Composition Operator for Credal Sets Reconsidered

Jiřina Vejnarová

Abstract This paper is the second attempt to introduce the composition operator, already known from probability, possibility, evidence and valuation-based systems theories, also for credal sets. We try to avoid the discontinuity which was present in the original definition, but simultaneously to keep all the properties enabling us to design compositional models in a way analogous to those in the above-mentioned theories. These compositional models are aimed to be an alternative to Graphical Markov Models. Theoretical results achieved in this paper are illustrated by an example.

1 Introduction

In the second half of 1990s a new approach to efficient representation of multidimensional probability distributions was introduced with the aim to be alternative to Graphical Markov Modeling. This approach is based on a simple idea: a multidimensional distribution is *composed* from a system of low-dimensional distributions by repetitive application of a special composition operator, which is also the reason why such models are called *compositional models*.

Later, these compositional models were introduced also in possibility theory [7, 8] (here the models are parameterized by a continuous t-norm) and almost ten years ago also in evidence theory [3, 4]. In all these frameworks the original idea is kept, but there exist some slight differences among these frameworks.

In [9] we introduced a composition operator for credal sets, but due to the problem of discontinuity it needed a revision. After a thorough reconsideration we decided to present a new proposal avoiding this discontinuity. The goal of this paper is to show that the revised composition operator keeps the basic properties of its counterparts in other frameworks, and therefore it will enable us to introduce compositional models for multidimensional credal sets.

J. Vejnarová (✉)
Institute of Information Theory and Automation of the Czech Academy of Sciences,
Pod Vodárenskou věží 4, Prague, Czech Republic
e-mail: vejnar@utia.cas.cz

© Springer International Publishing Switzerland 2017
M.B. Ferraro et al. (eds.), *Soft Methods for Data Science*, Advances in Intelligent Systems and Computing 456, DOI 10.1007/978-3-319-42972-4_60

This contribution is organized as follows. In Sect. 2 we summarise the basic concepts and notation. The new definition of the operator of composition is presented in Sect. 3, which is devoted also to its basic properties and an illustrative example.

2 Basic Concepts and Notation

In this section we will briefly recall basic concepts and notation necessary for understanding the contribution.

2.1 *Variables and Distributions*

For an index set $N = \{1, 2, \ldots, n\}$ let $\{X_i\}_{i \in N}$ be a system of variables, each X_i having its values in a finite set $\mathbf{X}_i$ and $\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n$ be the Cartesian product of these sets.

In this paper we will deal with groups of variables on its subspaces. Let us note that X_K will denote a group of variables $\{X_i\}_{i \in K}$ with values in $\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i$ throughout the paper.

Any group of variables X_K can be described by a *probability distribution* (sometimes also called *probability function*)

$$P : \mathbf{X}_K \longrightarrow [0, 1],$$

such that

$$\sum_{x_K \in \mathbf{X}_K} P(x_K) = 1.$$

Having two probability distributions P_1 and P_2 of X_K we say that P_1 is *absolutely continuous* with respect to P_2 (and denote $P_1 \ll P_2$) if for any $x_K \in \mathbf{X}_K$

$$P_2(x_K) = 0 \Longrightarrow P_1(x_K) = 0.$$

This concept plays an important role in the definition of the composition operator.

2.2 *Credal Sets*

A *credal set* $\mathcal{M}(X_K)$ describing a group of variables X_K is usually defined as a closed convex set of probability measures describing the values of this variable. In order to simplify the expression of operations with credal sets, it is often considered [5] that a credal set is the set of probability distributions associated to the probability measures in it. Under such consideration a credal set can be expressed as a *convex hull* (denoted by CH) of its extreme distributions (ext)

$$\mathcal{M}(X_K) = \mathrm{CH}\{ext(\mathcal{M}(X_K))\}.$$

Consider a credal $\mathcal{M}(X_K)$. For each $L \subset K$ its *marginal credal set* $\mathcal{M}(X_L)$ is obtained by element-wise marginalization, i.e.

$$\mathcal{M}(X_L) = \mathrm{CH}\{P^{\downarrow L} : P \in \mathrm{ext}(\mathcal{M}(X_K))\}, \quad (1)$$

where $P^{\downarrow L}$ denotes the marginal distribution of P on $\mathbf{X}_L$.

