# COMBINATION AND COMPOSITION IN PROBABILISTIC MODELS

Radim Jiroušek and Prakash P. Shenoy

**Abstract** In probability theory, as well as in other alternative uncertainty theories, the existence of efficient processes for the multidimensional model construction is a basic assumption making the application of the respective theory to practical problems possible. Most of the approaches are based on the idea that a multidimensional model is set up from a great number of smaller parts representing pieces of local knowledge. Such a process is called *knowledge integration.* In the probabilistic framework, it means that a multidimensional probability distribution is aggregated from a number of low-dimensional (possibly conditional) ones.

Historically, two different operators of aggregation were designed for this purpose: the operator of *combination*, and the operator of *composition*. This paper, using the simplest possible framework of discrete probability theory, answers some natural questions like: *What is the difference between these operators? Is there a need for both of them? Are there situations when they can be mutually interchanged?*

*Keywords:* Discrete probability, aggregation of distributions, factorization, algebraic properties, idempotency.

## 1 Introduction

Broad application of probability theory in artificial intelligence that took place in the last decades of the last century was facilitated by the development of new tools and models that were incorporated into the basic theoretical gear of artificial intelli-

Radim Jiroušek
University of Economics, Jindřichův Hradec & Inst. of Information theory and Automation, Czech Academy of Sciences, Prague, Czech Republic
e-mail: radim@utia.cas.cz

Prakash P. Shenoy
University of Kansas, Lawrence, KS, USA
e-mail: pshenoy@ku.edu

gence. There are many of them, though not all of them are as famous as the Bayesian networks [7], or more generally, graphical Markov models [13]. Some of them were developed not only in the theoretical framework of probability theory but also in the framework of other uncertainty theories like possibility [6] or belief functions theories [4, 17]. This holds true also for two operators of aggregation that belong among the concepts of several uncertainty theories. The goal of this paper is to make clear the difference between these two operators of aggregation and to show that both of them have their indisputable role for uncertain knowledge modeling. Namely, they are widely used in the process of knowledge integration, the process aiming at the construction of big knowledge bases of intelligent systems. To make the exposition as clear as possible we restrict our consideration only to the best-known probability theory, we will study aggregation of (discrete) probability distributions (measures).

The basic idea of the knowledge integration process copies a human-like behavior. Nobody is able to express/comprehend knowledge that is too complex. Therefore, it should be formulated in small pieces of local knowledge, and the pieces of local knowledge are then aggregated to form a complete knowledge of the area of interest. Analogously, a probabilistic model of a knowledge base should be integrated from a great number of pieces of local knowledge, which are represented by small dimensional probability distributions. This way of knowledge base representation has also an additional advantage. For such models, there exist efficient computational procedures that can be applied for making inferences [14, 15, 16, 1, 11].

In probability theory there are many ways and purposes why two or more probability distributions are aggregated; see, e.g. "Aggregating Probability Distributions" by Clemen and Winkler [3]. In this paper we restrict our attention only to two ways that can be both considered as an aggregation of knowledge in AI applications: *Combination* and *Composition*.

**Combination.**

The purpose of the combination of probability distributions can hardly be described better than it was done by Arthur Dempster in [4]: *A probability measure may be regarded as defining degrees of belief which quantify a state of partial knowledge. ... A mechanism for combining such sources of information is a virtual necessity for a theory of probability oriented to statistical inference. The mechanism adopted here assumes* independence *of the sources, a concept whose real world meaning is not so easily described as its mathematical definition.* So, by the operation of combination we understand in this paper a proper way to combine *independent* sources of information. In agreement with the Dempster's words, with the stress on the notion of independence.

As an example, consider a situation when data files are the main source of information. Let two data files describe patients from hospitals A and B, respectively. Then, a natural way of combining these two sources of information is to join the records into one file. The respective estimation of the probability distribution $\pi$ corresponding to the joint data file can be got as a weighted sum of the estimations

of probability distributions $\pi_A$ and $\pi_B$ corresponding to the data files from hospitals A and B, respectively. So, in spite of the fact that in the described situation we do not have any objections against the employment of distribution $\pi$ received as a weighted sum distributions $\pi_A$ and $\pi_B$ (representing two sources of information), we should not consider the weighted sum of distributions a *combination* operation for probability distributions. This is because the computation of a weighted sum is not appropriate when the sources are independent. Naturally, data collected in different hospitals cannot be considered independent. They are samples from different populations, or from two disjoint parts of a population.

**Composition.**

As the term suggests, the operation of the composition is an inverse operation to decomposition. By decomposition, we understand the result of a process that, with the goal of simplification, divides an original object into its sub-objects. Thus, for example, a problem is decomposed into two (or more) simpler sub-problems. General properties of such decomposition can be viewed on the example familiar to everybody: decomposition of a positive integer into prime numbers. In this case, an elementary decomposition is a decomposition of an integer into two factors, the product of which gives the original integer. For this example, we see that

- the result of decomposition are two objects of the same type as the decomposed object – an integer is decomposed into two integers;
- both these subobjects are simpler (smaller) than the original object – both factors are smaller that the original integer, we do not consider $1 \times n$ to be a decomposition of $n$;
- not all objects can be decomposed – prime numbers cannot be decomposed;
- there exists an inverse operation (we will call it a composition) yielding the original object from its decomposed parts – the composition of two integers is their product.

It can easily be deduced from the above-presented properties that the process of a repeatedly performed decomposition of an arbitrary (finite) object into elementary subobjects that cannot be further decomposed is always finite.

As another example, let us note that a decomposition is studied also in graph theory. A simple[1] graph $G = (V,E)$ is decomposed into two simple graphs $G_1 = (V_1,E_1)$ and $G_2 = (V_2,E_2)$ if

- $V_1 \cup V_2 = V$, $V_1 \neq V \neq V_2$,
- both $G_1$ and $G_2$ are induced subgraphs of $G$ (i.e., $E_j = \{(u-v) \in E \ : \ \{u,v\} \subseteq V_j\}$),
- $E_1 \cup E_2 = E$.

