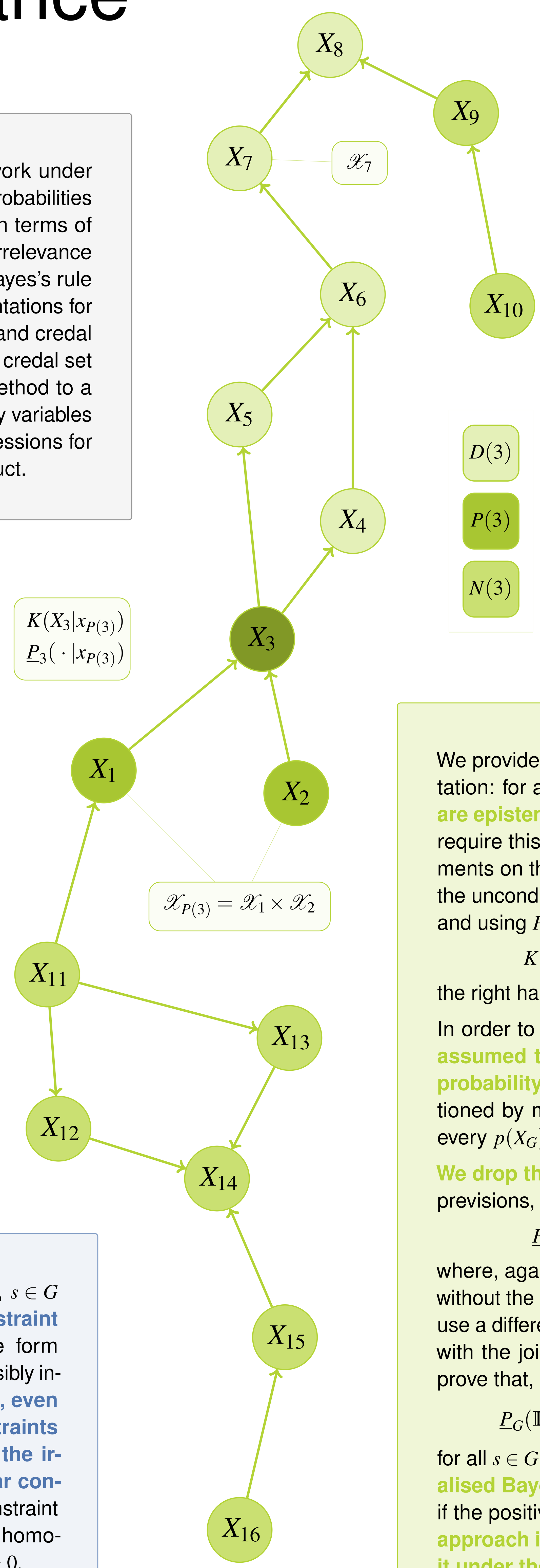
Allowing for probability zero in credal networks under epistemic irrelevance

Jasper De Bock and Gert de Cooman
{Jasper.DeBock,Gert.deCooman}@UGent.be
SYSTeMS, Ghent University, Belgium

Abstract of the paper

We generalise Cozman's (2000) concept of a credal network under epistemic irrelevance to the case where lower (and upper) probabilities are allowed to be zero. Our main definition is expressed in terms of coherent lower previsions (BMT) and imposes epistemic irrelevance by means of strong coherence rather than element-wise Bayes's rule (I, II & III). We also present a number of alternative representations for the resulting joint model, both in terms of lower previsions and credal sets, amongst which an intuitive characterisation of the joint credal set by means of linear constraints (IV). We then apply our method to a simple case: the independent natural extension for two binary variables (V). This allows us to, for the first time, find analytical expressions for the extreme points of this special type of independent product.



Basic modelling tools (BMT)

We will model a subject's beliefs about the value that a variable X assumes in some set $\mathscr{X}$ by means of two different, although mathematically equivalent, imprecise-probabilistic methods. The approach that is perhaps best known is to use a **credal set** $K(X)$, defined as a closed convex subset of $\Sigma_{\mathscr{X}}$, which is the set containing all probability mass functions on $\mathscr{X}$. The second approach is to use the associated **coherent lower prevision** $\underline{P}$ on $\mathscr{G}(\mathscr{X})$, where $\mathscr{G}(\mathscr{X})$ is the set of all gambles on $\mathscr{X}$. It is given by

$$\underline{P}(f) = \min\{P_p(f) \colon p(X) \in K(X)\} \text{ for all } f \in \mathscr{G}(\mathscr{X}),$$

where P_p is the expectation operator (prevision) for the probability mass function $p(X)$. The credal set of such a coherent lower prevision is given by

$$K(X) = \{p \in \Sigma_{\mathscr{X}} \colon (\forall f \in \mathscr{G}(\mathscr{X}))\, P_p(f) \geq \underline{P}(f)\},$$

thereby establishing the mathematical equivalence.