Besides marginalization we will also need the opposite operation, called vacuous extension. *Vacuous extension* of a credal set $\mathcal{M}(X_L)$ describing X_L to a credal set $\mathcal{M}(X_K) = \mathcal{M}(X_L)^{\uparrow K}$ $(L \subset K)$ is the maximal credal set describing X_K such that $\mathcal{M}(X_K)^{\downarrow L} = \mathcal{M}(X_L)$.

Having two credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ describing X_K and X_L, respectively (assuming that $K, L \subseteq N$), we say that these credal sets are *projective* if their marginals describing common variables coincide, i.e. if

$$\mathcal{M}_1(X_{K\cap L}) = \mathcal{M}_2(X_{K\cap L}).$$

Let us note that if K and L are disjoint, then $\mathcal{M}_1$ and $\mathcal{M}_2$ are always projective, as $\mathcal{M}_1(X_\emptyset) = \mathcal{M}_2(X_\emptyset) \equiv 1$.

2.3 Strong Independence

Among the numerous definitions of independence for credal sets [1] we have chosen strong independence, as it seems to be the most appropriate for multidimensional models.

We say that (groups of) variables X_K and X_L (K and L disjoint) are *strongly independent* with respect to $\mathcal{M}(X_{K\cup L})$ iff (in terms of probability distributions)

$$\mathcal{M}(X_{K\cup L}) = \mathrm{CH}\{P_1 \cdot P_2 : P_1 \in \mathcal{M}(X_K), P_2 \in \mathcal{M}(X_L)\}.$$

Again, there exist several generalizations of this notion to conditional independence, see e.g. [5], but as the following definition is suggested by the authors as the most appropriate for the marginal problem, it seems to be a suitable concept also in our case, since the operator of composition can also be used as a tool for solution of a marginal problem, as shown (in the framework of possibility theory) e.g. in [8].

Given three groups of variables X_K, X_L and X_M (K, L, M be mutually disjoint subsets of N, such that K and L are nonempty), we say that X_K and X_L are *conditionally independent* given X_M under global set $\mathcal{M}(X_{K\cup L\cup M})$ (to simplify the notation we will denote this relationship by $K \perp\!\!\!\perp L|M$) iff

$$\mathcal{M}(X_{K\cup L\cup M}) \\ = \mathrm{CH}\{(P_1 \cdot P_2)/P_1^{\downarrow M} : P_1 \in \mathcal{M}(X_{K\cup M}), P_2 \in \mathcal{M}(X_{L\cup M}), P_1^{\downarrow M} = P_2^{\downarrow M}\}.$$

This definition is a generalisation of stochastic conditional independence: if $\mathcal{M}(X_{K\cup L\cup M})$ is a singleton, then $\mathcal{M}(X_{K\cup M})$ and $\mathcal{M}(X_{L\cup M})$ are also (projective) singletons and the definition reduces to the definition of stochastic conditional independence.

3 Composition Operator

In this section we will introduce a new definition of composition operator for credal sets. The concept of the composition operator is presented first in a precise probability framework, as it seems to be useful for better understanding to the concept.

3.1 *Composition Operator of Probability Distributions*

Now, let us recall the definition of composition of two probability distributions [2]. Consider two index sets $K, L \subset N$. We do not put any restrictions on K and L; they may be but need not be disjoint, and one may be a subset of the other. Let P_1 and P_2 be two probability distributions of (groups of) variables X_K and X_L; then

$$(P_1 \triangleright P_2)(X_{K\cup L}) = \frac{P_1(X_K) \cdot P_2(X_L)}{P_2(X_{K\cap L})}, \tag{2}$$

whenever $P_1(X_{K\cap L}) \ll P_2(X_{K\cap L})$; otherwise, it remains undefined.

It is specific property of composition operator for probability distributions—in other settings the operator is always defined [3, 8].