Note that the graphs that cannot be decomposed are called *prime-graphs*.

[1] An undirected graph containing no loops and no multiple edges.

What is a decomposition of a finite probability distribution? Consider a two-dimensional distribution $\pi(X,Y)$. Simpler sub-objects are just one-dimensional distribution $\pi(X)$ and $\pi(Y)$. Generally, the process of marginalization is unique, but, with the exception of a degenerate distribution, we cannot unambiguously reconstruct the original two-dimensional distribution from its one-dimensional marginals. To bypass this fact, we restrict the decomposition of two-dimensional distributions $\pi(X,Y)$ into their one-dimensional marginals only for the case of independence (to denote that variables $X$ and $Y$ are independent for distribution $\pi$ we use symbol $X \perp\!\!\!\perp Y[\pi]$ – for a precise definition see the next section). In this case, $\pi(X,Y)$ can easily be reconstructed from its marginals $\pi(X)$ and $\pi(Y)$: $\pi(X,Y) = \pi(X) \cdot \pi(Y)$, where "$\cdot$" denotes pointwise multiplication, i.e., $\pi(X,Y)(x,y) = \pi(X)(x)\,\pi(Y)(y)$ for all values $x$ of $X$ and $y$ of $Y$.

Analogously, three-dimensional distribution $\pi(X,Y,Z)$ can be decomposed into two simpler probability distributions (marginals of $\pi(X,Y,Z)$) only if either a couple of variables (say $X,Y$) is independent of the remaining third variable (in this case $Z$), or, if two variables (say $X$ and $Z$) are conditionally independent given the remaining third variable (in this case $Y$):

- $\{X,Y\} \perp\!\!\!\perp Z[\pi]$, then $\pi(X,Y,Z)$ can be reconstructed from $\pi(X,Y)$ and $\pi(Z)$,
- $X \perp\!\!\!\perp Z|Y[\pi]$, then $\pi(X,Y,Z)$ can be reconstructed from $\pi(X,Y)$ and $\pi(Y,Z)$.

Thus, the composition considered in this paper will be an inverse operation to the following general operation of decomposition: *Probability distribution* $\pi(X_1,\ldots,X_n)$ *can be decomposed into* $\kappa(\{X_i\}_{i\in K})$ *and* $\lambda(\{X_i\}_{i\in L})$ *if*

1. $K \cup L = \{1,2,\ldots,n\}$;
2. $K \neq \{1,2,\ldots,n\}$, $L \neq \{1,2,\ldots,n\}$;
3. $\pi(\{X_i\}_{i\in K\cup L}) \cdot \lambda(\{X_i\}_{i\in K\cap L}) = \kappa(\{X_i\}_{i\in K}) \cdot \lambda(\{X_i\}_{i\in L})$.

Notice that in this case the original distribution $\pi(X_1,X_2,\ldots,X_n)$ can be uniquely reconstructed from distributions $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$.

The formal definitions of both combination and composition operators as well as the notation used in the paper form the content of the next section. The main part of the paper is Section 3 where we show what are the common properties of the studied operators (Section 3.1) and in what way they differ from each other (Section 3.2). The last section concludes the paper referring to the relation of the presented results with other uncertainty theories.

## 2 Basic Definitions and Notation

In this text we deal with finite-valued random variables denoted by upper case characters of Latin alphabet $X,Y,Z$, with possible indices. The respective finite (nonempty) sets of values of variables $X,Y,Z$ will be denoted by $\mathbb{X},\mathbb{Y},\mathbb{Z}$, respectively. Therefore, the values of $(X,Y)$ are from the Cartesian product $\mathbb{X}\times\mathbb{Y}$. In the case of a subset of variables $\{X_i\}_{i\in K} \subset \{X_1,X_2,\ldots,X_n\}$, we will use a simplified

notation:

$$\mathbb{X}_K = \times_{i\in K}\mathbb{X}_i$$

to denote the set of values of $\{X_i\}_{i\in K}$.

Distributions of subsets of variables will be denoted by lower-case Greek alphabets $\pi, \kappa, \lambda, \mu, \delta$ (again with possible indices). Thus, $\pi(X_1, X_2, \ldots, X_n)$ denotes an $n$-dimensional probability distribution. It's *marginal distribution* for $K \subseteq \{1, \ldots, n\}$ will be denoted $\pi(\{X_i\}_{i\in K})$, or, more often simply $\pi^{\downarrow K}$. Analogously, for $x \in \mathbb{X}_{\{1,\ldots,n\}}$, $x^{\downarrow K}$ denote the *projection* of $x$ into $\mathbb{X}_K$. When considering marginal distributions we do not exclude situations when $K = \emptyset$. In this case, we assume that $\mathbb{X}_\emptyset = \{\blacklozenge\}$, and naturally, $\pi^{\downarrow\emptyset}(\blacklozenge) = 1$.

In what follows we will also need a symbol for conditional probability distribution. For disjoint $L, M \subseteq K$, $\pi^{\downarrow L|M}$ denote the conditional probability distribution of variables $\{X_i\}_{i\in L}$ given variables $\{X_i\}_{i\in M}$, i.e., if the marginal $\pi^{\downarrow M}$ is positive then

$$\pi^{\downarrow L|M} = \frac{\pi^{\downarrow L\cup M}}{\pi^{\downarrow M}}.$$

In a general case, for each $x \in \mathbb{X}_M$, $\pi^{\downarrow L|M}(\{X_i\}_{i\in L}|x)$ is a probability distribution of variables $\{X_i\}_{i\in L}$ such that

$$\pi^{\downarrow L\cup M}(\{X_i\}_{i\in L}, \{X_i\}_{i\in M} = x) = \pi^{\downarrow L|M}(\{X_i\}_{i\in L}|x) \cdot \pi^{\downarrow M}(x).$$

Thus, the conditional probability distribution $\pi^{\downarrow L|M}$ is always defined, though sometimes ambiguously (in case that $\pi^{\downarrow M}(x) = 0$ for some $x \in \mathbb{X}_M$).