Local uncertainty models (I)

With every node s of a finite **directed acyclic graph (DAG)**, we associate a variable X_s taking values in some finite, non-empty set $\mathscr{X}_s$. The set of all nodes is denoted by G. For every subset $S \subseteq G$, the joint variable X_S takes values in $\mathscr{X}_S := \times_{s \in S} \mathscr{X}_s$. For every $s \in G$, we denote by $P(s)$ the set consisting of the parent nodes of s. Similar to what is done in classical Bayesian networks, **we attach local uncertainty models to the nodes of the network, conditional on the value of their parents.** For all $s \in G$ and every instantiation $x_{P(s)} \in \mathscr{X}_{P(s)}$ of $X_{P(s)}$, we require a **credal set** $K(X_s|x_{P(s)})$ or, equivalently, a **coherent lower prevision** $\underline{P}_s(\cdot|x_{P(s)})$ on $\mathscr{G}(\mathscr{X}_s)$ (BMT).

Imposing epistemic irrelevance (II)

We provide the graphical structure of the network with the following interpretation: for any node $s \in G$, its **non-parent non-descendant variables** $X_{N(s)}$ **are epistemically irrelevant** to X_s conditional on $X_{P(s)}$. (In our paper, we also require this for subsets of $N(s)$. We do not impose these additional assessments on this poster because we have recently discovered that, at least for the unconditional joint model, they are redundant.) Put more mathematically, and using $PN(s)$ as a shorthand notation for $P(s) \cup N(s)$, we require that

$$K(X_s|x_{PN(s)}) = K(X_s|x_{P(s)}) \text{ for all } s \in G \text{ and } x_{PN(s)} \in \mathscr{X}_{PN(s)}, \tag{1}$$

the right hand side of these equations being provided by the local models (I).

In order to translate this into a property of a joint model $K(X_G)$, **it is often assumed that** for every $p(X_G) \in K(X_G)$, **all events have strictly positive probability** (Cozman 2000). Under this assumption, $K(X_G)$ can be conditioned by means of **element-wise Bayes's rule** (applying Bayes's rule to every $p(X_G) \in K(X_G)$), thereby making it possible to impose Eq. (1).

We drop this positivity assumption by using an approach based on lower previsions, replacing Eq. (1) by the equivalent (BMT) statement that

$$\underline{P}_s(\cdot|x_{PN(s)}) = \underline{P}_s(\cdot|x_{P(s)}) \text{ for all } s \in G \text{ and } x_{PN(s)} \in \mathscr{X}_{PN(s)}, \tag{2}$$

where, again, the right hand side is provided by the local models (I). Since, without the positivity assumption, **conditioning is not uniquely defined**, we use a different method for making the conditional models in Eq. (2) consistent with the joint model $\underline{P}_G$: we require them to be **(strongly) coherent**. We prove that, in our particular case, this is equivalent to requiring that

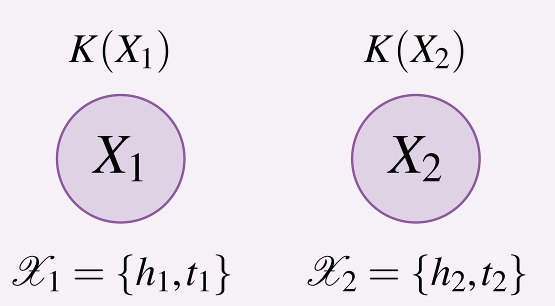
$$\underline{P}_G(\mathbb{I}_{x_{PN(s)}}[g - \underline{P}_s(g|x_{P(s)})]) = 0 \text{ and } \underline{P}_G(\mathbb{I}_{x_{P(s)}}[g - \underline{P}_s(g|x_{P(s)})]) = 0$$

