman. Inference in credal networks using multilinear programming. In E. Onaindia and S. Staab, editors, *Proceedings of the Second Starting AI Researchers’ Symposium (STAIRS)*, pages 50–61, Amsterdam, The Netherlands, 2004. IOS Press.
- [16] L. de Campos and S. Moral. Independence concepts for convex sets of probabilities. In P. Besnard and S. Hanks, editors, *XI Conference on Uncertainty in Artificial Intelligence*, pages 108–115, San Francisco, California, 1995. Morgan Kaufmann.
- [17] E. Fagioli and M. Zaffalon. 2U: An exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107, 1998.
- [18] D. Geiger, T. Verma, and J. Pearl. Identifying independence in Bayesian networks. *Networks*, 20:507–534, 1990.
- [19] 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.
- [20] W. Gochet and Y. Smeers. A branch-and-bound method for reversed geometric programming. *Operations Research*, 27(5):983–996, September/October 1979.
- [21] P. Hansen and B. Jaumard. Probabilistic satisfiability. Technical Report G-96-31, Les Cahiers du GERAD, École Polytechnique de Montréal, 1996.
- [22] R. Horst and H. Tuy. *Global Optimization: Deterministic Approaches*. Springer-Verlag, 1995.

- [23] V. P. Kuznetsov. *Interval Statistical Methods*. Radio i Svyaz Publ., (in Russian), 1991.
- [24] C.D. Maranas and C.A. Floudas. Global optimization in generalized geometric programming. *Computers and Chemical Engineering*, 21(4):351–370, 1997.
- [25] S. Moral and A. Cano. Strong conditional independence for credal sets. *Annals of Mathematics and Artificial Intelligence*, 35:295–321, 2002.
- [26] S. Moral. Epistemic irrelevance on sets of desirable gambles. In G. de Cooman, T. L. Fine, and T. Seidenfeld, editors, *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 247–254, Netherlands, 2001. Shaker Publishing Co.
- [27] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.
- [28] H.D. Sherali and W.P. Adams. *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*. Kluwer Academic Publishers, 1999.
- [29] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search (second edition)*. MIT Press, Cambridge, Massachusetts, 2000.
- [30] M. Studeny. Semigraphoids and structures of probabilistic conditional independence. *Annals of Mathematics and Artificial Intelligence*, 21(1):71–98, 1997.
- [31] H. Tuy. *Convex Analysis and Global Optimization*, volume 22 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 1998.
- [32] B. Vantaggi. The L-separation criterion for description of cs-independence models. *International Journal of Approximate Reasoning*, 29(3):291–316, 2002.
- [33] B. Vantaggi. Conditional independence structures and graphical models. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(5):545–571, 2003.
- [34] B. Vantaggi. Graphical representation of asymmetric graphoid structures. In *Third International Symposium on Imprecise Probabilities and Their Applications*, pages 560–574. Carleton Scientific, 2003.
- [35] P. Vicig. Epistemic independence for imprecise probabilities. *International Journal of Approximate Reasoning*, 24:235–250, 2000.
- [36] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- [37] P. Walley, R. Pelesoni, P. Vicig. Direct algorithms for checking consistency and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference*, 126(1):119–151, 2004.