In Section 1, we used the symbol $\perp\!\!\!\perp$ to denote the independence of variables. Let us, now, introduce it more formally. Consider a distribution $\pi(\{X_i\}_{i\in N})$, and three disjoint subset $K, L, M \subset N$, $K \neq \emptyset$, $L \neq \emptyset$. We say that for distribution $\pi$ variables $\{X_i\}_{i\in K}$ and $\{X_i\}_{i\in L}$ are *conditionally independent given variables* $\{X_i\}_{i\in M}$, if for all $x \in \mathbb{X}_{K\cup L\cup M}$

$$\pi^{\downarrow K\cup L\cup M}(x) \cdot \pi^{\downarrow M}(x^{\downarrow M}) = \pi^{\downarrow K\cup M}(x^{\downarrow K\cup M}) \cdot \pi^{\downarrow L\cup M}(x^{\downarrow L\cup M}).$$

This independence will be denoted $\{X_i\}_{i\in K} \perp\!\!\!\perp \{X_i\}_{i\in L}|\{X_i\}_{i\in M}[\pi]$. If $M = \emptyset$ the independence simplifies to (unconditional - some authors say also marginal) independence $\{X_i\}_{i\in K} \perp\!\!\!\perp \{X_i\}_{i\in L}[\pi]$.

Suppose $K$ and $L$ are subsets of $\{1, \ldots, n\}$. Two distributions $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$ are said to be *consistent* if their joint marginals coincide: $\kappa^{\downarrow K\cap L} = \lambda^{\downarrow K\cap L}$. Notice that if $K \cap L = \emptyset$ then $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$ are always consistent.

Having two distributions defined for the same set of variables $\pi(\{X_i\}_{i\in K})$ and $\kappa(\{X_i\}_{i\in K})$, we say that $\kappa$ *dominates* $\pi$ (in symbol $\pi \ll \kappa$) if for all $x \in \mathbb{X}_K$

$$\kappa(x) = 0 \implies \pi(x) = 0.$$

### Combination

Here we adopt (and adapt to the introduced notation) the definition introduced by Dempster in [4].

**Definition 1.** For arbitrary two distributions $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$ their *combination* is for each $x \in \mathbb{X}_{(L\cup K)}$ given by the following formula

$$(\kappa \oplus \lambda)(x) = Const^{-1}\, \kappa(x^{\downarrow K})\, \lambda(x^{\downarrow L}),$$

where *Const* is the normalization constant given by:

$$Const = \sum_{x\in\mathbb{X}_{K\cup L}} \kappa(x^{\downarrow K})\, \lambda(x^{\downarrow L}).$$

In the case where $Const = 0$, we say that distributions $\kappa$ and $\lambda$ are in *total conflict*, and, for this case, their combination is undefined.

### Composition

The following definition was first introduced in [8].

**Definition 2.** For arbitrary two distributions $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$, for which $\kappa^{\downarrow K\cap L} \ll \lambda^{\downarrow K\cap L}$ their *composition* is for each $x \in \mathbb{X}_{(L\cup K)}$ given by the following formula[2]

$$(\kappa \triangleright \lambda)(x) = \frac{\kappa(x^{\downarrow K})\lambda(x^{\downarrow L})}{\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L})}.$$

In case that $\kappa^{\downarrow K\cap L} \not\ll \lambda^{\downarrow K\cap L}$ the composition remains undefined.

The reader certainly noticed that the presented definition slightly extends the notion of composition discussed in Section 1. We do not require that both $K$ and $L$ are proper subsets of $K \cup L$. There are two reasons for this. First, we are going to compare the two operations, and combination was basically defined for the distributions defined for the same variable sets. Second, this generalization makes the formulation of some theoretical properties simpler.

## 3 Properties of Combination and Composition

As already said above, the two operators were designed for different purposes, and so it is not surprising that they possess different properties. Nevertheless, from a

[2] Define $\frac{0\cdot 0}{0} = 0$.

formal point of view, they manifest some similar, or even identical, properties. And, it is the purpose of this section to show what the similarities and dissimilarities between the two operators are.

## *3.1 Common Properties*

**Theorem 1.** *Suppose* $\kappa(\{X_i\}_{i\in K})$ *and* $\lambda(\{X_i\}_{i\in L})$ *are probability distributions.*

1. (Domain of combination): *If* $\kappa\oplus\lambda$ *is defined, then* $\kappa\oplus\lambda$ *is a probability distribution for* $\{X_i\}_{i\in K\cup L}$*, and*

$$\kappa\oplus\lambda=\left(\kappa^{\downarrow K\cap L}\oplus\lambda^{\downarrow K\cap L}\right)\cdot\kappa^{\downarrow K\setminus L|K\cap L}\cdot\lambda^{\downarrow L\setminus K|K\cap L}. \tag{1}$$

2. (Domain of composition): *If* $\kappa\triangleright\lambda$ *is defined, then* $\kappa\triangleright\lambda$ *is a probability distribution for* $\{X_i\}_{i\in K\cup L}$*, and*

$$\kappa\triangleright\lambda=\left(\kappa^{\downarrow K\cap L}\triangleright\lambda^{\downarrow K\cap L}\right)\cdot\kappa^{\downarrow K\setminus L|K\cap L}\cdot\lambda^{\downarrow L\setminus K|K\cap L}. \tag{2}$$