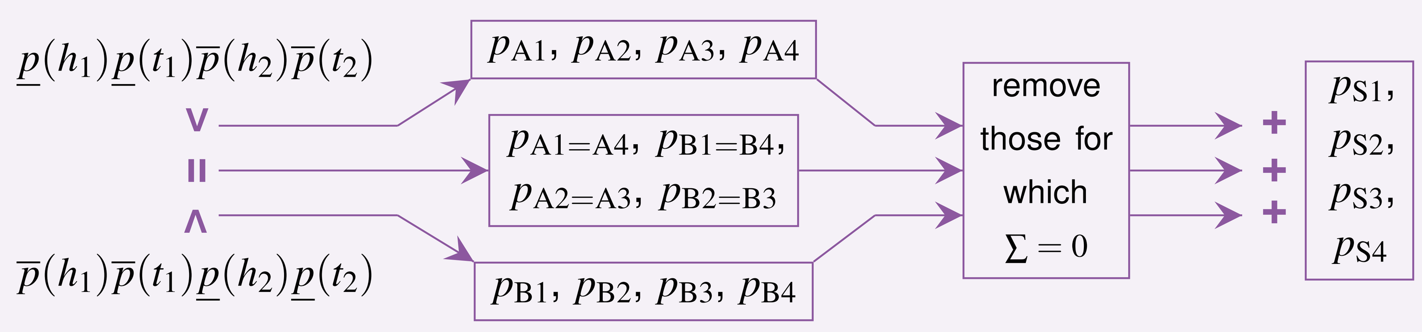
for all $s \in G$, $x_{PN(s)} \in \mathscr{X}_{PN(s)}$ and $g \in \mathscr{G}(\mathscr{X}_s)$. This formula is known as **Generalised Bayes's Rule** (GBR) and is equivalent to element-wise Bayes's rule if the positivity assumption is satisfied. It should therefore be clear that **our approach is an extension of the one by Cozman (2000), coinciding with it under the positivity assumption.**

Linear constraints (IV)

It is well known that **each local credal set** $K(X_s|x_{P(s)})$ **(I)**, $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, **is the solution set to a local unitary constraint and a set of linear homogeneous inequalities** of the form $\sum_{z_s \in \mathscr{X}_s} p(z_s|x_{P(s)})\gamma(z_s) \geq 0$, where γ takes values in some (possibly infinite, but often finite) set $\Gamma(s, x_{P(s)}) \subseteq \mathscr{G}(\mathscr{X}_s)$. **We show that, even without the positivity assumption (II), these local constraints can be used to derive an intuitive characterisation of the irrelevant natural extension** $K^{\mathrm{irr}}(X_G)$ **(III) in terms of linear constraints.** $K^{\mathrm{irr}}(X_G)$ is the solution set to the global unitary constraint and, for all $s \in G$, $x_{PN(s)} \in \mathscr{X}_{PN(s)}$ and $\gamma \in \Gamma(s, x_{P(s)})$, a linear homogeneous inequality $\sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p(x_{PN(s)}, z_s, z_{D(s)})\gamma(z_s) \geq 0$.

Independent natural extension for two binary variables (V)





For the simple credal network above, $K^{\mathrm{irr}}_{\{1,2\}}$ is the so-called **independent natural extension** of $K(X_1)$ and $K(X_2)$. Every $K(X_i)$, with $i \in \{1,2\}$, is fully determined by the lower probability $\underline{p}(h_i)$ and upper probability $\overline{p}(h_i)$ of 'heads'. The probability of 'tails' is bounded by $\underline{p}(t_i) := 1 - \overline{p}(h_i)$ and $\overline{p}(t_i) := 1 - \underline{p}(h_i)$. Using the linear constraints in (IV), we have derived **analytical expressions for the extreme points** of $K^{\mathrm{irr}}_{\{1,2\}}$. They can be found using the table and diagram to the right; see our paper for more details.