3.2 *Definition and Example*

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be credal sets describing X_K and X_L, respectively. Our goal is to define a new credal set, denoted by $\mathcal{M}_1 \triangleright \mathcal{M}_2$, which will be describing $X_{K\cup L}$ and will contain all of the information contained in $\mathcal{M}_1$ and, as much as possible, in $\mathcal{M}_2$.

The required properties are met by Definition 1 in [9].[1] However, that definition exhibits a kind of discontinuity and was thoroughly reconsidered. Here we decided to propose the following one.

[1]Let us note that the definition is based on Moral's concept of conditional independence with relaxing convexity.

Definition 1 For two credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ describing X_K and X_L, their *composition* $\mathcal{M}_1 \triangleright \mathcal{M}_2$ is defined as a convex hull of probability distributions P obtained as follows. For each couple of distributions $P_1 \in \mathcal{M}_1(X_K)$ and $P_2 \in \mathcal{M}_2(X_L)$ such that $P_2^{\downarrow K\cap L} \in argmin\{Q_2 \in \mathcal{M}_2(X_{K\cap L}) : d(Q_2, P_1^{\downarrow K\cap L})$, distribution P is obtained by one of the following rules:

[a] if $P_1^{\downarrow K\cap L} \ll P_2^{\downarrow K\cap L}$

$$P = \frac{P_1 \cdot P_2}{P_2^{\downarrow K\cap L}},$$

[b] otherwise

$$P \in \text{ext}\{P_1^{\uparrow K\cup L}\}.$$

Function d used in the definition is a suitable distance function (e.g. Kullback-Leibler divergence, total variation or some other f-divergence [6]).

Let us note, that this definition of composition operator does not differ from the original one [9] in case of projective credal sets, as in this case the only distributions in $\mathcal{M}_1 \triangleright \mathcal{M}_2$ are those satisfying $P = (P_1 \cdot P_2)/P_2^{\downarrow K\cap L}$, where $P_1^{\downarrow K\cap L} = P_2^{\downarrow K\cap L}$. However, it differs in the remaining cases. Let us illustrate the application of the operator in case [a] by an example.

Example 1 Let

$$\mathcal{M}_1(X_1X_2) = \text{CH}\{[0.2, 0.8, 0, 0], [0.1, 0.4, 0.1, 0.4], [0.25, 0.25, 0.25, 0.25], [0, 0, 0.5, 0.5]\},$$

and

$$\mathcal{M}_2(X_2X_3) = \text{CH}\{[0, 0.3, 0, 0.7], [0.2, 0.1, 0.4, 0.3], [0.5, 0, 0.5, 0], [0.2, 0.3, 0.2, 0.3]\},$$

be two credal sets describing binary variables X_1X_2 and X_2X_3, respectively. These two credal sets are not projective, as $\mathcal{M}_1(X_2) = \text{CH}\{[0.2, 0.8], [0.5, 0.5]\}$, while $\mathcal{M}_2(X_2) = \text{CH}\{[0.3, 0.7], [0.5, 0.5]\}$. Therefore $\mathcal{M}_2(X_2) \subset \mathcal{M}_1(X_2)$. Definition 1 in this case leads (using total variation) to

$$\begin{aligned}
&(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_1X_2X_3)\\
&= \text{CH}\{[0, 0.3, 0, 0.7, 0, 0, 0, 0], [0.2, 0.1, 0.4, 0.3, 0, 0, 0, 0],\\
&\quad [0, 0.1, 0, 0.3, 0, 0.2, 0, 0.4], [0.07, 0.03, 0.17, 0.13, 0.13, 0.07, 0.23, 0.17],\\
&\quad [0.25, 0, 0.25, 0, 0.25, 0, 0.25, 0], [0.1, 0.15, 0.1, 0.15, 0.1, 0.15, 0.1, 0.15],\\
&\quad [0, 0, 0, 0, 0.5, 0, 0.5, 0], [0, 0, 0, 0, 0.2, 0.3, 0.2, 0.3]\\
&\quad [0, 0.2, 0, 0.8, 0, 0, 0, 0], [0.13, 0.07, 0.46, 0.34, 0, 0, 0, 0],\\
&\quad [0, 0.1, 0, 0.4, 0, 0.1, 0, 0.4], [0.07, 0.03, 0.23, 0.17, 0.07, 0.03, 0.23, 0.17]\}.
\end{aligned}$$