3. (Disjoint domains of arguments): *If* $K\cap L=\emptyset$*, then both* $\kappa\oplus\lambda$ *and* $\kappa\triangleright\lambda$ *are defined and* $\kappa\oplus\lambda=\kappa\triangleright\lambda$.
4. (Simple marginalization): *Let* $(K\cap L)\subseteq M\subseteq K\cup L$*. If* $\kappa\oplus\lambda$ *is defined, then*

$$(\kappa\oplus\lambda)^{\downarrow M}=\kappa^{\downarrow K\cap M}\oplus\lambda^{\downarrow L\cap M}.$$

   *If* $\kappa\triangleright\lambda$ *is defined, then*

$$(\kappa\triangleright\lambda)^{\downarrow M}=\kappa^{\downarrow K\cap M}\triangleright\lambda^{\downarrow L\cap M}.$$

5. (Conditional independence): *Let* $K\setminus L\neq\emptyset\neq L\setminus K$*. If* $\kappa\oplus\lambda$ *is defined, then*

$$\{X_i\}_{i\in K\setminus L}\perp\!\!\!\perp\{X_i\}_{i\in L\setminus K}|\{X_i\}_{i\in K\cap L}[\kappa\oplus\lambda],$$

   *and if* $\kappa\triangleright\lambda$ *is defined, then*

$$\{X_i\}_{i\in K\setminus L}\perp\!\!\!\perp\{X_i\}_{i\in L\setminus K}|\{X_i\}_{i\in K\cap L}[\kappa\triangleright\lambda],$$

*Proof.* **Ad. 1.** Consider probability distributions $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$, such that their combination is well defined. Then, for each $x\in\mathbb{X}_{K\cup L}$:

$$
\begin{aligned}
(\kappa \oplus \lambda)(x) &= \mathit{Const}^{-1}\ \kappa(x^{\downarrow K})\lambda(x^{\downarrow L}) \\
&= \mathit{Const}^{-1}\ \Big(\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})\kappa^{\downarrow K\setminus L|K\cap L}(x^{\downarrow K\setminus L}|x^{\downarrow K\cap L})\Big) \\
&\qquad\qquad \cdot \Big(\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L})\lambda^{\downarrow L\setminus K|K\cap L}(x^{\downarrow L\setminus K}|x^{\downarrow K\cap L})\Big) \\
&= \mathit{Const}^{-1}\ \Big(\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})\cdot\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L})\Big) \\
&\qquad\qquad \cdot\kappa^{\downarrow K\setminus L|K\cap L}(x^{\downarrow K\setminus L}|x^{\downarrow K\cap L})\lambda^{\downarrow L\setminus K|K\cap L}(x^{\downarrow L\setminus K}|x^{\downarrow K\cap L}),
\end{aligned}
$$

which yields Equation (1), because the constant *Const* in the definition of $\kappa\oplus\lambda$ and $\kappa^{\downarrow K\cap L}\oplus\lambda^{\downarrow K\cap L}$ is the same as shown below:

$$
\begin{aligned}
\sum_{x\in\mathbb{X}_{K\cup L}} \kappa(x^{\downarrow K})\lambda(x^{\downarrow L}) &= \sum_{x\in\mathbb{X}_{K\cup L}} \Big(\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})\kappa^{\downarrow K\setminus L|K\cap L}(x^{\downarrow K\setminus L}|x^{\downarrow K\cap L})\Big) \\
&\qquad\qquad \cdot \Big(\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L})\lambda^{\downarrow L\setminus K|K\cap L}(x^{\downarrow L\setminus K}|x^{\downarrow K\cap L})\Big) \\
&= \sum_{x\in\mathbb{X}_{K\cap L}} \kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})\cdot\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L}) \\
&\qquad\qquad \cdot \left(\sum_{y\in\mathbb{X}_{K\setminus L}} \kappa^{\downarrow K\setminus L|K\cap L}(y|x^{\downarrow K\cap L})\right)\left(\sum_{z\in\mathbb{X}_{L\setminus K}} \lambda^{\downarrow L\setminus K|K\cap L}(z|x^{\downarrow K\cap L})\right) \\
&= \sum_{x\in\mathbb{X}_{K\cap L}} \kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})\cdot\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L}).
\end{aligned}
$$

The last equality holds true because both the expressions in parentheses equal 1.

**Ad. 2.** Equation (2) can be proven analogously to the first part of this proof.

**Ad. 3.** The assertion follows immediately from Equations (1) and (2), because $\kappa^{\downarrow\emptyset}\oplus\lambda^{\downarrow\emptyset} = \kappa^{\downarrow\emptyset}\triangleright\lambda^{\downarrow\emptyset} = 1$.

**Ad. 4.** Assume $(K\cap L)\subseteq M\subseteq K\cup L$, and $\kappa\oplus\lambda$ is defined. Then

$$
\begin{aligned}
(\kappa\oplus\lambda)^{\downarrow M} &= \Big(\Big(\kappa^{\downarrow K\cap L}\oplus\lambda^{\downarrow K\cap L}\Big)\cdot\kappa^{\downarrow K\setminus L|K\cap L}\cdot\lambda^{\downarrow L\setminus K|K\cap L}\Big)^{\downarrow M} \\
&= \Big(\kappa^{\downarrow K\cap L}\oplus\lambda^{\downarrow K\cap L}\Big)\cdot\Big(\kappa^{\downarrow K\setminus L|K\cap L}\cdot\lambda^{\downarrow L\setminus K|K\cap L}\Big)^{\downarrow M} \\
&= \Big(\kappa^{\downarrow K\cap L}\oplus\lambda^{\downarrow K\cap L}\Big)\cdot\kappa^{\downarrow(K\cap M)\setminus L|K\cap L}\cdot\lambda^{\downarrow(L\cap M)\setminus K|K\cap L} = \kappa^{\downarrow K\cap M}\oplus\lambda^{\downarrow L\cap M}.
\end{aligned}
$$

The respective assertion for the operator of composition was formulated (and proven) as Lemma 5.10 in [10], however, using Equation 2 it can easily be proven analogously to the preceding part of the proof.