	$p(h_1,h_2)\Sigma$	$p(h_1,t_2)\Sigma$	$p(t_1,h_2)\Sigma$	$p(t_1,t_2)\Sigma$	Σ
$p_{\mathrm{S}1}$	$\underline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(t_2)$	$\overline{p}(t_1)\underline{p}(h_2)$	$\overline{p}(t_1)\overline{p}(t_2)$	1
$p_{\mathrm{S}2}$	$\underline{p}(h_1)\overline{p}(h_2)$	$\underline{p}(h_1)\underline{p}(t_2)$	$\overline{p}(t_1)\overline{p}(h_2)$	$\overline{p}(t_1)\underline{p}(t_2)$	1
$p_{\mathrm{S}3}$	$\overline{p}(h_1)\underline{p}(h_2)$	$\overline{p}(h_1)\overline{p}(t_2)$	$\underline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_2)$	1
$p_{\mathrm{S}4}$	$\overline{p}(h_1)\overline{p}(h_2)$	$\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(h_2)$	$\underline{p}(t_1)\underline{p}(t_2)$	1
$p_{\mathrm{A}1}$	$\underline{p}(h_1)\overline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(h_1)\overline{p}(t_2)$	$\overline{p}(t_1)\overline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(t_2)$	$\underline{p}(h_1)\overline{p}(t_2)+\overline{p}(h_1)\underline{p}(h_2)$
$p_{\mathrm{A}2}$	$\underline{p}(h_1)\overline{p}(h_1)\overline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)$	$\overline{p}(t_1)\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(h_1)\overline{p}(h_2)+\overline{p}(h_1)\underline{p}(t_2)$
$p_{\mathrm{A}3}$	$\overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\underline{p}(h_1)\overline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_1)\overline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_2)+\overline{p}(t_1)\underline{p}(h_2)$
$p_{\mathrm{A}4}$	$\underline{p}(t_1)\underline{p}(h_1)\overline{p}(h_2)$	$\overline{p}(h_1)\overline{p}(t_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_1)\overline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(h_2)+\overline{p}(t_1)\underline{p}(t_2)$
$p_{\mathrm{B}1}$	$\underline{p}(h_2)\overline{p}(h_2)\underline{p}(h_1)$	$\overline{p}(t_2)\overline{p}(h_2)\underline{p}(h_1)$	$\underline{p}(h_2)\overline{p}(h_2)\overline{p}(t_1)$	$\underline{p}(h_2)\underline{p}(t_2)\overline{p}(t_1)$	$\underline{p}(h_2)\overline{p}(t_1)+\overline{p}(h_2)\underline{p}(h_1)$
$p_{\mathrm{B}2}$	$\overline{p}(h_2)\overline{p}(t_2)\underline{p}(h_1)$	$\underline{p}(t_2)\overline{p}(t_2)\underline{p}(h_1)$	$\underline{p}(t_2)\underline{p}(h_2)\overline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_2)\overline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_1)+\overline{p}(t_2)\underline{p}(h_1)$
$p_{\mathrm{B}3}$	$\underline{p}(h_2)\overline{p}(h_2)\overline{p}(h_1)$	$\underline{p}(h_2)\underline{p}(t_2)\overline{p}(h_1)$	$\underline{p}(h_2)\overline{p}(h_2)\underline{p}(t_1)$	$\overline{p}(t_2)\overline{p}(h_2)\underline{p}(t_1)$	$\underline{p}(h_2)\overline{p}(h_1)+\overline{p}(h_2)\underline{p}(t_1)$
$p_{\mathrm{B}4}$	$\underline{p}(t_2)\underline{p}(h_2)\overline{p}(h_1)$	$\underline{p}(t_2)\overline{p}(t_2)\overline{p}(h_1)$	$\overline{p}(h_2)\overline{p}(t_2)\underline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_2)\underline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(h_1)+\overline{p}(t_2)\underline{p}(t_1)$

Irrelevant natural extension (III)

The properties that we impose on our network (I & II) can be satisfied by multiple coherent lower previsions $\underline{P}_G$ on $\mathscr{G}(\mathscr{X}_G)$. However, amongst them, there is a unique most conservative (pointwise smallest) one. We call it the **irrelevant natural extension** of the network and denote it by $\underline{P}^{\mathrm{irr}}_G$. We show that $\underline{P}^{\mathrm{irr}}_G$ is the pointwise smallest coherent lower prevision on $\mathscr{G}(\mathscr{X}_G)$ such that for all $s \in G$, $x_{PN(s)} \in \mathscr{X}_{PN(s)}$ and $g \in \mathscr{G}(\mathscr{X}_s)$

$$\underline{P}_G(\mathbb{I}_{x_{PN(s)}}[g - \underline{P}_s(g|x_{P(s)})]) = 0.$$

We also prove the following **simple characterisation of the corresponding credal set** $K^{\mathrm{irr}}(X_G)$: it consists of all probability mass function $p(X_G)$ on $\mathscr{X}_G$ for which for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$ there are a real number $\lambda \geq 0$ and a $p(X_s|x_{P(s)}) \in K(X_s|x_{P(s)})$ such that

$$\textstyle\sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p(x_{PN(s)}, X_s, z_{D(s)}) = \lambda\, p(X_s|x_{P(s)}),$$

where we use $D(s)$ to denote the set consisting of the descendants of the node s.

We believe that most of **the marginalisation and graphoid properties that are presented on our other poster can be translated to the current framework.** Combined with linear programming (IV), this might allow for efficient inference algorithms.