On the other hand

$$\begin{aligned}
&(\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_1X_2X_3) \\
&= \mathrm{CH}\{[0, 0.3, 0, 0.7, 0, 0, 0, 0], [0.2, 0.1, 0.4, 0.3, 0, 0, 0, 0], \\
&\quad [0, 0.1, 0, 0.3, 0, 0.2, 0, 0.4], [0.07, 0.03, 0.17, 0.13, 0.13, 0.07, 0.23, 0.17], \\
&\quad [0.25, 0, 0.25, 0, 0.25, 0, 0.25, 0], [0.1, 0.15, 0.1, 0.15, 0.1, 0.15, 0.1, 0.15], \\
&\quad [0, 0, 0, 0, 0.5, 0, 0.5, 0], [0, 0, 0, 0, 0.2, 0.3, 0.2, 0.3]\},
\end{aligned}$$

which differs from $(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_1X_2X_3)$. ◇

This difference deserves an explanation. $\mathcal{M}_2 \triangleright \mathcal{M}_1$ is smaller (more precise) than $\mathcal{M}_1 \triangleright \mathcal{M}_2$, which corresponds to the idea that we want $\mathcal{M}_2 \triangleright \mathcal{M}_1$ to keep all the information contained in $\mathcal{M}_2$. Therefore, we do not consider those distributions from $\mathcal{M}_1$ not corresponding to any from $\mathcal{M}_2$, although these distributions are taken into account when composing $\mathcal{M}_1 \triangleright \mathcal{M}_2$.

This is an example of a typical property of the operator of composition—it is not commutative. The next subsection is devoted to other basic properties.

3.3 Basic Properties

In the following lemma we prove that this composition operator possesses basic properties required above.

Lemma 1 *For two credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ describing X_K and X_L, respectively, the following properties hold true:*

1. *$\mathcal{M}_1 \triangleright \mathcal{M}_2$ is a credal set describing $X_{K\cup L}$.*
2. *$(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_K) = \mathcal{M}_1(X_K)$.*
3. *$\mathcal{M}_1 \triangleright \mathcal{M}_2 = \mathcal{M}_2 \triangleright \mathcal{M}_1$ iff $\mathcal{M}_1(X_{K\cap L}) = \mathcal{M}_2(X_{K\cap L})$.*

Proof 1. To prove that $\mathcal{M}_1 \triangleright \mathcal{M}_2$ is a credal set describing $X_{K\cup L}$ it is enough to take into consideration that it is the convex hull of probability distributions on $\mathbf{X}_{K\cup L}$, which is obvious from both [a] and [b] of Definition 1.

2. As marginalization of a credal set is element-wise, it is enough to prove that for any $P \in (\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L})$, $P^{\downarrow K} = P_1 \in \mathcal{M}_1(X_K)$ holds. But it immediately follows in case [a] from the results obtained for precise probabilities (see e.g. [2]). In case [b] it is obvious, as any P belongs to a vacuous extension of $P_1 \in \mathcal{M}_1(X_K)$ to $X_{K\cup L}$.
3. First, let us assume that

$$(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L}) = (\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_{K\cup L}).$$

Then also its marginals must be identical, particularly

$$(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cap L}) = (\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_{K\cap L}).$$

Taking into account 2. of this lemma we have

$$\begin{aligned}(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cap L}) &= \left(((\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L}))^{\downarrow K}\right)^{\downarrow K\cap L} \\ &= ((\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_K))^{\downarrow K\cap L} \\ &= (\mathcal{M}_1(X_K))^{\downarrow K\cap L} = \mathcal{M}_1(X_{K\cap L})\end{aligned}$$

and similarly

$$(\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_{K\cap L}) = \mathcal{M}_2(X_{K\cap L}),$$

from which the desired equality immediately follows.