**Ad. 5.** Assume $\kappa\oplus\lambda$ is defined. Then due to already proven Property 4 (Simple marginalization)

$$(\kappa \oplus \lambda)^{\downarrow K} = \kappa \oplus \lambda^{\downarrow K\cap L}$$
$$(\kappa \oplus \lambda)^{\downarrow L} = \kappa^{\downarrow K\cap L} \oplus \lambda$$
$$(\kappa \oplus \lambda)^{\downarrow K\cap L} = \kappa^{\downarrow K\cap L} \oplus \lambda^{\downarrow K\cap L},$$

and therefore one can easily verify the validity of the equality defining the required conditional independence: For each $x \in \mathbb{X}_{K\cup L}$

$$\begin{aligned}(\kappa \oplus \lambda)(x) \cdot (\kappa \oplus \lambda)^{\downarrow K\cap L}(x^{\downarrow K\cap L}) \\ &= Const^{-1}\ \kappa(x^{\downarrow K})\lambda(x^{\downarrow L}) \cdot Const^{-1}\ \kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L}) \\ &= Const^{-1}\ \kappa(x^{\downarrow K})\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L}) \cdot Const^{-1}\ \kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})\lambda(x^{\downarrow L}) \\ &= (\kappa \oplus \lambda)^{\downarrow K} \cdot (\kappa \oplus \lambda)^{\downarrow L},\end{aligned}$$

which proves the required property for the operator of combination (notice that $Const = \sum_{x\in\mathbb{X}_{K\cup L}} \kappa(x^{\downarrow K})\lambda(x^{\downarrow L}) = \sum_{y\in\mathbb{X}_{K\cap L}} \kappa^{\downarrow K\cap L}(y)\lambda^{\downarrow K\cap L}(y)$).

The respective assertion for the operator of composition was formulated (and proven) as Lemma 5.2 in [10], however, it can be proven analogously to the preceding part of the proof. □

## Conditioning

Consider arbitrary two distributions $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$. From formulae (1) and (2) it immediately follows that $\kappa \oplus \lambda = \kappa \triangleright \lambda$ if and only if $\kappa^{\downarrow K\cap L} \oplus \lambda^{\downarrow K\cap L} = \kappa^{\downarrow K\cap L} \triangleright \lambda^{\downarrow K\cap L}$. As expressed in Property 3 of the previous theorem, it holds if $K \cap L = \emptyset$. However it holds also in other situations. In this paragraph we are going to show that this happens also in the case when one of these distributions is a degenerate one-dimensional distribution expressing certainty. Consider variable $X_k$ and its value $a \in \mathbb{X}_k$. The probability distribution $\delta_a(X_k)$ expressing for certain that variable $X_k = a$ is defined for each $x \in \mathbb{X}_k$ as

$$\delta_a(x) = \begin{cases} 1, \text{ if } x = a; \\ 0, \text{ otherwise.} \end{cases}$$

Let us now show that, using the respective definitions, $\delta_a(X_k) \oplus \lambda(\{X_i\}_{i\in L}) = \delta_a(X_k) \triangleright \lambda(\{X_i\}_{i\in L})$. In case that $k \notin L$ this equality holds because of Property 3 of Theorem 1. Therefore, consider the case when $k \in L$. Then for each $x \in \mathbb{X}_L$

$$(\delta_a \oplus \lambda)(x) = Const^{-1}\delta_a(x^{\downarrow\{k\}})\lambda(x) = \begin{cases} Const^{-1}\lambda(x), & \text{if } x^{\downarrow\{k\}} = a; \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where

$$Const = \sum_{x\in\mathbb{X}_L} \delta_a(x^{\downarrow\{k\}})\lambda(x) = \sum_{x\in\mathbb{X}_L : x^{\downarrow\{k\}}=a} \lambda(x) = \lambda^{\downarrow\{k\}}(a). \tag{4}$$

Notice that $\delta_a \oplus \lambda$ is a probability distribution when $\lambda^{\downarrow\{k\}}(a)$ is positive; otherwise $\delta_a$ and $\lambda$ are in total conflict and their combination is not defined.

Using Equation (4) we can rewrite formula (3) into the form

$$(\delta_a \oplus \lambda)(x) = \frac{\delta_a(x^{\downarrow\{k\}})\lambda(x)}{\lambda^{\downarrow\{k\}}(a)} = \begin{cases} \frac{\delta_a(x^{\downarrow\{k\}})\lambda(x)}{\lambda^{\downarrow\{k\}}(x^{\downarrow\{k\}})}, & \text{if } x^{\downarrow\{k\}} = a; \\ 0, & \text{otherwise,} \end{cases}$$

which obviously equal $(\delta_a \triangleright \lambda)(x)$ in case that $\lambda^{\downarrow\{k\}}(a)$ is positive (otherwise $\lambda^{\downarrow\{k\}} \not\gg \delta_a$ and the composition is not defined), because

$$(\delta_a \triangleright \lambda)(x) = \frac{\delta_a(x^{\downarrow\{k\}})\lambda(x)}{\lambda^{\downarrow\{k\}}(x^{\downarrow\{k\}})} = \begin{cases} \frac{\delta_a(x^{\downarrow\{k\}})\lambda(x)}{\lambda^{\downarrow\{k\}}(x^{\downarrow\{k\}})}, & \text{if } x^{\downarrow\{k\}} = a; \\ 0, & \text{otherwise.} \end{cases}$$

So, let us summarize the proven equality along with what was proven about the composition operator in Theorem 2.3 in [2].

**Theorem 2.** *Consider a distribution $\lambda(\{X_i\}_{i\in L})$, variable $X_k \in \{X_i\}_{i\in L}$, its value $a \in \mathbb{X}_k$, and $K \subseteq L \setminus \{k\}$. If $\lambda^{\downarrow\{k\}}(a) > 0$, then the corresponding conditional distribution $\lambda(\{X_i\}_{i\in K}|X_k = a)$ can be computed*

$$\lambda(\{X_i\}_{i\in K}|X_k = a) = (\delta_a(X_k) \oplus \lambda)^{\downarrow K} = (\delta_a(X_k) \triangleright \lambda)^{\downarrow K}.$$

It is worth mentioning that this assertion formally justifies what is often called "Dempster conditioning". Recall that, for example, Dubois and Denœux describe it in [5] as *a special case of Dempster rule of combination,* which, *widely used in evidence theory, can be viewed as a revision process, understood as a prioritized merging of a sure piece of information with an uncertain one.*

## *3.2 Differences*

### Commutativity

From Definition 1 it is obvious that the operator of combination is commutative. On the other hand, it is equally evident that, generally, the operator of composition is not commutative. To show it, it is enough to consider a pair of distributions $\kappa(\{X_i\}_{i\in K})$, $\lambda(\{X_i\}_{i\in L})$, for which $\kappa^{\downarrow K\cap L} \neq \lambda^{\downarrow K\cap L}$. Let us express the respective property precisely in the following assertion, the proof of which can be found in [10].

**Theorem 3.** *For arbitrary two probability distributions $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$, for which either $\lambda^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$, or $\lambda^{\downarrow K\cap L} \gg \kappa^{\downarrow K\cap L}$, it holds that $\kappa$ and $\lambda$ are consistent if and only if $\kappa \triangleright \lambda = \lambda \triangleright \kappa$.*

### Associativity

The associativity of the operator of combination is again obvious from the definition. The corresponding properties of the operator of composition are expressed in the following assertion.

**Theorem 4.** *Consider three probability distributions* $\kappa(\{X_i\}_{i\in K})$, $\lambda(\{X_i\}_{i\in L})$ *and* $\mu(\{X_i\}_{i\in M})$.

1. (Non-associativity): *In general,* $(\kappa \triangleright \lambda) \triangleright \mu \neq \kappa \triangleright (\lambda \triangleright \mu)$.
2. (Associativity under RIP): *Let* $\kappa \triangleright (\lambda \triangleright \mu)$ *be defined. If* $K \supset (L \cap M)$, *or* $L \supset (K \cap M)$, *then* $(\kappa \triangleright \lambda) \triangleright \mu = \kappa \triangleright (\lambda \triangleright \mu)$.
3. (Exchangeability): *If* $K \supset (L \cap M)$, *then* $(\kappa \triangleright \lambda) \triangleright \mu = (\kappa \triangleright \mu) \triangleright \lambda$.

*Proof.* **Ad. 1.** To show non-associativity, it is enough to consider $\kappa(X), \lambda(Y)$, and $\mu(X,Y)$, such that $X \not\perp\!\!\!\perp Y[\mu]$. Then, it is easy to show that $X \perp\!\!\!\perp Y[(\kappa \triangleright \lambda) \triangleright \mu]$ (this is because it follows from the definition that $(\kappa(X) \triangleright \lambda(Y)) \triangleright \mu(X,Y) = \kappa(X) \triangleright \lambda(Y)$), and $X \not\perp\!\!\!\perp Y[\kappa \triangleright (\lambda \triangleright \mu)]$. The latter relation follows from the fact that for the given choice of distributions

$$X \perp\!\!\!\perp Y[\kappa \triangleright (\lambda \triangleright \mu)] \iff X \perp\!\!\!\perp Y[\lambda \triangleright \mu] \iff X \perp\!\!\!\perp Y[\mu].$$

**Ad. 2.** This property was proven in [10] as Theorems 7.2. and 7.3.
**Ad. 3.** This property was proven in [10] as Lemma 5.7. □

Notice that from the commutativity and associativity of the operator of combination it follows that $(\kappa \oplus \lambda) \oplus \mu = (\kappa \oplus \mu) \oplus \lambda$ holds always true. Thus, the exchangeability property holds for the combination operator trivially.

### Idempotency of composition

The following assertion summarizes the basic properties of the operator of composition, neither of which, generally, hold for the operator of combination. The respective proofs can be found in [10, 9].

**Theorem 5.** *Suppose* $\kappa(\{X_i\}_{i\in K})$ *an* $\lambda(\{X_i\}_{i\in L})$ *are probability distributions such that* $\kappa^{\downarrow K\cap L} \ll \lambda^{\downarrow K\cap L}$. *Then the following statements hold true:*

1. (Extension): *If* $M \subseteq K$ *then,* $\kappa^{\downarrow M} \triangleright \kappa = \kappa$.
2. (Composition preserves first marginal): $(\kappa \triangleright \lambda)^{\downarrow K} = \kappa$.
3. (Reduction): *If* $L \subseteq K$ *then,* $\kappa \triangleright \lambda = \kappa$.
4. (Perfectization): $\kappa \triangleright \lambda = \kappa \triangleright (\kappa \triangleright \lambda)^{\downarrow L}$.
5. (Stepwise composition): *If* $(K \cap L) \subseteq M \subseteq L$ *then,* $(\kappa \triangleright \lambda^{\downarrow M}) \triangleright \lambda = \kappa \triangleright \lambda$.

All these properties are, in a way, connected with the fact that the operator of composition is *idempotent*. This fact supports the explanation of the difference between the combination and the composition.

The composition assembles (composes) pieces of knowledge that are supposed to have their origin by decomposition of global knowledge. So it corresponds, for example, to the reconstruction of a picture that was torn into pieces. Having one piece of the picture twice does not help us to reconstruct the picture better than if we have this very piece of picture only once.