Let, on the other hand, $\mathcal{M}_1(X_{K\cap L}) = \mathcal{M}_2(X_{K\cap L})$. In this case only [a] of Definition 1 is applied and for any distribution P of $(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L})$ there exist $P_1 \in \mathcal{M}_1(X_K)$ and $P_2 \in \mathcal{M}_2(X_L)$ such that $P_1^{\downarrow K\cap L} = P_2^{\downarrow K\cap L}$ and $P = (P_1 \cdot P_2)/P_2^{\downarrow K\cap L}$. But simultaneously (due to projectivity) $P = (P_1 \cdot P_2)/P_1^{\downarrow K\cap L}$, which is an element of $(\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_{K\cup L})$. Hence

$$(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L}) = (\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_{K\cup L}),$$

as desired. □

The following theorem, proven in [9], expresses the relationship between strong independence and the operator of composition. It is, together with Lemma 1, the most important assertion enabling us to introduce multidimensional models.

Theorem 1 *Let $\mathcal{M}$ be a credal set describing $X_{K\cup L}$ with marginals $\mathcal{M}(X_K)$ and $\mathcal{M}(X_L)$. Then*

$$\mathcal{M}(X_{K\cup L}) = (\mathcal{M}^{\downarrow K} \triangleright \mathcal{M}^{\downarrow L})(X_{K\cup L})$$

iff

$$(K \setminus L) \perp\!\!\!\perp (L \setminus K) | (K \cap L).$$

This theorem remains valid also for this, revised definition, as $\mathcal{M}(X_K)$ and $\mathcal{M}(X_L)$ are marginals of $\mathcal{M}(X_{K\cup L})$, and therefore only [a] (for projective distributions) is applicable.

4 Conclusions

We presented revised version of composition operator for credal sets. This definition seems to be satisfactory from the theoretical point of view; it satisfies the basic required properties and, in contrary to the original one, it avoids discontinuity.

It seems to be a reasonable tool for construction of compositional multidimensional models. Nevertheless, many problems should be solved in the near future. From the theoretical point of view it is the relationship to probabilistic and evidential compositions operators. From the practical viewpoint it is the problem of effective finding of the nearest probability distributions (if there is no projective).

Acknowledgments The support of Grant GAČR 16-12010S is gratefully acknowledged.

References

1. Couso I, Moral S, Walley P (1999) Examples of independence for imprecise probabilities. In: De Cooman G, Cozman FG, Moral S, Walley P (eds) Proceedings of ISIPTA'99. Ghent, pp 121–130
2. Jiroušek R (1997) Composition of probability measures on finite spaces. In: Geiger D, Shenoy PP (eds) Proceedings of UAI'97. Morgan Kaufmann Publishers, San Francisco, California, pp 274–281
3. Jiroušek R, Vejnarová J, Daniel M (2007) Compositional models for belief functions. In: De Cooman G, Vejnarová J, Zaffalon M (eds) Proceedings of ISIPTA'07. Mat-fyz Press, Praha, pp 243–252
4. Jiroušek R, Vejnarová J (2011) Compositional models and conditional independence in evidence theory. Int J Approx Reason 52:316–334
5. Moral S, Cano A (2002) Strong conditional independence for credal sets. Ann Math Artif Intell 35:295–321
6. Vajda I (1989) Theory of statistical inference and information. Kluwer Academic Publishers, Dordrecht
7. Vejnarová J (1998) Composition of possibility measures on finite spaces: preliminary results. In: Bouchon-Meunier B, Yager RR (eds) Proceedings of IPMU'98. Editions E.D.K. Paris, pp 25–30
8. Vejnarová J (2007) On possibilistic marginal problem. Kybernetika 43(5):657–674
9. Vejnarová J (2013) Operator of composition for credal sets. In: Cozman FG, Denoeux T, Destercke S, Seidenfeld T (eds) Proceedings of ISIPTA'13, pp 355–364