In contrast, the combination operator combines pieces of knowledge from *independent* sources. So, in this case, one cannot have the same piece of knowledge twice. Though the two pieces of knowledge can (formally) be expressed in the same way, being from independent sources they are not (they cannot be) identical. Learning from two independent physicians that I am healthy makes me feel better than when I hear this message just from one of them. Hearing the same message repeatedly from independent sources decreases my uncertainty, and this is exactly the property, which is expressed in the following assertion. Using Shannon entropy of a probability distribution $\kappa(X)$

$$H(\kappa) = -\sum_{x\in\mathbb{X}} \kappa(x)\log_2 \kappa(x)$$

as a measure of uncertainty connected with the distribution $\kappa$, the following assertion says that getting the same amount of information from another independent source decreases our uncertainty. This, in a way, corresponds to what is often understood by the Latin proverb "Repetitio est mater studiorum".

**Theorem 6.** *For an arbitrary probability distribution* $\pi$

$$H(\pi\oplus\pi) \le H(\pi). \tag{5}$$

*Proof.* In the proof, we will use an obvious property of Shannon entropy: For two probability distributions $\kappa(X)$ and $\lambda(X)$, such that $\kappa(x) = \lambda(x)$ for all $x \in \mathbb{X}\setminus\{a,b\}$, $\kappa(a) - \lambda(a) = \lambda(b) - \kappa(b) > 0$, and $\kappa(a) \le \kappa(b)$ it holds that

$$H(\kappa) > H(\lambda). \tag{6}$$

This property can be proven by the following simple consideration. Denote $\varepsilon = \kappa(a) - \lambda(a)$, and

$$\begin{aligned} f(\varepsilon) = H(\kappa) - H(\lambda) = &-\kappa(a)\log_2(\kappa(a)) - \kappa(b)\log_2(\kappa(b)) \\ &+(\kappa(a)-\varepsilon)\log_2(\kappa(a)-\varepsilon) + (\kappa(b)+\varepsilon)\log_2(\kappa(b)+\varepsilon). \end{aligned}$$

Since $f(0) = 0$, and

$$f'(\alpha) = \log_2\left(\frac{\kappa(b)+\alpha}{\kappa(a)-\alpha}\right)\cdot(\ln(2))^{-1}$$

is nonnegative for all $\alpha \in [0,\varepsilon]$, it is clear that $f(\varepsilon) > 0$, and therefore strict inequality (6) holds true.

To prove inequality (5) for $\pi(X)$ notice that

$$(\pi \oplus \pi)(x) = (\pi(x))^2 \cdot Const^{-1},$$

where $Const = \sum_{x \in \mathbb{X}} (\pi(x))^2$, and therefore

$$\begin{array}{l} (\pi \oplus \pi)(x) < \pi(x) \ \text{ iff } \ \pi(x) < Const, \\ (\pi \oplus \pi)(x) > \pi(x) \ \text{ iff } \ \pi(x) > Const. \end{array} \tag{7}$$

To finish the proof we will construct a finite sequence of probability distributions, such that $\pi = \pi_0, \pi_1, \pi_2, \ldots, \pi_k = \pi \oplus \pi$, and $H(\pi_i) < H(\pi_{i-1})$ for all $i = 1, 2, \ldots, k$.

Consider $\pi_i$ (starting with $\pi_0$), and denote $\mathbb{A}_i = \{x \in \mathbb{X} : \pi_i(\{x\}) \neq (\pi \oplus \pi)(x)\}$. Let $a$ be the element of $\mathbb{A}_i$, for which the difference between $\pi_i(x)$ and $(\pi \oplus \pi)(x)$ is minimal, i.e.,

$$|\pi_i(a) - (\pi \oplus \pi)(a)| \leq |\pi_i(x) - (\pi \oplus \pi)(x)| \ \ \forall x \in \mathbb{A}_i. \tag{8}$$

Naturally, there must exist $b \in \mathbb{A}_i$ such that

$$\text{sign}(\pi_i(b) - (\pi \oplus \pi)(b)) = -\text{sign}(\pi_i(a) - (\pi \oplus \pi)(a)), \tag{9}$$

and, because of (8), $|\pi_i(b) - (\pi \oplus \pi)(b)| \geq |\pi_i(a) - (\pi \oplus \pi)(a)|$. Therefore we can define distribution $\pi_{i+1}$:

$$\begin{array}{l} \pi_{i+1}(a) = (\pi \oplus \pi)(a), \\ \pi_{i+1}(b) = \pi_i(b) + (\pi_i(a) - (\pi \oplus \pi)(a)), \\ \pi_{i+1}(x) = \pi_i(x) \ \text{ for all } x \in \mathbf{X} \setminus \{a, b\}. \end{array}$$

We immediately see that $|\mathbb{A}_{i+1}| > |\mathbb{A}_i|$, and therefore the sequence $\pi = \pi_0, \pi_1, \pi_2, \ldots, \pi_k = \pi \oplus \pi$ must be finite. We also can see that, because of inequalities (7), each pair $\pi_i$ and $\pi_{i+1}$ meets the assumptions of the property presented at the beginning of this proof. Therefore $H(\pi_{i+1}) < H(\pi_i)$, which completes the proof. □

### Factorization/Decomposition

Property 5 of Theorem 1 says that a relation of conditional independence holds for distributions that are created as a combination or composition of two probability distributions $\kappa(\{X_i\}_{i \in K})$ and $\lambda(\{X_i\}_{i \in L})$. The following assertion expresses the fact that for the operator of composition the assertion may be strengthen. For this operator it can be formulated in a form of equivalence that was proven as Corollary 5.3. in [10].

**Theorem 7.** *For arbitrary probability distribution $\mu(\{X_i\}_{i \in M})$ and $K, L \subset M$ such that $K \setminus L \neq \emptyset \neq L \setminus K$*

$$\{X_i\}_{i \in K \setminus L} \perp\!\!\!\perp \{X_i\}_{i \in L \setminus K} | \{X_i\}_{i \in K \cap L} [\mu]$$

*if and only if $\mu^{\downarrow K \cup L} = \mu^{\downarrow K} \triangleright \mu^{\downarrow L}$.*

Let us conclude this section by saying that combining Property 5 of Theorem 1 with Theorem 7 we get the following assertion, which casts a new viewpoint to the relation between the two studied operators.

**Corollary.** *Assume that $K \setminus L \neq \emptyset \neq L \setminus K$. For any two probability distribution $\kappa(\{X_i\}_{i\in K})$ and $\lambda(\{X_i\}_{i\in L})$, which are not in total conflict, their combination $\kappa \oplus \lambda$ can be expressed in the following way*

$$\kappa \oplus \lambda = (\kappa \oplus \lambda)^{\downarrow K} \triangleright (\kappa \oplus \lambda)^{\downarrow L}.$$

## 4 Summary & Conclusions

In this paper, we studied the properties of the operators of combination and composition. Though these operators were designed to solve different tasks, both of them may be used in the process of knowledge integration, and both of them were defined not only in classical probability theory but also in some alternative uncertainty theories like the possibility and belief functions theories. Since both these operators were introduced in Shenoy's valuation-based system framework[18, 12], it means that they can directly be applied in other uncertainty theories such as Spohn's epistemic belief theory [20], Dempster-Shafer theory [17], and others.

The purpose, for which the two operators were designed, is explained in Section 1. The remainder of the paper studies the formal (algebraic) properties of these operators stressing their common features and differences. To make the presentation as simple as possible, we restricted the exposition to classical probability theory. The readers familiar with alternative uncertainty theories should keep in mind that analogous results can be formulated in all the theories that meet the axioms of Shenoy's valuation-based systems [18, 19], and also in a general possibility theory [21]. It concerns also the important formulae (1) and (2), which help us to reveal surprisingly many similarities of the studied operators, the similarities that may explain why some researchers do not properly distinguish between the combination and composition. From the formal point of view, an interesting result is formulated as a corollary in the last section. It says that if a distribution is a combination of two independent distributions then it may be expressed also as a composition of its marginals.

**Acknowledgements** This work has been supported in part by funds from grant GAČR 15-00215S to the first author, and from the Ronald G. Harper Distinguished Professorship at the University of Kansas to the second author.

## References


1. Bína, V., Jiroušek, R. Marginalization in multidimensional compositional models. *Kybernetika* 42, 4, 405–422, 2006.
2. Bína, V., Jiroušek, R. On computations with causal compositional models. *Kybernetika* 51, 3, 525–539, 2015.
3. Clemen, R. T., Winkler, R. L. Aggregating Probability Distributions. In: Edwards, Ward; Miles JR, Ralph F.; von Winterfeldt, Detlof (eds.). Advances in Decision Analysis: from foundations to applications. 154–176, Cambridge University Press, 2007.
4. Dempster, A. P. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38, 2, 325-339, 1967.
5. Dubois, D. and Denœux, T. Conditioning in Dempster-Shafer theory: prediction vs. revision. In: Belief Functions, 385-392, 2012.
6. Dubois, D., Prade, H. *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* Plenum Press, New York, 1988.
7. Jensen, F. V. *Bayesian Networks and Decision Graphs.* IEEE Computer Society Press, New York, 2001.
8. Jiroušek, R. Composition of probability measures on finite spaces. Uncertainty in Artificial Intelligence: Proceedings of the 13th Conference (UAI-97). D. Geiger and Prakash P. Shenoy (eds.), Morgan Kaufmann, 274–281, 1997.
9. Jiroušek, R. Decomposition of multidimensional distributions represented by perfect sequences. *Annals of Mathematics and Artificial Intelligence* 35, 1-4, 215-226, 2002.
10. Jiroušek, R. Foundations of compositional model theory. *International Journal of General Systems* 40, 6, 623–678, 2011.
11. Jiroušek, R. Local Computations in Dempster-Shafer Theory of Evidence. *Int. J. Approx. Reason.* 53, 8, 1155–1167, 2012.
12. Jiroušek, R., Shenoy, P. P. Compositional Models in Valuation-Based Systems. *Int. J. Approx. Reason.*, 55, 1, 277-293, 2014.
13. Lauritzen, S. L. *Graphical models.* Oxford University Press, 1996.
14. Lauritzen, S. L., Spiegelhalter, D. Local computation with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society series B* 50, 157–224, 1988.
15. Malvestuto, F. M. Equivalence of Compositional Expressions and Independence Relations in Compositional Models. *Kybernetika* 50, 3, 322–362, 2014.
16. Malvestuto, F. M. Marginalization in Models Generated by Compositional Expressions. *Kybernetika* 51, 4, 541–570, 2015.
17. Shafer, G. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, 1976.
18. Shenoy, P. P.: A valuation-based language for expert systems. *Int. J. Approx. Reason.*, 3, 5, 383-411, 1989.
19. Shenoy, P. P., Shafer, G. Axioms for probability and belief-function propagation. Classic Works of the Dempster-Shafer Theory of Belief Functions. Springer Berlin Heidelberg, 499-528, 2008.
20. Spohn, W. A general non-probabilistic theory of inductive reasoning. In: R. D. Shachter and T. S. Levitt and J. F. lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4 (UAI-90)*, North Holland, 274–281, 1990.
21. Vejnarová, J. Composition of possibility measures on finite spaces: preliminary results. Proceedings of 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems IPMU'98. B. Bouchon-Meunier, R.R. Yager (eds.), Editions E.D.K. Paris, 25–30, 1998.